EVALUATION OF DECAF METHODS USING AN EXISTING EIGHT-YEAR FIXED ACOUSTIC MONITORING AND LOCALIZATION DATASET, DEPLOYED DURING E&P ACTIVITIES ALONG THE ARCTIC CONTINENTAL SHELF

A report prepared by Greeneridge Sciences Inc for the Joint Industry Programme (JIP) on E&P Sound and Marine Life

JIP Topic - Mitigation and Monitoring

E&P **SOUND & MARINE LIFE** PROGRAMME

May 2022

About the E&P Sound & Marine Life Programme

The ocean is filled with a wide variety of natural and man-made sounds. Since the early 1990s, there has been increasing environmental and regulatory focus on man-made sounds in the sea and on the effects these sounds may have on marine life. There are now many national and international regimes that regulate how we introduce sound to the marine environment. We believe that effective policies and regulations should be firmly rooted in sound independent science. This allows regulators to make consistent and reasonable regulations while also allowing industries that use or introduce sound to develop effective mitigation strategies.

In 2005, a broad group of international oil and gas companies and the International Association of Geophysical Contractors (IAGC) committed to form a Joint Industry Programme under the auspices of the International Association of Oil and Gas Producers (IOGP) to identify and conduct a research programme that improves understanding of the potential impact of exploration and production sound on marine life. The Objectives of the programme were (and remain):

1. To support planning of E&P operations and risk assessments
2. To provide the basis for appropriate operational measures that are protective of marine life
3. To inform policy and regulation.

The members of the JIP are committed to ensuring that wherever possible the results of the studies it commissions are submitted for scrutiny through publication in peer-reviewed journals. The research papers are drawn from data and information in the contract research report series. Both contract reports and research paper abstracts (and in many cases full papers) are available from the Programme's web site at www.soundandmarinelife.org.

Disclaimer:

This publication is an output from the IOGP Joint Industry Programme on E&P Sound and Marine Life ("the JIP"). Whilst every effort has been made to ensure the accuracy of the information contained in this publication, neither IOGP nor any of participants in the JIP past, present or future, nor the Contractor appointed to prepare this study warrants its accuracy or will, regardless of its or their negligence, assume liability for any foreseeable use made thereof, whether in whole or in part, which liability is hereby excluded. Consequently such use is at the recipient's own risk on the basis that any use by the recipient constitutes agreement to the terms of this disclaimer. The recipient is obliged to inform any subsequent recipient of such terms.

GREENERIDGE SCIENCES, INC.
5142 HOLLISTER AVE, STE 283 • SANTA BARBARA, CA 93111 • T 805-967-7720 • F 805-683-8577

Evaluation of DECAF Methods Using an Existing Eight-Year Fixed Acoustic Monitoring and Localization Dataset, Deployed During E&P Activities Along the Arctic Continental Shelf

Final report by:

Katherine H. Kim, Susanna B. Blackwell
Greeneridge Sciences, Inc.

Aaron M. Thode
Scripps Institution of Oceanography, UCSD

Tiago A. Marques, Danielle Harris, Len Thomas
University of St. Andrews

Prepared for:

Joint Industry Programme on E&P Sound and Marine Life
International Association of Oil & Gas Producers

GSI Report 526-1
May 2022

Table of Contents

1. Introduction 3
Overview 3
Relationship Between Project Objectives and Manuscripts 4
References 9

2. Comparison of Density Estimation (DE) Methods 10
A comparison of three methods for estimating call densities of migrating bowhead whales using passive acoustic monitoring 10

3. Relative Abundance: Lombard Effect 48
Roaring vs. repeating: how bowhead whales adjust their call density and source level (Lombard effect) in the presence of natural and seismic airgun survey noise 48

4. Relative Abundance: Multi-year Trends 69
Estimating relative abundance of bowhead whale activity between two locations and across multiple years 69

5. Absolute Abundance: Cue Rate Estimation 119
Estimating acoustic cue rates in bowhead whales, *Balaena mysticetus*, during their fall migration through the Alaskan Beaufort Sea 119

Appendix: Double-Difference Tracking 135
Double-difference tracking of bowhead whales using autonomous vector sensors in the Beaufort Sea 135

1. Introduction

This document serves as a final report for the International Association of Oil and Gas Producers (IOGP) Sound and Marine Life (SML) Joint Industry Programme (JIP)-sponsored study entitled, "Evaluation of DECAF Methods Using an Existing Eight-Year Fixed Acoustic Monitoring and Localization Dataset, Deployed During E&P Activities Along the Arctic Continental Shelf" (Contract Reference JIP22 III-15-14). The structure of this report consists of the following overview and four additional chapters, with the four latter chapters formatted as peer-reviewed publications. This format was agreed upon with David Hedgeland of BP, supervisor of the contract. Three of the chapters were submitted and accepted for peer-review publication, and one may be submitted for peer review after submission of this report.

This introductory chapter will review the objectives of the original proposal, as well as summarize and discuss how the following chapters relate to those objectives.

Overview

Over a period of eight years, on behalf of the Shell Exploration and Production Company (SEPCO), Greeneridge Sciences, Inc. deployed a large array of acoustic recorders, then collected and analyzed the dataset to study the response of migrating bowhead whales to various E&P (Exploration & Production) activities (Blackwell et al. 2015). These deployments took place along a 280 km swath of the Alaskan Arctic continental shelf, during the whales' annual fall migration from Canadian waters to the Bering Sea. Over 2.4 million bowhead calls were detected and localized on 35 Directional Autonomous Seafloor Acoustic Recorders (DASARs), which have the unique capability of measuring the arrival azimuth of sounds produced by localized sources. Groups of seven DASARs were arranged in triangular grids, or "sites," which allowed whale calls to be triangulated within and surrounding the grid perimeter. Automated detection and localization software for processing this enormous dataset was developed, evaluated, and published in the peer-reviewed literature (Thode et al. 2012). Previous analyses of this dataset have demonstrated that sound production rates of bowhead whales change in response to sound from seismic airgun surveys, but that the rates can either decrease or increase, depending on the received levels of the airgun pulses (Blackwell et al. 2013, 2015). Other industrial activities monitored using this dataset are drilling activities and the associated vessel movements of the drilling logistical support (Blackwell et al. 2017).

The enormous sample sizes of high-quality acoustic localizations in this dataset, combined with independent visual survey estimates of migrating bowhead whale spatial densities across multiple years, have allowed practical evaluations of several key questions about the efficacy and value of applying acoustic density estimation (DE) methods to passive acoustic monitoring (PAM) data. These questions included: (1) whether localization of individual calls is a necessary step in achieving accurate call density estimates, (2) whether long-term trends in population growth or decline in E&P regions can be estimated without knowledge of baseline cue rates (calling rates) of individuals, and (3) whether individual bowhead cue rates could be estimated with reasonable uncertainties, by combining passive acoustic and visual survey datasets.

Relationship Between Project Objectives and Manuscripts

Mirroring these questions, the project had three main objectives. Each objective is listed below in italics, followed by a summary of the manuscript(s) that address that objective.

(1) *Compare call production spatial densities estimated from DASAR array acoustic data using four different density estimation (DE) techniques, checking for self-consistency between the methods' predictions, and whether some degree of localization is an essential ingredient for density estimation.*

Chapter 2 by Oedekoven et. al addressed this objective and resulted in a manuscript entitled "A comparison of three methods for estimating call densities of migrating bowhead whales using passive acoustic monitoring". The University of St. Andrews was the lead institution on this effort, which compared three density estimation methods using manual analyses of multiple deployments across multiple years: *direct acoustic census* (DAC), *empirical point transect* (EPT), and *spatially explicit capture recapture* (SECR). A fourth method, *modeled point transect*, was not evaluated, but the modeled detection functions required for this method were derived in a separate effort (Chapter 3). The DAC and EPT methods require knowledge of the distance of a call from a sensor (i.e., some localization is required), while SECR requires only that the same call be identified correctly across multiple sensors. [Thomas and Marques (2012) and Marques et al. (2013) provide reviews of these DE methods and their application to PAM data.]

This manuscript found that DAC and EPT gave consistent results for density estimation, but SECR results (and the resulting detection function estimate) were substantially and unrealistically different. The study concluded that the way the manual analysis was conducted violated one of the key assumptions required for the SECR method, namely, detections of calls on individual sensors must be statistically independent of each other. In reality, human analysts often used a strong detection on one instrument as a signal to search more closely for weaker detections occurring at a similar time on other instruments. In addition, the automated analysis dataset could not be used with SECR because the technique requires incorporating data where some signals are detected on only one sensor: a "singleton". Unfortunately, the automated processing algorithm produces high false detection rates when detecting calls on single sensors; only when potential call candidates are matched across sensors does the false alarm rate fall to practical levels. The study thus concluded that localization is a necessary feature when conducting density estimation for this dataset, and methods that do not require localization (e.g., SECR) would require more stringent procedures for both manual and automated analyses.

This chapter was submitted and accepted for peer-reviewed publication (Oedekoven et al. 2022).

(2) *Evaluate whether PAM can estimate relative trends in population growth or decline over several years, without the need to estimate absolute animal densities or abundance. The derived trends will be compared with decadal-long population trends observed from multi-year independent visual censuses.*

This objective was split into two sub-objectives. The first sub-objective examined whether relative abundance estimates were consistent when measured between two independent sites

separated by over 100 km. The second sub-objective investigated whether multi-year trends in relative abundance matched visual estimates of animal abundance trends.

Two chapters addressed this objective: "Roaring vs. repeating: how bowhead whales adjust their call density and source level (Lombard effect) in the presence of natural and seismic airgun survey noise" by Thode et al. (2020; Chapter 3) and "Estimating relative abundance of bowhead whale activity between two locations and across multiple years" by Thode et al. (Chapter 4). The Scripps Institution of Oceanography (SIO) was the lead institution on these manuscripts.

The first manuscript examined how bowhead whale vocal activity (call density) and source level vary with diffuse noise levels caused by wind and seismic survey airgun noise. The manuscript found that bowhead whales became more vocally active as ambient noise levels increase and as weak levels of airgun noise first appear. The whales became less vocally active as airgun noise levels increase further. The whales also increased their call source level with increasing levels of ambient noise (i.e., the Lombard effect, also known as "the cocktail party effect"), but they did not change their source level as the cumulative dose of airgun sound increases.

The first manuscript also discussed how to use point transect theory to correct call density measurements for calls masked by noise, and how to adjust point transect theory to handle situations where multiple widely-separated (kilometer-scale baseline) sensors are used to locate a sound. (Conventional point transect theory assumes that the range to a sound can be derived from a single point.)

The second manuscript then took the aforementioned results and attempted to adjust raw call counts to account for behavioral changes caused by diel effects and responses to natural and anthropogenic ambient noise, in order to make a more accurate relative abundance estimate in terms of cumulative call counts over each season. This manuscript found that multi-year relative abundance trends were consistent between both sites, and the correlations improved when behavioral corrections were added. However, the corrections only had a substantial effect on one year (2009), an unusual season during which a large percentage of the season's calls occurred over a very short time period, so the large-scale efficacy of applying behavioral corrections remains inconclusive. The study found that relative call abundance decreased from 2008 through 2011, then climbed steadily between 2011 and 2014, such that the relative call abundance in 2014 was nearly the same as 2008. The interpretation of these results is difficult, as these variations in relative abundance can arise from shifts in the timing or position of the migration, changes in vocal behavior across seasons, or changes in population abundance.

The first manuscript was submitted and accepted for peer-reviewed publication (Thode et al. 2020). We feel the second manuscript would require additional analyses to be a valid peer-reviewed publication.

(3) *Derive absolute abundance (density) estimates by using three different approaches for estimating the calling rates of individual bowhead whales, as a function of environmental and behavioral conditions. Similarities between these three cue rate estimates will be interpreted as evidence that estimating absolute abundance is feasible. The three methods included attempting to track individual whales acoustically ("bottom-up"), dividing the cumulative call counts across sites by the estimated population size ("top-down"), and applying acoustic tags.*

This objective was the most speculative, and only one of the methods, "top-down", turned out to be promising. This method was analyzed in the manuscript "Estimating acoustic cue rates in bowhead whales, *Balaena mysticetus*, during their fall migration through the Alaskan Beaufort Sea" by Blackwell et al. (2021; Chapter 5), which combined visual censuses of the total bowhead whale migration population, aerial surveys of the migration corridor width, and three passive acoustic datasets in order to place bounds on the "average" calling rate of an individual whale. The paper concluded that despite the wide range of uncertainties of the variables involved, estimated cue rates for migrating whales derived across multiple arrays and multiple years were within an order of magnitude of each other. This chapter did not use the behavioral corrections employed in the previous chapters, as the impact of these factors was generally small for all years except 2009. The fall migration median call rate was only 1.3 calls/whale/hour (and interquartile range of 0.5–5.4 calls/whale/hour), a rate much lower than what has been reported from acoustic studies of the spring migration.

This chapter was submitted and accepted for peer-reviewed publication (Blackwell et al. 2021).

The low individual call rate of bowhead whales is the primary reason why the second, "bottom-up", analysis eventually failed. Given the relatively short amount of time a whale spends swimming through a DASAR array (1–2 hours), a typical whale would call perhaps once or twice, and it was very difficult to link a sequence of localizations into a track. Some such sequences were found, but they were unusual and rare. Aaron Thode and his graduate student Ludovic Tenorio analyzed one sequence in detail (Fig. 1), in which 50 calls were measured over 1.7 hours (~29 calls/hour); however, we felt that several individuals swimming in tandem might have generated the track. The precision of the triangulation technique was only about 300 m for these calls, so it was possible that multiple whales could have been calling. Tenorio did some research into whether the tracks could be refined using a "double-difference" technique adopted from seismology (Wilcock 2012; Tenorio-Hallé et al. 2017) and gave two scientific presentations on the topic at Acoustical Society of America (ASA) meetings in 2017 and 2018, where he tested the technique on calibrated playbacks generated by known sources at known locations, and then applied it to the track in question. He found that applying additional relative timing information from the sensors did not improve the tracking results because measuring the relative arrival times between low-frequency calls in a shallow-water environment was not very precise, due to strong attenuation of many call components. Adding relative bearing information did improve the resolution, but not enough where we could rule out the possibility of multiple calling animals. Given that (1) acoustic tracks were the result of rare outliers, and (2) we could not distinguish individuals within a potential group, we decided that the "bottom-up" approach was not worth further effort. A copy of Tenorio's 2018 ASA presentation is included in the Appendix of this report.

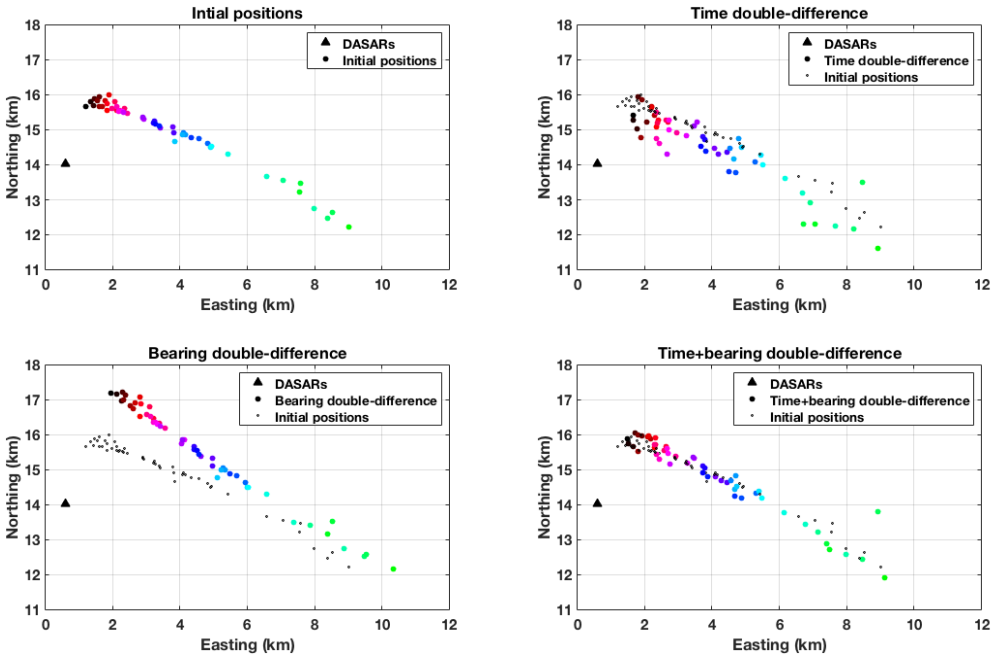


FIGURE 1. Acoustic tracks of an unusually vocal whale or group of whales between 23:12 AKDT (30 Aug 2010) and 00:55 AKDT (31 Aug 2010). Upper left: raw track obtained by DASAR triangulation; upper right: track refined using relative arrival times between sensors and the "double-difference" technique; bottom left: track refined using relative bearing information and "double-difference" technique; bottom right: track refined using both relative timing and bearing information.

The third and final approach to cue rate estimation, acoustic tagging, relied on collaborating with other institutions with the relevant bowhead tagging experience and permits. These two groups had been attempting to apply Acousondes™ to bowhead whales: one in the Beaufort Sea (Lori Quakenbush and colleagues) and another in Disko Bay in West Greenland (Mads Peter Heide-Jørgensen and colleagues). For various reasons out of our control, including an unsuccessful tagging season and a cancelled tagging season, we were not able to obtain such data over the timeline of the project. Bowhead whales had been tagged successfully in separate studies which took place in 2013 and 2016 in Disko Bay, West Greenland. Counting only records that exceeded a few hours in length, a total of six records were obtained, lasting 8–25 h. Calls or other sounds made by the tagged whales were not the focus of the analyses in these studies and were not fully quantified, but it is clear that they were exceedingly rare, on the order of 1 or 2 per record.

In summary, this project identified at least two consistent density estimation methods (DAC and EPT), documented how migrating bowhead whales adjust their vocal activity in response to various environmental factors, established that relative call abundance can be consistently predicted between two independent deployment sites, and placed surprisingly small bounds on the cue rates of migrating bowhead whales by combining multiple independent visual and acoustic datasets. This study's findings regarding DE methods are applicable to other E&P locations, irrespective of species of interest and acoustic propagation characteristics (although such issues affect selection of acoustic recorder specifications and recorder

deployment arrangement, respectively), and suggest that localization is necessary for robust density estimation.

The results of this project have been widely shared among the marine mammal, acoustics, and statistical ecology scientific communities. Numerous presentations have been given at professional conferences: Society for Marine Mammalogy biennial conference (2017); Acoustical Society of America meetings (2017, 2018); National Centre for Statistical Ecology meetings (2017, 2019); Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics workshop (2018); Effects of Sound in the Ocean on Marine Mammals / IOGP E&P Sound & Marine Life Joint Industry Programme meeting (2018); International Statistical Ecology Conference (2018, 2020); Effects of Noise on Aquatic Life conference (2019); World Marine Mammal Conference (2019); among others. Furthermore, four potential peer-reviewed manuscripts have been produced, of which three have been submitted and published (Table 1).

TABLE 1. Publications and manuscripts resulting from the JIP study "Evaluation of DECAF Methods Using an Existing Eight-Year Fixed Acoustic Monitoring and Localization Dataset, Deployed During E&P Activities Along the Arctic Continental Shelf".

Title	Authors	Journal	Status
A comparison of three methods for estimating call densities of migrating bowhead whales using passive acoustic monitoring	Cornelia S. Oedekoven, Tiago A. Marques, Danielle Harris, Len Thomas, Aaron M. Thode, Susanna B. Blackwell, Alexander S. Conrad, Katherine H. Kim	*Environmental and Ecological Statistics (EEST)*	Submitted: September 2019 1st Revision: May 2020 2nd Revision: August 2020 Resubmitted: February 2021 Revised: May 2021 Published Online: June 2021 Issue Date: March 2022
Roaring vs. repeating: how bowhead whales adjust their call density and source level (Lombard effect) in the presence of natural and seismic airgun survey noise	Aaron M. Thode, Susanna B. Blackwell, Alexander S. Conrad, Katherine H. Kim, Tiago Marques, Len Thomas, Cornelia S. Oedekoven, Danielle Harris, Koen Bröker	*Journal of the Acoustical Society of America (JASA)*	Submitted: September 2019 Revised: February 2020 Published: March 2020
Estimating relative abundance of bowhead whale activity between two locations and across multiple years	Aaron M. Thode, Susanna B. Blackwell, Alexander S. Conrad, Katherine H. Kim	*Journal of the Acoustical Society of America (JASA)*	Manuscript in Preparation
Estimating acoustic cue rates in bowhead whales, *Balaena mysticetus*, during their fall migration through the Alaskan Beaufort Sea	Susanna B. Blackwell, Aaron M. Thode, Alexander S. Conrad, Megan C. Ferguson, Catherine L. Berchok, Kathleen M. Stafford, Tiago A. Marques, Katherine H. Kim	*Journal of the Acoustical Society of America (JASA)*	Submitted: November 2020 Revised: April 2021 Published: May 2021

References

Blackwell, S.B., C.S. Nations, T.L. McDonald, C.R. Greene Jr., A.M. Thode, M. Guerra and A.M. Macrander. 2013. Effects of airgun sounds on bowhead whale calling rates in the Alaskan Beaufort Sea. **Marine Mammal Science** 29(4):E342–E365. https://doi.org/10.1111/mms.12001

Blackwell, S.B., C.S. Nations, T.L. McDonald, A.M. Thode, D. Mathias, K.H. Kim, C.R. Greene Jr. and A.M. Macrander. 2015. Effects of airgun sounds on bowhead whale calling rates: evidence for two behavioral thresholds. **PLoS ONE** 10(6): e0125720, 29p. https://doi.org/10.1371/journal.pone.0125720

Blackwell, S.B., C.S. Nations, A.M. Thode, M.E. Kauffman, A.S. Conrad, R.G. Norman and K.H. Kim. 2015. Effects of tones associated with drilling activities on bowhead whale calling rates. **PLoS ONE** 12(11): e0188459, 21p. https://doi.org/10.1371/journal.pone.0188459

Thomas, L., T.A. Marques. 2012. Passive acoustic monitoring for estimating animal density. **Acoustics Today** 8(3):35–44.

Marques, T.A., L. Thomas, S.W. Martin, D.K. Mellinger, J.A. Ward, D.J. Moretti, D. Harris, P.L. Tyack. 2013. Estimating animal population density using passive acoustics. **Biological Reviews** 88:287–309. https://doi.org/10.1111/brv.12001

Tenorio-Hallé, L., A.M. Thode, J. Sarkar, C. Verlinden, J. Tippmann, W.S. Hodgkiss and W.A. Kuperman. 2017. A double-difference method for high-resolution acoustic tracking using a deep-water vertical array. **J. Acoust. Soc. Am.** 142(6):3474–3485.

Thode, A.M., K.H. Kim, S.B. Blackwell, C.R. Greene Jr. and A.M. Macrander. 2012. Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys. **J. Acoust. Soc. Am.** 131(5):3726–3747.

Thode, A.M., S.B. Blackwell, A.S. Conrad, K.H. Kim, T. Marques, L. Thomas, C.S. Oedekoven, D. Harris and K. Bröker. 2020. Roaring and repetition: how bowhead whales adjust their call density and source level (Lombard effect) in the presence of natural and seismic airgun survey noise. **J. Acoust. Soc. Am.** 147(3):2061–2080.

Wilcock, W. 2012. Tracking fin whales in the northeast Pacific Ocean with a seafloor seismic network. **J. Acoust. Soc. Am.** 132(4):2408–2419.

2. Comparison of Density Estimation (DE) Methods

A comparison of three methods for estimating call densities of migrating bowhead whales using passive acoustic monitoring

by

Cornelia S. Oedekoven, Tiago A. Marques, Danielle Harris, Len Thomas, Aaron M. Thode, Susanna B. Blackwell, Alexander S. Conrad, Katherine H. Kim

---

Submitted: *Environmental and Ecological Statistics (EEST),* September 2019
1st Revision: *Environmental and Ecological Statistics (EEST),* May 2020
2nd Revision: *Environmental and Ecological Statistics (EEST),* August 2020
Resubmitted: *Environmental and Ecological Statistics (EEST),* February 2021
Revised: *Environmental and Ecological Statistics (EEST),* May 2021
Published Online: *Environmental and Ecological Statistics (EEST),* June 2021
Issue Date: *Environmental and Ecological Statistics (EEST),* March 2022

Environmental and Ecological Statistics
https://doi.org/10.1007/s10651-021-00506-3

A comparison of three methods for estimating call densities of migrating bowhead whales using passive acoustic monitoring

Cornelia S. Oedekoven, et al. ***[full author details at the end of the article]***

Received: 9 February 2021 / Revised: 12 May 2021 / Accepted: 20 May 2021
© The Author(s) 2021

Abstract
Various methods for estimating animal density from visual data, including distance sampling (DS) and spatially explicit capture-recapture (SECR), have recently been adapted for estimating call density using passive acoustic monitoring (PAM) data, e.g., recordings of animal calls. Here we summarize three methods available for passive acoustic density estimation: plot sampling, DS, and SECR. The rs t two require distances from the sensors to calling animals (which are obtained by triangulating calls matched among sensors), but SECR only requires matching (not localizing) calls among sensors. We compare via simulation what biases can arise when assumptions underlying these methods are violated. We use insights gleaned from the simulation to compare the performance of the methods when applied to a case study: bowhead whale call data collected from arrays of directional acoustic sensors at v e sites in the Beaufort Sea during the fall migration 20072014. Call detections were manually extracted from the recordings by human observers simultaneously scanning spectrograms of recordings from a given site. The large discrepancies between estimates derived using SECR and the other two methods were likely caused primarily by the manual detection procedure leading to non-independent detections among sensors, while errors in estimated distances between detected calls and sensors also contributed to the observed patterns. Our study is among the rs t to provide a direct comparison of the three methods applied to PAM data and highlights the importance that all assumptions of an analysis method need to be met for correct inference.

Keywords Distance sampling N on-independent detections P lot sampling
Spatially explicit capture-recapture

Handling Editor: Luiz Duczmal.

Published online: 15 June 2021

1 Introduction

Passive acoustic monitoring (PAM) is a non-invasive method for monitoring animals in their natural environment that involves recording the sounds that the animals produce (e.g., calls, songs and echolocation clickshe reafter generically referred to as calls). It has proven to be an important tool for monitoring wildlife populations, including both aquatic animals (e.g., shrimp, s h and cetaceans, Lammers and Munger 2016; sharks, Heupel et al. 2004) and terrestrial animals (e.g., birds and amphibians, Acevedo and Villanueva-Rivera 2016; elephants, Wrege et al. 2017; primates, Kalan et al. 2015; frogs, Stevenson et al. 2015). PAM is gaining importance for mitigation management and the protection of endangered species (e.g., Van Parijs et al. 2009; Hildebrand et al. 2015; Brunoldi et al. 2016; Jaramillo-Legorreta et al. 2017).

Using PAM data for monitoring wildlife populations generally involves using acoustic data to estimate either absolute animal density (number of animals per unit area), or some index of relative animal density such as call density (number of calls per unit area per unit time) or call counts (number of calls per unit time detected on a sensor). In general, estimating animal density from PAM requires additional a priori information about the average sound production rate by the individual animals during different behavioral states, which may not be available. For this reason, this paper does not attempt to compare absolute density estimates. We focus instead on methods for estimating relative density from PAM data, specic ally for estimating call density. Call densities have advantages over simple call counts in that they can account for variation in detectability over time or space, avoiding the need to assume that detectability is constant when interpreting any observed pattern. However, additional data and analyses are required to account for detectability.

The main statistical methods for estimating animal density include spatially explicit capture-recapture (SECR, e.g., Borchers and Efford 2008; Dawson and Efford 2009; Marques et al. 2012; Martin et al. 2013; Stevenson et al. 2015), distance sampling (DS, e.g., Buckland et al. 2015) and plot sampling (PS, e.g., Vilchis et al. 2006). Each of these have been adapted for use in estimating call density from PAM data (see reviews by Thomas and Marques 2012; Marques et al. 2013). In this paper, we present a comparison of PS, DS and SECR and examine their relative performance when applied to the same real-world PAM dataset.

Each method requires different assumptions and also demands different capabilities from the PAM system in terms of the ability to localize detected calls. Both PS and DS require explicit distances between sensors and calls, while SECR only requires matching detected calls among acoustic sensors. In some specic cases, it is possible to estimate distances to calls from a single instrument (e.g., using echoes, Tiemann et al. 2006, or modal sound separation, Marques et al. 2011); however, for most PAM data, this typically requires analyzing the relative time-of-arrival of the call among multiple sensors. For moving PAM systems, e.g., a hydrophone array towed by a ship, distances can be obtained by triangulating multiple call positions from a static source using a moving baseline (Barlow and Taylor 2005; Lewis et al. 2017). Here we focus on x ed PAM systems (sensors mounted or moored on the

sea oor), and use a dataset of call detections obtained from recordings made with directional sensors that enabled us to triangulate individual call positions simply and quickly without precise time-of-arrival estimates. All these techniques require some form of cross-sensor matching, where the same call is recognized on different sensors with a relative timing precision, the scale of which depends on the distance between the sensors (Thode et al. 2012).

A comparison between the three density estimation methods has rarely been undertaken with PAM data, as in most cases the data have limited the analyses to a particular method. For example, Phillips (2016) compared DS and SECR estimates of animal density from a combination of PAM and focal follow data against a small population of known size and concluded that they generally produced similar results. In this study, we use a single large PAM dataset (without auxiliary data from a different source) consisting of > 680,000 bowhead whale calls in the Beaufort Sea, collected by Greeneridge Sciences, Inc. (Santa Barbara, California), on behalf of Shell Exploration and Production Company (SEPCO) over an 8-year monitoring period (20072014). The required data for each methoddis tances to the calls for DS and PS and matched calls across sensors for SECRw ere all available from this single dataset. Since the same dataset was used for all three estimation methods, any discrepancies in results must arise from violations of one or more of the assumptions for the respective methods. We also created a simulation tool to examine the robustness of the density estimation methods to various violations of underlying assumptions.

We rs t describe how to estimate call densities from PAM data using PS, DS, and SECR, and the assumptions underlying these methods. We test via simulation what biases may arise when these assumptions are violated. We then analyze the bowhead whale data with each method, compare the resulting call densities and detection functions (where applicable) and discuss the observed discrepancies between the methods using insights gleaned from the simulation. Lastly, we discuss the implications of our ndings in the wider context of density estimation with PAM.

2 Methods for estimating call densities

We focus on calls in this study, although the methods apply to other sounds produced by the animals as well. Marques et al. (2013) described four steps for estimating call densities from PAM data:

1. Identify calls produced by animals of the target population that relate to animal density, i.e., calls that are produced by a known proportion of the population (e.g., adult males) with some regularity following a mean call production rate (given, e.g., as the number of calls produced by an individual per day).
2. Collect a sample of detections of calls using a well-designed survey protocol (e.g., the calls detected in the acoustic recordings in Fig. 1).
3. Estimate the false positive rate i.e., the rate of incorrectly classifying a detected sound as the call of the target species.
4. Estimate the average probability of detecting a call within the search area.

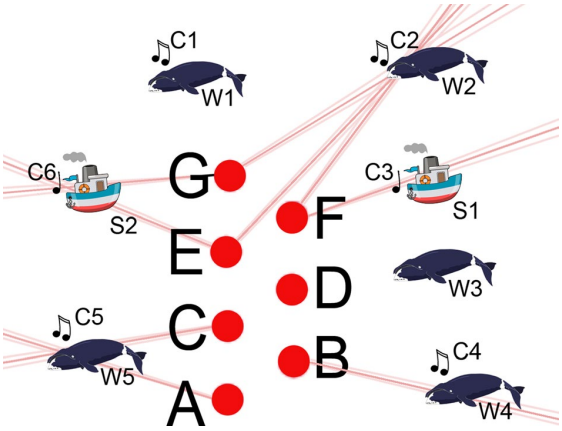


Fig. 1 Example PAM survey design based on the bowhead whale study using a congur ation of seven acoustic sensors moored to the seaoor (red dots *A G*) where each sensor represents a vertex in regular triangles with 7 km edges. Pink lines are azimuths to the call locations with uncertainty (light pink). *W* whale, *S* ship, *C* sounds produced by whales or ships identie d as calls

More than one method is available for each step. While each of these four steps is necessary to estimate call density and relies on the previous steps, this paper focuses on a comparison of different methods for step 4, i.e., estimating the probability of detecting a call. In order to estimate animal density from PAM data, we would need to convert call densities into animal densities in a f th step which requires obtaining a conversion factor (e.g., the mean call production rate per individual, Marques et al. 2013).

To familiarize the reader with the four steps of PAM call density estimation and the bowhead whale dataset used in this study, we present a simple hypothetical example in Fig. 1. We are interested in estimating call density of bowhead whales during their fall migration from Canada into the Bering and Chukchi Seas, hence we use all calls produced by bowhead whales (Mathias et al. 2008) (step 1). For step 2, we moor seven sensors at our study site in shallow waters (approximately 50 m), each capable of measuring the azimuth to the sounds they record. For reasons explained below related to localizing calls, the spacing of the sensors should be chosen that calls produced near one sensor (e.g., sensor A in Fig. 1) have a high probability of being detected at neighboring sensors (B and C). While the sensors are recording, bowhead whales migrate through the area and make calls (e.g., whale W1 produces call C1) or not (e.g., silent whale W3). Some calls are not detected (e.g., C1), while others are detected by one sensor (e.g., C4) or multiple sensors (e.g., C2). Other sounds might also be detected by the sensors and falsely classi ed as whale calls (e.g., C3).

As part of step 2, the recordings are analyzed for acoustic detections using either (i) a manual search protocol where human observers scan the recordings for calls, e.g., visually screening spectrograms of the acoustic data, or (ii) an automated, computer-based, detection and classi cation algorithm. The latter

Table 1 Capture history of detections at sensors AG based on the example from Fig. 1

Object	Call	A	B	C	D	E	F	G	Total
W1	C1	0	0	0	0	0	0	0	0
W2	*C2*	*0*	*0*	*0*	*0*	*1*	*1*	*1*	*3*
S1	*C3*	*0*	*0*	*0*	*0*	*0*	*1*	*0*	*1*
W3		0	0	0	0	0	0	0	0
W4	*C4*	*0*	*1*	*0*	*0*	*0*	*0*	*0*	*1*
W5	*C5*	*1*	*0*	*1*	*0*	*0*	*0*	*0*	*2*
S2	*C6*	*0*	*0*	*0*	*0*	*1*	*0*	*1*	*2*

As we do not detect the objects themselves in a real scenario but only the sounds they produce, we only observe the Call column and the columns to the right of it and row entries that are italic. 1: detected, 0: not detected. Total: number of sensors on which each call (row) is detected

Table 2 Distances (in km) between localized calls and sensors AG , following the example from Fig. 1 and the capture history in Table 1

Object	Call	A	B	C	D	E	F	G	Total
W1	C1								0
W2	*C2*	*ñ*	*ñ*	*ñ*	*ñ*	*21.3*	*15.1*	*17.7*	*3*
S1	*C3*	*ñ*	*ñ*	*ñ*	*ñ*	*ñ*	*ñ*	*ñ*	*0*
W3									0
W4	*C4*	*ñ*	*ñ*	*ñ*	*ñ*	*ñ*	*ñ*	*ñ*	*0*
W5	*C5*	*11.2*	*ñ*	*10.5*	*ñ*	*ñ*	*ñ*	*ñ*	*2*
S2	*C6*	*ñ*	*ñ*	*ñ*	*ñ*	*12.6*	*ñ*	*11.1*	*2*

Total: number of distances that can be used for a DS analysis

generally requires that a false positive rate be estimated (see step 3 above), typically by comparing the automatic detections with detections acquired by a human observer (as in (i)). In the case of large datasets, not all automated detections need to be veri ed to estimate a false positive rate (Marques et al. 2009). A systematic-random sample can be taken instead, where even spacing occurs between samples and a random starting detection is selected, to ensure both a representative and random sample, e.g., every 100th detection starting at the 32nd detection. After automated or manual detection, calls are matched across sensors, leading to a capture history similar to that illustrated in Table 1.

Calls detected on multiple sensors are localized using the callís azimuths from the sensors (Fig. 1). Pomerleau et al. (2011) showed that the mean dive depth of bowhead whales does not exceed 100 m. As the difference between this and the sensor depth is much smaller than the distance that bowhead whales can be detected from (e.g., Thode et al. 2020), we ignore depth and use horizontal space in the following (Barlow and Taylor 2005). For these localized calls, the distance to the detecting sensors can be determined (Table 2). This process naturally results in that only those calls that are easier to detect at greater distances can be

localized. Consider, e.g., a call produced 4 km south of sensor A. Even though this call may be close enough to sensor A to be detected with high probability, in order for it to be localized, it has to be detected by at least one more sensor, e.g., B or C at 9.6 km or 11 km distance to the call, respectively.

In this hypothetical example, the data used for SECR analyses would be the capture histories from Table 1, while the data used for DS analyses would be the distances from Table 2. Although for PS we do not use distances for model tting, we use these distances to limit the analyses to counts of calls within a de ned search radius. Calls only detected by one sensor cannot be localized; they therefore lack distance estimates and are not included in the PS or DS analyses. We refer to these single-detector calls as singletons in the following. SECR is the only method that includes singletons in the analysis.

2.1 Analyses methods and assumptions

Here we summarize the formulas and assumptions for the three density estimation methods in the context of PAM. More complete descriptions of these methods in the context of PAM can be found in Marques et al. (2013) and, in general, for PS in Borchers et al. (2002), for DS in Buckland et al. (2015) and for SECR in Borchers and Efford (2008) and Borchers (2012). Using the notation from the four steps above, i.e., the , and p, the estimator for call density D_c in its most basic form is (Marques et al. 2013):

$$\hat{} \quad \frac{-\hat{}}{\hat{}} \tag{1}$$

where A is the total search area covered by sensors and T is the duration of the recording. While and are the same for PS, DS and SECR, each method dene s D_c, , , A and p differently. Hence, we use subscript notation for these quantities in the following.

The average detection probability p within the search area is generally modelled using two main components: the absolute detection probability at zero distance from the sensor , which is the probability that a call made at zero horizontal distance from the sensor is detected by the sensor, and a detection function that describes the decay in detection probabilities with increasing distance y from the sensor relative to . Depending on the method, either component is assumed or estimated from the data where applicable (see below). A frequently used detection function is the half-normal:

$$\left(-\frac{}{\sigma}\right)\ \sigma \tag{2}$$

Equation (2) contains one parameter, the scale parameter σ, which needs to be estimated. Note that = 1. Larger σ values yield detection functions with

high detection probabilities out to larger distances. In the following we use these components, and , to compare for the three different methods.

2.1.1 Plot Sampling (PS) for PAM data

PS is the simplest of the three estimation methods, but places the most demands on the PAM localization capability. PS limits the search area to the K circles with radius around the sensors, each circle with area , and includes only the calls localized within . Here, is the sum of the number of detections within radius around each sensor, counting any duplicates of a given call caused by overlapping circles twice. The total search area equals .

PS assumes that all calls produced within the individual are detected with certainty. To meet this assumption, the search area is typically limited to a relatively small radius . We can therefore assume that and . Hence, this method does not require estimating a detection function, at the cost of rejecting large numbers of detections that originate outside . As we need to determine which calls originated from within , a successful PS application requires that all calls produced within around any sensor are localized hence, the required sensor spacing described above. Further assumptions are listed in Table 3.

The false positive rate for calls within is dene d as the proportion of all sounds localized within around the sensors that were falsely identi ed as calls of interest. It can be estimated as described above in Sect. 2, limiting the representative sample to calls localized within .

2.1.2 Distance sampling (DS) for PAM data

Here, each sensor represents a point in a point transect survey, which is a form of DS (e.g., Buckland et al. 2001, chapter 5; Buckland 2006). In comparison to PS, we expand the search radius from to a larger radius, and include all call detections within (the circular area around a sensor with radius) from each of the K sensors. Like PS, DS assumes that all calls at (or near) the sensor are detected with certainty, i.e.: However, we no longer assume that all calls within the area around each sensor are detected with certainty. Instead, we t a detection function to the distances between the sensors and the detected calls (e.g., as in Table 2) and use it to estimate the average detection probability within :

$$\frac{}{} \int \qquad (3)$$

An estimate of can be obtained using Eq. (3), replacing with ^ (Buckland et al. 2015). One sees that PS is a limiting case of DS when the search radius is shrunken to values small enough that becomes 1. Similar to PS, refers to the sum of the number of detections that fall within the search areas

Table 3 Summary of assumptions and their importance for PS, DS and SECR in the PAM context (based on Buckland et al. 2001; Buckland 2006; Borchers and Efford 2008; Marques et al. 2013; Thomas and Marques 2012)

#	Assumption	Description	PS	DS	SECR
1	Adequate survey design representative of study area	Placement of an adequate number of sensors according to a (systematic-)random design	Highly important	Highly important	Highly important
2	No un-modelled heterogeneity in detection probabilities	Detection probabilities only depend on distance to the sensor	Not important as long as	Not important if moderate heterogeneity	Highly important
3	$= 1$	Detection probability at/very near the sensor equals 1	Highly important	Highly important	Not required as it is estimated from the data
4	The false positive rate is estimated accurately	$\hat{}$ in Eq. (1) accounts for all false positives in	Highly important	Highly important	Highly important
5	No mis-associations of calls across sensors	Detections of the same call are not falsely identi ed as separate calls	Problematic if this leads to singletons within or to an in ated	Problematic if this leads to singletons on/near any of the sensors or to an in ated	Highly important
6	No lumping of calls	Calls detected on different sensors are not falsely identi ed as the same call	Problematic if this leads to negatively biased	Problematic if this leads to negatively biased	Highly important
7	Independent detection of calls across sensors	The probability of detecting a call on one sensor does not affect the probability of detecting it on other sensors	Not important	Robust to violation	Highly important
8	Accurate distances between sensors and calls	Distances between sensors and calls are measured accurately	Important to obtain accurate	Important. Minor errors can be alleviated by using binned distances for model tting	Does not apply

We note that additional methods not considered here are available to accommodate violations of some of these assumptions for the respective methods

of the sensors, and any call that falls within overlapping search areas is counted towards for each time it was detected by a sensor along with the distance to the respective sensor. While this may seem to artic ially ina te , the reasoning again arises from the requirement that the total search area is , i.e., no subtraction of any overlapping areas occurs.

Multiplication of the search area around a sensor with yields a quantity called the effective area ν , which is the area around a sensor within which as many calls were missed as were detected outside. It can also be expressed in terms of the detection function (Buckland et al. 2015):

$$\nu \quad \pi \int \tag{4}$$

An estimate of the effective area, $\hat{\nu}$ can be obtained using Eq. (4), replacing with $\hat{}$. We can substitute $\hat{\nu}$ for $\hat{}$ in Eq. (1) for estimating call densities.

Another critical assumption for DS is that distances between the sensor and the calls are measured accurately, just like for PS. Uncertainty in localizations and, hence, in the distances, leads to bias in $\hat{}$ and the estimated call densities (e.g., Borchers et al. 2010). The inue nce of minor random distance errors can be alleviated by tting the detection function to binned distances, where the bin width is set to equal the distance error (Buckland et al. 2015). As only localized calls are included in tting the detection function (as opposed to any detected call), the detection function describes the probability of localizing a call with increasing distance from the sensor (as opposed to the probability of detecting a call). It follows that the detection function in the PAM context considered here is a localization function rather than a detection function. Generally, we expect to decrease with increasing distance from the sensor and, although singletons are not localized, an increasing proportion of singletons with increasing distance from the sensor. Further assumptions are listed in Table 3.

The false positive rate for calls within is estimated as the proportion of all sounds localized within around the sensors that were falsely identi ed as calls of interest. It can be estimated as described above in Sect. 2, limiting the representative sample to calls localized within .

2.1.3 Spatially explicit capture-recapture (SECR) for PAM data

For SECR we estimate the probability of detecting a call at distance zero as well as the detection function from the capture histories of the calls (e.g., Borchers and Efford 2008; Borchers 2012). Hence, in comparison to PS or DS, we are not required to assume that all calls at/near the sensor are detected and we do not require call distances or locations. Furthermore, the data are not truncated by a search radius, i.e., . All detected calls, along with their detection histories, are included in the analysis, regardless of the number of sensors they were detected on. Theoretically, with , the total search area and approaches zero. Hence, in practice, we use a different approach where the search area around each sensor

only extends out to a de ned distance beyond which it is safe to assume that no calls can be detected (Efford 2019). Nonetheless, we do not estimate an average detection probability within . Instead, we use estimates of and to obtain an estimate of the effective area ν . As for DS, it is estimated using a combination of the search area and the detection probabilities. However, in contrast to DS, the effective area ν is de ned as the whole area surrounding the K sensors within which as many calls were missed as were detected beyond. It is estimated using the following steps (e.g., Stevenson et al. 2015). First we estimate the probability that a call produced at location (this location is unobserved) is detected by the kth sensor using:

$$( ) \tag{5}$$

where is the distance between and the kth sensor. The probability that the call was detected on at least one sensor becomes:

$$- \prod [-] \tag{6}$$

The effective area is obtained by integrating over . In practice this is done by dividing into grid cells, each with size , where the represent the center points of the grid cells:

$$\nu \quad \sum ( ) \tag{7}$$

The estimate $\hat{\nu}$ obtained using Eq. (7) replaces $\hat{}$ from Eq. (1) for estimating call density with SECR. Also in contrast to PS or DS, refers to the total number of unique calls included in the analyses for SECR and each call contributes to only once, regardless of how many sensors detected it (as opposed to or which refer to the number of detections for PS and DS, respectively).

This method assumes that calls are matched reliably across sensors, detections are made independently between sensors, no un-modelled heterogeneity in detection probabilities exists (i.e., the call detection function depends only on the distance to the sensor, or other appropriate covariates are included in the detection function model, e.g., Singh et al. 2014). Further assumptions are listed in Table 3. The assumption of independent detections between sensors, which emerges as a key factor in this study, means that the detection of a call on one sensor does not in uence the probability of detecting a call on another sensor.

The false positive rate is estimated as the proportion of *all* calls detected on any sensor that were falsely identi ed as calls. In general, we expect the false positive rate to be higher for SECR compared to PS and DS, because the SECR analysis incorporates *all* call detections including singletons, and not just localized calls. In comparison, for PS and DS both the truncation and the inclusion of localized calls only, potentially eliminate a lot of false detections from the analysis.

3 Simulation study

3.1 Methods

We developed a simulation tool to investigate the effect of violating the assumptions from Table 3 on call density estimates from PAM data using PS, DS and SECR. The full description of the tool is given in Appendix 1; here we summarize the key ndings. The simulation results allowed us to understand and diagnose the causes for potential discrepancies in the results between methods in the bowhead whale data.

Each simulation consisted of 1,000 iterations. For each iteration we generated random call detections in a rs t step, using the same sensor con guration as shown in Fig. 1 and a known number = 10,000 calls produced at known locations over a x ed recording time T throughout a dene d study area; hence, call density was known. These calls were detected by each sensor with probability , with known and using a half-normal key function (Eq. (2)) for with known scale parameter σ . Any call detected on multiple sensors was considered as matched correctly between sensors. In a second stage, we analyzed the call detection data using each of the three methods. We rs t tested the methods performed if all of the assumptions from Table 3 were met in a baseline simulation. We then expanded these tests to scenarios where one of the assumptions from Table 3 was violated. To identify potential biases, we used the following diagnostics:

a. Comparisons of estimated call density with true call density ;
b. Comparisons of the estimated with true probabilities of detection (DS and SECR only) using visual tools the detection function plot as shown in Fig. 2 and comparisons of (SECR only) and σ estimated with the respective method vs the true values and σ ;
c. Plotting proportions of calls detected by one, two, three, etc. sensors (de ned here as *proportion plots*) as shown in Fig. 2. Following the hypothetical example, these

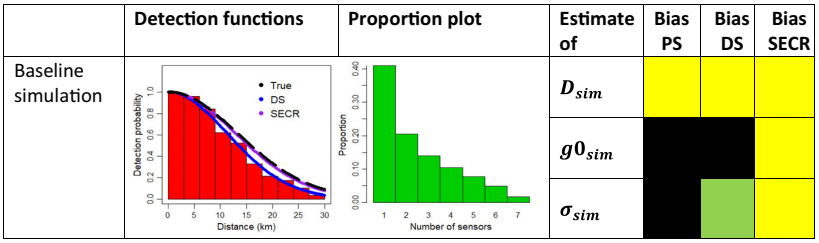


Fig. 2 Results from the baseline simulation. Red, scaled histogram of distances (km) to calls detected by two or more sensors within km, overlain with the true (black line) and estimated DS (blue line) and SECR (purple line) detection functions. Green proportion plot depicting the proportion of calls detected by 1 7 sensors. Color code for median-biases in the estimates as a percentage: none to minor negative or positive bias of < 10%; positive biases: ≥ 10%; ≥ 20%; ≥ 50%; negative biases: ≥ 10%; ≥ 20%; ≥ 50%; NA. Numerical results are given in Appendix 1

Fig. 3 Results from simulation where one assumption from Table 3 was violated. Red: scaled histogram ▶ of distances (km) to calls detected by two or more sensors, overlain with the true (black line) and estimated DS (blue line) and SECR (purple line) detection functions. Green proportion plot depicting the proportion of calls detected by 17 sensors. Color code for median biases in the estimates as a percentage: none to minor negative or positive bias of < 10%; positive biases: ≥ 10%; ≥ 20%; ≥ 50%; negative biases: ≥ 10%; ≥ 20%; ≥ 50%; NA. Numerical results of biases are given in Appendix 1

were produced using the total number of detections for each call (i.e., as presented in the Total column from Table 1).

Note that for both a. and b. we estimated bias with respect to the median of the 1000 estimates, and not the mean, because the non-linear transformations created by the use of the detection function generated long tails in the distribution that disproportionately impacted the mean (McHugh 2003).

3.2 Baseline simulation

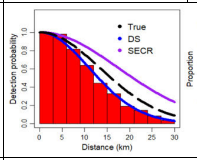
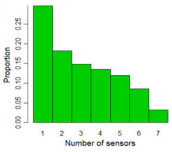
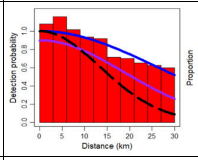
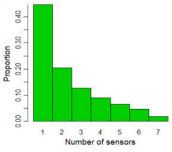
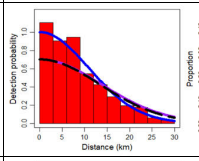
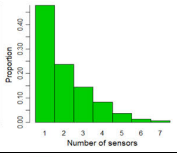
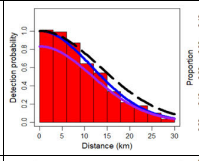
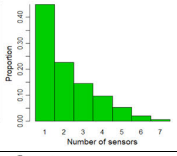
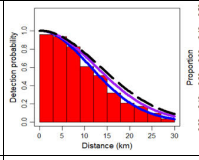
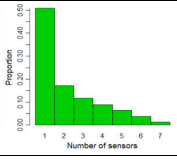
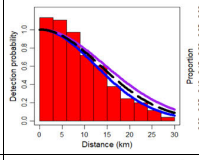
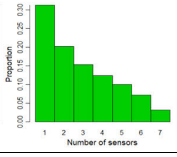
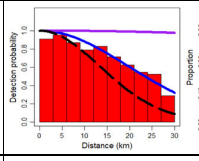
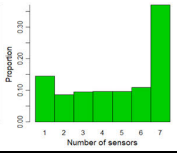
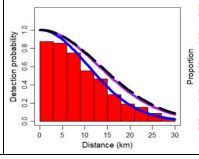
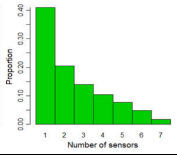
For the baseline simulation, we ensured that all of the assumptions from Table 3 were met. As the acoustic sensors (Fig. 1) were not randomly placed throughout the entire study area (thus violating assumption 1: adequate survey design representative of the entire study area), we randomly distributed calls using a uniform distribution in the simulation in order to preserve the assumption.

Average biases were minor for the call density estimates for each of the three methods, as well as for the parameters pertaining to detection probabilities obtained with SECR (σ and) (Fig. 2). The estimates of the DS scale parameter were negatively biased and the DS detection function declined more quickly with increasing distance than the true or SECR detection functions; however, as the estimated DS detection function actually represented a localization function where each call needed to be detected by two or more sensors, this bias was expected. The fact that tte d better to the histogram of distances than was caused by the increasing number of singletons with increasing distance which were not included in the histogram. Hence, although the DS detection function was negatively biased, the missing singletons meant that, overall, the DS detection function tte d the distances to the measured distances well and estimated call densities only had minor biases (Fig. 2).

The proportion plots consistently revealed the pattern shown in Fig. 2 i.e., the largest proportion of calls (~ 0.42) detected on only one sensor and decreasing proportions towards the maximum possible number of sensors.

3.3 Simulating Violations of Underlying Assumptions

We ran eight simulations of 1,000 iterations each, where in a given simulation one of the eight assumptions listed in Table 3 was violated. Appendix 1 details how these violations were modelled. Almost every violation introduced various biases. Inadequate survey design caused strong bias in call density estimates for each method

#	Violated Assumption	Detection functions	Proportion plots	Est. of	Bias PS	Bias DS	Bias SECR
1	Adequate survey design			D_{sim}			
				$g0_{sim}$			
				σ_{sim}			
2	No un-modelled heteroge-neity in detection probabilities			D_{sim}			
				$g0_{sim}$			
				σ_{sim}			
3	$g0 = 1$			D_{sim}			
				$g0_{sim}$			
				σ_{sim}			
4	The false positive rate is estimated accurately			D_{sim}			
				$g0_{sim}$			
				σ_{sim}			
5	No mis-associations of calls across sensors			D_{sim}			
				$g0_{sim}$			
				σ_{sim}			
6	No lumping of calls			D_{sim}			
				$g0_{sim}$			
				σ_{sim}			
7	Independent detection of calls across sensors			D_{sim}			
				$g0_{sim}$			
				σ_{sim}			
8	Accurately measured distances between sensors and calls			D_{sim}			
				$g0_{sim}$			
				σ_{sim}			

(row 1, Fig. 3). Un-modelled heterogeneity in detection probabilities (row 2) and inaccurate estimates of the false positive rate (row 4) only caused severe biases in call densities estimated with SECR; mis-associations (row 5) and lumping of calls (row 6) also caused larger biases for SECR compared to PS and DS. Setting <1 (row 3) and introducing distance errors (row 8), on the other hand, only caused biases for PS and DS estimates. However, violation of the independence assumption (row 7) lead to the largest discrepancies from the baseline, both in terms of the SECR detection functionw hich was nearly horizontal within the 30 km displayed in Fig. 3a nd the proportion plot. This was the only scenario in which the pattern differed from the decreasing proportions with the increasing number of sensors from the baseline simulation in Fig. 2.

4 Case study

4.1 Data description

Greeneridge Sciences, commissioned by SEPCO, collected acoustic data in the Beaufort Sea during 20072014 to monitor potential effects of oil exploration on bowhead whales. Data were collected using DASARs (Directional Autonomous Seaoor Acoustic Recorders), whose directional capability allowed localization of calls through triangulation (Greene et al. 2004). Each year during the bowhead whale migration westward through the Beaufort Sea (e.g., Harwood et al. 2017), up to a total of 40 DASARs were deployed at v e sites (Fig. 4) in July or August and retrieved in September or October, obtaining continuous acoustic recordings. The geometry of the normal congur ation at each site was seven DASARs arranged in a triangular grid with 7 km spacing between sensors (Figs. 1, 5), although some sites had as few as three and as many as 13 sensors during some years (Appendix 2).

The recordings were analyzed for whale calls using both manual detection by observers and an automated detector (Blackwell et al. 2013, 2015), although the manual analysis was only performed on a subset of the monitoring period. In this paper we restricted the analyses to the manually detected calls, under the assumption

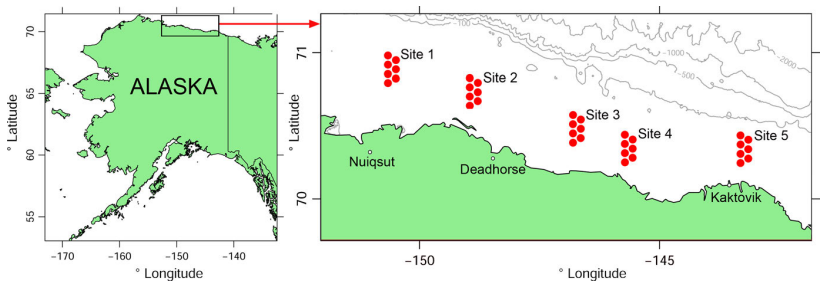


Fig. 4 Study area and DASARs in their normal congur ation at sites 15 shown in red, land shown in green, depth contour lines in grey (100 m, 500 m, 1000 m and 2000 m)

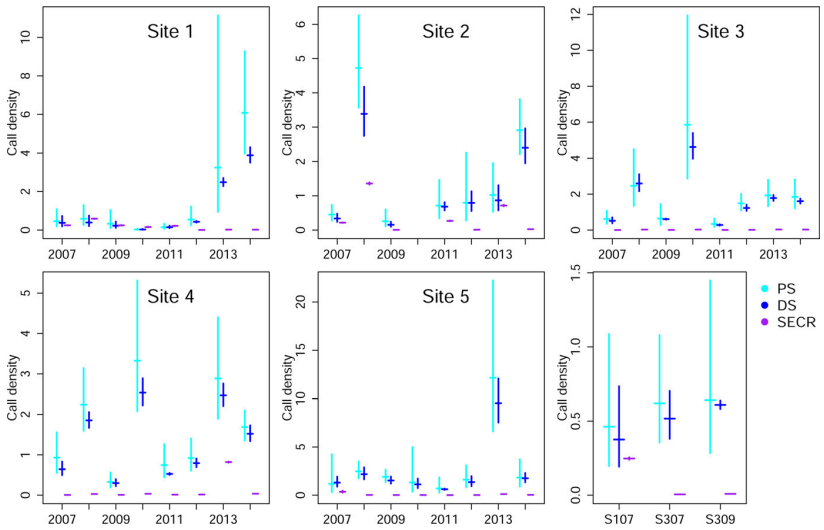


Fig. 5 Bowhead whale call density (number of calls/km^2/day) in 20072014 for each site estimated with PS (light blue), DS (dark blue), and SECR (purple); horizontal lines represent the estimates, vertical lines the 95% CIs. Bottom right plot displays the three focal site-year combinations

that the false positive rate should be nearly zero; hence, we were able to set the false positive rate for each of the methods, , and , to zero. The dataset of automated call detections had much higher fractions of singletons and we suspected that these singletons contained a large proportion of false positives, making them unsuitable for SECR analysis.

Manual detection involved observers visually inspecting 1-min spectrograms from all DASARs at a site simultaneously on a single screen. When a call was detected, the observer examined each spectrogram individually to mark the time and frequency range of the call on each DASAR on which it was visible. These detections generated the call detection histories (similar in format to Table 1) used in the SECR analyses. For a call detected on at least two DASARs, we used the estimated angles between the call and the DASARs to triangulate the location of the source (Thode et al. 2012). As there was some uncertainty in the angles, there was uncertainty about the localization. For those calls that could be localized, we calculated the distances between the call and each of the DASARs that detected the call (similar in format to Table 2).

In 2007 the entire season was inspected manually, whereas in 20082012 and 2014, 59 full days (midnight to midnight) spread throughout the season in the respective year were inspected. The chosen days were judged to be representative of the varying levels of natural and anthropogenic noise each year, as well as the varying numbers of whale calls detected. In 2013 all data from six selected days were inspected for sites 1 and 2, but due to the huge numbers of whale calls detected, a modie d inspection regimen was adopted for sites 3, 4, and 5. Each hour of a

day was divided into four quarters, each 15 min in duration, and only the rst and third quarters were inspected. For sites 3 and 5, this resulted in 50% of each of the six days being inspected. For site 4, the new protocol was applied after the manual inspection for that site had begun, so some hours were inspected at 50%, some at 75%, and some at 100%. In 2010 only two DASARs were deployed at site 2 which could also not be calibrated due to too much ice at the site. Hence, we excluded data from site 2 in 2010 from the analyses.

4.2 Analysis

We estimated call densities using the manually detected bowhead whale call data with the three density estimation methods using functions from the R libraries Distance (Miller 2017; Miller et al. 2019) and secr (Efford 2019). For PS and DS, data were truncated at = 4 km and = 30 km, respectively. Previous work (Blackwell et al. 2015; Thode et al. 2020) concluded that 80% of all whale calls are detected within 3.5 km radius of a sensor, regardless of their source level (how loud they were). We assumed that no calls could be detected from beyond 200 km and, hence, set 200 km. For DS and SECR we tte d one-parameter half-normal detection functions (Eq. (2)) without modelling potential heterogeneity in detection probabilities, i.e., detection probabilities were assumed to depend only on the distance to the sensor, but not on other factors. For DS, distances were binned into ten bins of 3 km each to mediate potential biases due to distance errors. Separate analyses were conducted for each site and each year with the exception of site 2 in 2010, which had insufficient data, yielding 39 different site-year combinations.

For PS, no detection function was tte d; therefore, estimates of uncertainty (95% conde nce intervals (CIs)) represent only variance due to encounter rate. This was estimated using the P3 estimator from Fewster et al. (2009), which is the standard encounter rate variance estimator and the default method of the Distance::ds function for DS point transect analyses (Miller 2017). For DS, the uncertainty from the detection function, estimated using the Distance::ds function, also contributed to the estimate of the uncertainty of call density and both components were combined using the delta method (Buckland et al. 2001, p. 76). Log-normal con dence intervals for call density were produced for PS and DS using methods described in Buckland et al. (2001, pp. 7778) which take into account the small number of samplers (DASARs). For SECR, density in general or call density in the PAM context is one of the model parameters and, hence, asymptotic estimates of uncertainty are based on the inverse of the information matrix from maximizing the unconditional likelihood and are reported as outputs by the secr function of the secr R library (Borchers and Efford 2008).

We used the same assessments for comparing results of the three methods as for the baseline simulation, except that here true values were unknown and estimates could only be compared between the three methods. In the following, we focus on three representative site-year combinations (denoted by, e.g., S107 for site 1 in 2007), but full comparisons for all 39 site-year combinations are included in Appendix 2.

Table 4 Number of DASAR deployments (DASARs) at the sites, recording times (in days), number of calls and number of detections included in the respective analyses across all years or site-year-specic for the three examples

				PS		DS		SECR	
Site	Year	DASARs	Days	Calls	Detections	Calls	Detections	Calls	Detections
15	200714	278	444	171,252	185,010	470,594	2,151,254	686,192	3,091,842
1	2007	5	49	5216	5696	8034	23,376	13,017	29,603
3	2007	7	46	9467	10,076	22,625	97,613	32,473	130,071
3	2009	7	8	1654	1807	6082	30,527	9634	50,244

4.3 Results

The analyzed dataset included a total of 444 recording days and 686,192 calls across all sites and years combined (Table 4). The number of calls and number of detections included in the analyses varied between methods due to different truncation distances. Sample sizes were large even when broken down into site-year specic counts and truncated at 4 km for PS. For any given site-year combination, the computational time for tting models was longest for SECR and shortest for PS due to inherent methods and different sample sizes included in the analyses (Table 4 and Appendix 2). To t models, for example, to data from S107 on an Intel(R) Core(TM) i7 processor with 2.60 GHz CPU and 16.0 GB RAM took 6 s, 22 s and 6 min 47 s for PS, DS and SECR, respectively.

Estimated call densities per site and year were generally similar for PS and DS although slightly higher for PS compared to DSbut typically much lower for SECR (Fig. 5). The latter were on average more than 60 times lower than PS estimates across all site-year combinations and, on average, more than 50 times lower than DS estimates. These discrepancies between SECR estimates and PS or DS estimates were unexpected, as the analyses were based on the same detection data. Even though singletons were included only for SECR, we expected that estimated detection probabilities should be slightly smaller for DS and, hence, correct for the reduced number of detections, yielding similar call density estimates.

Uncertainty in the estimated call densities was generally the largest for PS and lowest for SECR (Fig. 5). Often CIs were too narrow, particularly for SECR, to be visible in Fig. 5 on the scale required for the comparison between the three methods. 95% CIs were wider for PS than for DS due to the larger encounter rate variances for PS. They always overlapped for DS and PS while only in very few cases did they overlap between PS and SECR or DS and SECR (e.g., S107, S108, S213).

The three speci c site-year combinations that we focus on in the following to investigate these discrepancies in call density estimates were sites 1 and 3 in 2007 and site 3 in 2009. This selection included one case (S107) where observed patterns were similar to the baseline simulation in Fig. 2 and two cases (S307 and S309) that showed substantially divergent patterns. For S307 and S309, call density estimates were much smaller for SECR compared to PS and DS (Figs. 4, 5), while for S107 the SECR estimate was lower but within the 95% CIs of PS and DS. The comparison

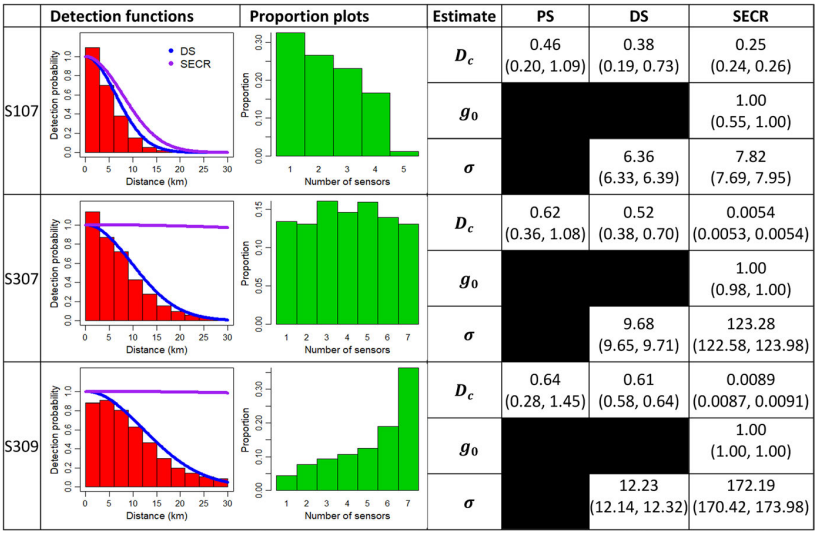


Fig. 6 Results from analyzing the bowhead whale data including estimates and 95% CIs of call density , and parameters and σ estimated with the three methods for three example sites. ■ indicates NA. Red, scaled histogram of distances (km) to calls detected by two or more sensors, overlain with the estimated DS (blue line) and SECR (purple line) detection functions. Green proportion plots depicting the proportion of calls detected by 17 s ensors

of the DS and SECR detection functions for S107 looked as expected, i.e., similar to Fig. 2 where the SECR detection function was slightly wider and $\hat{\sigma}$ slightly larger than $\hat{\sigma}$ (Fig. 6). For S307 and S309, the pattern was different in that the SECR detection functions were unreasonably a tunr easonable because we do not expect to detect bowhead whale calls at 30 km with probability ~ 0.9 (Thode et al. 2020)a nd estimates of σ were extremely large. The proportion plot for S107 (Fig. 6) showed decreasing proportions with increasing sensors, while for the other two site-years this pattern was reversed (S309) and the majority of calls were detected on all DASARs, or proportions were similar across number of DASARs (S307).

4.4 Comparison with simulation study

Results from our case study revealed the following discrepancies between call density estimation methods for most site-year combinations:

1. SECR call density estimates were much lower than PS or DS density estimates;
2. SECR detection functions yielded estimated detection probabilities that were unreasonably large out to large distances;
3. Nearly equal or increasing proportions of calls detected with increasing numbers of DASARs.
4. Slightly higher call density estimates for PS compared to DS.

Of these, we consider the rs t three to be major discrepancies. The simulations conducted determined that the only scenario that recreated the same three major discrepancies was one that modelled non-independent detections across sensors. Only then did the simulation results for SECR show strong negative biases in call density, strong positive biases in the scale parameter estimates and a very wide detection function. This was also the only violation that caused the highest proportion of detected calls to be in the all-sensor category (Fig. 3, row 7), while for all other violations, the highest proportion of detected calls was in the single sensor category (similar to the pattern revealed by the baseline simulation in Fig. 2). The simulation results also revealed that PS estimates were unaffected by non-independence violations and DS estimates were slightly positively biased by it.

The only simulated scenario for which PS estimates were higher than DS estimates, the fourth, minor discrepancy listed above, was when error in the distance measurement was introduced. These results conr med our suspicion that measurement error existed in the case study due to call localization uncertainties.

5 Discussion

5.1 How non-independence of detections affects SECR

We believe the non-independent detections originated from the manual detection process during which observers visually scanned 1-min spectrograms of all DASARs at a given site simultaneously and logged each detected call on each channel. It is likely that a detection made on one DASAR cued the observer into searching for the same call on the other DASARs and, hence, artic ially increased the detection probabilities for this call on the other DASARs. Proportion plots proved to be a key tool for revealing the non-independence issue. As a result, many more calls than expected were detected on more than one DASAR (Fig. 6 and Appendix 2). In 18 cases of the 39 site-year combinations analyzed, the highest proportion of calls

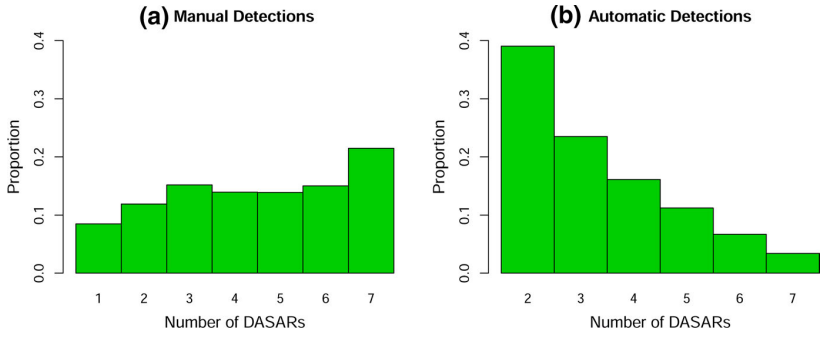


Fig. 7 Proportion plots of the bowhead whale detections for all sites-years with normal DASAR congu - ration combined: **a** for the manually detected calls and **b** for the automatically detected calls (singletons not shown, see text)

were detected on all DASARs (Appendix 2). Only in two manually-analyzed cases was the highest proportion on one DASAR, which is the pattern expected from a fully independent detection process.

Additional evidence that the manual detection process caused the non-independence between detections was provided by the following observation: the proportion plots for the automatically and manually detected data were very different from each other (Fig. 7). In fact, the automatic detections showed the distribution we would expect, i.e., decreasing proportions with the increasing number of DASARs (Thode et al. 2012). Singletons are not shown in Fig. 7 for the automatic detections because they dominated the proportion of detected calls. Because standard formulations of SECR rely on the presence of accurate counts of singletons, applying SECR analysis to the automated results would have required either improving the automated algorithm or developing an SECR estimator that requires calls to be detected by at least two sensors instead of at least one.

We suspect that modic ations to the manual analysis protocol could reduce the dependence between detections on different DASARs. One simple but very labor-intensive option would have observers scanning the spectrograms for each DASAR separately in a rst round, marking the detected calls and, in a second round, matching the marked calls across sensors. Here it would be essential that during the second round, observers would not add any new detections as a result of referring to detections made during the rs t round, which would increase the probability of detection for these new detections in the second round. This second round matching could also be done automatically using a customized algorithm (Thode et al. 2012).

Possible ways for dealing with non-independence in the data for SECR, which could be considered in future studies, include developing a new estimator which accommodates non-independent detections (e.g., Stevenson et al. unpublished data). Non-independence can be alleviated if the process that caused independence can be incorporated in the model. For our case study, this may be as simple as including a covariate in the SECR detection model indicating which sensor the call was detected on rs t by the observer. Here, we would expect that the detection function would drop off relatively quickly for call-sensor combinations detected rst by the observers and be much a tter for the remaining call-sensor combinations. This information was not available for our case study.

5.2 Relationships between PS and DS

Another interesting feature in our case study was that call density estimates were consistently slightly higher for PS compared to DS. For our simulation, this pattern was generally the opposite, i.e., slightly higher density estimates for DS compared to PS (Fig. 3). The only simulated scenario where PS density estimates were higher compared to DS estimates was when errors in the observed distances were introduced. Bochers et al. (2010) showed that random error in distance measurements causes positive biases in DS estimates, more so for point transects compared to line transects. In a sense, the issue is comparable to biases caused by random movement of animals before detection where animals are not detected at their original snapshot

location but at some distance from it. Glennie et al. (2015) showed that for most animal speeds, the movement of animals causes larger biases for strip transects (a type of PS) than for line transects (a type of DS). Hence, we assume that for our case study, the higher estimates for PS compared to DS may have been caused by the uncertainty in localization and the resulting random error in distance measurements. This casts doubt on whether the use of binned distances for the DS analyses was sufficient to mediate any potential issues. Further simulations would be needed to determine the exact amount of bias in call densities for each of the three methods.

5.3 Estimated vs assumed detection probability at or near the DASAR

Lastly, we note that ^ was not considered sufficient proof that the assumptions of or were met as the latter two require certain detection by at least one more DASAR for localization. However, we assumed that a violation of this assumption would yield smaller call density estimates for PS and DS in comparison to SECR. Further research will be conducted to investigate this using a mark-recapture DS (MRDS, Borchers 2012) approach where is estimated for each DASAR individually using the detections of the other DASARs at the same site as trials (Oedekoven et al., unpublished data). For S307, for example, of DASAR A is estimated using the detections made by the other DASARs at S307, i.e., DASARs BG , as trials for A.

6 Conclusion

While passive acoustic density estimation is becoming a widely used alternative to visual methods, our ndings highlight the importance of satisfying key assumptions behind the various methods to avoid substantial bias. In particular, our study has highlighted some fundamental problems in implementing SECR in PAM datasets. First, the strong requirement for independent detections across sensors implies that a rigorous manual inspection protocol needs to be implemented, ensuring that manual reviewers cannot consult multiple data streams simultaneously to enhance the detection of weak calls. Implementing such a protocol would likely slow down the rate of analysis and increase the risk of missing weaker calls.

While most large-scale automatic detection algorithms do satisfy the independence assumption, and would thus seem to be suited to SECR analysis, automated detectors also tend to have a relatively large false detection rate for detections based on one sensor. When detections are compared between multiple sensors, the automated false detection rate tends to drop considerably, which ensures that the distribution of *localized* calls is accurate but produces inaccurate samples of calls *detected* on just a single sensor. Since DS relies on localized calls only, this method is unconcerned with high false detection rates on a single sensor; but for current SECR implementations, this high false detection rate on singletons would be fatal. Practical implementation of SECR on large-scale PAM datasets will therefore either require improvements in manual detection (without incorporating contextual

information from earlier times or other sensors to ensure statistical independence), or require further theoretical development of SECR algorithms that can exclude singletons, thus making automatic detections with high false positive rates among singletons suitable for SECR analysis.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10651-021-00506-3.

Acknowledgements We thank the vessels, their captains and crew for eight seasons of DASAR deployments in the Beaufort Sea: the *R/V Norseman II*, the *R/V Alpha Helix*, and the *R/V Westward Wind.* We also thank Shell Exploration and Production Company for allowing the use of the call data in this study, as well as CR Greene, Jr, and RG Norman for their vision and contributions in the development of the DASAR recorders. The analyses of the data were funded by the Joint Industry Program (JIP contract reference number JIP22 III-15-14). We further thank David Borchers, Murray Efford, Ben Stevenson, Richard Glennie and Douglas Gillespie for fruitful discussions on this topic. TAM thanks partial support by Centro de Estatistica e Aplica e s, Universidade de Lisboa (funded by FCTF unda o para a Ci ncia e a Tecnologia, Portugal, through the project UID/MAT/00006/2013).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the articleís Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the articleís Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Acevedo MA, Villanueva-Rivera JT (2006) Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. Wildl Soc Bull 34(1):211214

Barlow J, Taylor B (2005) Estimates of sperm whale abundance in the northeastern temperate Pacic from a combined acoustic and visual survey. Mar Mamm Sci 21:429445

Blackwell SB, Nations CS, McDonald TL, Greene CR Jr, Thode AM, Guerra M, Macrander AM (2013) Effects of airgun sounds on bowhead whale calling rates in the Alaskan Beaufort Sea. Marine Mammal Sci 29(4):E342E365

Blackwell SB, Nations CS, McDonald TL, Thode AM, Mathias D, Kim KH, Greene CR Jr, Macrander AM (2015) Effects of airgun sounds on bowhead whale calling rates: evidence for two behavioral thresholds. PLoS ONE 10(6):e0125720

Borchers DL, Buckland ST, Zucchini W (2002) Estimating animal abundance: closed populations. Springer, New York

Borchers DL, Efford M (2008) Spatially explicit maximum likelihood methods for capture-recapture studies. Biometrics 64:377438

Borchers DL (2012) A non-technical overview of spatially explicit capturer ecapture models. J Ornithol 152:435444

Bochers DL, Marques TA, Gunnlaugsson T, Jupp P (2010) Estimating distance sampling detection functions when distances are measured with errors. J Agric Biol Environ Stat 15(3):346361

Brunoldi M, Bozzini G, Casale A, Corvisiero P, Grosso D, Magnoli N, Alessi J, Bianchi CN, Mandich A, Morri C, Povero P, Wurtz M, Melchiorre C, Viano G, Cappanera V, Fanciulli G, Bei M, Stasi N, Taiuti M (2016) A permanent automated real-time passive acoustic monitoring system for bottlenose dolphin conservation in the Mediterranean Sea. PLoS ONE 11(1): e0145362

Buckland ST (2006) Point-transect surveys for songbirds: robust methodologies. Auk 123:345357

Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) Introduction to distance samplinge stimating abundance of biological populations. Oxford University Press, Oxford

Buckland ST, Rexstad E, Marques TA, Oedekoven CS (2015) Distance sampling: methods and applications. Methods in statistical ecology. Springer, New York

Dawson DK, Efford MG (2009) Bird population density estimated from acoustic signals. J Appl Ecol 46:12011209

Efford MG (2017) secrlinear: Spatially explicit capture-recapture for linear habitats. R Package Version 1(1):1

Efford MG (2019) secr: Spatially explicit capture-recapture models. R package version 3.2.0

Fewster RM, Buckland ST, Burnham KP, Borchers DL, Jupp PE, Laake JL, Thomas L (2009) Estimating the encounter rate variance in distance sampling. Biometrics 65:225236

Glennie R, Buckland ST, Thomas L (2015) The effect of animal movement on line transect estimates of abundance. PLoS ONE 10(3):e0121333

Greene CR Jr, McLennan MW, Norman RG, McDonald TL, Jakubczak RS, Richardson WJ (2004) Directional frequency and recording (DIFAR) sensors in sea oor recorders to locate calling bowhead whales during their fall migration. J Acoust Soc Am 116(2):799813

Harwood LA, Quakenbush LT, Small RJ, George JC, Pokiak J, Pokiak C, Heide-Jr gensen MP, Lea EV, Brower H (2017) Movements and inferred foraging by bowhead whales in the Canadian Beaufort Sea during August and September, 200612. A rctic 70(2):161176

Heupel MR, Simpfendorfer CA, Hueter RE (2004) Estimation of shark home ranges using passive monitoring techniques. Environ Biol Fishes 71:135142

Hildebrand JA, Baumann-Pickering S, Frasier KE, Trickey JS, Merkens KP, Wiggins SM, McDonald MA, Garrison LP, Harris D, Marques TA, Thomas L (2015) Passive acoustic monitoring of beaked whale densities in the Gulf of Mexico. Scientic Reports, 5: Article number: 16343

Jaramillo-Legorreta A, Cardenas-Hinojosa G, Nieto-Garcia E, Rojas-Bracho L, Ver Hoef J, Moore J, Tregenza N, Barlow J, Gerrodette T, Thomas L, Taylor B (2017) Passive acoustic monitoring of the decline of Mexicoís critically endangered vaquita. Conserv Biol 31:183191

Kalan AK, Mundry R, Wagner OJJ, Heinicke S, Boesch C, Khl HS (2015) Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. Ecol Ind 54:217226

Lammers MO, Munger LM (2016) From shrimp to whales: biological applications of passive acoustic monitoring on a remote Pacic Coral reef. In: Au W, Lammers M (eds) Listening in the ocean. Modern acoustics and signal processing. Springer, New York

Lewis T, Gillespie D, Lacey C, Matthews J, Danbolt M, Leaper R, McLanaghan R, Moscrop A (2007) Sperm whale abundance estimates from acoustic surveys of the Ionian Sea and Straits of Sicily in 2003. J Mar Biol Assoc UK 87:353357

Marques TA, Thomas L, Ward J, DiMarzio N, Tyack PL (2009) Estimating cetacean population density using x ed passive acoustic sensors: an example with beaked whales. J Acoust Soc Am 125:19821994

Marques TA, Munger L, Thomas L, Wiggins S, Hildebrand JA (2011) Estimating North Pacic right whale *Eubalaena japonica* density using passive acoustic cue counting. Endangered Species Research 13:1631 72

Marques TA, Thomas L, Martin SW, Mellinger DK, Jarvis S, Morrissey RP, Ciminello C, DiMarzio N (2012) Spatially explicit capture recapture methods to estimate minke whale abundance from data collected at bottom mounted hydrophones. J Ornithol 152:445455

Marques TA, Thomas L, Martin SW, Mellinger DK, Ward JA, Moretti DJ, Harris D, Tyack PL (2013) Estimating animal population density using passive acoustics. Biol Rev 88:287 309

Martin SW, Marques TA, Thomas L, Morrissey RP, Jarvis S, DiMarzio N, Moretti D, Mellinger DK (2013) Estimating minke whale (*Balaenoptera acutorostrata*) boing sound density using passive acoustic sensors. Mar Mamm Sci 29:142158

Mathias D, Thode A, Blackwell S, Greene C (2008) Computer-aided classic ation of bowhead whale call categories for mitigation monitoring. *New Trends for Environmental Monitoring Using Passive Systems*, Hyeres, French Riviera: 16

McHugh ML (2003) Descriptive statistics, part II: most commonly used descriptive statistics. J Spec Pediatr Nurs 8(3):111116

Miller DL (2017) Distance: distance sampling detection function and abundance estimation. R package version 0.9.7.

Miller DL, Rexstad E, Thomas L, Marshall L, Laake J (2019) Distance sampling in R. J Stat Softw 89(1):128

Phillips GT (2016) Passive acoustics: a multifaceted tool for marine mammal conservation. PhD Thesis. Duke University, Durham, NC

Pomerleau C, Patterson T, Luque S, Lesage V, Ferguson S (2011) Bowhead whale *Balaena mysticetus* diving and movement patterns in the eastern Canadian Arctic: implications for foraging ecology. Endanger Spec Res 15:167177

Singh R, Qureshi Q, Sankar K, Krausman PR, Goyal SP (2014) Evaluating heterogeneity of sex-specic capture probability and precision in camera-trap population estimates of tigers. Wildl Soc Bull 38:791796

Stevenson BC, Borchers DL, Altwegg R, Swift RJ, Gillespie DM, Measey GJ (2015) A general framework for animal density estimation from acoustic detections across a x ed microphone array. Methods Ecol Evol 6:3848

Thode AM, Kim KH, Blackwell SB, Greene CR, Nations CS, McDonald TL, Macrander AM (2012) Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys. J Acoust Soc Am 131(5):37263747

Thode AM, Blackwell SB, Conrad AS, Kim KH, Marques T, Thomas L, Oedekoven CS, Harris D, Brk er K (2020) Roaring and repetition: how bowhead whales adjust their call density and source level (Lombard effect) in the presence of natural and seismic airgun survey noise. J Acoust Soc Am 147:20612080

Thomas L, Marques TA (2012) Passive acoustic monitoring for estimating animal density. Acoust Today 8(3):3544

Van Parijs SM, Clark CW, Sousa-Lima RS, Parks SE, Rankin S, Risch D, Van Opzeeland IC (2009) Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales. Mar Ecol Prog Ser 395:2136

Vilchis LI, Ballance LT, Fiedler PC (2006) Pelagic habitat of seabirds in the eastern tropical Pacic : effects of foraging ecology on habitat selection. Mar Ecol Prog Ser 315:279292

Wrege PH, Rowland ED, Keen S, Shiu Y, Matthiopoulos J (2017) Acoustic monitoring for conservation in tropical forests: examples from forest elephants. Methods Ecol Evol 8:12921301

Cornelia S. Oedekoven is a Senior Research Fellow at the Centre for Research into Ecological and Environmental Modelling at the University of St Andrews, where she develops methods for quantitative wildlife population assessment. She obtained her doctoral degree in Statistics from the University of St Andrews in 2013.

Tiago A. Marques got his PhD in Statistics in 2007 and is currently a Senior Research Fellow working in CREEM at the University of St Andrews, interested in estimating animal abundance with a focus on passive acoustic density estimation. He holds an invited professor position at the Universidade de Lisboa where he teaches statistics to biologists.

Danielle Harris received her PhD in Biology and Statistics from the University of St Andrews in 2012 and is currently a Senior Research Fellow at the Centre for Research into Ecological and Environmental Modelling (University of St Andrews). DHís research interests focus on estimating animal abundance, particularly using passive acoustic recordings and novel technologies such as autonomous vehicles.

Len Thomas is Professor of Statistics at the University of St Andrews where he works on statistical methods for estimating wildlife population size, trends and responses to human-caused stressors. He obtained his PhD from the University of British Colombia in 1997.

Aaron M. Thode received his PhD in Oceanography in 1999 and is currently a Research Scientist at the Scripps Institution of Oceanography, University of California San Diego. His research interests include passive acoustic monitoring and signal processing.

Susanna B. Blackwell has a Ph.D. in Biology from the University of California at Santa Cruz, where she is currently a Research Associate. She has been a Senior Scientist at Greeneridge Sciences, Inc. since 2000, where she works on understanding the effects of anthropogenic sounds on marine mammals in the Arctic.

Environmental and Ecological Statistics

Alexander S. Conrad received an MS in Ocean Engineering at Florida Atlantic University in 2010. He began working for Greeneridge Sciences Inc. as an Ocean Engineer in the summer of 2011 where he collects and analyzes underwater acoustic data to assess human effects on marine life.

Katherine H. Kim received her Ph.D. from the Scripps Institution of Oceanography & Dept. of Electrical Engineering, University of California, San Diego in 2002 and is currently Senior Research Scientist and President of Greeneridge Sciences, Inc., where her research includes passive acoustic monitoring of marine mammals, signal processing of bioacoustic and anthropogenic sounds, and ocean acoustic propagation modeling.

Authors and Affiliations

Cornelia S. Oedekoven[1] ¥Tiago A. Marques[1,2] ¥Danielle Harris[1] ¥ Len Thomas[1] ¥Aaron M. Thode[3] ¥Susanna B. Blackwell[4] ¥Alexander S. Conrad[4] ¥ Katherine H. Kim[4]

✉ Cornelia S. Oedekoven
cso2@st-andrews.ac.uk

[1] Centre for Research Into Ecological and Environmental Modelling, University of St Andrews, The Observatory, Buchanan Gardens, St Andrews KY16 9LZ, Scotland

[2] Departamento de Biologia Animal, Centro de Estats tica E Aplica e s, Faculdade de Ci ncias, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal

[3] Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

[4] Greeneridge Sciences Inc, 5266 Hollister Avenue, Suite 107, Santa Barbara, CA 93111, USA

A comparison of three methods for estimating call densities of migrating bowhead whales using passive acoustic monitoring data

Cornelia S. Oedekoven[1], Tiago A. Marques[1,2], Danielle Harris[1], Len Thomas[1], Aaron M. Thode[3], Susanna B. Blackwell[4], Alexander S. Conrad[4] and Katherine H. Kim[4]

[1] Centre for Research into Ecological and Environmental Modelling / University of St Andrews, The Observatory, Buchanan Gardens, St Andrews, KY16 9LZ, Scotland

[2] Departamento de Biologia Animal, Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal

[3] Scripps Institution of Oceanography / University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093 USA

[4] Greeneridge Sciences Inc., 5266 Hollister Avenue, Suite 107, Santa Barbara, CA 93111 USA

[*]cso2@st-andrews.ac.uk

6.1 Appendix 1: Simulation study

6.1.1 Baseline simulation

To determine potential biases between different methods for estimating densities, we developed a simulation tool that allows comparing estimated call densities between three methods: plot sampling (PS), distance sampling (DS), and spatially-explicit capture-recapture (SECR). Each simulation employed 1,000 iterations, with each iteration consisting of the following steps:

Step 1: Study area, call distribution and sensor placement

The objective of the baseline simulation was to test how the methods perform when all the assumptions from Table 3 were met. Here, we placed calls in a study area with dimension 200km x 200km according to a uniform call distribution (Fig. 8). We used the normal configuration of seven sensors from the bowhead whale study with 7km spacing in two offset north-south lines (Figs. 1, 8). All calls were produced within a single day during which the sensors were recording, i.e. $T = 1$day (see equation (1)).

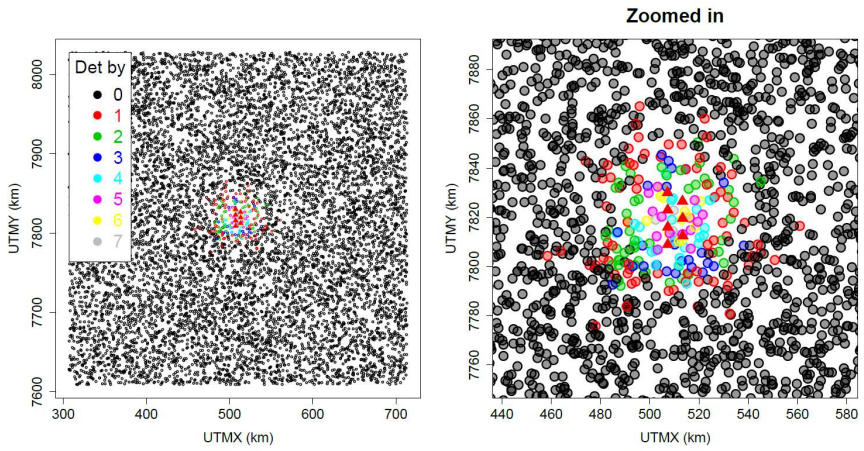


Fig. 8 *Example of the simulated study area with sensors (red triangles) and uniformly distributed calls (circles) color-coded by the number of sensors they were detected by (see legend) for the baseline simulation. Right plot is a zoom of the left plot*

Step 2: Detection of calls, call history and distances between sensors and calls

The sensors detected each of the calls according to probability $g_{0_{sim}} g_{sim}(y)$, where $g_{0_{sim}}$ was set to 1, $g_{sim}(y)$ a half-normal (equation (2)) with scale parameter σ_{sim} = 13.5km for the baseline simulation and y is the distance between the call and the sensor. Detection was determined by drawing a random binomial sample from $Bin(1, g_{0_{sim}} g_{sim}(y))$, where an outcome of 1 meant detected and 0 not detected. This calculation was completed for each call-sensor combination creating the detection histories (similar in format to Table 1) used for the SECR analyses.

Building on the detection histories, we determined which calls were detected on two or more sensors and calculated distances between the call and those sensors that detected the call (similar to Table 2). These localized calls including their detection distances to the sensors were used for the DS and PS analyses.

Step 3: Estimating call density and call abundance

For the baseline PS analysis, we only used localized calls within a radius of 4 km. Call density was estimated using equation (1).

For the distance sampling analysis, we truncated the data at 30km. We used the ddf function from the Distance R package (Miller 2017) to fit a half-normal detection function to the detection distances and to estimate the average detection probability within the search area (equation (5)). The ddf function also returns an estimate of call densities based on the methods described in Section 2.2.2.

For the SECR analysis, we fitted the models with a half-normal detection function using the secr.fit function from the secr R package (Effort 2019). These functions use methods as described in Section 2.2.3.

For each method, call abundance estimates for the study area were obtained by multiplying the estimated call density with the size of the study area and with T. We present the results in terms of call abundance instead of call density as, in this case, it is easier to recognize the size of the bias of the

estimates on the scale of abundance, e.g. comparing a call abundance estimate of 9,669 calls against a true value of 10,000 calls, as opposed to a call density estimate of 0.057 calls per km^2 per day against the true value of 0.059 calls per km^2 per day.

Summarizing the results for 1,000 iterations

We determined biases in the median of the estimated abundances and parameters by comparing these with the true abundances and true parameter values (e.g. using $(median(\hat{N}_{sim}) - N_{sim})/N_{sim}$). Table 5 shows the numerical values that correspond to the biases presented in Fig. 2. For the baseline scenario, estimated abundances were, on average, slightly biased, <5% negatively for SECR and PS and ~6% positively for DS (Table 2). The variability in the estimates was largest for SECR and smallest for DS. The scale parameter estimates also showed minor biases, ~1% positive for SECR and ~11% negative for DS. This negative bias for DS was expected as $g_{DS}(y)$ describes the probability of localizing calls, whereas $g_{SECR}(y)$ describes the detection probability for calls being detected at a single sensor.

Table 5 *Estimated median and interquartile range of estimates of abundance from 1,000 iterations of the baseline simulation*

N_{sim}							
#	**PS Median**	**PS IQR**	**DS Median**	**DS IQR**	**SECR Median**	**SECR IQR**	**True**
0	9,669	8,219 - 11,119	10,625	9,490 – 11,654	9,882	8,400 – 11,308	10,000

Table 6 *Median and interquartile range of estimates of the detection probability parameters from the baseline simulation*

σ_{sim}						$g_{0_{sim}}$		
#	**DS Median**	**DS IQR**	**SECR Median**	**SECR IQR**	**True**	**SECR Median**	**SECR IQR**	**True**
0	12.01	11.62 - 12.44	13.67	12.67 - 14.86	13.5	1.00	0.96 – 1.00	1.00

6.1.2 Deviations from the baseline

In comparison to the baseline simulation described above, we investigated how the methods performed when the assumptions from Table 3 were violated. The methods for implementing these violations are listed in Table 7. We note that the amount of bias caused should only be compared between methods for a given violation and not for a given method between violations.

In Table 7, rows 1-3 pertain to potential environmental conditions encountered in a real scenario, e.g. how animals distribute themselves in their environment, the properties of the calls they produce or how sound travels across the underwater landscape; rows 2 and 3 could also be caused by sensor properties. Rows 4-8 relate to potential errors introduced during the call analysis. Each assumption violation was investigated in an individual simulation (i.e. without combining potential issues) with 1,000 iterations each.

Table 7 *Description of how the assumption violation was implemented in the simulation*

#	Violated assumption	Implementation of violation in the simulation
1	Adequate survey design representative of entire study area	We introduced a gradient of call densities across study area (both a North-South gradient and concentration of calls in a horizontal band near sensors).
2	No un-modelled heterogeneity in detection probabilities	We used σ_{sim} that varied between sensors for generating the call detections ranging between 1.35 and 54.00.
3	$g_0 = 1$	We set $g_{0_{sim}}$ to 0.7 for generating the call detections.
4	The false positive rate is estimated accurately	40% of the total number of detected calls were added as singleton detections allocated randomly to the seven sensors.
5	No mis-associations of calls across sensors	Call detections were incorrectly matched between sensors: for a proportion, 0.2, of calls detected on 2+ sensors, one detection was set to not-detected and a new call created as a singleton for that sensor.
6	No lumping of calls	A proportion, 0.2, of calls detected on 1+ sensors were falsely identified as the same call.
7	Independent detection of calls across sensors	The probability that a call was detected by a sensor depended on whether it was detected by a different sensor: a proportion, 0.9, of detected calls, were logged as detected on all sensors for which the true detection probability was ≥0.1, where the true detection probability was calculated using the detection function defined in 6.1.1, step 2.
8	Accurate distances between sensors and calls	Systematic and random errors in distances between the calls and sensors were introduced using $y_{err} \sim N(1.1 \times y_{true}, y_{true})$ where y_{true} are the true distances between the sensors and the calls and y_{err} are the distances with errors.

Results

The worst biases (>50%) in median call abundance estimates for SECR occurred for the simulations with a gradient in call distribution across the study area (Table 7, row 1), heterogeneity in detection probabilities (row 2), lumping of calls across sensors (row 6) and non-independence (row 7). We note, however, that analyses methods exist for SECR for modelling non-uniform distributions of animals within the study region (e.g. Efford 2017) or heterogeneity in detection probabilities (e.g. Effort 2019). For DS and PS striking biases occurred for the simulation with a gradient in call distribution (row 1), which was impossible to pick up with the survey design as all sensors were placed in the center of the study area. This was less of a problem for SECR. For PS, introducing errors in distances also caused striking biases in call abundance estimates (row 8), much more so than for PS. This was likely due to the fact that random errors in distances have a much larger effect on small circular plots (4km radius for PS vs 30km for DS). This resulted in drastically overestimating densities within the search area and inferring abundances.

The most striking bias for the scale parameter of the detection function existed for SECR in the case of non-independence (Table 7, row 7) where the median of the estimates was almost 10× σ_{sim}. This resulted in a much wider detection function for SECR than the true detection function (Fig. 3, row 7).

For SECR, n_{SECR} is the total number of detected calls (Section 2.2.3), which did not change in the case of non-independence. On the other hand, as a result of the positively biased scale parameter, the effective area estimates were positively biased, thus leading to a median abundance estimate with >97% negative bias. For DS, non-independence still introduced ~57% positive bias in the median of the scale parameter estimates. However, as here n_{DS} is the number of call detections, which increased in the case of non-independence, the positive bias was mostly alleviated. PS was unaffected by non-independence.

Positive bias in σ_{SECR} also existed in three further cases, i.e. when the assumptions of adequate survey design, no un-modelled heterogeneity in detection probabilities or no lumping of calls were violated (Fig. 3, rows 1, 2, 6). This lead to biases in the abundance estimates that were positive in the former two cases and negative only in the latter case. The most indicative diagnostic tool for non-independence was the proportion plot. The proportion plots consistently resembled the equivalent plot from the baseline simulation for all simulations, except for the one with non-independence (compare Fig. 3, row 7 with Fig. 2). Hence, we conclude that non-independence was likely the main issue causing the discrepancies in the results of the case study between PS, DS and SECR.

Table 8 *Median and interquartile range (IQR) of call abundance estimates using PS, DS and SECR from the simulations. Medians with >50% bias highlighted in red. A: reference number to assumptions listed in Tables 3, 7*

	N_{sim}						
A	**PS Median**	**PS IQR**	**DS Median**	**DS IQR**	**SECR Median**	**SECR IQR**	**True**
1	45,927	42,543 – 49,311	48,545	45,820 – 51,227	15,039	890 – 17,071	10,000
2	9,669	8,219 – 11,603	9,731	8,884 – 10,601	44,864	40,705 – 49,018	10,000
3	6,768	5,318 – 8,219	7,313	6,471 – 8,081	9,774	7,486 – 11,991	10,000
4	9,669	8,219 – 11,119	10,509	9,572 – 11,622	20,148	17,918 – 22,325	10,000
5	9,185	7,735 – 10,636	10,182	9,068 – 11,181	11,379	9,654 – 13,080	10,000
6	7,735	6,285 – 9,185	8,077	7,175 – 8,950	4,880	301 – 6,060	10,000
7	9,669	8,219 – 11,603	11,699	10,645 – 12,615	252	236 – 272	10,000
8	46,411	42,543 – 50,762	12,204	11,087 – 13,420	9,882	8,400 – 11,308	10,000

Table 9 *Median and interquartile range (IQR) of estimates of σ_{sim} using DS and SECR and $g0_{sim}$ using SECR from the simulations. Medians with >50% bias highlighted in red. A: reference number to assumptions listed in Tables 3, 7*

	σ_{sim}					$g0_{sim}$		
A	**DS Median**	**DS IQR**	**SECR Median**	**SECR IQR**	**True**	**SECR Median**	**SECR IQR**	**True**
1	10.70	10.53 – 10.90	18.76	17.51 – 87.00	13.5	1.00	0.95 – 1.00	1.00
2	31.83	27.81 – 37.87	18.58	17.72 – 19.59	1.35 – 54.00	0.90	0.88 – 0.91	1.00
3	11.57	11.19 – 12.04	13.66	12.16 – 15.64	13.5	0.70	0.66 – 0.74	0.70
4	12.01	11.62 – 12.44	12.05	11.69 – 12.49	13.5	0.96	0.92 – 1.00	1.00
5	11.80	11.43 – 12.21	13.69	12.62 – 14.74	13.5	0.92	0.88 – 0.96	1.00
6	13.54	12.97 – 14.21	17.58	15.79 – 77.54	13.5	1.00	0.97 – 1.00	1.00
7	21.23	19.94 – 22.88	132.99	124.65 – 139.22	13.5	1.00	1.00 – 1.00	1.00
8	10.06	9.78 – 10.37	13.67	12.67 – 14.86	13.5	1.00	0.96 – 1.00	1.00

A comparison of three methods for estimating call densities of migrating bowhead whales using passive acoustic monitoring data

Cornelia S. Oedekoven[1], Tiago A. Marques[1,2], Danielle Harris[1], Len Thomas[1], Aaron M. Thode[3], Susanna B. Blackwell[4], Alexander S. Conrad[4] and Katherine H. Kim[4]

[1] Centre for Research into Ecological and Environmental Modelling / University of St Andrews, The Observatory, Buchanan Gardens, St Andrews, KY16 9LZ, Scotland

[2] Departamento de Biologia Animal, Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal

[3] Scripps Institution of Oceanography / University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093 USA

[4] Greeneridge Sciences Inc., 5266 Hollister Avenue, Suite 107, Santa Barbara, CA 93111 USA

[*]cso2@st-andrews.ac.uk

6.2 Appendix 2: Case study

6.2.1 Data

The normal configuration as depicted in Fig. 1 was present in 25 of the 39 site-year combinations analyzed; variations of this configuration with 3 – 13 DASARs per site were used in others (Fig. 9).

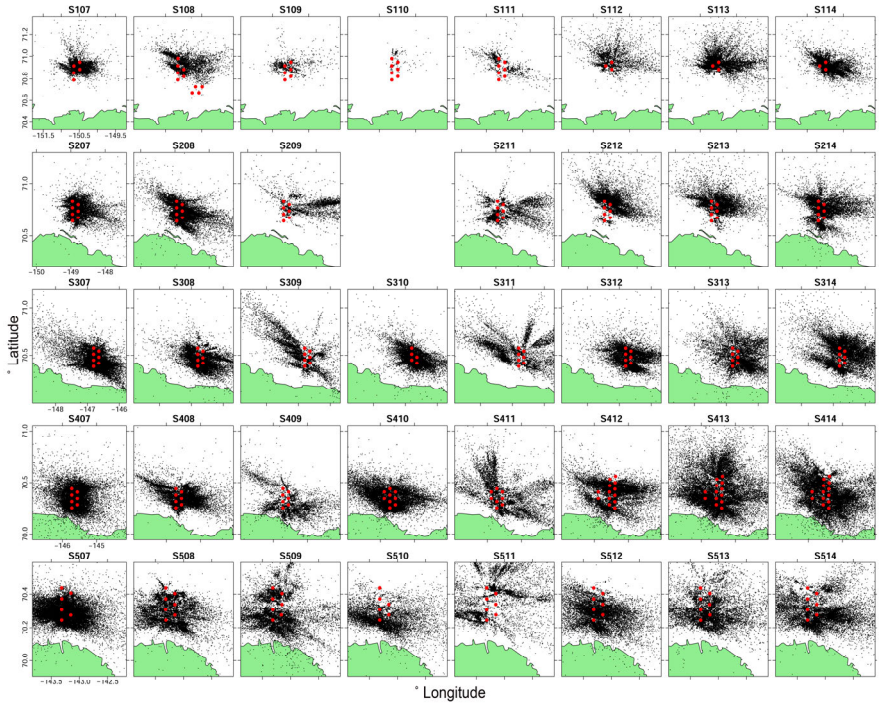


Fig. 9 *Specific configurations of DASARs (red dots) and locations of calls (black dots) in degrees of latitude and longitude detected by two or more DASARs at each site and in each year*

6.2.2 Analyses

We analyzed the bowhead whale call data with the three methods described in the main manuscript, PS, DS and SECR. Analyses were conducted separately for each of the 39 site-year combination. For PS and DS, data were truncated at 4km and 30km, respectively. For DS and SECR we fitted one-parameter half-normal detection functions without modelling potential heterogeneity in detection probabilities (Buckland et al. 2015). For DS, distances were binned into ten bins of 3km each to mediate potential biases due to distance errors – which generally pose the largest problem for point transects at distances near zero.

6.2.3 Results

The analyses included 278 deployments of DASARs, 444 recording days from the individual sites and 686,192 calls (Table 10).

Table 10 *Number of DASAR deployments (DASARs), recording times (in days) for each site-year combination and as a total across all sites and years. Number of calls and number of detections included in the analyses varied between methods due to different truncation distances (4km, 30km and 200km for PS, DS and SECR, respectively)*

				PS		DS		SECR	
Site	**Year**	**DASARs**	**Days**	**Calls**	**Detections**	**Calls**	**Detections**	**Calls**	**Detections**
1	2007	5	49	5,216	5,696	8,034	23,376	13,017	29,603
1	2008	10	6	1,637	1,766	5,241	17,887	9,413	26,652
1	2009	6	8	760	797	1,385	4,260	2,031	5,449
1	2010	7	6	61	65	179	474	645	994
1	2011	7	8	405	413	1,353	5,062	2,745	8,059
1	2012	3	8	628	661	3,067	7,916	5,914	13,478
1	2013	3	6	2,748	2,941	9,805	25,694	17,253	39,594
1	2014	3	5	4,241	4,583	9,852	26,822	13,471	32,540
2	2007	7	49	7,166	7,734	12,244	44,964	18,543	58,831
2	2008	7	6	9,035	9,980	17,263	71,403	24,685	93,329
2	2009	7	8	625	718	1,917	7,461	6,077	23,668
2	2011	7	8	1,858	2,011	5,198	23,988	7,754	33,599
2	2012	7	8	2,162	2,244	8,494	39,601	11,035	50,131
2	2013	7	6	1,976	2,159	7,607	32,392	11,606	44,150
2	2014	7	5	4,636	5,116	11,574	50,617	15,999	63,841
3	2007	7	46	9,467	10,076	22,625	97,613	32,473	130,071
3	2008	7	5	4,481	4,806	15,124	74,821	20,211	94,934
3	2009	7	8	1,654	1,807	6,082	30,527	9,634	50,244
3	2010	5	6	8,234	8,834	17,082	61,515	22,490	74,374
3	2011	7	8	827	960	5,310	26,836	10,940	61,045
3	2012	7	9	4,339	4,718	11,798	57,624	14,850	73,545
3	2013	7	3	1,863	2,039	8,786	46,542	13,483	75,254
3	2014	7	5	3,001	3,260	14,204	70,311	19,527	102,889
4	2007	7	49	14,736	15,972	28,065	110,767	41,511	150,416
4	2008	7	6	4,415	4,732	13,862	71,469	22,998	108,663
4	2009	7	8	846	920	4,656	18,281	8,413	33,255
4	2010	9	6	8,434	9,027	19,505	103,740	27,003	144,675
4	2011	8	8	2,238	2,402	7,413	36,384	13,296	74,188
4	2012	13	9	4,931	5,414	14,194	92,180	20,148	126,636
4	2013	13	4	7,001	7,506	23,863	150,573	32,461	245,587
4	2014	13	5	5,078	5,510	19,077	120,384	25,667	185,237
5	2007	6	29	13,693	14,534	32,704	120,038	44,919	155,537
5	2008	7	6	4,874	5,224	13,265	59,101	17,704	76,272
5	2009	7	8	4,834	5,354	14,105	64,181	18,861	82,483
5	2010	6	6	2,308	2,393	9,096	34,812	13,721	49,921
5	2011	7	8	1,819	1,957	8,168	40,292	13,007	69,004
5	2012	7	8	4,284	4,644	12,831	54,415	15,942	66,025
5	2013	7	3	11,695	12,818	31,786	160,051	46,563	236,151
5	2014	7	5	3,046	3,219	13,780	66,880	20,182	101,518
Total:		**278**	**444**	**171,252**	**185,010**	**470,594**	**2,151,254**	**686,192**	**3,091,842**

Comparison of call density estimates

Estimated call densities per site and year were generally similar for PS and DS – although slightly higher for PS compared to DS – but often much lower for SECR (Fig. 5, Table 11). These discrepancies between SECR estimates and PS or DS estimates were unexpected as the analyses were based on the same detection data (although singletons were included for SECR and excluded for PS and DS). We note that the use of only three sensors at site 1 in 2012-2014 may have likely increased the uncertainty in localizations (as opposed to using seven sensors in the normal configuration) and getting accurate distances (Thode et al. 2012), which may have been part of the reason for the strong discrepancies between estimated call densities for all three methods in 2013 and 2014.

Table 11 *Estimated call density (calls per km^2 per day) and 95% confidence intervals for each site-year combination of the bowhead whale study*

		PS		DS		SECR	
Site	**Year**	**Estimate**	**95%CIs**	**Estimate**	**95%CIs**	**Estimate**	**95%CIs**
1	2007	0.46	0.20 - 1.09	0.38	0.19 - 0.73	0.2478	0.2398 - 0.2562
1	2008	0.59	0.26 - 1.30	0.39	0.20 - 0.75	0.5908	0.5756 - 0.6065
1	2009	0.33	0.10 - 1.04	0.23	0.12 - 0.44	0.2452	0.2281 - 0.2637
1	2010	0.03	0.01 - 0.08	0.03	0.01 - 0.06	0.1584	0.1410 - 0.1779
1	2011	0.15	0.06 - 0.33	0.15	0.10 - 0.23	0.2183	0.2058 - 0.2316
1	2012	0.55	0.24 - 1.23	0.43	0.39 - 0.47	0.0059	0.0057 - 0.006
1	2013	3.25	0.95 - 11.15	2.49	2.28 - 2.71	0.0229	0.0225 - 0.0232
1	2014	6.08	3.98 - 9.29	3.88	3.50 - 4.31	0.0211	0.0207 - 0.0215
2	2007	0.45	0.28 - 0.74	0.34	0.24 - 0.48	0.2163	0.2114 - 0.2214
2	2008	4.73	3.57 - 6.26	3.39	2.74 - 4.18	1.3607	1.3279 - 1.3942
2	2009	0.26	0.11 - 0.60	0.15	0.10 - 0.24	0.0057	0.0056 - 0.0059
2	2011	0.71	0.35 - 1.47	0.68	0.58 - 0.81	0.2675	0.2559 - 0.2796
2	2012	0.80	0.28 - 2.26	0.79	0.56 - 1.13	0.0102	0.0100 - 0.0104
2	2013	1.02	0.53 - 1.96	0.87	0.58 - 1.31	0.7174	0.6950 - 0.7405
2	2014	2.91	2.22 - 3.82	2.40	1.95 - 2.96	0.0244	0.0240 - 0.0248
3	2007	0.62	0.36 - 1.08	0.52	0.38 - 0.70	0.0054	0.0053 - 0.0054
3	2008	2.47	1.35 - 4.51	2.60	2.17 - 3.11	0.0272	0.0268 - 0.0275
3	2009	0.64	0.28 - 1.45	0.61	0.58 - 0.64	0.0089	0.0087 - 0.0091
3	2010	5.86	2.87 - 11.95	4.63	3.97 - 5.39	0.0287	0.0284 - 0.0291
3	2011	0.34	0.18 - 0.65	0.28	0.25 - 0.32	0.0101	0.0099 - 0.0103
3	2012	1.49	1.10 - 2.02	1.23	1.07 - 1.41	0.0122	0.0120 - 0.0124
3	2013	1.93	1.33 - 2.80	1.78	1.63 - 1.95	0.0331	0.0326 - 0.0337
3	2014	1.85	1.21 - 2.83	1.61	1.48 - 1.75	0.0288	0.0284 - 0.0292
4	2007	0.93	0.56 - 1.56	0.64	0.50 - 0.83	0.0066	0.0065 - 0.0067
4	2008	2.24	1.60 - 3.14	1.85	1.67 - 2.06	0.0285	0.0281 - 0.0288
4	2009	0.33	0.19 - 0.56	0.30	0.23 - 0.40	0.0080	0.0078 - 0.0081
4	2010	3.33	2.08 - 5.31	2.53	2.22 - 2.89	0.0334	0.0330 - 0.0338
4	2011	0.75	0.44 - 1.27	0.53	0.50 - 0.55	0.0122	0.0120 - 0.0124
4	2012	0.92	0.60 - 1.40	0.79	0.69 - 0.91	0.0160	0.0158 - 0.0162
4	2013	2.89	1.89 - 4.40	2.47	2.21 - 2.76	0.8195	0.8062 - 0.8329
4	2014	1.69	1.35 - 2.10	1.52	1.34 - 1.73	0.0361	0.0357 - 0.0365
5	2007	1.67	0.74 - 3.76	1.30	0.88 - 1.92	0.0119	0.0117 - 0.0120
5	2008	2.47	1.75 - 3.49	2.18	1.64 - 2.89	0.0220	0.0217 - 0.0224
5	2009	1.90	1.36 - 2.65	1.52	1.19 - 1.93	0.0176	0.0174 - 0.0179
5	2010	1.32	0.35 - 4.99	1.12	0.73 - 1.71	0.0174	0.0171 - 0.0177
5	2011	0.70	0.26 - 1.84	0.61	0.55 - 0.68	0.0120	0.0118 - 0.0122
5	2012	1.61	0.84 - 3.09	1.36	0.93 - 1.97	0.0146	0.0143 - 0.0148
5	2013	12.14	6.63 - 22.24	9.54	7.53 - 12.08	0.1146	0.1136 - 0.1156
5	2014	1.83	0.90 - 3.72	1.75	1.33 - 2.30	0.0300	0.0294 - 0.0303

Comparison of detection functions

Visual inspection of the detection functions showed that, despite a peak in distances near zero, the fit was generally good for the DS models but poor in most cases for SECR (Fig. 10). However, in several cases the detection functions seemed to underfit near distance zero (e.g. for S109). These spikes in detection distances near zero may have been caused by localization uncertainty and resulting random distance error. They were also often the cause for that chi-square goodness-of-fit test statistics for the DS models tended to be large and tests significant (Table 7), indicating a poor fit. For the SECR models, detection probabilities seemed to decay unreasonably slowly with increasing distance from the sensor. Note, that when DS and SECR detection functions were similar (e.g. site 1 2007:2011 or site 2 year 2007 and 2009), call density estimates were similar for DS and SECR as well (Table 11, Fig. 5).

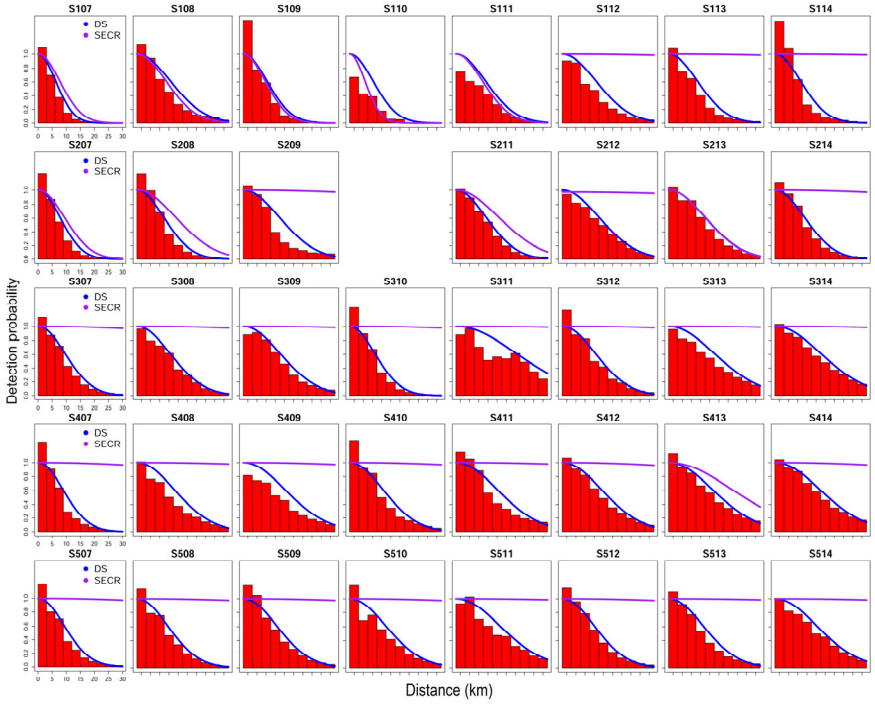


Fig. 10 *Scaled histograms of distances (km) detected by two or more DASARs per site and year with DS (blue line) and SECR (purple line) half-normal detection functions*

Table 12 *Chi-square test statistics rounded to the nearest integer for half-normal detection functions fitted with DS. All tests were not significant at the 0.05 level*

Site	2007	2008	2009	2010	2011	2012	2013	2014
1	8360	1192	803	129	107	253	1972	5501
2	8688	161	1083	241	865	29	33	2396
3	4199	371	619	6726	703	1337	157	165
4	12,214	1020	202	5270	1803	676	1145	625
5	7060	987	1391	468	421	1349	4881	159

Comparison of proportion plots

The proportion plots showed strong variability among the different site-year combinations (Fig. 11). If all our assumptions were met, we would expect these to be similar to the proportion plot shown in Fig. 2, i.e. the highest proportion of singletons and decreasing proportions with increasing number of DASARs. However, in 18 of the 39 site-year combinations, this pattern was reversed and highest proportions were in the all-DASAR category.

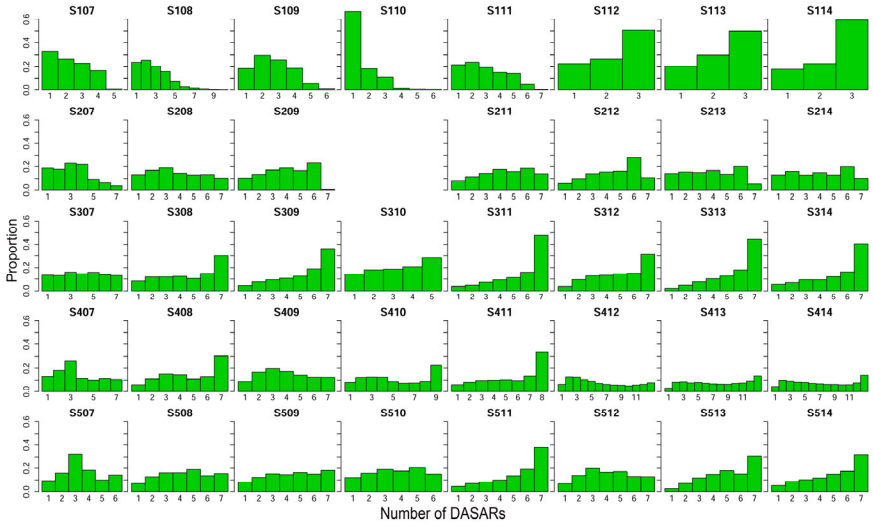


Fig. 11 *Proportion plots for each site-year combination included in the case study: proportion of calls detected by the respective number of DASARs*

We further investigated if at least some of this variability in the proportion plots form Fig. 11 may have been caused by the variation in DASAR configurations. To this end, we conducted 39 baseline simulations, one for each site-year combination, as in Section 6.1.1 (Appendix 1) but with the specific DASAR configuration of the site-year combination. The decreasing proportions similar to that in Fig. 2 was the consistent pattern in the simulations across all site-years regardless of DASAR configuration (Fig. 12).

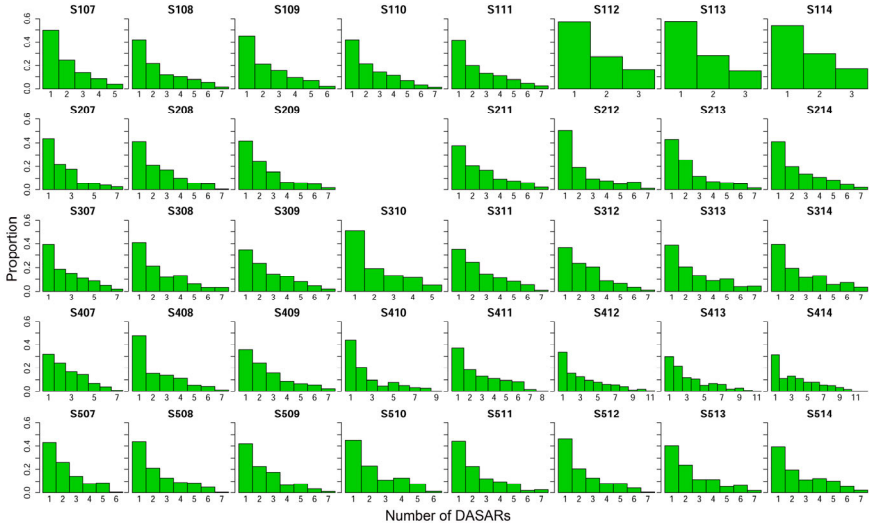


Fig. 12 *Proportion plots for simulated call detections at the 39 site-year combinations using all baseline simulation settings except for the DASAR configuration for which we used the specific configurations from case study (Fig. 9)*

3. Relative Abundance: Lombard Effect

Roaring vs. repeating: how bowhead whales adjust their call density and source level (Lombard effect) in the presence of natural and seismic airgun survey noise

by

Aaron M. Thode, Susanna B. Blackwell, Alexander S. Conrad, Katherine H. Kim, Tiago Marques, Len Thomas, Cornelia S. Oedekoven, Danielle Harris, Koen Bröker

---

Submitted: *Journal of the Acoustical Society of America (JASA)*, September 2019
Revised: *Journal of the Acoustical Society of America (JASA)*, February 2020
Published: *Journal of the Acoustical Society of America (JASA)*, March 2020

JASA ARTICLE

Roaring and repetition: How bowhead whales adjust their call density and source level (Lombard effect) in the presence of natural and seismic airgun survey noise[a)]

Aaron M. Thode,[1,b)] Susanna B. Blackwell,[2] Alexander S. Conrad,[2] Katherine H. Kim,[2] Tiago Marques,[3] Len Thomas,[4] Cornelia S. Oedekoven,[4] Danielle Harris,[4] and Koen Bröker[5]

[1]*Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238, USA*

[2]*Greeneridge Sciences, Inc., 90 Arnold Place, Suite D, Santa Barbara, California 93117, USA*

[3]*Departamento de Biologia Animal, Centro de Estatística e Aplicações, Faculdade de Ciências, Building C2-Level 2, Campo Grande, 1749-016 Lisbon, Portugal*

[4]*CREEM/University of St Andrews, Fife KY16 9LZ, Scotland*

[5]*Shell Exploration and Production Company, 3601 C Street, Anchorage, Alaska 99503, USA*

ABSTRACT:

Over 500 000 automated and manual acoustic localizations, measured over seven years between 2008 and 2014, were used to examine how natural wind-driven noise and anthropogenic seismic airgun survey noise influence bowhead whale call densities (calls/km^2/min) and source levels during their fall migration in the Alaskan Beaufort Sea. Noise masking effects, which confound measurements of behavioral changes, were removed using a modified point transect theory. The authors found that mean call densities generally rose with increasing continuous wind-driven noise levels. The occurrence of weak airgun pulse sounds also prompted an increase in call density equivalent to a 10–15 dB change in natural noise level, but call density then dropped substantially with increasing cumulative sound exposure level (cSEL) from received airgun pulses. At low in-band noise levels the mean source level of the acoustically-active population changed to nearly perfectly compensate for noise increases, but as noise levels increased further the mean source level failed to keep pace, reducing the population's communication space. An increase of >40 dB cSEL from seismic airgun activity led to an increase in source levels of just a few decibels. These results have implications for bowhead acoustic density estimation, and evaluations of the masking impacts of anthropogenic noise. © *2020 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).*

https://doi.org/10.1121/10.0000935

(Received 10 September 2019; revised 27 February 2020; accepted 2 March 2020; published online 31 March 2020)

[Editor: Rebecca A. Dunlop] Pages: 2061–2080

I. INTRODUCTION

After summering in the eastern Beaufort Sea, the Bering–Chukchi–Beaufort (BCB) population of bowhead whales (*Balaena mysticetus*) typically begins its autumn westward migration in late August (Moore and Reeves, 1993). Unlike the spring migration, the autumn migration takes place relatively close to the northern shores of Alaska (Moore and Reeves, 1993). During their travels the animals produce a wide variety of signals that often defy simple classification into specific call types (Ljungblad *et al.*, 1982; Clark and Johnson, 1984; Cummings and Holliday, 1987; Moore *et al.*, 2006; Blackwell *et al.*, 2007), but past work has roughly divided calls between "simple" frequency-modulated (FM) calls and "complex" calls (Blackwell *et al.*, 2007). These calls are distinct from more extended sequences defined as "song," produced during the winter season at more southern latitudes (Blackwell *et al.*, 2007; Stafford *et al.*, 2008; Delarue *et al.*, 2009; Tervo *et al.*, 2009; Tervo *et al.*, 2011). While bowhead song appears to serve a reproductive purpose, the functional purposes of the call repertoire used during the summer remain largely unknown, although it is suspected that long-range communication plays one important role.

Each year from 2007 through 2014, the Shell Exploration and Production Company (SEPCO) commissioned Greeneridge Sciences, Inc. to deploy at least 35 Directional Autonomous Seafloor Acoustic Recorders (DASARs, model C) [(Greene *et al.*, 2004)], divided unequally among five sites in the coastal Beaufort Sea. The motivation behind the effort was to evaluate the potential impact of airgun and other industrial sounds on bowhead whale behavior during their westward fall migration in the relatively shallow Arctic waters off Alaska (Blackwell *et al.*, 2013; Blackwell *et al.*, 2015; Blackwell *et al.*, 2017). Over that entire period, over one

[a)]This paper is part of a special issue on The Effects of Noise on Aquatic Life.

[b)]Electronic mail: athode@ucsd.edu

https://doi.org/10.1121/10.0000935 JASA

million bowhead whale calls were recorded during the fall migrations (Blackwell *et al.*, 2015). To our knowledge, no bowhead song was recorded.

The scale of the dataset, combined with a need for timely analysis, motivated the development of methods for automatically detecting, classifying, and localizing bowhead whale sounds, while exploiting the directional localization capabilities of the DASAR packages (Thode *et al.*, 2012). A team of experienced analysts also manually processed a subset of these data from all years, to serve as a consistency check on the automated results. These combined analyses have previously been used to track seismic airgun activity around the Beaufort Sea (Thode *et al.*, 2010), to determine that the population changes its calling rate in response to both airguns (Blackwell *et al.*, 2015) and industrial activities (Blackwell *et al.*, 2017), to establish source levels and the depth distributions of calling animals during the migration (Thode *et al.*, 2016), and to demonstrate that over the span of seven seasons, the distribution of minimum call frequency decreased from a mean of 94 to 84 Hz (Thode *et al.*, 2017). The previous source level study did not include background noise level as a covariate in the analysis.

Here, this seven-year automatically-analyzed dataset and the manually-analyzed subset are used to examine how both the source level and the spatial density of bowhead whale calls, or "call density" (calls generated per unit area per unit time[1]) vary with changes in continuous natural ambient noise levels and seismic airgun activity. For over a century it has been known that humans increase their speech amplitude in response to increases in background noise levels, an effect pithily dubbed the "Lombard effect," after Eugene Lombard, who first observed the phenomenon in 1911 (Lombard, 1911; Brumm and Zollinger, 2011; Hotchkin and Parks, 2013). This effect has also been reported in multiple terrestrial species (Hotchkin and Parks, 2013) and several marine mammal species, including humpback (*Megaptera novaeangliae*) (Dunlop *et al.*, 2014), right (*Eubalaena glacialis* (Parks *et al.*, 2011; Parks *et al.*, 2012; Parks *et al.*, 2016), and killer whales (*Orcinus orca*) (Holt *et al.*, 2009; Holt *et al.*, 2011). Other studies have also noted changes in call production rate in response to changes in anthropogenic noise levels (Castellote *et al.*, 2012; Melcon *et al.*, 2012; Risch *et al.*, 2012), with calling rates generally decreasing even in the presence of low levels of noise, but sometimes also increasing (Blackwell *et al.*, 2015; Blackwell *et al.*, 2017; Di Iorio and Clark, 2010). Little literature exists on how marine mammals adjust calling rate in response to natural ambient noise fluctuations, but it is now accepted that many species of marine and terrestrial animals respond to changes in background noise levels by varying their source level, call production rate, or call structure/frequency (Bradbury and Vehrencamp, 1998).

Beaufort Sea ambient noise levels are currently dominated by wind-driven sources, since at present shipping and other persistent human activities minimally impact the overall noise environment. The data analyzed in this study thus provide an opportunity to measure how natural variations in an ambient acoustic environment without anthropogenic noise sources could affect the "communication space" of an entire baleen whale population, which is defined by (Clark *et al.*, 2009) as "space over which an individual animal can be heard by other conspecifics, or a listening animal can hear sounds from other conspecifics." These data can also provide insight into how a baleen whale population, in aggregate, could adjust its vocal behavior to compensate for such variations, in an attempt to maintain a fixed communication space.

Over the seven-year period analyzed, several seismic airgun surveys occurred at various distances from the study area. These surveys provide an additional opportunity to directly compare the acoustic strategies used by a baleen whale population to compensate for natural and artificial noise interference, allowing its acoustic response to human industrial noise to be placed within the context of its natural noise response.

Demonstrating behavioral changes in source levels and calling rates in a marine environment is tricky, because changing background noise levels also affect the likelihood that a passive sensor detects a sound (Helble *et al.*, 2013). Higher noise levels "mask" weaker calls, shifting the observed source level distribution upward. As a result, both the measured source level and measured call density distributions become correlated with background noise level, even if a population has no actual underlying behavioral response to these factors. For this reason, this paper will use the term "measured call density" when discussing raw (potentially masked) measurements of call density, while the term "call density" will always refer to the true (unmasked) underlying call density produced by the population.

Section II develops the theory used in this paper to account for masking effects, using distance sampling theory with noise-related covariates. Section III then describes the geography of the field site, the equipment and deployments, methods for automated and manual call detection and localization, sample selection criteria, procedures for measuring continuous noise and airgun exposure levels, and procedures for statistical regression. Section IV presents the conditional probabilities of call density and source level as a function of background noise level and analysis type (i.e., manual or automated), as well as regression analyses of source level and call density vs background noise and airgun cumulative sound exposure levels (cSEL). Finally, Sec. V discusses the similarities and differences between the population-level response to natural and anthropogenic noise, and outlines the relevance of these observations to passive acoustic density estimation.

II. MODIFIED POINT TRANSECT THEORY FOR REMOVING MASKING EFFECTS

A. Definition of localization probability P_a

An appropriate measure of a population's behavioral response to noise is the *conditional* probability density function (PDF) that a source level SL is generated, given a fixed

JASA

https://doi.org/10.1121/10.0000935

noise level *NL*: $p(SL|NL)$, which is defined here as the *behavioral response distribution*. If the population does not exhibit a Lombard effect, then over sufficiently long measuring times its source level distribution becomes independent of the noise distribution, such that $p(SL|NL)=p(SL)$ and $p(SL,NL)= p(SL)\,p(NL)$. This behavioral response distribution is not directly measured from data; instead, one observes the joint PDF $p(+, SL, NL|R_{max})$ of the *measured* bowhead source levels and associated noise levels, where $p(\text{“+”})$ indicates the probability that a call is detected AND localized, and R_{max} is the maximum range from the closest sensor from which localized calls are accepted.[2] This *observed*, or *masked*, distribution represents the probability density of measuring a given noise level *NL* and localizing a call with source level *SL* within distance R_{max}. This observed distribution is thus weighted by the probability that a given noise level $p(NL)$ occurs. The joint PDF can be estimated from an appropriately normalized two-dimensional histogram of all measured call samples. The *SL* and *NL* of the joint PDF are always computed using the same units.

Basic probability theory yields

$$p(SL|NL,R_{max}) = \frac{p(+,SL,NL|R_{max})}{p(NL)p(+|SL,NL,R_{max})}, \tag{1}$$

where we have explicitly retained a potential dependence of the derived conditional distribution on R_{max}, even though the true underlying behavioral response distribution should be independent of R_{max}. Equation (1) shows that two correction factors must be applied to the observed distribution to obtain the underlying behavioral response distribution. The first factor, $p(NL)$, the observed distribution of noise throughout all seasons, converts the observed joint probability into a probability conditioned on *NL*, and is readily estimated from the data. Following Buckland *et al.*, 2012, the second factor, $p(+|SL,NL,R_{max})$, can be rewritten as a *localization probability* $P_a(SL,NL,R_{max})$, which represents the average probability that a call within radius R_{max} of the closest sensor is both detected *and* localized by the system, given that the call's source level is *SL* and the background noise level is *NL*. P_a depends on source level, noise level, and the value of R_{max} selected. In principle, the azimuth of a call with respect to a sensor should also affect P_a, since the ability to locate a call depends on the relative location of other sensors. We found, however, that if all DASARs distributed at a site were incorporated into estimating $p(+, SL, NL|R_{max})$, the resulting distribution showed no azimuthal variation.

There are several potential approaches to estimating P_a. The simplest approach, and the one used here for analyzing call rates, only uses samples that lie within a small value of R_{max}, so that any sound generated within that radius is assumed detectable and localizable ($P_a \sim 1$), regardless of the call's source level or ambient noise conditions. Past work on this dataset effectively took this approach by concluding that calls generated within 2 km of the nearest DASAR were always localizable (Blackwell *et al.*, 2015).

The second approach, used in this source level analysis, takes advantage of the relatively flat bathymetry and simple propagation environment surrounding the DASAR sensors to apply a modified point transect analysis, a particular version of distance sampling theory (Buckland *et al.*, 2012). This approach empirically estimates the localization probability of a call as a function of both its range to the nearest sensor and its *source level-to-noise* ratio (SLNR), thus providing a means of correcting masking effects out to ranges of at least 30 km. Among other assumptions, appropriate use of distance sampling theory requires that the mean spatial density of calling animals (call density), when measured over long enough intervals, is independent of distance from a sensor, such that any apparent variation in the measured spatial density with range can be attributed to changes in the detectability of calls. Since the DASAR sensors were deployed in the middle of the migration corridor of bowhead whales, and the dataset spans multiple years, it is reasonable to assume that the true call density of animals across the study site averages out to a constant value, a conclusion that is supported by the resulting analysis.

A more general approach to treating masking, not used here, is to conduct Monte-Carlo-type modeling that combines simulated source signals, propagation modeling, and bootstrapped noise samples and then passes the resulting synthesized time series through a detector (human or automated) to numerically estimate a detection probability (Küsel *et al.*, 2011; Helble *et al.*, 2013). While this approach can handle more complex propagation environments and situations where call densities are heterogeneous, this approach is also time consuming, assumes considerable knowledge about the propagation environment, and can be difficult to evaluate when multiple humans have been involved in analyzing the original dataset, due to various biases human operators display when choosing what to detect and localize (Urazghildiiev and Clark, 2007; Moyer-Horner *et al.*, 2012).

B. Modified point transect theory

Given these assumptions, distance-sampling theory states that the average localization probability P_a can be computed as follows:

$$\begin{aligned} P_a(SL,NL,R_{max}) &\equiv p(+|SL,NL,R_{max}) \\ &= \int_0^{R_{max}} p(+|SL,NL,r)\pi(r)dr \equiv \int_0^{R_{max}} g(SL,NL,r)\pi(r)dr. \end{aligned} \tag{2}$$

Here r is the range to the closest sensor, and $g(SL,NL,r)$, the *localization function*, is the probability that an animal is localized, given a source level *SL*, noise level *NL* and range r. Note that most distance sampling literature defines $g(r)$ as the “detection function” and P_a as the “probability of detection,” but here $g(r)$ and P_a are defined, respectively, as a “localization function” and “probability of localization,” to emphasize that these quantities reflect the probability that

https://doi.org/10.1121/10.0000935 JASA

a call is detected on two or more sensors, and can thus be localized.

$\pi(r)$, the probability that an animal is present at range r, is defined in the distance sampling literature (Buckland *et al.*, 2012) as "the distribution of distances available for detection," an unwieldy moniker that is abbreviated to "availability function" for the rest of the paper. Under the assumption of uniform call density, $\pi(r)$ becomes proportional to the geometric perimeter defined by points that lie at range r from the sensor.[3] The scenario where an animal's range is measured from a single location is defined as a *point transect* and, under that particular geometry, $\pi(r) = 2\pi r/\pi R_{max}^2 = 2r/R_{max}^2$. For a simple point transect the availability function is proportional to the circular perimeter with radius r. However, the majority of passive acoustic systems that can measure range (including the configurations to be discussed here) require at least two spatially distributed sensors to obtain a distance estimate from the closest sensor. The requirement for distributed tracking arrays requires modifications of standard point transects, and as a result $\pi(r)$ becomes a more complex function that depends on the number and spacing of sensors in the tracking system. The Appendix provides these modified availability function definitions. We also re-emphasize that as a consequence of using a distributed tracking array $g(r)$ *actually* represents a probability of *localization*, instead of the more standard probability of *detection*, and for that reason we will continue to refer to $g(r)$ as a *localization function* instead of the more standard "detection function" terminology.

The localization function is generally modeled as a parameterized "key" function with the property that $g(0) = 1$ and $g(\infty) = 0$. A variety of standard functions exist, including the half-normal and uniform, but the best-fit model was found to be the hazard-rate (Buckland, 1992), since it provides two adjustable parameters:

$$g(SL, NL, r) = 1 - \exp\left[-\left(\frac{r}{\sigma(SL, NL)}\right)^{-b(SL,NL)}\right]. \quad (3)$$

Here σ is defined as the *scale* parameter, as it defines the range scale over which the localization function begins to rapidly decrease. Larger values of σ indicate higher detectability levels at greater ranges. Meanwhile b, the *shape* parameter, defines the "sharpness" of the drop-off in localization probability with range, with larger values of b indicating a sharper transition. Both parameters are assumed to vary with source and noise level.

From Eq. (2), the probability $f(SL,NL,r)$ of observing a call at range r becomes

$$f(SL, NL, r) = \pi(r)g(SL, NL, r)/P_a(SL, NL, R_{max}). \quad (4)$$

If M calls with the same source level are observed under the same noise conditions, each at a range r_m from the sensor, then the log-likelihood function of these observations is proportional to $\sum_{m=0}^{M} \log(f(r_m))$. Applying maximum likelihood methods to Eqs. (2)–(4) then yields best-fit estimates of the localization function's scale and shape parameters for all samples that share the same SL and NL. The **R** package **Distance** provides standard software for obtaining the maximum-likelihood solution (Thomas *et al.*, 2010) and was applied here. For reasons detailed in the Appendix, localization ranges were binned with 250 m resolution before maximizing the likelihood.

Despite the large sample sizes available from this dataset, the sheer number of possible combinations of source and noise levels meant that obtaining sufficient sample sizes could be challenging for many SL and NL combinations. While "multi-covariate" distance sampling (MCDS) techniques have been developed to handle situations where a detection function with multiple covariates must be estimated from limited data (Marques and Buckland, 2004), in this situation one can exploit the sonar equation to reduce the number of covariates needed for the localization function. Specifically, the sonar equation gives the signal-to-noise ratio (SNR) of a signal with source level SL received on a sensor at range r as (expressed in dB units) SNR = SL-$TL(r)$-NL = [SL-NL]-$TL(r)$ = SLNR-$TL(r)$, where SLNR is the source-level-to-noise ratio (or SL-NL in dB terms) and $TL(r)$ is the *transmission loss* arising from the sound propagating a distance r through the environment. The sonar equation demonstrates that signals that originate at the same location and share a common SLNR will generate the same set of SNR values across a distributed group of sensors. If we assume that SNR is the dominant factor in determining the detectability for low-frequency, low-directionality signals, then signals that share the same SLNR should share the same detection probabilities across all sensors, and thus share the same localization function: $g(SL, NL, r) = g(SL\text{-}NR, r) = g(SLNR, r)$. Thus, the SLNR, which can also be interpreted as a source level "normalized" by background noise level, becomes the only relevant covariate for $g(r)$; consequently, all data samples that share the same SLNR can be lumped together to estimate the localization function. We found that samples arranged into 2 dB SLNR bins between 52 and 90 dB provided sufficient sample sizes to fit a localization function. Below 52 dB there were so few localized calls that we felt a localization function could not be fitted, and thus calls below 52 dB SLNR were not incorporated into the analysis.

III. METHODS

A. Equipment and deployment configuration

DASARs are autonomous acoustic recording packages equipped with an omnidirectional acoustic pressure sensor (sensitivity of −149 dB re V/1 μPa) and two horizontal directional sensors capable of measuring the north–south and east–west components of acoustic particle velocity. This arrangement permits the azimuth of received sounds, such as bowhead whale calls, to be measured from individual DASARs. Each time series is sampled at 1 kHz with a maximum usable acoustic frequency of 450 Hz due to antialiasing filter roll off. Coincident bearings to calls detected on different DASARs are combined via triangulation to yield

JASA

https://doi.org/10.1121/10.0000935

two-dimensional call positions, from which the range of each call to every DASAR can be estimated (Greene *et al.*, 2004). This ability to measure bearing from a single point allows a location to be estimated using only two DASARs (instead of three to four nondirectional sensors), but DASARs still require modifications of point transect theory (Appendix).

From August to October, 2008 to 2014, between 35 and 40 DASARs were deployed across a 280 km swath off the Alaskan North Slope, on the continental shelf in water depths between 20 and 53 m. The deployments were grouped into "Sites," labeled 1–5 traveling from west to east (Fig. 1).

Most sites included seven DASARs, deployed in a triangular grid with 7 km separation and labeled "A" to "G" from south to north. The analysis presented here merged data collected at Sites 3 and 5, as these sites had identical layouts. The analysis excludes data from the first year of the study (2007), when a different type of sensor was used in the DASARs.

Bowhead whale calls in the raw acoustic data were post-processed two ways: by a team of human analysts, and by a six-stage automated detection and localization program. Both approaches have been extensively described and evaluated in other publications (Thode *et al.*, 2012; Thode *et al.*, 2016; Thode *et al.*, 2017). Regardless of the particular approach used, each detected call event on every DASAR was assigned a start time, duration, frequency bandwidth, and range. Call events matched between DASARs yielded both a 2-D location estimate and uncertainties in azimuth and range.

B. Sample selection criteria

Call detections and localizations varied in quality, so three selection criteria were applied to determine whether a particular call was included in subsequent analyses:

(1) A call's localized range to the closest DASAR had to be less than a threshold value R_{max}. Two values of R_{max}—3.5 and 30 km—are examined in Sec. IV. The 3.5 km threshold was selected because the distance sampling analysis showed that DASARs are effective for detecting and localizing these calls for most source level values, regardless of ambient noise conditions (Blackwell *et al.*, 2013). These data are thus assumed to be unaffected by noise masking, but yield smaller sample sizes. The 30 km range cutoff permits larger sample sizes, but must be corrected for masking effects.

(2) The frequency band covered by the call's fundamental component had to lie between 20 and 170 Hz, a low-frequency cutoff enforced in order to justify the assumption of an omnidirectional call directivity.

(3) For analyses involving call source level (but not call density), the call's estimated source level had to be within 6 dB of the source level computed from any other DASAR detecting the same call, a metric dubbed the "discrepancy" in Thode *et al.* (2016). This procedure provides a safeguard against the possibility that the automated algorithm captured only a small fragment of a call, generating an inappropriate source level value. Had only a fragment of a call been captured on one or more DASARs by the automated detection process, then the estimated source level would vary between the DASARs and the discrepancy of the call would be high. A more restrictive 3 dB discrepancy criteria was found to lower the sample size substantially, but not change the statistical regression results.

Filtering call samples by their range uncertainty had little impact on the estimated localization function, even though calls generated at ranges greater than 20 km from the DASAR site displayed substantial range uncertainties. Thus, call samples were included in the two sets (manual and automated) regardless of their range uncertainty.

C. Metrics for continuous in-band noise and airgun survey exposure

Calls that passed the above criteria were assigned the source level (SL) and noise level (NL) values derived from

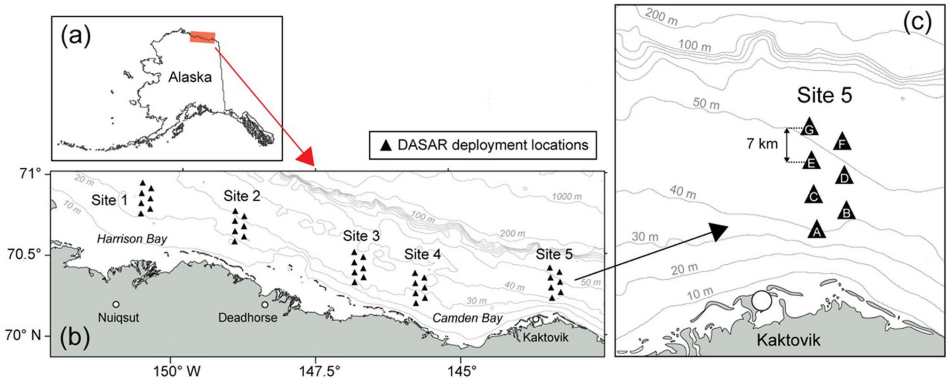


FIG. 1. (Color online) Locations of passive acoustic deployments. (a) North Slope of Alaska; (b) DASAR deployments. (c) Closeup of Site 5 deployment and bathymetry.

https://doi.org/10.1121/10.0000935

JASA

the DASAR closest to the call's position. We assumed that the noise level measured at a DASAR was the same as the noise level experienced by a whale within 30 km range, because noise levels between DASARs at the same site generally varied by only a few decibels over the same time interval.

Noise levels associated with each call were computed by extracting a time series that had the same duration and bandwidth as the call, but starting 3 s before the start of the call sample. A 3 s time shift was chosen because most bowhead whale calls have less than 2 s duration, and an extra buffer second was added to reduce the possibility that any signal energy from a bowhead call might contaminate the noise sample. The noise sound exposure level (SEL), root-mean-square (RMS) sound pressure level (SPL), and peak power spectral density (PSD) were computed by band-pass filtering the noise sample over the same bandwidth as the call. The possible presence of airgun signals or other nonstationary transients in the noise sample was checked by comparing noise metrics from the first half of the noise sample with the second half. Metrics that differed by more than 3 dB led to that particular sample (noise and associated call) being rejected from further consideration. Repeating these analyses with noise samples taken *after* each call yielded no change in the results.

Noise levels were also calculated over a fixed bandwidth between 20 and 170 Hz for each call, in order to allow the "*in-band*" noise levels to be compared with more familiar fixed-bandwidth measurements. We found that RMS fixed-bandwidth measurements were generally 15 dB greater than a typical in-band RMS measurement. For R_{max} =3.5 km the in-band noise sample bandwidth distribution was skewed, with mean, median, and mode values of 52, 46, and 32 Hz, respectively, with 25 Hz standard deviation. For the 30 km limit the distribution values shift slightly lower (45, 38, 24 Hz; 24 Hz standard deviation), reflecting the fact that more distant calls display narrower received bandwidths due to propagation attenuation. Thus the 15 dB difference between the in-band and fixed 150 Hz bandwidth noise samples must arise partially from the reduction in noise sample bandwidth [$10\log_{10}(50\,\text{Hz}/150\,\text{Hz}) \sim 5$ dB] with the remainder arising from larger noise levels at lower frequencies (a spectral tilt exists in the noise spectrum).

Seismic airgun activity was detected using the same automated algorithm described in detail by Thode *et al.* (2012) and Blackwell *et al.* (2015). Seismic activity was designated as "present" for a given call on any DASAR if at least one airgun pulse was detected on DASAR G at the same site within 5 min of the detected call. The motivation for using DASAR G data was that it was the deepest location at each site, and thus offered the best probability of detecting the greatest number of airgun pulses of all the sensors at the site. Using detection criteria from the deepest location thus acted as a safeguard against missing pulse detections on shallower DASARs. If airgun pulses were associated with a call, then the cumulative sound exposure level (cSEL; dB re 1 μPa^2-s) was measured over the 10-min window centered around the call, the same metric used by Blackwell *et al.* (2015). However, that previous study measured cSEL over fixed nonoverlapping 10-min intervals, while here the cSEL associated with each call was integrated over all airgun pulses detected within 5 min of a given call. The integration only included time windows when an airgun pulse was deemed present; bowhead whale calls and ambient wind-driven noise were thus excluded from the cSEL calculation. The cSEL calculations were also computed over the entire 10–450 Hz bandwidth, consistent with Blackwell *et al.* (2015).

D. Call density analysis

Blackwell *et al.* (2015) previously analyzed the relationship between call density and seismic airgun activity, but did not include ambient noise level as a predictor variable. That study computed call density by counting call localizations that occur within contiguous, nonoverlapping, 10-min windows that start at the top of the hour. Localizations were only counted if they occurred within 2 km of the closest DASAR, entirely sidestepping the issue of masking effects by assuming a localization probability of one for all calls within 2 km range. The call counts within these "cell-time intervals" were then used as the dependent variable in estimating the impact of seismic airgun sounds on call density (and thus underlying call production rates).

The main issue of applying this method to the present study is that it cannot distinguish between times when animals are present but silent, and when animals are not present at all, so the resulting cell-time interval samples have many zero values that bias the subsequent measured call rate distribution. There are also boundary artifacts, in that the measured density assigned to calls that occur at the start of a new time window are not influenced by the presence of calls detected at the end of the previous window. Furthermore, it is challenging to define ambient noise level for a collection of calls with different frequency content. For this reason, a modified version of the analysis was applied here. For a given call localized within 3.5 km of a particular DASAR, all other calls detected within 5 min of that given call (and also located within 3.5 km of the same particular DASAR) were tabulated to assign a measured call density to the given call, along with its associated noise level. While the frequency range of the calls remained restricted to values between 20 and 170 Hz, the calls' discrepancies or other localization features were not used to filter calls, because the discrepancy criteria is relevant only for evaluating source level and not whether a call was generated close to a DASAR. By selecting R_{max} equal to 3.5 km, we assumed that all calls generated within that radius were successfully localized, a strategy similar to that used for Blackwell *et al.* (2015). That particular study used only a 2 km radius; the choice of a larger 3.5 km radius for this analysis will be justified in Sec. IV B.

Assigning a unique call density to each sample ensured that call density was only measured whenever animals were

JASA

https://doi.org/10.1121/10.0000935

present, and avoided the boundary artifacts mentioned above. However, centering a time window on each individual call created samples that were not statistically independent; the auto-covariance between call density measurements 50 samples apart was 0.5, which only fell to 0.1 at 250-sample lag separation. Given a median time separation of 3 min and 23 s between samples, these lag scales are the time equivalent 3 and 14 h, respectively.

A preliminary regression analysis using generalized estimation equations (GEE) (Dobson and Barnett, 2008) with a first-order autoregressive covariance structure found that although measured call density samples were correlated in time, the impact of nonindependent samples could be neglected. Although call densities for adjacent samples were highly correlated, the regression results from GEE were virtually the same as a conventional generalized linear model (GLM), because the time scale used for the complete analysis (months) was much greater than the correlation window in the data (hours). The statistical analysis thus focused on the GLM regressions.

Distributions of call density had a long tapering tail to the right (higher densities), but a normal probability plot showed that the logarithm of call density fit a normal distribution, so measured distributions were formulated in terms of logarithms. The relationship between the logarithm of calling density, noise level, and cSEL was computed with a GLM using a polynomial fit up to fourth-order with interactive terms permitted between noise and cSEL, and assuming a normal distribution for the response variable. Calls (and their associated call rates) were separated into sets where airgun activity was either present or absent. Separate models were fitted to each nonoverlapping dataset, in order to implement an efficient dose-response model for airgun activity for the airgun-present dataset. For calls detected without airgun presence, in-band noise level was the only predictor variable, while the calls detected during airgun presence used in-band noise level and cSEL as predictor variables. The Bayes Information Criterion (BIC) was used to establish the highest-order terms permitted in the model. The resulting residuals were examined to confirm the fit was a normal distribution.

E. Source level analysis

Call received levels and positions were combined with an acoustic propagation model to derive the estimated source level of the call within the 20–170 Hz frequency band, under the assumption that the low-frequency acoustic radiation propagating from the animal was omnidirectional (i.e., the source level would be the same regardless of the animal's aspect relative to the sensor). Three different propagation models were tested: a 15log$R$ power-law transmission loss model, a Pekeris waveguide model, and a normal mode propagation model that incorporated source depth, sound speed profile, water depth, and bottom sediment profile (Thode *et al.*, 2016). All three models yielded results within three dB of each other in terms of source level distribution estimates, so the simple power-law transmission model was retained for the rest of the analysis.

Source levels were computed using four metrics: sound exposure level (SEL; dB re 1 μPa^2-s @ 1 m), root-mean-square sound pressure level (SPL; dB re 1 μPa @ 1 m), and median and maximum power spectral density (PSD; dB re 1 μPa^2/Hz @ 1 m). The last two metrics were estimated by computing a spectrogram of the call using a 512-point fast Fourier transform (FFT) with 90% overlap, collating all time-frequency cells (Δf = 1.95 Hz; Δt = 51.2 ms) that lie within the "bounding box" of the call localization, and extracting the median and maximum PSD values from the resulting distribution. The call duration was simply defined as the duration of the bounding box. Calls that had noise samples contaminated by airgun signals were rejected.

The regression analysis of the masked source level distribution followed a procedure similar to that of the call rate analysis. For each combination of localization method (manual; automated) and R_{max} (3.5, 30 km), calls were divided according to whether airgun activity was absent or present. Calls belonging to the airgun-absent set were fit using a normal GLM to up to a fourth-order polynomial regression using in-band noise level as the only predictor variable, and using the BIC to set the maximum model order. Calls in the airgun-present set included cSEL as an additional, potentially interactive, parameter.

The regression analysis for the unmasked source level distribution required some additional steps. The original raw call samples were binned according to airgun cSEL level, with the bins being defined as 0 (no airgun presence) and within the 90–140 dB range, with 2 dB increments. For every cSEL bin k, the N_k joint observations of source level SL and noise level NL that existed in that bin were used to construct a normalized two-dimensional histogram estimate of the joint PDF $p(+, SL, NL|R_{max}, cSEL)$. The SLNR of each histogram bin was then calculated, and the appropriate value of $P_a(R_{max}, SLNR)$, as computed from point transect theory, was divided into the bin value to generate the unnormalized unmasked joint probability $p(SL, NL|R_{max}, cSEL)$. (We thus assumed that the presence and intensity of airgun survey activity did not influence call detectability). After every histogram bin was readjusted, the entire distribution was renormalized (so that integrating the joint distribution over all values of NL and SL yielded one). We then numerically resampled this unmasked joint distribution N_k times to generate N_k new joint estimates of SL and NL for cSEL bin k. The complete set of resynthesized observations was then applied to the same regression analysis as the original masked samples.

Two additional statistics were also computed: the behavioral response distribution $p(SL|NL)$ (defined in Sec. II A) and the derivative of the regression curves with respect to in-band noise level, defined here as the "*behavioral sensitivity*." The behavioral response distribution was obtained by calculating the marginal distribution $p(NL)$ for both masked and unmasked distributions, and then using Eq. (1).

The behavioral sensitivity reveals how the mean source level of the calling population increases with respect to a

https://doi.org/10.1121/10.0000935

JASA

TABLE I. Sample sizes used for analyses of Sites 3 and 5.

Dataset criteria	Method	Max range (km)	Site 3	Site 5	% with airgun	Total
All samples (Sites 3 and 5)	Manual	—	49 940	64 207		114 147
	Automated	—	361 149	402 379		
Short range	Manual	3.5	4727	5381	44.11	10 108
	Automated	3.5	42 662	46 915	21.73	89 577
Long range	Manual	30	18 673	21 865	45.95	40 538
	Automated	30	198 144	206 328	21.45	404 472

unit increase in noise level (ΔSL/ΔNL), and thus indicates how "sensitive" a population's acoustic behavioral response is to ambient noise fluctuations at various in-band noise levels. For example, a ΔSL/ΔNL value of one at all noise levels would indicate that the population's mean source level increases 1 dB for every 1 dB increase in ambient noise level, regardless of the original ambient noise level involved: a perfect Lombard adjustment. By contrast, a sensitivity value of zero at all noise levels would indicate the population's source level distribution is completely indifferent to ambient noise changes.

IV. RESULTS

A. Sample sizes

Table I shows the call sample sizes available from the four different datasets. Sites 3 and 5 contributed roughly equal numbers of samples to the analysis, even when the range, frequency, and discrepancy restrictions are applied. Although the manual datasets are only ~10% of the sample size of the automated results, they had nearly double the percentage of calls associated with airgun pulses when compared with the automated datasets. The reason for this is that the manual analyses were originally intended for use in evaluating the impact of airguns on bowhead whale behavior, so the subset of days selected for manual analysis was not sampled randomly.

B. Distance sampling analysis

Figure 2 displays histograms (normalized in terms of probability density) of the distribution of localized calls from the automated long-range dataset, as a function of range from the closest sensor and for three different values of SLNR: 62, 72, and 82 dB. Overlaid on the histograms are the best-fit estimates of Eq. (4), $f(SLNR,r)$, using the hazard rate model of Eq. (3) and the seven-sensor availability function defined in Eq. (A2), Appendix.

Each curve matches the observed distribution well. As the SLNR increases, the most probable range for localizing a call increases: a natural consequence of point transect theory, since larger numbers of calls are available at greater ranges, and calls become more detectable as their relative source level increases. The good fit between the data histograms and the theoretical curves justifies the assumption of an uniform spatial distribution of the migration coordinator, which accumulated across multiple years.

The top row of Fig. 3 replots this observed and modeled $f(r|\mathrm{SLNR})$ as two-dimensional images with respect to both range and SLNR (in 2 dB increments), illustrating that calls with 52 dB SLNR seem to be the lower limit for localization capability with the 7-km-spaced array at this study site.

The middle row of Fig. 3 displays the corresponding empirical and modeled localization function $g(SLNR, r)$ as a function of range and SLNR. We also estimate g empirically [Fig. 3(c)] by dividing the observed $f(r|SLNR)$ by the availability function $\pi(r)$ [Eqs. (4) and (A2)] and setting the maximum value attained to 1.

Even with large sample sizes, relatively few high-SLNR signals occur at small ranges, so the upper-left region in Fig. 3(c) is undersampled simply because of the low

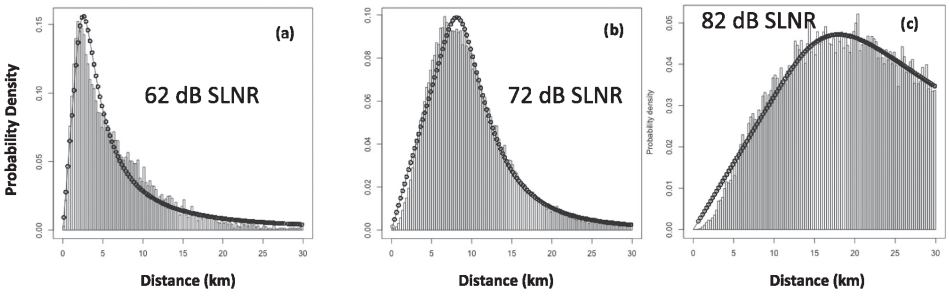


FIG. 2. Observed call range distribution $f(SLNR,r)$ at Sites 3 and 5 combined, using long-range automated analysis (Table I), evaluated at three different source level-to-noise ratios (SLNR): (a) 62 dB; (b) 72 dB; (c) 82 dB. Histograms have been normalized to approximate probability density functions (PDFs). Dark circles represent modeled fit of f [Eq. (4)] assuming a hazard-rate localization function [Eq. (3)] and a distributed sensor availability function (A2). For each localization at Sites 3 and 5, only the range to the closest DASAR is used.

JASA
https://doi.org/10.1121/10.0000935

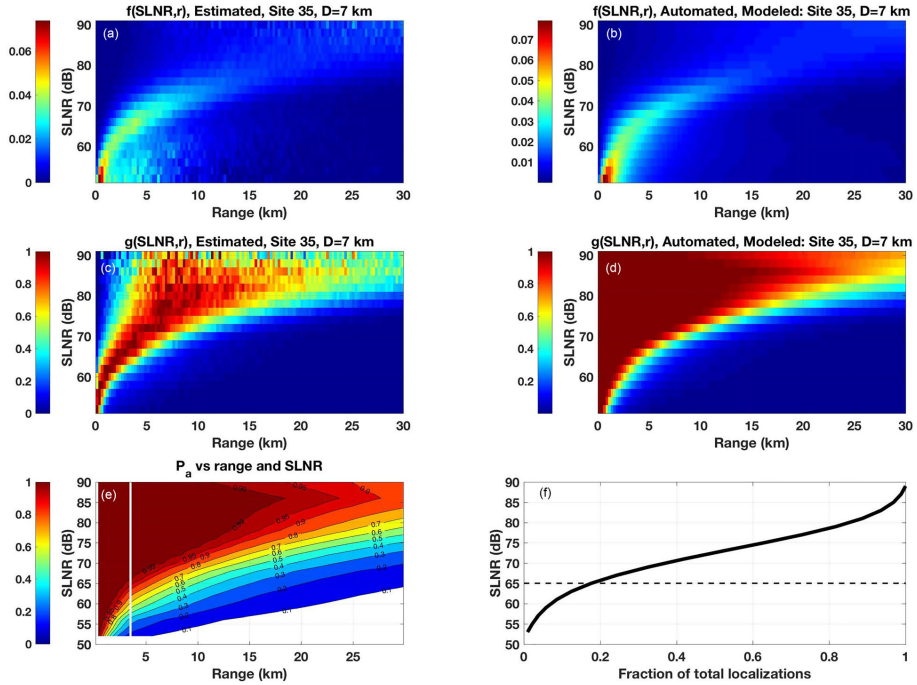


FIG. 3. (Color online) Computation of localization probability P_a via modified point transect theory, using automated data. (a) Observed $f(SLNR, r)$ imaged in grid steps of 2 dB SLNR and 250 m range; (b) best-fit modeled $f(SLNR,r)$; (c) estimated $g(SLNR,r)$ $[f(SLNR,r)/\pi(r)]$; (d) best-fit modeled $g(SLNR,r)$; (e) probability of detecting a call within radius r (white line =3.5 km), as a function of $SLNR$ [Eq. (2)]; (f) cumulative distribution of calls as a function of SLNR (dashed line = 65 dB, the SLNR with 99% localization probability at 3.5 km range). The figure illustrates that over 80% of calls have $SLNR$ levels that exceed 65 dB.

probability that a relatively rare loud call is produced at close range. This undersampling thus generates artificially low values of the estimated $g(r)$ in this region. The modeled form of $g(r)$ in Eq. (3) [Fig. 3(d)] is not influenced by undersampling, since Eq. (3) forces the model into maintaining monotonic decreases in detectability with increasing range. Both the estimated and modeled localization functions display similar behavior at greater ranges, with the shoulder of the localization function (point at which the localization function begins decreasing substantially) increasing with larger SLNR, as expected. Figure 3(e) shows the values of $P_a(SLNR,R_{max})$ derived from the localization function using Eqs. (2) and (A2), while Fig. 3(f) displays the cumulative distribution of SLNR values from the observed data. Subplots (e) and (f), when taken together, support earlier conclusions by Blackwell *et al.* (2015) that calls located within 2 km of the closest sensor are virtually always detectable: for example, the two subplots combined demonstrate that calls within 2 km range and with SLNR >60 dB (which comprise 90% of all calls) have a >0.99 probability of being detected. If R_{max}=3.5 km [vertical line in Fig. 3(e)], then calls with 65 dB SLNR or higher, which comprise over 80% of the calls [dashed line in Fig. 5(f)], have a 99% probability of localization. The choice of R_{max} = 3.5 km, instead of 2 km, for the call density analysis in Sec. IV C, was made in an attempt to balance sample size against potential masking effects. While using locations less than 2 km would have eliminated all possibility of masking, only 2748 samples would have been analyzed, too few to allow a robust regression of source level vs noise level. Using 3.5 km as an upper limit quadrupled the sample size (Table I) while only raising the prospect of masking for the weakest calls.

Figure 4(a) shows the best-fit values of the scale (σ) parameter in Eq. (3) as a function of range and analysis type, and confirms that the "shoulder" of the localization function increases with increasing SLNR, rising from less than a kilometer at 52 dB to nearly 30 km for an 85 dB SLNR.[4]

For a fixed SLNR, the manual analysis generally yields a larger scale parameter than the automated analysis, implying that the human analysts are able to localize weaker signals. As discussed in Thode *et al.* (2012), the automated call localization algorithm sets a RMS detection threshold of 8 dB over 50 Hz bandwidth, so it is not surprising that the

https://doi.org/10.1121/10.0000935

JASA

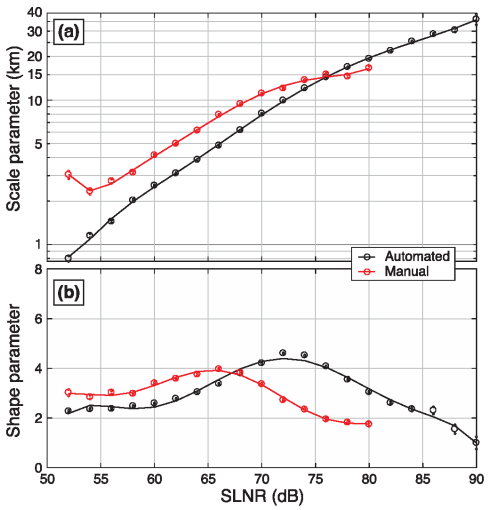


FIG. 4. (Color online) Scale "σ" (a) and shape "b" (b) parameters of hazard rate localization function vs SLNR for both automated (black) and manual (red) data analysis. Small dots indicate standard error of parameter fits.

effective localization range of the automated procedure is a few kilometers smaller than the manual analysts', who can detect calls much weaker than 8 dB SNR in spectrograms. Beginning at 70 dB SLNR, the manual scale parameter begins to taper off, crossing below the automated scale at 76 dB SLNR. We interpret this tapering off as an artifact arising from the relatively few samples available at high SLNR for the smaller manual dataset.

Figure 4(b) shows the variation of the shape parameter b, which determines the sharpness of the localization cutoff. The results suggest that, at low SLNR values, human analysts have a sharper cutoff in their localizing ability (i.e., they can localize most calls out to the range defined by the scale parameter, then drop off quickly beyond that). However, once the SLNR increases past 68 dB, the automated algorithm achieves a sharper cutoff. The decrease in the shape parameter for the manual analysis at high SLNR levels may be a sample size artifact, and not a fundamental measurement of human localization performance at high SLNR values.

C. Call density vs continuous in-band noise levels and seismic airgun exposure

As a reminder, the term "call density" in this section refers to a call density estimate where noise masking effects have either been removed or deemed negligible.

Background in-band noise levels and airgun cSEL levels were correlated with Pearson correlation coefficients of 0.23 and 0.26 for the respective manual and automated short-range datasets, because only airgun pulses with higher cSEL can be detected at high noise levels. This correlation is the reason why a fourth-order polynomial fit was the highest possible for the regression analysis, because attempts to fit higher orders became numerically unstable.

Figure 5 (left column) displays the log-normal regression prediction results of call density vs background noise intensity, for situations where seismic activity is nonexistent (red dashed line), present at small levels (cSEL of 100 dB re 1 uPa^2-s; green solid line), and present at moderate/heavy levels (cSEL of 120 dB re 1 uPa^2-s; gray dashed line). The right column shows the predicted call rate vs airgun cSEL level at a fixed in-band SPL noise level of 90 dB. The top and bottom rows represent the manual and automated analyses.[5]

Restricting our attention to the automated regression analysis results [Figs. 5(c), 6(c), and 5(d)], we find that whenever airgun activity is absent [dashed gray line in Fig. 5(c)], call density increases from roughly 0.2 to 0.4 calls/min within 3.5 km range as the ambient noise levels increase from 65 to 105 dB (40 dB), although the response tapers off beyond 95 dB.

Replicating previous research on the same dataset (Blackwell *et al.*, 2015), Fig. 5(c) shows that the presence of even low levels of seismic activity results in an increase in call density, given the same fixed background noise level. For example, at weak airgun exposures of 100 dB cSEL, the call density normally produced at background noise levels of 90 dB RMS [vertical black line in Fig. 5(c)] increases 31% from 0.38 to 0.5 calls/min. In order to return to the baseline call density, ambient noise levels would have to decrease 12 dB to 78. The presence of airgun pulses therefore boosts call density by 23% to 40%, depending on the initial ambient noise level.

In contrast, as seismic survey cSEL levels continue to increase, call densities are gradually suppressed. Although not shown in Fig. 5, at 115 dB cSEL call densities match those produced at ambient baseline levels. At higher levels call densities become suppressed below baseline states, as can be seen for the red curve in Fig. 5(c) for 120 dB cSEL. If the in-band noise level is 90 dB, an increase in the cSEL from 100 to 135 dB roughly halves the call density from 0.5 to 0.25 calls/min. Thus a 40 dB SPL increase in in-band ambient noise level prompts roughly the same response as a 35 dB decrease in airgun cSEL (when the airgun cSEL does starts at high levels).

As discussed previously, if noise levels are measured over a fixed bandwidth of 20–170 Hz, the resulting noise levels are roughly 15 dB higher than the in-band levels reported in the previous paragraph.

When airgun presence is treated as a simple categorical variable (no cSEL value incorporated), the regression finds no significant relationship between call production rate and airgun presence.

D. Call source level vs continuous in-band noise levels and seismic airgun exposure

Figures 6 and 7 display various joint and conditional probabilities of source and background noise levels,[6] displayed on a logarithmic scale (log-10), for the various

JASA

https://doi.org/10.1121/10.0000935

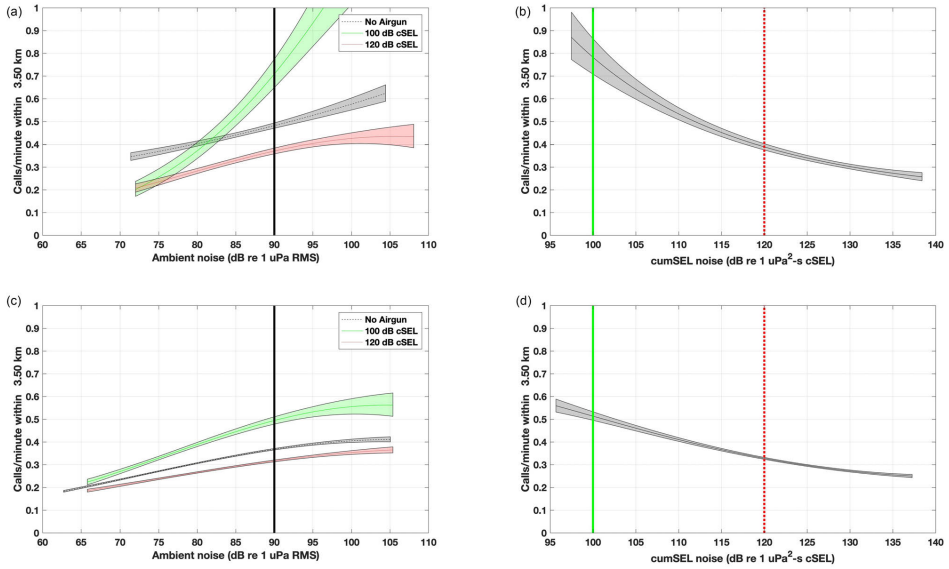


FIG. 5. (Color online) Regression analysis of call density vs in-band ambient noise level [SPL (RMS)] and airgun cSEL. The top and bottom rows display manual and automated analyses, respectively. (a) Call density vs ambient noise level when no seismic activity is present (black, dashed), when seismic activity generates 100 dB re 1 Pa^2-s cSEL (green, dashed), and when seismic activity is 120 dB (red, dashed). (b) Call density vs seismic cSEL, with ambient noise level fixed at 90 dB SPL [black vertical line in subplot (a)]. The green and red vertical lines indicate the cSEL levels held fixed in subplot (a). Subplots (c) and (d) show the corresponding automated regressions. Shaded regions represent 5%–95% confidence intervals for nonsimultaneous bounds, and the horizontal span of the curves cover the 1st through 99th percentiles. All call density values have been transformed back from the log-transformed regression models.

datasets provided in Table I. Figures 6 shows manual locations, restricted to ranges less than 3.5 km from the closest sensor. Figure 7 uses locations up to 30 km range from the automated dataset.[5] All units are expressed in terms of dB sound pressure level (RMS SPL); results computed in terms of sound exposure level (SEL) and maximum power spectral density (mPSD) generate the same overall patterns and are not reproduced here. All figures follow the same format, with joint probabilities in the left column, probabilities conditioned on noise level in the right column, and marginal distributions of noise and source levels in the middle column.

The top rows [subplots (a) and (c)] in Figs. 6 and 7 show the observed distributions, uncorrected for masking effects, while the bottom rows [subplots (d) and (f)] display the results of applying the P_a values from the distance sampling methods presented in Sec. III E, thus correcting for noise masking effects. A comparison between the masked (top row) and unmasked (bottom row) distributions of Fig. 6 (R_{max} = 3.5 km) shows little difference in the distributions, validating the argument that using R_{max} = 3.5 km removes most masking effects.

Subplot f on the lower right of both figures shows the final estimated behavioral response distribution $p(SL|NL)$. Some of these behavioral response distributions contain SLNR values less than 52 dB, and so have not been corrected with a P_a value, generating a 45° spurious sharp cutoff in the figure [e.g., Fig. 7(f)]. Figure 7 also shows evidence of artifacts at very high source and noise levels, where small sample sizes have been inflated by very low Pa values. When these artifacts are ignored, one sees a strong Lombard effect in both the masked and unmasked data, with the mean source level increasing with background noise level.

Figure 8 displays regressions of both source level (top row) and behavioral sensitivity (ΔSL/ΔNL; middle row) for data samples when seismic airgun survey noise is present. The bottom row plots the modeled relationship between source level and airgun survey cSEL level. The modeled SPLs are shown for the masked and unmasked behavioral response distributions (left and right columns) as a function of background noise level, analysis type, and R_{max}.[5]

V. DISCUSSION

A. Relationship between source level, ambient noise, and airgun exposure

Figure 8(a) demonstrates a clear Lombard effect for all four datasets, with the mean source level rising 20 to 25 dB over a 30–40 dB increase in noise. However, the mean source level regressed from the long-range dataset (R_{max} = 30 km) is several dB greater than that of the short-range

https://doi.org/10.1121/10.0000935

JASA

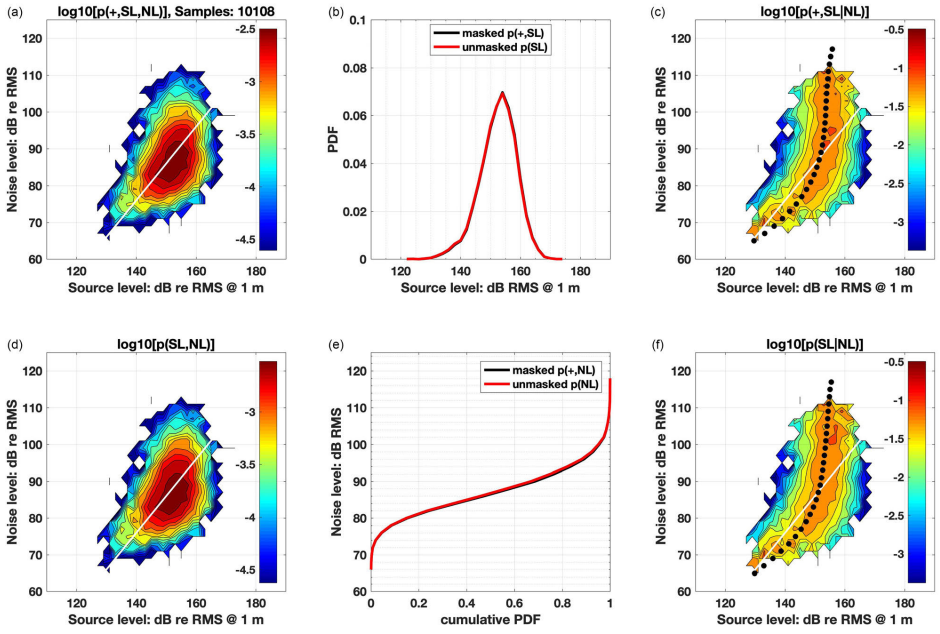


FIG. 6. (Color online) Various stages of computing the underlying conditional probability density $p(SL|NL)$ from the observed joint distribution of source level and noise level $p(+,SL,NL)$, using the "short range" manually analyzed data ($N = 10\,108$; $R_{max} = 3.5$ km) from Table I. (a) Observed joint distribution $p(+,SL,NL)$; (b) marginal distribution of observed source level [$p(+, SL)$; black] and underlying unmasked source level [$p(SL)$; red]; (c) observed source level distribution conditioned on noise level $p(+,SL|NL) = p(+,SL,NL)/p(NL)$; (d) underlying unmasked joint distribution of $p(SL,NL) = p(+,SL,NL)/P_a(SLNR)$; (e) observed (black) and underlying unmasked (red) cumulative background noise distributions, where each noise sample is computed over a unique bandwidth that matches the call; (f) underlying unmasked source level distribution conditioned on noise level [$p(SL|NL)$; Eq. (1)]. Small black circles indicate mean polynomial regression of source level with respect to noise level. All units are in terms of sound pressure level (dB re 1 𝜇Pa, RMS). Diagonal white lines illustrate points that share a constant SLNR of 64 dB (for manual analyses) and 68 dB (for automated analyses).

set. The reason behind this difference arises from masking effects on weaker calls, and not due to errors in the transmission loss function used to estimate source level. Permitting call samples to be collected from a larger region biases the samples toward louder calls: the effective sampling area for weaker calls will be smaller than that for louder calls, as weaker calls cannot be detected out to the cutoff range R_{max}. By applying the masking correction factors from Fig. 3(e) and recalculating the regression from the resampled distribution, the discrepancy between the sampling ranges becomes substantially reduced [Fig. 8(b)]. The distance sampling unmasking approach described in Sec. II B is thus validated.

The behavioral sensitivity for both the masked and unmasked regressions (Fig. 8, middle row) shows that at low in-band noise values, the short-range analyses display a nearly perfect compensation for ambient noise variations, with sensitivities just above 1. For the same noise values the corresponding long-range sensitivities are lower, lying between 0.6 and 0.9. As noise levels increase, the sensitivities for all analyses decline steadily, corresponding with a gradual shrinking of the population's communication space. While the exact details of the decline vary between the analyses, the sensitivities of the masked regressions fall to nearly zero at high noise levels between 95 and 105 dB, indicating that the population is no longer able to adjust (increase) its source level at these higher noise levels, which correspond to the 90–99th percentiles of the noise distribution [Fig. 7(e)]. Unmasking the distribution still reveals this sensitivity decrease, although at very high noise levels the sensitivities do not fall to zero, which may be a spurious artifact from low sample sizes.

By contrast, the population barely changes its source level in response to increasing airgun activity [Figs. 8(e) and 8(f)]: over a 40 dB increase of cSEL the mean source level increases by just a few dB, yielding a low behavioral sensitivity with respect to airguns. This result is not initially puzzling, since seismic airgun signals are impulsive, and one might expect that the animals would not need to raise their source levels to avoid the noise, but would simply increase their call production rate in order to transmit a call during times when the airguns are silent. However, seismic survey noise is not completely impulsive; previous work (Guerra *et al.*, 2011) has shown that continuous reverberation can exist between the airgun pulses during a seismic survey, and

JASA
https://doi.org/10.1121/10.0000935

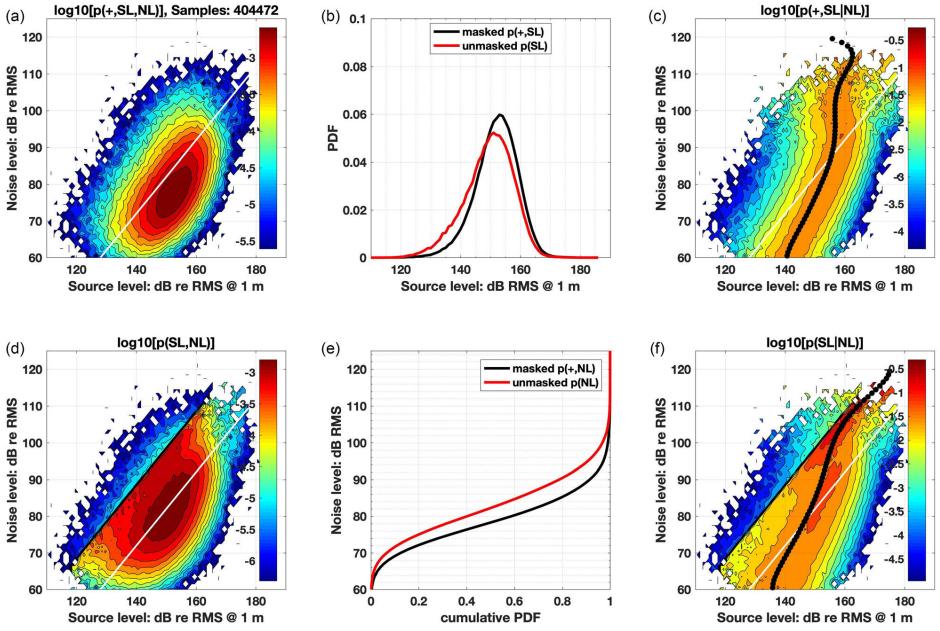


FIG. 7. (Color online) Same as Fig. 6, except displaying long-range automated data (N = 404 472; R_{max} = 30 km) from Table I.

thus one would expect the animals to raise their source levels as the diffuse reverberation levels rise. Perhaps that is the explanation for the mild increases in source level as the cSEL airgun noise level increases.

B. Implications for passive acoustic population density estimation

This work has two implications when applying passive acoustic monitoring to density estimation: first, when converting measurements of call production density into estimates of underlying animal density and abundance; and second, when trying to estimate call density over short intervals within a season, a situation where the "pooling robustness" assumption of distance sampling is violated.

1. Converting call production density into animal density

While a variety of methods exist for estimating the true (unmasked) underlying call density of animals, translating this density into an animal density is difficult because an individual's call production rate depends heavily on the animal's behavioral state as well as other environmental, ecological, and contextual factors that influence behavior (Ellison *et al.*, 2012). The work presented here lists another behavioral factor—natural and anthropogenic noise—which should be considered when estimating long-term abundance or abundance trends.

For example, to learn whether the bowhead whale population off Alaska is increasing over time, one must compare call density estimates across years in order to obtain relative trends. This work shows that in order to measure accurate population trends using passive acoustics only, two correction factors should be applied to raw counts of measured call density: a noise masking correction, and then a behavioral correction to adjust call production rates for Lombard effects. The former correction can be achieved by standard distance sampling methods, if one fitted a separate localization function for each season, but the latter correction would still need to be applied as well, because localization distance functions can only correct for masking effects, and not behavioral effects.

In principle the behavioral correction would not be required if the underlying noise distributions (natural and seismic) between seasons were the same, but in reality, seismic survey activity varies widely across years. Even mean ambient noise level percentiles varied by up to 4 dB across the study's lifetime (Thode *et al.*, 2017), which translates into changes of call density of 0.3 to 0.4 calls/min per unit area: a nearly 33% shift. Thus, call rates should be corrected for both masking *and* behavioral noise effects if passive acoustics is used to precisely estimate long-term, multi-seasonal trends.

2. Addressing breakdowns in pooling robustness

The explicit incorporation of noise levels into distance sampling theory becomes relevant in situations where call

https://doi.org/10.1121/10.0000935

JASA

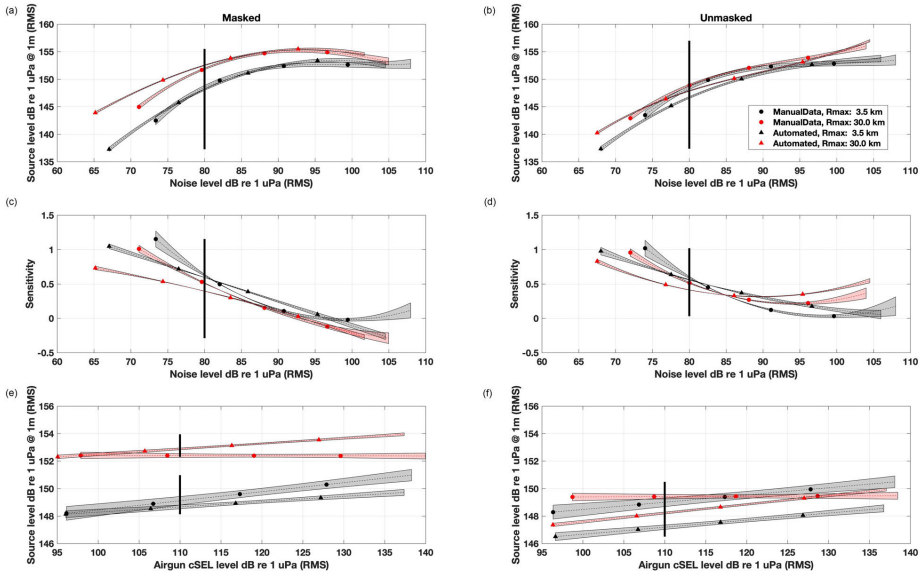


FIG. 8. (Color online) Regression analysis of source level against in-band noise level, when seismic airgun activity is present. Top row shows predicted source vs noise level for masked (a) and unmasked (b) distributions, when airgun cSEL is fixed at 110 dB re 1 uPa (black vertical line in bottom row); middle row shows behavioral sensitivities for (c) masked and (d) unmasked distributions; bottom row shows predicted source level vs airgun cSEL, when in-band noise level is fixed to 80 dB (vertical lines in top and middle rows). Manual analyses are represented by circles, automated analyses by triangles, short-range ($R_{max} = 3.5$ km) analyses by black, and long-range ($R_{max} = 30$ km) by red. Shading represents 95% confidence intervals, and horizontal extent of curves covers the 1st to 99th percentiles of measured noise level.

density is estimated over relatively short intervals, such as if one were to estimate a time series of relative animal abundance throughout a single season. Under these circumstances, the so-called "pooling robustness" property of distance sampling becomes invalid, and it becomes useful to explicitly incorporate noise levels as a covariate into the localization function estimate.

Pooling robustness is often invoked to explain why distance sampling theory generally neglects the impacts of noise on acoustic density estimation. The concept is explained by Burnham *et al.* (2004, p. 19):

"In reality, detection probability does not depend on distance only. It may depend on the ability of the surveyor, the characteristics of the individual animals, environmental or weather conditions, and a host of other factors. However, when animals at zero distance are detected with certainty [g (0) = 1], then providing that the fitted detection function model is flexible enough, distance sampling estimators of abundance and density are unbiased even though all things other than distance are ignored in estimating detection probability. This property, known as 'pooling robustness', is a very powerful feature of distance sampling methods."

Thus, noise masking effects, one of these "other factors," can often be neglected if enough data are sampled from enough circumstances to reproduce the underlying noise statistics.

Buckland *et al.* (2015), however, provide a warning about blindly applying pooling robustness (p. 55): "If a survey region is stratified into two habitats, and detectability is lower in one habitat than the other, the stratum-specific abundance estimates will again be biased, if we assume that the same detection function [will be applied] to both habitats. Total abundance across habitats will only have the pooling robustness property if effort is in proportion to stratum area. For example, if one stratum is twice the size of the other, it should have twice the survey effort."

Expressing this caution in terms of call detectability in noise, PAM distance sampling estimates that ignore noise are unbiased only if calls are sampled in a way that reflects the true underlying distribution of SLNR, which in turn relies on the true underlying distribution of noise levels. If the noise conditions associated with a set of call samples are not representative of the overall long-term noise distribution used to create the localization function, then the resulting call density estimates will be biased. If the call samples are collected under unusually quiet conditions, then the density estimates will be biased high. In particular, if one is trying to compute call densities at weekly intervals, it is risky to use a localization function computed using all samples collected across the season, because it is not guaranteed that the noise conditions experienced over one week's time are representative of noise levels (or seismic cSEL activity) captured over

JASA
https://doi.org/10.1121/10.0000935

an entire season. At the very least, when constructing a call abundance time series the temporal and spatial consistency of ambient noise statistics should be confirmed.

If noise statistics are not stationary (consistent), at least two approaches exist to rectify this situation. The first is to subdivide the call samples so that each set shares similar detectability conditions, and then assign a separate localization function to each. An example of this strategy is creating a separate localization function for every season of data. Unfortunately, this approach is often not practical when computing short-term call density estimates. There are simply not enough call samples collected over enough ranges in the course of a week to derive a weekly updated localization function.

Another approach, recommended here, is to build a localization function out of the full seasonal dataset, but use SLNR (and thus noise) as a covariate. Calls detected over a short time window can then be sorted by SLNR, and measured call densities for each SLNR value can then be calculated and adjusted by the appropriate SLNR masking factor P_a. In this manner, an accurate, short-term time series can be reconstructed. This particular approach only works if calls with different SLNR are statistically independent; e.g., no call sequences with alternating high- and low-source levels are generated.

Marques and Buckland (2004) discuss additional situations where covariates like SLNR need to be considered explicitly in distance sampling density estimation.

VI. CONCLUSION

Bowhead whales respond to differing ambient noise conditions by increasing their rate of calling and increasing their call source level (Lombard effect). Here we show that call density increases with increasing in-band continuous, natural ambient noise and in the presence of weak seismic survey activity. The effect of weak seismic survey activity is roughly similar to a 10–15 dB change in continuous noise levels. At higher exposures to seismic activity noise, individual call rates become suppressed, and the measured call density decreases. As the 10-min exposure reaches 115 dB re 1 μPa^2-s cSEL, call density returns to baseline (no airgun) levels, and densities are further suppressed at higher levels.

Distance sampling, using the SLNR as a covariate, was successfully applied to address masking effects. This technique may be viable to other situations where one can assume a uniform or other *a priori* animal distribution in azimuthally symmetric propagation conditions. Both the raw and unmasked behavioral response distributions show that calling bowhead whales display a strong Lombard effect by increasing their call source levels in the presence of in-band, continuous noise, producing a nearly 20 dB change in mean source level between the 1st and 99th noise percentiles. At low noise levels individual whales can adjust their mean source level to completely compensate for ambient noise level changes, but they steadily loses their ability to adjust as noise levels increase, until most calling individuals in a population can no longer compensate for increasing noise levels. We postulate that this apparent reduction in or loss of behavioral sensitivity at high noise exposure levels arises from physiological limits to sound production and does not necessarily represent a loss of behavioral sensitivity. By contrast, the population's mean call source level increases by just a few dB when seismic survey acoustic conditions increase noise cSEL conditions by 40 dB. This apparent insensitivity may arise from the impulsive nature of this noise, which might allow whales to communicate using their baseline source level during the times between airgun pulses, provided that they call more frequently. Increases of reverberation levels with increasing cSEL may explain why a weak source level response does exist to seismic airgun noise.

These results illustrate the importance of using behavioral responses to natural noise fluctuations to place anthropogenic responses in context, and may provide insight into how to translate call density estimates into animal densities. The modified distance sampling technique derived here may also have applications in correcting measured call density estimates collected over relatively short timescales.

Both humans and animals display other responses to ambient noise shifts, including changes in the temporal and spectral structure of signals (Brumm and Zollinger, 2011). For example, Parks *et al.* (2012) has identified shifts in call minimum frequency in right whale calls in response to increasing low frequency noise. This dataset is ripe for further such investigations, including incorporating long-term changes in call spectral structure into the analysis (e.g., Thode *et al.*, 2017).

ACKNOWLEDGMENTS

The authors wish to thank SEPCO for permission to use this dataset, as well as the crew of the *Norseman II*, *Alpha Helix*, and *Westward Wind*, particularly deck chief Scotty Hameister, for consistently safe and professional deployments and retrievals of the DASARs over the project life. A. Michael Macrander provided strong leadership and guidance to the project, while Sheyna Wisdom and her colleagues at Fairweather LLC provided high-quality fieldwork logistic support. Bob Norman of Greeneridge provided helpful support on DASAR hardware and recording details. Finally, none of this work would have existed if Charles R. Greene, Jr. had not had the vision to deploy DASARs off the Beaufort coast nearly two decades ago and to recruit all the authors to the project. Analyses of these data were supported by the North Pacific Research Board (Project 1610) and the Sound and Marine Life Joint Industry Program (JIP Grant No. JIP22 III-15-14) of the Oil and Gas Producers Association (OGP).

APPENDIX: DISTANCE SAMPLING AVAILABILITY FUNCTIONS FOR DISTRIBUTED PAM ARRAYS

The distance sampling Eqs. (1)–(4) are derived assuming that the distance of a detected object can be determined using measurements from a single point. Thus, the standard

https://doi.org/10.1121/10.0000935
JASA

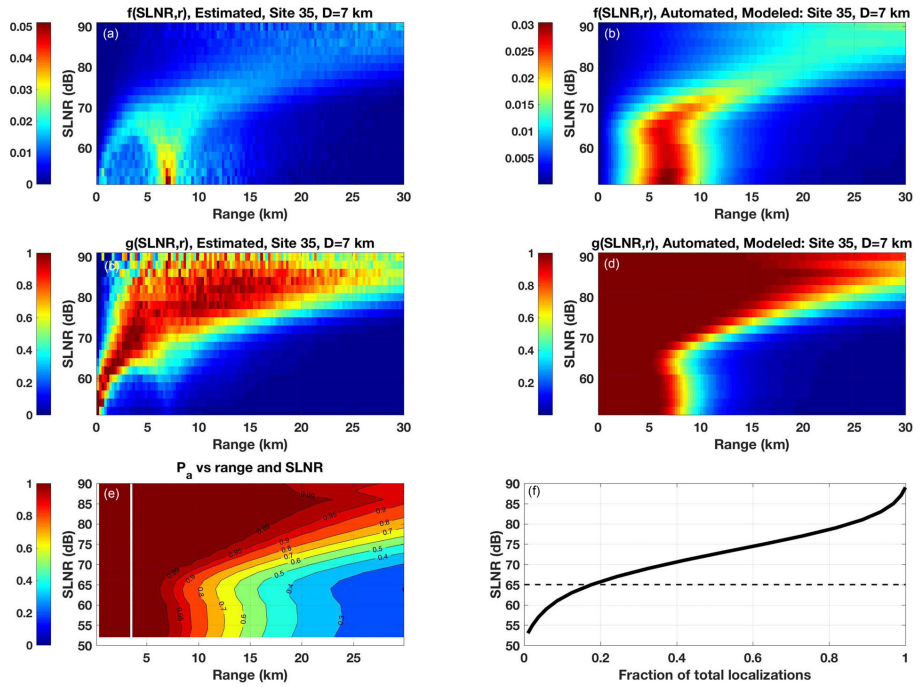


FIG. 9. (Color online) Same as Fig. 3, except all ranges to a localization (not just closest range) have been incorporated into the distance analysis, using a standard circular availability function.

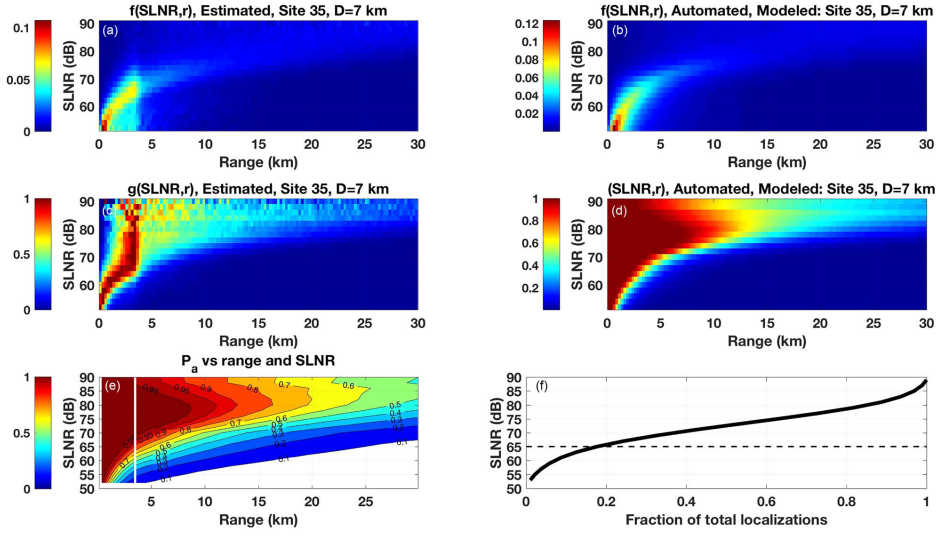


FIG. 10. (Color online) Figure 9 recomputed using only the closest range for each localization, using the standard circular availability function.

JASA

https://doi.org/10.1121/10.0000935

form of the availability function, π(r), is a circle normalized by the total monitoring area A: $2\pi r/A$ (Marques *et al.*, 2013b; Buckland *et al.*, 2012). However, most methods for localizing underwater sounds require detecting the same signal on multiple sensors, using either relative arrival time differences on hydrophones or triangulation on directional sensors to achieve the localization. As a result, a given localization yields multiple ranges to different sensors, raising the question as to whether all ranges measured from all sensors should be treated as independent samples when computing the maximum-likelihood fit for the localization function, or whether only the closest range to each position should be used. In this appendix we demonstrate that the latter choice yields an answer that is clean and intellectually consistent for this study, but requires a more complex availability function than a simple circle.

Two issues were found when using all ranges to a localization: first, their use artificially inflates the scale parameter in the localization function [σ in Eq. (3)], since using all ranges to a localization, and not just the closest, biases the observed distribution toward larger ranges. For example, if a call is detected at three DASARs at 300, 1000, and 2000 m range, then the mean range of the three samples (1100 m) will be larger than using the minimum range (300 m). A more serious issue can be seen in Fig. 9, which displays the localization function results in a format similar to that of Fig. 3. One sees that for SLNR values below 70 dB the observed range distribution $f(SLNR,r)$ in subplot 9(a) (top left plot) bifurcates into a bimodal distribution, with the most prominent peak emerging at 7 km, i.e., the separation between DASARs. Our interpretation of this result is that it is a quirk of the triangulation algorithm that arises when an analyst tries to estimate a bearing for a weak (low SNR) call. The bearing uncertainty for weak calls rises substantially, and when applying the maximum-likelihood robust triangulation scheme of Lenth (1981) we observe a tendency of the algorithm to "cluster" calls with high azimuthal uncertainty around the closest DASAR, when then generates a set of distances close to 7 km from the other DASARs. Numerical simulations found that this effect only occurs when a call is generated within a range $\delta D\phi$ of the closest DASAR, where D is the sensor separation (7 km), and $\delta\phi$ is

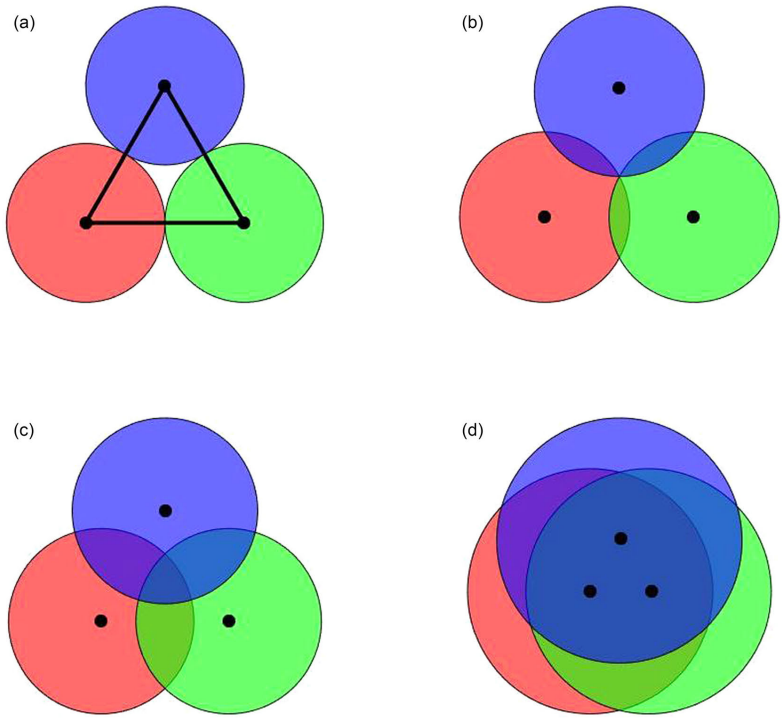


FIG. 11. (Color online) Illustration of how the availability function changes for a three-sensor grid. The availability function is all points on the circles' perimeter that do not penetrate other circles. (a) $r < D/2 : \pi(r) = 6\pi r$, (b) $D/2 < r < D/\sqrt{3} : \pi(r) = 12\alpha r$, (c) $D/\sqrt{3} < r : \pi(r) = (6\alpha + 2\pi)r$, (d) $D/\sqrt{3} \ll r : \pi(r) = 3D + 2\pi r$. $\alpha = \sin^{-1}(\mathrm{D}/2\mathrm{R})$. For (d) one sees that at large ranges the availability function for the distributed array converges toward the standard availability function of a circle.

https://doi.org/10.1121/10.0000935

JASA

the azimuthal uncertainty in radians. At weak signal-to-noise ratios (below 5 dB) the uncertainty can reach up to 10°, so the clustering effect occurs for weak calls 1 km or closer to the nearest DASAR.

The number of calls affected by this situation is small (20% of the sample), so using all ranges to construct a localization function can be viable if one were simply trying to build a general localization function without using SLNR as a covariate, and had an ability to measure the source level of a call (so low SLNR calls could be rejected). However, given the desire to generate an accurate image of the behavioral response $p(SL|NL)$, this 7 km clustering in Fig. 9(a) was unacceptable, and we are forced to fit a localization curve using only the range to the closest DASAR that contributed to a given localization.

However, using only the closest range to a call generates a new problem, as revealed for the resulting estimates of $f(SLNR,r)$ and $g(SLNR,r)$, plotted in Figs. 10(a) and 10(c), which show strong discontinuities in the localization function at 3.5 km for any SLNR above 68 dB, e.g., Fig. 10(c).

The reason for this artifact is that when only the range to the closest DASAR (in a triangulating grid of DASARs) is used, the availability function is no longer a circle, but becomes a more complex locus of points whose structure changes with increasing range. Figure 11 illustrates this point for a simple three-sensor triangular grid, with sensor separation D. When calls are generated at a range $r < D/2$ (upper left, a), then the loci of points that satisfy this condition define three nonoverlapping circles with cumulative perimeter of $6\pi r$, and thus the perimeter function is proportional to this value.

As r increases past $D/2$, the circles of radius r overlap, and only points outside the perimeter of any other circle satisfy the condition of r being the closest range to a DASAR. As r grows much larger than $D/2$, the circles nearly completely overlap and the perimeter converges toward the value of a single circle with circumference $2\pi r$, shown in Fig. 11(d).

Thus, for a three-sensor localization grid as in Fig. 11, the availability function $\pi(r)$ can be shown to be:

$$\begin{aligned} &r < \frac{D}{2} : \pi(r) = 6\pi r; \quad \frac{D}{2} < r < \frac{D}{\sqrt{3}} : 12\alpha r; \\ &\frac{D}{\sqrt{3}} < r : \pi(r) = (6\alpha + 2\pi) r; \\ &\frac{D}{\sqrt{3}} \ll r : \pi(r) = 3D + 2\pi r, \end{aligned} \tag{A1}$$

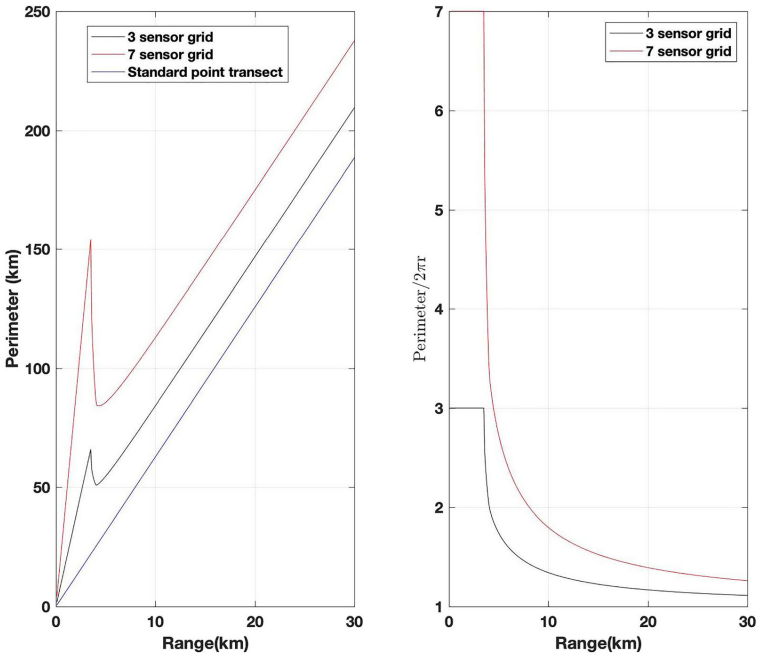


FIG. 12. (Color online) (a) Availability function for a three-sensor array with 7 km separation (black), seven-sensor array with 7 km separation (red), and standard point transect (blue). (b) Availability function for three-sensor (black) and seven-sensor (red) distributed array, normalized by standard point transect availability function.

JASA

https://doi.org/10.1121/10.0000935

where $\alpha = \sin^{-1}(D/2R)$. For seven sensors in a triangular grid, which represents the actual Site 3 and 5 geometry, the availability function becomes

$$\begin{aligned} r < \frac{D}{2} &: \pi(r) = 14\pi r, \\ \frac{D}{2} < r < \frac{D}{\sqrt{3}} &: \pi(r) = (44\alpha - 8\pi)r, \\ \frac{D}{\sqrt{3}} < r &: \pi(r) = (14\alpha + 2\pi)r, \\ \frac{D}{\sqrt{3}} \ll\sim r &: \pi(r) = 7D + 2\pi r. \end{aligned} \tag{A2}$$

Figure 12 plots the three- and seven-DASAR availability functions vs range for $D = 7$ km. The right subplot shows the availability function divided by the standard distance function $2\pi r$ and clearly reveals the availability functions' discontinuous slope at D/2 (3.5 km) range. At larger ranges the ratio in subplot b rapidly decreases and then asymptotically approaches 1, thus converging into the standard availability function when the range is much larger than the dimensions of the distributed array.

The discontinuity at 3.5 km is what introduces the vertical artifact visible in Figs. 10(a) and 10(c). This is an important point because if *point transect theory is applied to a distributed tracking array with spacing D using only one range (closest range) per localized call, then the resulting localization function will appear to fall off sharply at a range equal to D/2*. This artifact could easily be interpreted as a true localization function when, in reality, it arises from using an incorrect availability function $\pi(r)$.

Equation (A2), combined with Eq. (4), yields Fig. (3). While standard distance sampling software packages, e.g., the DISTANCE package (Thomas *et al.*, 2010) developed for the statistical software language R, do not allow the specification of arbitrary availability functions, one can select the range binning option and then duplicate samples at each range according to the normalized availability function (e.g., right side of Fig. 12) to effectively apply any availability function desired. Other approaches for modeling nonstandard availability functions are provided in Marques *et al.* (2010, 2013a) for point and linear transects, respectively.

[1]Technically this term should be called "call intensity" or "call flux," as it describes a quantity that is a rate per unit area, but we opted for "call density" to avoid confusion with acoustic intensity and to be consistent with other animal density estimation literature.

[2]We emphasize at the outset that our definition of p("+") and other upcoming quantities includes the probability of not only detection, but also of *localization*, a process that requires a call to be detected on additional sensors beyond just the closest sensor (which is the origin for R_{max}).

[3]It is unfortunate that the distance sampling literature uses the Greek symbol π for both the availability function and the famous irrational number, but we chose to respect that [irrational] convention here by always writing the availability function with an explicit dependence on r: $\pi(r)$.

[4]At first glance, this relationship implies an effective transmission loss of $\sim 20\log_{10}(R)$, which is considerably harsher attenuation than the $15\log_{10}(R)$ used to model the source levels. Note, however, that Fig. 4 shows parameters from a *localization*, and not a true *detection*, function, and that at least two detections on at least two DASARs are required for a localization. A simple simulation demonstrates that if a call is equidistant from two sensors, and if the detection probabilities at both locations are statistically independent, then the probability of localization is equal to the square of the probability of detection. If a signal's detectability vs SNR is modeled as a sigmoid, one finds that a $15\log_{10}(R)$ propagation model will generate a localization function that displays an 18 to $20\log_{10}(R)$ relationship between the scale parameter and SLNR, as shown.

[5]See supplementary material at https://doi.org/10.1121/10.0000935 for contour plots of the probability distributions of call density vs in-band ambient SPL noise level; for displays of the joint and conditional probabilities for other combinations of R_{max} and analysis method; and for displays of regressions of both source level and behavioral sensitivity for data samples collected in the absence of seismic survey noise.

[6]As stated in Sec. II B, the noise levels displayed here are computed from bandwidths that vary from call to call. A regression of broadband noise levels against these "floating bandwidth" levels found that adding 12 dB to the noise scales shown here makes them approximately equal to the RMS sound pressure level (SPL) measured over a fixed bandwidth of 20–170 Hz.

Blackwell, S. B., Nations, C. S., McDonald, T. L., Greene, C. R., Jr., Thode, A. M., Guerra, M., and Macrander, A. M. (**2013**). "Effects of airgun sounds on bowhead whale calling rates in the Alaskan Beaufort Sea," Mar. Mammal Sci. **29**, E342–E365.

Blackwell, S. B., Nations, C. S., McDonald, T. L., Thode, A. M., Mathias, D., Kim, K. H., Greene, C. R., Jr., and Macrander, A. M. (**2015**). "The effects of airgun sounds on bowhead whale calling rates: Evidence for two behavioral thresholds," PLoS One **10**, e0125720.

Blackwell, S. B., Nations, C. S., Thode, A. M., Kauffman, M. E., Conrad, A. S., Norman, R. G., and Kim, K. H. (**2017**). "Effects of tones associated with drilling activities on bowhead whale calling rates," PLoS One **12**, e0188459.

Blackwell, S. B., Richardson, W. J., Greene, C. R., Jr., and Streever, B. (**2007**). "Bowhead whale (*Balaena mysticetus*) migration and calling behaviour in the Alaskan Beaufort sea, Autumn 2001–04: An acoustic localization study," Arctic **60**, 255–270, available at: https://www.jstor.org/stable/40512894.

Buckland, S. T. (**1992**). "Fitting density functions with polynomials," J. R. Stat. Soc.: Ser. C (Appl. Stat.) **41**, 63–76.

Buckland, S. T., Anderson, D. R., Burnham, K. P., and Laake, J. L. (**2012**). *Distance Sampling: Estimating Abundance of Biological Populations* (Springer Science & Business Media, New York).

Buckland, S. T., Rexstad, E. A., Marques, T. A., and Oedekoven, C. S. (**2015**). *Distance Sampling: Methods and Applications* (Springer Science & Business Media, New York).

Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer Associates, Sunderland, MA).

Brumm, H., and Zollinger, S. A. (**2011**). "The evolution of the Lombard effect: 100 years of psychoacoustic research," Behaviour **148**, 1173–1198.

K. P. Burnham, S. T. Buckland, J. L. Laake, D. L. Borchers, T. A. Marques, J. R. B. Bishop, and L. Thomas (**2004**). "Further topics in distance sampling," in *Advanced Distance Sampling*, edited by S. T. Buckland (Oxford University Press, New York).

Castellote, M., Clark, C. W., and Lammers, M. O. (**2012**). "Acoustic and behavioural changes by fin whales (*Balaenoptera physalus*) in response to shipping and airgun noise," Biol. Conserv. **147**, 115–122.

Clark, C. W., Ellison, W. T., Southall, B. L., Hatch, L., Van Parijs, S. M., Frankel, A., and Ponirakis, D. (**2009**). "Acoustic masking in marine ecosystems: Intuitions, analysis, and implication," Mar. Ecol.:Prog. Ser. **395**, 201–222.

Clark, C. W., and Johnson, J. H. (**1984**). "The sounds of the bowhead whale, *Balaena mysticetus*, during the spring migrations of 1979 and 1980," Can. J. Zool. **62**, 1436–1441.

Cummings, W. C., and Holliday, D. V. (**1987**). "Sounds and source levels from bowhead whales off Pt. Barrow, Alaska," J. Acoust. Soc. Am. **82**, 814–821.

Delarue, J., Laurinolli, M., and Martin, B. (**2009**). "Bowhead whale (*Balaena mysticetus*) songs in the Chukchi Sea between October 2007 and May 2008," J. Acoust. Soc. Am. **126**, 3319–3328.

https://doi.org/10.1121/10.0000935

JASA

Di Iorio, L., and Clark, C. W. (**2010**). "Exposure to seismic survey alters blue whale acoustic communication," Biol. Lett. **6**, 51–54.

Dobson, A. J., and Barnett, A. (**2008**). *An Introduction to Generalized Linear Models* (CRC, New York).

Dunlop, R. A., Cato, D. H., and Noad, M. J. (**2014**). "Evidence of a Lombard response in migrating humpback whales (*Megaptera novaeangliae*)," J. Acoust. Soc. Am. **136**, 430–437.

Ellison, W., Southall, B., Clark, C., and Frankel, A. (**2012**). "A new context-based approach to assess marine mammal behavioral responses to anthropogenic sounds," Conserv. Biol. **26**, 21–28.

Greene, C. R., McLennan, M. W., Norman, R. G., McDonald, T. L., Jakubczak, R. S., and Richardson, W. J. (**2004**). "Directional frequency and recording (DIFAR) sensors in seafloor recorders to locate calling bowhead whales during their fall migration," J. Acoust. Soc. Am. **116**, 799–813.

Guerra, M., Thode, A., Blackwell, S. B., and Macrander, M. A. (**2011**). "Quantifying seismic survey reverberation off the Alaskan North Slope," J. Acoust. Soc. Am. **130**, 3046–3058.

Helble, T. A., D'Spain, G. L., Hildebrand, J. A., Campbell, G. S., Campbell, R. L., and Heaney, K. D. (**2013**). "Site specific probability of passive acoustic detection of humpback whale calls from single fixed hydrophones," J. Acoust. Soc. Am. **134**, 2556–2570.

Holt, M. M., Noren, D. P., and Emmons, C. K. (**2011**). "Effects of noise levels and call types on the source levels of killer whale calls," J. Acoust. Soc. Am. **130**, 3100–3106.

Holt, M. M., Noren, D. P., Veirs, V., Emmons, C. K., and Veirs, S. (**2009**). "Speaking up: Killer whales (*Orcinus orca*) increase their call amplitude in response to vessel noise," J. Acoust. Soc. Am. **125**, EL27–EL32.

Hotchkin, C., and Parks, S. (**2013**). "The Lombard effect and other noise-induced vocal modifications: Insight from mammalian communication systems," Biol. Rev. **88**, 809–824.

Küsel, E. T., Mellinger, D. K., Thomas, L., Marques, T. A., Moretti, D., and Ward, J. (**2011**). "Cetacean population density estimation from single fixed sensors using passive acoustics," J. Acoust. Soc. Am. **129**, 3610–3622.

Lenth, R. V. (**1981**). "On finding the source of a signal," Technometrics **23**, 149–154.

Ljungblad, D. K., Thompson, P. O., and Moore, S. E. (**1982**). "Underwater sounds recorded from migrating bowhead whales, *Balaena mysticetus*, in 1979," J. Acoust. Soc. Am. **71**, 477–482.

Lombard, E. (**1911**). "The signs of increasing the voice," Ann. Malad. l'Oreille Larynx **37**, 101–119.

Marques, F. F. C., and Buckland, S. T. (**2004**). "Covariate models for the detection function," in *Advanced Distance Sampling*, edited by S. T. Buckland (Oxford University Press, New York).

Marques, T., Buckland, S., Borchers, D., Tosh, D., and McDonald, R. (**2010**). "Point transect sampling along linear features," Biometrics **66**, 1247–1255.

Marques, T. A., Buckland, S. T., Bispo, R., and Howland, B. (**2013a**). "Accounting for animal density gradients using independent information in distance sampling surveys," Stat. Methods Appl. **22**, 67–80.

Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Moretti, D. J., Harris, D., and Tyack, P. L. (**2013b**). "Estimating animal population density using passive acoustics," Biol. Rev. **88**, 287–309.

Melcon, M. L., Cummins, A. J., Kerosky, S. M., Roche, L. K., Wiggins, S. M., and Hildebrand, J. A. (**2012**). "Blue whales respond to anthropogenic noise," PLoS One **7**, e32681.

Moore, S. E., and Reeves, R. R. (**1993**). "Distribution and movement," in *The Bowhead Whale*, edited by J. Burns, J. Montague, and C. Cowles (Allen Press, Lawrence, Kansas), pp. 313–386.

Moore, S. E., Stafford, K. M., Mellinger, D. K., and Hildebrand, J. A. (**2006**). "Listening for large whales in the offshore waters of Alaska," Bioscience **56**, 49–55.

Moyer-Horner, L., Smith, M. M., and Belt, J. (**2012**). "Citizen science and observer variability during American pika surveys," J. Wildlife Manage. **76**, 1472–1479.

Parks, S. E., Groch, K., Flores, P., Sousa-Lima, R., and Urazghildiiev, I. R. (**2016**). "Humans, fish, and whales: How right whales modify calling behavior in response to shifting background noise conditions," in *Effects of Noise on Aquatic Life II*, edited by A. N. Popper, and A. Hawkins (Springer, New York), pp. 809–813.

Parks, S. E., Johnson, M., Nowacek, D., and Tyack, P. L. (**2011**). "Individual right whales call louder in increased environmental noise," Biol. Lett. **7**, 33–35.

Parks, S. E., Johnson, M. P., Nowacek, D. P., and Tyack, P. L. (**2012**). "Changes in vocal behavior of North Atlantic right whales in increased noise," in *Effects of Noise on Aquatic Life*, edited by A. N. Popper, and A. Hawkins (Springer, New York), pp. 317–320.

Risch, D., Corkeron, P. J., Ellison, W. T., and Van Parijs, S. M. (**2012**). "Changes in humpback whale song occurrence in response to an acoustic source 200 km away," PLoS One **7**, e29741.

Stafford, K. M., Moore, S. E., Laidre, K. L., and Heide-Jørgensen, M. P. (**2008**). "Bowhead whale springtime song off West Greenland," J. Acoust. Soc. Am. **124**, 3315–3323.

Tervo, O. M., Parks, S. E., Christoffersen, M. F., Miller, L. A., and Kristensen, R. M. (**2011**). "Annual changes in the winter song of bowhead whales (*Balaena mysticetus*) in Disko Bay, Western Greenland," Mar. Mammal Sci. **27**, E241–E252.

Tervo, O. M., Parks, S. E., and Miller, L. A. (**2009**). "Seasonal changes in the vocal behavior of bowhead whales (*Balaena mysticetus*) in Disko Bay, Western-Greenland," J. Acoust. Soc. Am. **126**, 1570–1580.

Thode, A. M., Blackwell, S. B., Conrad, A. S., Kim, K. H., and Michael Macrander, A. (**2017**). "Decadal-scale frequency shift of migrating bowhead whale calls in the shallow Beaufort Sea," J. Acoust. Soc. Am. **142**, 1482–1502.

Thode, A. M., Blackwell, S. B., Seger, K. D., Conrad, A. S., Kim, K. H., and Macrander, A. M. (**2016**). "Source level and calling depth distributions of migrating bowhead whale calls in the shallow Beaufort Sea," J. Acoust. Soc. Am. **140**, 4288–4297.

Thode, A. M., Kim, K. H., Blackwell, S. B., Greene, C. R., and Macrander, M. A. (**2012**). "Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys," J. Acoust. Soc. Am. **131**, 3726–3747.

Thode, A. M., Kim, K., Greene, C. R., and Roth, E. H. (**2010**). "Long range transmission loss of broadband seismic pulses in the Arctic under ice-free conditions," J. Acoust. Soc. Am. **128**, EL181–EL187.

Thomas, L., Buckland, S. T., Rexstad, E. A., Laake, J. L., Strindberg, S., Hedley, S. L., Bishop, J. R., Marques, T. A., and Burnham, K. P. (**2010**). "Distance software: Design and analysis of distance sampling surveys for estimating population size," J. Appl. Ecol. **47**, 5–14.

Urazghildiiev, I. R., and Clark, C. W. (**2007**). "Detection performances of experienced human operators compared to a likelihood ratio based detector," J. Acoust. Soc. Am. **122**, 200–204.

4. Relative Abundance: Multi-year Trends

Estimating relative abundance of bowhead whale activity between two locations and across multiple years

by

Aaron M. Thode, Susanna B. Blackwell, Alexander S. Conrad, Katherine H. Kim

---

Manuscript in Preparation

1

Estimating relative abundance of bowhead whale calls between two locations and across multiple years

Aaron M. Thode[1], Susanna B. Blackwell[2], Alexander S. Conrad[2], Katherine H. Kim[2]

1: athode@ucsd.edu, Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238.

2: Greeneridge Sciences, Inc., 5266 Hollister Ave., Ste. 107, Santa Barbara, California 93111, U.S.A.

Manuscript submission date: TBD

Running title: bowhead whale relative abundance

2

ABSTRACT

Seven years of passive acoustic data from the Beaufort Sea were used to estimate the cumulative number of calls detected each season across two sites spaced 132 km apart along the bowhead whale fall migration corridor, to determine whether both sites generate consistent multi-year trends in call abundance (cumulative call counts). Raw call localizations, generated using automated methods, were adjusted three ways. First, call counts were restricted to those generated within 3.5 km of an acoustic sensor, in order to remove noise masking effects. Second, call densities were adjusted for population-scale behavioral effects that include diel responses, geographic latitude variations, and responses to changes in ambient noise levels and seismic airgun survey activity. Finally, multiple sensor malfunctions during two seasons required additional adjustments. Both sites indicated a multiyear decrease in cumulative calls between 2008 and 2011, followed by a steady increase in cumulative call rates between 2011 and 2014. Adjusting for behavioral effects was only significant during 2009, when the migration was highly clumped and call densities very high. This analysis suggests that passive acoustic monitoring can generate consistent long-term call abundance estimates at two different sites, but it remains inconclusive whether localization restrictions and behavioral adjustments are needed in order to do so.

3

I. INTRODUCTION

Challenges in passive acoustic marine mammal abundance estimation arise from both variations in call detectability and variations in animal behavior, which can be difficult to separate (Marques *et al.* 2013). Several approaches exist for handling call detectability. One straightforward approach is to restrict samples to those generated at sufficiently close distances to sensors, which allows call detection rates to be interpreted as call production rates (Blackwell *et al.,* 2013, 2015; Thode *et al.* in review).

Adjusting raw call counts for behavioral variations is more challenging, and requires prior estimates of "cue rates" of individual (Marques et al. 2013, Warren et al. 2017, Harris et al. 2018). The cue rate, or the mean number of calls generated per individual per unit time, depends on a large number of factors, including behavioral state, environmental factors such as ambient noise levels and time-of-day, and demographic factors like sex and age (Marques *et al.*, 2013). Cue rate measurements are typically conducted using acoustic tag studies (Baumgartner *et al.,* 2015; Simon *et al.* 2009; Johnson *et al.*, 2014), which suffer from large logistical burdens, limited sample sizes, short observation times, and sampling biases with regards to particular demographic groups and behavioral states. Passive acoustic data from fixed recorders can be combined with independent data on population sizes and distribution to place bounds on cue rates (Blackwell *et al.* in prep), but passive acoustic methods alone cannot determine cue rates or absolute abundance. They can measure *call abundance* (i.e. cumulative counts of call localizations measured over a fixed time interval), which could still provide

insight into relative long-term temporal trends and spatial distributions within a population.

Unfortunately, even this more limited goal faces challenges similar to absolute abundance estimation in terms of variations in call detectability (masking effects) and behavioral state. In certain circumstances one can assume the observed population is in the same general behavioral state (e.g. migration), or that the relative proportions of time the population spends in different behavioral states remains steady over time. But even with these assumptions call production rates can still vary in response to changes in various environmental factors, including diel effects (Goold, 2000; Baumgartner and Fratantoni, 2008; Blackwell et al. 2007), and the presence of both natural and anthropogenic noise sources (Blackwell *et al.,* 2015; Thode *et al.,* in review). Fortunately, passive acoustic measurements can provide insight into these potential environmental responses, without the need for additional external data, provided that the population can be assumed to be under the same general behavioral state, and that the demographic composition of the population doesn't change both within and between seasons. However, to date little work exists on how to use passive acoustics to correct raw call abundance measurements for behavioral variations in response to environmental factors, and whether these corrections would have a significant impact on call abundance.

Here we derive relative abundance estimates of migrating bowhead whales over seven years, using a large acoustic dataset collected between 2008 and 2014 along the continental shelf of the Alaskan North Slope, in order to document the westward fall migration of bowhead whales (Blackwell *et al.* 2011). Our analysis is simplified by

assuming that the population remains in a single behavioral state—migration—during the August-September field season of each sampled year. The data are collected on clusters, or "sites," of special directional acoustic sensors that have the ability to both detect and localize calls, whenever calls are produced at sufficiently close range to the site. Previous work on this dataset has determined how to avoid masking effects using so-called "direct census" methods, and has also derived a statistical regression model that predicts how population-scale call production rates shift in response to two types of acoustic noise: natural wind-driven and seismic airgun survey (Thode *et al.*, in review). Here we expand the statistical analysis to include year as a categorical variable, and to incorporate diel effects, distance from the shoreline, and malfunctioning sensors. This behavioral model then provides weighting, or "behavioral adjustment," factors that are applied to raw call counts in order to compensate for potential differences in these environmental factors between sites.

We evaluate the resulting cumulative call abundances by comparing them between two sites 132 km apart, before and after compensating for environmental factors and call masking. We also compare the multiyear trends in call abundance between the sites. We hypothesize that as the whale migration passes through both sites, the cumulative call abundances and multi-year trends should be similar between the sites, and that any compensating adjustments should improve the match further.

Section II reviews the location, equipment, and deployments used to collect the data, the automated detection and localization methods used to obtain cumulative call abundance, and the statistical regression and factor adjustment procedures. Section III

displays examples of the predictive model and adjustment factors, before reviewing cumulative call counts between two sites. Section IV discusses the relative importance of call masking, environmental factor adjustment, complications arising from defective sensors, and why one particular year-2009-was especially sensitive to environmental factor adjustements.

II. METHODS

A. Equipment and deployment configuration

Each year from 2007 through 2014, the Shell Exploration and Production Company (SEPCO) commissioned Greeneridge Sciences, Inc., to deploy at least 35 Directional Autonomous Seafloor Acoustic Recorders (DASARs, model C) (Greene *et al.*, 2004), divided among five sites in the coastal Beaufort Sea. The motivation behind the effort was to evaluate the potential impact of airgun and other industrial sounds on bowhead whale behavior during their westward fall migration in the relatively shallow Arctic waters off Alaska (Blackwell *et al.*, 2013, 2015, 2017). Over that entire period, nearly 2.3 million bowhead whale calls were localized during the fall migrations.

DASARs are autonomous recording packages equipped with an omnidirectional acoustic pressure sensor (sensitivity of -149 dB re V/ 1 μPa) and two horizontal directional sensors capable of measuring the north-south and east-west components of acoustic particle velocity. This arrangement permits the azimuth of received sounds, such as bowhead whale calls, to be measured from individual DASARs. Each time series is sampled at 1 kHz with a maximum usable acoustic frequency of 450 Hz due to

antialiasing filter roll-off. Coincident bearings to calls detected on different DASARs are combined via triangulation to yield two-dimensional call positions, from which the range of each call to every DASAR can be estimated (Greene *et al.*, 2004). This ability to measure bearing from a single point allows a location to be estimated using only 2 DASARs (instead of 3-4 time-aligned nondirectional sensors). Occasionally, the directional sensor on a DASAR malfunctions while the omnidirectional sensor keeps working, so while the package can aid other DASARs in detecting a nearby call, it cannot assist in localizing it. Instruments that have only a functional omnidirectional sensor will be labeled as "ASARs" in the rest of the text, as they have an important role to play in our analysis.

From August to October, 2008 to 2014, multiple DASARs were deployed across a 280 km swath off the Alaskan North Slope, on the continental shelf in water depths between 20 and 53 m. The deployments were grouped into "Sites", labeled 1–5 traveling from west to east (Blackwell et al. 2015). In this paper we focus only on Sites 3 and 5, which consisted of seven DASARs each, deployed in a triangular grid with 7 km separation and labeled 'A' to 'G' from south to north (Figure 1). These sites had identical layouts and thus presumably similar localization footprints. They were also sufficiently distant from each other (132 km) to provide a reasonable basis for independent comparison. Cumulative call abundance at sites 3 and 5 were computed between 24 August and 28 September of each season, as those dates were the ones during which both sites were operational across all seasons.

8

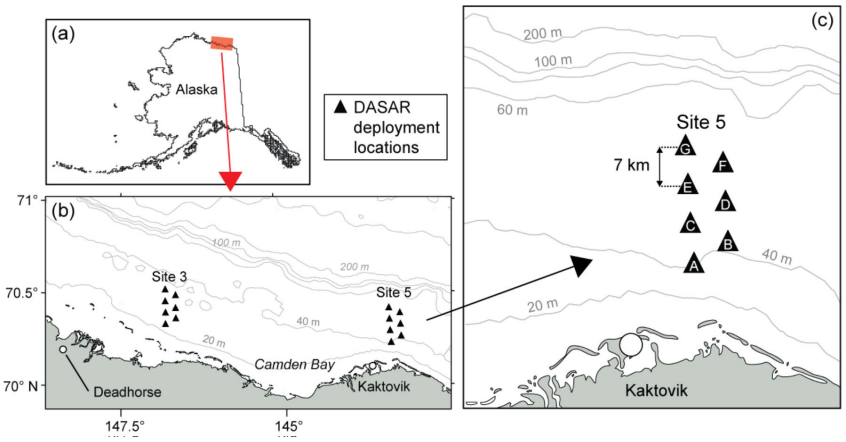


Figure 1: Location and configuration of DASAR sensors. Cumulative call abundance was computed at Sites 3 and 5. Details of the latter site are shown in subplot (c), along with the naming convention for individual DASARs at a site.

Bowhead whale calls in the raw acoustic data were post-processed two ways: by a team of human analysts, and by a six-stage automated detection and localization program. Both approaches have been extensively described and evaluated in other publications (Thode *et al.*, 2012, 2016, 2017), but only the automated analysis is used here, because it was the only approach that processed all deployment days and thus permitted cumulative counts for an entire season. Each detected call event on every DASAR was assigned a start time, duration, frequency bandwidth, and range. Call events matched between DASARs yielded both a 2-D location estimate and positional uncertainty. Malfunctioning DASARs that could not localize ("ASARs") could still participate in matching calls between two fully functioning DASARs, by using the omnidirectional sensor. Allowing this participation allowed more robust linkages between distant DASARs for improved localization. During a few deployments multiple DASARs failed at a single site, with the worst situation arising in 2013, when DASARs

A, B, and E lost directional ability at Site 3, and DASARs C, D, and E lost directional ability at Site 5 (Table I). The potential seriousness of this situation prompted the inclusion of a DASAR 'state' in the statistical regression model described below. We also explored the effect of removing call localizations associated with ASARs before fitting the statistical model.

Table I: List of DASARS with localization or complete malfunctions. Regular text: ASAR deployment. ***Bold Italic*:** Complete malfunction or DASAR loss.

Year	***Site 3***	***Site 5***
2008	E	
2009	G	
2010	***A[1], F***	***F***
2011	B, G	
2012		
2013	A, B, E	C, D, E
2014		F

[1] DASAR not deployed

B. Removing masking effects

As discussed in the Introduction, both masking and behavioral response effects may bias call abundance estimates. In order to remove masking as a factor, two approaches were examined.

The first approach was to only incorporate a particular call into a call abundance count if it occurred less than a distance R_{max} from the closest DASAR at a site. Previous work (Thode *et al.*, in review) found that when R_{max} is set to 3.5 km (half the distance between adjacent DASARs), over 80% of the calls generated had a >99% probability of

being localized by the system. These results were derived from calls whose frequency range was restricted to between 20 and 170 Hz, in order to permit the assumption that whale calls had no directivity, and the same restriction was applied here. By restricting call localizations to those less than 3.5 km from the nearest DASAR (and between 20 and 170 Hz), one could assume that the call detection rate was equal to the underlying call production rate in that region, removing any masking effect.

The second approach was to simply include background noise estimates into the subsequent statistical model, with the expectation that noise masking effects would be automatically incorporated into the behavioral adjustment factors. To that end R_{max} was set to 500 km for certain analysis scenarios, an artificially large value that effectively permitted all call samples into the model, regardless of location.

C. Statistical regression models for call density

In order to derive correction factors that adjust for behavioral responses to environmental factors, we extended the regression analysis in Thode *et al.* (2017, in review) to determine how six potential prediction variables affected observed call density, once masking effects were removed. We constructed four different scenarios that differed in the number of predictor variables used, how call masking effects were treated, and how ASAR data were handled. Sites 3 and 5 were fitted with separate and independent regression models, regardless of scenario.

For all models and all scenarios the dependent variable – **CallDensity**- was computed for every DASAR at both sites. It is defined here as the number of calls detected over a 10-minute window within a range R_{max} of a particular DASAR. Each

localized call is assigned a call density, computed using a ten-minute window centered on the detection time of that call. As discussed above, the frequency range of calls was restricted between 20 and 170 Hz.

The complete set of environmental factors, or predictor variables, associated with each call was defined as follows:

TimeOfDay: Diel effects have been observed in numerous mysticete and odontocete species, including bowheads (Blackwell *et al.,* 2007; Goold, 2000; Baumgartner and Fratantoni, 2008), so a categorical variable was defined that is '0' for calls generated at night and '1' for calls generated during daylight. Daylight is defined as any time that lies between local sunrise and sunset. A preliminary analysis of the potential impact of dawn or dusk on call rate (defined as anytime within one hour of sunrise or sunset, respectively), found no effect, so the final values assigned were either '0 or '1'.

DistanceFromShore: This factor accounts for potential differences in animal abundance with distance from shore, such as a density gradient. The value assigned to this continuous variable was simply '1' if the call was generated closest to DASAR A, '2' if closest to DASAR B, etc. Thus larger numerical values were associated with further distances from shore.

Including this variable was an attempt to compensate for the fact that detections on the northern and southern edges of the tracking array (i.e., **DistanceFromShore** values 1 and 7, which correspond to DASARs A and G) may not be successfully localized as effectively as calls generated in the middle of the array, because fewer

DASARs are available to contribute to a localization at the array edges. So this factor is technically a detectability adjustment, and not a behavioral adjustment.

AmbientNoiseRMS: Noise levels associated with each call were computed by extracting a noise sample beginning 3 s before the start of the call sample, and then computing the root-mean-square (RMS) sound pressure level (SPL) using the same time duration and bandwidth as the call. A 3 s time shift was chosen because most bowhead whale calls are shorter than 2 s, and an extra buffer second was added to reduce the possibility that any signal energy might contaminate the noise sample. The noise samples used here are thus identical to those used in (Thode *et al.*, in review).

AirgunCSEL: Seismic airgun activity was detected using the same automated algorithm described in detail by Thode *et al.* (2012), and modified by Thode *et al.* (in review). Seismic activity was designated as "present" for a given call on any DASAR if at least one airgun pulse was detected on DASAR G at the same site within 5 min of the detected call. The motivation for using DASAR G data was that it was at or near the deepest location at each site, and thus detected many more airgun pulses than sensors in shallower water. Using detection criteria from a deep location thus acted as a safeguard against missing pulse detections on shallower DASARs. If airgun pulses were associated with a call, then the cumulative sound exposure level (cSEL; dB re 1 µPa^2-s) was calculated over the 10-minute window centered around the call, the same metric used by Blackwell *et al.* (2015). However, that previous study measured cSEL over fixed non-overlapping 10-minute intervals, while here the cSEL associated with each call was integrated over all airgun pulses detected within 5 min of a given call. Thus each call

detected was assigned a unique cSEL value, as well as a binary designation of 'present' or 'absent'. During the analysis it was found that during 2009 precisely-timed sequences of bowhead calls occasionally falsely triggered the airgun detector, so the detector output was manually reviewed and corrected for all years.

DASARstatus: The fact that multiple DASARs lost their directional capability across the seasons raised concerns as to whether the resulting "ASARs" would experience altered call densities in their vicinity, since the existence of these positions would arise from triangulation from more distant sensors. To this end some scenarios defined a categorical variable where '0' designates an ASAR and '1' a fully functional DASAR. In other scenarios, calls associated with ASARs were removed altogether and the model did not use this variable. The disadvantage of this latter approach is that it eliminates large numbers of calls during 2013.

Year: Finally, in recognition that some of the factors above may vary by year (e.g. **DistanceFromShore**), in most scenarios the year the call was detected was included as a categorical variable, with '0' corresponding to 2008, etc. One scenario pooled data across all years, thus eliminating **Year** as a factor.

The behavioral adjustment model consisted of two generalized linear models (GLM) per site, which were derived by regressing the logarithm of call density against various combinations of the six factors listed above, assuming a log-normal distribution for call rate. (A log-normal GLM yielded residuals that better fit a normal distribution than those from a Poisson regression on call rate). The two GLMs per site were fitted from call samples that were produced in the presence or absence of seismic airgun

activity, respectively. For the submodel where seismic airgun activity was absent, **AirgunCSEL** was removed as a predictor variable. The Bayes Information Criterion (BIC) was used to build a stepwise regression model from a constant value, allowing up to third order terms and cross-terms between the factors. Thus for each scenario listed below four GLM models were fit to cover every combination of site and airgun presence/absence.

D. Call correction weights

Compensating for environmental influences on calling behavior involved defining a "reference" environment that contained fixed values for up to five environmental factors (excluding **Year**). The underlying logic of the following steps is that if a call is detected under conditions associated with call densities lower than those measured during reference conditions, then had the environmental conditions actually been at "reference" conditions one would expect additional calls to have been generated (due to the animals' shift in calling behavior), and thus measured call density estimates from that time should be revised upwards. Similarly, if calls are detected under conditions associated with higher call densities than those generated under reference conditions, then call density estimates from that time should be revised downwards. It's important to re-emphasize that this adjustment strategy is a separate issue from correcting call detectability: it provides an adjustment for behavioral variations produced in response to changes in environmental conditions (although the **DistanceFromShore** may also incorporate detectability variations as well).

Call correction factors, or *weights*, are generated from predictions from the GLM models. Each GLM uses the environmental factors co-measured with each detected call to predict the "*expected*" call density from the GLM, using those specific factor values as inputs. Applying the reference values for these same factors as inputs to same GLM produces a "*reference*" density that remains constant for every call sample localized during a given year. The ratio of the reference to the expected call density is defined as the "*call correction weight*" for that call, and quantifies how heavily a given call will contribute to the final cumulative call abundance. Specifically, if $d(\underline{F}, Y)$ represents the expected call density predicted by a particular GLM derived during Year Y using environmental factors $\underline{F}$, then the call correction weight w for a specific call is defined as

$$w(\underline{F}, Y) = d\left(\underline{F_{fixed}}, Y\right) / d\left(\underline{F_{obs}}, Y\right) \quad (1)$$

where F_{fixed} represents the reference values of the environmental factors, and F_{obs} represents the environmental factors actually observed with the call[1]. If no environmental factors are adjusted to their reference values, so that $F_{fixed} = F_{obs,}$ then the weights will have a value of 1, since the expected and reference call densities are equal. A *raw* cumulative call abundance estimate effectively assigns each detected call a weight of one (so that the cumulative call abundance is simply the sum of all calls measured across the seasons), while an *adjusted* call abundance sums over the call correction weights, so that calls generated during conditions associated with low call densities (call

[1] One might be tempted to define $w(\underline{F}, Y) = d\left(\underline{F_{fixed}}, 2008\right) / d\left(\underline{F_{obs}}, Y\right)$ in order to compensate for shifts in behavioral responses *between* years as well as within a single season. Unfortunately this approach also eliminates any long-term changes in call density arising from increased abundance.

densities below the reference call density) result in larger correction weights, and thus have higher contributions to the final call abundance.

To determine the relative impact a particular environmental factor has on the adjusted cumulative call abundance, one can evaluate the correction weights that arise from assigning the reference value to only that factor in the GLM, while retaining the values of the other factors actually measured with the call.

The net impact of these correction factors depends on how the migration is distributed over time. If environmental conditions fluctuate randomly throughout a migration characterized by a steady progression of whales, then the net effect of the corrections on the cumulative call abundance estimate may be small, since an increase in adjusted call density at one time can be offset by an adjusted call density decrease at another time. By contrast, if the migration is heavily concentrated or clustered over specific time intervals within the season, then the environmental conditions present during these periods can have a disproportionate impact on the cumulative estimate. These considerations become important when interpreting the results below.

In principle, if the behavioral model of the population does not vary across years (e.g. the population's cumulative response to changes in ambient noise levels does not depend on **Year**), then the exact values used for the reference environment should be irrelevant for determining the *relative* cumulative abundances, as long as the correction factors are applied consistently across sites and seasons. The values used in the scenarios below were '1'(daytime) for **TimeOfDay**, 3 (DASAR C) for **DistanceFromShore**, 85 dB re 1 µPa (SPL, 20-170 Hz)) for **AmbientNoiseRMS**, 115 dB re 1 μPa^2-s for

AirgunCSEL, and '1' for **DASARstatus (**when applicable to a scenario). The reference value for **AmbientNoiseRMS** was the median value of the multi-year noise distribution, while the **AirgunCSEL** value was selected to be the value at which the GLM predicts call densities with the same values as those when airguns are absent. (At lower airgun exposure values call densities are higher than when airguns are absent). As mentioned in the previous subsection, GLM models were fit for circumstances when airgun signals were present or absent, and the reference value for **AirgunCSEL** was only applied to the GLM model derived from calls generated when airgun signals were present. Alternate reference values were also analyzed to test whether behavioral models derived during different seasons were similar.

E. Final scenarios

Four particular scenarios captured the essential features of all the possible analysis combinations discussed above:

NoLocLimit: Here no adjustments or localization restrictions (R_{max} = 500 km) were made on calls when computing cumulative call abundance. Calls associated with ASARs were not removed. **Year** was incorporated as a categorical variable.

LocAdj: Calls were restricted to 3.5 km range from the nearest sensor. Calls associated with ASARs were retained, so **DASARstatus** was included as a predictor variable. **Year** is incorporated as a categorical variable. This scenario incorporated the largest number of predictor variables and most stringent localization restrictions, and thus examined to what degree, if any, call localization improves consistency of call abundance estimates between sites.

LocRemove: Same as previous scenario, but calls associated with ASARs are removed, and **DASARstatus** is not included as a predictor variable.

LocPooled: Same as ***LocAdj*,** but **Year** is not incorporated as a variable in the predictive GLM (and thus data are pooled across years).

As a reminder, for every scenario each site was independently fitted with two separate submodels, one for when airgun activity was present, and one for when it was absent.

F. Metrics for comparing call abundance between sites

Once the raw and adjusted cumulative call abundances N_3 and N_5 were computed for sites 3 and 5 for each year under each scenario above, two metrics were defined to compare the degree of similarity shared by both multi-year time series. The first metric is simply the Pearson's correlation coefficient R, computed from the seven-sample time series from each site. The closer the coefficient is to 1, the more correlated the multiyear time series. The second metric, the "mean percentage spread" (MPS), is a normalized measure of the difference between the cumulative abundances, averaged across all years, and is defined as follows:

$$MPS = 100 * \overline{\frac{|N_3 - N_5|}{(N_3 + N_5)/2}} \quad (1)$$

where the average is across all seasons.

While the use of different reference values for the environment will change the absolute values of N_3 and N_5, both metrics R and MPS should remain the same regardless

of the reference environment chosen, if the behavioral model of the population remains consistent across years.

III. RESULTS

A. Statistical regressions of call density

Figures 2 and 3 shows examples of the slices of the GLM prediction surface of the ***LocAdj*** scenario, which incorporates the full set of predictor variables. Figure 2 displays slices from the 2013 model, for times when airgun activity is absent, and Fig. 3 displays slices from the same year, but for times when airgun activity is present (and thus **AirgunCSEL** is included as an additional predictor variable). The 2013 season was chosen as an example because several DASARs lost directional capability that season, and so the **DASARstatus** factor was relevant for that year. Both Figures also show how the predicted mean call density varies with **Year,** when all other environmental factors are fixed at reference values.

Although most variables were accepted into the model using a BIC criterion, the effect size of some factors is small. For example, the categorical variable **TimeOfDay** generally has a small effect, as is the case here when DASAR C is fixed as the reference **DistanceFromShore** value. The effect size of this diel factor varies as a function of **DistanceFromShore**, **Site,** and **Year**. By contrast, the **AmbientNoiseRMS** factor is always significant, and follows a pattern previously reported by Thode *et al.* (in review), where increasing wind-driven noise levels are associated with higher calling densities. It's worth noting that in the ***NoLocLimit*** scenario, where no restrictions are made on a

call's range to a sensor, the resulting prediction model shows a *decrease* in call density with increasing noise levels, revealing how noise masking effects counteract the populations' behavioral response.

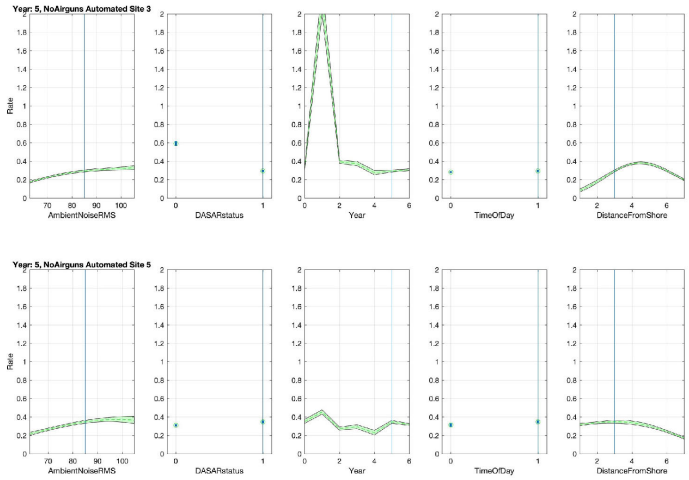


Figure 2: Slices of prediction surface of GLM model for ***LocAdj*** scenario, for situations when no airgun activity is present. The top row plots Site 3 predictions; the bottom row plots Site 5. Vertical blue lines indicate values that are held fixed when computing other plot slices. For example, for the first four subplots the reference DASAR is set to 'C' (i.e., **DistanceFromShore** is fixed at 3). Regressions are shown for Year 5 (2013), except for the **Year** subplot.

A significant nonlinear relationship exists between call density and **DistanceFromShore** that is strongly site- and year-dependent. **DASARstatus** reveals that ASARs have a much bigger impact on Site 3 than Site 5, with call densities actually being predicted to have higher values near ASARs, despite their inability to contribute to the call localization. This result is counterintuitive, but was consistent across all scenarios and models.

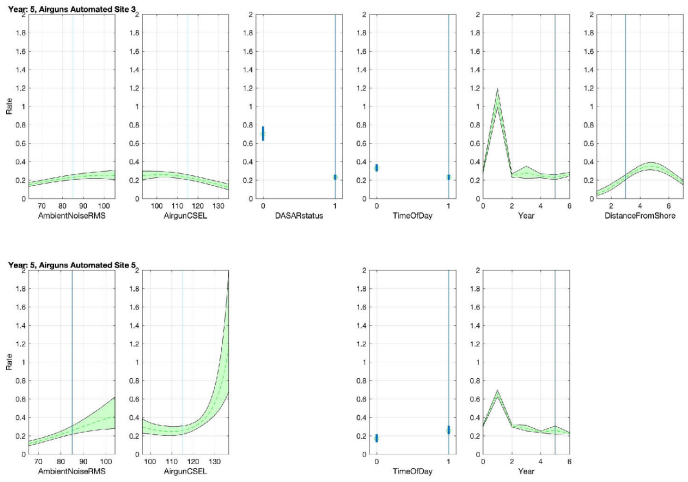


Figure 3: Same as previous, except regression is applied to 2013 data where airgun activity is present, allowing another predictor variable (**AirgunCSEL**) to be incorporated. Neither **DASARstatus** or **DistanceFromShore** were found to be significant predictors for Site 5.

Figure 3 also displays predictions that are partially consistent with previous analyses (Blackwell et al., 2015; Thode et al., in review), in that increasing cSEL levels are generally associated with decreasing call densities (although high call densities were correlated with high exposure levels at Site 5 in 2013). The correlation between call density and other factors during seismic activity in Fig. 3 is generally similar to that when airguns are absent (Fig. 2), although the final Site 5 airgun-present model does not incorporate **DASARstatus** or **DistanceFromShore** as significant factors at all.

Figures 2 and 3 also show how **Year** impacts mean predicted call density when other factors are fixed to their reference values. The choice of year has little effect on mean call density, with the prominent exception of 2009 (Year 1), where the predicted mean densities at Sites 5 and 3 are roughly three to five times that of the other years,

respectively. The unusually high call densities predicted by the 2009 GLM model had repercussions throughout the subsequent analyses.

Figure 4 shows the resulting prediction curves under the ***LocPooled*** scenario when airguns are present, where data have been pooled regardless of the particular year. The no-airgun model associated with ***LocPooled*** (not shown) shows nearly identical predictions as Fig. 4.

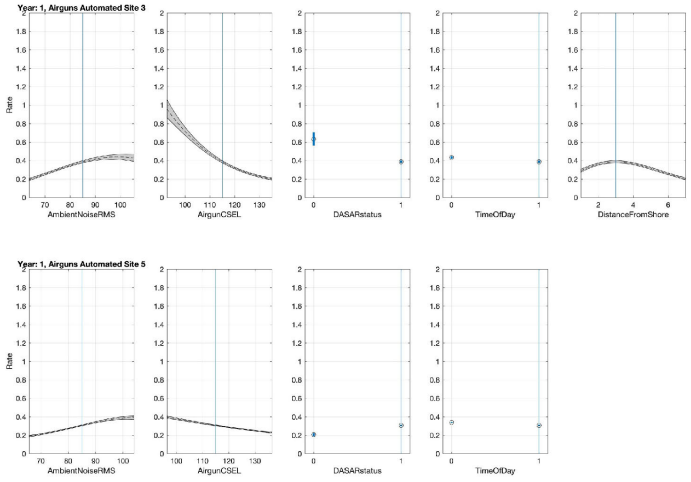


Figure 4: Prediction slices when data are pooled across all seasons under the ***LocPooled*** scenario, in the presence of sound from airgun pulses.

B. Call correction weight examples

The next three figures illustrate three examples of call correction weights from the ***LocAdj*** scenario, plotted at the dates and times when call samples were detected at Site 3 during airgun activity. By systematically fixing different factors, one gains insight into

the relative impact of the factor on the adjusted call abundance. For example, Fig. 5 illustrates how the weights shift when a single factor, **TimeOfDay,** is held fixed at 'daytime' values. If a particular call was detected during daylight conditions anyway, then fixing the **TimeOfDay** factor to 'daytime' has no impact on the predicted call density, and the weight factor for that call is simply one. If call rates are higher during the nighttime, the resulting correction weights will be below one.

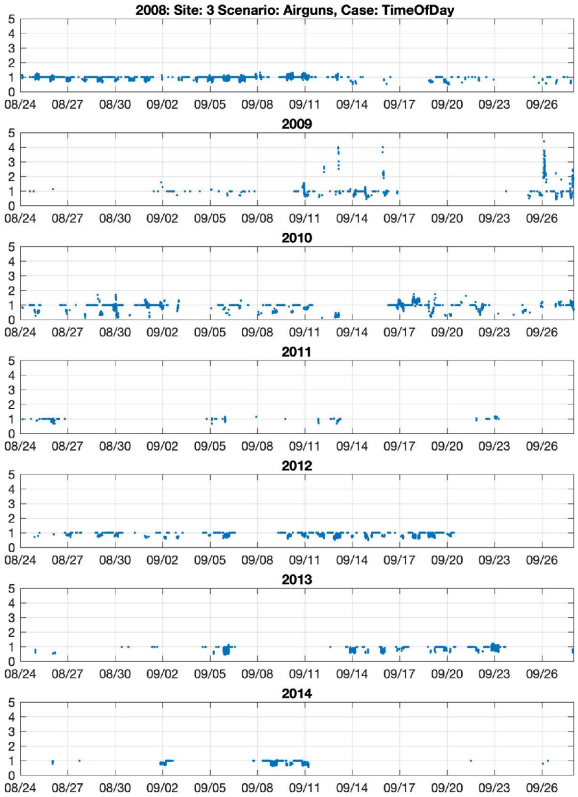


Figure 5: Call correction weights for all calls at Site 3 during airgun activity for the ***LocAdj*** scenario, when **TimeOfDay** is fixed at "daytime" levels.

The results indicate that the diel effect is generally weak, and during the last few years of the study call densities were higher at night.

Figure 6 illustrates the effect of fixing **AmbientNoiseRMS** to 85 dB for the ***LocAdj*** scenario with airgun sounds present. Comparing this figure with Fig. 5, one sees that ambient noise has a more dramatic adjustment effect than the relatively minor changes caused by diel adjustments. In particular, one finds that during 2009 individual calls are commonly given correction weights approaching 3; that is, a single call is often weighted as three calls in the adjusted cumulative count. Other years show less dramatic correction factors.

Figure 7 shows the resulting correction weights when all factors (except **Year**) are held at the "reference" values listed in Section II.C during the ***LocAdj*** scenario. These are the actual correction factors used to estimate the cumulative abundance in the following section. In this situation every call undergoes some adjustment, but once again one sees very high correction weights assigned to calls in 2009.

25

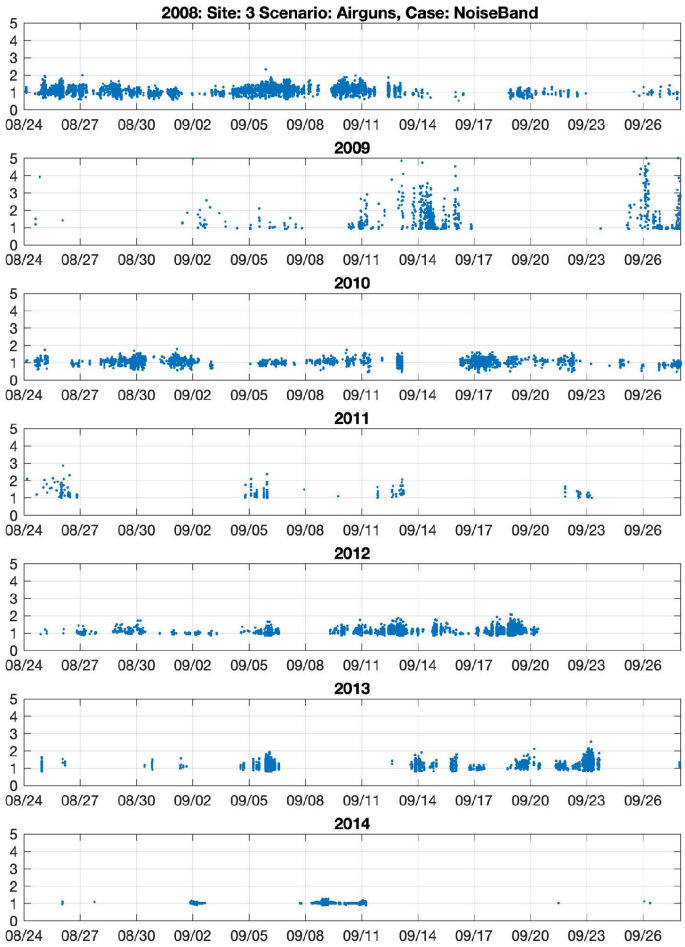


Figure 6: Call correction weights for all calls at Site 3 during airgun activity for the ***LocAdj*** scenario, ambient noise adjustments only.

26

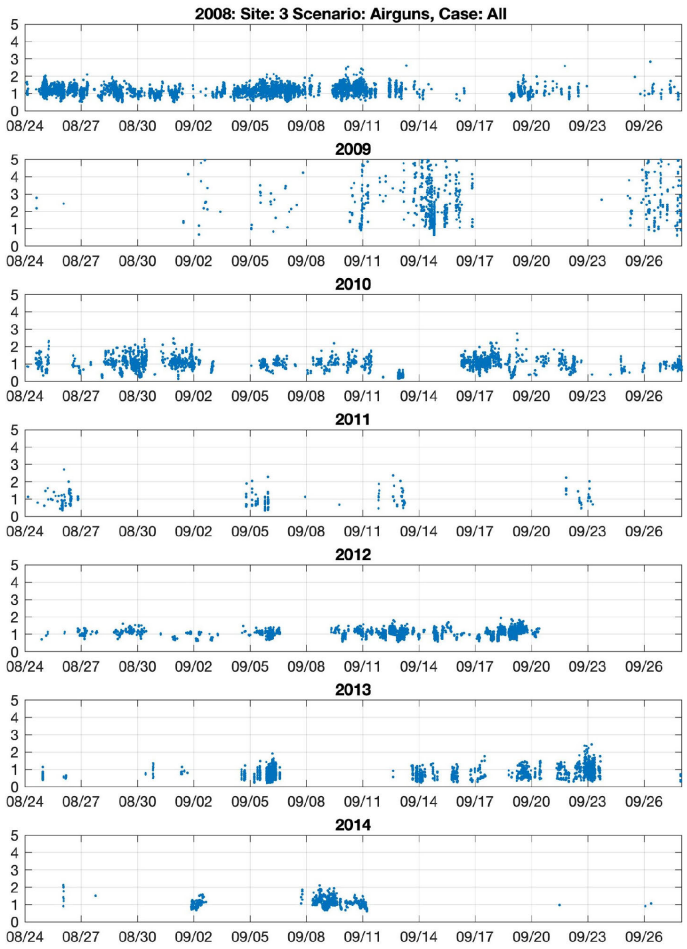


Figure 7: Call correction weights for Site 3, airguns present, when all six factors are incorporated into the ***LocAdj*** scenario. Note large correction factors for 2009.

C. Raw and adjusted cumulative call abundance

1. NoLocLimit scenario

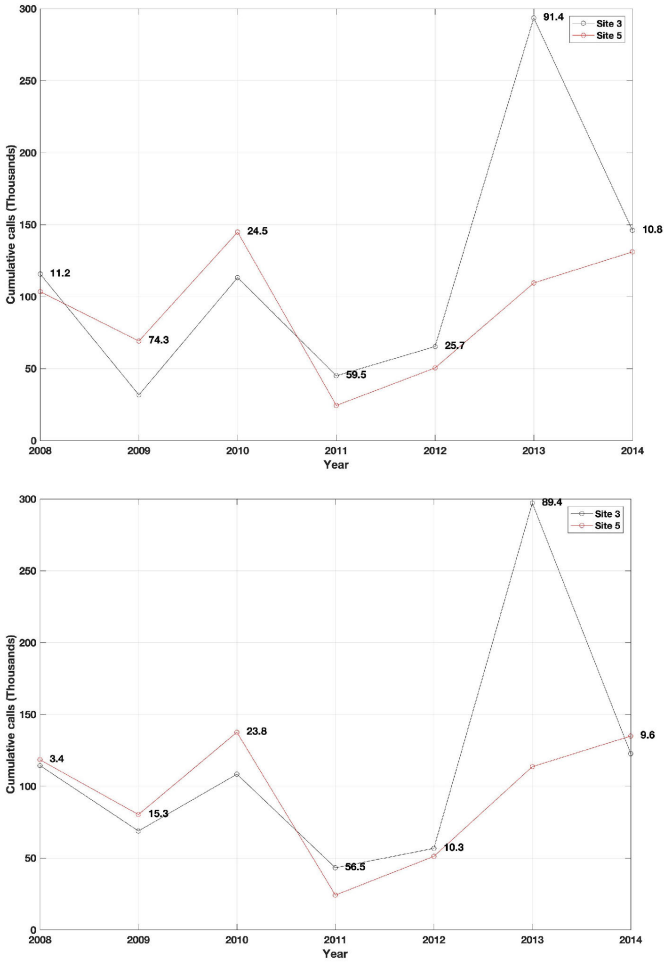


Figure 8: Time series of cumulative raw call abundance, in thousands for ***NoLocLimit*** scenario, with red points indicating Site 5, and black points Site 3. The bold labels next to the curves give the percent spread between the sites for each year. Top: raw cumulative call abundance. Bottom: adjusted call abundance after applying call correction weights.

Figure 8 shows the raw and adjusted cumulative call abundance (in thousands of counts) measured at both Sites 3 and 5 over seven years, for the ***NoLocLimit*** scenario. Table II shows the corresponding metrics between the sites when various factors are adjusted. The MPS decreases from 42 to 30 percent when the cumulative abundance is adjusted, with **AmbientNoiseRMS** being the factor with the greatest individual contribution to the reduction. Figure 8 suggests that the improvement in MPS arises primarily from changes in 2009, where the percentage spread falls from 74 to 15 percent. In contrast with the improvements in the MPS, the correlation coefficient between sites changes very little after factor adjustments, and the 95% confidence intervals include a value of zero, or no correlation, as a possibility.

Table II: Metrics of cumulative call abundance consistency between Sites 3 and 5 for ***NoLocLimit*** scenario. Each row shows the effect of adjusting for a single predictive factor. "All" indicates the effect of adjusting all factors. MPS=Mean percentage spread between sites [Eq. (1)]. *R*=Pearson's correlation coefficient. [Brackets are 95% confidence intervals].

Term	*MPS*	*R [CI]*	*MPS (2013 excluded)*	*R [CI] (2013 excluded)*
No Adjustment	42	0.56 [-0.34/0.92]	34	0.84 [0.09/0.98]
All	30	0.52 [-0.38/0.92]	20	0.97 [0.74/1.0]
AmbientNoiseRMS	33	0.57 [-0.32/0.93]	23	0.88 [0.25/0.99]
AirgunCSEL	41	0.53 [-0.37/0.92]	33	0.86 [0.14/0.98]
DASARstatus	41	0.59 [-0.29/0.93]	33	0.87 [0.21/0.99]
TimeOfDay	39	0.62 [-0.24 0.94]	32	0.87 [0.21/0.99]
DistanceFromShore	39	0.59 [-0.29/0.93]	30	0.89 [0.27/0.99]

29

Site 3 exhibits an unusually high cumulative abundance in 2013, a situation that also occurred at all other sites except Site 5 that year (Blackwell et al., in prep). High call counts were also obtained in another study, a DASAR deployed over the same years northeast of Northstar, an oil production island west of Prudhoe Bay, in a location near the southern edge of the bowhead migration corridor (Kim *et al*. 2014). Additionally, an annual aerial survey in the Beaufort Sea sighted very high numbers of bowhead whales (Clarke *et al*. 2014). Table II thus includes recomputed metrics, where 2013 is excluded as an outlier, an action that leads to significant improvements in both MPS and *R* after factor adjustments. With this exclusion the behavioral factor adjustments still improve the metrics, with **AmbientNoiseRMS** being the dominant single factor.

30

2. *LocAdj scenario*

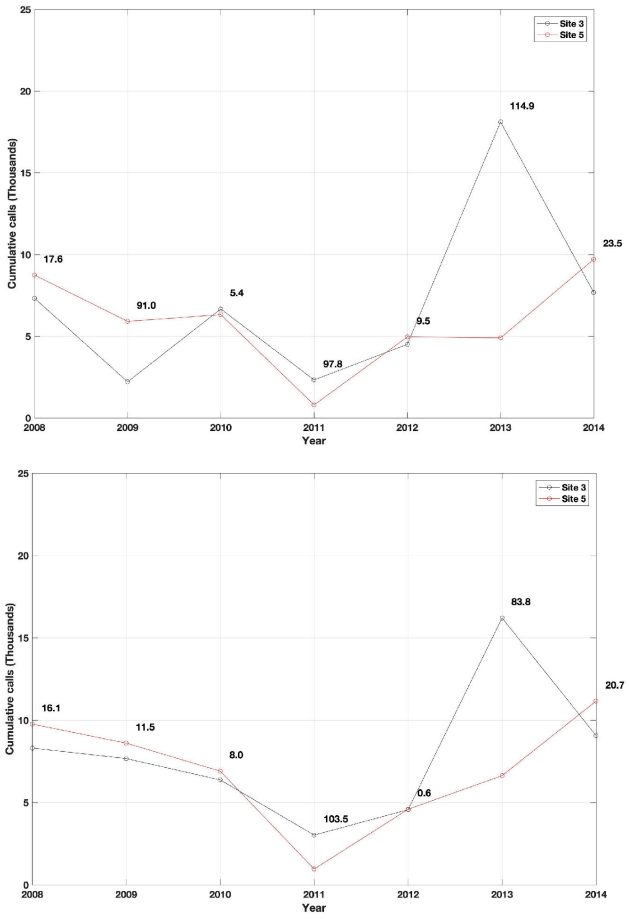


Figure 9: Time series of cumulative call abundance for ***LocAdj*** scenario, using format identical with Fig. 8. Call counts are limited to those generated within 3.5 km of a DASAR at a given site. Top row: raw call abundance. Bottom: adjusted call abundance, using the correction weights shown in Fig. 7.

Figure 9 displays the results of the ***LocAdj*** scenario, which only differs from the previous scenario by enforcing a localization limit of 3.5 km range from the nearest sensor, causing cumulative abundance values to drop by an order of magnitude. Table III lists the associated performances metrics, including those excluding 2013. As with the ***NoLocLimit*** scenario, incorporating call correction weights improves the two metrics, with the year 2009 remaining the primary reason behind the improvements. Unlike the ***NoLocLimit*** scenario, the ***LocAdj*** scenario also demonstrates substantial improvements in *R* when correction weights are applied, even when the outlier year 2013 is retained. The confidence intervals, however, still include 0 if all years are included.

The next two scenarios examine perturbations of this scenario, which will be the baseline scenario in the Discussion.

Table III: Metric evaluations of consistency of cumulative call counts between Sites 3 and 5, for the ***LocAdj*** scenario. MPS=Mean percentage spread between sites [Eq. (1)]. *R*=Pearson's correlation coefficient.

Term	*MPS*	*R [CI]*	*MPS (2013 excluded)*	*R [CI] (2013 excluded)*
No Adjustment	51	0.19 [-0.65/0.83]	41	0.81 [0-0.98]
All	34	0.46 [-0.45/0.90]	27	0.99 [0.95 – 1.00]
AmbientNoiseRMS	45	0.24 [-0.63/0.84]	34	0.85 [0.13-0.98]
AirgunCSEL	53	0.15 [-0.68/0.81]	43	0.77 [-0.12/0.97]
DASARstatus	48	0.39 [-0.52/0.88]	41	0.82 [0.02/0.98]
TimeOfDay	51	0.19 [-0.66/0.82]	41	0.85 [0.11/0.98]
DistanceFromShore	49	0.29 [-0.59/0.86]	37	0.85 [0.14/0.98]

3. *LocRemove scenario*

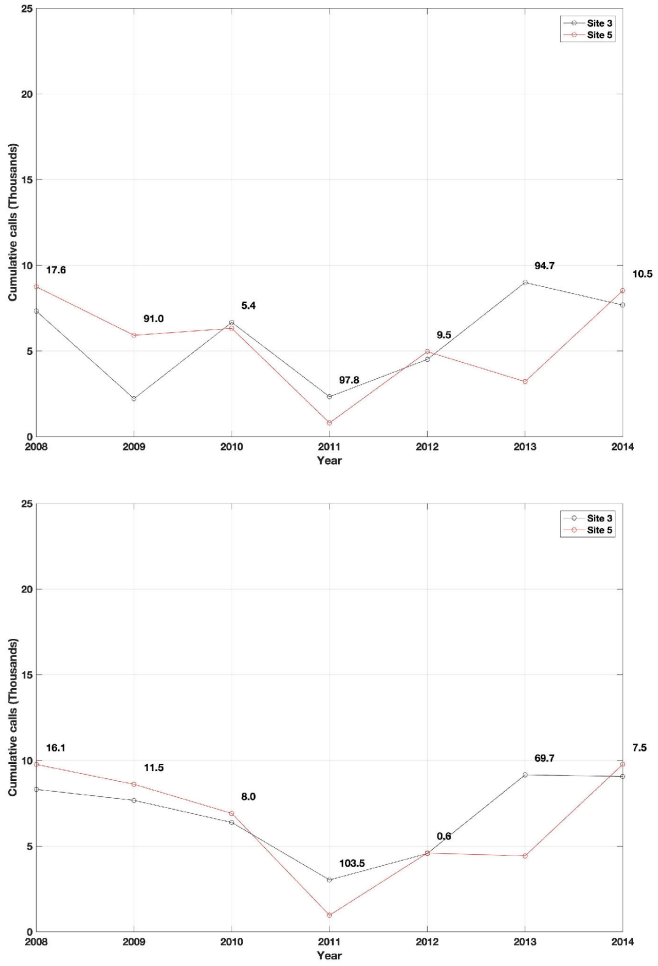


Figure 10: Raw and adjusted cumulative call abundance for ***LocRemove*** scenario, where calls close to ASAR positions have been removed.

Figure 10 shows the resulting raw and adjusted cumulative call abundances for the ***LocRemove*** scenario, which differs from the ***LocAdj*** scenario in how ASARs are treated. In this scenario, calls close to ASAR positions are simply removed before applying the statistical model. A comparison of Figs. 9 and 10 shows that the major effect of this different treatment is to remove large numbers of calls from Site 3 in 2013, with little difference from ***LocAdj*** in other years. Given the large number of ASARs present at both sites in 2013 (Table I), this improvement is not surprising.

Table IV: Metric evaluations of consistency of cumulative call counts between Sites 3 and 5, for the ***LocRemove*** scenario. MPS=Mean percentage spread between sites [Eq. (1)]. *R*=Pearson's correlation coefficient.

Term	*MPS*	*R [CI]*	*MPS (2013 excluded)*	*R [CI] (2013 excluded)*
No Adjustment	47	0.43 [0.48/0.89]	39	0.79 [-0.05/0.98]
All	31	0.74 [0.02/0.96]	25	0.99 [0.87/1.0]
AmbientNoiseRMS	40	0.45 [-0.46/0.90]	32	0.84 [0.09/0.98]
AirgunCSEL	48	0.39 [-0.52/0.88]	41	0.74 [-0.18/0.97]
TimeOfDay	46	0.43 [-0.48/0.89]	39	0.83 [0.07/0.98]
DistanceFromShore	40	0.66 [-0.19/0.94]	35	0.85 [0.11/0.98]

Table IV demonstrates how the reduction of calls in 2013 improves the *R* metric substantially, with only mild improvement to the MPS metric, when compared with ***LocAdj*** in Table III. The improvement from excluding 2013, while still extant, is not as

dramatic. Interestingly, **DistanceFromShore** becomes the dominant factor for improving the correlation between sites.

4. LocPooled scenario

Finally, the ***LocPooled*** scenario examines the effect of pooling all call localization data when deriving the call correction weights, by excluding **Year** as a predictor variable in all models. The raw cumulative call abundance remains the same as the top row of Fig. 9, and Fig. 11 shows the resulting adjusted call abundance from the pooled data. Unlike the previous scenarios, the adjustments only mildly affect the 2009 and 2013 samples, so that the metrics in Table V show hardly any improvement over the raw call abundances.

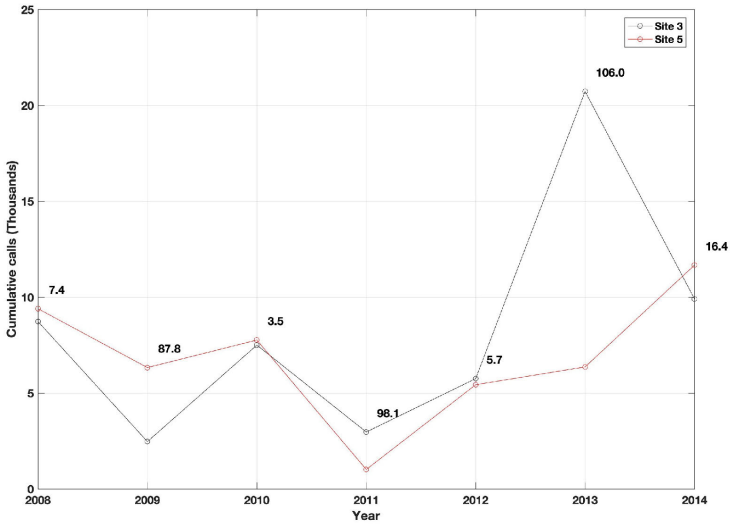


Figure 11: Adjusted cumulative call abundance for ***LocPooled*** scenario, where the correction weights are derived from pooled data, regardless of year recorded. Compare with top figure in Fig. 9.

Table V: Metric evaluations of consistency of cumulative call counts between Sites 3 and 5, for the ***LocPooled*** scenario (data pooled over years). MPS=Mean percentage spread between sites [Eq. (1)]. *R*=Pearson's correlation coefficient.

Term	*MPS*	*R [CI]*	*MPD (2013 excluded)*	*R [CI] (2013 excluded)*
No Adjustment	51	0.19 [-0.65/0.83]	41	0.81 [0/0.98]
All	46	0.32 [-0.57/0.87]	36	0.84 [0.10/0.98]
AmbientNoiseRMS	46	0.25 [-0.62/0.84]	35	0.84 [0.08/0.98]
AirgunCSEL	51	0.22 [-0.64/0.83]	41	0.83 [0.04/0.98]
DASARstatus	50	0.28 [-0.60/0.85]	41	0.82 [0.02/0.98]
TimeOfDay	51	0.22 [-0.64/0.83]	42	0.82 [0.02/0.98]
DistanceFromShore	51	0.17 [-0.67/0.82]	38	0.83 [0.05/0.98]

IV. DISCUSSION

A. Comparison between adjustment scenarios

When comparing all four scenarios across Figs. 8-11 and between Tables II-V, some general trends are visible. At first glance it appears that employing call correction factors for environmental variations improves the comparison metrics between sites. For example, for the ***LocAdj*** scenario the MPS falls from 51 to 34 percent, and the correlation rises from 0.19 to 0.47 when call abundance is adjusted, and these numbers improve further when 2013 is excluded: the MPS falls from 41 to 27 percent, and the correlation increases from 0.81 [0-0.98 confidence interval] to 0.99 [0.95-1.00 CI]. Background noise levels are consistently a dominant factor in improving the comparison metrics

between sites, but multiple factor adjustments are required to reduce the confidence intervals of the correlation. Applying call correction weights still leads to some improvements in the comparison metrics, even if call locations are not restricted to short ranges in order to mitigate call masking issues (***NoLocLimit***; Fig. 8). Indeed, localization restrictions have relatively minor effects on the metrics.

The only situation where adjusting call counts causes little to no improvement between sites occurs when call density data are pooled across seasons (Fig. 11). Creating a separate statistical regression model for each season seemed crucial to improving agreement between sites. The ***LocRemove*** scenario is much more successful than the ***LocAdj*** approach in reducing the discrepancy between the 2013 cumulative call abundances at Sites 3 and 5, but this may simply be a coincidence: 2013 happened to be the year when an enormous number of whale calls were present, but also when many DASARs lost localization capability at sites 3 and 5, so removing ASARs automatically removed large numbers of calls.

Although these improvements in the comparison metrics appear encouraging, a closer look at Figs. 8-11 shows that during most seasons, applying call correction weights had little impact on the MPS. Figure 7 provides some insight into why this is the case: calls seem roughly likely to be assigned a weight either greater than or less than one, so that the mean value of the correction weight distribution remains close to one, and that the sum of the weights is relatively close to the number of original samples. This observation seems closely related to the concept of "pooling robustness" in distance sampling theory (Buckland *et al.*, 2004), which states that variations in call detectability arising from various environmental factors need not be explicitly modeled, as long as the

calls are sampled under a variety of environmental conditions that mirror the underlying probability distributions of these conditions. To provide one example, since sample calls are associated with a mixture of ambient noise levels that fluctuate above and below the reference value of 85 dB, noise effects should end up having little impact on the cumulative call count.

Why then, is there an improvement in the MPS and correlation call metrics when environmental corrections are incorporated? The answer is that a single year shows a dramatic response to environmental adjustments: 2009. During that year in the ***LocAdj*** scenario, the percentage spread between the Site 3 and 5 cumulative abundance fell from 91 to 11.5 percent, once call correction weights were applied (Fig. 9, bottom). To understand whether the behavioral adjustments are valid, we need to examine how 2009 is unique.

B. Why is 2009 unique?

At first glance, 2013 seems to be the most unusual year on record in terms of cumulative call counts, in that the total number of calls collected at Site 3 dwarfs that at Site 5 and both sites in other years. Given that most of the failed DASARs occurred in 2013, one might naturally wonder whether the automated localization procedure was breaking down. For example, could the algorithm be dividing a single call event into multiple localizations, simply because call detections at different sensors could not be successfully linked together? Evidence against this concern lies in the fact that high numbers of whales were observed that year from both aerial surveys and manual analysis

of contemporary acoustic data (Kim *et al.*, 2014). But while the 2013 abundance is large, it remains relatively unaffected by the environmental adjustments.

Instead, 2009 turns out to be the key year in the analysis, in that its cumulative abundance is very sensitive to the behavioral adjustments. The reason for this is not immediately obvious. The distributions of the various factors, like ambient and airgun noise, are not dramatically different between 2009 and other years. However, Figs. 2 and 3 (**Year** subplot) show that if all other factors are held fixed, 2009 displays much higher predicted call densities than other years. Figure 12 provides more direct evidence of this, by plotting the cumulative distribution functions (CDFs) of call densities assigned to the call sampled, broken down by season. Unusually high call densities characterize the 2009 CDF. Figures 13 and 14 show the impact of this distribution on the GLM prediction slices for 2009: the predicted call densities at Site 3 become much more sensitive to changes in the predictive factors, especially the noise-related variables. When airgun activity is absent (Fig. 13) **TimeOfDay** shows a strong diel effect towards calls occurring during daytime hours, a reversal of the usual expectation.

39

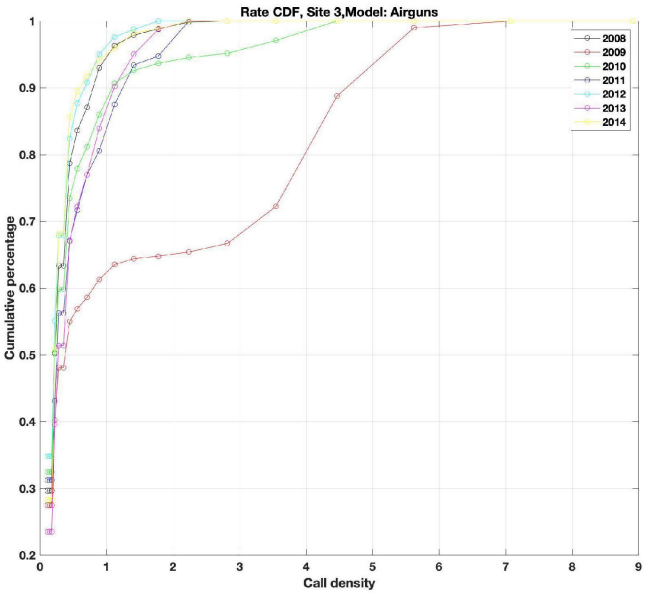


Figure 12: Cumulative distributions of call density for Site 3, when airgun activity is present. Note how nearly 40% of the 2009 call samples are associated with densities greater than 1 call/minute within 3.5 km range.

The results of this sensitivity can be seen in Fig. 7, which shows large call correction weights for 2009 (values of 4 or greater), compared with those from other years.

From where does this sensitivity to environmental factors arise? Figure 7 also indicates that call counts in 2009 are tightly grouped in clusters, a fact that Fig. 15 shows more clearly, by plotting the cumulative call abundance for both the raw and adjusted densities over time. During some years (e.g. 2008, 2014) calls accumulate steadily over time, but during other years a large chunk of the total call abundance occurs over less

than a week, and 2009 represents an extreme case. Over a period of just 4 hours over 900 calls accumulated at Site 3: 41% of the 2200 calls detected for the total season.

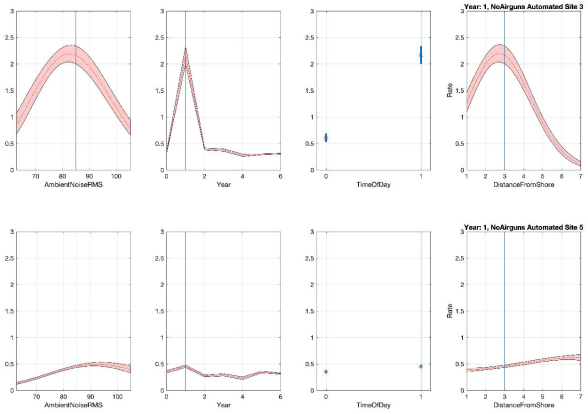


Figure 13: Slices of prediction surface of GLM model for ***LocAdj*** scenario, for 2009 (vs. 2013 in Fig. 3), when airguns are absent. Note the difference in the vertical scale for call density, which reflects the higher call densities present during 2009.

41

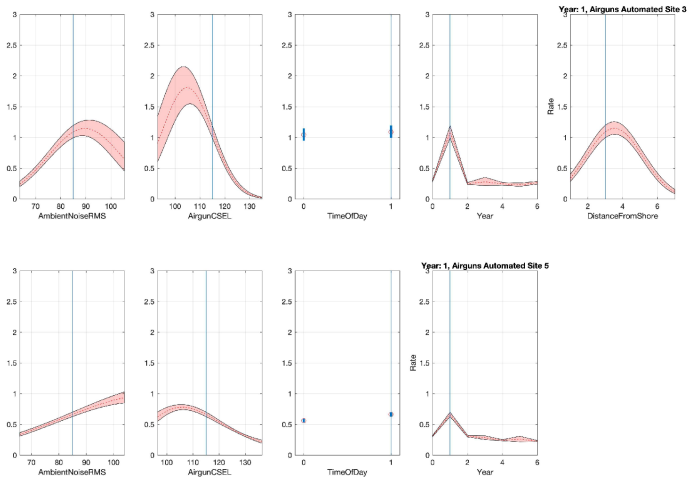


Figure 14: Slices of prediction surface of GLM model for ***LocAdj*** scenario, for 2009 (vs. 2013 in Fig. 3), when airguns are present.

Figure 16 shows a spectrogram of calls detected during this busy time frame, indicating that a regular sequence of calls is being repeated in the presence of seismic airgun activity. These *call sequences* have been documented elsewhere (e.g., Stafford *et al.*, 2012), but are not a common feature of the bowhead whale repertoire during autumn. The year 2009 may therefore be dominated by whales in a behavioral state different than other seasons. Complicating matters further, a storm moved over sites 3 and 5 shortly after the times shown in Fig. 16, lasting for nearly a week and thus reducing the cumulative abundance at both sites. Seismic airgun activity was present during most of the season as well.

42

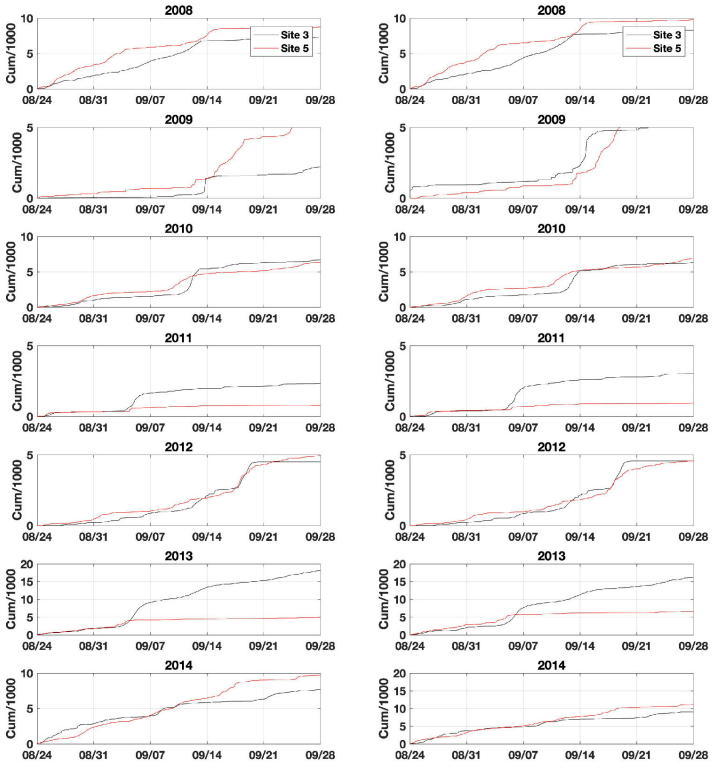


Figure 15: Cumulative call abundance over time, for Site 3 (black) and Site 5 (red). Each row represents a different year. The left column displays cumulative raw counts, while the right column displays the cumulative adjusted counts.

This potentially different behavioral state of the animals is responsible for the abnormally high call density, and the intense clustering of a substantial portion of the total cumulative abundance over such a short time window may have amplified the

impact of the environmental adjustments, as compared with a more typical years, where calls are more evenly distributed over time. In the latter case, calls are recorded under a variety of environmental conditions, and if the true animal density is uncorrelated with these factors, then the variations in call density caused by these fluctuations will average out. This is why summing over the call correction weights in Fig. 7 yields a cumulative abundance close to the raw call count, and why most years are relatively unaffected by the environmental adjustments.

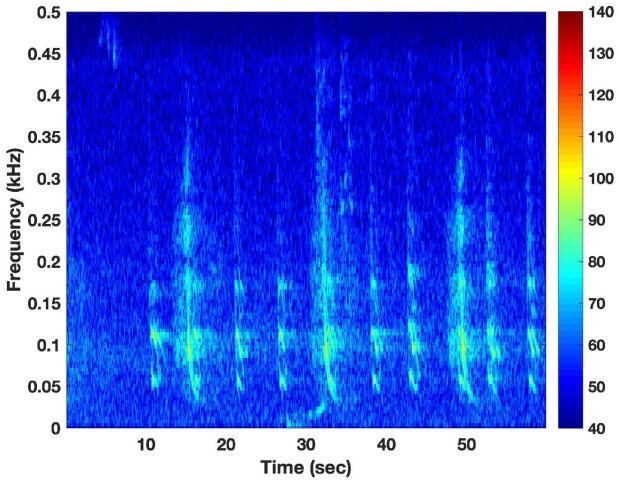


Figure 16: Call sequence made by an individual bowhead whale during Sept. 13, 18:58 local time, on DASAR D at site 3, during a time of high call density. The broadband pulses at 18, 32, and 50 sec are airgun pulses.

However, if calls arrive in dense clusters over short time intervals, and if these clusters arrive during conditions substantially different than the seasonally-averaged reference conditions, then the call correction weights can be substantial, and the adjusted

cumulative call abundance can display large shifts from the raw call abundance. That is what seemed to happen in 2009: the bulk of the call sequences took place over times when airgun activity was present, but at relatively low (105-110 dB) exposures. As a result, the predicted call density at low cSEL levels became very high (Fig. 14). Calls generated outside the call sequences happened to be detected when airgun activity exposures were higher (115 dB), resulting in very high correction factors and thus large behavioral adjustments. Similar effects happened for other factors: when a call occurred outside the call sequence it tended to be assigned a large correction weight, because the environmental factors associated with that call were typically different than those measured during the call sequence. Since the call density associated with the call sequence was high, the resulting correction weights became very high. Interestingly, although 2011 also displays evidence of sounds arriving in clusters, the statistical model makes no substantial adjustments to the call abundance.

The improvements in the site comparison metrics shown in Tables II-V thereby hinge on large call adjustments during a single year, which in turn seem to arise at least partially from an unusual set of behavioral circumstances that took place over only a few hours during 2009. The improvement in the metrics may simply be a statistical fluke; for example, the adjusted cumulative call abundances for 2009 in Fig. 15 (second row, right column) always show more accumulated calls at site 3 instead of site 5, even though one would expect a western migration to cause the easternmost site 5 to lead in the cumulative call count. There exist multiple years, however, where both the raw and adjusted call counts at site 3 lead site 5. The fact that the adjusted cumulative abundances between the sites match so well, despite being fit with completely

independent models, remains intriguing. Unfortunately, the small number of site/year data points preclude definite conclusions about the efficacy of adjusting call abundance for behavioral adjustments.

C. Multi Year trends in adjusted call abundance

If we tentatively allow that the adjusted call abundances shown for the ***LocAdj*** or ***LocRemove*** scenarios are valid, then both sites showed a consistent multiyear trend in the call abundance (Fig. 9, bottom). After a steady decline in abundance from 2008 through a nadir in 2011, (when the call abundance of 2,500 was a quarter of the 10,000 call abundance count of 2008), call abundances then increased steadily during the last three years of the study. A 74% decrease over four years is then followed by a 300% increase over three years, until the adjusted cumulative call abundance in 2014 has returned to nearly the same value (~10,000) as in 2008, the start of the study. Even the raw call abundance estimates show a consistent increase in abundance after 2011.

These are dramatic changes, and multiple potential explanations exist beyond simple changes in actual animal abundance within the range of our acoustic recorders. The migration corridor is known to shift from year to year, the behavioral responses of whales to environmental factors may be changing with time, or the timing of the migration onset may have shifted. Other work by Blackwell et al. (2014) has shown that the migration continues into October and early November, so that DASAR deployments only cover a fraction of the total migration duration. Were the timing of the migration to advance or retreat by a few weeks, the cumulative call abundance could change

substantially, even if the population levels remain steady. It would be interesting to collect further data from this area in upcoming years, for longer periods of time.

V. CONCLUSION

Cumulative call abundances of bowhead whale calls are computed and compared between two sites over seven years, to test the hypothesis that the relative abundance between sites and across seasons should be similar. Two adjustments to the raw call counts are tested: (1) restricting permissible call ranges to reduce masking effects, and (2) using a statistical model to compensate counts for variations in animal behavior in response to diel effects, distance from shore, and two kinds of ambient noise. The analysis was complicated by multiple failures of acoustic instruments in 2013, which led to that year being treated as an outlier.

Even without these adjustments the correlation between sites was substantial (if 2013 was treated as an outlier). Applying the adjustments further increased the match between the long-term trends at both sites (e.g. Figs. 9 and 10, bottom plots), which show call abundance falling by a factor of 3.5 between 2008 and 2011, and then nearly returning to the original levels by 2014. Shifts in the migration corridor and timing, along with potential changes in acoustic behavioral state, are all possible explanations for this multi-year shift in call abundance, as well as population size changes.

The effect of the environmental compensations fell disproportionately on one year, 2009. A close examination of that year showed that a significant fraction of the cumulative call count occurred over less than a day, due to a small group of whales

exhibiting an usual behavioral state that consisted of long stretches of call sequences. This situation, along with a long stretch of time where ambient noise levels reached very high levels, resulted in acoustic activity being very clustered and patchy in 2009. This uneven distribution of calls over time then amplified the call correction weights for calls generated outside those call sequences.

Unfortunately, the question as to whether these behavioral corrections truly improved relative call abundance estimates remains inconclusive, due to the heavy emphasis on a single year that may be a statistical fluke. Despite this uncertainty, it is clear that the cumulative call abundances at Sites 3 and 5 are highly correlated for six out of the seven years of the study, and that this correlation is visible even without the improvements provided by call masking and behavioral response corrections.

VI. REFERENCES

Baumgartner, M. F., and Fratantoni, D. M. (**2008**). "Diel periodicity in both sei whale vocalization rates and the vertical migration of their copepod prey observed from ocean gliders," Limnology and Oceanography **53**, 2197-2209.

Baumgartner MF, Hammar T, Robbins J. 2015. Development and assessment of a new dermal attachment for short-term tagging studies of baleen whales. Methods in Ecology and Evolution **6**:289–297.

Blackwell, S. B., Richardson, W. J., Greene Jr., C. R., and Streever, B. (**2007**). "Bowhead whale (Balaena mysticetus) migration and calling behaviour in the Alaskan Beaufort sea, Autumn 2001-04: An acoustic localization study," Arctic **60**, 255-270.

Blackwell SB, Nations CS, Griswold J, Kim KH, Greene CR, Thode AM, Norman RG. (**2011**). Beaufort Sea acoustic monitoring program. (Chapter 8) *In*: Funk, D.W., C.M. Reiser, D.S. Ireland, R. Rodrigues, and W.R. Koski (eds.). 2011. Joint Monitoring Program in the Chukchi and Beaufort seas, 2006–2010. LGL Alaska Draft Report P1213-1, Report from LGL Alaska Research Associates, Inc., LGL Ltd., Greeneridge Sciences, Inc., and JASCO Research, Ltd., for Shell Offshore, Inc. and Other Industry Contributors, and National Marine Fisheries Service, U.S. Fish and Wildlife Service. 592 p. + Appendices.

Blackwell SB, Nations CS, McDonald TL, Greene CR Jr, Thode AM, Guerra M, Macrander AM. (**2013**). Effects of airgun sounds on bowhead whale calling rates in the Alaskan Beaufort Sea. Marine Mammal Science 29:E342–E365.

Blackwell SB, Conrad AS, Thode AM. (**2014**). Characterizing underwater sounds at six locations in the U.S. Beaufort Sea during winter months. Appendix C *in* Joint Monitoring Program in the Chukchi and Beaufort Seas, 2012. LGL Alaska Final Report P1272-2 for Shell Offshore, Inc ION Geophysical, Inc, and Other Industry Contributors, National Marine Fisheries Service, and US Fish and Wildlife Service. 320 p + Appendices.

Blackwell SB, Nations CS, McDonald TL, Thode AM, Mathias D, Kim KH, Greene CR Jr, Macrander AM (**2015**). Effects of airgun sounds on bowhead whale calling rates: evidence for two behavioral thresholds. PLoS ONE 10(6): e0125720.

Blackwell SB, Nations CS, Thode AM, Kauffman ME, Conrad AS, Norman RG, Kim KH. (**2017**). Effects of tones associated with drilling activities on bowhead whale calling rates. PLoS ONE 12(11): e0188459.

Blackwell, S. B., Conrad, A. S., Thode, A. M., Berchok, C. L., Stafford, K. M., and Kim, K. H. (**in prep**). "Estimating acoustic cue rates in bowhead whales, Balaena mysticetus, during the fall migration".

Buckland, S. T. (**2004**). Advanced distance sampling (Oxford University Press), Section 11.12.

Clark CW, Berchok CL, Blackwell SB, Hannay DE, Jones J, Ponirakis D, Stafford KM. (**2015**). A year in the acoustic world of bowhead whales in the Bering, Chukchi, and Beaufort seas. Progress in Oceanography 136:223–240.

Clarke JT, Brower AA, Christman CL, Ferguson MC. 2014. Distribution and relative abundance of marine mammals in the Northeastern Chukchi and Western Beaufort seas, 2013. Annual report, OCS Study BOEM 2014-018. National Marine Mammal Laboratory, Alaska Fisheries Science Center, NMFS, NOAA, 7600 Sand Point Way NE, F/AKC3, Seattle, WA 98115-6349.

Givens GH, Edmondson SL, George JC, Suydam R, Charif RA, Rahaman A, Hawthorne D, Tudor B, DeLong RA, Clark CW (**2013**). Estimate of 2011 abundance of the Bering-Chukchi-Beaufort Seas bowhead whale population. Paper SC/65a/BRG01 presented to the Scientific Committee of the International Whaling Commission. Korea, June 2013; 2013. 30 p. Available from: https://iwc.int/sc65adocs

Goold, J. C. (**2000**). "A diel pattern in vocal activity of short-beaked common dolphins, Delphinus delphis," Marine Mammal Science **16**, 240-244.

Greene CR, Jr, McLennan MW, Norman RG, McDonald TL, Jakubczak RS, Richardson WJ. (**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.

Harris DV, Miksis-Olds JL, Vernon JA, Thomas L. (**2018**). Fin whale density and distribution estimation using acoustic bearings derived from sparse arrays.

Journal of the Acoustical Society of America 143:2980–2993.

Johnson, M. (**2014**). On-animal methods for studying echolocation in free-ranging animals. Pages 195–229 *in* Surlykke A, Nachtigall PE, Fay RR, Popper AN, Editors. Biosonar. Springer Handbook of Auditory Research 51, DOI 10.1007/978-1-4614-9146-0_7, Springer Verlag, New York.

Kim KH, Norman RG, Burgess WC, Blackwell SB, Greene CR Jr. (**2014**). Acoustic monitoring of bowhead whale migration, autumn 2013. Chapter 2 In Richardson WJ and Kim KH (eds.). 2014. *Monitoring of industrial sounds, seals, and bowhead whales near BP's Northstar Oil Development, Alaskan Beaufort Sea, 2013: Annual Summary Report.* LGL Rep. P1259B. Rep. from LGL Alaska Res. Assoc. (Anchorage, AK), Greeneridge Sciences Inc. (Santa Barbara, CA) and Applied Sociocult. Res. (Anchorage, AK) for BP Explor. (Alaska) Inc., Anchorage, AK.

Marques, T.A., L. Thomas, S.W. Martin, D.K. Mellinger, J.A. Ward, D.J. Moretti, D. Harris, and P.L. Tyack. (**2013**). Estimating animal population density using passive acoustics. Biological Reviews 88:287–309.

Simon M, Johnson M, Tyack P, Madsen PT. (**2009**). Behaviour and kinematics of continuous ram filtration in bowhead whales (*Balaena mysticetus*). Proceedings of the Royal Society B, DOI: 10.1098/rspb.2009.1135.

Stafford KM, Moore SE, Berchok CL, Wiig Ø, Lydersen C, Hansen E, Kalmbach D, Kovacs KM. (**2012**). Spitsbergen's endangered bowhead whales sing through the polar night. Endangered Species Research 18:95–103.

Thode A.M., Kim KH, Blackwell SB, Greene CR Jr, Nations CS, McDonald TL, Macrander AM. (**2012**). "Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys," J. Acoust. Soc. Am **131**:3726–3747.

Thode, A. M., Blackwell, S. B., Seger, K. D., Conrad, A. S., Kim, K. H., and Macrander, A. M. (**2016**). "Source level and calling depth distributions of migrating bowhead whale calls in the shallow Beaufort Sea," J. Acoust. Soc. Am **140**, 4288-4297.

Thode, A. M., Blackwell, S. B., Conrad, A. S., Kim, K. H., and Michael Macrander, A. (**2017**). "Decadal-scale frequency shift of migrating bowhead whale calls in the shallow Beaufort Sea," J. Acoust. Soc. Am **142**, 1482-1502.

Thode, A. M., Blackwell, S. B., Conrad, A. S., Kim, K. H., Marques, T., Thomas, L., Oedekoven, C. S., Harris, D., and Bröker, K. (**in review**). "Roaring vs. repeating: How bowhead whales adjust their call density and source level (Lombard effect) in the presence of natural and seismic airgun survey noise," J. Acoust. Soc. Am. **in review**.

Warren VE, Marques TA, Harris D, Thomas L, Tyack PL, Aguilar de Soto N, Hickmott LS, Johnson MP. (**2017**). "Spatio-temporal variation in click production rates of beaked whales: implications for passive acoustic density estimation," J. Acoust. Soc. Am. **141**:1962–1974.

5. Absolute Abundance: Cue Rate Estimation

Estimating acoustic cue rates in bowhead whales, *Balaena mysticetus*, during their fall migration through the Alaskan Beaufort Sea

by

Susanna B. Blackwell, Aaron M. Thode, Alexander S. Conrad, Megan C. Ferguson, Catherine L. Berchok, Kathleen M. Stafford, Tiago A. Marques, Katherine H. Kim

---

Submitted: *Journal of the Acoustical Society of America (JASA)*, November 2020
Revised: *Journal of the Acoustical Society of America (JASA)*, April 2021
Published: *Journal of the Acoustical Society of America (JASA)*, May 2021

JASA ARTICLE

Estimating acoustic cue rates in bowhead whales, *Balaena mysticetus*, during their fall migration through the Alaskan Beaufort Sea

Susanna B. Blackwell,[1,a)] Aaron M. Thode,[2] Alexander S. Conrad,[1] Megan C. Ferguson,[3] Catherine L. Berchok,[3] Kathleen M. Stafford,[4] Tiago A. Marques,[5,b)] and Katherine H. Kim[1]

[1]*Greeneridge Sciences, Inc., 5266 Hollister Avenue, Santa Barbara, California 93111, USA*

[2]*Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA*

[3]*Marine Mammal Laboratory, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 7600 Sand Point Way Northeast, Seattle, Washington 98115, USA*

[4]*Applied Physics Laboratory, University of Washington, Seattle, Washington 98105, USA*

[5]*Centre for Research into Ecological and Environmental Modelling, School of Mathematics and Statistics, University of St Andrews, The Observatory, Buchanan Gardens, St Andrews, Fife KY16 9LZ, Scotland*

ABSTRACT:

Eight years of passive acoustic data (2007–2014) from the Beaufort Sea were used to estimate the mean cue rate (calling rate) of individual bowhead whales (*Balaena mysticetus*) during their fall migration along the North Slope of Alaska. Calls detected on directional acoustic recorders (DASARs) were triangulated to provide estimates of locations at times of call production, which were then translated into call densities (calls/h/km^2). Various assumptions were used to convert call density into animal cue rates, including the time for whales to cross the arrays of acoustic recorders, the population size, the fraction of the migration corridor missed by the localizing array system, and the fraction of the seasonal migration missed because recorders were retrieved before the end of the migration. Taking these uncertainties into account in various combinations yielded up to 351 cue rate estimates, which summarize to a median of 1.3 calls/whale/h and an interquartile range of 0.5–5.4 calls/whale/h.

© *2021 Acoustical Society of America*. https://doi.org/10.1121/10.0005043

(Received 4 November 2020; revised 30 April 2021; accepted 5 May 2021; published online 26 May 2021)

[Editor: Klaus Lucke] Pages: 3611–3625

I. INTRODUCTION

A. Motivation and general concept

In passive acoustic monitoring (PAM), an acoustic *cue* is an identifiable sound made by an animal of interest. When studying marine mammals, examples of cues include tonal calls, echolocation clicks, and pulsed calls such as feeding buzzes. In their most simple PAM use, these cues indicate the presence of animals within the detection range of a recorder. The *cue rate* expresses the mean number of cues produced per animal per unit time, for example, the average number of whistles produced by a dolphin per hour. If one of the goals of a PAM project is the estimation of animal densities via a cue-based method, then knowledge of cue rates is indispensable (Marques *et al.*, 2013; Warren *et al.*, 2017; Harris *et al.*, 2018). The lack of reliable knowledge on acoustic behavior and relevant cue rates for many species is a factor that prevents broader use of passive acoustic density estimation.

A seemingly straightforward way to obtain cue rates is by the use of acoustic tags that are attached to individual animals for periods of up to a few days. After retrieval of a tag and processing of the acoustic information it contains, cue rates can be calculated for the various types of sounds produced by the individual who carried the tag. With enough deployments on animals of different sex and age classes and during different behavioral states, population-level variable cue rates should be obtainable.

In practice, obtaining cue rates from tags is not easy. Tag deployments require substantial logistical effort, expense, and expertise, so obtaining a sufficient sample size of tagged whales is challenging. Records also need to be long enough to be representative of the normal behavior of the whale, after the initial effects of the tagging operations have subsided (e.g., Warren *et al.*, 2020). Fortunately, due to rapid advances in tag retention, deployments lasting several days are now possible (e.g., Calambokidis *et al.*, 2019). Another issue is that only the cues produced by the tagged individual should be included in the cue-rate calculation. Depending on the types of sounds produced and the species studied, differentiating sounds made by the tagged animal from those by other nearby untagged individuals can be challenging (Johnson, 2014; Goldbogen *et al.*, 2014).

[a)]Electronic mail: susanna@greeneridge.com, ORCID: 0000-0001-7158-6852.

[b)]Also at Centro de Estatística e Aplicações, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Portugal.

https://doi.org/10.1121/10.0005043 JASA

Finally, assumptions must be made about how representative the acoustic behavior of a particular tagged individual is compared to the overall population's acoustic behavior. Bowhead whales (*Balaena mysticetus*) have been tagged with acoustic tags in only a few studies, and not enough to provide information on cue rates. In Simon *et al.* (2009), for example, the deployments were of short duration (<3 h), and no vocalizations were detected from the tagged whales.

Here, we propose a different approach for estimating cue rates in the Bering-Chukchi-Beaufort (BCB) population of bowhead whales during their westward fall migration. The general principle is simple: imagine a population of known size N_{pop} that summers in a well-defined location, such as a bay. In autumn, the population migrates out of the bay, and in doing so it passes over an array of seafloor recorders. Assume that it takes the average whale T_{array} hours to swim across that array and that every whale call produced within the array is detected and counted. The mean cue rate (CR) of this population, in calls per whale per hour for the types of calls produced at that particular time of the year, is calculated as follows:

$$CR = \frac{N_{calls}}{N_{pop} \cdot T_{array}}, \tag{1}$$

where N_{calls} is the total number of detected calls within the array.

In actuality, the procedure described in this paper for estimating BCB bowhead whale cue rates is not as straightforward as the hypothetical example above. First, the summering area of the BCB population is not a bay but a borderless area of the Beaufort Sea. Second, we rely on localized whale calls obtained with five arrays of recorders in an area that only covers a fraction ($f_{corridor}$) of the geographical width of the migration corridor. Furthermore, the hydrophone arrays capture only a fraction ($f_{migration}$) of the entire migration season, in part because not all whales migrate at the same time. As a result, these two additional factors need to be added to Eq. (1) to account for this incomplete spatial and temporal coverage,

$$CR = \frac{(N_{calls} / f_{corridor})}{(N_{pop} \cdot f_{migration}) \cdot T_{array}}. \tag{2}$$

None of these five factors are known or can be estimated with high precision, but by appropriately defining the factors and their ranges, lower and upper bounds for population cue rates during the fall migration can be estimated. Such information is currently lacking for BCB bowheads or any other population of bowhead whales.

B. Bowhead autumn migration

The majority of the BCB population of bowhead whales typically summers in the eastern Beaufort Sea, in areas such as Amundsen Gulf, around Banks Island, and north of the Mackenzie River Delta, Canada (e.g., Moore and Reeves, 1993). Beginning in late August and continuing into October and November, whales travel westward along the North Slope of Alaska, heading for their overwintering grounds in the Bering Sea. Unlike the eastward spring migration, when whales follow open-water leads that are often far from shore, the fall migration corridor in the Beaufort Sea is generally close and parallel to shore, mostly in water depths of 20–50 m (Würsig and Clark, 1993; Moore *et al.*, 2000; Quakenbush *et al.*, 2012; Citta *et al.*, 2015; Clarke *et al.*, 2018). Aerial surveys over many years (e.g., Miller *et al.*, 1996; Clarke *et al.*, 2018) have confirmed the generally westward movement of the migrating whales, but whales will opportunistically continue feeding during the migration. As a result, some individuals may linger or wander, with some whales doubling back to briefly travel eastward (e.g., Harwood *et al.*, 2017).

II. METHODS

Sections II A–II E explain in detail how we obtain estimates for the components of Eq. (2) above, including necessary assumptions and approximations. Section II A deals with N_{calls}, as obtained using passive acoustic recorders over eight field seasons. Section II B explains the methods used to bound $f_{corridor}$, while Sec. II C explains the methods used to bound $f_{migration}$. Section II D addresses the variable N_{pop}, while Sec. II E addresses the array crossing time T_{array}. All of this information is combined to calculate bounds on cue rates in migrating bowhead whales, which are presented in Sec. III.

A. Whale call database (obtaining N_{calls})

Between 2007 and 2014, as part of their exploration activities in the Beaufort Sea, Shell Exploration and Production Company implemented an acoustic monitoring program to study the effects of industrial activities on bowhead whales (see Blackwell *et al.*, 2013; Blackwell *et al.*, 2015; Blackwell *et al.*, 2017; Thode *et al.*, 2012; Thode *et al.*, 2016; Thode *et al.*, 2020). Directional autonomous seafloor acoustic recorders (DASARs) were deployed at five sites (where each site consists of an array of DASARs) in the central Beaufort Sea between Kaktovik and Harrison Bay, Alaska, over an east-west distance spanning 280 km (Fig. 1; latitude range 70.2°–71° N, longitude range 143.1°–150.7° W). Each array was arranged as a grid of equilateral triangles with 7 km spacing between adjacent DASARs. There was some variation between years in the number of DASARs per array. For the calculations presented in this paper, we considered four arrays (sites 2–5) with seven DASARs each and one array (site 1) with three DASARs, as shown in Fig. 1. Site 4 had two different (flipped west to east) configurations over the years, each with seven DASARs: the western configuration (blue + red DASARs in Fig. 1, 2007–2011) or the eastern configuration (red + yellow DASARs in Fig. 1, 2012–2014). Site 2 could not be deployed in 2010 because of pack ice. Note, however, that 2010 deployments took place 2–3 weeks before the onset of the migration (see below), and all years included in this study were considered low-ice years (see National Snow and Ice Data Center, 2021).

JASA
https://doi.org/10.1121/10.0005043

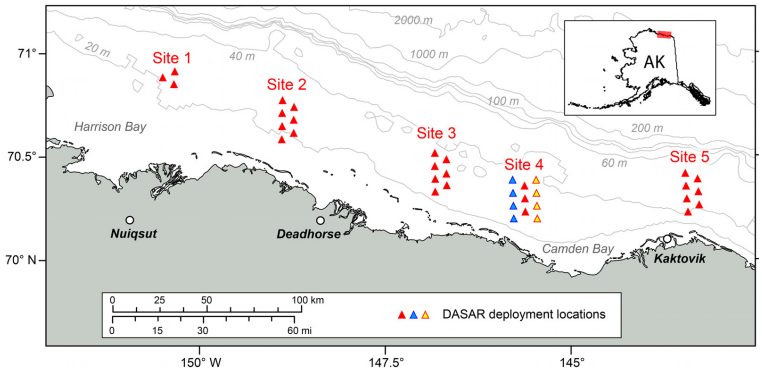


FIG. 1. (Color online) Locations of the five DASAR sites (arrays, sites 1–5) in the Beaufort Sea, 2007–2014. The inset shows the location of the map on the north coast of Alaska. Blackwell *et al.* (2015) includes DASAR deployment positions.

In the analysis that follows, each of the five arrays served the role of a "gate" through which migrating whales traveled during their westward migration. Specifically, we aimed to count calls in a series of adjacent circles covering the area of each array, as illustrated in Fig. 2. Sections II A 1–II A 3 provide details on the time period (the migration) over which the calls were counted as well as how call numbers were tallied.

1. Defining the migration sampling period

Bowhead whales are generally omnipresent in the shallow (<50 m) waters of the Canadian Beaufort Sea during the summer and fall (Harwood *et al.*, 2017; Ferguson *et al.*, 2021). Because they are traveling around during this time, presumably looking for food, there is usually not a clearly identifiable start to the fall migration based on acoustic detections in the arrays. (DASAR deployment dates varied by year, ranging from 30 July to 26 August.) We therefore relied on local traditional knowledge from whale hunters, who place the start of the fall migration, i.e., the time when whales are consistently heading westward, near the end of August or early September (Moore and Reeves 1993; Huntington and Quakenbush, 2009; Clarke *et al.*, 2018). With the goal of being conservative (i.e., miss the fewest migrating whales possible), the start date of the migration was arbitrarily set at 27 August for the easternmost site (site 5, see Fig. 1), 28 August for the central sites (sites 4 and 3), and 29 August for the westernmost sites (sites 2 and 1). This staggering of days accounts for the fact that at a mean speed of 5 km/h (see Sec. II E), a bowhead whale could cover the 280 km between sites 5 and 1 in 56 h, or 2.3 days, though they likely take longer (Olnes *et al.*, 2020). The end of data collection varied between sites and years, occurring between 28 September and 12 October (Table I).

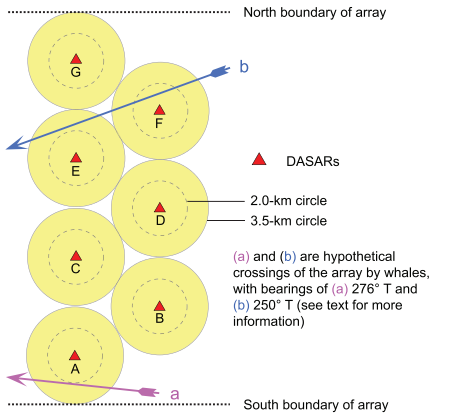


FIG. 2. (Color online) DASAR array of seven adjacent DASAR circles (A–G), each of radius 3.5 km (2-km circles are shown with dashed lines). Dotted lines show the northern and southern boundaries of each array (when all DASARs are functional).

2. Localized call counts at each DASAR

Bowhead whale calls were identified with an automated call detector (Thode *et al.*, 2012), which used triangulation to localize any whale call detected simultaneously on two or more DASARs within the same array. [For more information on localization methods, see Greene *et al.* (2004), Blackwell *et al.* (2007), Blackwell *et al.* (2013), Blackwell *et al.* (2015), and Blackwell *et al.* (2017).] The detectability of calls and the accuracy of localizations decrease with increasing distance from the arrays (Greene *et al.*, 2004; Thode *et al.*, 2012; Thode *et al.*, 2020). Two previous studies (Blackwell *et al.*, 2015; Blackwell *et al.*, 2017) have shown that within 2 km of a DASAR, there was insignificant variation in call detectability with ambient noise conditions. Here, however, calls need to be tallied in larger circles, of radius 3.5 km, to meet the requirements of the analysis, i.e., continuous monitoring of whale calls over the north-south

https://doi.org/10.1121/10.0005043

JASA

TABLE I. Periods of data collection at each site each year. The start date for tallying call localizations, i.e., the start of the migration, was the same across years: 27 August at the easternmost site 5, 28 August at the central sites 4 and 3, and 29 August at the westernmost sites 2 and 1. Site 2 could not be deployed in 2010 due to the presence of pack ice.

Year	Start/End	Site 1	Site 2	Site 3	Site 4	Site 5
All years	Start	29 Aug.	29 Aug.	28 Aug.	28 Aug.	27 Aug.
2007	End	12 Oct.[a]	11 Oct.	8 Oct.	10 Oct.	9 Oct.
2008	End	7 Oct.	6 Oct.	5 Oct.	4 Oct.	2 Oct.
2009	End	4 Oct.	5 Oct.	1 Oct.	2 Oct.	5 Oct.
2010	End	30 Sep.	—	1 Oct.	3 Oct.	4 Oct.
2011	End	5 Oct.	4 Oct.	3 Oct.	1 Oct.	30 Sep.
2012	End	3 Oct.	4 Oct.	6 Oct.	6 Oct.	5 Oct.
2013	End	3 Oct.	2 Oct.	1 Oct.	30 Sep.	29 Sep.
2014	End	28 Sep.	29 Sep.	30 Sep.	1 Oct.	2 Oct.

[a]Whale call localization ended at site 1 on 12 October 2007, but the three DASARs actually continued recording until late November 2007 (see Sec. II C).

(N-S) width of the "gate" (Fig. 2). A circle of radius 3.5 km has about thrice the area of a circle of radius 2 km (~38.5 versus 12.6 km^2). All else being equal, one would therefore expect about 3 times the number of calls in the larger circles. A comparison of the number of calls localized within 3.5-km circles versus 2-km circles, at each site and year combination ($n = 39$) and over the entire season, showed a mean ratio ± standard deviation (SD) of 2.53 ± 0.2 instead of the predicted 3.06, meaning that on average, about 17% of calls were missed due to masking.[1] Consequently, for each site and year combination, the number of localized calls was tallied within 2-km circles around each DASAR, starting on the late August dates listed above and ending when the recorders were retrieved. These values were multiplied by 3.06 to get estimated numbers of whale calls in the 3.5-km circles shown in Fig. 2. This extrapolation assumption is supported by the relatively uniform distribution of whale calls across a DASAR array when viewed over an entire season.[1]

Knowledge of the distribution of hourly call localizations at individual DASARs helps when later interpreting results, so this variability was quantified within 2 km circles.[1] Overall, of 197 640 h of monitoring data at individual DASARs across all years, 78% of sampled hours were devoid of calls. The remaining 22% of sampled hours (42 849 h) included one or more calls, indicating the nearby presence of at least one whale. Of these hours with calls present, 70% included 1–3 calls, 92% had 10 or fewer calls, and over 99% had fewer than 30 calls.

3. Compensating for non-functional or missing DASARs

Adjustments had to be made for the fact that sites did not always include a full complement of functional DASARs. For example, in 2009, DASAR 3 G gave unreliable bearings that could not be used in localizations, and in 2010, DASAR 3 A could not be deployed due to ice.[1] Call densities could be quite different between DASARs on a particular day, but over an entire season, the densities smoothed out.[1] Therefore, if the missing DASAR was the northernmost or southernmost of an array, call counts obtained at the DASAR with the nearest latitude (within the same array) were used. If the missing DASAR was in the middle of an array, the average call count from its northern and southern neighbors was used. Table II shows the estimated numbers of whale calls for each site in each year, as adjusted for masking and missing DASAR data; they total 561 001 calls over the years 2007–2014.

B. Correcting call counts for spatial undersampling: Compensating for the N-S width of the migration corridor (obtaining $f_{corridor}$)

The ~28 km N-S span of adjacent circles extending offshore at each site did not cover the full geographical width of the bowhead migration corridor, and therefore it is likely that not all whales swam through the arrays. To allow estimation of mean cue rate, the call counts shown in Table II thus need to be corrected for this incomplete spatial coverage, to account for calls generated north and south of our defined array boundaries (Fig. 2). A failure to account for these calls would lead to an undercount of the true number of whale calls being produced within the east/west boundaries of a site and a corresponding underestimation of call rate. To correct this bias, the factor $f_{corridor}$ was introduced in Eq. (2). $f_{corridor}$ requires

TABLE II. Number of localized calls at each site, each year, adjusted for calls missed due to masking and for missing data at certain sites and years (see the text), with totals in boldface. Each value is the sum of the call localizations obtained in 3.5-km circles around each DASAR of an array, over the date ranges specified in Table I. Site 2 was not deployed in 2010 due to ice. Sites are listed from west (site 1) to east (site 5).

Array size No. of DASARs	Site 1 3	Site 2 7	Site 3 7	Site 4 7	Site 5 7	Total
2007	5119	11 585	8329	10 682	20 074	**55 790**
2008	6141	30 361	18 290	25 470	19 045	**99 308**
2009	2335	4250	3859	3926	15 802	**30 172**
2010	2111	—	21 975	28 008	14 197	**66 292**
2011	2292	6099	5073	5935	1285	**20 684**
2012	2938	8501	10 906	10 491	12 283	**45 118**
2013	18 834	42 871	32 288	62 213	12 861	**169 067**
2014	5025	11 230	14 673	24 844	18 799	**74 571**
TOTAL	**44 795**	**114 897**	**115 393**	**171 570**	**114 346**	**561 001**

JASA
https://doi.org/10.1121/10.0005043

independent information on the whales' spatial distribution during the autumn migration, so line-transect aerial survey data were used for this purpose. The Appendix includes complete technical details about the analysis, while a summary of the methods is presented below.

The percentage of the migration corridor covered by the DASAR arrays was estimated using bowhead whale sighting and survey effort data from the Aerial Surveys of Arctic Marine Mammals (ASAMM) project (Alaska Fisheries Science Center, 2021; Clarke *et al.*, 2018). From these data, 90 899 km of transect effort and 719 bowhead whale sightings were concurrent with the years of our study (2007–2014). The analysis involved a three-step process: (i) constructing spatially explicit models of bowhead whale relative abundance based on ASAMM bowhead whale sightings from September 2007–2014 (refer to whale sightings and the resultant relative abundances in Fig. 3); (ii) applying the relative abundance model to predict the expected number of bowhead whales in every cell of a 5×5 km grid overlying the migration corridor (see the Appendix); and (iii) using the predicted number of bowhead whales in each cell to compute, for each site, $f_{corridor}$, i.e., the proportion of whales expected to be within the latitudinal range of the site (shown with white polygons in Fig. 3).

The predicted number of whales within each cell is based on the assumption of uniform survey effort throughout the study area, thereby eliminating apparent variability in bowhead whale distribution due only to spatial heterogeneity in survey effort. For each site, $f_{corridor}$ was calculated as the predicted number of whales within the north/south boundaries of the array ($\mu_{TOT,1}$, corresponding to the number of whales within the white polygons of Fig. 3; see the Appendix), divided by the predicted number of whales passing through the full north/south span of the migration corridor at the longitude of the array ($\mu_{TOT,2}$, corresponding to the number of whales within the black dashed lines of Fig. 3). Values of $f_{corridor}$ are shown in Table III; for each site, the call counts in Table II were adjusted using these $f_{corridor}$ values to yield estimated call counts, as if the entire migration corridor had been monitored at each site.

C. Correcting raw call counts for temporal undersampling: Compensating for the duration of the migration season (obtaining $f_{migration}$)

Another bias in the raw call counts is that they are not measured over the entire duration of the migration season. For logistical reasons, the DASAR recorders were removed in late September to early October, right before the onset of ice freeze-up. While the bulk of the bowhead migration is thought to occur from late August to late September, it is known to continue in October and into early November (e.g., Blackwell *et al.*, 2014; Ferguson *et al.*, 2021). For the calculations presented in this document, we need to estimate the fraction of the population missed due to removal of the recorders prior to the end of the migration (i.e., $1 - f_{migration}$). We relied on three different passive acoustic datasets to help us estimate $f_{migration}$.

1. Dataset (1)

Year-round acoustic data were collected by an AURAL-M2 recorder ~87 km north-northwest of site 1 [recorder

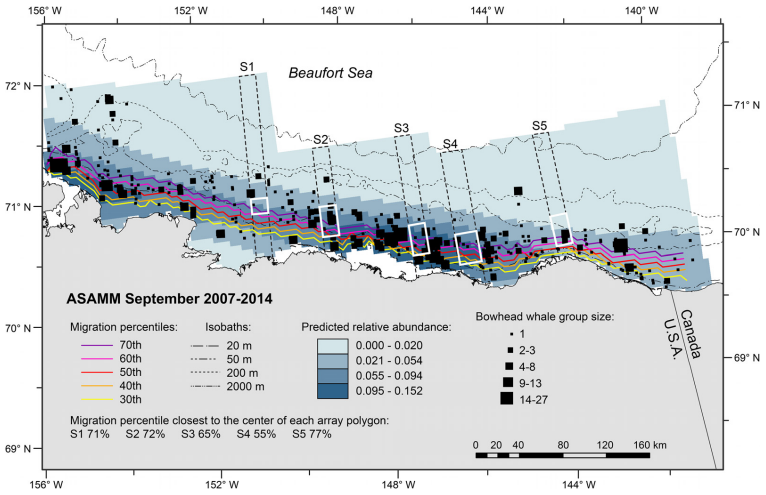


FIG. 3. (Color online) Determination of the proportion of bowhead whales migrating through the area covered by each array. ASAMM bowhead whale sighting data from the month of September in 2007–2014 (black squares) were used to construct spatially explicit models of bowhead whale relative abundance, shown with shaded areas (blue online). The parallel lines show the 30th–70th percentiles of the migration's distance to shore. The white polygons overlay each of the five sites, S1–S5 and have a width of 15 km (sites 1, 2, 3, and 5) or 20 km (for the wider site 4, see Fig. 1). The estimated number of bowhead whales in these white polygons, as a fraction of the estimated number of bowhead whales over the entire width of the migration (as delimited by the dashed black lines), defines $f_{corridor}$ (see the Appendix).

https://doi.org/10.1121/10.0005043

JASA

TABLE III. Call counts at each site for each year, adjusted to include the full N-S width of the migration corridor, with totals in boldface. $f_{corridor}$ is the fraction of the bowhead migration corridor covered by the DASAR arrays during 2007–2014, as determined by aerial surveys (all years combined). For each site and year, call counts from Table II were adjusted using the listed $f_{corridor}$ value.

	Site 1	Site 2	Site 3	Site 4	Site 5	
$f_{corridor}$	0.221	0.523	0.567	0.556	0.531	Total
2007	23 147	22 140	14 681	19 228	37 838	**117 035**
2008	27 768	58 024	32 237	45 845	35 900	**199 774**
2009	10 557	8123	6801	7067	29 786	**62 333**
2010	9547	—	38 733	50 414	26 761	**125 455**
2011	10 363	11 655	8942	10 683	2423	**44 066**
2012	13 282	16 246	19 222	18 884	23 153	**90 787**
2013	85 158	81 930	56 910	111 982	24 243	**360 223**
2014	22 718	21 462	25 862	44 719	35 436	**150 197**
TOTAL	**202 540**	**219 580**	**203 389**	**308 823**	**215 540**	**1 149 872**

152W, University of Washington (UW), blue dot in Fig. 4(a)] for the years 2008–2009 and 2011–2013. Data collection was duty-cycled at 30% (9 min of recordings every half hour). Presence/absence of bowhead calls was assessed for each 9-min file and then expressed as a daily percent time with bowhead calls present. The daily percentage of time with bowhead whales present was then expressed as a cumulative percentage for each of the 5 years [dashed blue lines, Fig. 4(b)] as well as an average across all years [blue dots, Fig. 4(b)].

2. Dataset (2)

Year-round acoustic data were collected by an AURAL-M2 recorder northeast of Utqiaġvik [recorder BF2, National Oceanic and Atmospheric Administration (NOAA), yellow dot in Fig. 4(a)] for the years 2007–2009 and 2011–2014. Different duty cycles were used over the years, with between 27% and 45% coverage. Presence/absence of bowhead calls was assessed for every 10 min of recorded data and expressed as a daily percent time with bowhead calls present. Data were plotted as cumulative percentages for each of the 7 years [yellow lines, Fig. 4(b)] as well as an average across all years [yellow triangles, Fig. 4(b)].

3. Dataset (3)

On 12 October 2007, inclement weather forced us to abandon the three site 1 DASARs [1D, 1E, and 1F, red dot in Fig. 4(a); see the Appendix] and retrieve them in August 2008. (Note that in 2007, these DASARs were part of a larger array, which was retrieved on 12 October.) These DASARs recorded continuously until late November 2007. Whale calls were manually analyzed on these three

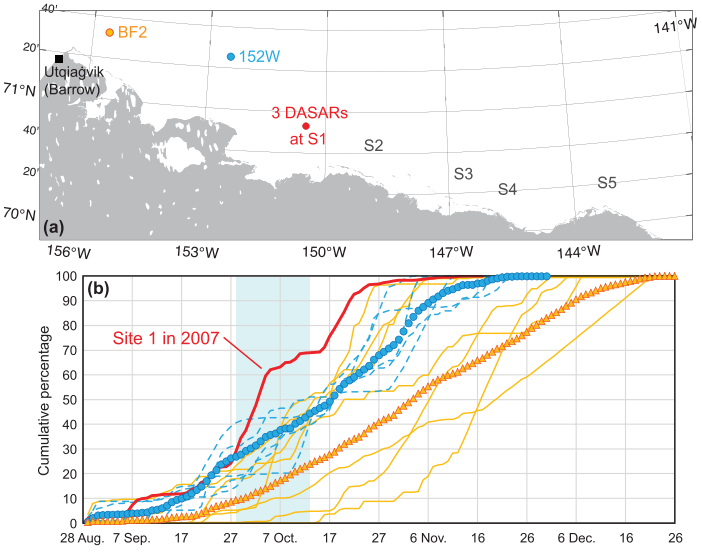


FIG. 4. (Color online) Assessing the timing of the bowhead migration. (a) Map showing the locations of recorders BF2 and 152W, in addition to the three DASARs at site 1, all in relation to the other DASAR sites, indicated with S2–S5. Recorder BF2, northeast of Utqiaġvik (Barrow) is ~195 km from site 1 and ~470 km from site 5, while recorder 152W is ~87 km north-northwest of site 1. (b) Daily cumulative percentage of intervals with bowhead whale detections for recorder BF2 (triangles, light-colored lines), 152W (dots, dashed lines), and the three DASARs at site 1 in 2007 (dark thick line). Lines represent data from individual years, while the symbols represent multi-year averages for each site. The shading indicates the range of retrieval dates for site 1 over the 8 years of the study.

JASA

https://doi.org/10.1121/10.0005043

DASARs for the entire migration. To allow comparison with the other acoustic datasets presented above, presence/absence of bowhead calls was assessed for the first 9 min of every half hour throughout the season, i.e., emulating a 30% duty cycle. Data were plotted as a cumulative percentage, shown with the solid red line in Fig. 4(b).

Site 1 in 2007 is the only site for which we have information on the progression of the migration. The blue shading in Fig. 4(b) shows the range of retrieval dates at site 1 during 2007–2014. When compared to the 2007 (red) cumulative percentage line, this range of retrieval dates corresponds to a fraction of missed migration ($1 - f_{migration}$) of between 70% (for a 28 September retrieval) and 30% (for a 12 October retrieval). Figure 4(b) shows that the timing of the migration at locations 152W and BF2 varies considerably from one year to the next, as is also known from other studies in the area covered by S1–S5 (e.g., Blackwell *et al.*, 2007). Our goal in this paper is to estimate a range of likely cue rates produced by the whales, using the best available information for the unknowns in Eq. (2). Therefore, considering the factors above and our wish not to bound our cue rate estimates too narrowly, we settled on the assumption that the DASAR deployments missed between 25% ($f_{migration} = 0.75$) and 75% ($f_{migration} = 0.25$) of the migrating bowhead whales. These values are used in the calculation of cue rates in Sec. III.

D. Estimating BCB bowhead population size (estimating N_{pop})

Givens *et al.* (2013) estimated the abundance of the BCB population of bowhead whales in 2011 to be 16 892 individuals, from a combination of visual sightings and acoustic locations [95% confidence interval (CI): 15 704–18 928]. They also calculated the rate of increase in the population by combining the 2011 population estimate with a time-series of visual abundance estimates, which started in 1978. As a result, they obtained an annual rate of increase in 3.7% (95% CI: 2.8%–4.7%). The 2011 abundance estimate and 3.7% yearly rate of increase were therefore used to estimate a population size for each year from 2007 to 2014 (Table IV).

E. Whale travel speed and direction (estimating T_{array})

Migrating whales require a certain time to traverse the east-west boundaries of each site, and this amount of time needs to be quantified for the calculation of cue rate. The time to traverse [T_{array} in Eq. (2)] depends on the speed of travel and the pathway (angle) across the array. Speed of travel in migrating bowhead whales in the Beaufort and Chukchi seas has been measured in several studies and has been shown to average about 5 km/h (Mate *et al.*, 2000; Rugh and Cubbage, 1980), with minimum and maximum values of 3.1 km/h (Braham *et al.*, 1980) and 7 km/h (Zeh *et al.*, 1993; Citta *et al.*, 2015). The latter is considered a maximum observed migration speed of bowheads not fleeing vessels or assisted by currents.

TABLE IV. Estimated size of the BCB populations of bowhead whales (N_{pop}) for the years 2007–2014, based on Givens *et al.* (2013). Values for all years but 2011 were estimated assuming an annual rate of increase of 3.7%. The three rightmost columns represent the population sampled when $f_{migration} = 75\%$, 50%, or 25%, where $(1 - f_{migration})$ is the fraction of the migration missed in October and early November, after the DASAR recorders have been retrieved. The columns thus display possible values for the term $N_{pop} f_{migration}$ in Eq. (2).

Year	Estimated abundance (N_{pop})	Population sampled for $f_{migration}$ values of 75%	50%	25%
2007	14 607	10 955	7303	3652
2008	15 147	11 361	7574	3787
2009	15 708	11 781	7854	3927
2010	16 289	12 217	8144	4072
2011	16 892	12 669	8446	4223
2012	17 517	13 137	8758	4379
2013	18 165	13 624	9082	4541
2014	18 837	14 128	9418	4709

To estimate the angular distribution of migration paths across the array, we used ASAMM aerial survey data collected between longitudes 142° W and 152° W (which covers all our sites) and compiled the orientations of 120 groups (of 1–5 individuals) of whales that were seen during the month of September in 2007–2014. Groups were oriented in all cardinal directions, but the mode of the distribution was toward the west, 300°, with a circular mean orientation of 307° T, i.e., somewhat south of northwest. However, roughly 1/3 of the observed orientations had an easterly component, resulting in a wide circular SD of 85°. To model a distribution of migration headings, only orientations toward the west were retained and then used to construct a cumulative empirical distribution. The implications of removing eastern orientations from the heading estimates are examined in Sec. IV.

To estimate bounds on T_{array}, putative crossing paths of whales across the array were then simulated (see two examples in Fig. 2). First, the aforementioned heading distribution was randomly sampled 10 000 times to simulate a set of migration headings that whales would take across an array. For each heading generated, a grid of parallel, putative whale paths was constructed, evenly distributed in space across the entire area of both types of arrays (7- and 3-DASAR), with the paths separated by 100 m and all orientated along the selected heading. For each simulated path, the distance required to cross the array was calculated. (Gaps between circles were skipped, since calls localized there were not counted.) The process was then repeated for another sampled heading. The simulations tested 3 323 681 possible crossing paths for the large (7-DASAR) arrays (sites 2–5) and 2 028 344 possible crossing paths for the site 1 array.[1] Because of the wide spread of headings observed and the irregular shape of the arrays, possible crossing distances varied widely, from a minimum near 0 if a crossing happened to barely intersect the circle surrounding a single DASAR to 27.5 km for sites 2–5 and 14 km for site 1. The

https://doi.org/10.1121/10.0005043

JASA

TABLE V. Estimated values of T_{array} from Eq. (2), using three different swimming speeds (3.1, 4.7, and 7 km/h) combined with a "short" and "long" crossing path through each type of array, as defined in the supplementary material (see Footnote 1). See text for more information.

	Crossing distances			
	7 DASARs (sites 2, 3, 4, and 5)		3 DASARs (site 1)	
	Short, 4.4 km	Long, 27.5 km	Short, 2.8 km	Long, 13.9 km
	Time to cross (h)			
Travel speed				
3.1 km/h	1.4	8.9	0.90	4.5
4.7 km/h	0.93	5.8	0.60	3.0
7 km/h	0.63	3.9	0.40	2.0
Mean	3.6		1.9	
SD	3.3		1.6	
Crossing time				
Long duration (mean + SD)	6.9		3.5	
Short duration (mean – SD)	0.33		0.30	

5th and 95th percentiles of each distribution[1] were used as representative "short" and "long" crossing distances.

Table V combines the three travel speed estimates and the two "short" and "long" crossing distances to yield six crossing times for each type of array, which were then averaged. This method was used instead of simply using the mean or median crossing distance from the simulations in order to establish the broadest feasible spread of crossing durations, which in turn is used to place bounds on the maximum and minimum possible cue rates. The mean crossing time ± SD for sites 2–5 is 3.6 ± 3.3 h, and that for site 1 is 1.9 ± 1.6 h. In addition to these means, the values (mean + SD) and (mean – SD) were used in the cue-rate calculations as the "long duration" and "short duration" crossing times for each type of array (Table V).

III. RESULTS

Cue rates were calculated by combining Eq. (2) with the corrected call counts ($N_{calls}/f_{corridor}$) from Table III, the three estimated sizes of the population available to the DASAR arrays ($N_{pop} \cdot f_{migration}$) from Table IV, and three crossing times (T_{array}) from Table V (bottom section: mean duration, long duration crossing, and short duration crossing). Altogether, these combinations yield 351 different estimates (5 sites × 8 years × 3 population estimates × 3 crossing durations, minus missing data for site 2 in 2010), with a median cue rate of 1.3 calls/whale/h and an interquartile range (IQR) of 0.5–5.4 calls/whale/h. Eighty percent of these estimates lie between 0.3 and 14.5 calls/whale/h.

Data are summarized graphically in Figs. 5 and 6. In Fig. 5, only the two extreme crossing times ("long duration" and "short duration" at the bottom of Table V) were used in the calculation of cue rates. The values in Fig. 5 therefore collectively represent upper and lower boundaries of our estimates, shown as a function of site [Fig. 5(a)] and array crossing time [Figs. 5(b) and 5(c)], for all years combined. We did this to bracket our cue rate estimates as much as possible, despite the uncertainty in several parameters that enter into these estimates.

Figure 6 shows cue rates calculated using the mean array crossing time (Table V: 3.6 h for sites 2–5, 1.9 h for site 1), while illustrating the effect of $f_{migration}$ on the cue rate estimates: for each year and site, cue rates are shown for $f_{migration}$ values of 75% (A), 50% (B), and 25% (C). The thick purple line is placed at the median value for all cue rates shown in the figure ($n = 117$), 0.96 calls/whale/h, and the shaded area shows the IQR, 0.5–1.7 calls/whale/h.

IV. DISCUSSION

The main finding of this analysis, illustrated in Figs. 5 and 6, is that bowhead calling rates during the migration are only, on average, a few calls per whale per hour in the late summer and fall. Table II shows large variations in the numbers of localized whale calls among sites in the same year and from one year to the next, with differences of up to an order of magnitude. There are also substantial uncertainties in the temporal coverage, and, to a lesser extent, the spatial coverage, of the bowhead migration by the DASAR arrays. In addition, between 2007 and 2014, the bowhead population is thought to have increased by 25%–30% (Table IV). Nevertheless, despite these sources of variation and ambiguity, mean or median cue rates (e.g., dots and squares in Fig. 5) are surprisingly consistent for most of the site/year combinations. The overall median cue rate, which includes all 351 estimates, came to 1.3 calls/whale/h, with half the estimates between 0.5 and 5.4 calls/whale/h. In the more conservative summary in Fig. 6, calculated assuming a mean crossing time through the arrays, the median cue rate was somewhat lower, 0.96 calls/whale/h, with 99% of the estimates below 6.6 calls/whale/h.

One might wonder whether this relatively low median cue rate per animal arises from long periods of time during the migration when no whales are present, which are then occasionally punctuated by the passage of whales with an intrinsically higher cue rate. Short intervals with high cue-rate animals, divided over the entire season, could produce an artificially low mean cue rate across the entire season. In actuality, the low cue rates presented in this study are

JASA
https://doi.org/10.1121/10.0005043

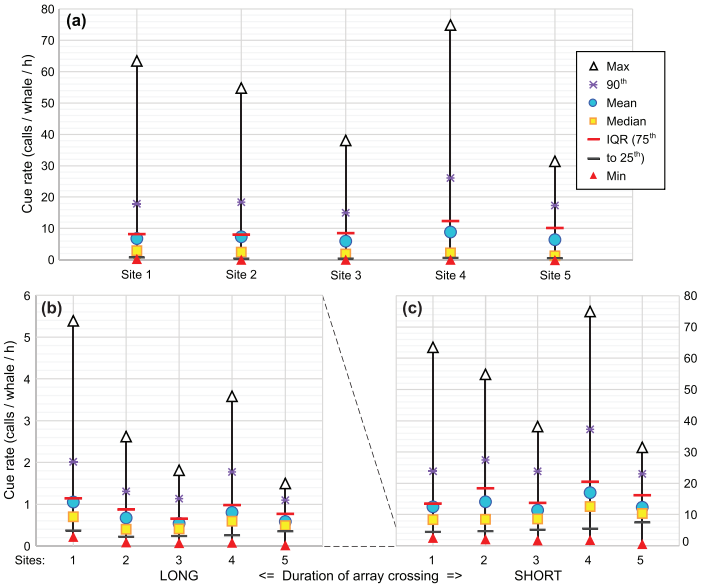


FIG. 5. (Color online) Estimated cue rates, i.e., minima and maxima, means and medians, 90th percentiles, as well as IQRs (25th–75th percentiles), for the five sites and all years combined, summarized by site and by the time taken to cross the arrays. (a) Overall summary as a function of site. Number of estimates included: 48 each for sites 1 and 3–5 and 42 for site 2. (b) and (c) Same data, shown as a function of crossing time. A long crossing time (b) is 6.9 h for sites 2–5 and 3.5 h for site 1, while a short crossing time (c) is 0.33 h for sites 2–5 and 0.30 h for site 1. Number of estimates included for each summary in (b) and (c): 24 each for sites 1 and 3–5 and 21 for site 2. Note the change in the y axis scale between (b) and (c).

consistent with both aerial survey observations and the distribution of hourly localized call counts at individual DASARs.[1] Of all hours with at least one call localized within 2 km of a DASAR, 81% had fewer than five calls. Meanwhile, migrating whales most commonly travel singly or in small groups of a few individuals (e.g., Ashjian *et al.*, 2010; Okkonen *et al.*, 2018). At an average travel speed of 4.7 km/h, it would take up to 0.85 h to cross a circle of radius 2 km, and if calling at a rate of 1.3 calls/whale/h, four transiting whales could thereby produce ~4.4 calls during their crossing (i.e., fewer than 5 calls/h). This simple reality check links aerial survey-based behavioral observations of group sizes with raw call counts at individual DASARs to demonstrate that the cue rates of individual animals must be on the order of only a few per hour.

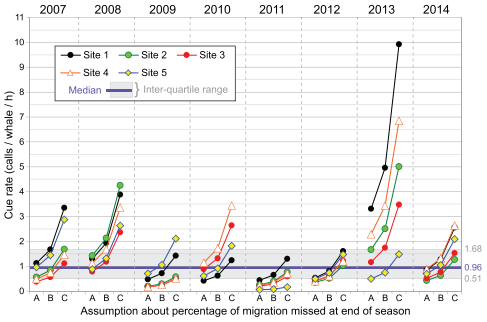


FIG. 6. (Color online) Estimated cue rates as a function of site and year, using the estimated mean crossing time through the arrays (see Table V). For each year and site combination, three cue rates were estimated (x axis labels): (A) assuming 25% of the bowhead migration was missed at the end of the season ($f_{migration} = 0.75$), (B) assuming 50% was missed ($f_{migration} = 0.5$), and (C) assuming 75% was missed ($f_{migration} = 0.25$). See Table IV for details. The dark line shows the median cue rate for the combined 117 estimates included in the figure, and the shading shows the IQR (25th–75th percentiles).

Bowhead cue rates obtained here are similar to other published values for mysticetes. For example, Marques *et al.* (2011) obtained 1.7 calls/whale/h for North Pacific right whales (*Eubalaena japonica*), but, understandably, this value was based on a very small sample size. Martin *et al.* (2013) obtained a cue rate for minke whale (*Balaenoptera acutorostrata*) "boing" calls of 6 calls/whale/h, based on a single individual tracked over ~12 h. Finally, in a study combining visual sightings, acoustic recordings, and infrared camera video, Guazzo *et al.* (2019) obtained average cue rates for migrating Eastern North Pacific gray whales (*Eschrichtius robustus*) of 7.5 calls/whale/day, which corresponds to 0.31 calls/whale/h.

It is important to remember that this first attempt at estimating bowhead whale cue rates during the fall migration is, by necessity, coarse. We have pooled all age and sex classes and all call types produced by the whales. We have also assumed that all whales were in the same behavioral state (migrating). Recent satellite telemetry studies covering 2006–2018 (spanning the years of this study) have shown that 64%–78% of location estimates in the areas of our five sites were classified as "transiting" (Olnes *et al.*, 2020). Therefore, our recordings could also have included sounds from whales that were lingering and/or feeding, presumably

https://doi.org/10.1121/10.0005043 JASA

with different calling rates for these activities (e.g., May-Collado and Quiñones-Lebrón, 2014). We have also not taken cohort segregation into account, whereby different sex and age classes migrate at different times (Koski and Miller, 2009; Huntington and Quakenbush, 2009) and may vocalize at different rates (e.g., mother-calf pairs versus adult females without a calf). Note, however, that because this cohort segregation is not clear-cut and varies between years, our sample will likely have included whales of all sizes and both sexes, including cow-calf pairs. We have also assumed that individual cue rates do not change over the course of the autumn migration (e.g., Guazzo *et al.*, 2019) and do not vary with whale density (e.g., Noad *et al.*, 2017). Finally, in addition to the effects of man-made sounds on calling rates, discussed in Sec. IV B, wind-driven ambient noise levels—a natural factor—also influence call production rates (Thode *et al.*, 2020), but no adjustment was made for this. Fortunately, the statistical distribution of ambient noise levels was consistent across all years.

A. Assessing our uncertainties

Understanding how our uncertainties affect the calculated cue rates helps increase our confidence in the values. For example, whale call counts are the primary factor that the cue rates are based upon, so what would happen if we had missed half of the calls produced within 2 km of all the DASARs—an extremely unlikely scenario? The median cue rate given above would simply double to 2.6 calls/whale/h, still a very low value.

When estimating bounds on T_{array}, aerial survey data were used to estimate the migration headings of the animals. To be consistent with the assumption of Eq. (2), all easterly orientations (0°–180°) were removed when simulating paths across a DASAR array. What would be the impact on the results if instead some whales were allowed to temporarily migrate eastward across the arrays, as shown by the aerial survey data? We note first that if all group orientations are included in the simulations described in Sec. II E, the distribution of single-crossing T_{array} times[1] hardly changes. However, the 120 whale bearings from the aerial surveys show 22% of migrating groups heading toward the eastern quadrant (45°–135°). Under an extreme-case scenario, nearly a quarter of the population (25%), after initially crossing an array from east to west, could double back and travel back across the array west to east, before eventually doubling back once more and crossing the array for a third time. One can continue this logic and deduce that 25% of those whales that doubled back once (or 6.25% of the total population) will double back yet again and end up crossing an array five times in total, etc. If the mean time to cross the array once is T_{array}, then the effective mean crossing time for the entire population T_{eff} becomes

$$T_{eff} = T_{array}\left[\frac{3}{4} + 3*\left(\frac{1}{4}\right) + 5*\left(\frac{1}{4}\right)^2 + 7*\left(\frac{1}{4}\right)^3 \cdots\right] \sim T_{array}(0.75 + 1.22) \sim 2T_{array}, \quad (3)$$

and the mean traversal time across an array for the population effectively doubles ($2T_{array}$). Equation (2) then shows that under this extreme scenario, the initial cue rates presented in the figures would need to be halved. In this case, the cue rates provided in Sec. III can be considered as an upper bound, which again emphasizes the low values of cue rates in migrating bowhead whales. Further analyses of bowhead whale migration directions would help determine whether the scenario discussed here is realistic.

It is important to remember that the most extreme values in the spread of cue rates shown in Figs. 5 and 6 result from a combination of our most extreme assumptions. For example, the maxima in Figs. 5(a) and 5(c) (empty triangles) were all obtained assuming 75% of the migration was missed and array crossing time was extremely short (0.33 and 0.3 h, Table V) in peak calling years, 2013 for sites 1–4 and 2007 for site 5. Similarly, the minima in Figs. 5(a) and 5(b) (red triangles) were all obtained assuming 25% of the migration was missed and array crossing time was extremely long (6.9 and 3.5 h, Table V) in years with low calls counts (2009–2011). Table V shows that for the 7-DASAR and 3-DASAR arrays, respectively, the "long duration" array crossing times were nearly 21 and 12 times greater than the "short duration" times. This variability is reflected directly in the calculated cue rates in Fig. 5, yet 77% and 90% of the 234 estimates included in the overall summary [Fig. 5(a)] are below 10 and 20 calls/whale/h, respectively.

B. Can differences between sites be explained by variable levels of man-made sounds?

During 2007–2014, our study area included a wide range of anthropogenic activities concurrent with data collection by the DASARs. Considering there are known dose-dependent effects of certain man-made sounds—such as airgun pulses and machinery tones—on bowhead whale calling rates (Blackwell *et al.*, 2015; Blackwell *et al.*, 2017; Thode *et al.*, 2020), it is worth investigating whether these external factors may have led to predictable differences in cue rates at different sites within the same year:

- In **2007**, two seismic surveys (using 3147 and 20 in.^{3} arrays) took place between sites 3 and 4, between mid-September and early October. Blackwell *et al.* (2013) and Blackwell *et al.* (2015) showed that proximity to seismic operations represses calling in bowhead whales, while at greater distances, they call more than in the absence of airgun sounds. Considering the relative distances of the sites to the seismic operations, we would expect sites 3 and 4 to have lower calling rates than the other sites, which is what is shown in Fig. 6.
- In **2008**, three seismic surveys (using a variety of arrays or single guns: 3147, 880, 440, and 20 in.^{3}) took place near site 1 and between sites 3 and 4 (Blackwell *et al.*, 2015). The two sites where heightened calling rates would be expected based on received levels of sound from the

JASA
https://doi.org/10.1121/10.0005043

airguns are sites 2 and 5, but only site 2 shows such a trend in Fig. 6.

- In **2009**, there were no seismic surveys near or within our study area. There was, however, a seismic survey in Canadian waters, about 300 km east of site 5, with numerous airgun pulses detected at site 5 (and many fewer at other sites). The higher cue rates at site 5 in 2009 matches our prediction.
- The **2010** deployment season was the only one with plentiful nearshore ice in the DASAR deployment area, particularly west of Prudhoe Bay. As a result, site 2 was not deployed, and site 1 had ice coverage longer than any other site, which could explain the low site 1 cue rates.
- In **2011**, cue rates at all sites were low with little variation between sites. Seismic exploration was present to the north, but very distant. If any effects on calling rates were present, they should have had similar consequences on all sites.
- In **2012**, Shell Exploration and Production Company performed exploratory drilling at Sivulliq, located between sites 3 and 4. Using data collected during these activities, Blackwell *et al.* (2017) showed a clear effect of industrial tones from vessels and other machinery on bowhead whale calling rates. Nevertheless, these hour-to-hour or day-to-day shifts would not be visible in Fig. 6, particularly considering that the largest source of tones was vessels, which during the season repeatedly transited through or near sites 1–4, while other unidentified industrial operations took place near site 5.
- The year **2013** yielded exceptionally high call counts, with more than 8 times the number of call localizations obtained in 2011 and 1.7 times the number of calls obtained in 2008, the second-highest year (Table II; average numbers per site were compared, since only four sites were deployed in 2010). Results from another study (Kim *et al.*, 2014) conducted over the same time period between sites 2 and 3 also showed high call counts in 2013. Additionally, the annual ASAMM aerial survey in the Beaufort Sea sighted high numbers of bowhead whales in 2013 (Clarke *et al.*, 2014). Cue rates for scenario C in 2013 in Fig. 6 (75% of the migration missed at the end of the season) were particularly high for sites 1, 2, and 4. Nevertheless, with such large numbers of calls, it seems unlikely that as much as 75% of the migration was missed that year. It is also possible that the whales were in a different behavioral state, for example, feeding instead of migrating. If that were the case, we would expect whales to be meandering around looking for food, increasing the likelihood of them crossing an array multiple times, which in turn would have the same effect of overestimating cue rates. If we therefore ignore scenario C, the remaining estimates for 2013 (scenarios A and B) are much closer to values obtained in previous years.
- In **2014**, there were no particular industrial or other activities known to have occurred in the vicinity of our study area; obtained cue rates were consistent across sites.

The above assessment suggests specific trends in calling rates as a result of industrial operations, but the comparison remains qualitative.

C. Seasonal specificity of calculated cue rates

The cue rates estimated in this paper are only valid for migrating bowheads in September and early October, when the whales travel westward along the Alaskan Beaufort Sea shelf. Late summer and early fall, during the autumn migration, may be the time of the year when bowhead whales call the least. In the thousands of days of data analyzed since Greeneridge Sciences started using DASAR-based monitoring in the Beaufort Sea in 2001 (e.g., Blackwell *et al.*, 2007), detection of a bowhead call "here and there" has been the norm, while consecutive minutes with many calls have been rare. In 2011–2012 (Blackwell *et al.*, 2014) and 2012–2013 (unpublished), overwintering recorders were deployed in the locations of the DASAR arrays. The data obtained illustrate the changes that occur in bowhead calling at the end of the open-water season. For example, in early November, chorusing was detected, when several whales sang concurrently and nearly continuously for minutes at a time, a situation never encountered in summer recordings. Delarue *et al.* (2009) have reported bowhead song continuously in the Chukchi Sea in November and December, while whales were migrating southward to the Bering Sea, and then in April and May, during their return toward the Chukchi Sea. In Fram Strait, in the North Atlantic, Stafford *et al.* (2018) recorded complex song or call sequences nearly every hour during November to April in 2008–2014—an acoustic detection density very different from that in the summer. Finally, in the spring, as the whales pass Utqiaġvik (Barrow), authors have reported both song and simpler frequency modulation (FM)-sweep calls, as whales transition back to their summer repertoire (Würsig and Clark, 1993; Johnson *et al.*, 2015).

V. SUMMARY AND CONCLUSION

In this paper, a first step was taken toward estimating cue rates in bowhead whales off the North Slope of Alaska during the autumn migration. This is important because cue rates are the vital link between PAM and density estimation. Four independent datasets were combined: DASAR localization data, other PAM presence/absence data, ASAMM aerial survey data, and population estimates from Utqiaġvik spring whale counts. We present cue rates obtained for eight consecutive years but believe the focus should be on the overall summary values, i.e., the median and IQR, 1.3 calls/whale/h and 0.5–5.4 calls/whale/h, respectively. For several of the variables considered, such as the PAM call presence/absence data and the aerial survey data, data from multiple years had to be pooled, despite knowing that both the timing and pathway of the fall migration can vary between years. The amount of exposure to airgun pulses was also quite variable in our study area during 2007–2014. Differences in cue rates at a site over two or more consecutive years could

https://doi.org/10.1121/10.0005043 JASA

therefore simply be due to these external factors. Despite these complexities relating to bowhead behavior, and despite substantial uncertainties in some of our assumptions, it is encouraging that reasonable bounds on cue rates can be obtained through this approach and that these bounds are consistent across five sites and eight years. In the future, cue rates presented here may be used to monitor trends in abundance of the increasing BCB bowhead population.

While median cue rates are a good first step, knowledge of how a species' cue rates vary by season, with sex and age classes, and with behavioral state is fundamental to the application of reliable PAM density estimation. Further work on several variables, including the spread of migration headings and the fraction of the migration season missed, would reduce the spread of these bounds further. A more sophisticated analysis than the one presented here could also be envisioned using a Monte Carlo simulation approach that propagates uncertainties in all the relevant intermediary factors to the final estimates, producing a probability distribution of cue rates in migrating bowhead whales.

ACKNOWLEDGMENTS

The idea of using DASAR data to estimate cue rates was hatched years ago during a discussion between S.B.B. and Christopher S. Nations (formerly at Western EcoSystems Technology, Inc.). We thank Shell Exploration and Production Company for funding the DASAR data collection (2007–2014) and encouraging continued mining of the datasets. We also thank analysts Stephanie L. Grassia, Megan A. Wood, Kristin Otte, and Sara Tennant for help with whale call analyses and the captains and crews of the R/V *Norseman II*, R/V *Alpha Helix*, R/V *Westward Wind*, F/V *Alaskan Enterprise*, F/V *Mystery Bay*, and F/V *Aquila*. Funding for the ASAMM project is from Bureau of Ocean Energy Management (BOEM) Interagency Agreements M07RG13260 and M11PG00033 (C. Monnett, R. Denton, and C. Fairfield). Funding for the NOAA mooring and analysis is from BOEM: CHAOZ–M09PG00016 (C. Monnett and H. Crowley) and ARCWEST–M12PG00021 (C. Monnett, J. Denton, and H. Crowley). Funding for the UW mooring and analysis is from National Oceanographic Partnership Program (NOPP) Grant No. N00014-08-1-0311 and National Science Foundation (NSF) Grant Nos. ARC-0856244 and ARC-855828 to K.M.S. from the Office of Polar Programs of the NSF. We thank J. T. Clarke for providing us with guidance and whale heading information for the years of our study, C. S. Oedekoven and D. Harris for comments on an early version of the manuscript, and two anonymous reviewers for comments that improved the manuscript. T.A.M.'s time was covered by the Living Marine Resources (LMR) ACCURATE project (Contract No. N3943019C2176) and partial support by the Center of Statistics and its Applications (CEAUL) (funded by FCT–Fundação para a Ciência e a Tecnologia, Portugal, through Project No. UIDB/00006/2020). Reference to trade names does not imply endorsement by the National Marine Fisheries Service or NOAA. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect those of NOAA or the Department of Commerce.

APPENDIX: METHOD USED TO ESTIMATE $F_{corridor}$ THE PROPORTION OF THE CORRIDOR COVERED BY THE DASAR ARRAYS

Bowhead whale sighting and survey effort data from the ASAMM project (e.g., Clarke *et al.*, 2018) concurrent with the years of our study (2007–2014) were used to estimate the percentage of the migration corridor covered by the DASARs. This analysis involved a three-step process: (1) constructing spatially explicit models of bowhead whale relative abundance based on ASAMM bowhead whale sightings from September in each year from 2007 to 2014; (2) applying the relative abundance model to predict the expected number of bowhead whales in every cell of a grid overlying the migration corridor; and (3) using the predicted number of bowhead whales in each cell to compute $f_{corridor}$ for each site, the proportion of whales expected to be within the northern and southern boundaries of the site.

This analysis was based on bowhead whale sightings made during transect effort by primary observers (Fig. 3). The analysis did not account for availability or perception bias because we were interested in only the proportion of the whales traveling through the region that were within acoustic detection range of each array. The analysis was conducted in R version 3.6.2 (R Core Team, 2019) using packages *sp* (Pebesma and Bivand, 2005; Bivand *et al.*, 2013), *maptools* (Bivand and Lewin-Koh, 2019), *raster* (Hijmans, 2020), *rgeos* (Bivand and Rundel, 2019), *rgdal* (Bivand, *et al.*, 2019), and *mgcv* (Wood, 2017).

To begin, the migration corridor was partitioned into a 5-km × 5-km grid of cells.[1] This grid resolution was chosen as a compromise between having adequate survey effort and sightings in each cell to construct models and maximizing the spatial resolution of the data.

All geospatial data were projected into an equidistant conic projection [false easting: 0.0; false northing: 0.0; central meridian: –148.0°; latitude of origin: 70.75°; standard parallels: 69.9°, 71.6°; linear unit: meter (1.0)]. Data extracted for each cell included the total number of whales sighted and the projected x and y coordinates of the midpoint of each cell. Bowhead whale relative abundance was modeled as a generalized additive model, parameterized by a Tweedie distribution (Tweedie, 1984; Dunn and Smith, 2005) with a natural logarithmic link function. Negative binomial models were also considered, but examination of model residuals (Ver Hoef and Boveng, 2007) suggested that the Tweedie distribution provided a better fit to the data. The model formula may be represented as

$$\ln\left(E(W_i)\right) = \ln\left(\mu_i\right) = \alpha + s(X_i, Y_i) + \text{offset}(\ln(L_i)),$$

JASA
https://doi.org/10.1121/10.0005043

where W_i is a random variable for the number of individual bowhead whales in cell i, with W_i referring to the associated observations and $E(W_i)$ the expected value (mean) of W_i; μ_i is the number of individual bowhead whales expected to be observed in cell i; α is the intercept; X_i is the projected (equidistant conic) longitude of the midpoint of cell i; Y_i is the projected (equidistant conic) latitude of the midpoint of cell i; $s()$ is the smooth function (Wood, 2017) of location covariates used to describe bowhead whale relative abundance (this function is parameterized in the model-fitting process); and L_i is the length (km) of transect effort in cell i, which was incorporated into the model as a constant (an offset) to account for spatially heterogeneous survey effort throughout the study area.

The proportion of migrating bowhead whales expected to be within the latitudinal range of each array (between the dashed boundary lines in Fig. 2) during September of each year (2007–2014), $f_{corridor}$, was estimated using the spatial model to predict the number of whales in two polygons: (1) a strip 15 km (sites 1, 2, 3, and 5) or 20 km wide (site 4, due to the two configurations; see Fig. 1), centered on the axis of each site, bounded on the north and south by the array boundaries (these areas are shown in Fig. 3 as white polygons), and (2) a strip of the same width as above, centered on the axis of each site, bounded on the north and south by the expected northern and southern limits of the bowhead migration corridor in September. These larger areas are delimited with black dashed lines in Fig. 3, while the migration corridor is depicted in shades of blue. The number of migrating bowheads within each polygon was calculated as (see supplementary material[1])

$$\mu_{TOT,j} = \sum_{i=1}^{n} \frac{a_i}{A_i} \mu_i,$$

where j is the polygon index, $j = 1$ for the strip bounded by the array and $j = 2$ for the strip bounded by the expected bowhead whale migration corridor; n is the total number of cells intersected by polygon j; a_i is the area of cell i contained in polygon j; A_i is the total area of cell i, which is 25 km^2 for all cells; and $\mu_{TOT,j}$ is the expected total number of whales in polygon j.

The predicted number of whales within each cell (μ_i) was based on the assumption of uniform survey effort (constant L_i for all i) throughout the study area. The magnitude of L_i used to predict μ_i does not affect the resulting value of $f_{corridor}$ as long as L_i is constant across all cells, thereby eliminating apparent variability in bowhead whale distribution due only to spatial heterogeneity in survey effort. For each site, $f_{corridor}$ was calculated as $\mu_{TOT,1}/\mu_{TOT,2}$ (Table II).

[1]See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0005043 for ratios of the number of calls localized within 3.5 km versus 2 km of a DASAR, an example distribution of whale calls at a site and year combination, percentage distribution of the (raw) number of localized calls detected within 2 km of DASARs, a table of sites and years with missing data, plots of the distribution of crossing paths across 7-DASAR and 3-DASAR arrays, and additional information pertaining to the calculation of $f_{corridor}$.

Alaska Fisheries Science Center (**2021**). "AFSC/MML: Marine mammal aerial surveys in the Bering, Chukchi, and Beaufort Seas, and Amundsen Gulf, 1979-2019," https://www.fisheries.noaa.gov/inport/item/17338 (Last viewed 17 May 2021).

Ashjian, C. J., Braund, S. R., Campbell, R. G., George, J. C., Kruse, J., Maslowski, W., Moore, S. E., Nicolson, C. R., Okkonen, S. R., Sherr, B. F., Sherr, E. B., and Spitz, Y. H. (**2010**). "Climate variability, oceanography, bowhead whale distribution, and Iñupiat subsistence whaling near Barrow, Alaska," Arctic **63**, 179–194.

Bivand, R. S., Keitt, T., and Rowlingson, B. (**2019**). "rgdal: Bindings for the 'geospatial' data abstraction library," R package version 1.4-8, https://CRAN.R-project.org/package=rgdal (Last viewed 1 March 2021).

Bivand, R., and Lewin-Koh, N. (**2019**). "maptools: Tools for reading and handling spatial objects," R package version 0.9-9, https://CRAN.R-project.org/package=maptools (Last viewed 1 March 2021).

Bivand, R. S., Pebesma, E. J., and Gomez-Rubio, V. (**2013**). *Applied Spatial Data Analysis with R* (Springer, New York).

Bivand, R., and Rundel, C. (**2019**). "rgeos: Interface to geometry engine–open source ('GEOS')," R package version 0.5-2, https://CRAN.R-project.org/package=rgeos (Last viewed 1 March 2021).

Blackwell, S. B., Conrad, A. S., and Thode, A. M. (**2014**). "Characterizing underwater sounds at six locations in the U.S. Beaufort Sea during winter months," Joint Monitoring Program in the Chukchi and Beaufort Seas, 2012, LGL Alaska Final Report P1272-2 for Shell Offshore, Inc., ION Geophysical, Inc., and Other Industry Contributors (National Marine Fisheries Service and U.S. Fish and Wildlife Service, Washington, DC), Appendix C.

Blackwell, S. B., Nations, C. S., McDonald, T. L., Greene, C. R., Jr., Thode, A. M., Guerra, M., and Macrander, A. M. (**2013**). "Effects of airgun sounds on bowhead whale calling rates in the Alaskan Beaufort Sea," Mar. Mamm. Sci **29**, E342–E365.

Blackwell, S. B., Nations, C. S., McDonald, T. L., Thode, A. M., Mathias, D., Kim, K. H., Greene, C. R., Jr., and Macrander, A. M. (**2015**). "Effects of airgun sounds on bowhead whale calling rates: Evidence for two behavioral thresholds," PLoS One **10**, e0125720.

Blackwell, S. B., Nations, C. S., Thode, A. M., Kauffman, M. E., Conrad, A. S., Norman, R. G., and Kim, K. H. (**2017**). "Effects of tones associated with drilling activities on bowhead whale calling rates," PLoS One **12**, e0188459.

Blackwell, S. B., Richardson, W. J., Greene, C. R., Jr., and Streever, B. (**2007**). "Bowhead whale (*Balaena mysticetus*) migration and calling behaviour in the Alaskan Beaufort Sea, autumn 2001–04: An acoustic localization study," Arctic **60**, 255–270.

Braham, H. W., Fraker, M. A., and Krogman, B. D. (**1980**). "Spring migration of the western Arctic population of bowhead whales," Mar. Fish. Rev. **42**, 36–46.

Calambokidis, J., Fahlbusch, J. A., Szesciorka, A. R., Southall, B. L., Cade, D. E., Friedlaender, A. S., and Goldbogen, J. A. (**2019**). "Differential vulnerability to ship strikes between day and night for blue, fin, and humpback whales based on dive and movement data from medium duration archival tags," Front. Mar. Sci. **6**, 543.

Citta, J. J., Quakenbush, L. T., Okkonen, S. R., Druckenmiller, M. L., Maslowski, W., Clement-Kinney, J., George, J. C., Brower, H., Small, R. J., Ashjian, C. J., Harwood, L. A., and Heide-Jørgensen, M. P. (**2015**). "Ecological characteristics of core-use areas used by Bering–Chukchi–Beaufort (BCB) bowhead whales, 2006–2012," Prog. Oceanogr. **136**, 201–222.

Clarke, J. T., Brower, A. A., Christman, C. L., and Ferguson, M. C. (**2014**). "Distribution and relative abundance of marine mammals in the Northeastern Chukchi and Western Beaufort seas, 2013: Annual report," OCS Study BOEM 2014-018 (U.S. Department of the Interior, Anchorage, AK).

Clarke, J. T., Ferguson, M. C., Willoughby, A. L., and Brower, A. A. (**2018**). "Bowhead and beluga whale distributions, sighting rates, and habitat associations in the western Beaufort Sea in summer and fall 2009–16, with comparison to 1982–91," Arctic **71**, 115–138.

Delarue, J., Laurinolli, M., and Martin, B. (**2009**). "Bowhead whale (*Balaena mysticetus*) songs in the Chukchi Sea between October 2007 and May 2008," J. Acoust. Soc. Am. **126**, 3319–3328.

Dunn, P. K., and Smith, G. K. (**2005**). "Series evaluation of Tweedie exponential dispersion model densities," Stat. Comput. **15**, 267–280.

https://doi.org/10.1121/10.0005043

JASA

Ferguson, M. C., Clarke, J. T., Brower, A. A., Willoughby, A. L., and Okkonen, S. R. (**2021**). "Ecological variation in the Western Beaufort Sea," in *The Bowhead Whale Balaena mysticetus: Biology and Human Interactions*, edited by J. C. George and J. G. M. Thewissen (Elsevier, London), pp. 365–379.

Givens, G. H., Edmondson, S. L., George, J. C., Suydam, R., Charif, R. A., Rahaman, A., Hawthorne, D., Tudor, B., DeLong, R. A., and Clark, C. W. (**2013**). "Estimate of 2011 abundance of the Bering-Chukchi-Beaufort Seas bowhead whale population," Report SC/65a/BRG01 to the Scientific Committee of the International Whaling Commission (International Whaling Commission, Cambridge, UK).

Goldbogen, J. A., Stimpert, A. K., DeRuiter, S. L., Calambokidis, J., Friedlaender, A. S., Schorr, G. S., Moretti, D. J., Tyack, P. L., and Southall, B. L. (**2014**). "Using accelerometers to determine the calling behavior of tagged baleen whales," J. Exp. Biol. **217**, 2449–2455.

Greene, C. R., Jr., McLennan, M. W., Norman, R. G., McDonald, T. L., Jakubczak, R. S., and Richardson, W. J. (**2004**). "Directional frequency and recording (DIFAR) sensors in seafloor recorders to locate calling bowhead whales during their fall migration," J. Acoust. Soc. Am. **116**, 799–813.

Guazzo, R. A., Weller, D. W., Europe, H. M., Durban, J. W., D'Spain, G. L., and Hildebrand, J. A. (**2019**). "Migrating eastern North Pacific gray whale call and blow rates estimated from acoustic recordings, infrared camera video, and visual sightings," Sci. Rep. **9**, 12517.

Harris, D. V., Miksis-Olds, J. L., Vernon, J. A., and Thomas, L. (**2018**). "Fin whale density and distribution estimation using acoustic bearings derived from sparse arrays," J. Acoust. Soc. Am. **143**, 2980–2993.

Harwood, L. A., Quakenbush, L. T., Small, R. J., George, J. C., Pokiak, J., Pokiak, C., Heide-Jørgensen, M. P., Lea, E. V., and Brower, H. (**2017**). "Movements and inferred foraging by bowhead whales in the Canadian Beaufort Sea during August and September, 2006–12," Arctic **70**, 161–176.

Hijmans, R. J. (**2020**). "raster: Geographic data analysis and modeling," R package version 3.0.2, http://CRAN.R-project.org/package=raster (Last viewed 1 March 2021).

Huntington, H. P., and Quakenbush, L. T. (**2009**). "Traditional knowledge of bowhead whale migratory patterns near Kaktovik and Barrow, Alaska," Final report to the Barrow and Kaktovik Whaling Captains Associations, the Alaska Eskimo Whaling Commission, ConocoPhillips, and the Minerals Management Service (Alaska Department of Fish and Game, Juneau, AK).

Johnson, H. D., Stafford, K. M., George, J. C., Ambrose, W. G., Jr., and Clark, C. W. (**2015**). "Song sharing and diversity in the Bering-Chukchi-Beaufort population of bowhead whales (*Balaena mysticetus*), spring 2011," Mar. Mamm. Sci. **31**, 902–922.

Johnson, M. (**2014**). "On-animal methods for studying echolocation in free-ranging animals," in *Biosonar*, edited by A. Surlykke, P. E. Nachtigall, R. R. Fay, and A. N. Popper (Springer-Verlag, New York), pp. 195–229.

Kim, K. H., Norman, R. G., Burgess, W. C., Blackwell, S. B., and Greene, C. R., Jr. (**2014**). "Acoustic monitoring of bowhead whale migration, autumn 2013," in *Monitoring of Industrial Sounds, Seals, and Bowhead Whales near BP's Northstar Oil Development, Alaskan Beaufort Sea, 2013: Annual Summary Report*, edited by W. J. Richardson and K. H. Kim (BP Exploration (Alaska) Inc., Anchorage, AK).

Koski, W. R., and Miller, G. W. (**2009**). "Habitat use by different size classes of bowhead whales in the central Beaufort Sea during late summer and autumn," Arctic **62**, 137–150.

Marques, T. A., Munger, L., Thomas, L., Wiggins, S., and Hildebrand, J. A. (**2011**). "Estimating North Pacific right whale *Eubalaena japonica* density using passive acoustic cue counting," Endang. Species Res. **13**, 163–172.

Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Moretti, D. J., Harris, D., and Tyack, P. L. (**2013**). "Estimating animal population density using passive acoustics," Biol. Rev. **88**, 287–309.

Martin, S. W., Marques, T. A., Thomas, L., Morrissey, R. P., Jarvis, S., DiMarzio, N., Moretti, D., and Mellinger, D. K. (**2013**). "Estimating minke whale (*Balaenoptera acutorostrata*) boing sound density using passive acoustic sensors," Mar. Mamm. Sci. **29**, 142–158.

Mate, B. R., Krutzikowsky, G. K., and Winsor, M. H. (**2000**). "Satellite-monitored movements of radio-tagged bowhead whales in the Beaufort and Chukchi seas during the late-summer feeding season and fall migration," Can. J. Zool. **78**, 1168–1181.

May-Collado, L. J., and Quiñones-Lebrón, S. G. (**2014**). "Dolphin changes in whistle structure with watercraft activity depends on their behavioral state," J. Acoust. Soc. Am. **135**, EL193.

Miller, G. W., Elliott, R. E., and Richardson, W. J. (**1996**). "Marine mammal distribution, numbers and movements," in *Northstar Marine Mammal Monitoring Program, 1995: Baseline Surveys and Retrospective Analyses of Marine Mammal and Ambient Noise Data from the Central Alaskan Beaufort Sea* (BP Exploration (Alaska) Inc., Anchorage, AK), pp. 3–72.

Moore, S. E., DeMaster, D. P., and Dayton, P. K. (**2000**). "Cetacean habitat selection in the Alaskan Arctic during summer and autumn," Arctic **53**, 341–480.

Moore, S. E., and Reeves, R. R. (**1993**). "Distribution and movement," in *The Bowhead Whale*, edited by J. J. Burns, J. J. Montague, and C. J. Cowles (Allen Press, Lawrence, KS), pp. 313–386.

National Snow and Ice Data Center (**2021**). https://nsidc.org/ (Last viewed 17 May 2021).

Noad, M. J., Dunlop, R. A., and Mack, A. K. (**2017**). "Changes in humpback whale singing behavior with abundance: Implications for the development of acoustic surveys of cetaceans," J. Acoust. Soc. Am. **142**, 1611–1618.

Okkonen, S. R., Clarke, J. T., and Potter, R. A. (**2018**). "Relationships among high river discharges, upwelling events, and bowhead whale (*Balaena mysticetus*) occurrence in the central Alaskan Beaufort Sea," Deep-Sea Res. Pt. II **152**, 195–202.

Olnes, J., Citta, J. J., Quakenbush, L. T., George, J. C., Harwood, L. A., Lea, E. V., and Heide-Jørgensen, M. P. (**2020**). "Use of the Alaskan Beaufort Sea by bowhead whales (*Balaena mysticetus*) tagged with satellite transmitters, 2006–18," Arctic **73**, 278–291.

Pebesma, E. J., and Bivand, R. S. (**2005**). "Classes and methods for spatial data in R," R News, http://cran.r-project.org/doc/Rnews/ (Last viewed 1 March 2021).

Quakenbush, L., Citta, J., George, J. C., Heide-Jørgensen, M. P., Small, R., Brower, H., Harwood, L., Adams, B., Brower, L., Tagarook, G., Pokiak, C., and Pokiak, J. (**2012**). "Seasonal movements of the Bering-Chukchi-Beaufort stock of bowhead whales: 2006–2011 satellite telemetry results," Report SC/64/BRG1 to the scientific committee of the International Whaling Commission (International Whaling Commission, Cambridge, UK).

R Core Team (**2019**). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/ (Last viewed 1 March 2021).

Rugh, D. J., and Cubbage, J. C. (**1980**). "Migration of bowhead whales past Cape Lisburne, Alaska," Mar. Fish. Rev. **42**(9–10), 46–51.

Simon, M., Johnson, M., Tyack, P., and Madsen, P. T. (**2009**). "Behaviour and kinematics of continuous ram filtration in bowhead whales (*Balaena mysticetus*)," Proc. R. Soc. B **276**, 3819–3828.

Stafford, K. M., Lydersen, C., Wiig, Ø., and Kovacs, K. M. (**2018**). "Extreme diversity in the songs of Spitsbergen's bowhead whales," Biol. Lett. **14**, 20180056.

Thode, A. M., Blackwell, S. B., Conrad, A. S., Kim, K. H., Marques, T., Thomas, L., Oedekoven, C. S., Harris, D., and Bröker, K. (**2020**). "Roaring and repetition: How bowhead whales adjust their call density and source level (Lombard effect) in the presence of natural and seismic airgun survey noise," J. Acoust. Soc. Am. **147**, 2061–2080.

Thode, A. M., Blackwell, S. B., Seger, K. D., Conrad, A. S., Kim, K. H., and Macrander, A. M. (**2016**). "Source level and calling depth distributions of migrating bowhead whale calls in the shallow Beaufort Sea," J. Acoust. Soc. Am. **140**, 4288–4297.

Thode, A. M., Kim, K. H., Blackwell, S. B., Greene, C. R., Jr., Nations, C. S., McDonald, T. L., and Macrander, A. M. (**2012**). "Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys," J. Acoust. Soc. Am. **131**, 3726–3747.

Tweedie, M. C. K. (**1984**). "An index which distinguishes between some important exponential families," in *Statistics: Applications and New Directions: Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, edited by J. K. Ghosh and J. Roy (Indian Statistical Institute, Calcutta, India), pp. 579–604.

Ver Hoef, J. M., and Boveng, P. L. (**2007**). "Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data?," Ecology **88**, 2766–2772.

Warren, V. E., Marques, T. A., Harris, D., Thomas, L., Tyack, P. L., Aguilar de Soto, N., Hickmott, L. S., and Johnson, M. P. (**2017**). "Spatio-temporal variation in click production rates of beaked

JASA
https://doi.org/10.1121/10.0005043

whales: Implications for passive acoustic density estimation," J. Acoust. Soc. Am. **141**, 1962–1974.
Warren, V. E., Miller, P. J. O., and Tyack, P. L. (**2020**). "Short-term responses of sperm whales *Physeter macrocephalus* to the attachment of suction cup tags," Mar. Ecol. Prog. Ser. **645**, 219–234.
Wood, S. N. (**2017**). *Generalized Additive Models: An Introduction with R* (CRC, Boca Raton, FL).
Würsig, B., and Clark, C. (**1993**). "Behavior," in *The Bowhead Whale*, edited by J. J. Burns, J. J. Montague, and C. J. Cowles (Allen Press, Lawrence, KS), pp. 157–199.
Zeh, J. E., Clark, C. W., George, J. C., Withrow, D., Carroll, G. M., and Koski, W. R. (**1993**). "Current population size and dynamics," in *The Bowhead Whale*, edited by J. J. Burns, J. J. Montague, and C. J. Cowles (Allen Press, Lawrence, KS), pp. 409–489.

Appendix: Double-Difference Tracking

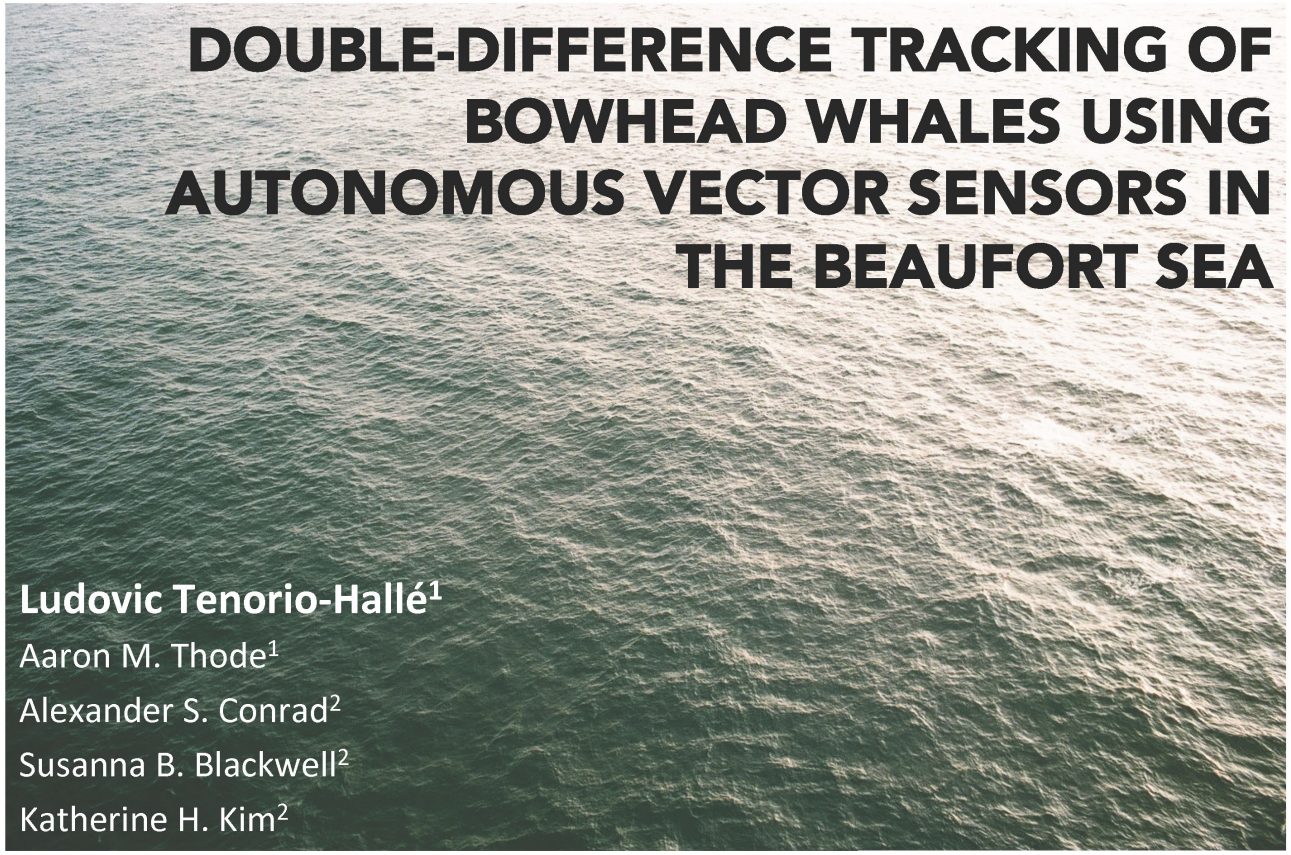
Double-difference tracking of bowhead whales using autonomous vector sensors in the Beaufort Sea

by

Ludovic Tenorio-Hallé, Aaron M. Thode, Alexander S. Conrad, Susanna B. Blackwell, Katherine H. Kim

---

Submitted: *Acoustical Society of America (ASA) 176th Meeting*, May 2018
Presented: *Acoustical Society of America (ASA) 176th Meeting*, November 2018


DOUBLE-DIFFERENCE TRACKING OF BOWHEAD WHALES USING AUTONOMOUS VECTOR SENSORS IN THE BEAUFORT SEA
Ludovic Tenorio-Hallé[1]
Aaron M. Thode[1]
Alexander S. Conrad[2]
Susanna B. Blackwell[2]
Katherine H. Kim[2]
SCRIPPS INSTITUTION OF OCEANOGRAPHY
1
2
GREENERIDGE SCIENCES

OUTLINE

BACKGROUND:

- Data set
- Vector sensors & current localization approach
- Double-difference methods

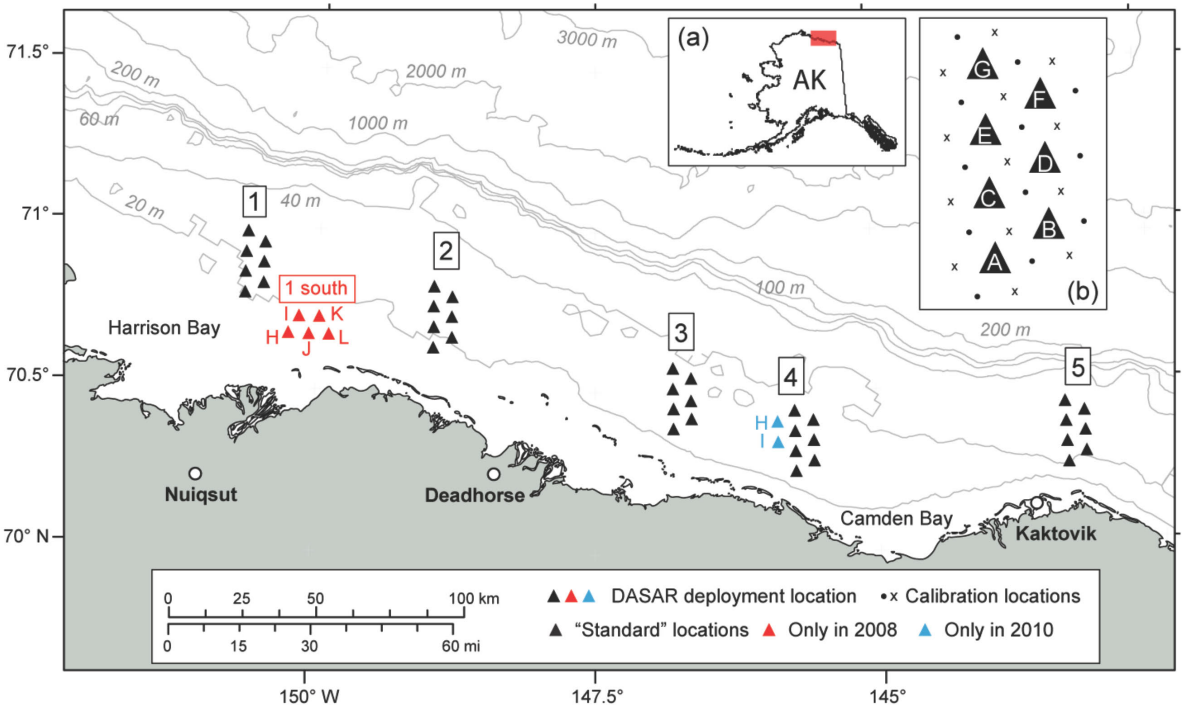
THEORY:

- Time double-difference
- Bearing double-difference

RESULTS:

- Simulations
- Calibration signals (real data)
- Whale track preliminary results (real data)

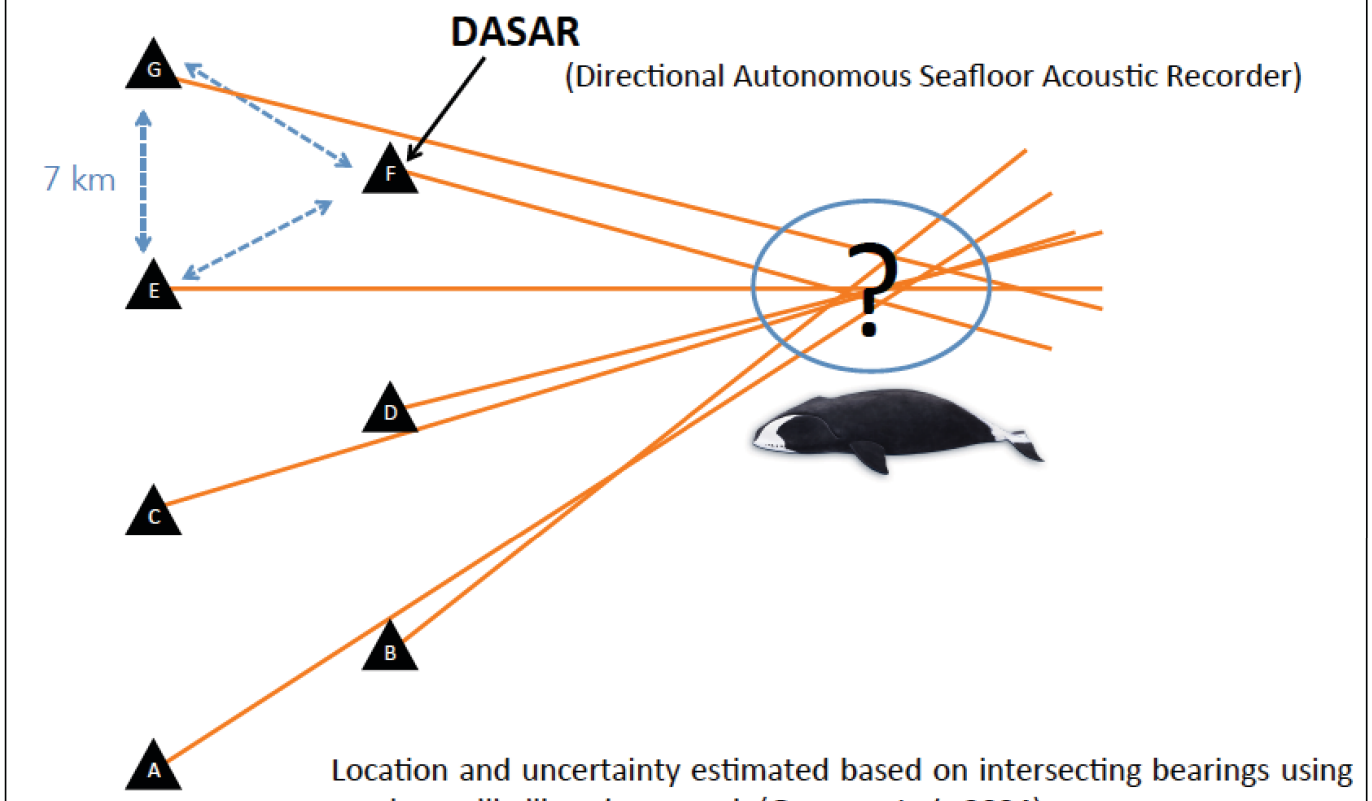
Passive acoustic monitoring during fall migration (2007-2014)



(Blackwell et al., 2015)

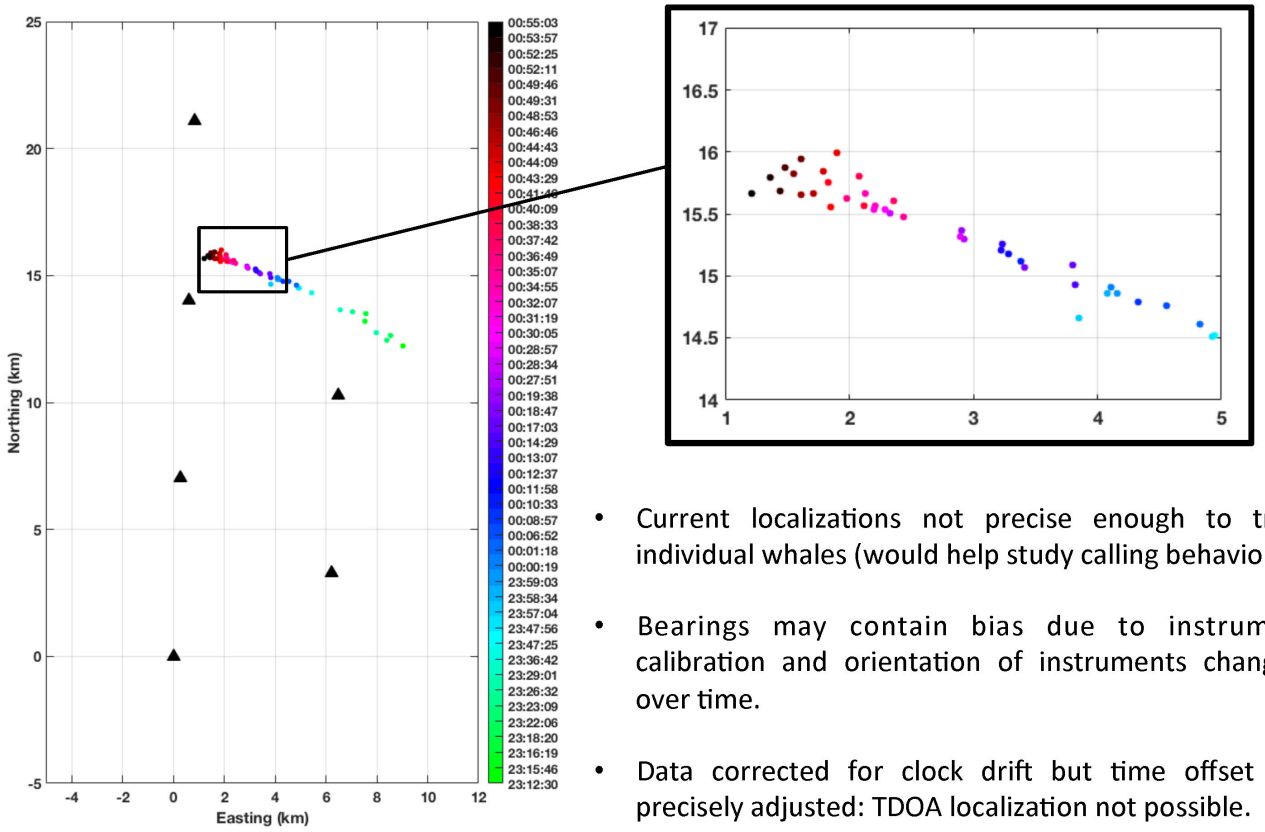
BACKGROUND

Current localization method using vector sensors



BACKGROUND

Current localizations produce low resolution tracks



- Current localizations not precise enough to track individual whales (would help study calling behavior).
- Bearings may contain bias due to instrument calibration and orientation of instruments changing over time.
- Data corrected for clock drift but time offset not precisely adjusted: TDOA localization not possible.

BACKGROUND

Double-difference methods

- Used in seismology **to improve relative location of earthquakes** in the presence of **systemic errors** (i.e. timing error due to clock offsets and/or unmodeled velocity structure from two nearby earthquakes).

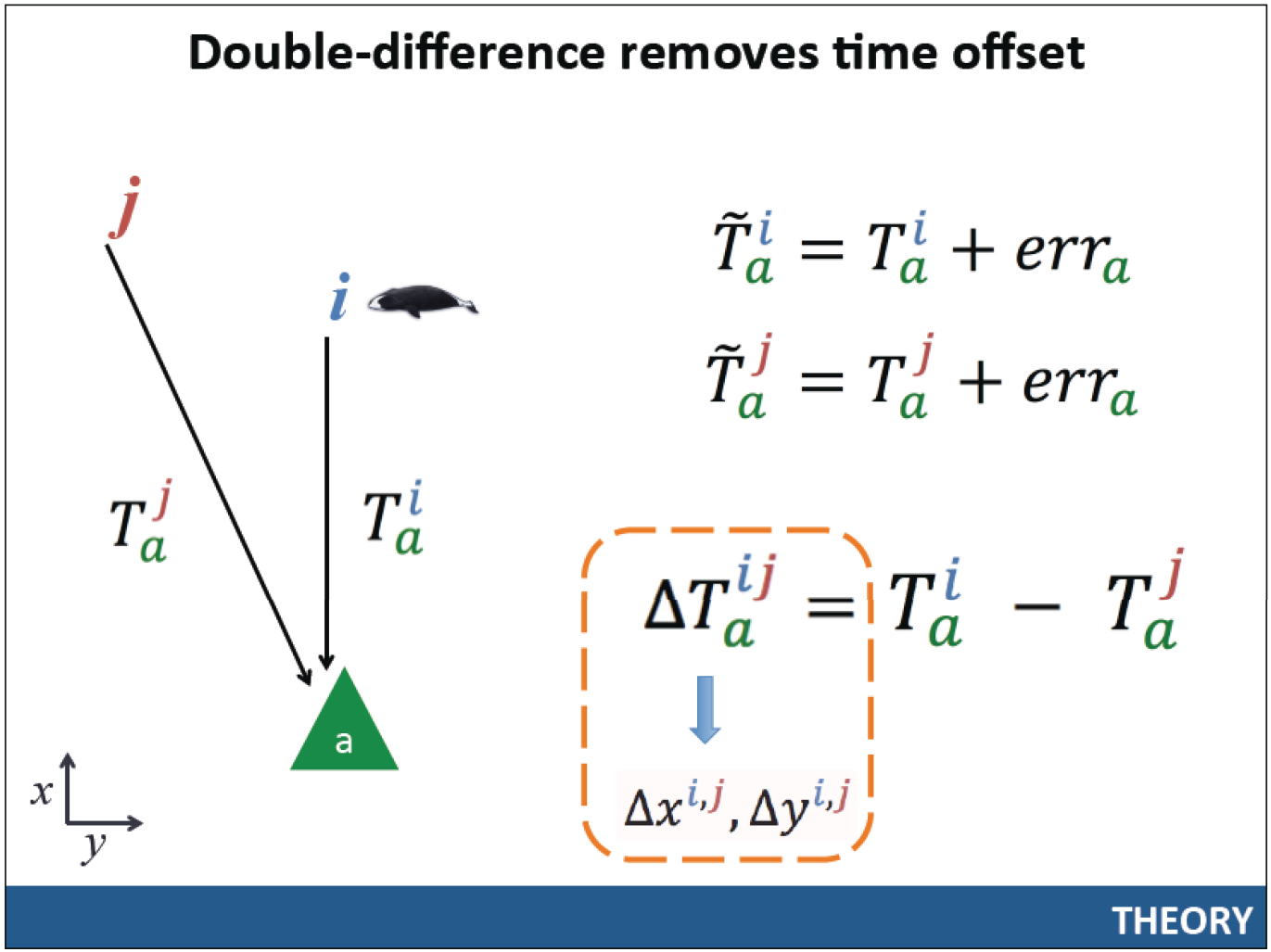
(Waldhauser & Ellsworth, 2000)

For pair of seismic events:

Change in travel time ⟺ **Change in position**

- OBJECTIVE: implement double-difference approach to **vector sensors** in order to improve resolution of **bowhead whale call localization**.

BACKGROUND


Double-difference removes time offset
j
i
a
x
y
T_a^j
T_a^i
$\tilde{T}_a^i = T_a^i + err_a$
$\tilde{T}_a^j = T_a^j + err_a$
$\Delta T_a^{ij} = T_a^i - T_a^j$
$\Delta x^{i,j}, \Delta y^{i,j}$
THEORY

Double-difference equation: Time

(Waldhauser & Ellsworth, 2000)

Model Measurements Model

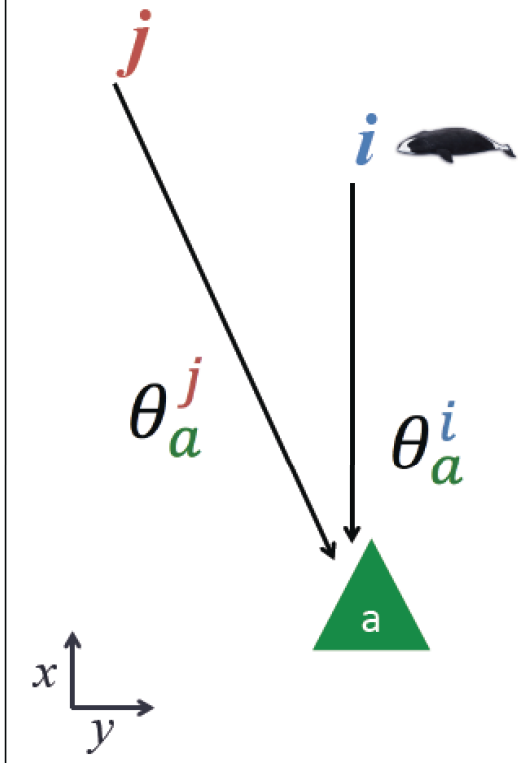
$$\frac{\partial T_a^i}{\partial \boldsymbol{m}} \Delta \boldsymbol{m}^i - \frac{\partial T_a^j}{\partial \boldsymbol{m}} \Delta \boldsymbol{m}^j = \underbrace{\left(T_a^i - T_a^j\right)^{meas} - \left(T_a^i - T_a^j\right)^{mod}}_{\text{Residual}}$$

Adjustments : $\Delta \boldsymbol{m}^i = (\Delta x^i, \Delta y^i, \Delta \tau^i)$

Initial modeled positions: bearing triangulation

THEORY

Double-difference removes bearing bias



$$\tilde{\theta}_a^i = \theta_a^i + err_a$$

$$\tilde{\theta}_a^j = \theta_a^j + err_a$$

$$\Delta\theta_a^{ij} = \theta_a^i - \theta_a^j$$

$$\Delta x^{i,j}, \Delta y^{i,j}$$

(Chen *et al.*, 2016)

THEORY

Double-difference equation: Bearing

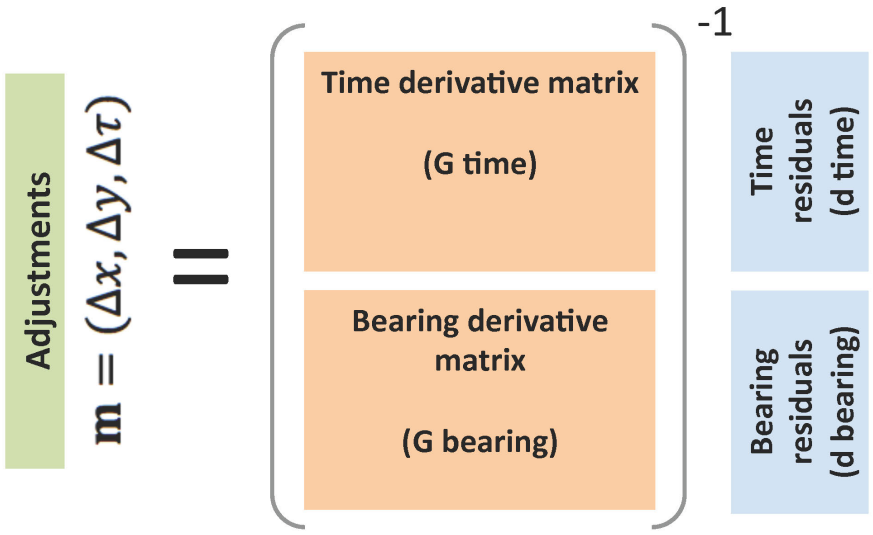
$$\frac{\partial \theta_a^i}{\partial \boldsymbol{m}} \Delta \boldsymbol{m}^i - \frac{\partial \theta_a^j}{\partial \boldsymbol{m}} \Delta \boldsymbol{m}^j = \left(\theta_a^i - \theta_a^j\right)^{meas} - \left(\theta_a^i - \theta_a^j\right)^{mod}$$

$$\Delta \boldsymbol{m}^{\boldsymbol{i}} = (\Delta x^i, \Delta y^i)$$

THEORY

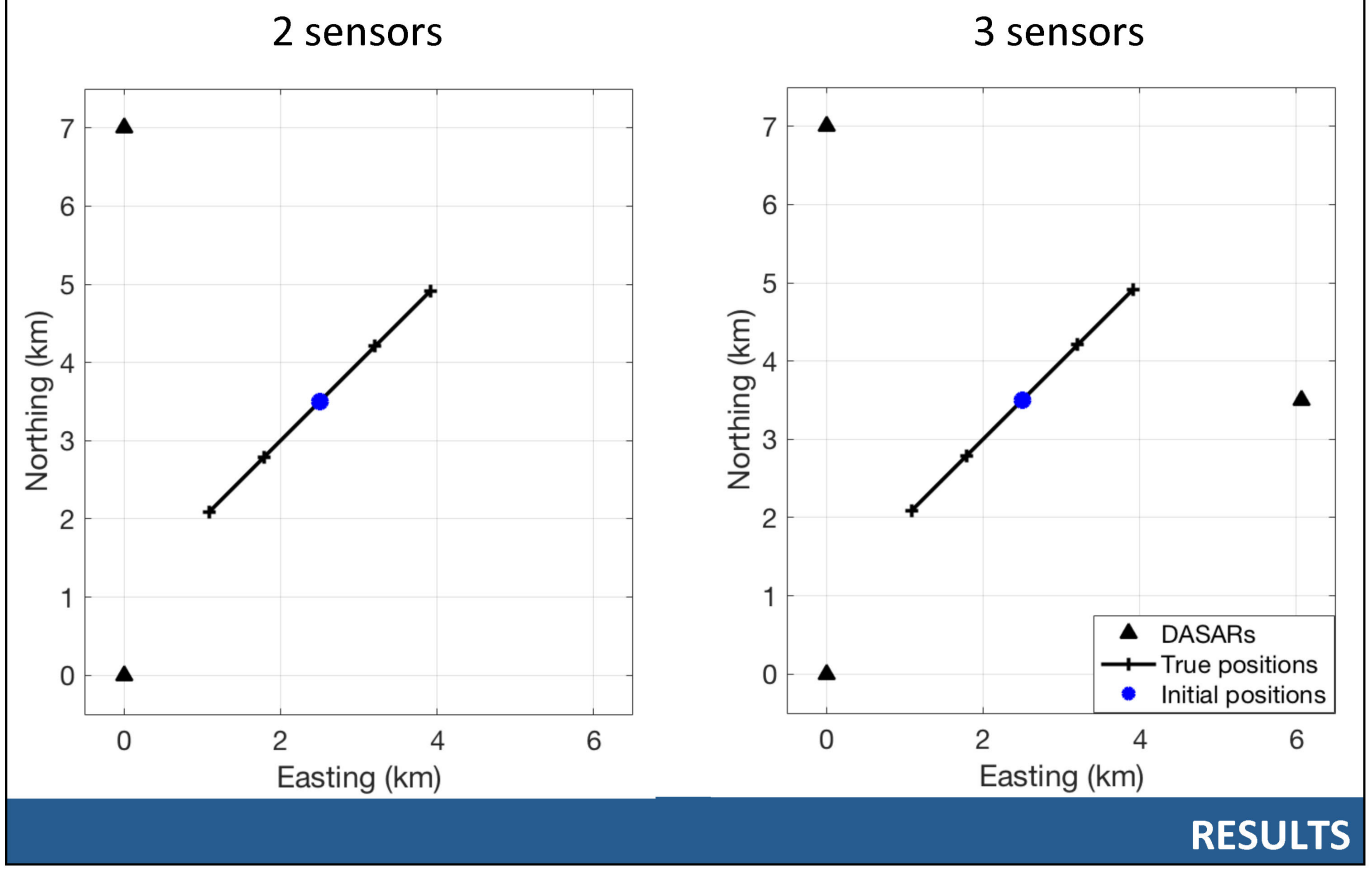
Double-difference equations

$$\mathbf{Gm} = \mathbf{d} \longrightarrow \mathbf{m} = \boldsymbol{G}^{-1}\mathbf{d}$$

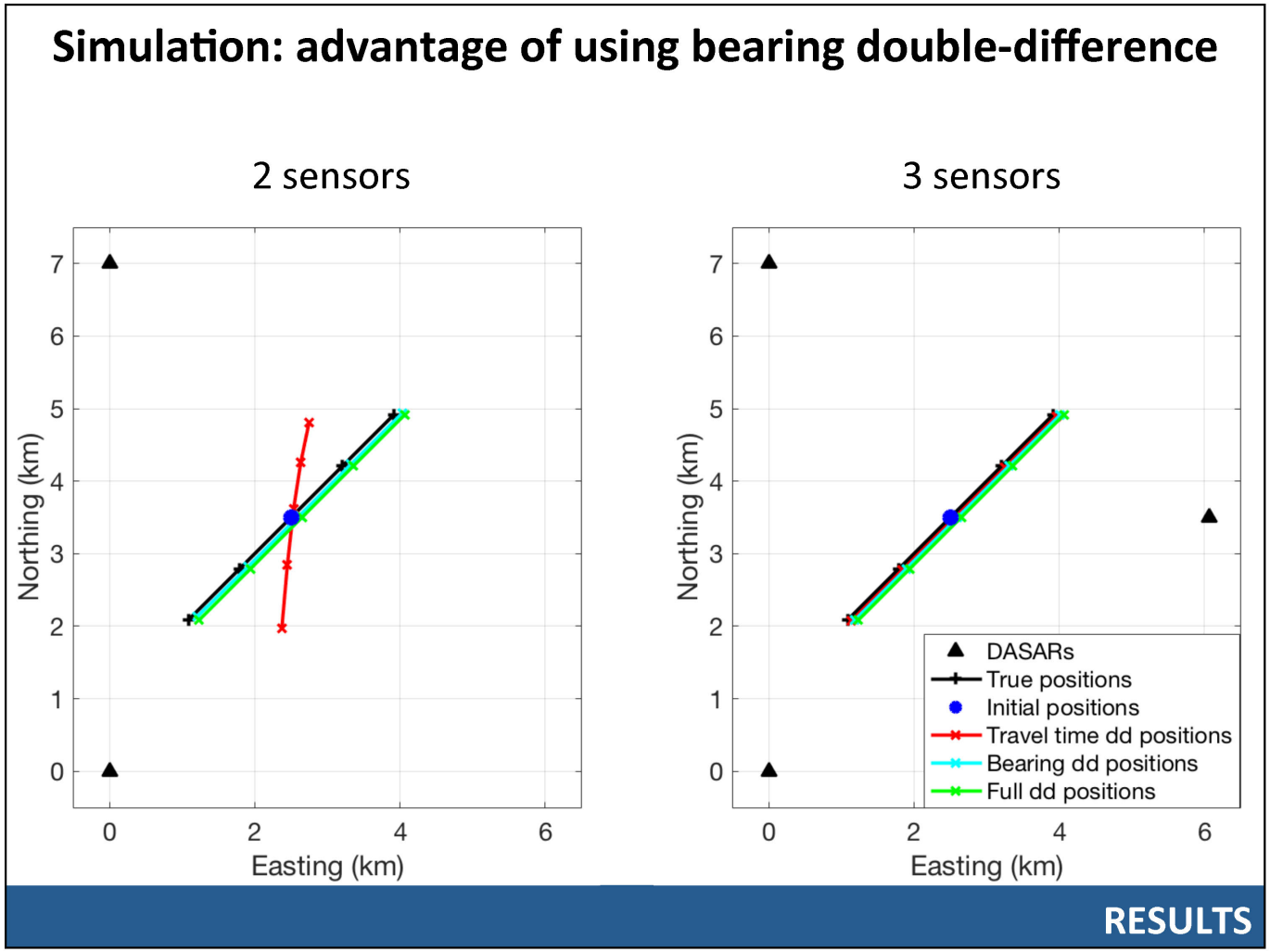


THEORY

Simulation: advantage of using bearing double-difference



Simulation: advantage of using bearing double-difference

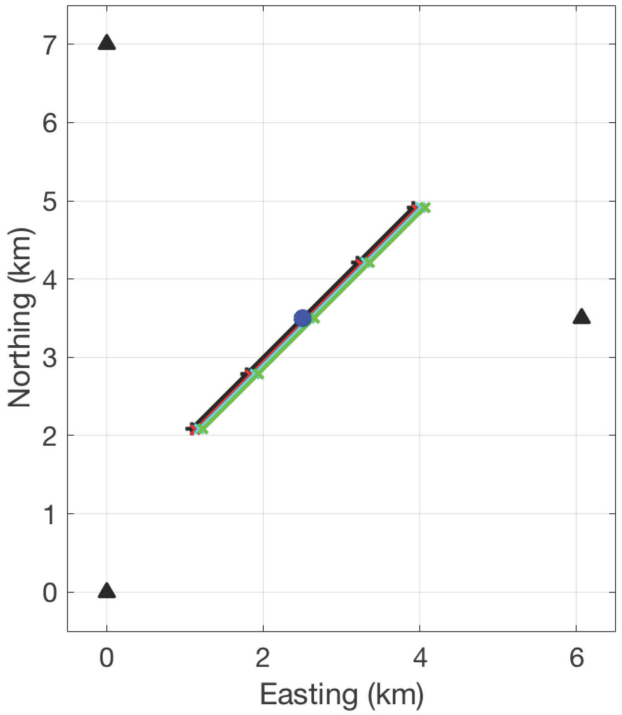


RESULTS

Simulation: Double-difference removes T & θ biases

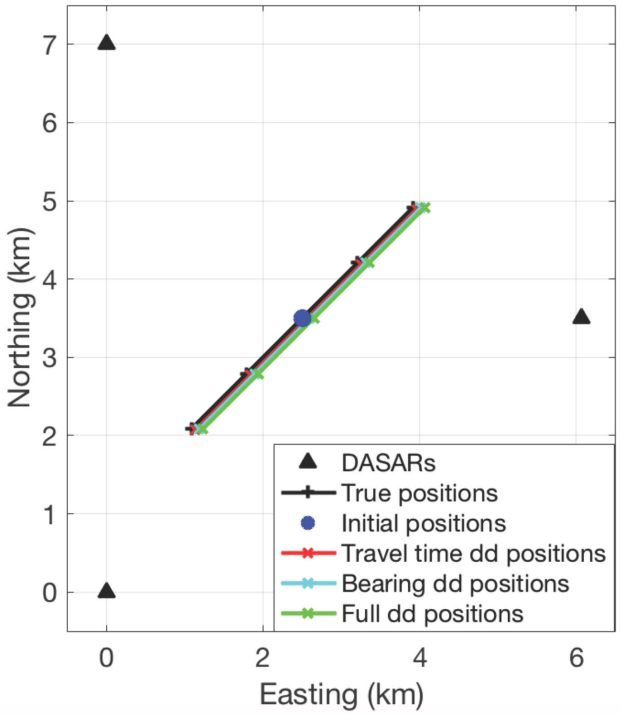
Biases:

$\sigma_T = 0.1s \quad \sigma_\theta = 1°$



Biases:

$\sigma_T = 0.5s \quad \sigma_\theta = 5°$

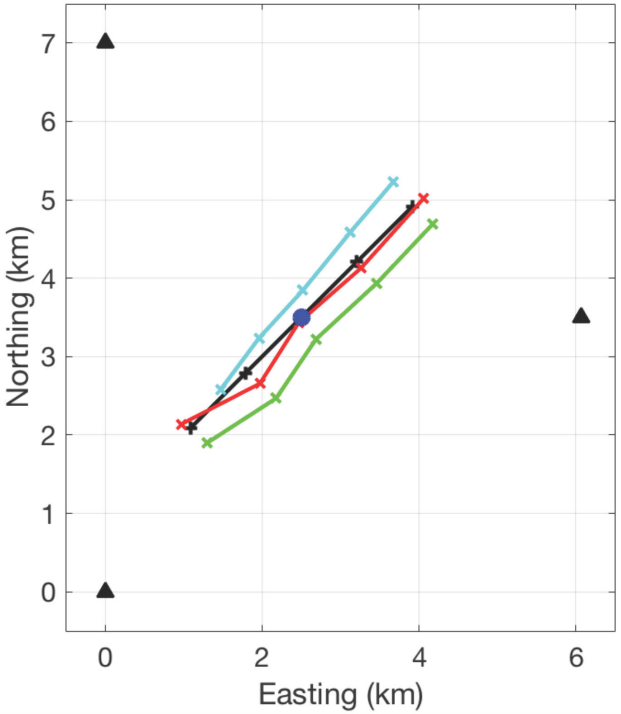


RESULTS

Simulation: effect of sporadic error on double-difference

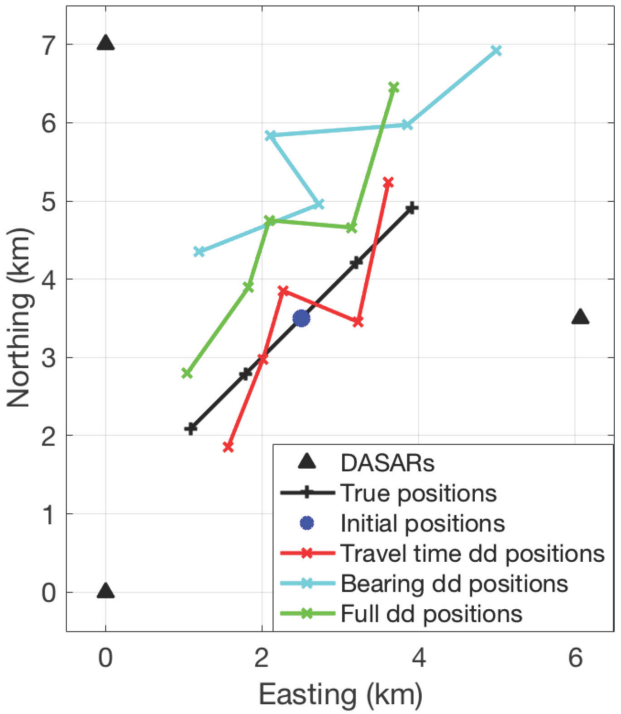
Sporadic error:

$\sigma_T = 0.1s \quad \sigma_\theta = 1°$



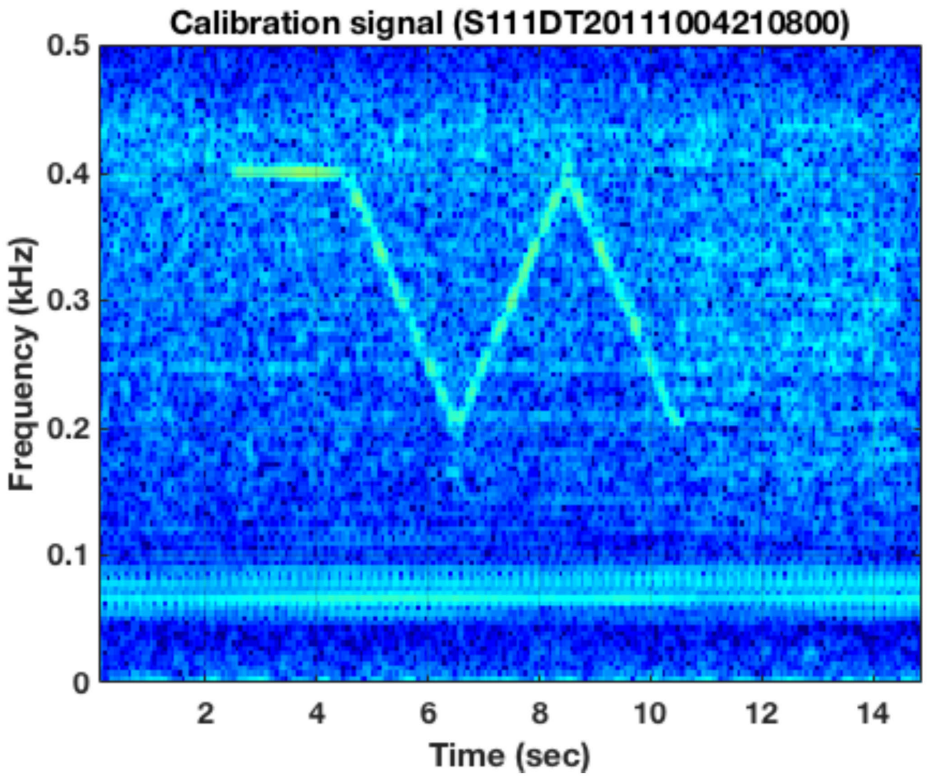
Sporadic error:

$\sigma_T = 0.5s \quad \sigma_\theta = 5°$



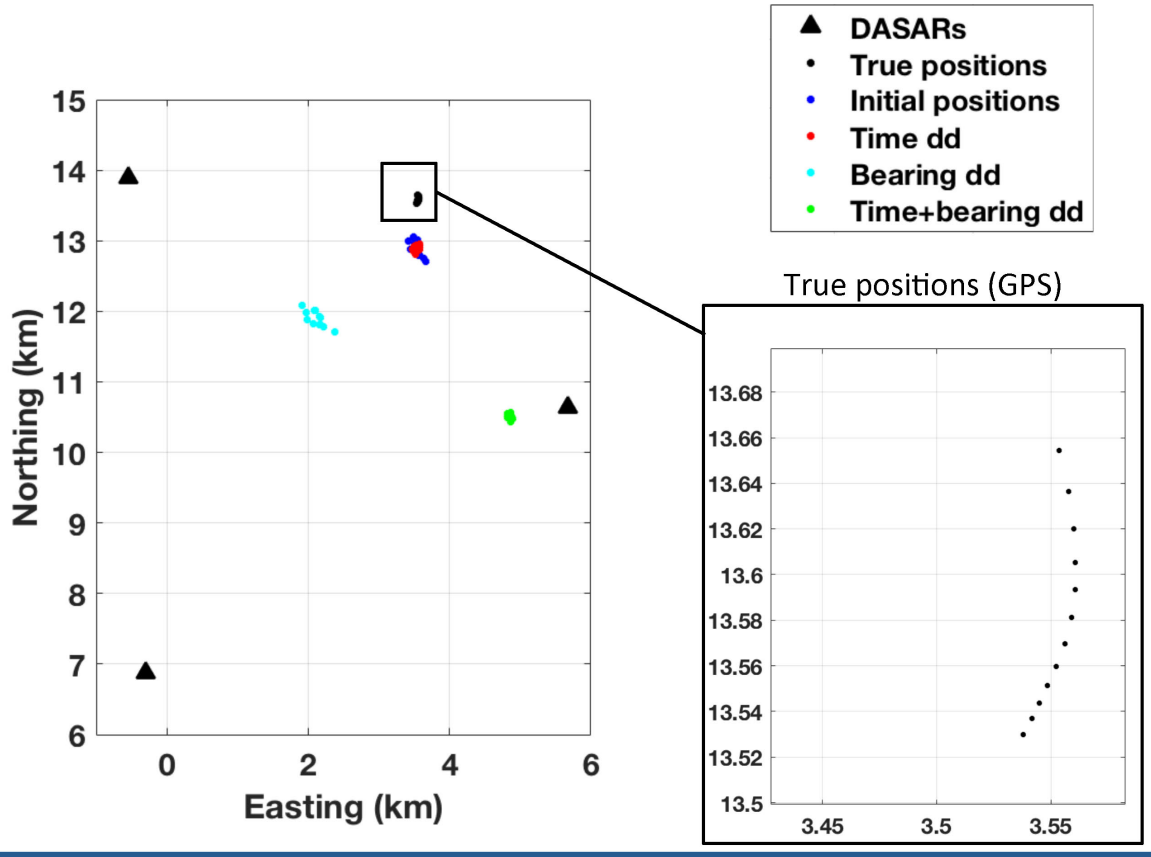
RESULTS

Calibration signals



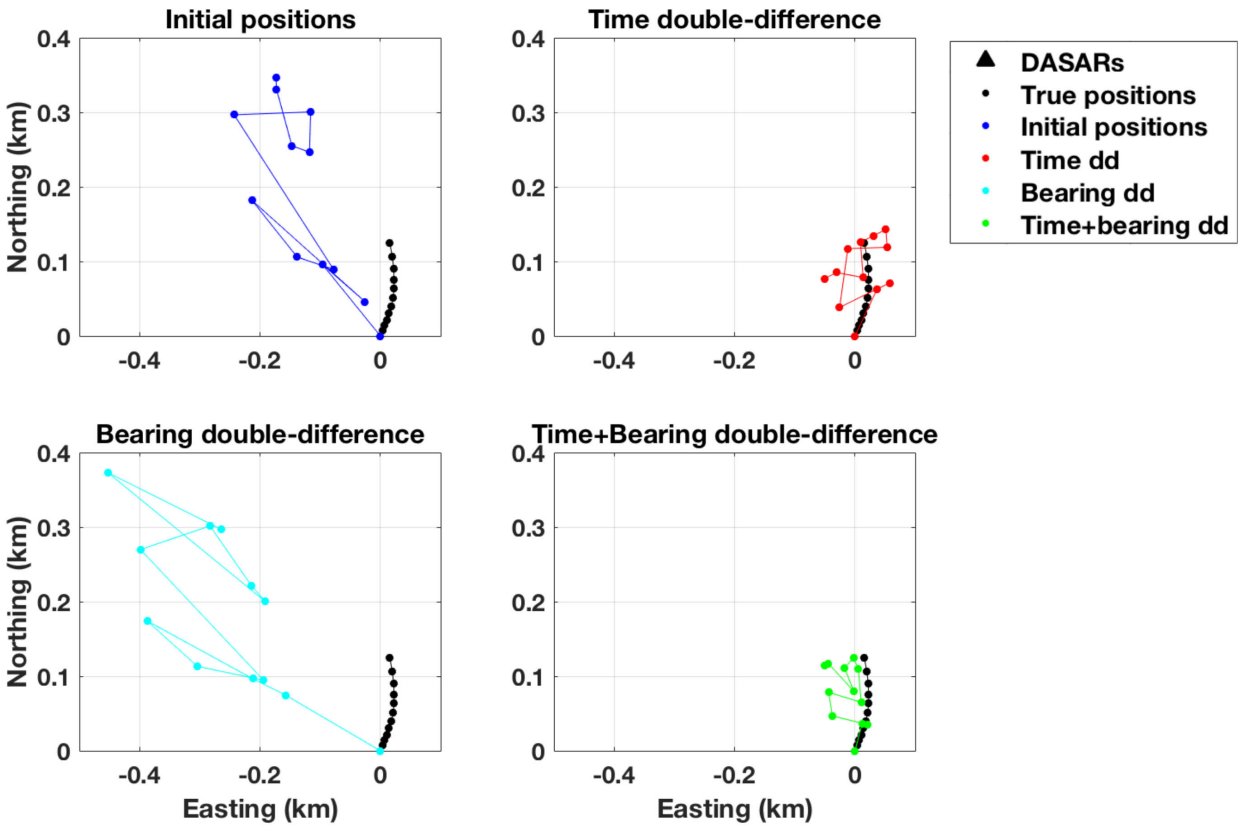
RESULTS

Calibration signals



RESULTS

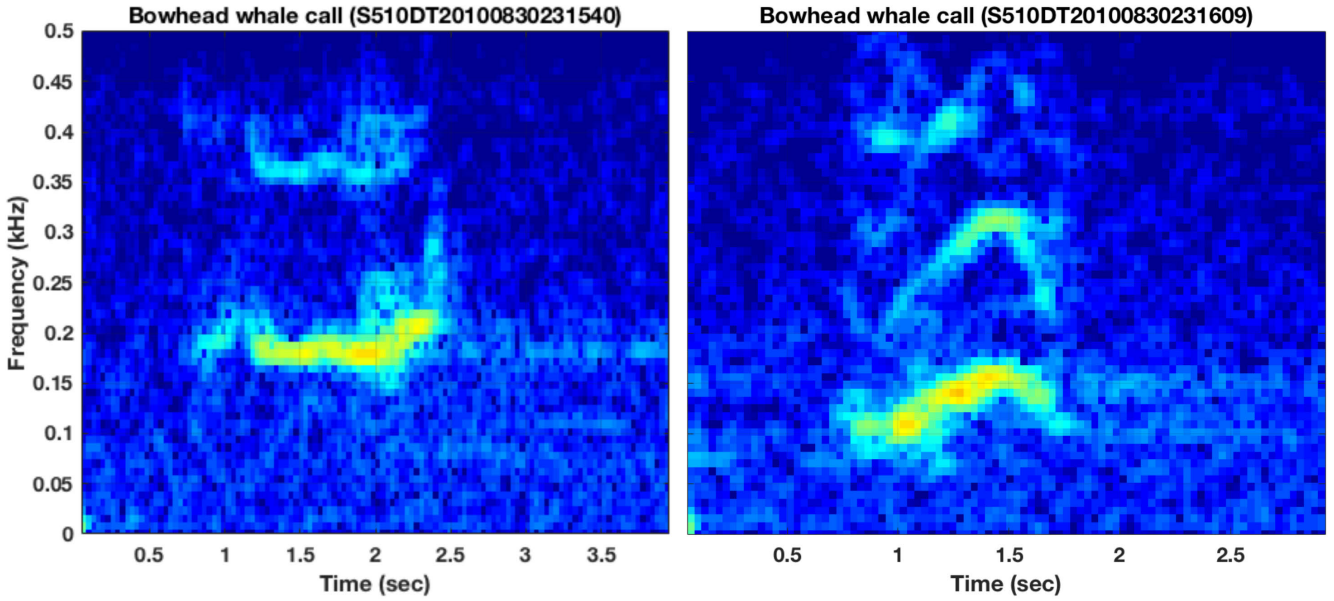
Calibration signals



RESULTS

Bowhead whale sample calls

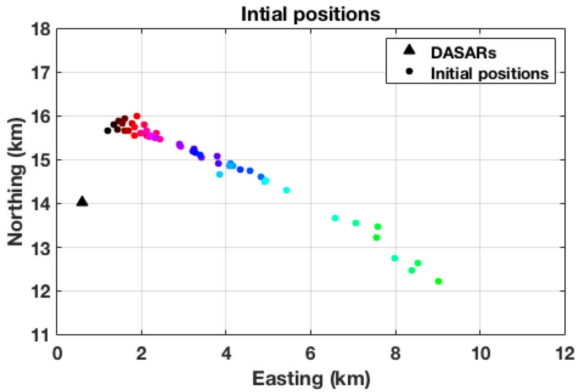
Two calls measured on same sensor:

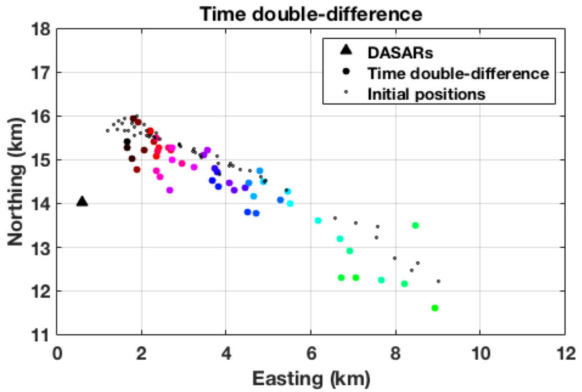


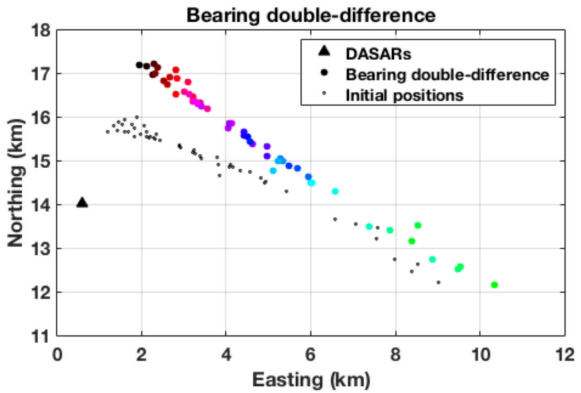
Note: hard to pick arrival time precisely.

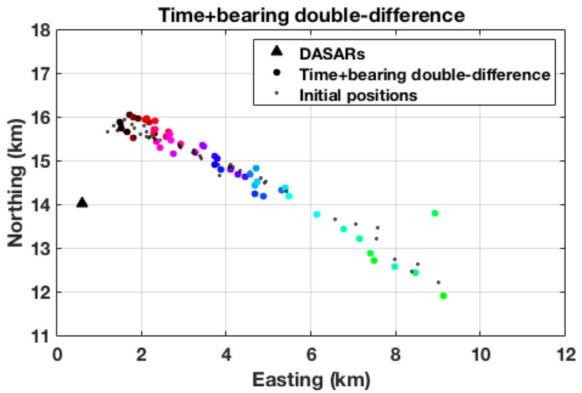
RESULTS

Bowhead whale track (preliminary results)









RESULTS

CONCLUSIONS

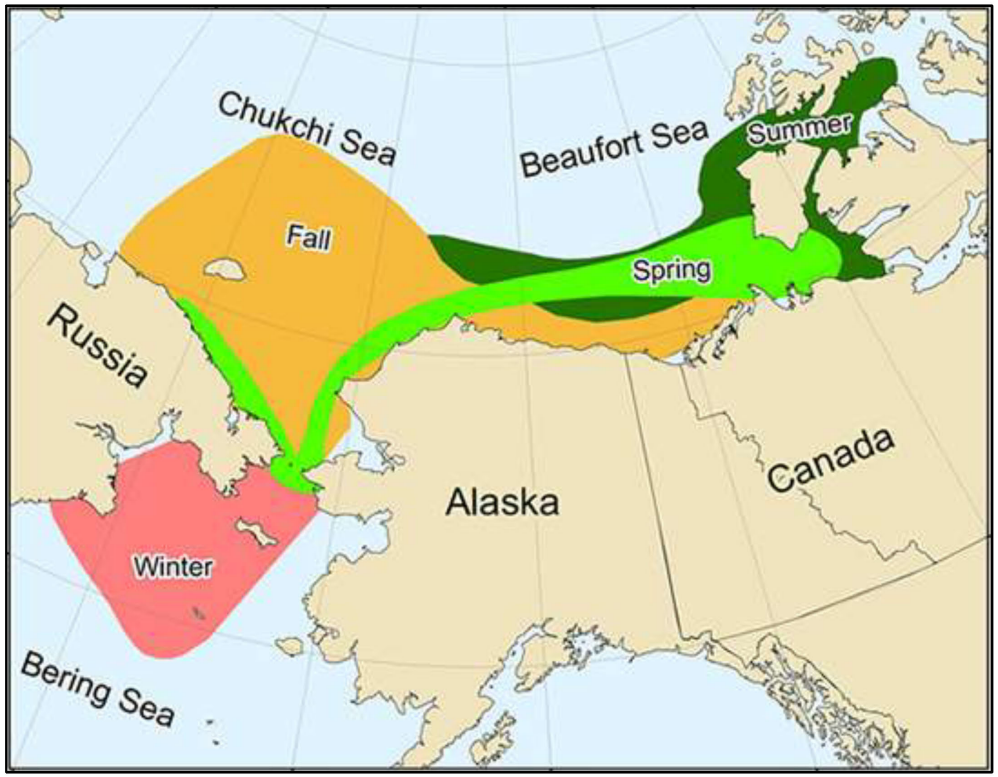
- **Double-difference** approach can be formulated to use **both time & bearing information** from vector sensors.
- Simulations show promising results for **reducing systemic errors** (time offset + bearing bias).
- Double-difference shown **to improve localization precision of calibration signals**.
- Preliminary bowhead whale results show difficulty with time double-difference.

Future work

- Understand bearing double-difference results on calibration signals.
- Estimate double-difference error to determine localization precision.
- Reformulate double-difference to use TDOA instead of arrival times for bowhead whale calls (to improve time measurements).

QUESTIONS ?
SCRIPPS INSTITUTION OF OCEANOGRAPHY
GREENERIDGE SCIENCES

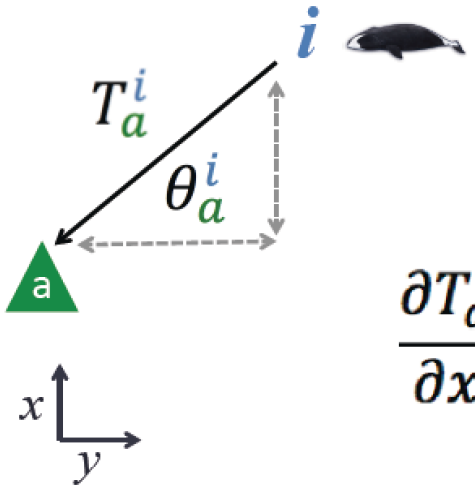
Bowhead whale migration (Pacific population)



(Quakenbush et al., 2013)

BONUS SLIDE

Double-difference derivatives

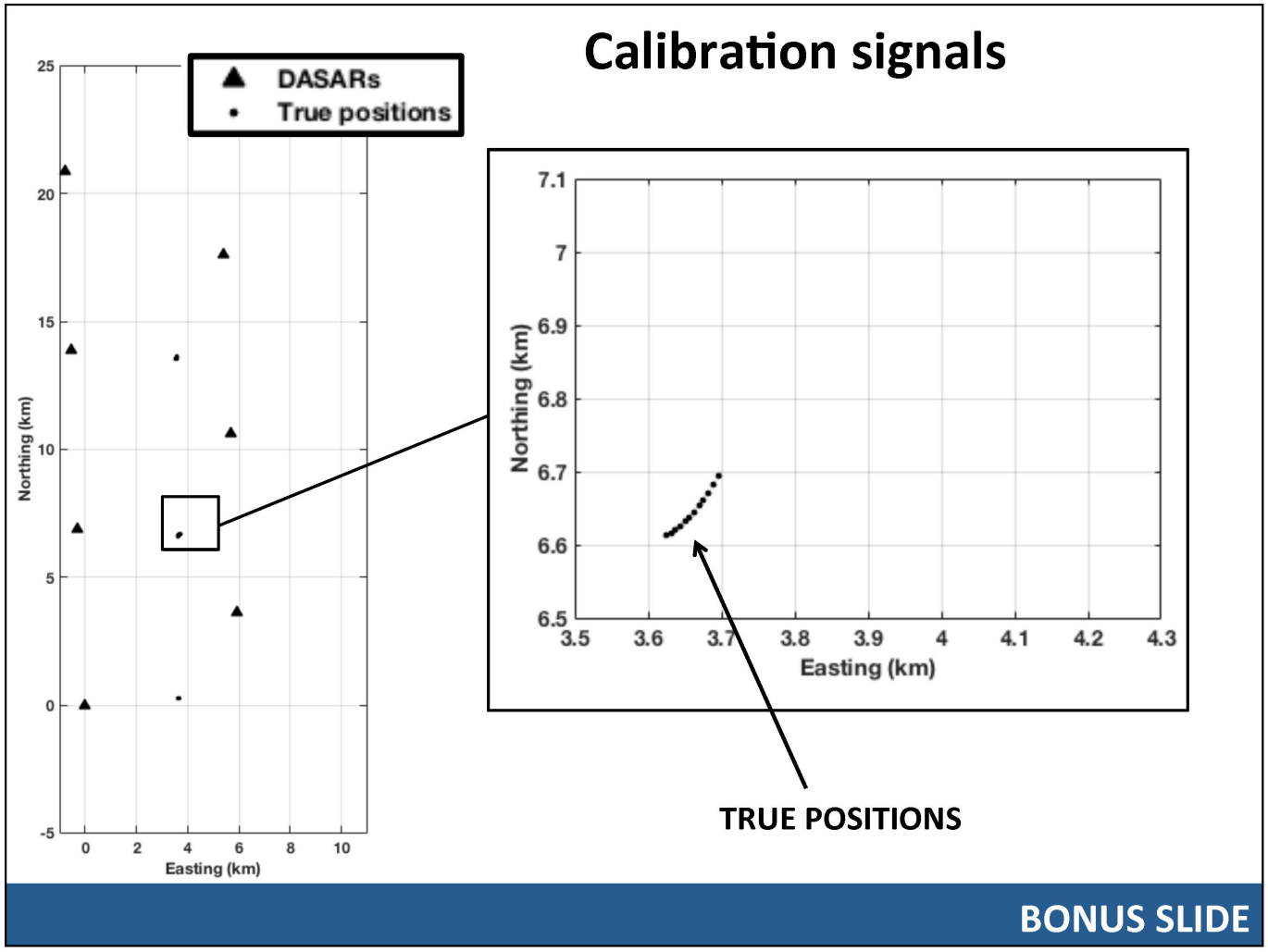


$$\frac{\partial T_a^i}{\partial x} = \frac{cos\theta_a^i}{c} \qquad \frac{\partial \theta_a^i}{\partial x} = -\frac{sin\theta_a^i}{{R_a^i}^2}$$

$$\frac{\partial T_a^i}{\partial y} = \frac{sin\theta_a^i}{c} \qquad \frac{\partial \theta_a^i}{\partial y} = \frac{cos\theta_a^i}{{R_a^i}^2}$$

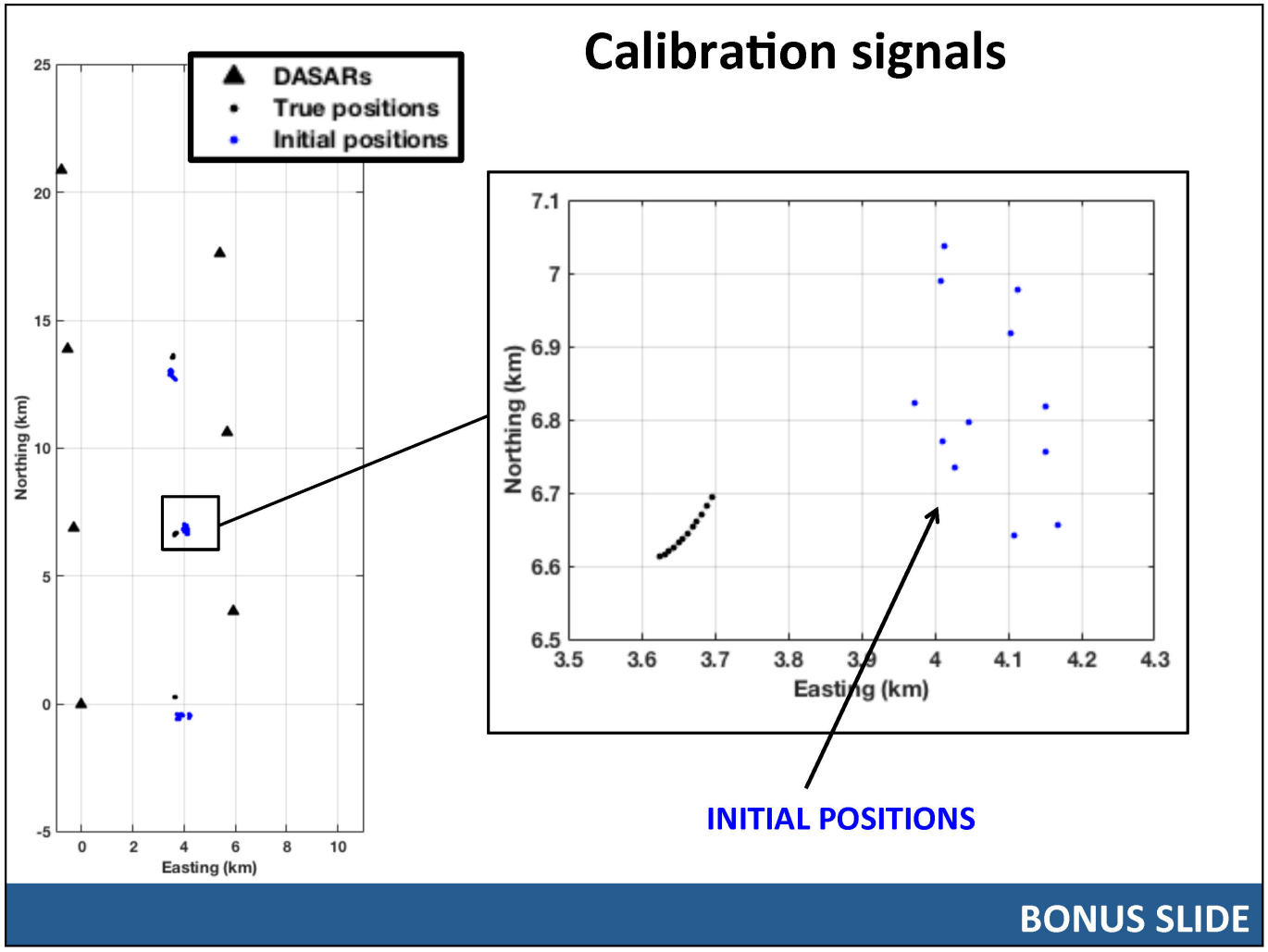
BONUS SLIDE

Calibration signals

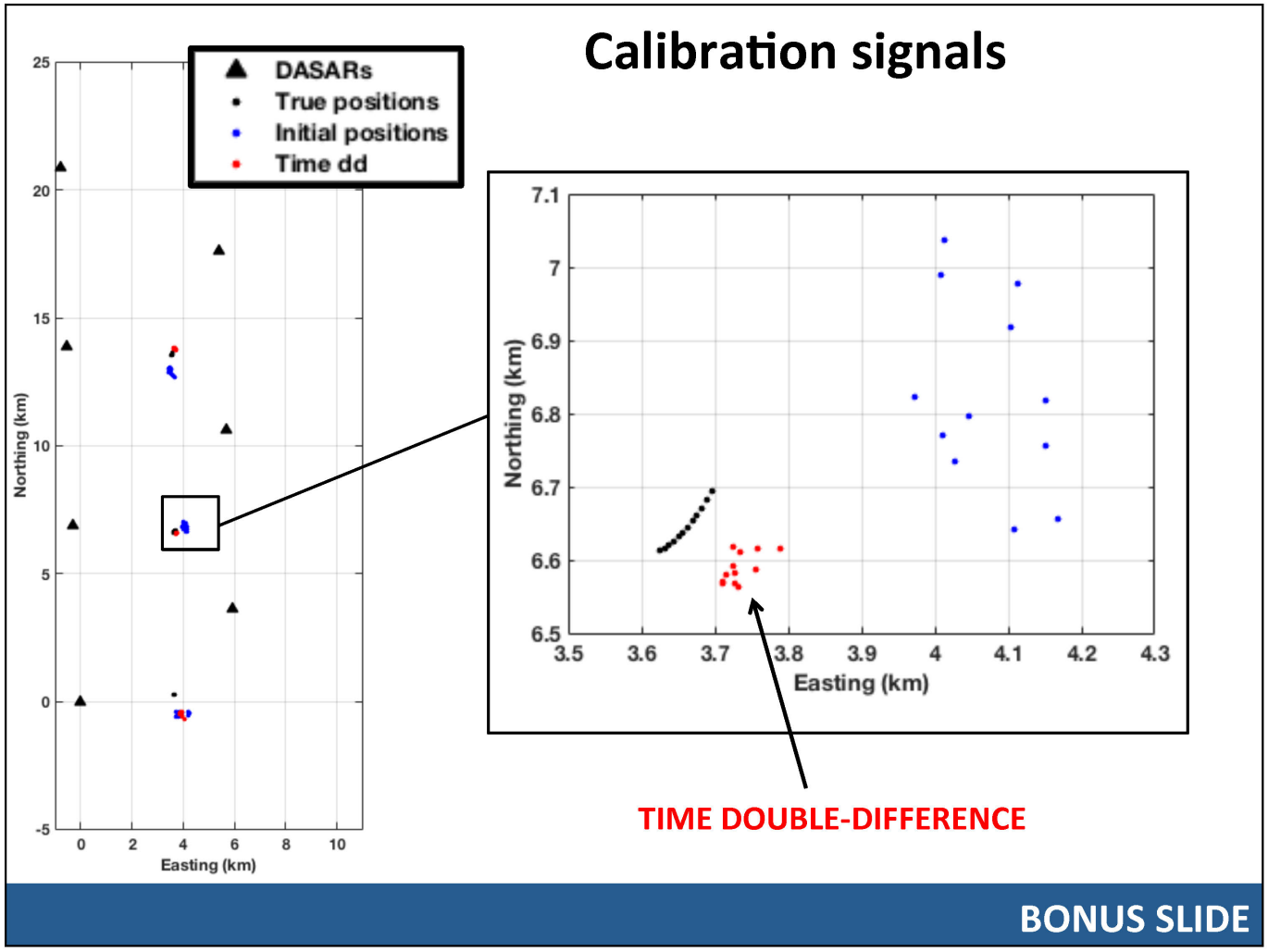


BONUS SLIDE

Calibration signals

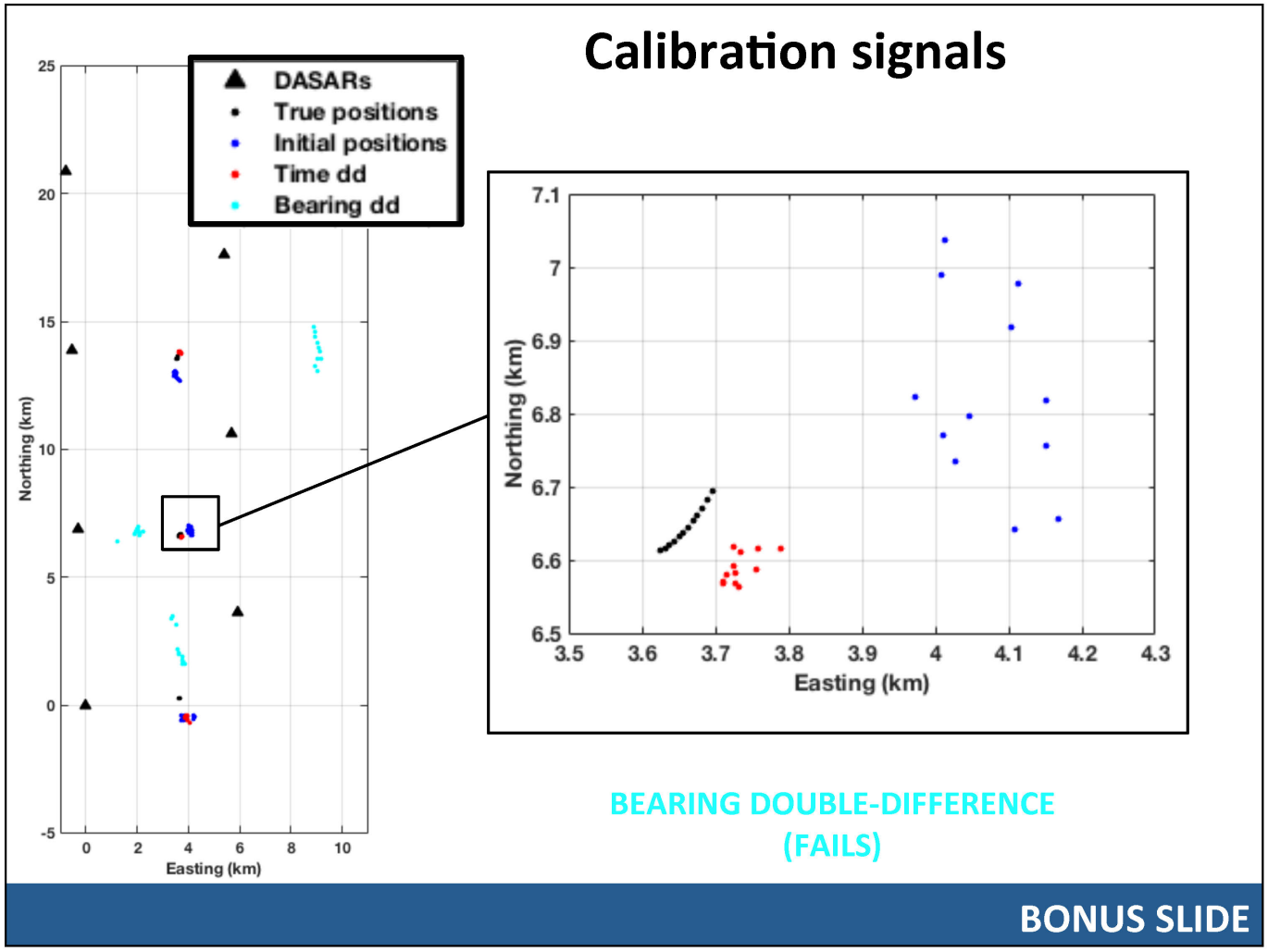


Calibration signals



BONUS SLIDE

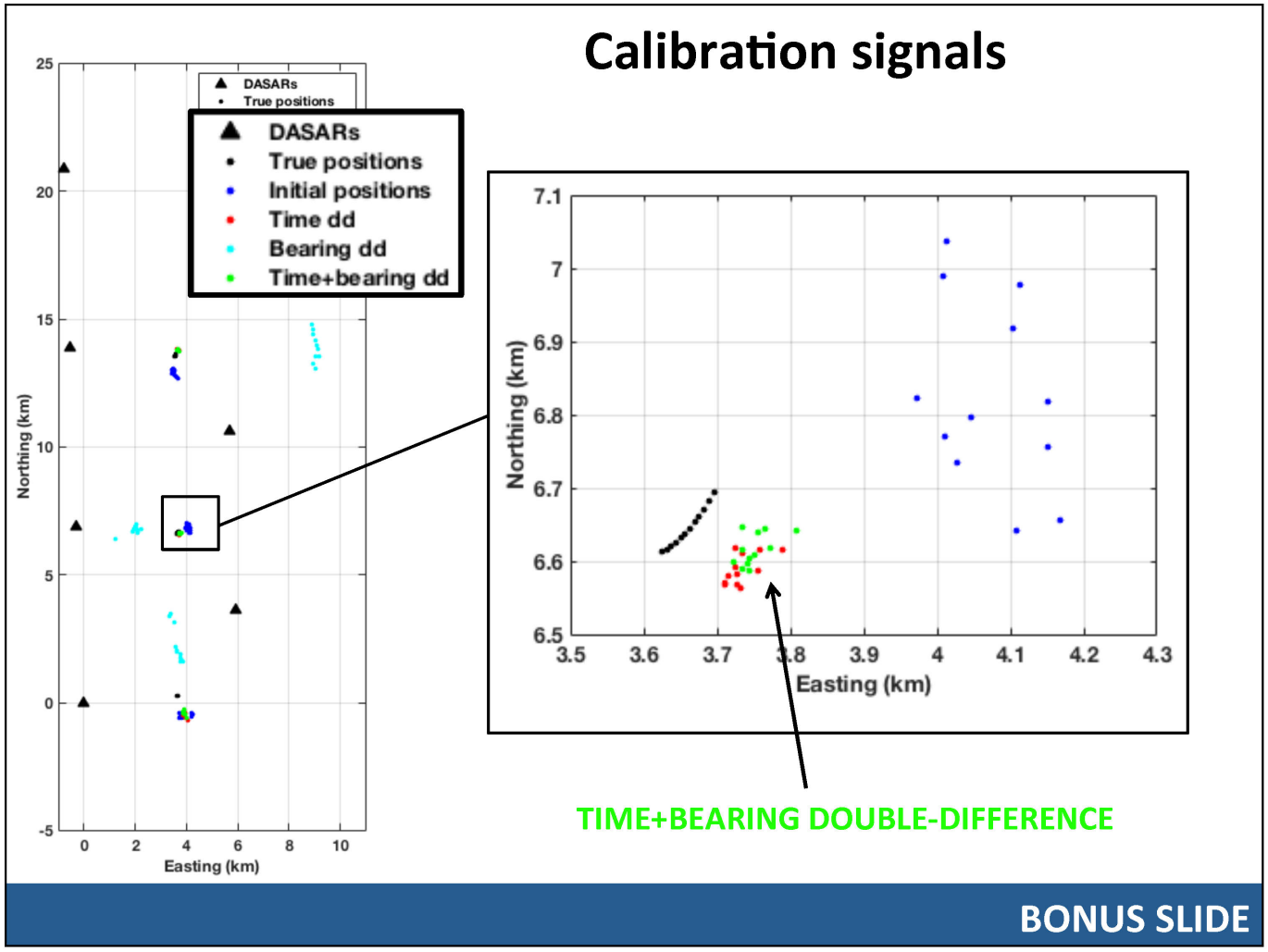
Calibration signals



BEARING DOUBLE-DIFFERENCE (FAILS)

BONUS SLIDE

Calibration signals



BONUS SLIDE

Contact us

General enquiries

info@soundandmarinelife.org

Media enquiries

press@soundandmarinelife.org
+44 (0) 20 7413 3416

E&P **SOUND & MARINE LIFE** PROGRAMME