UC San Diego
UC San Diego Previously Published Works

Title

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

Permalink

Journal

ISSN

Authors

Thode, Aaron M
Blackwell, Susanna B
Conrad, Alexander S
et al.

Publication Date

DOI

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

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

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

2062   J. Acoust. Soc. Am. **147** (3), March 2020

Thode *et al.*

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 $p("+")$ 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

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

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:

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

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

J. Acoust. Soc. Am. **147** (3), March 2020

Thode *et al.*   2063

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

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
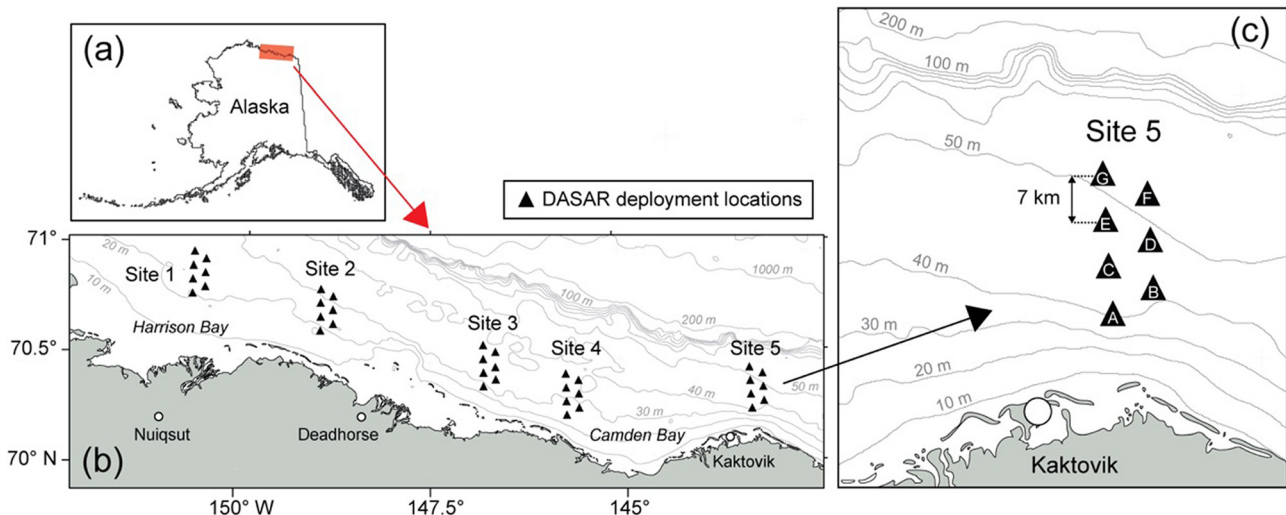

FIG. 1. (Color online) Locations of passive acoustic deployments. (a) North Slope of Alaska; (b) DASAR deployments. (c) Closeup of Site 5 deployment and bathymetry.

J. Acoust. Soc. Am. **147** (3), March 2020

Thode *et al.* 2065

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

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

J. Acoust. Soc. Am. **147** (3), March 2020

Thode *et al.*   2067

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 ($\Delta SL/\Delta NL$), 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 $\Delta SL/\Delta NL$ 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 $f(r|SLNR)$ 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
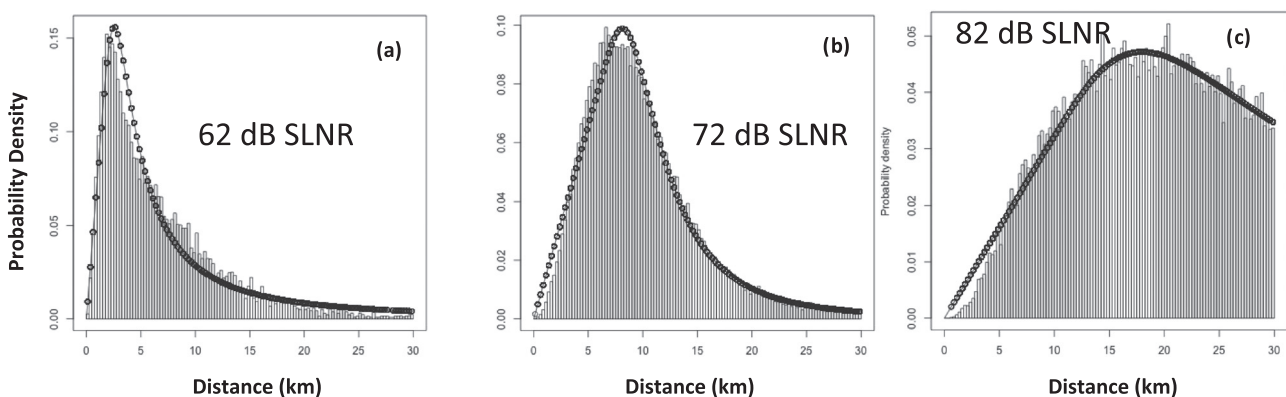

FIG. 2. Observed call range distribution $f(SLNR,r)$ at Sites 3 and 5 combined, using long-range automated analysis (Table I), evaluated at three different source level-to-noise ratios (SLNR): (a) 62 dB; (b) 72 dB; (c) 82 dB. Histograms have been normalized to approximate probability density functions (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.
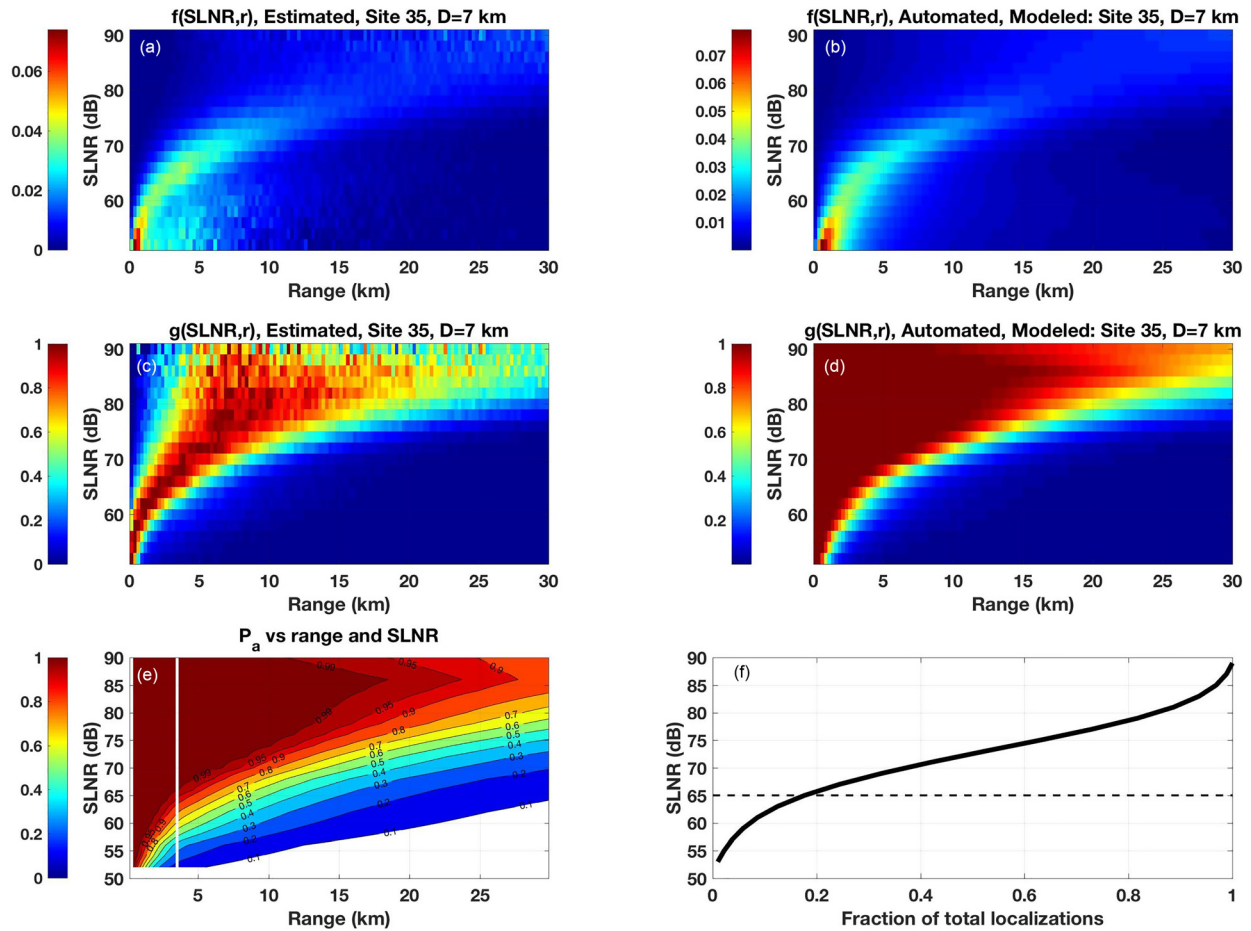
FIG. 3. (Color online) Computation of localization probability P_a via modified point transect theory, using automated data. (a) Observed $f(SLNR, r)$ imaged in grid steps of 2 dB SLNR and 250 m range; (b) best-fit modeled $f(SLNR,r)$; (c) estimated $g(SLNR,r)$ [$f(SLNR,r)/\pi(r)$]; (d) best-fit modeled $g(SLNR,r)$; (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
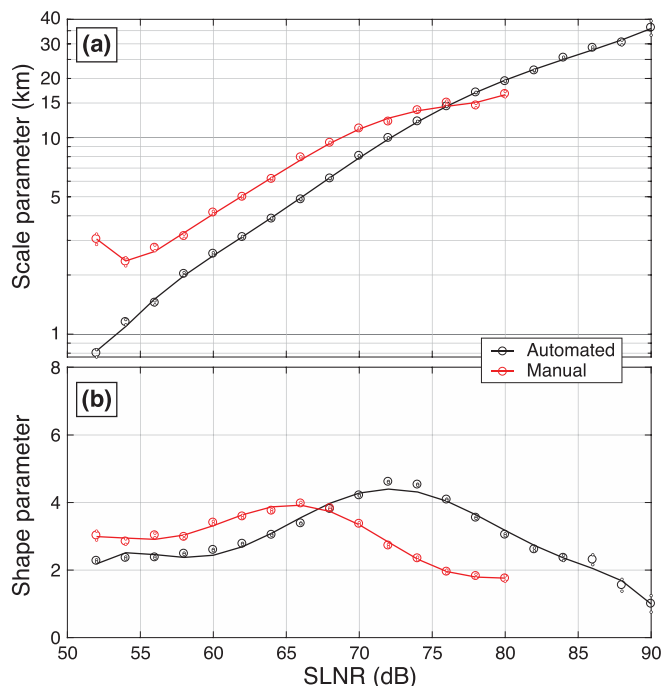
FIG. 4. (Color online) Scale "σ" (a) and shape "b" (b) parameters of hazard rate localization function vs SLNR for both automated (black) and manual (red) data analysis. Small dots indicate standard error of parameter fits.

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
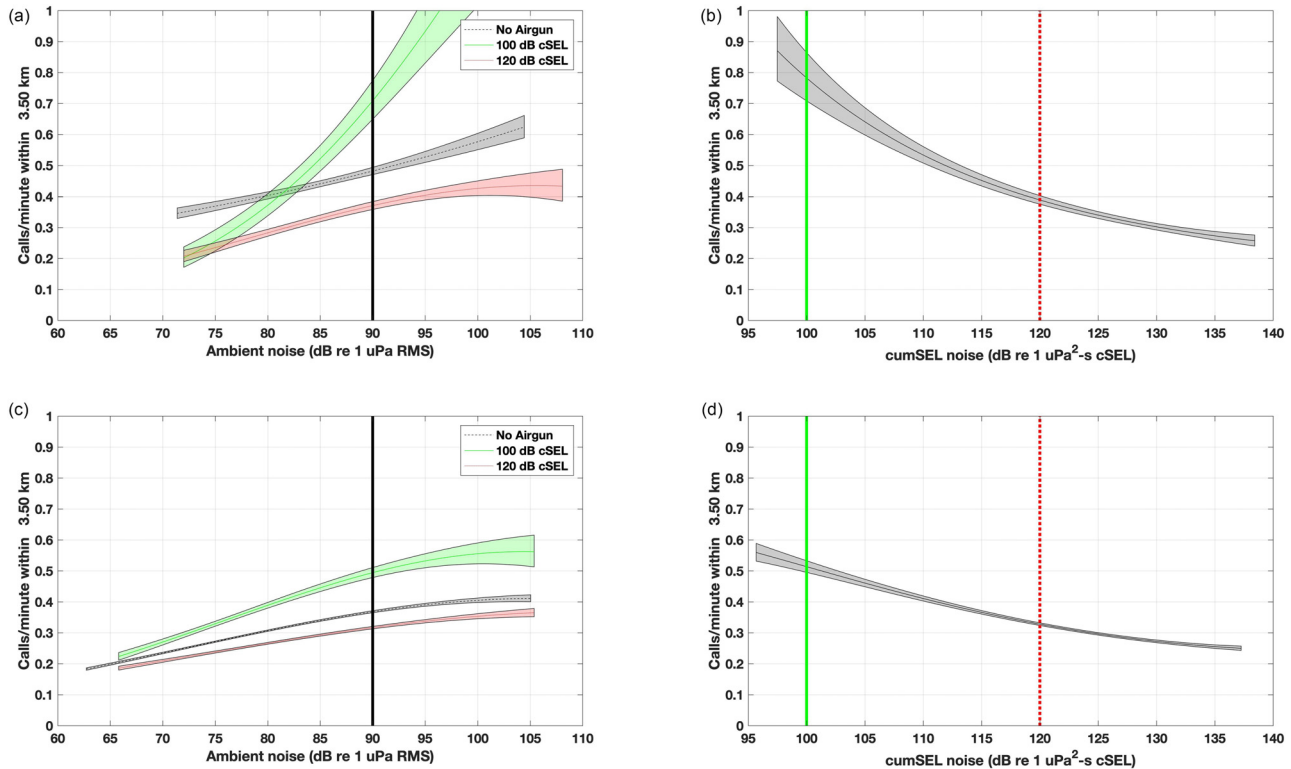
FIG. 5. (Color online) Regression analysis of call density vs in-band ambient noise level [SPL (RMS)] and airgun cSEL. The top and bottom rows display manual and automated analyses, respectively. (a) Call density vs ambient noise level when no seismic activity is present (black, dashed), when seismic 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

J. Acoust. Soc. Am. **147** (3), March 2020
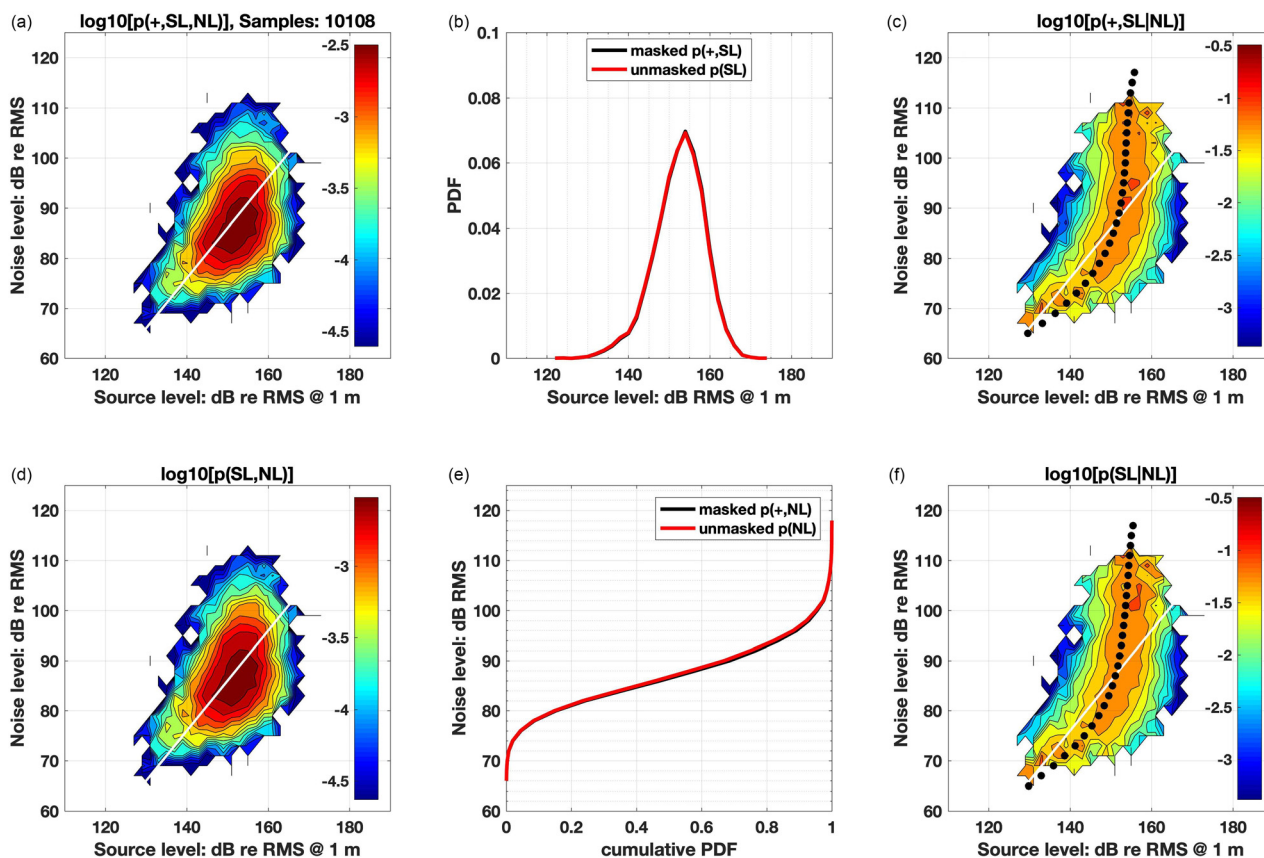
Thode *et al.*   2071

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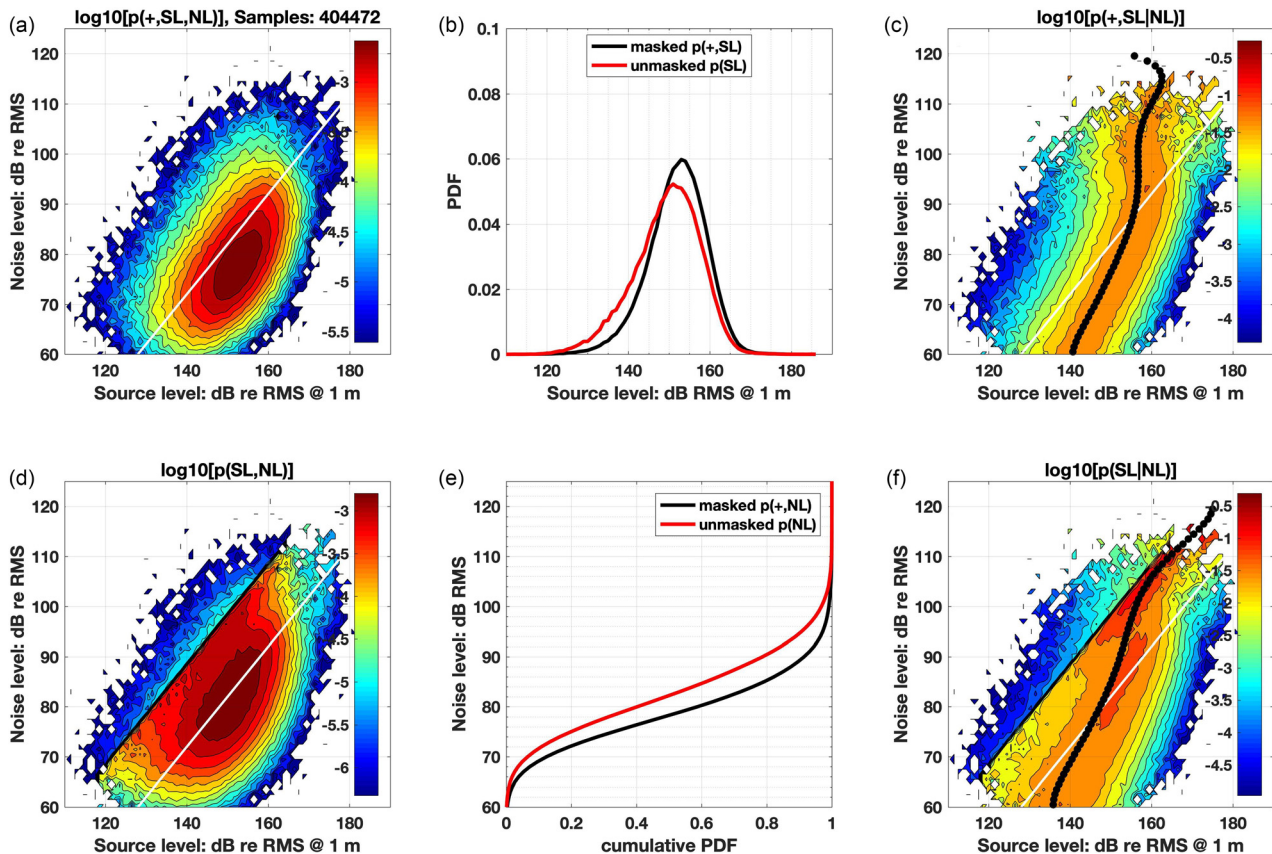
FIG. 7. (Color online) Same as Fig. 6, except displaying long-range automated data (N = 404 472; R_{max} = 30 km) from Table I.

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

J. Acoust. Soc. Am. **147** (3), March 2020
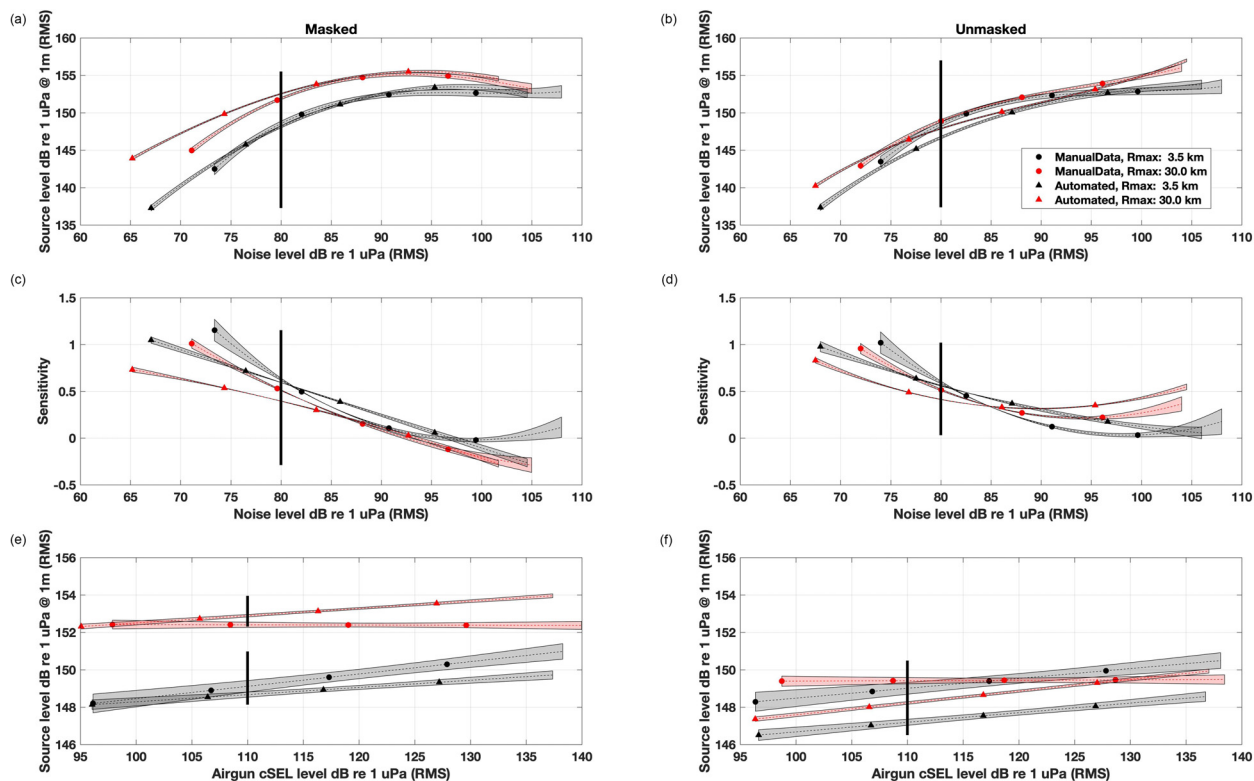
Thode *et al.*   2073

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

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).

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

J. Acoust. Soc. Am. **147** (3), March 2020
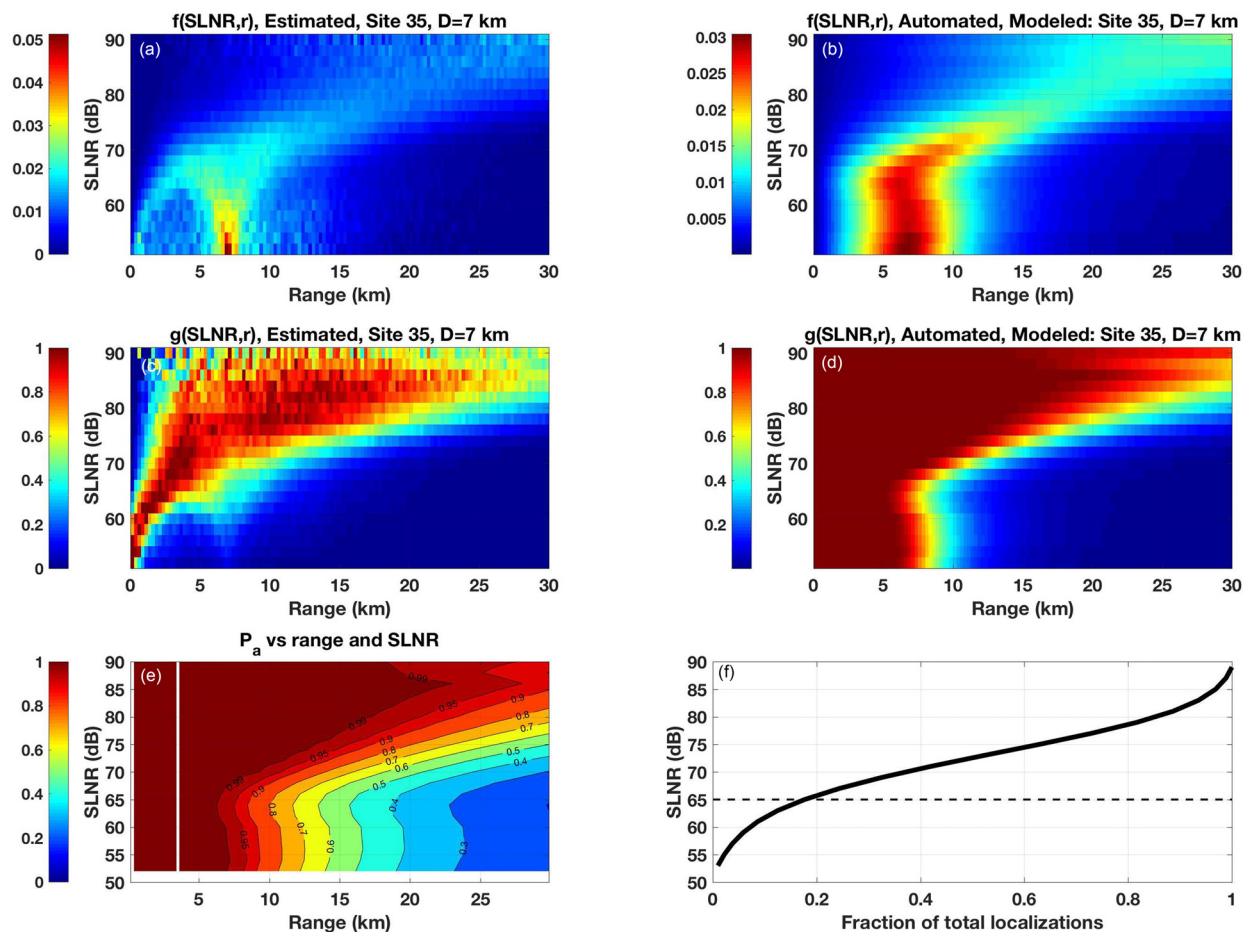
Thode *et al.*   2075

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.
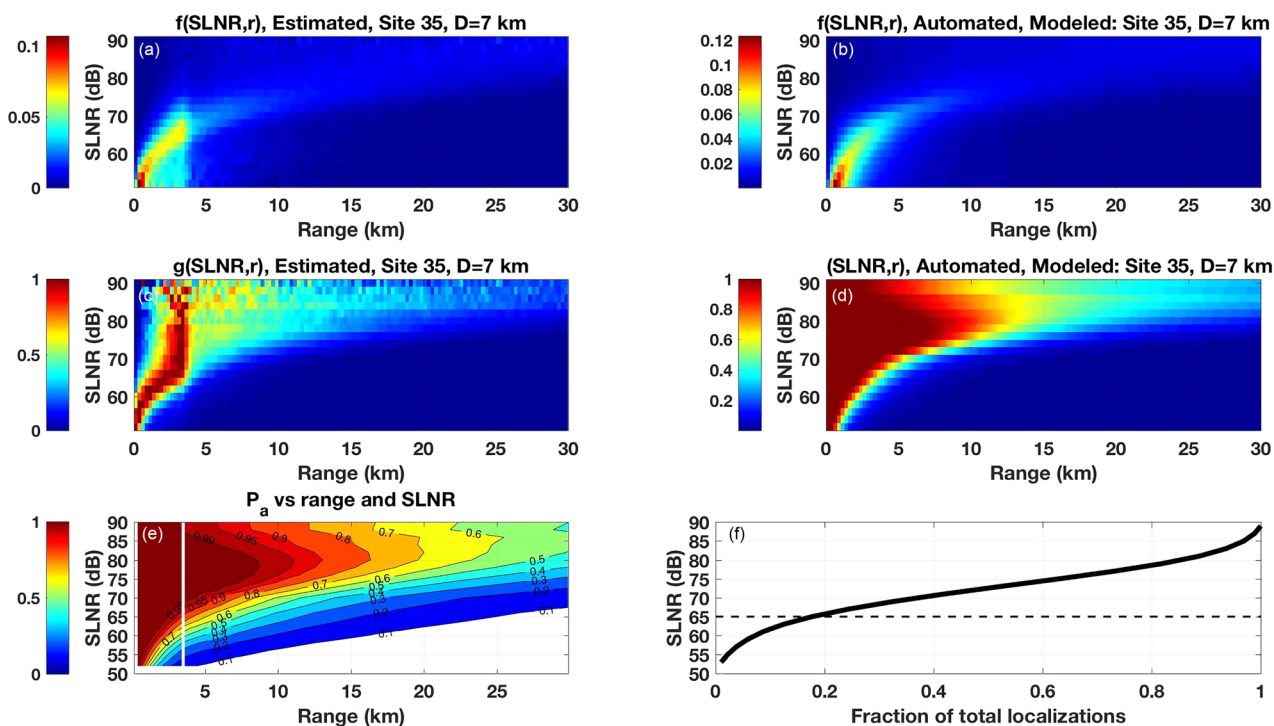

FIG. 10. (Color online) Figure 9 recomputed using only the closest range for each localization, using the standard circular availability function.

form of the availability function, $\pi(r)$, 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
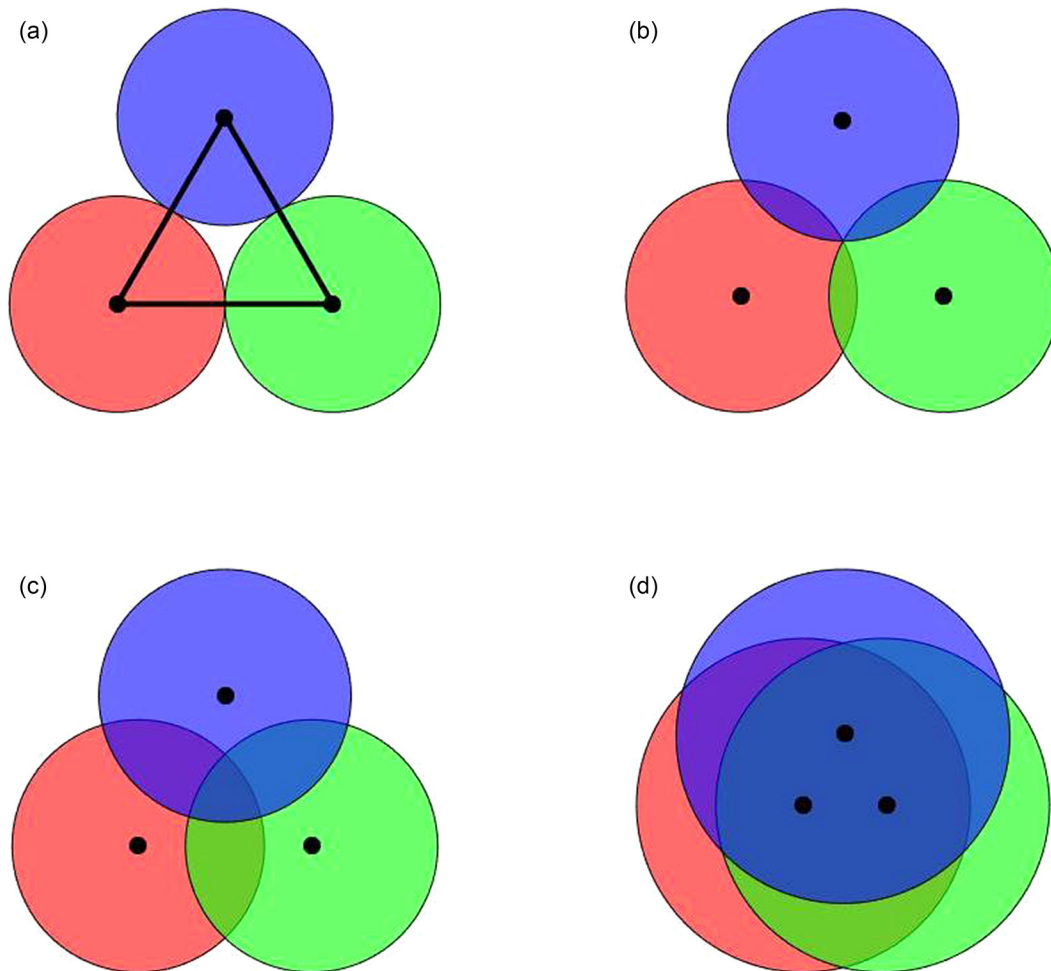

FIG. 11. (Color online) Illustration of how the availability function changes for a three-sensor grid. The availability function is all points on the circles' perimeter that do not penetrate other circles. (a) $r < D/2 : \pi(r) = 6\pi r$, (b) $D/2 < r < D/\sqrt{3} : \pi(r) = 12\alpha r$, (c) $D/\sqrt{3}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

J. Acoust. Soc. Am. **147** (3), March 2020

Thode *et al.*   2077

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:

$$r < \frac{D}{2} : \pi(r) = 6\pi r; \quad \frac{D}{2} < r < \frac{D}{\sqrt{3}} : 12\alpha r;$$

$$\frac{D}{\sqrt{3}} < r : \pi(r) = (6\alpha + 2\pi)r;$$

$$\frac{D}{\sqrt{3}} \ll r : \pi(r) = 3D + 2\pi r, \tag{A1}$$

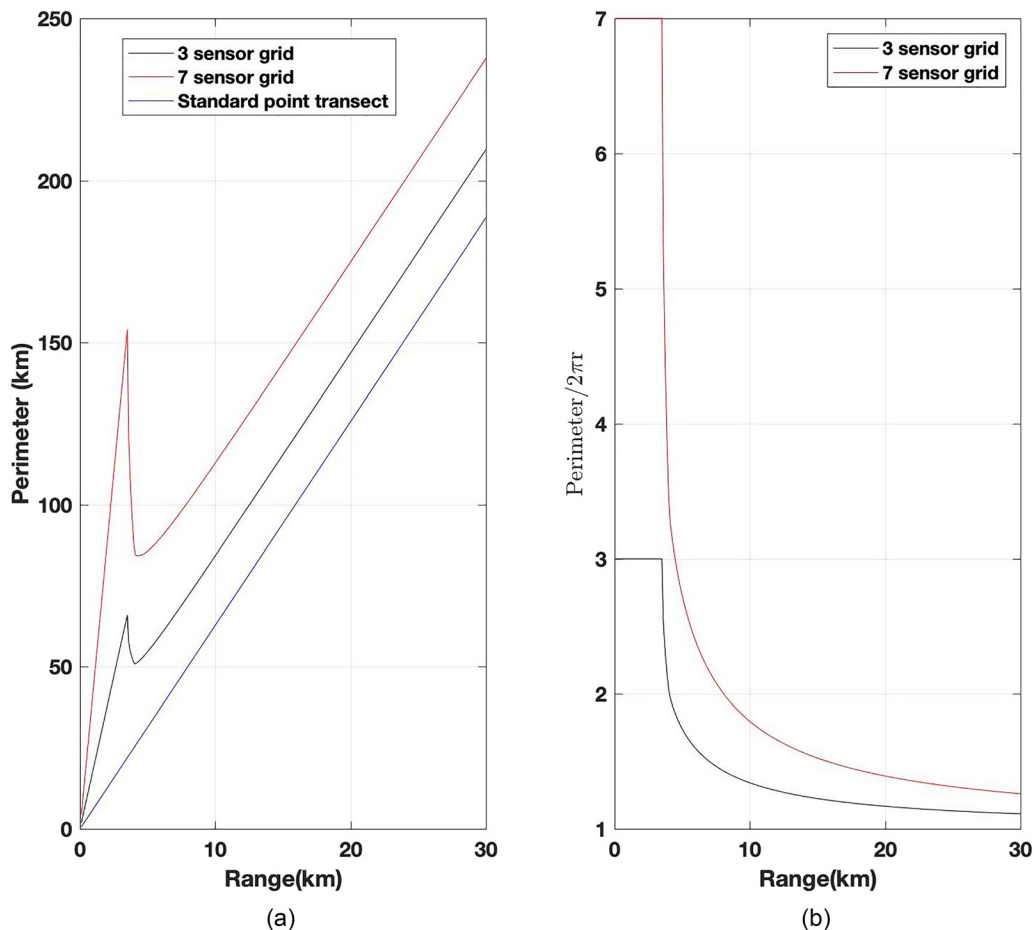FIG. 12. (Color online) (a) Availability function for a three-sensor array with 7 km separation (black), seven-sensor array with 7 km separation (red), and standard point transect (blue). (b) Availability function for three-sensor (black) and seven-sensor (red) distributed array, normalized by standard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

2078   J. Acoust. Soc. Am. 147 (3), March 2020

Thode *et al.*

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

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

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 air-gun 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.

J. Acoust. Soc. Am. **147** (3), March 2020

Thode *et al.*   2079

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.