



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

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

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

Abstract

Various methods for estimating animal density from visual data, including distance sampling (DS) and spatially explicit capture-recapture (SECR), have recently been adapted for estimating call density using passive acoustic monitoring (PAM) data, e.g., recordings of animal calls. Here we summarize three methods available for passive acoustic density estimation: plot sampling, DS, and SECR. The first two require distances from the sensors to calling animals (which are obtained by triangulating calls matched among sensors), but SECR only requires matching (not localizing) calls among sensors. We compare via simulation what biases can arise when assumptions underlying these methods are violated. We use insights gleaned from the simulation to compare the performance of the methods when applied to a case study: bowhead whale call data collected from arrays of directional acoustic sensors at five sites in the Beaufort Sea during the fall migration 2007–2014. Call detections were manually extracted from the recordings by human observers simultaneously scanning spectrograms of recordings from a given site. The large discrepancies between estimates derived using SECR and the other two methods were likely caused primarily by the manual detection procedure leading to non-independent detections among sensors, while errors in estimated distances between detected calls and sensors also contributed to the observed patterns. Our study is among the first to provide a direct comparison of the three methods applied to PAM data and highlights the importance that all assumptions of an analysis method need to be met for correct inference.

Keywords Distance sampling · Non-independent detections · Plot sampling · Spatially explicit capture-recapture

Handling Editor: Luiz Duczmal.

1 Introduction

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

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

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

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

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

A comparison between the three density estimation methods has rarely been undertaken with PAM data, as in most cases the data have limited the analyses to a particular method. For example, Phillips (2016) compared DS and SECR estimates of animal density from a combination of PAM and focal follow data against a small population of known size and concluded that they generally produced similar results. In this study, we use a single large PAM dataset (without auxiliary data from a different source) consisting of > 680,000 bowhead whale calls in the Beaufort Sea, collected by Greeneridge Sciences, Inc. (Santa Barbara, California), on behalf of Shell Exploration and Production Company (SEPCO) over an 8-year monitoring period (2007–2014). The required data for each method—distances to the calls for DS and PS and matched calls across sensors for SECR—were all available from this single dataset. Since the same dataset was used for all three estimation methods, any discrepancies in results must arise from violations of one or more of the assumptions for the respective methods. We also created a simulation tool to examine the robustness of the density estimation methods to various violations of underlying assumptions.

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

2 Methods for estimating call densities

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

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

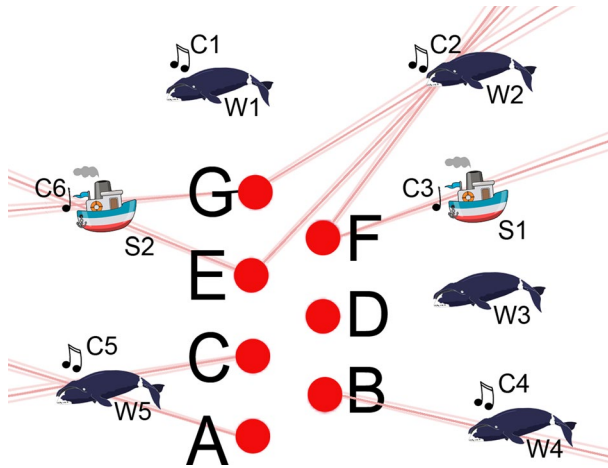


Fig. 1 Example PAM survey design based on the bowhead whale study using a configuration of seven acoustic sensors moored to the seafloor (red dots A–G) where each sensor represents a vertex in regular triangles with 7 km edges. Pink lines are azimuths to the call locations with uncertainty (light pink).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

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

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

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

Table 1 Capture history of detections at sensors A–G based on the example from Fig. 1

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

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

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

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

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

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

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

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

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

2.1 Analyses methods and assumptions

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

$$\hat{D}_c = \frac{n(1 - \hat{f})}{A\hat{p}T}, \quad (1)$$

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

The average detection probability p within the search area is generally modelled using two main components: the absolute detection probability at zero distance from the sensor g_0 , which is the probability that a call made at zero horizontal distance from the sensor is detected by the sensor, and a detection function $g(y)$ that describes the decay in detection probabilities with increasing distance y from the sensor relative to g_0 . Depending on the method, either component is assumed or estimated from the data where applicable (see below). A frequently used detection function is the half-normal:

$$g(y) = \exp\left(-\frac{y^2}{2\sigma^2}\right), \sigma > 0. \quad (2)$$

Equation (2) contains one parameter, the scale parameter σ , which needs to be estimated. Note that $g(y = 0) = 1$. Larger σ values yield detection functions with

high detection probabilities out to larger distances. In the following we use these components, g_0 and $g(y)$, to compare p for the three different methods.

2.1.1 Plot Sampling (PS) for PAM data

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

PS assumes that all calls produced within the individual a_{PS} are detected with certainty. To meet this assumption, the search area is typically limited to a relatively small radius w_{PS} . We can therefore assume that $g_{0PS} = 1$ and $p_{PS} = 1$. Hence, this method does not require estimating a detection function, at the cost of rejecting large numbers of detections that originate outside w_{PS} . As we need to determine which calls originated from within w_{PS} , a successful PS application requires that all calls produced within w_{PS} around any sensor are localized—hence, the required sensor spacing described above. Further assumptions are listed in Table 3.

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

2.1.2 Distance sampling (DS) for PAM data

Here, each sensor represents a point in a point transect survey, which is a form of DS (e.g., Buckland et al. 2001, chapter 5; Buckland 2006). In comparison to PS, we expand the search radius from w_{PS} to a larger radius, w_{DS} and include all n_{DS} call detections within a_{DS} (the circular area around a sensor with radius w_{DS}) from each of the K sensors. Like PS, DS assumes that all calls at (or near) the sensor are detected with certainty, i.e.: $g_{0DS} = 1$. However, we no longer assume that all calls within the area a_{DS} around each sensor are detected with certainty. Instead, we fit a detection function $g_{DS}(y)$ to the distances between the sensors and the detected calls (e.g., as in Table 2) and use it to estimate the average detection probability within a_{DS} :

$$p_{DS} = \frac{2}{w_{DS}^2} \int_0^{w_{DS}} yg_{DS}(y)dy. \tag{3}$$

An estimate of p_{DS} can be obtained using Eq. (3), replacing $g_{DS}(y)$ with $\hat{g}_{DS}(y)$ (Buckland et al. 2015). One sees that PS is a limiting case of DS when the search radius w_{DS} is shrunken to values small enough that p_{DS} becomes 1. Similar to PS, n_{DS} refers to the sum of the number of detections that fall within the search areas a_{DS}

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

#	Assumption	Description	PS	DS	SECR
1	Adequate survey design representative of study area	Placement of an adequate number of sensors according to a (systematic)-random design	Highly important	Highly important	Highly important
2	No un-modelled heterogeneity in detection probabilities	Detection probabilities only depend on distance to the sensor	Not important as long as $p_{PS} = 1$	Not important if moderate heterogeneity	Highly important
3	$g_0 = 1$	Detection probability at/very near the sensor equals 1	Highly important	Highly important	Not required as it is estimated from the data
4	The false positive rate is estimated accurately	\hat{f} in Eq. (1) accounts for all false positives in n	Highly important	Highly important	Highly important
5	No mis-associations of calls across sensors	Detections of the same call are not falsely identified as separate calls	Problematic if this leads to singletons within $K_{D_{PS}}$ or to an inflated n_{PS}	Problematic if this leads to singletons on/near any of the sensors or to an inflated n_{DS}	Highly important
6	No lumping of calls	Calls detected on different sensors are not falsely identified as the same call	Problematic if this leads to negatively biased n_{PS}	Problematic if this leads to negatively biased n_{DS}	Highly important
7	Independent detection of calls across sensors	The probability of detecting a call on one sensor does not affect the probability of detecting it on other sensors	Not important	Robust to violation	Highly important
8	Accurate distances between sensors and calls	Distances between sensors and calls are measured accurately	Important to obtain accurate n_{PS}	Important. Minor errors can be alleviated by using binned distances for model fitting	Does not apply

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

of the K sensors, and any call that falls within overlapping search areas is counted towards n_{DS} for each time it was detected by a sensor along with the distance to the respective sensor. While this may seem to artificially inflate n_{DS} , the reasoning again arises from the requirement that the total search area A_{DS} is Ka_{DS} , i.e., no subtraction of any overlapping areas occurs.

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

$$v_{DS} = 2\pi \int_0^{w_{DS}} yg_{DS}(y)dy. \tag{4}$$

An estimate of the effective area, \hat{v}_{DS} can be obtained using Eq. (4), replacing $g_{DS}(y)$ with $\hat{g}_{DS}(y)$. We can substitute $K\hat{v}_{DS}$ for $A\hat{p}$ in Eq. (1) for estimating call densities.

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

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

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

For SECR we estimate the probability g_{0SECR} of detecting a call at distance zero as well as the detection function $g_{SECR}(y)$ from the capture histories of the calls (e.g., Borchers and Efford 2008; Borchers 2012). Hence, in comparison to PS or DS, we are not required to assume that all calls at/near the sensor are detected and we do not require call distances or locations. Furthermore, the data are not truncated by a search radius, i.e., $w_{SECR} = Inf$. All detected calls, along with their detection histories, are included in the analysis, regardless of the number of sensors they were detected on. Theoretically, with $w_{SECR} = Inf$, the total search area $A_{SECR} = Inf$ and p_{SECR} approaches zero. Hence, in practice, we use a different approach where the search area a_{SECR} around each sensor

only extends out to a defined distance w_{SECR} beyond which it is safe to assume that no calls can be detected (Efford 2019). Nonetheless, we do not estimate an average detection probability within a_{SECR} . Instead, we use estimates of g_{0SECR} and $g_{SECR}(y)$ to obtain an estimate of the effective area v_{SECR} . As for DS, it is estimated using a combination of the search area and the detection probabilities. However, in contrast to DS, the effective area v_{SECR} is defined as the whole area surrounding the K sensors within which as many calls were missed as were detected beyond. It is estimated using the following steps (e.g., Stevenson et al. 2015). First we estimate the probability $p_k(X)$ that a call produced at location X (this location is unobserved) is detected by the k th sensor using:

$$p_k(X) = g_{0SECR}g_{SECR}(y_X), \tag{5}$$

where y_X is the distance between X and the k th sensor. The probability $p.(X)$ that the call was detected on at least one sensor becomes:

$$p.(X) = 1 - \prod_{k=1}^K [1 - p_k(X)], \tag{6}$$

The effective area is obtained by integrating $p.(X)$ over A_{SECR} . In practice this is done by dividing A_{SECR} into I grid cells, each with size a_i , where the X_i represent the center points of the grid cells:

$$v_{SECR} = \sum_{i=1}^I p.(X_i)a_i. \tag{7}$$

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

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

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

3 Simulation study

3.1 Methods

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

Each simulation consisted of 1,000 iterations. For each iteration we generated random call detections in a first step, using the same sensor configuration as shown in Fig. 1 and a known number $N_{sim} = 10,000$ calls produced at known locations over a fixed recording time T throughout a defined study area; hence, call density D_{sim} was known. These calls were detected by each sensor with probability $g_{0sim}g_{sim}(y)$, with known g_{0sim} and using a half-normal key function (Eq. (2)) for $g_{sim}(y)$ with known scale parameter σ_{sim} . Any call detected on multiple sensors was considered as matched correctly between sensors. In a second stage, we analyzed the call detection data using each of the three methods. We first tested the methods performed if all of the assumptions from Table 3 were met in a baseline simulation. We then expanded these tests to scenarios where one of the assumptions from Table 3 was violated. To identify potential biases, we used the following diagnostics:

- Comparisons of estimated call density with true call density D_{sim} ;
- Comparisons of the estimated with true probabilities of detection (DS and SECR only) using visual tools—the detection function plot as shown in Fig. 2—and comparisons of g_0 (SECR only) and σ estimated with the respective method vs the true values g_{0sim} and σ_{sim} ;
- Plotting proportions of calls detected by one, two, three, etc. sensors (defined here as *proportion plots*) as shown in Fig. 2. Following the hypothetical example, these

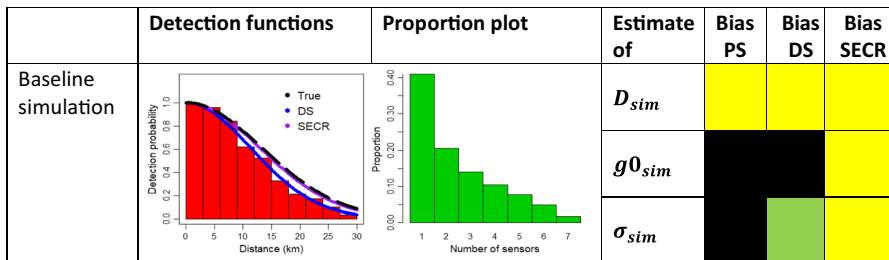


Fig. 2 Results from the baseline simulation. Red, scaled histogram of distances (km) to calls detected by two or more sensors within $w_{PS} = 30$ km, overlain with the true (black line) and estimated DS (blue line) and SECR (purple line) detection functions. Green proportion plot depicting the proportion of calls detected by 1–7 sensors. Color code for median-biases in the estimates as a percentage: Yellow none to minor negative or positive bias of < 10%; positive biases: Orange $\geq 10\%$; Red $\geq 20\%$; Dark Red $\geq 50\%$; negative biases: Light Green $\geq 10\%$; Green $\geq 20\%$; Blue $\geq 50\%$; Black NA. Numerical results are given in Appendix 1

Fig. 3 Results from simulation where one assumption from Table 3 was violated. Red: scaled histogram of distances (km) to calls detected by two or more sensors, overlain with the true (black line) and estimated DS (blue line) and SECR (purple line) detection functions. Green proportion plot depicting the proportion of calls detected by 1–7 sensors. Color code for median biases in the estimates as a percentage: ■ none to minor negative or positive bias of < 10%; positive biases: ■ ≥ 10%; ■ ≥ 20%; ■ ≥ 50%; negative biases: ■ ≥ 10%; ■ ≥ 20%; ■ ≥ 50%; ■ NA. Numerical results of biases are given in Appendix 1

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

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

3.2 Baseline simulation

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

Average biases were minor for the call density estimates for each of the three methods, as well as for the parameters pertaining to detection probabilities obtained with SECR (σ_{sim} and g_{0sim}) (Fig. 2). The estimates of the DS scale parameter were negatively biased and the DS detection function declined more quickly with increasing distance than the true or SECR detection functions; however, as the estimated DS detection function actually represented a localization function where each call needed to be detected by two or more sensors, this bias was expected. The fact that $g_{DS}(y)$ fitted better to the histogram of distances than $g_{SECR}(y)$ was caused by the increasing number of singletons with increasing distance which were not included in the histogram. Hence, although the DS detection function was negatively biased, the missing singletons meant that, overall, the DS detection function fitted the distances to the measured distances well and estimated call densities only had minor biases (Fig. 2).

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

3.3 Simulating Violations of Underlying Assumptions

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

#	Violated Assumption	Detection functions	Proportion plots	Est. of	Bias PS	Bias DS	Bias SECR
1	Adequate survey design			D_{sim}	Red	Red	Red
				g_{0sim}	Black	Black	Yellow
				σ_{sim}	Black	Green	Red
2	No un-modelled heterogeneity in detection probabilities			D_{sim}	Yellow	Black	Red
				g_{0sim}	Black	Black	Light Green
				σ_{sim}	Black	Red	Orange
3	$g_0 = 1$			D_{sim}	Green	Green	Yellow
				g_{0sim}	Black	Black	Yellow
				σ_{sim}	Black	Light Green	Yellow
4	The false positive rate is estimated accurately			D_{sim}	Yellow	Black	Red
				g_{0sim}	Black	Black	Yellow
				σ_{sim}	Black	Yellow	Light Green
5	No mis-associations of calls across sensors			D_{sim}	Yellow	Yellow	Orange
				g_{0sim}	Black	Black	Yellow
				σ_{sim}	Black	Light Green	Yellow
6	No lumping of calls			D_{sim}	Green	Light Green	Blue
				g_{0sim}	Black	Black	Yellow
				σ_{sim}	Black	Yellow	Red
7	Independent detection of calls across sensors			D_{sim}	Yellow	Orange	Blue
				g_{0sim}	Black	Black	Yellow
				σ_{sim}	Black	Red	Red
8	Accurately measured distances between sensors and calls			D_{sim}	Red	Red	Yellow
				g_{0sim}	Black	Black	Yellow
				σ_{sim}	Black	Green	Yellow

(row 1, Fig. 3). Un-modelled heterogeneity in detection probabilities (row 2) and inaccurate estimates of the false positive rate (row 4) only caused severe biases in call densities estimated with SECR; mis-associations (row 5) and lumping of calls (row 6) also caused larger biases for SECR compared to PS and DS. Setting $g_0 < 1$ (row 3) and introducing distance errors (row 8), on the other hand, only caused biases for PS and DS estimates. However, violation of the independence assumption (row 7) lead to the largest discrepancies from the baseline, both in terms of the SECR detection function—which was nearly horizontal within the 30 km displayed in Fig. 3—and the proportion plot. This was the only scenario in which the pattern differed from the decreasing proportions with the increasing number of sensors from the baseline simulation in Fig. 2.

4 Case study

4.1 Data description

Greeneridge Sciences, commissioned by SEPCO, collected acoustic data in the Beaufort Sea during 2007–2014 to monitor potential effects of oil exploration on bowhead whales. Data were collected using DASARs (Directional Autonomous Seafloor Acoustic Recorders), whose directional capability allowed localization of calls through triangulation (Greene et al. 2004). Each year during the bowhead whale migration westward through the Beaufort Sea (e.g., Harwood et al. 2017), up to a total of 40 DASARs were deployed at five sites (Fig. 4) in July or August and retrieved in September or October, obtaining continuous acoustic recordings. The geometry of the normal configuration at each site was seven DASARs arranged in a triangular grid with 7 km spacing between sensors (Figs. 1, 5), although some sites had as few as three and as many as 13 sensors during some years (Appendix 2).

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

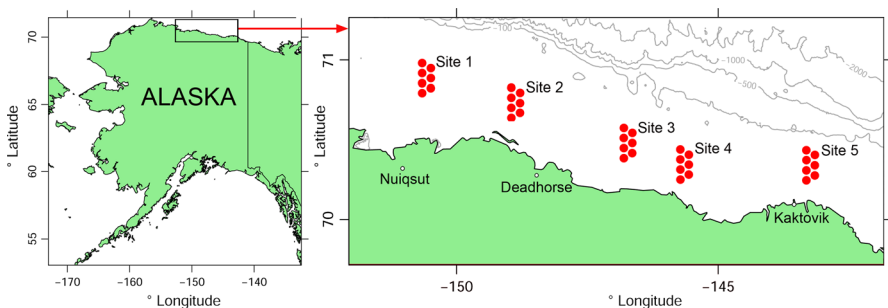


Fig. 4 Study area and DASARs in their normal configuration at sites 1–5 shown in red, land shown in green, depth contour lines in grey (100 m, 500 m, 1000 m and 2000 m)

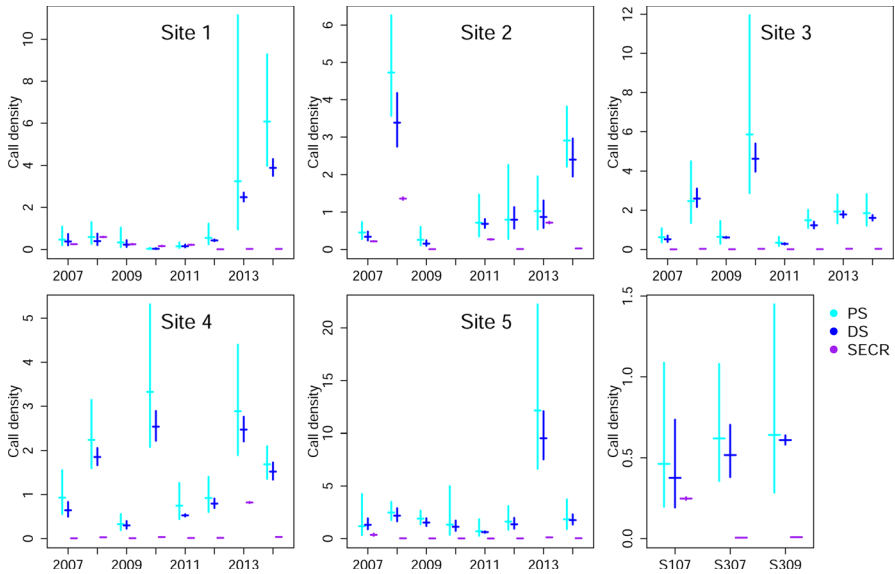


Fig. 5 Bowhead whale call density (number of calls/km²/day) in 2007–2014 for each site estimated with PS (light blue), DS (dark blue), and SECR (purple); horizontal lines represent the estimates, vertical lines the 95% CIs. Bottom right plot displays the three focal site-year combinations

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

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

In 2007 the entire season was inspected manually, whereas in 2008–2012 and 2014, 5–9 full days (midnight to midnight) spread throughout the season in the respective year were inspected. The chosen days were judged to be representative of the varying levels of natural and anthropogenic noise each year, as well as the varying numbers of whale calls detected. In 2013 all data from six selected days were inspected for sites 1 and 2, but due to the huge numbers of whale calls detected, a modified inspection regimen was adopted for sites 3, 4, and 5. Each hour of a

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

4.2 Analysis

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

For PS, no detection function was fitted; therefore, estimates of uncertainty (95% confidence intervals (CIs)) represent only variance due to encounter rate. This was estimated using the P3 estimator from Fewster et al. (2009), which is the standard encounter rate variance estimator and the default method of the *Distance::ds* function for DS point transect analyses (Miller 2017). For DS, the uncertainty from the detection function, estimated using the *Distance::ds* function, also contributed to the estimate of the uncertainty of call density and both components were combined using the delta method (Buckland et al. 2001, p. 76). Log-normal confidence intervals for call density were produced for PS and DS using methods described in Buckland et al. (2001, pp. 77–78) which take into account the small number of samplers (DASARs). For SECR, density in general or call density in the PAM context is one of the model parameters and, hence, asymptotic estimates of uncertainty are based on the inverse of the information matrix from maximizing the unconditional likelihood and are reported as outputs by the *secr* function of the *secr* R library (Borchers and Efford 2008).

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

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

Site	Year	DASARs	Days	PS		DS		SECR	
				Calls	Detections	Calls	Detections	Calls	Detections
1–5	2007–14	278	444	171,252	185,010	470,594	2,151,254	686,192	3,091,842
1	2007	5	49	5216	5696	8034	23,376	13,017	29,603
3	2007	7	46	9467	10,076	22,625	97,613	32,473	130,071
3	2009	7	8	1654	1807	6082	30,527	9634	50,244

4.3 Results

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

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

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

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

	Detection functions	Proportion plots	Estimate	PS	DS	SECR	
S107			D_c	0.46 (0.20, 1.09)	0.38 (0.19, 0.73)	0.25 (0.24, 0.26)	
			g_0	■			1.00 (0.55, 1.00)
			σ	■			6.36 (6.33, 6.39)
S307			D_c	0.62 (0.36, 1.08)	0.52 (0.38, 0.70)	0.0054 (0.0053, 0.0054)	
			g_0	■			1.00 (0.98, 1.00)
			σ	■			9.68 (9.65, 9.71)
S309			D_c	0.64 (0.28, 1.45)	0.61 (0.58, 0.64)	0.0089 (0.0087, 0.0091)	
			g_0	■			1.00 (1.00, 1.00)
			σ	■			12.23 (12.14, 12.32)

Fig. 6 Results from analyzing the bowhead whale data including estimates and 95% CIs of call density D_c , and parameters g_0 and σ estimated with the three methods for three example sites. ■ indicates NA. Red, scaled histogram of distances (km) to calls detected by two or more sensors, overlain with the estimated DS (blue line) and SECR (purple line) detection functions. Green proportion plots depicting the proportion of calls detected by 1–7 sensors

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

4.4 Comparison with simulation study

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

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

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

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

5 Discussion

5.1 How non-independence of detections affects SECR

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

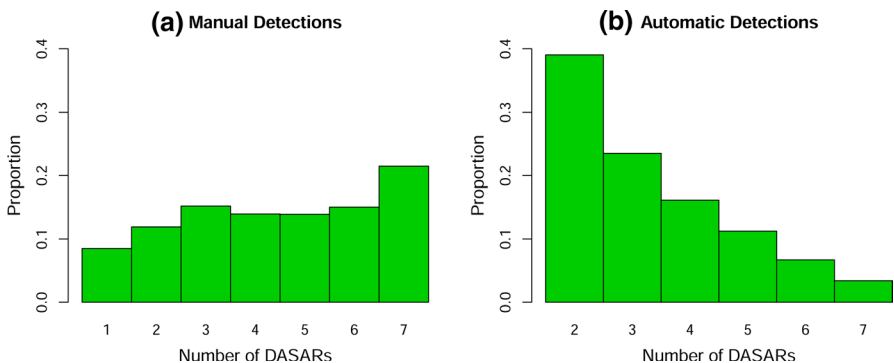


Fig. 7 Proportion plots of the bowhead whale detections for all sites-years with normal DASAR configuration combined: **a** for the manually detected calls and **b** for the automatically detected calls (singletons not shown, see text)

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

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

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

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

5.2 Relationships between PS and DS

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

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

5.3 Estimated vs assumed detection probability at or near the DASAR

Lastly, we note that $\hat{g}_{0SECR} = 1$ was not considered sufficient proof that the assumptions of $g_{0PS} = 1$ or $g_{0DS} = 1$ were met as the latter two require certain detection by at least one more DASAR for localization. However, we assumed that a violation of this assumption would yield smaller call density estimates for PS and DS in comparison to SECR. Further research will be conducted to investigate this using a mark-recapture DS (MRDS, Borchers 2012) approach where g_{0MRDS} is estimated for each DASAR individually using the detections of the other DASARs at the same site as trials (Oedekoven et al., unpublished data). For S307, for example, g_{0MRDS} of DASAR A is estimated using the detections made by the other DASARs at S307, i.e., DASARs B–G, as trials for A.

6 Conclusion

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

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

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

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

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

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

References

- Acevedo MA, Villanueva-Rivera JT (2006) Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. *Wildl Soc Bull* 34(1):211–214
- Barlow J, Taylor B (2005) Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey. *Mar Mamm Sci* 21:429–445
- Blackwell SB, Nations CS, McDonald TL, Greene CR Jr, Thode AM, Guerra M, Macrander AM (2013) Effects of airgun sounds on bowhead whale calling rates in the Alaskan Beaufort Sea. *Marine Mammal Sci* 29(4):E342–E365
- Blackwell SB, Nations CS, McDonald TL, Thode AM, Mathias D, Kim KH, Greene CR Jr, Macrander AM (2015) Effects of airgun sounds on bowhead whale calling rates: evidence for two behavioral thresholds. *PLoS ONE* 10(6):e0125720
- Borchers DL, Buckland ST, Zucchini W (2002) Estimating animal abundance: closed populations. Springer, New York
- Borchers DL, Efford M (2008) Spatially explicit maximum likelihood methods for capture-recapture studies. *Biometrics* 64:377–438
- Borchers DL (2012) A non-technical overview of spatially explicit capture–recapture models. *J Ornithol* 152:435–444
- Bochers DL, Marques TA, Gunnlaugsson T, Jupp P (2010) Estimating distance sampling detection functions when distances are measured with errors. *J Agric Biol Environ Stat* 15(3):346–361
- Brunoldi M, Bozzini G, Casale A, Corvisiero P, Grosso D, Magnoli N, Alessi J, Bianchi CN, Mandich A, Morri C, Povero P, Wurtz M, Melchiorre C, Viano G, Cappanera V, Fanciulli G, Bei M, Stasi N, Taiuti M (2016) A permanent automated real-time passive acoustic monitoring system for bottlenose dolphin conservation in the Mediterranean Sea. *PLoS ONE* 11(1): e0145362
- Buckland ST (2006) Point-transect surveys for songbirds: robust methodologies. *Auk* 123:345–357

- Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) Introduction to distance sampling—estimating abundance of biological populations. Oxford University Press, Oxford
- Buckland ST, Rexstad E, Marques TA, Oedekoven CS (2015) Distance sampling: methods and applications. Methods in statistical ecology. Springer, New York
- Dawson DK, Efford MG (2009) Bird population density estimated from acoustic signals. *J Appl Ecol* 46:1201–1209
- Efford MG (2017) secrlinear: Spatially explicit capture-recapture for linear habitats. R Package Version 1(1):1
- Efford MG (2019) secr: Spatially explicit capture-recapture models. R package version 3.2.0
- Fewster RM, Buckland ST, Burnham KP, Borchers DL, Jupp PE, Laake JL, Thomas L (2009) Estimating the encounter rate variance in distance sampling. *Biometrics* 65:225–236
- Glenne R, Buckland ST, Thomas L (2015) The effect of animal movement on line transect estimates of abundance. *PLoS ONE* 10(3):e0121333
- Greene CR Jr, McLennan MW, Norman RG, McDonald TL, Jakubczak RS, Richardson WJ (2004) Directional frequency and recording (DIFAR) sensors in seafloor recorders to locate calling bowhead whales during their fall migration. *J Acoust Soc Am* 116(2):799–813
- Harwood LA, Quakenbush LT, Small RJ, George JC, Pokiak J, Pokiak C, Heide-Jørgensen MP, Lea EV, Brower H (2017) Movements and inferred foraging by bowhead whales in the Canadian Beaufort Sea during August and September, 2006–12. *Arctic* 70(2):161–176
- Heupel MR, Simpfendorfer CA, Hueter RE (2004) Estimation of shark home ranges using passive monitoring techniques. *Environ Biol Fishes* 71:135–142
- Hildebrand JA, Baumann-Pickering S, Frasier KE, Trickey JS, Merckens KP, Wiggins SM, McDonald MA, Garrison LP, Harris D, Marques TA, Thomas L (2015) Passive acoustic monitoring of beaked whale densities in the Gulf of Mexico. *Scientific Reports*, 5: Article number: 16343
- Jaramillo-Legorreta A, Cardenas-Hinojosa G, Nieto-Garcia E, Rojas-Bracho L, Ver Hoef J, Moore J, Tregenza N, Barlow J, Gerrodette T, Thomas L, Taylor B (2017) Passive acoustic monitoring of the decline of Mexico's critically endangered vaquita. *Conserv Biol* 31:183–191
- Kalan AK, Mundry R, Wagner OJJ, Heinicke S, Boesch C, Kühl HS (2015) Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. *Ecol Ind* 54:217–226
- Lammers MO, Munger LM (2016) From shrimp to whales: biological applications of passive acoustic monitoring on a remote Pacific Coral reef. In: Au W, Lammers M (eds) *Listening in the ocean. Modern acoustics and signal processing*. Springer, New York
- Lewis T, Gillespie D, Lacey C, Matthews J, Danbolt M, Leaper R, McLanaghan R, Moscrop A (2007) Sperm whale abundance estimates from acoustic surveys of the Ionian Sea and Straits of Sicily in 2003. *J Mar Biol Assoc UK* 87:353–357
- Marques TA, Thomas L, Ward J, DiMarzio N, Tyack PL (2009) Estimating cetacean population density using fixed passive acoustic sensors: an example with beaked whales. *J Acoust Soc Am* 125:1982–1994
- Marques TA, Munger L, Thomas L, Wiggins S, Hildebrand JA (2011) Estimating North Pacific right whale *Eubalaena japonica* density using passive acoustic cue counting. *Endangered Species Research* 13:163–172
- Marques TA, Thomas L, Martin SW, Mellinger DK, Jarvis S, Morrissey RP, Ciminello C, DiMarzio N (2012) Spatially explicit capture recapture methods to estimate minke whale abundance from data collected at bottom mounted hydrophones. *J Ornithol* 152:445–455
- Marques TA, Thomas L, Martin SW, Mellinger DK, Ward JA, Moretti DJ, Harris D, Tyack PL (2013) Estimating animal population density using passive acoustics. *Biol Rev* 88:287–309
- Martin SW, Marques TA, Thomas L, Morrissey RP, Jarvis S, DiMarzio N, Moretti D, Mellinger DK (2013) Estimating minke whale (*Balaenoptera acutorostrata*) boing sound density using passive acoustic sensors. *Mar Mamm Sci* 29:142–158
- Mathias D, Thode A, Blackwell S, Greene C (2008) Computer-aided classification of bowhead whale call categories for mitigation monitoring. *New Trends for Environmental Monitoring Using Passive Systems*, Hyeres, French Riviera: 1–6
- McHugh ML (2003) Descriptive statistics, part II: most commonly used descriptive statistics. *J Spec Pediatr Nurs* 8(3):111–116
- Miller DL (2017) Distance: distance sampling detection function and abundance estimation. R package version 0.9.7.
- Miller DL, Rexstad E, Thomas L, Marshall L, Laake J (2019) Distance sampling in R. *J Stat Softw* 89(1):1–28

- Phillips GT (2016) Passive acoustics: a multifaceted tool for marine mammal conservation. PhD Thesis. Duke University, Durham, NC
- Pomerleau C, Patterson T, Luque S, Lesage V, Ferguson S (2011) Bowhead whale *Balaena mysticetus* diving and movement patterns in the eastern Canadian Arctic: implications for foraging ecology. *Endanger Spec Res* 15:167–177
- Singh R, Qureshi Q, Sankar K, Krausman PR, Goyal SP (2014) Evaluating heterogeneity of sex-specific capture probability and precision in camera-trap population estimates of tigers. *Wildl Soc Bull* 38:791–796
- Stevenson BC, Borchers DL, Altwegg R, Swift RJ, Gillespie DM, Measey GJ (2015) A general framework for animal density estimation from acoustic detections across a fixed microphone array. *Methods Ecol Evol* 6:38–48
- Thode AM, Kim KH, Blackwell SB, Greene CR, Nations CS, McDonald TL, Macrander AM (2012) Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys. *J Acoust Soc Am* 131(5):3726–3747
- Thode AM, Blackwell SB, Conrad AS, Kim KH, Marques T, Thomas L, Oedekoven CS, Harris D, Bröker K (2020) Roaring and repetition: how bowhead whales adjust their call density and source level (Lombard effect) in the presence of natural and seismic airgun survey noise. *J Acoust Soc Am* 147:2061–2080
- Thomas L, Marques TA (2012) Passive acoustic monitoring for estimating animal density. *Acoust Today* 8(3):35–44
- Van Parijs SM, Clark CW, Sousa-Lima RS, Parks SE, Rankin S, Risch D, Van Opzeeland IC (2009) Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales. *Mar Ecol Prog Ser* 395:21–36
- Vilchis LI, Ballance LT, Fiedler PC (2006) Pelagic habitat of seabirds in the eastern tropical Pacific: effects of foraging ecology on habitat selection. *Mar Ecol Prog Ser* 315:279–292
- Wrege PH, Rowland ED, Keen S, Shiu Y, Matthiopoulos J (2017) Acoustic monitoring for conservation in tropical forests: examples from forest elephants. *Methods Ecol Evol* 8:1292–1301

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

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

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

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


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

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

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

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

Authors and Affiliations

Cornelia S. Oedekoven¹  · **Tiago A. Marques**^{1,2} · **Danielle Harris**¹ · **Len Thomas**¹ · **Aaron M. Thode**³ · **Susanna B. Blackwell**⁴ · **Alexander S. Conrad**⁴ · **Katherine H. Kim**⁴

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

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

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

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

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