Fin whale density and distribution estimation using acoustic bearings derived from sparse arrays

Authors: Danielle Harris[a], Jennifer Miksis-Olds[b,c], Julia Vernon[b], Len Thomas[a]

Affiliations:

(a) Centre for Research into Ecological and Environmental Modelling, The Observatory, Buchanan Gardens, University of St. Andrews, St. Andrews, Fife, KY16 9LZ, United Kingdom.

(b) Applied Research Laboratory, The Pennsylvania State University

(c) School of Marine Science and Ocean Engineering, University of New Hampshire

Date resubmitted: 20 October 2017

Running title: Fin whale density and distribution estimation.

Abstract

Passive acoustic monitoring of marine mammals is common, and it is now possible to estimate absolute animal density from acoustic recordings. The most appropriate density estimation method depends on how much detail about animals' locations can be derived from the recordings. Here, a method for estimating cetacean density using acoustic data is presented, where only horizontal bearings to calling animals are estimable. This method also requires knowledge of call signal-to-noise ratios (SNR), as well as auxiliary information about call source levels, sound propagation, and call production rates. Results are presented from simulations, and from a pilot study using recordings of fin whale (*Balaenoptera physalus*) calls from Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) hydrophones at Wake Island in the Pacific Ocean. Simulations replicating different animal distributions showed median biases in estimated call density of less than 2%. The estimated average call density during the pilot study period (December 2007 - February 2008) was 0.02 calls.hr^{-1}.km^{2} (coefficient of variation, CV: 15%). Using a tentative call production rate, estimated average animal density was 0.54 animals/1000 km^{2} (CV: 52%). Calling animals showed a varied spatial distribution around the northern hydrophone array, with most detections occurring at bearings between 90 and 180 degrees.

I. INTRODUCTION

Using acoustic data to estimate animal density has been demonstrated for both terrestrial and marine species (e.g., Buckland, 2006; Marques *et* al., 2013, Stevenson *et al*., 2015). A suite of density estimation methods exist that can be applied to different types of acoustic survey data. The most appropriate density estimation method depends on how much detail about animals' locations can be derived from the recordings, which is often determined by the number and configuration of deployed instruments. At best, three-dimensional locations of calling animals can be estimated from acoustic data; conversely some recordings can yield little to no information about animals' locations.

Distance sampling (Buckland *et al*., 2001) and spatially explicit capture-recapture (SECR; e.g., Borchers, 2012) are methods that estimate the probability of detecting animals (a key parameter of any animal density estimation method) using spatial data collected during the survey. Specifically, distance sampling can be used when the horizontal range between an instrument and a calling animal can be estimated (e.g., Marques *et al*., 2011), which, for marine animals, typically requires animal depth to be estimable (or assumed). SECR requires that the same acoustic event is matched across multiple recorders, creating "capture histories" of acoustic events. Indirect information about the location of calling animals can be inferred from these capture histories by assessing which recorders (with known locations) detected the acoustic events. Although SECR does not need measured ranges, SECR analyses can be supplemented with data relating to animals' locations such as direction, received sound level and time of arrival (Borchers *et al.*, 2015). Given their data requirements, both distance sampling and SECR require arrays of recorders to estimate detection probability (though horizontal ranges to calling animals can, in some particular scenarios, be estimated from single instruments, e.g., Harris *et al.*, 2013; Marques *et al.*, 2011; Tiemann *et al*, 2004).

Conversely, when no spatial information can be estimated from recorded data (e.g., in most scenarios where single instruments are deployed), detection probability can be estimated using some form of auxiliary data. Marques *et al.* (2013) consider two types of auxiliary information: (1) a sample of measured animal locations in relation to a recorder either from animal-borne tags (e.g., Marques *et al*., 2009) or combined visual and acoustic trials using focal animals (e.g., Kyhn *et al*., 2012); (2) acoustic modeling using elements of the passive sonar equation (Urick, 1983) including information about the target species' call source level, transmission loss, ambient noise levels, and the efficiency of the detection and classification process. This latter information can be combined to estimate the probability of detection using a simulation-based framework (e.g., Küsel *et* al., 2011). Monte Carlo simulations have been implemented for a range of cetacean species (Küsel *et al*, 2011; Harris, 2012; Helble *et al*., 2013; Frasier *et al*., 2016) but rely on accurate simulation inputs. One such input is the distribution of simulated animals; however, there are often no *a priori* data about what this distribution should be. This is a key limitation of the Monte Carlo simulation approach.

Here, a new method is presented for estimating cetacean density using acoustic data, for cases where horizontal bearings to calling animals are estimable. This approach is suitable for scenarios where neither distance sampling nor SECR can be implemented, due to lack of recorders (note that SECR survey design is an ongoing area of research but, to date, the minimum number of recorders used for acoustic capture histories has been three, Kidney *et al*., 2016). The new method is related to the Monte Carlo simulation methodology as it uses the passive sonar equation; measured call signal-to-noise ratios (SNR) are required, as well as auxiliary information about call source levels, sound propagation, and call production rates. However, the additional bearing data give some empirical information about animal distribution, conferring an advantage over the standard Monte Carlo simulation. Another advantage of this method is that it produces a spatial map of estimated abundance (or density), allowing inferences about spatial habitat preferences of acoustically active animals.

The paper is structured as follows. Section II presents a background to density estimation using acoustic data, and a description of the new method. Details about the motivating case study – fin whales recorded in the Pacific Ocean by Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) hydrophones – are given in Section III (including details of all the required auxiliary analyses). Simulations are presented, which investigate method performance under different known spatial animal distributions (Section IV). The method is then applied to three months of recordings from Wake Island between December 2007 and February 2008 (Section V). This analysis forms a pilot study prior to applying the method to long-term CTBTO datasets from Wake Island and Diego Garcia in the Indian Ocean. Finally, Section VI presents a discussion of the approach, including its limitations, benefits, and potential implementations.

II. DENSITY ESTIMATION USING ACOUSTIC DATA

A general estimator of animal density using acoustic cues (e.g., animal calls) from static instruments was presented by Marques *et al.* (2009) (Eqn. 1):

$$\hat{D} = \frac{n_c(1-\hat{c})}{K\pi w^2 \hat{P}_a T\hat{r}} \qquad \text{(Eqn. 1)}$$

where $\hat{D}$ = call density, n_c = number of detected signals, $\hat{c}$ = false positive proportion, K = number of monitoring points, w = maximum detection range, $\hat{P}_a$ = average probability of detection of an animal within radius w of the sensor, T = total monitoring time and $\hat{r}$ = cue production rate . This equation can be decomposed into three components:

$$\widehat{D} = \frac{n_c(1-\hat{c})}{\hat{P}_a} \times \frac{1}{K\pi w^2} \times \frac{1}{T\hat{r}} \quad \text{(Eqn. 2)}$$

where $\widehat{N}_c = n_c(1-\hat{c})/\hat{P}_a$ is the estimated abundance of cues, $K\pi w^2$ is the area monitored, so that dividing the abundance of cues by the area monitored gives a density of cues, and $1/T\hat{r}$ converts the density of cues to the density of animals.

The average probability of detection, $\hat{P}_a$, can be estimated in several ways, as shown by the variety of available density estimation methods (Marques et al. 2013). Each method has various assumptions that must be met to produce an unbiased detection probability and hence density. One key assumption in distance sampling is that the distribution of animals' distances from samplers (i.e., transect lines in a line transect survey, or monitoring points in a point transect survey) is known. This is achieved by random placement of multiple samplers within the study area so that, on average, animals are distributed uniformly in horizontal space. For a survey using many fixed monitoring points with circular detection areas, this assumed average distribution of animal distances is specifically a triangular distribution due to the linear increase in area with increasing incremental horizontal distance from each sample point (Buckland *et* al., 2001). However, when single acoustic stations are used, it may not be reasonable to assume animal distances from that single station follow a triangular distribution, and standard distance sampling should not be used to estimate $\hat{P}_a$ (even if ranges to animals can be estimated). Therefore, an alternative approach to estimating detection probability is required. In the method developed here, cue abundance is estimated using a Horvitz-Thompson-like estimator (after terminology used by Borchers & Burnham, 2004). These estimators are based on seminal work by Horvitz & Thompson (1952), who showed that when sampling at random from a population where each individual, *i*, has probability P_i of being sampled, then an unbiased estimator of population size is given by the sum over detected individuals of $1/P_i$. One can think of each detection "representing", on average,

$1/P_i$ objects in the population. In animal density estimation methods, individual detection probabilities for every detection can be estimated (rather than estimating an average detection probability as shown in Eqn. 1) and combined to give $\hat{N}_c = \sum_{i=1}^{n_c} 1/\hat{P}_i$. However, the detection probabilities, P_i, are estimated, not known (hence "Horvitz-Thompson-like"). Horvitz-Thompson-like estimators are not unbiased; the bias is typically small unless estimated probabilities are highly uncertain or close to zero (Borchers *et al*., 2002). The key advantage of this approach in the current case is that the individual detection probabilities can be estimated without requiring any assumption about the distribution of animals with respect to the samplers.

Other key assumptions that apply to this new method are that (1) all data measurements and derived parameters are accurate and (2) detections are independent of one another. It is highly improbable that recorded whale calls are produced independently of each other, given that one animal may produce many calls. However, violation of the independence assumption should not produce severe bias, though variance estimation can be affected (Marques *et al*., 2013). Another assumption of any density estimation method is that parameters used in the estimator are accurate for the time and place of the main survey. A frequent limitation of auxiliary data used in density estimation analyses is that the additional experiments (e.g., to estimate cue production rate) may have been conducted in a limited part of the study area (or in a different location) and/or at a different time as the main survey, which may lead to bias in the estimated parameters. Therefore, as many auxiliary analyses should be undertaken using data from the main survey region and time period as possible.

A. Method overview

It is assumed that acoustic data have been recorded at known locations for a known time and then processed using an automated detection and classification algorithm.

Estimation proceeds in the following stages, described in more detail in the next subsection.

1. Characterize the automatic detection process to estimate the probability of detecting a call as a function of SNR ($P(SNR)$). The resulting fitted "detection characterization curve" is used to estimate the detection probability for each detected signal.

2. Determine the monitored area: for each of a set of discrete bearings, use the assumed call source level (SL) and the measured noise level (NL) distributions with a transmission loss (TL) model to determine a set of ranges at which calls are almost certain to be masked (i.e., the resulting SNR is so low that probability of detection is very low) and exclude these areas from further analysis.

3. Estimate the distribution of possible ranges for each detection. Use the measured received level (RL) and bearing of each detection, together with the assumed SL distribution and TL model to estimate the probability density function (pdf) of ranges for that detection. A probabilistic approach is required because (a) source level for each detection is not assumed known, but is assumed to come from a probability distribution; (b) even if source level were assumed known, the TL does not increase monotonically with range and hence a detected signal with a given RL can correspond to more than one range.

4. Estimate the range-specific distribution of number of signals corresponding to each detection, i.e., scale each detection by its associated detection probability to account for undetected signals. Using the Horvitz-Thompson-like estimator, each detection, i, on average corresponds to $1/P(SNR_i)$ signals within the area monitored.

5. Estimate spatial density of signals by summing over the estimated number of signals at each bearing and range to yield an empirical spatially-explicit abundance of signals. Then smooth this using a Generalized Estimation Equation (GEE) spatial model.

6. Estimate animal density: use additional multipliers i.e., false positive proportion, time spent monitoring (excluding periods of high ambient noise that cause masking) and cue rate (Eqn. 1). Also potentially restrict inference to areas where detection probability is higher and hence inference more reliable.

B. Further details

Stage 1: Characterize the automatic detector. Detector characterization is performed using a sample of manually-detected calls. To ensure the sample is representative, a systematic random subset of recordings (i.e., short sections equally spaced in time – see Section III for an example) should be analysed manually. SNR is measured for a sample of manually-detected calls, as well as noting whether or not each call was detected by the automatic detector. Logistic regression with automated detection/non-detection as the response and SNR as the explanatory variable is used to model the probability of detecting a call as a function of SNR. A Generalized Additive Model (GAM, Wood, 2006) is used to allow a smooth, nonlinear relationship between probability of detection and SNR. The fitted detector characterization curve is then used to predict probability of detection, $P(SNR)$, for each detection (over the entire monitoring period), $\hat{P}_i = \hat{P}(SNR_i)$.

If bearings cannot be estimated for all detections, one of two approaches can be taken: the detector characterization curve can be estimated where a successful detection is defined as either (1) any detected fin whale call (regardless of whether it had an associated bearing or not), or (2) detected fin whale calls that had an associated bearing measurement. The choice of detector characterization approach will affect the value used for n_c in the estimator (Eqn.

1). Under the first definition, n_c will be the number of detections (with or without measured bearings); under the second definition, n_c will be the number of detections with measured bearings only. In both cases, an assumption is made that the measured bearings represent the spatial distribution of all detected signals, including those for which bearings could not be estimated.

Stage 2. Determine area monitored. This stage is analogous to identifying the maximum detection range, w, in Eqn. 1, although a set of bearing-specific ranges are derived, allowing TL to vary in different directions, and be non-monotonic with increasing range. Hence the area monitored does not have to be circular or continuous.

SL is assumed to follow a normal distribution; so it is theoretically possible to detect calls from implausibly large (or even infinite) ranges in Stage 3. Therefore, a pragmatic cut-off is used that ensures detections from outside the area monitored will be very rare. The assumed SL and NL distributions are evaluated at the 90th and 10th percentiles, respectively, to represent a loud call in low noise. These values are used in the passive sonar equation along with TL to calculate the SNR of the hypothetical call at various range and bearing steps around the hydrophone (SL – TL – NL = SNR). The detection probability of the call at all locations is evaluated from the detector characterization curve. Locations where the call has a detection probability of equal to or less than 0.1 are considered to be acoustically masked. The lowest TL associated with a masked location is used as a TL threshold to define acoustically masked areas, which are then excluded from the remainder of the analysis.

Stage 3. Estimate distribution of possible ranges for each detection. Given a detection with measured RL and bearing θ, the SL of the detection if the source was at range r can be derived from the (simplified) passive sonar equation as

$$SL(r,\theta) = RL + TL(r,\theta) \qquad \text{(Eqn. 3)}$$

where $TL(r, \theta)$ is range- and bearing-specific transmission loss. An SL distribution is required, which is assumed to follow a normal distribution with mean μ and standard deviation σ. In this analysis, SL could be estimated from a subsample of localized calls at short ranges. Then, the pdf of range is

$$f(r|RL, \theta) = \frac{r}{\nu} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(SL(r,\theta)-\mu)^2}{2\sigma^2}} \quad \text{(Eqn. 4)}$$

where ν is a normalizing constant to ensure *f* is a proper pdf:

$$\nu = \int_{r=0}^{w} \frac{r}{\sqrt{2\pi\sigma^2}} e^{-\frac{(SL(r,\theta)-\mu)^2}{2\sigma^2}} dr \quad \text{(Eqn. 5)}$$

The need for an *r* in the denominator of Eqn. 4 is explained by viewing the analysis as analogous to distance sampling with measurement error on the distances. In this case, the geometry of a circular detection area means that random measurement error (in this case, uncertainty in location) will result in underestimation of detections' true locations (discussed in Buckland *et al*., 2015), leading to biased density estimates at closer ranges.

In practice, range is discretized into a fixed set of range intervals, with midpoints $\{R\}$. TL is calculated at these ranges, and it is assumed that the TL values apply to each corresponding interval. Then, the probability a detection comes from interval *k* is

$$\Pr(k|RL, \theta) = \frac{f(R_k|RL,\theta)}{\sum_{R_j \in R} f(R_j|RL,\theta)} \quad \text{(Eqn. 6)}$$

Stage 4. Estimate range-specific distribution of number of signals corresponding to each detection. SNR for each detected signal is calculated from the RL and NL measurements associated with each signal (SNR = RL – NL). Detection probabilities of each detected signal are estimated using the detector characterization curve and the range-specific distribution for each detection is divided by the estimated detection probability. Using the

Horvitz-Thompson-like approach, the estimated number of signals in the population "represented" by a signal detected with a given SNR is $1/P(SNR)$. Hence, the range-specific distribution of number of signals corresponding to a particular detection is given by

$$N_c(k|RL, NL, \theta) = \frac{\Pr(k|RL,\theta)}{P(SNR)} \quad \text{(Eqn. 7)}$$

Stage 5. Estimate spatial density of signals. At each bearing and range interval, the estimated number of signals are summed. This yields a spatial abundance surface, but one that is not necessarily smooth because of random variation in detections. Given a long monitoring period, the true distribution of calls around the sensor likely is smooth, so precision can be gained by smoothing the raw estimates using a GEE model (Hardin & Hilbe, 2012), which accounts for spatial autocorrelation. The response variable is the estimated signal abundance, assuming an overdispersed quasipoisson error distribution and using a log link function. Explanatory variables are the location of the centre of the bearing and range interval in (*x*,*y*) space (2-dimensional Cartesian coordinates). To account for the fact that intervals at larger ranges represent a larger area, the area of each interval is included as an offset in the model. To account for spatial autocorrelation, spatial blocks of 100 km x 100 km are created through the study area and an independent working correlation structure implemented; model residuals can therefore be correlated within blocks but are assumed to be independent between spatial blocks. The spatial GEE is fitted using CReSS (Complex Region and Spatial Smoother, Scott-Hayward *et* al, 2014) and SALSA (Spatially Adaptive Local Smoothing Algorithm, Walker *et al*., 2011) methods, allowing a flexible surface with spatially-varying smoothness to be modeled. Model fit is assessed using concordance correlation and marginal R squared values (in both cases, values close to 1 indicates good fit). A predicted density surface is created by predicting abundance on a regular (*x*,*y*) grid, and dividing by the area of each grid cell.

Stage 6. Estimate animal density. The predicted density surface of signals is converted to a predicted animal density surface by multiplying by $(1-\hat{c})/T\hat{r}$, where *c* is the false positive proportion, *T* is monitoring time, and *r* the cue production rate. False positive proportion is estimated from the manually-validated sample of data. Monitoring time should be known as part of the survey protocol. Furthermore, the NL measurements of the detections can be compared to ambient NL measured throughout the dataset to determine a NL threshold, above which total acoustic masking is likely to occur. Time periods of data where ambient NL exceeds the maximum NL associated with a detection are omitted from the monitoring time, *T*. Cue production rate must come from auxiliary information and is often not known, in which case density of calls can be estimated but not density of animals.

Average density can be computed by taking the average across the prediction surface. To increase robustness, grid cells far from the sensor, where detection probability is low, may be excluded from this averaging. A Horvitz-Thompson-like estimator is known to produce positively biased estimates, particularly when some of the $\hat{P}_i$ values are small (Borchers *et al.* 2002) as is the case for more distant calls. To mitigate this, a simulation study can be used to determine at what range bias may be minimised and this can be used to truncate the range over which average density is inferred.

C. Variance estimation

The delta method (Seber, 1982) is used to combine the coefficients of variation (CVs) for each random variable used in the density estimator to estimate the overall CV for the resulting density estimate. Note that the encounter rate also contributes to the overall variance of a density estimate, and is denoted by $CV(n_c)$ in Eqn. 8. All other density estimator inputs such as *K*, *T* and *w* are known constants and therefore do not have an associated variance.

$$CV(\hat{D})^2 = CV(n_c)^2 + CV(\hat{c})^2 + CV(\hat{P}_a)^2 + CV(\hat{r})^2 \qquad \text{(Eqn. 8)}$$

where: $\hat{P}_a$ = overall mean probability of detection, defined as

$$n_c / \left(\sum_{i=1}^{n} 1/\hat{P}_i \right) \qquad \text{(Eqn. 9)}$$

In surveys with multiple samplers (i.e., monitored line or points), between-sampler variance in encounter rate is usually estimated. With only one monitoring point as in this study, there is no spatial variance in encounter rate and, instead, variance in encounters is linked only to the detection process. Following guidance in Buckland *et* al., (2001), the encounters are assumed to follow an overdispersed Poisson distribution. Therefore, encounter variance can be estimated using the Poisson expression for variance (multiplied by a factor of 2 to acknowledge assumed aggregation in the encounters) (Eqn. 10), which can then be used to calculate the CV:

$$var(n_c) = 2n_c \qquad \text{(Eqn. 10)}$$

The false positive proportion and call production rate have weighted means (see Section III for details) so variance is estimated using Cochran's approximation (Cochran, 1997, recommended by Gatz and Smith, 1995). Detection probability variance is estimated using parametric bootstrapping of the SL and NL distributions, the coefficients of both the logistic regression and GEE spatial models, then taking the empirical variance of the resulting bootstrapped signal densities. As these signal densities are uncorrected for false positives, the only parameter used in their estimation is $\hat{P}_i$, and so the signal density CV will be equivalent to $CV(\hat{P}_a)$.

III. CASE STUDY - FIN WHALES IN THE PACIFIC OCEAN

The pilot study focused on fin whale calls recorded in the Pacific Ocean. Fin whales, the second largest cetacean, occur globally and are currently listed as "Endangered" in the IUCN

Red List (Reilly *et al*., 2013). Fin whales produce a low-frequency pulsed call, the "20-Hz" call (Watkins *et al*., 1987), which has been widely utilized to investigate fin whales' distribution and density through passive acoustic monitoring (e.g., Širović *et al*., 2015). In particular, a study of fin whales near Oahu, Hawaii, was an early example of using passive acoustic data to estimate density (McDonald & Fox, 1999). Multipath arrivals and the passive sonar equation were both used to estimate ranges to calling animals. However, neither detection probability nor non-calling animals were explicitly accounted for, so the resulting estimates were interpreted as a minimum number of animals (McDonald & Fox, 1999).

Data from the CTBTO IMS station at Wake Island (station identifier: H11) in the Equatorial Pacific Ocean were used (1) as a basis for simulation studies to test the efficacy of the method and (2) to demonstrate a pilot analysis using fin whale 20 Hz calls. Data from peak seasonal detections from Dec. 1, 2007 to Feb. 29, 2008 were used, and details of data processing and auxiliary analyses are given throughout the rest of this section.

A. Wake Island CTBTO IMS station

The Wake Island station is composed of two 3-element triangular arrays with 2.5 km spacing between elements, with three hydrophones located to the north of the island (Fig. 1) and three to the south. These cabled hydrophones are suspended in the deep sound channel. The three-month pilot study used data from the northern array (hydrophone depths were 731 m, 732 m, and 729 m). The average water depth at the array was 1068 m (estimated from Amante & Eakins, 2009). Sound levels were recorded continuously at a 250 Hz sampling rate and 24 bit A/D resolution. The hydrophones were calibrated individually prior to initial deployment in January 2002 and re-calibrated while at sea in 2011. All hydrophones had a flat (within 3 dB) frequency response from 8-100 Hz. Information from individual hydrophone response curves was applied to the data to obtain absolute values over the full frequency spectrum (5-115 Hz). Data less than 5 Hz and from 115-125 Hz were not used due to the steep frequency response roll-off at these frequencies.

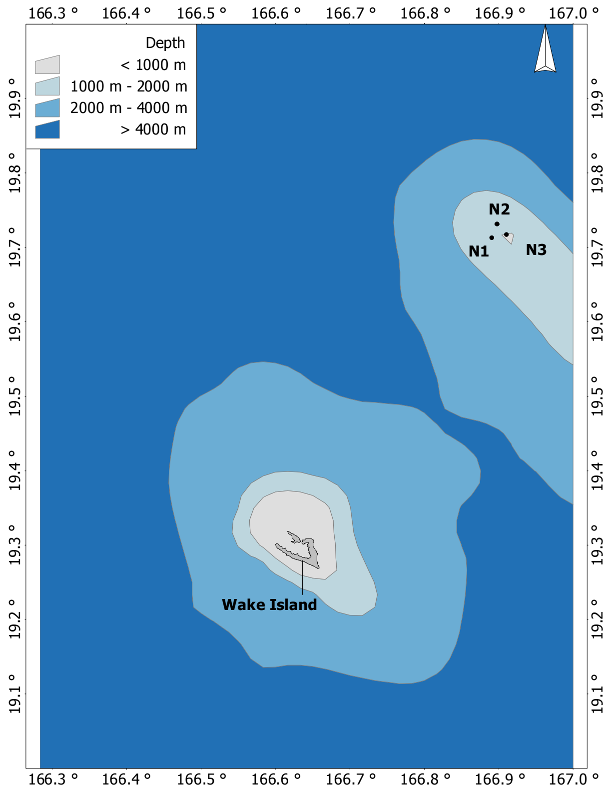

Figure 1. Map showing the location of Wake Island (coordinates: 19.30, 166.63) and the northern hydrophone array. Water depth contours (1000 m, 2000m and 4000 m) are also depicted.

B. Transmission loss of a fin whale call

The transmission loss due to range-dependent propagation between a vocalizing whale using a 20 Hz call and one of the northern hydrophone receivers (labelled N1) at 731 m depth was modelled along 360 bearings at 1^{o} resolution using the OASIS Peregrine parabolic equation model out to 1000 km from N1 (Heaney & Campbell, 2016) (Fig. 2). The transmission loss was modelled at 1 km range steps over the three-month study using seasonal sound speed profiles obtained from The World Ocean Atlas (https://www.nodc.noaa.gov/OC5/indprod.html). It was assumed that the source was at a depth of 15 m, in keeping with results about fin whale calling behavior (Stimpert *et al.*, 2015). The bathymetry was taken from the global bathymetry database ETOPO1 (Amante & Eakins, 2009). Surface loss was negligible due to the low frequency of signals. Sea floor parameters of soft sand sediment were used representing a global average of deep ocean sediment. Details of the geoacoustics parameters in the specific Wake Island region are not known but should not affect propagation in this environment due to direct path/sound channel propagation.

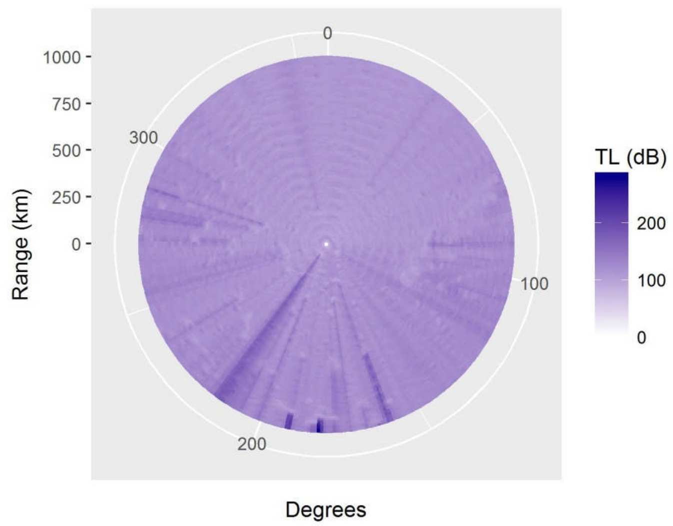

Figure 2. Transmission loss of a 20 Hz signal propagating from Wake Island N1 at a depth of 15 m. The model was run for every bearing between 0 and 359 degrees at 1 km range steps. In this plot, 0 degrees indicates north.

C. Ambient noise levels

Mean spectral levels within the 10-30 Hz band were calculated for each minute of the three-month dataset, resulting in spectral levels with units of dB re 1 $\mu Pa^2/Hz$. Ambient noise levels were calculated in the targeted 10-30 Hz band to directly overlap with the frequency range of the fin whale 20-Hz pulse. Mean spectral levels were calculated using a Hann windowed 15,000 point Discrete Fourier Transform with no overlap to produce sequential 1-min power spectrum estimates. Note that these measurements included fin whale calls, where present; it was important that the noise levels reflected all noise sources that each fin whale call could be exposed to, which included calls by conspecifics.

D. Source level estimation

A sample of fin whale calls were localized using the northern array so that a source level distribution could be estimated. Source level (SL) estimates of detected fin whale vocalizations were computed using the passive sonar equation (Eqn. 11) that incorporated environmental noise levels present at the time of the call within the received level (RL) of the vocalization.

$$SL = RL + TL - DI + DT - PG \qquad \text{(Eqn. 11)}$$

As the low-frequency calls are omnidirectional, the directivity index (DI) was set to zero. Processing gain (PG) and detection threshold (DT) are accounted for in the calibration of the recording system. Received levels were calculated for individual vocalizations recorded at N1 using a custom MATLAB (Mathworks, 2016) code. Spectrograms were calculated using a 512-point FFT and 93.75% overlap. Calls were then manually detected, with a human analyst selecting the upper and lower frequency and time bounds of an individual call. The rms (root-mean- square) RL of the call was then calculated from the selected spectral data.

The TL of a signal of a given frequency is dependent on the range, bearing, and depth of the vocalizing animal. The time difference of arrival (TDOA) between each hydrophone pair was found by cross-correlation of received signals and was supplemented with manual inspection due to dispersed waveforms. 2D hyperbolic localization was then used to find the range and bearing of the vocalizing animal. Location information was then input into the site-specific, seasonal transmission loss models to back calculate the SL of each identified vocalization. The depths of the sources were unknown but assumed to be at a depth of 15 m following results from Stimpert *et al*. (2015). For comparison, source levels of the same sample of calls were also calculated using simple spherical spreading instead of the more complex Peregrine transmission loss model.

E. Automated fin whale call detection

Fin whale calls were detected from the N1 hydrophone using the automatic detection feature of Ishmael, an open-access bioacoustic analysis software package (Mellinger, 2002). The spectrogram correlation method was utilized for the full three-month dataset, cross-correlating the spectrogram of the dataset with a synthetic call kernel. The kernel is a template that indicates the time and frequency endpoints of the desired call. To prepare the dataset for autodetection, time-waveform data were first passed through a 10-30 Hz bandpass filter. Spectral data were then calculated using a 512-point FFT with a 93% overlap, and a 22-14 Hz one-second downsweep call kernel was applied.

Results from the automatic detector were compared with the manually detected calls from a subset of data. The three-month dataset was divided into six-hour sections, and a systematic random sample of these sections was taken. Every 11^{th} six-hour section was selected under the sampling scheme, resulting in 32 six-hour sections. All calls within the 32 selected sections were manually detected, and a receiver-operator curve was generated for the automatic detector that compared the false positive proportion (the number of false positives divided by the total number of automatic detections) with the proportion of missed calls (the number of missed calls divided by the total number of manually detected calls, i.e., false negative proportion) for a range of detection thresholds. The ROC curve indicated that the optimal detection threshold had a 10% false positive proportion and a false negative proportion of 59%. The mean false positive proportion was weighted by the number of detections checked in each six-hour section.

F. Bearing measurements

Bearings were calculated using the TDOA of received signals. Using the known distances between receivers and the seasonal sound speed, an estimated bearing was calculated for each pair of hydrophones (Eqn. 12).

$$\varphi = \arcsin(\tau * c/d) \qquad \text{(Eqn. 12)}$$

where τ represents the TDOA of a signal between a hydrophone pair, d is the distance between a hydrophone pair, and c is the speed of sound.

Left-right ambiguity of each bearing estimate could be resolved by comparing with the other two estimates. The median bearing was then selected. An acceptable bearing is one where the three bearings resulting from the three pair combinations all produced bearings within 10 degrees of each other. TDOA between each pair of hydrophones (N1 and N2, N2 and N3, N3 and N1) were found through three different methods, as described in order of application below. If the cross-correlation method failed to produce an acceptable bearing, manual estimation was performed. When manual estimation using the start point of each call failed to produce an acceptable bearing, a band energy analysis was performed. The first step of all methods was to pass the signals through a 10-30 Hz band pass filter. Bearings were rounded to the nearest integer, to correspond with the resolution of the TL model.

(1) Cross-correlation

Once the data were filtered, a simple cross-correlation was performed in MATLAB to determine time delays. Characteristics of the environment cause dispersion in the waveforms traveling from distant ranges. As a result, a simple cross-correlation was not a viable option for many of the distant calls.

(2) Manual Estimation

TDOA was found by manually selecting the start of each call from the time waveform. Manual inspection eliminates the discrepancies that arise from the modal dispersion. Manual selection also provided reliable results for calls with a low (< 6 dB) signal-to-noise ratio (SNR), which is not always possible with automated methods. Manual detections were feasible for a limited pilot study, but this method would not be appropriate for large datasets.

(3) Band Energy Analysis

Filtered data from N1 were analyzed in 3 Hz bands with 1 Hz overlap, starting at 10 Hz, finding the peak in each band. The first band with a peak of at least 5 dB SNR was then selected. The time index of the first peak in this frequency band for each sensor was then noted and time delays were calculated from the identified time index.

G. Detector characterization

All calls were manually detected in the subsampled six-hour sections. The rms RL of each call was measured, and the SNR of the call was calculated using a noise level measured from the second of data preceding the call (in the same frequency bandwidth as the measured call rms RL). Whether or not the call was detected by the automatic detector was also noted. The detector characterization curve was modeled using the statistical analysis software, R version

3.3.1 (R Core Team, 2016). A GAM (Wood 2006) with a binary response and logit link function was fitted to the data.

H. Call production rate

No call production rate data were available for fin whales occurring near Wake Island, but call production rate data from the Southern California Bight in the North Pacific Ocean have been published (Stimpert *et al*., 2015). The fin whale data from southern California were collected in summer months, and so it is possible that this cue rate is biased for the fin whales calling near Wake Island in the winter months. Cue rates from Stimpert et al. (2015) were applied here as a proof of concept only, and resulting animal density estimates must be treated cautiously.

IV. SIMULATION STUDIES

A. Simulation overview and input data

The primary aim of the simulation studies was to investigate whether the method returned unbiased (1) detection probability estimates and (2) distribution maps under a range of scenarios. To that end, call density only was estimated in the simulations (i.e., a false positive proportion and call production rate were not considered).

Ambient noise and source level information, as well as the detector characterization curve, were measured directly from the Wake Island dataset. The source level distribution (assumed to be normally distributed) had a mean of 177.7 dB re 1 $\mu Pa^2/Hz$ @ 1m (standard deviation: 3.30, n = 79) using the Peregrine transmission loss model and 177.6 dB re 1 $\mu Pa^2/Hz$ @ 1m (standard deviation: 3.03) using spherical spreading to predict propagation loss. Further, estimated source level decreased significantly as a function of range when using the Peregrine model (linear regression coefficient = -2.20, p-value < 0.001, n = 76 due to the removal of three outlying data points using Cook’s distance measures). Estimated source levels assuming spherical spreading also decreased slightly with range, though not significantly (linear regression coefficient = -0.62, p-value = 0.27, n = 76) (Fig. 3). Given that the means and standard deviations of the two source level distributions were almost identical, the source level estimates using the more complex, bathymetry-dependent Peregrine model were used for all simulations and analyses (though see Section VI for a discussion of the regression results). The mean of the noise level distribution (also assumed to be normally distributed) measured in association with manually detected calls was 92.5 dB re 1 $\mu Pa^2/Hz$ (standard deviation: 2.74, n = 1484). The detector characterization curve was estimated using 1484 manually detected calls, which were found in 20 out of 32 manually checked six-hour sections (12 sections contained no calls). The mean SNR of automatically detected calls was 13.98 (standard devation: 7.09, n = 612) and the mean SNR of calls missed by the automatic detector was 4.45 (standard deviation: 1.59, n = 872). The fitted GAM predicted that the majority of calls with an SNR greater than 10 dB were certain to be detected (Fig. 4).

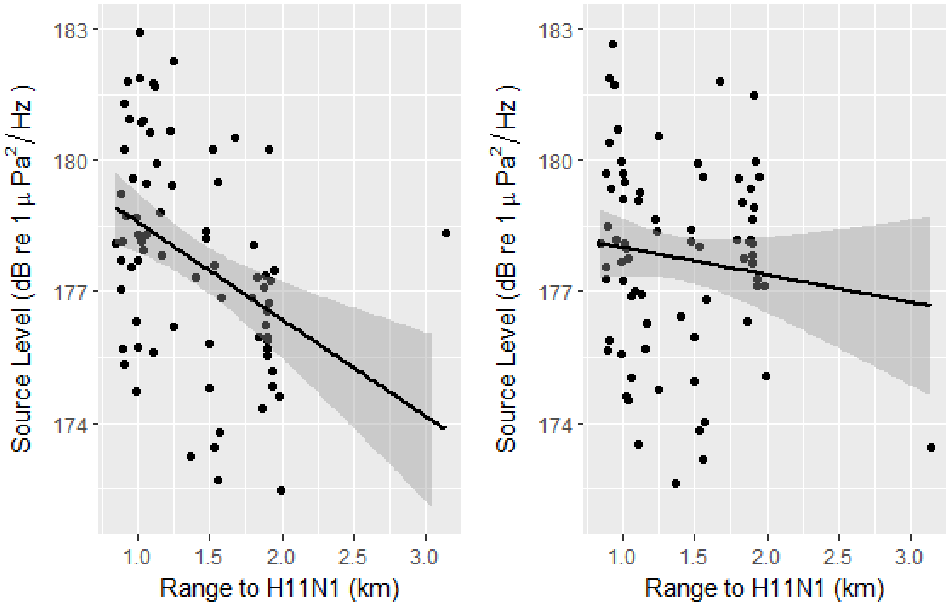

Figure 3. Source levels estimated from 79 calls using transmission loss derived from (left) the Peregrine model and (right) assuming spherical spreading. Both plots show a fitted linear regression model (black line), with associated 95% confidence intervals shaded in gray.

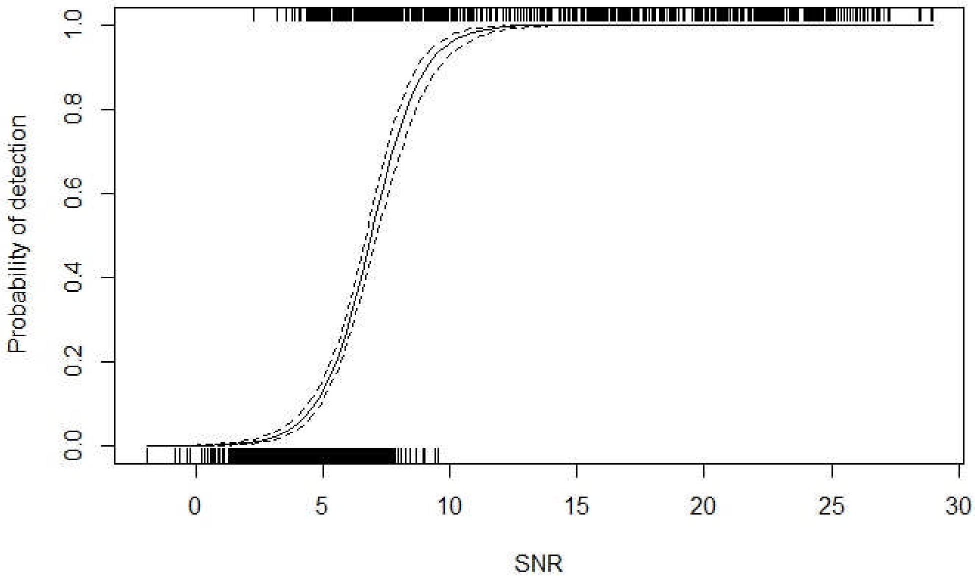

Figure 4. Detector characterization curve (with 95% confidence interval) predicting detection probability as a function of SNR for known fin whale calls (n = 1484).

Simulation TL data were based on TL data from Wake Island but were modified due to extreme TL encountered in the real Wake Island data (see Section V). Wake Island TL data were extracted at a depth of 15 m to reflect realistic fin whale calling behavior. TL ranged between 71.70 dB and 286.46 dB. For the simulation studies, the minimum TL value (71.70 dB) was subtracted from all TL values resulting in simulated TL values that ranged between 0 and 214.76 dB.

Three call spatial distributions were tested via simulation, designed to reflect differing calling animal distributions (Figure 5): calls were distributed (1) uniformly throughout the study area, (2) limited to the north-east, and (3) limited to the south of the hydrophone. The simulation was set up as follows:

(1) Calls were simulated through the study area; call distribution were changed by drawing x- and y-coordinates from either a uniform or scaled beta distribution, depending on the desired spatial call pattern (Fig. 5).

(2) Each simulated call was assigned an SNR based on the passive sonar equation; each call was assigned a source level (SL) and noise level (NL) by drawing values from Normal distributions with mean and standard deviations as measured from the Wake Island dataset, which were then combined with the bearing- and range-specific TL value for that call, taken from the modified TL data.

(3) Each call's detection probability was evaluated from the detector characterization curve and a Bernoulli trial was used to determine whether a given simulated call was detected or not.

(4)The TL value above which no calls are detected was determined using the approach described in Section II.

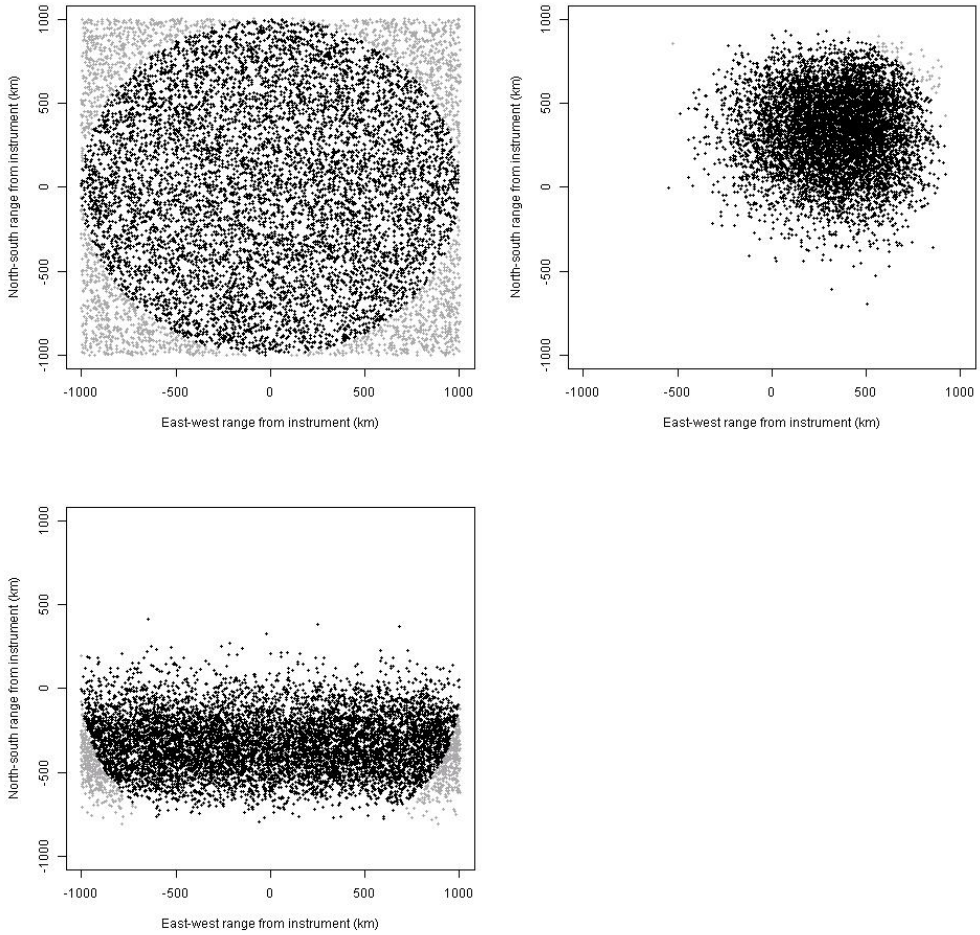

Figure 5. Examples of distributions of simulated signals (clockwise from top left: uniform, northeastern and southern distributions). The black dots denote signals within the 1000 km maximum detection radius. Gray dots show signals outside the maximum detection range.

All simulations were run 500 times in R. The maximum detection range of the recording system was specified as 1000 km in all cases. In both simulations and analyses, the maximum detection range is set as an upper limit for a given instrument but may be reduced when the monitored area is defined (Step 2, Section II.A). Call density or abundance (density could then be used to calculate abundance or *vice versa*) was also specified. Secondly, following the simulated detection process, the simulated RL, NL, and bearing values for each simulated detected call were used as inputs for analysis instead of using measurements from real recordings. In each of the three simulation scenarios, the initial abundance was altered so that the number of detected calls was similar across all scenarios. The estimated call density was compared to the known true value by calculating the median percentage bias (with associated 2.5% and 97.5% percentiles). Additionally, because the true number of simulated calls was known at increasing range steps from the array, the percentage bias as a function of range from the array could also be assessed by comparing the true number of simulated calls and the predicted number of calls within each range step. The maximum range at which the percentage bias of call density was minimised was calculated for every iteration (in some cases, the same minimal bias was calculated at multiple ranges, so the largest range was selected). The distribution of these ranges could then be assessed after all iterations were run to see whether there was an optimal prediction range, beyond which percentage bias was likely to became larger, decreasing the robustness of the final predicted density. This feature of the simulation algorithm may be useful for analysts to decide whether to restrict the area of inference following an analysis to potentially reduce bias in the reported density estimate. However, it is important to note that the simulation relies on an assumed distribution of animal calls, which is likely to be different from the true, and unknown, animal distribution, so a reduction in bias in analysis results is not guaranteed.

B. Simulation results

The simulations performed well – results from all scenarios had median percentage biases less than 2% (Table 1). Percentage bias did not exceed 5% in any of the simulations. In some scenarios, assessing the bias as a function of range showed that bias in call density estimates could be substantially reduced when call density was inferred over a reduced range. Bias was negligible for the uniform and southern distributions at median ranges of 678 km, and 360 km, respectively, suggesting that these ranges were the optimal prediction ranges for these scenarios. The NE distribution results were not improved by reducing the range of prediction. Spatial model fit across scenarios varied, with uniform distribution models displaying the poorest fit and the NE distribution producing spatial models with the best fit (median marginal R squared values: 0.51, 0.79 and 0.92; median concordance correlation values: 0.68, 0.88 and 0.96, for uniform, southern and NE distributions, respectively.). However, all spatial models produced density maps that replicated the initial distributions (Fig. 6).

Table 1: Simulation results from three scenarios with different call distributions. Simulations were run 500 times and all results report the median value, and the 2.5 and 97.5 percentiles in parentheses.

Scenario→	Uniform distribution	Southern distribution	NE distribution
Number of detections	7243 (7147, 7354)	7597 (7484, 7714)	7408 (7389, 7427)
Percentage bias	-1.52 (-3.13, 1.12)	-1.88 (-3.96, 0.97)	0.01 (-0.45, 0.86)
Minimised % bias	-1.93e-4 (-0.98, 0.32)	-0.02 (-0.67, 0.70)	-0.01 (-0.38, 0.32)
Range at which bias minimised (km)	678 (50, 993)	360 (235, 1000)	1000 (45, 1000)

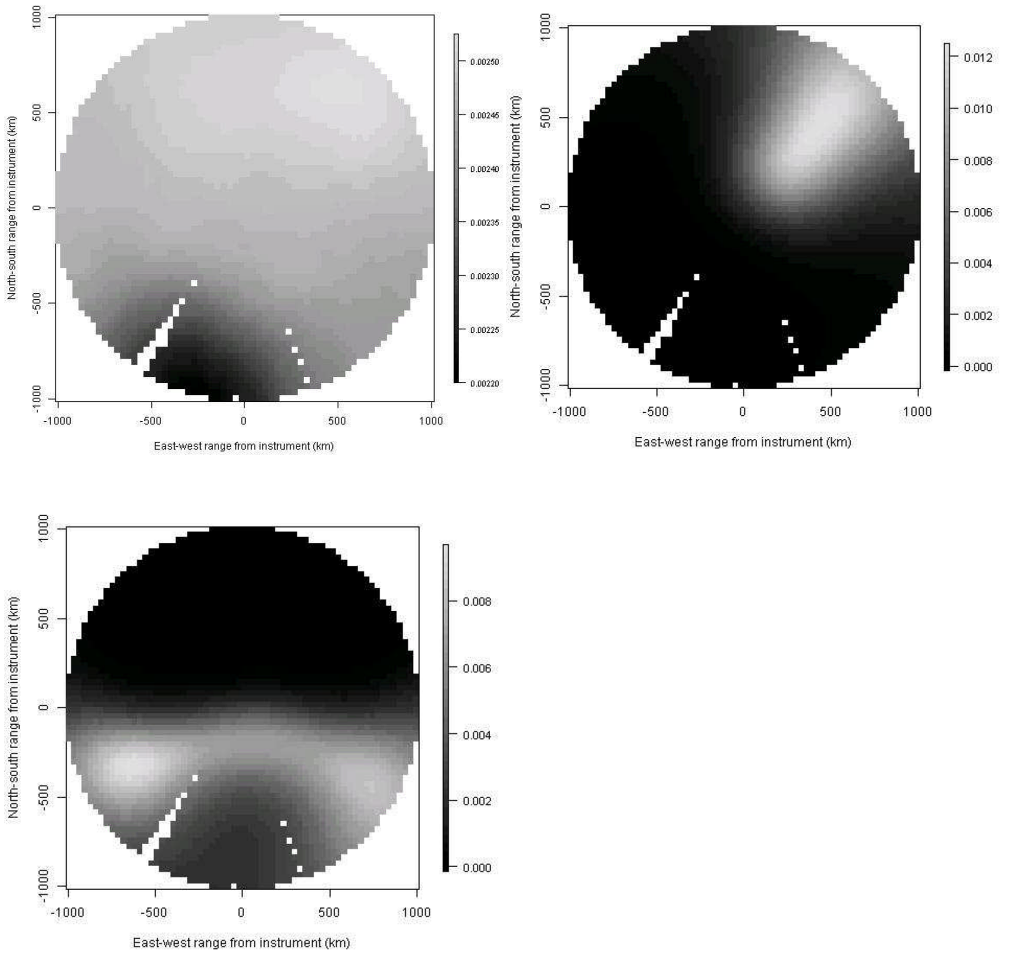

Figure 6. Distribution maps of signal density (signals/km^2) predicted by a Generalized Estimating Equation . Initial simulated distributions were, clockwise from top left, uniform, northeastern and southern distributions. The depicted maps are the median estimated surface from 500 simulations.

V. PILOT STUDY

A. Pilot study overview and input data

The pilot study analysis estimated fin whale density based on the detected calls (and associated SNR and bearing measurements) from three months of data. A simulation was also run to investigate the level of potential bias in the analysis results, and whether inferring density over a smaller area may reduce any bias (as discussed in Sections II.B and IV.A). Calls were uniformly distributed through the simulated study area and the steps of the simulation set-up were the same as those described in Section IV.A, except for the TL data used.

A key difference between the simulations described in Section IV.A and the pilot study analysis and simulation was that unmodified TL data were used in the pilot study, reflecting the true environmental conditions at Wake Island (Fig 7).

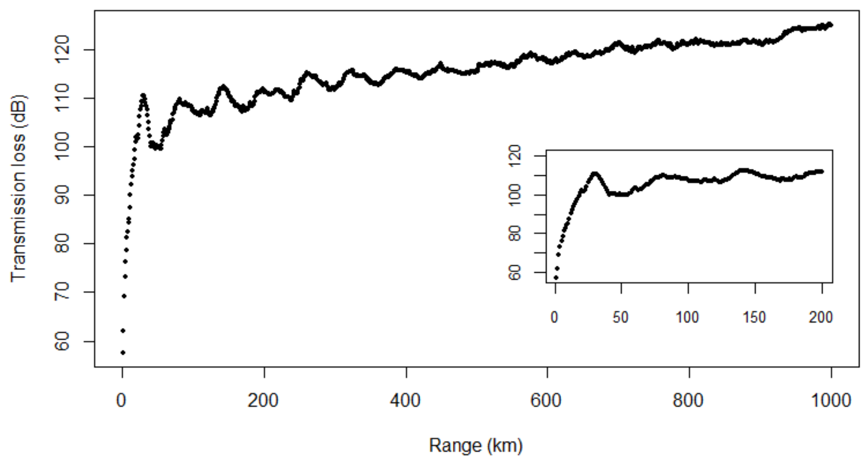

Figure 7. Transmission loss of a 20 Hz signal propagating from Wake Island N1 at a depth of 15 m, averaged across 360°. The main plot shows mean TL values up to the maximum range

without any unmeasurable infinite TL estimates (1231 km). The inset plot shows the same data plotted up to 200 km; this inset shows the decrease in TL at ~ 50 km.

Inputs for the analysis were the following: number of detections, *n*, was 6552. The automatic detector detected 6658 signals but the SNR of 106 signals fell below the lower SNR limit of detected calls in the detector characterisation analysis (2.24 dB) and so were removed to prevent model extrapolation when estimating detection probability using the detector characterization curve. Of the remaining detections, 3086 (47%) had measurable bearings, which ranged between 1.69 and 359.40 degrees (Fig. 8). While detections occurred at all bearings around N1, the quadrant with the greatest number of detections occurred between 90 and 180 degrees.

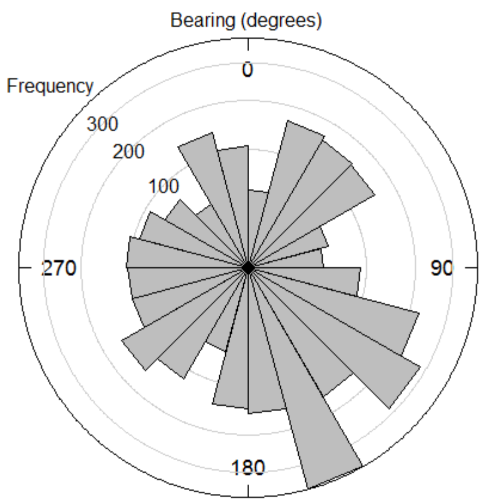

Figure 8. Histogram of measured bearings (in degrees) from the three-month pilot study dataset (n = 3086). In this plot, 0 degrees indicates north.

The highest NL associated with a detection was 123.89 dB re 1 $\mu Pa^2/Hz$. Of the 91 days of continuous monitoring, 27 mins had an average NL of 124 dB re 1 $\mu Pa^2/Hz$ or above. Therefore, it is possible that high noise levels in these minutes could have prevented any detections taking place, so these periods were considered "off effort" and were excluded from the time spent monitoring, *T*.

The false positive proportion, $\hat{c}$, was 0.097 (standard error: 0.05). The maximum detection radius, where detection probability was assumed to be negligible, was set to 1000 km and a total of 2183.55 hours were analysed (excluding 27 mins of recordings where ambient noise was assumed to be too high to successfully run the automatic detector).

Call production rate was determined from Stimpert *et al*. (2015). Deployment duration and number of calls recorded were reported for 18 digital acoustic recording tag (DTAGs, Johnson & Tyack, 2003) records. Ten animals were tagged with a version of the DTAG (v3) that enables calls from the tagged animal to be identified from other calls made by non-tagged conspecifics. It is crucial when estimating call production rate that only calls from the focal animal are included in the analysis, so the other 8 animals tagged with v2 DTAGs were omitted from the analysis. The v3 DTAGs were deployed between 1.60 and 6.30 hours. Six tags did not record any calls, while the number of calls produced by the remaining four tagged whales ranged between 23 and 942. The weighted mean call production rate was 45.08 calls.hr^{-1} (standard error: 22.31).

B. Pilot study results

The pilot study simulation was run 500 times assuming a uniform distribution with an initial starting abundance of 5e+6 calls, and a maximum detection range of 1000 km. The median number of observations was 238, and the resulting median percentage bias in estimated density was -56.37%, but decreased to -10.76% if density was only estimated up to a range step of 10 km. The median estimated density surface showed that the area within which the calls were predicted to originate was very restricted, compared to the detection area initially considered (~12 million km^2) and is fragmented (Fig. 9a).

The pilot study analysis estimated initial average call density over the three month period from Dec 2007 – Feb 2008 to be 0.014 $calls.hr^{-1}.km^2$ (CV: 0.15). Applying the call production rate from the Southern Californian Bight resulted in an average fin whale density of 0.32 animals.1000 km^2. The CV for the density estimate was 0.52. The overall monitored area for both the pilot study simulation and analysis (once spatial acoustic masking was taken into consideration) was 973 km^2 (Fig. 9b). Based on the results of the simulation, the pilot analysis results were re-analyzed with a range step restriction of 10 km. There was no way to determine which of the detections without bearings would have been detected within 10 km, so it was assumed that the relative abundances of the two detection types (which could be calculated from the initial analysis across the whole survey region) was not altered by making inference over a smaller area. Therefore, an additional multiplier, b, was used to scale the estimated density based on detections with bearings (b = 1.22). The resulting call density estimate was 0.02 $calls.hr^{-1}.km^2$ (CV: 0.15), which resulted in a density of 0.54 animals.1000 km^2 (95% confidence interval: 0.21 - 1.40 animals/1000 km^2). The CV associated with the density estimate was 0.52.

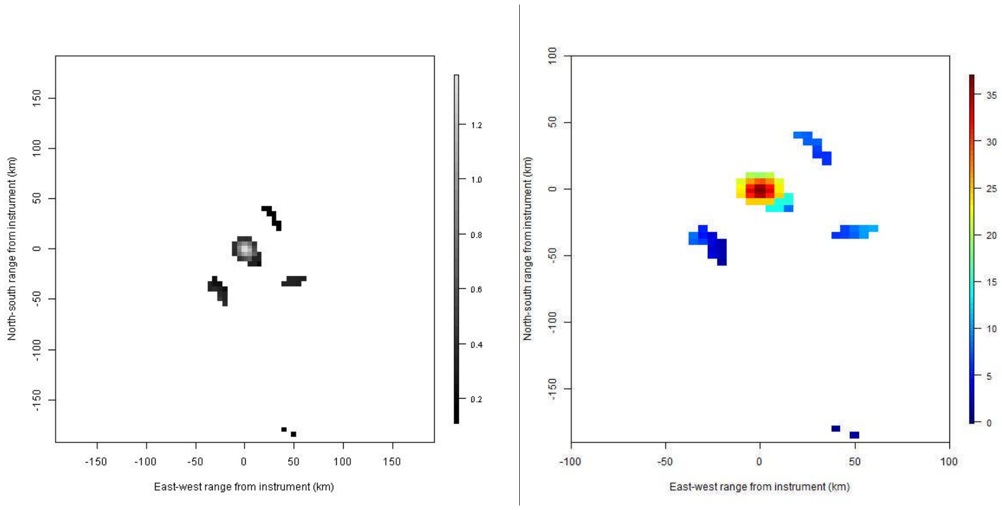

Figure 9. Distribution maps of signal density (signals/km^2) predicted by a Generalized Estimating Equation based on the pilot study data inputs. Fig 9a (left) the median estimated surface from 500 simulations. Fig 9b (right) the map from the analysis of fin whale calls from the three-month pilot study (signals/km^2).

VI. DISCUSSION

There are already several existing methods that can be used to estimate animal density from acoustic data. However, the large variety of acoustic hardware and instrument configurations continue to present new surveying challenges and require current density approaches to be adapted. The CTBTO dataset presents such a case; there are 6 hydroacoustic stations similar to Wake Island situated in the Pacific, Atlantic, and Indian Oceans (CTBTO, 2016), which have provided a wealth of baleen whale recordings (e.g., Stafford *et al*., 2011, Samaran *et al*., 2013; Le Bras *et al*., 2016). Each site is configured in a similar way to Wake Island, with two triads of cabled hydrophones, one located to the north and one to the south of a land-based station that collects data round the clock. However, to date, it has not been possible to utilize CTBTO data fully for cetacean density estimation. Distance sampling is not a suitable method for CTBTO data: only two monitoring points would be formed by the two triads at each site, which is too few for distance sampling (due to the animal distribution assumption discussed in Section I). In addition, the array spacing within triads only enables call localization using traditional time difference of arrival methods at close ranges, meaning that detections from greater distances would have to be omitted from an analysis. Given that the large detection ranges due to the deep sound channel moorings are an advantageous feature of CTBTO hydrophones, distance sampling would not be an optimal analysis method in cases where the majority of signals were originating from distant locations and could not be localized (recently, however, Le Bras *et al.* (2015) presented an alternative location methodology using bearing and amplitude information in a Bayesian framework to estimate calling animals' locations from CTBTO data, which may extend the localization capabilities of these arrays). The array design at each site is also not configured well for an SECR analysis. Although six hydrophones are available per site, acoustic masking is expected between the northern and southern arrays, creating an acoustic barrier (Pulli & Upton, 2001). Furthermore, the close spacing of the hydrophones in each triad would likely lead to many detections being recorded by all three instruments. SECR depends on a variety of capture histories to infer the location of calling animals; in this case, the array design may provide limited information (i.e., scenarios where all instruments are ensonified on each occasion yields little spatial information about the calling animals).

Therefore, data from the CTBTO arrays required a density estimation approach that used auxiliary data. Although Monte Carlo simulations have been used to estimate call density of blue whales in the Indian Ocean using CTBTO data (Harris, 2012), the method presented here used the additional distributional information available in the measured bearings. The more empirical data about animals' locations that can be collected during the acoustic survey, the fewer methodological assumptions are required during the analysis. Although this method was developed specifically for CTBTO data, there are other instrument systems that record similar information. For example, DIFAR (directional frequency analysis and recording) sonobuoys record bearings and have been used to detect blue whales at distances over 100 nautical miles (e.g., Miller *et al*., 2015).

The simulations demonstrated that the method performed well under the three different simulated animal distributions (though with less extreme propagation conditions as modelled at Wake Island). In two of the three cases, bias was further reduced when density was predicted over a smaller area than the detection radius originally set for the simulation. For example, in the median surface plot of the uniform distribution scenario, an area on the periphery of the detection radius has some negative bias (as shown by the darker region to the south of the array in Fig. 6a) and the simulation results recommended that density only be predicted out to 678 km. The same issue was also encountered during the pilot study. Running a simulation specifically for the pilot study suggested that the initial estimates were likely to be negatively biased and inference was restricted to a smaller area. In this case, restricting the area nearly doubled the point estimate (from 0.32 to 0.54 animals.1000 km^2). In summary, the simulation code provides a tool for users to explore optimal detection ranges for their given target species, survey location, and automated detection software. A natural extension to the work would be to incorporate more complex animal distributions into the simulation algorithm.

The pilot study analysis demonstrated how most of the required auxiliary data for this approach can be generated using subsampled data from the main three-month survey. It is crucial that all parameters in the density estimator have been estimated accurately for the time and place of the main survey, otherwise resulting density estimates may be biased. Source levels, noise levels, transmission loss, the proportion of false positives, and the detector characterization curve were all estimated specifically for the Wake Island dataset. The source level analysis suggested that, while the choice of transmission loss model made little difference to the source level distribution parameters used in the simulations and analyses, the negative relationship between estimated source level and range of the call from the hydrophone when using the Peregrine transmission loss model warrants further investigation. Parabolic equation models can have limitations at high incidence angles (i.e., small ranges in this case) (Jensen *et al*., 2000), which could result in the discrepancies seen between the two sets of source level results. Further, a fixed source depth of 15 m was assumed for all TL data used in both the simulations and analyses; an extension to this work would be to see whether changes in source depth (or using a distribution of source depths) significantly affects the Peregrine TL (and therefore SL) results. The one parameter that could not be estimated from the collected data was call production rate. In the absence of any other available data, call production rates from the Southern Californian Bight collected during summer months were applied to the estimated call densities. It is highly probable that the call production rates of fin whales around Wake Island and southern California are different; cue production rates do show spatiotemporal variation (e.g., Warren *et al*., 2017). Therefore, the fin whale densities estimated around Wake Island should be considered a "ballpark" estimate at best.

The pilot study also demonstrated the flexibility of density estimation methods. In this case, bearings could not be measured for all detections, but all detections (except those with SNR values below the lower SNR limit of the detector characterization curve) could still be incorporated into the analysis. It should be noted, however, that the estimated distribution map was based on those detections with measurable bearings only. In order to interpret the resulting distribution map as the predicted spatial distribution of calling fin whales, an assumption must be made that the measured bearings represent the spatial distribution of all detections. In any method that makes assumptions, it is important to assess whether the assumptions are reasonable, or whether they may have been violated. Therefore, consideration should be given as to whether there are any oceanographic or bathymetric features of the study area that may result in certain bearings being difficult, or impossible, to measure (other than high TL values, which are accounted for by identifying areas of acoustic masking at the start of the analysis). In these cases, the resulting map would not depict the distribution of all calling animals.

The most striking result of the pilot analysis was the fact that the monitored area at Wake Island for fin whale calls was much smaller than originally anticipated. Sirovic *et al*., (2007) estimated detection ranges of fin whale calls in the Antarctic Ocean up to 56 km, though their instruments were not moored in the deep sound channel. Previous work investigating detection range of blue whale calls at CTBTO sites in the Indian Ocean (Samaran *et al*., 2010, Harris, 2012) predicted that blue whale calls could be detected hundreds of kilometres away, facilitated by the deep sound channel. However, the pilot study results are supported by previous work that predicted detectability of low frequency signals at Wake Island to be lower than at Diego Garcia (Miksis-Olds *et al*., 2015). The results of all simulations and pilot analysis also demonstrated that the monitored area may be an irregular shape, or even fragmented, as seen in the pilot study. The fragmentation of the monitored area in the pilot study is most likely caused by fluctuations in TL with range; the TL decreases at approximately 50 km (Fig. 8, inset), which corresponds to the fragmented regions. Monitored areas with unusual shapes should not lead to biased density estimates, as long as the results are not extrapolated to areas outside the defined monitored area.

The pilot study has demonstrated the importance of quantifying the size and shape of the monitored area (by estimating detection probabilities of the target species) during acoustic surveys. The same site may show temporal variation in detection probability as oceanographic conditions change through the year. Geographic variability in detection probability between sites, caused by local bathymetric and ocean conditions should also be considered, even if the acoustic system is the same. Detection probability may also alter if the behavior of the target species changes e.g., if animals increase call source levels in certain behavioral contexts. Investigating such spatial and temporal variation in detection probabilities at Wake Island and another CTBTO site, Diego Garcia in the Indian Ocean, will comprise the next stage of this research. Another natural extension to this work would be to analyse the southern site at Wake Island to investigate whether the same monitoring conditions are present at a site ~ 200 km from the focal instrument in this initial study.

ACKNOWLEDGMENTS

DH and LT were funded by the Office of Naval Research (Award: N00014-14-1-0394). JMO and JV were funded under Award: N00014-14-1-0397 also from the Office of Naval Research. Kevin Heaney (OASIS, Inc.) generously contributed the TL modelling results for which we extend sincere thanks. Thanks also to Lindesay Scott-Hayward for her help with implementing CReSS and SALSA methods. The CTBTO data was accessed from the Air Force Tactical Applications Center/US National Data Center. Thanks are extended to James Neely (AFTAC), Richard Baumstark (AFTAC), Mark Prior (formerly CTBTO), and Andrew Forbes (CTBTO) for their assistance in data transfer and transfer of knowledge of CTBTO data. Thanks also to the CTBTO for making available the virtual Data Exploitation Center (https://www.ctbto.org/specials/vdec/). Finally, please note that the views expressed in this study are those of the authors and not necessarily represent the views of the CTBTO Preparatory Commission.

REFERENCES

Amante, C. and B.W. Eakins, (2009) ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. doi:10.7289/V5C8276M [02/01/2018].

Borchers, D.L. 2012. A non-technical overview of spatially explicit capture recapture models. *Journal of Ornithology* 152 (Suppl 2), S435-444.

Borchers, D. L., Buckland, S. T. & Zucchini, W. (2002) *Estimating Animal Abundance*. Springer, New York.

Borchers DL, Burnham KP. General formulation for distance sampling. In: Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L, editors. *Advanced Distance Sampling*. Oxford: Oxford University Press; 2004. pp. 6–30.

Borchers, D. L., B. C. Stevenson, D. Kidney, L. Thomas & T. A. Marques (2015) A Unifying Model for Capture–Recapture and Distance Sampling Surveys of Wildlife Populations, *Journal of the American Statistical Association*, 110:509, 195-204, DOI: 10.1080/01621459.2014.893884

Buckland, S. T. (2006) Point transect surveys for songbirds: robust methodologies. *The Auk* **123**: 345-345

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. & 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. DOI: 10.1007/978-3-319-19219-2

Cochran, W.G. (1977) Sampling Techniques. Wiley, New York, NY, USA.

Frasier, K. E., Wiggins, S. M., Harris, D., Marques, T. A., Thomas L, & Hildebrand J. A. (2016) Delphinid echolocation click detection probability on near-seafloor sensors *J. Acoust. Soc. Am.*, 140, 1918-1930.

Gatz, D.F. & Smith, L. (1995) The standard error of a weighted mean concentration – I. Bootstrapping vs other methods. *Atmospheric Environment*, 29, 1185–1193.

Hardin, J. W. and Hilbe, J. M. (2012). *Generalized Estimating Equations: 2nd Ed.* Chapman & Hall, CRC Press, Boca Raton, FL.

Harris, D. (2012) Estimating whale abundance using sparse hydrophone arrays. Ph.D. University of St Andrews.

Harris, D., L. Matias, L. Thomas, J. Harwood & W. Geissler. (2013). Applying distance sampling to fin whale calls recorded by single seismic instruments in the northeast Atlantic. *The Journal of the Acoustical Society of America* 134: 3522-3535.

Heaney, K.D., Campbell, R.L. (2016). Three-dimensional parabolic equation modeling of mesoscale eddy deflection. *Journal of the Acoustical Society of America* 139: 918-926.

Helble, T. A., D'Spain, G. L., Hildebrand, J. A., Campbell, G. S., Campbell, R. L. & Heaney, K. D. (2013) Site specific probability of passive acoustic detection of humpback whale calls from single fixed hydrophones. *The Journal of the Acoustical Society of America* 134: 2556-2570.

Horvitz, D.G., and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe,Journal of the American Statistical Association 47, 663-685.

Jensen, F. B., Kuperman, W. A, Porter, M. B. & Schmidt, H. (2000) *Computational Ocean Acoustics*, Springer-Verlag, New York.

Kidney, D., B.J. Rawson, D.L. Borchers, L. Thomas, T.A. Marques and B. Stevenson. (2016). An efficient acoustic density estimation method with human detectors, applied to gibbons in Cambodia. *PLOS One* 11(5): e0155066.

Kyhn L.A., J. Tougaard, L. Thomas, L.R. Duve, J. Steinback, M. Amundin, G. Desportes and J. Teilmann. (2012). From echolocation clicks to animal density - acoustic sampling of harbour porpoises with static dataloggers. *Journal of the Acoustical Society of America* 131: 550-560.

Küsel, E.T., D.K. Mellinger, L. Thomas, T.A. Marques, D.J. Moretti, and J. Ward. (2011). Cetacean population density from single fixed sensors using passive acoustics. *Journal of the Acoustical Society of America* 129: 3610-3622.

Le Bras, Ronan J., Kuzma, H., Sucic V., & Bokelmann, G (2016) Observations and Bayesian location methodology of transient acoustic signals (likely blue whales) in the Indian Ocean, using a hydrophone triplet. *Journal of the Acoustical Society of America*. 139, 2656; http://dx.doi.org/10.1121/1.4948758

Marques, T.A., L. Thomas, J. Ward, N. DiMarzio and P. L. Tyack. (2009). Estimating cetacean population density using fixed passive acoustic sensors: an example with beaked whales. Journal of the Acoustical Society of America 125: 1982-1994.

Marques, T.A., L. Thomas, L. Munger, S. Wiggins and J.A. Hildebrand. (2011). Estimating North Pacific right whale (*Eubalaena japonica*) density using passive acoustic cue counting. *Endangered Species Research* 13: 163-172.

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

Mathworks, 2016, The Mathworks Inc., Cambridge, U.K., accessed 28 September, http://www.mathworks.co.uk

Mellinger, D. K. (2002) Ishmael 1.0 User's Guide, NOAA Technical Memorandum OAR PMEL-120, NOAA/PMEL, Seattle, WA.

Mikisis-Olds, J. M., J. A Vernon and K. D. Heaney (2015) The Impact of Ocean Sound Dynamics on Estimates of Signal Detection Range. *Aquatic Mammals* 41(4): 444-454.

Miller BS, J Barlow, S Calderan, K Collins *et al*. (2015). Validating the reliability of passive acoustic localisation: a novel method for encountering rare and remote Antarctic blue whales. *Endang Species Res* 26:257-269

Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organisation.(CTBTO), 2016, Vienna International Centre, Austria, accessed 29 September, < http://www.ctbto.org>.

Pulli, J. J. & Upton, Z. M. (2001) Hydroacoustic blockage at Diego Garcia: models and observations. 23rd Seismic Research Review, Wyoming, U.S.A, October 2-5 2001.

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

Reilly, S.B., Bannister, J.L., Best, P.B., Brown, M., Brownell Jr., R.L., Butterworth, D.S., Clapham, P.J., Cooke, J., Donovan, G.P., Urbán, J. & Zerbini, A.N. (2013) *Balaenoptera physalus*. The IUCN Red List of Threatened Species 2013: e.T2478A44210520. http://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T2478A44210520.en. Downloaded on 28 September 2016.

Samaran, F., Adam, O. & Guinet, C. (2010) Detection range modeling of blue whale calls in Southwestern Indian Ocean. *Applied Acoustics* **71**: 1099-1106.

Samaran F, Stafford KM, Branch TA, Gedamke J, Royer J-Y, Dziak RP, et al. (2013) Seasonal and Geographic Variation of Southern Blue Whale Subspecies in the Indian Ocean. *PLoS ONE* 8(8): e71561. doi:10.1371/journal.pone.0071561

Scott Hayward, LAS, MacKenzie, ML, Donovan, CR, Walker, C & Ashe, E (2014) Complex Region Spatial Smoother (CReSS), *Journal of Computational and Graphical Statistics*, vol 23, no. 2, pp. 340-360. DOI: 10.1080/10618600.2012.762920

Širović, A., Hildebrand, J. A. & Wiggins, S. M. (2007) Blue and fin whale call source levels and propagation range in the Southern Ocean. *Journal of the Acoustical Society of America* **122(2)**: 1208–1215.

Seber, G. A. F. (1982) *The Estimation of Animal Abundance*, 2nd Ed. Griffin, London.

Širović, A, A Rice, E Chou, JA Hildebrand, SM Wiggins, MA Roch. (2015) Seven years of blue and fin whale call abundance in southern California. *Endangered Species Research*. 28:61-76. 10.3354/esr00676

Stevenson, B.C., Borchers, D.L., Altwegg, R., Swift, R.J., Gillespie, D.M., and Measey, G.J. (2015) A general framework for animal density estimation from acoustic detections across a fixed microphone array. *Methods in Ecology and Evolution*, **6** 38-48.

Stimpert, A. K., DeRuiter, S. L., Falcone, E. A., Joseph, J., Douglas, A. B., Moretti, D. J., Friedlaender, A.S., Calambokidis, J., Gailey, G, Tyack, P.L. and Goldbogen, J. A. (2015). Sound production and associated behavior of tagged fin whales (Balaenoptera physalus) in the Southern California Bight. *Animal Biotelemetry*, *3*(23), **DOI:** 10.1186/s40317-015-0058-3.

Tiemann, C. O., Thode, A. M., Straley, J., O'Connell, V. & Folkert, K. (2006) Three dimensional localization of sperm whales using a single hydrophone. *Journal of the Acoustical Society of America* **120(4):** 2355-2365.

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

Walker C, Mackenzie M, Donovan C, O-Sullivan M (2011) SALSA: a spatially adaptive local smoothing algorithm. *J Stat Comput Simul* 81:179–191. doi:10.1080/00949650903229041

Warren, VE, Marques, TA, Harris, D, Thomas, L, Tyack, PL, Aguilar de Soto, N, Hickmott, LS & Johnson, MP (2017), Spatio-temporal variation in click production rates of beaked whales: implications for passive acoustic density estimation. *Journal of the Acoustical Society of America*, vol 141, no. 3, pp. 1962-1974. DOI: 10.1121/1.4978439

Wood, S. N. (2006) *Generalised Additive Models: An Introduction with R*. Chapman & Hall, Boca Ranton, FL

Table 1: Simulation results from three scenarios with different call distributions. Simulations were run 500 times and all results report the median value, and the 2.5 and 97.5 percentiles in parentheses.

Scenario→	Uniform distribution	Southern distribution	NE distribution
Number of detections	7243 (7147, 7354)	7597 (7484, 7714)	7408 (7389, 7427)
Percentage bias	-1.52 (-3.13, 1.12)	-1.88 (-3.96, 0.97)	0.01 (-0.45, 0.86)
Minimised % bias	-1.93e-4 (-0.98, 0.32)	-0.02 (-0.67, 0.70)	-0.01 (-0.38, 0.32)
Range at which bias minimised (km)	678 (50, 993)	360 (235, 1000)	1000 (45, 1000)

FIGURE LEGENDS

Figure 1. Map showing the location of Wake Island (coordinates: 19.30, 166.63) and the northern hydrophone array. Water depth contours (1000 m, 2000m and 4000 m) are also depicted (color online).

Figure 2. Transmission loss of a 20 Hz signal propagating to Wake Island N1 at a depth of 15 m. The mod el was run for every bearing between 0 and 359 degrees at 1 km range steps. In this plot, 0 degrees indicates north (color online).

Figure 3. Source levels estimated from 79 calls using transmission loss derived from (left) the Peregrine model and (right) assuming spherical spreading. Both plots show a fitted linear regression model (black line), with associated 95% confidence intervals shaded in gray.

Figure 4. Detector characterization curve (with 95% confidence interval) predicting detection probability as a function of SNR for known fin whale calls (n = 1484).

Figure 5. Examples of distributions of simulated signals (clockwise from top left: uniform, northeastern and southern distributions). The black dots denote signals within the 1000 km maximum detection radius. Gray dots show signals outside the maximum detection range.

Figure 6. Distribution maps of signal density (signals/km^2) predicted by a Generalized Estimating Equation . Initial simulated distributions were, clockwise from top left, uniform, northeastern and southern distributions. The depicted maps are the median estimated surface from 500 simulations (color online).

Figure 7. Transmission loss of a 20 Hz signal propagating to Wake Island N1 at a depth of 15 m, averaged across 360°. The main plot shows mean TL values up to the maximum range without any unmeasurable infinite TL estimates (1231 km). The inset plot shows the same data plotted up to 200 km; this inset shows the decrease in TL at ~ 50 km.

Figure 8. Histogram of measured bearings (in degrees) from the three-month pilot study dataset (n = 3066). In this plot, 0 degrees indicates north.

Figure 9. Distribution maps of signal density (signals/km^2) predicted by a Generalized Estimating Equation based on the pilot study data inputs. Fig 9a (left) the median estimated surface from 500 simulations. Fig 9b (right) the map from the analysis of fin whale calls from the three-month pilot study (signals/km^2) (color online).

TABLE TITLES

Table 1: Simulation results from three scenarios with different call distributions. Simulations were run 500 times and all results report the median value, and the 2.5 and 97.5 percentiles in parentheses.