Inter-Research Science Publisher

Endangered Species Research
Volume 56:101–115

Published February 20, 2025
https://doi.org/10.3354/esr01384

Acoustic abundance estimation for Critically Endangered North Atlantic right whales in Cape Cod Bay, Massachusetts, USA

Marissa L. Garcia[1,2,3,*,#], Irina Tolkova[1,4,#], Shyam Madhusudhana[1,5], Ashakur Rahaman[1], C. Scott Baker[6], Charles A. Mayo[7], Christine A. Hudak[7], Holger Klinck[1,6]

[1]K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, NY 14850, USA
[2]Department of Natural Resources and the Environment, Cornell University, Ithaca, NY 14853, USA
[3]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
[4]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02134, USA
[5]Centre for Marine Sciences and Technology, Curtin University, Perth, Western Australia 6102, Australia
[6]Marine Mammal Institute, Hatfield Marine Science Center, Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Newport, OR 97365, USA
[7]Center for Coastal Studies, Provincetown, MA 02657, USA

ABSTRACT: With an estimated 372 individuals remaining, Critically Endangered North Atlantic right whales *Eubalaena glacialis* (henceforth, NARWs) embody New England's foremost marine conservation challenge. Every year, a large portion of the NARW population visits Cape Cod Bay (CCB), Massachusetts, USA, a critical foraging area. Traditionally, aerial surveys have documented the abundance and distribution of NARWs in CCB. In this work, we demonstrate abundance estimation through passive acoustic monitoring, utilizing recordings from an array of 5 marine autonomous recording units (MARUs) deployed from February to June 2019. We first trained, validated, and applied a deep-learning-based detector for NARW upcall vocalizations, achieving a precision of 0.857 and recall of 0.896. Next, we matched duplicate detections across the MARU array through time-difference-of-arrival association. Lastly, after calibrating acoustic cue counts to concurrent aerial surveys conducted by the Center for Coastal Studies, we estimated daily NARW abundance in CCB across the foraging season. We demonstrated diel and seasonal patterns in acoustic phenology consistent with prior studies. Upcall rates were higher at night, particularly just after sunset, than during daylight hours. We observed low presence of NARWs in late February, with presence rising in early March, peaking in late March and early April, and ultimately decreasing through mid-May. While many sources of uncertainty limit the precision of abundance estimates, PAM offers a cost-effective, generalizable, and largely automated approach for detecting NARW abundance trends applicable to both conservation management and ecological studies.

KEY WORDS: Abundance estimation · Automated detection · Cape Cod Bay, MA · Conservation bioacoustics · Convolutional neural network · North Atlantic right whales · Passive acoustic monitoring

1. INTRODUCTION

Globally, North Atlantic right whales *Eubalaena glacialis* (henceforth, NARWs) rank among the most endangered baleen whale species. Although NARWs received international protection from hunting over 80 yr ago, the species' recovery has been limited (Pace et al. 2017); the population is now estimated at 372 individuals (New England Aquarium 2024). Ship strikes and entanglement in fishing gear are the

[#]These authors contributed equally to this paper
*Corresponding author: mg2377@cornell.edu

Publisher: Inter-Research · www.int-res.com

primary causes of NARW mortality (Garrison et al. 2022, Knowlton et al. 2022). Furthermore, anthropogenic noise pollution and climate-driven changes in prey distribution further exacerbate the ability of NARWs to recover in a rapidly changing environment (Rolland et al. 2012, Meyer-Gutbrod et al. 2015, Corkeron et al. 2018).

Passive acoustic monitoring (PAM) is a well-established tool for detecting the presence and distribution of NARWs (e.g. Mellinger et al. 2007, Davis et al. 2017, Durette-Morin et al. 2022), a federally protected and Critically Endangered species (Cooke 2022). Upcalls — frequency-swept contact calls (50–350 Hz) typically 0.3 to 1.5 s in duration (Parks & Tyack 2005, Clark et al. 2007) — are regularly vocalized by NARWs of all ages and sexes and thus have been applied successfully as an acoustic proxy for presence (e.g. Morano et al. 2012, Durette-Morin et al. 2019, Charif et al. 2020). Indeed, PAM has detected NARW upcall presence on days when aerial surveys, the traditional means of monitoring, have not (Clark et al. 2010). Furthermore, PAM offers logistical convenience, being both cost-effective and safer for field observers. In a study reviewing occupational fatalities in the wildlife profession, airplanes and helicopter flights — one mode by which visual surveys for cetaceans are conducted — accounted for 66% of wildlife biologist mortalities (Sasse 2003). Additionally, since analyses of PAM data are often conducted in post-processing stages rather than in the immediacy of a field environment, observer skill and attention span are less likely to affect species detection and identification accuracy. Pertinently, acoustic recordings are opportune for analysis with rapidly advancing deep learning techniques. Since PAM creates an archival dataset, it can be re-analyzed in later work as more advanced signal processing tools emerge.

However, PAM data can amount to several terabytes in size, which can be cumbersome to analyze via manual annotations alone. To reduce the workload of human analysts and expedite analytical speed overall, deep learning can generalize the acoustic structure learned from manual annotations across large-scale PAM datasets to create automatic detectors. In particular, convolutional neural networks (CNNs) are a class of deep learning architectures designed for image analysis. CNNs can be effectively applied to PAM data by representing audio in spectrogram form, a visual representation of a signal in the time-frequency domain that the model can then treat as an image (Kirsebom et al. 2020). Over the last few years, CNNs have successfully facilitated automatic detection and classification of marine mammal vocalizations (Shiu et al. 2020, Kirsebom et al. 2020, Ibrahim et al. 2021). While achieving reliable performance of deep learning systems in field conditions can be challenging, the highly stereotyped structure of the NARW's upcall is especially well suited for automated classification.

Moreover, while PAM has helped uncover species presence and occurrence trends, its applications toward abundance and density estimation are still being advanced for multiple marine mammal species and have historically benefitted from supplementary visual techniques used to validate estimates (e.g. Barlow & Taylor 2005, Taylor et al. 2017, Roberts et al. 2024). Acoustic abundance estimation requires measuring acoustic 'cues' (for NARW, upcalls), with an understanding of the cue rate of a species, as well as the number of cues in each instance. Studies face the ongoing obstacle of accurately converting these measured cues into counts of individual animals (Marques et al. 2009, 2011). While individual call rates, measured *in situ* or published, can facilitate this conversion, these rates still vary across individuals (e.g. age, sex, behavioral state, social state), time scales (e.g. diel, seasonal, annual), and ecosystems. In recent years, acoustic abundance estimation techniques have been further developed for species for which cue rates are better understood, such as beaked whales, which continuously produce 2–3 echolocation pulses per second during deep foraging dives (Zimmer et al. 2005, Barlow et al. 2021, 2022). By comparison, NARW upcall rates are highly variable per individual (Franklin et al. 2022). Average call rates have been estimated to be 6 upcalls per hour, but can vary greatly based on social interactions (Parks et al. 2011). While a NARW is alone, call rates range from 0 to 200 upcalls per hour, whereas while 2 or more NARWs are present, call rates range from 0 to 333 upcalls per hour (Parks et al. 2011). With this in mind, we therefore focused on population-level abundance calibration to circumvent the challenges of extrapolating individual cue rates across a population.

In this work, we addressed each of these challenges to develop a passive acoustic pipeline for estimating the abundance of the NARW. First, we trained a CNN to detect upcalls (henceforth, detector) and evaluated its performance against manually annotated upcalls. Next, upcalls detected across multiple marine autonomous recording units (MARUs) were associated based on their times-of-arrival to attain the number of unique upcalls and avoid double-counting. Finally, we fitted a quasi-Poisson model to relate the number of unique upcalls with corresponding aerial survey whale counts, and then used this model to estimate daily

abundance throughout the acoustic survey (Fig. 1). We tested the pipeline in Cape Cod Bay (CCB), Massachusetts (MA), USA, for 3 hallmark reasons: the site's ecological significance for NARWs, peninsular landmass structure, and routine visual survey coverage. Every year, from late winter through spring, NARWs migrate to foraging and nursery areas, specifically the productive CCB waters fueled by nutrient-rich Gulf of Maine waters (Watkins & Schevill 1976, Clark et al. 2010, Mayo et al. 2018, Hudak et al. 2023). Further, the bay is semi-enclosed by the Cape Cod peninsula. This restricts the bay to a total area of only 1560 km^2, providing an opportunity for effective plot sampling. Additionally, the Center for Coastal Studies (CCS) has performed routine aerial surveys in CCB for over 25 yr. Each right whale observed during a survey flight is routinely photographed and matched to a NARW identification catalog, permitting analysts to refine the final count by eliminating duplicate sightings. Acoustic cue counts are later calibrated to these sightings to ultimately estimate abundance.

In this study, by calculating a population-level call rate for a species with highly variable individual call rates, we provided evidence that an acoustic abundance estimation can monitor NARW populations with reasonable accuracy, though we also identified the need for future work to improve precision of these estimates. Ultimately, this semi-automated pipeline, which reduces the labor burden of manual analysis, may facilitate greater adoption of PAM-derived abundance estimates for the conservation and management efforts of both NARWs on the US East Coast and other cetacean species globally.

2. MATERIALS AND METHODS

2.1. Data collection

2.1.1. Acoustic survey

Acoustic sampling provides an index of relative NARW abundance in CCB. Passive acoustic data was collected via an array of 5 bottom-moored MARUs (Calupca et al. 2000, Clark et al. 2002). A MARU, or 'pop-up', is composed of a spherical buoy and an attached hydrophone with a sensitivity of −168 dB re 1 V μPa^{-1}. The MARUs recorded data continuously from 17 February to 12 June 2019 for 116 d at a sampling rate of 2000 Hz. Audio was post-processed with a band-pass filter between 10 and 800 Hz. Distances between each corner MARU spanned between 12 and 14 km, while distances between each corner MARU and the center MARU spanned between 6 and 11 km. The latitude and longitude coordinates for the units, and corresponding depths, are provided in Table 1, and the array is visualized in Fig. 2A. To time-align the recorded passive acoustic data across the 5 MARUs, distinctive acoustic signals were recorded synchronously at the start and end of the survey period. Upon recovery, the recordings from the 5 MARUs were extracted, aligned, and concatenated to form 5-channel chronological sound files in audio interchange file format (AIFF).

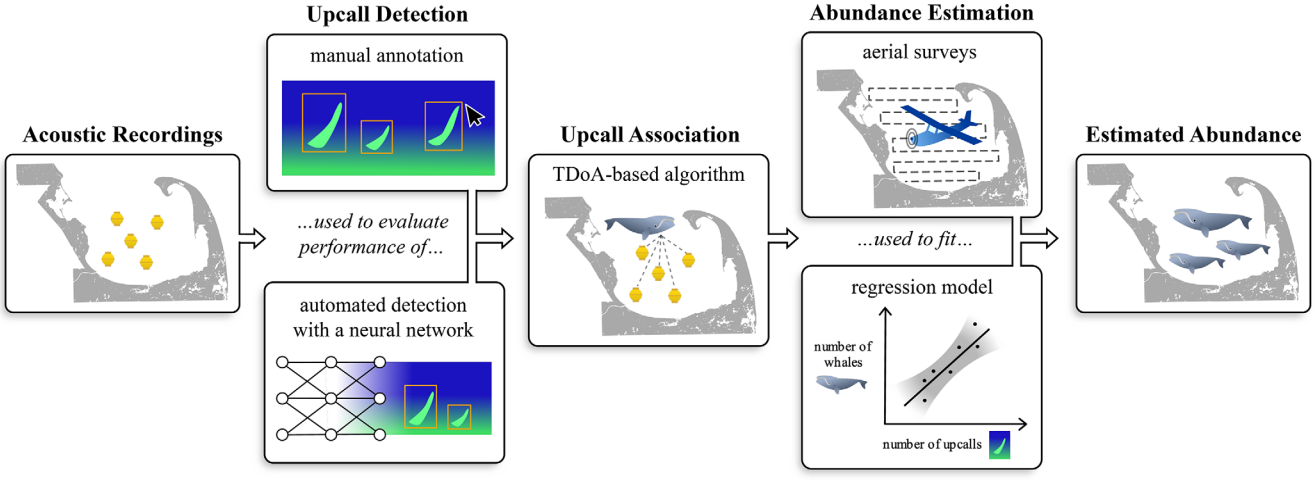

Fig. 1. The analysis pipeline consisted of a series of modular components. First, a trained analyst manually annotated upcalls in a subset of the recordings across 5 marine autonomous recording units (MARUs). The annotations were then used to evaluate the performance of an automated upcall detection and classification system. Next, the detections were associated across units through an automated time-difference-of-arrival (TDoA) procedure. Finally, the number of distinct upcalls was combined with visual observations from aerial surveys to fit a quasi-Poisson model and predict daily NARW abundance throughout the acoustic survey

Table 1. Latitude and longitude coordinates for the 5 marine autonomous recording units (MARUs) deployed for this study

MARU	Latitude	Longitude	Depth (m)
Unit 1	41.9337	−70.2286	33
Unit 2	41.9466	−70.4002	41
Unit 3	41.8413	−70.4693	20
Unit 4	41.8397	−70.3148	26
Unit 5	41.8949	−70.3592	33

2.1.2. Aerial surveys

Aerial surveys provide a conservative count of NARWs in CCB for each survey period. Throughout the acoustic survey period, the CCS conducted 21 aerial surveys over CCB using a Cessna Skymaster survey plane. During these surveys, observers were tasked with locating, counting, and photographing observed NARWs. The E–W survey track lines, with 2.8 km separation, allowed for 100% visual coverage of the sea surface to maximize the potential to detect NARWs (Fig. 2A). While the effective total swath of aerial surveys using Cessna airplanes was 4.2 km (Kenney et al. 1995, Nichols et al. 2008), the photographic identification of individuals reduced the duplication of sightings, decreasing the margin of error for total counts (Brown & Marx 1998, Nichols et al. 2008, Mayo et al. 2018).

When completed, aerial surveys covered approximately 500 km of track line, at a speed of 100 knots and an altitude of approximately 300 m. Aerial surveys could be terminated early, however, due to inclement weather or poor sea conditions (i.e. when Beaufort sea state was equal to or larger than 5). Of the 21 aerial surveys conducted throughout the acoustic survey period, 13 were completed in full. Of the 8 partial surveys, 5 were removed from further analysis due to substantially incomplete coverage (i.e. 10 or fewer track lines completed). The remaining 3 partial surveys covered at least 12 track lines, missing only peripheral track lines along either the mouth (i.e. track lines 1–2) or the southern edge (i.e. track lines 13–15) of CCB, and were deemed to have sufficient coverage for a high-fidelity aerial abundance estimate. In total, 16 aerial surveys were included in this study.

2.2. Data analysis

2.2.1. Upcall detection

We utilized the challenge dataset from the 2013 workshop on the Detection, Classification, Localization, and Density Estimation of Marine Mammals (DCLDE 2013; Gillespie 2019) as the primary source of labeled NARW upcall data for training the detector. We implemented our machine learning work-

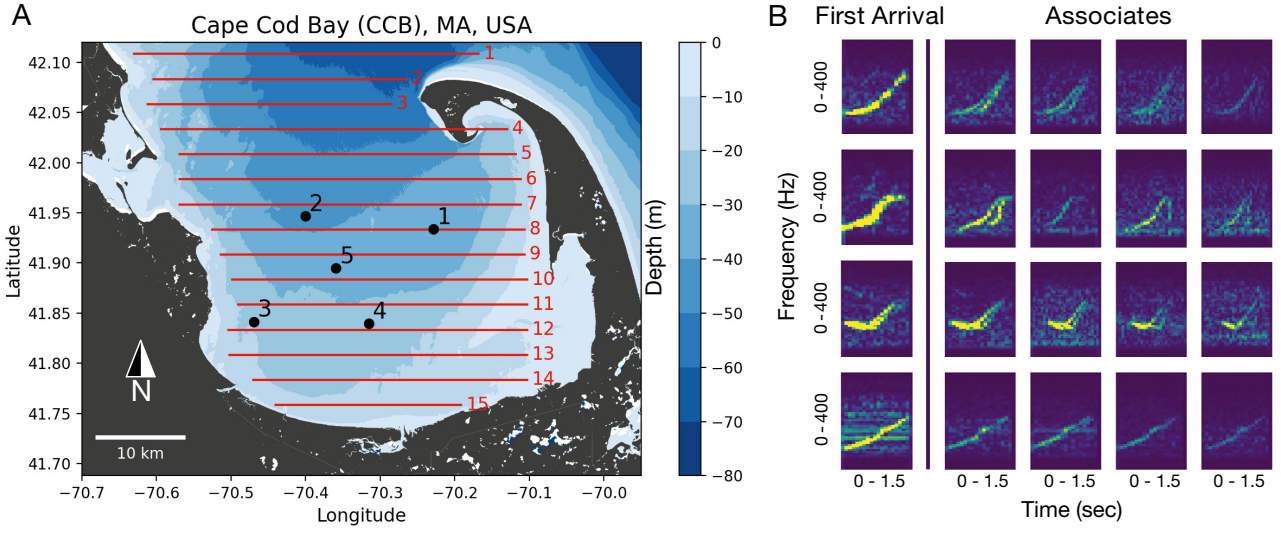

Fig. 2. (A) Cape Cod Bay (CCB) is enclosed by the peninsular structure of Massachusetts, USA, with a mouth extending into the Atlantic Ocean along the US East Coast. An array of 5 marine autonomous recording units (MARUs), indicated by the black points, were deployed in CCB from February to June 2019. During these months, aerial surveys were conducted by the Center for Coastal Studies, following the 15 track lines shown and numbered in red. (B) Each row demonstrates a set of detections associated across all 5 channels. The left column exhibits the spectrogram of a first arrival, and the 4 following columns show corresponding spectrograms of associates across the other channels. The time axes vary within 1–2 s in duration, and the frequency axis spans from 0 to 400 Hz

flow using a pre-release version of Koogu (www.shyamblast.github.io/Koogu/) and TensorFlow v1.13 (Abadi et al. 2016). We used a custom quasi-DenseNet architecture comprising 3 quasi-Dense blocks, each containing 2 convolutional layers with a growth rate of 2. The detector's frontend featured a Laplacian of Gaussian (LoG) layer (Madhusudhana et al. 2021) with scales set at 4, 8, and 16, followed by a 3×3 convolutional layer with 16 filters. A LoG operator highlights regions of rapid intensity changes in image-like inputs (Marr & Hildreth 1980). The multi-scale approach employed here was aimed at highlighting intensity ridges (corresponding to narrowband tonal signals) of different widths within input spectrograms. The classification head was preceded by a fully connected layer comprising 16 nodes. To improve robustness, we considered a dropout rate of 0.05. We trained the detector for 100 epochs, starting with an initial learning rate of 0.01 and successively diminishing it by a factor of 1/10 at epochs 30, 60, 80, and 90.

The detector's inputs comprised fixed-dimension spectrograms generated from 2.2-s-long audio segments. The DCLDE 2013 audio, which was sampled at 2000 Hz just as the data collected for this study, was first bandpass filtered using a 9th-order Butterworth filter with a passband of 43–395 Hz. This study's acoustic data (see Section 2.2) was also subject to similar bandpass filtering to ensure similarity in downstream processing between training, testing, and inferencing. The resulting waveforms were split into 2.2 s segments with a 0.4 s segment advance. These segments were then transformed into spectrograms using a 256-point discrete Fourier transform with a 75% frame overlap (resulting in time and frequency resolution of 32 ms and 7.8125 Hz, respectively). To focus on relevant frequency components, the resulting spectrograms were trimmed to encompass the bandwidth of 46–391 Hz. This preprocessing resulted in detector inputs (band-limited spectrograms) of dimensions 45×65. Binary class (positive, negative) association to each input was performed similarly to the approach described in Miller et al. (2023) and Madhusudhana et al. (2021) — segments that fully contained the temporal extents of one or more upcall annotations were considered as positive class (upcall) inputs. For further reference, the Koogu documentation details similar steps for data preparation, training, and performance assessment as employed in this work.

A 20 d subset from this study's acoustic survey, amounting to 480 h of audio, constituted our test set for assessing the model's recognition performance. A trained analyst visually/aurally reviewed spectrograms throughout the acoustic recordings concurrent with aerial surveys and manually annotated upcalls using Raven Pro 1.6.2 (K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology 2024). We processed the test set using the trained detector and analyzed its precision-recall characteristics by comparing the manual annotations to reported detections.

2.2.2. Upcall association

The geometry of the 5-unit MARU array was designed to provide acoustic survey coverage across the interior of the CCB, based on previously reported estimates of the reliable upcall detection range (9 km; Clark et al. 2010). In consequence, as this distance is comparable to the spacing between MARUs (6–14 km), it was likely that a single call would be heard on multiple recorders. For instance, a call closer to the center of the array may be heard on all 5 units, while a call closer to the edge may be detected on just one. Consequently, if all upcall detections are included in the acoustic cue count, the resulting cue counts would be spatially biased towards vocal activity in the center of the array. To produce an acoustic cue count more representative of the true number of unique upcall vocalizations produced in CCB, we applied an association procedure to match these duplicates (henceforth, associates). If the MARUs maintained high-precision time synchronization across channels, association of calls detected at 3 or more units could be performed through acoustic localization (Helble et al. 2015). However, though the units were synchronized at the start and end of the survey, relative clock drift due to temperature variation likely rendered localization infeasible (Marchetto 2015, Estabrook et al. 2022). Instead, to leverage time-difference-based association while allowing for some temporal imprecision, and to include calls detected on just 1 or 2 units, we adapted a common localization algorithm (Nosal 2007) into a flexible combinatorial association procedure.

The MARU closest to a vocalizing whale would record the first instance of the upcall, or its 'first arrival', at the 'time-of-arrival' (ToA). MARUs slightly farther from the vocalizing whale but still within the detection range would record that same upcall at time delays commensurate with the distance from the whale, or 'time-difference-of-arrival' (TDoA). The maximum TDoA of a vocalization detected at 2 units would be equal to the distance between the units divided by the sound speed (at most equal to 15.3 s between MARUs 1 and 3). We performed association

by iterating through the classifier output and determining sets of upcall detections with physically plausible TDoAs. For each upcall under consideration, we restricted the search for associates to neighboring detections within a 15.3-s window and iterated through all unique combinations of sets of 5 detections, looking for a set for which the relative times could correspond to a source location within CCB. If a set was found, it was saved and removed from further consideration; if not, it was retained for further matching. Once a full iteration over the dataset was completed with sets of 5 detections, we repeated this association process with sets of 4, 3, and 2. This combinatorial procedure therefore prioritized larger matches across all units before attempting matches across fewer units.

To verify whether a set of detections could be duplicates of the same upcall, we first pre-computed the ToAs and TDoAs across the region of interest. Specifically, the CCB interior was discretized into a grid, with a resolution of 0.005° in latitude and longitude (approximately 0.56 and 0.42 km, respectively). For each grid location $\vec{x}[i,j]$, we calculated the travel time to each hydrophone position $\vec{h_k}$, with k in {1, 2, 3, 4, 5}, yielding ToA grids T_k:

$$T_k[i,j] = \frac{1}{c}\left|\left|\vec{x}[i,j] - \vec{h_k}\right|\right| \quad (1)$$

where c is the speed of sound. Next, pairwise differences of the 5 ToA grids yielded 25 TDoA grids representing the difference in arrival times for each geographic location and each pair of hydrophones:

$$\Delta_{k,l}[i,j] = T_k[i,j] - T_l[i,j] \quad (2)$$

Note that $\Delta_{k,k}[i,j]$ is identically 0 for all k, and $\Delta_{k,l}[i,j] = -\Delta_{l,k}[i,j]$ for all pairs l and k.

If a call was detected on MARU k and MARU l with a measured time difference of Δ_t, we could isolate the geographic locations consistent with the calculated theoretical TDoA grids. In particular, we computed a mask indicating where the measured TDoA differed from the theoretical TDoA by no more than a chosen parameter τ, representing the acceptable temporal error:

$$\begin{aligned} M_{k,l}[i,j] &= 1 \text{ if } |\Delta_{k,l}[i,j] - \Delta t| < \tau \\ M_{k,l}[i,j] &= 0 \text{ otherwise} \end{aligned} \quad (3)$$

If a call was detected on MARU k but not MARU l, we inferred that the source was positioned closer to h_k than to h_l. This yielded a different geographic mask:

$$\begin{aligned} M_{k,l}[i,j] &= 1 \text{ if } \Delta_{k,l}[i,j] < 0 \\ M_{k,l}[i,j] &= 0 \text{ otherwise} \end{aligned} \quad (4)$$

Altogether, feasible locations for the sound source would be defined by an intersection of the geographic masks across all hydrophone pairs:

$$L[i,j] = \prod_{k=1}^{5}\prod_{l=1}^{5} M_{k,l}[i,j] \quad (5)$$

Lastly, we multiplied the output $L[i,j]$ by a final mask $W[i,j]$ indicating which of the elements in the grid are admissible source locations. Specifically, as the rectangular grid comprised of locations $\vec{x}[i,j]$ covers CCB along with part of the shoreline, we set $W[i,j] = 1$ if $\vec{x}[i,j]$ is in the water and $W[i,j] = 0$ if $\vec{x}[i,j]$ is on land. Altogether, we calculated the final output $G[i,j] = L[i,j]W[i,j]$ for each set of possible associated detections. If $G[i,j]$ contained nonzero values, we concluded that the selected set has a feasible source location and, therefore, may be associated with one distinct upcall. We recorded this association and removed the detections from further consideration. If not, we concluded that the set cannot be associated, and continued association with alternative combinations of detections. Lastly, we visualized associated detections from the acoustic survey to verify this procedure.

At a latitude of 41.89, a depth of 30 m, salinity of 35 ppt, and water temperature between 2°C and 8°C, we estimated sound speeds between 1458 and 1482 m s^{-1} via the National Physical Laboratory's online calculator for speed of sound in sea water; we proceeded to use an average value of 1470 m s^{-1} for this work. In considering anticipated temporal errors, we estimated the difference in sound travel time associated with water temperature variation to remain within ±0.1 s, and the temporal error of the automated bounding box returned by the detector to remain within ±0.2 s. Despite synchronization at the beginning and end of the survey, relative clock drift due to temperature variation could introduce additional temporal error of unknown magnitude. Altogether, we chose a parameter of τ = 0.5 s to represent admissible temporal errors, with the expectation that the association procedure would be conservative. Following detection and association, we calculated the number of upcalls within every hour throughout the acoustic survey to quantify and visualize seasonal and diel patterns in NARW vocal activity within CCB.

2.2.3. Abundance estimation

To estimate daily NARW abundance in CCB, we calibrated acoustic cue counts, an index of relative

abundance, to sightings observed during the 16 d of concurrent aerial surveys. First, we applied additional processing to match the acoustic and aerial survey data as closely as possible in both space and time. Spatially, the MARU array was designed to detect upcalls from across the interior of CCB, in agreement with the aerial surveys. While we did not evaluate the acoustic detection range of upcalls in this work, the statistics of call association (discussed in Section 3.2) supported the assumption that the acoustic survey covers the majority of the bay. To account for increased uncertainty of partial aerial surveys in the regression, each data point was weighted by the number of completed track lines: 13 points had a weight of 1, 2 points had a weight of 13/15 = 0.87, and one point had a weight of 12/15 = 0.8. Temporally, it is clear from aerial surveys that the number of observed individuals could vary significantly on consecutive days; for instance, 11 individuals were observed on 8 May and only 2 on 9 May. This suggests that NARWs may migrate into and out of CCB over the course of the day, motivating the analysis of upcalls occurring specifically within the time frame of the aerial surveys. Consequently, we included upcalls occurring between the median start time and median end time of the aerial surveys: from 09:12 to 15:09 h EST (henceforth, 'median flight interval'). Note that while using the actual flight intervals rather than the median would provide the closest match, such analysis would impede extrapolation to days without co-occurring flights.

Next, we statistically modeled the number of individual NARWs in CCB on a particular day based on the number of associated upcalls detected during the median flight interval of that day. Count-based data is typically modeled with Poisson regression, which represents the response variable (the number of whales) as a Poisson distribution around an expected value that grows as the natural exponent of a linear function of the independent variable (the number of upcalls). By design, Poisson regression enforces non-negativity of predictions and of corresponding uncertainty intervals. However, Poisson regression assumes that the variance of the response variable is equal to the mean, an assumption that is commonly violated in ecological data ('overdispersion'; Ver Hoef & Boveng 2007). Instead, quasi-Poisson regression can be applied, which adjusts the uncertainty of the model to account for higher variance in the data. Additionally, monotonic data transformation is a common technique in statistical modeling to represent more diverse relationships between the dependent and independent variables, and is particularly relevant in our case, as the exponential relationship in the standard (linear) model may not be representative of upcall production. Accordingly, after verifying that our data exhibited overdispersion, we applied quasi-Poisson regression, with a comparison of 3 transformations of the data — a linear model, a square-root model, and a logarithmic model:

$$\text{Linear: } \log(E[\text{whales}]) = \alpha + \beta\,[\text{upcalls}]$$
$$\text{Square-root: } \log(E[\text{whales}]) = \alpha + \beta\,\text{sqrt}[\text{upcalls}]$$
$$\text{Logarithmic: } \log(E[\text{whales}]) = \alpha + \beta\,\log[\text{upcalls}]$$

We applied these 3 models with several variations for the values of [upcalls]: the number of associated upcalls during the median flight interval, the number of detections across all units during the median flight interval, and the number of detections on the central unit (MARU 5) over the median flight interval. By comparing the models, we aimed to assess the effects of both association and the usage of a 5-unit array on abundance estimation. After fitting the models, we evaluated and compared quality-of-fit. Though Akaike's information criterion (AIC) is not well defined for quasi-distributions, we evaluated the AIC of the corresponding Poisson model, as the model comparison remains within the same distributional family. We additionally reported the root-mean-square error (RMSE) weighted by the proportion of track lines surveyed. Lastly, to evaluate generalizability, we performed leave-one-out cross-validation with each model and report root-mean-square error, also weighted by the proportion of track lines surveyed (indicated by CV RMSE). Once the best model was identified, we reported and interpreted the model parameters and overall fit.

Next, we estimated abundance throughout the acoustic survey by calculating the counts of associated upcalls within the median flight interval, passing these values through the model, and calculating the anticipated number of NARWs as well as the corresponding 95% prediction intervals. Uncertainty intervals were calculated through bootstrapping, as there is no analytic representation for prediction intervals for a quasi-Poisson distribution. These steps concluded the analysis pipeline and yielded daily estimates of NARW abundance within CCB from February to June 2019. We also calculated mean estimates of NARW abundance over half-month intervals. All pre-processing, analysis, and visualization, including the detection and association procedures, were implemented in Python (3.11), relying primarily on the *numpy*, *scipy*, *pandas*, *geopandas*, *matplotlib*, and *suntime* packages. The quasi-Poisson regression and predictions were calculated in R using the *stats* and *ci_tools* packages.

3. RESULTS

3.1. Upcall detection

To select a threshold score for the classifier, we manually validated a subset of 104 021 detections across all 5 recorders throughout the acoustic survey. An evaluation of precision (i.e. the percentage of detected upcalls that are truly upcalls), recall (i.e. the percentage of upcalls that were successfully detected), and false positive rate (i.e. the number of incorrect detections per hour) across confidence thresholds is provided in the Supplement at www.int-res.com/articles/suppl/n056p101_supp.pdf. To jointly optimize precision and recall, we selected a threshold of 0.75, corresponding to a precision of 0.857, a recall of 0.896, and an average of 4.8 false positives per hour. A total of 485 880 upcalls scoring at or above the confidence threshold were detected by the classifier across the 5 recorders throughout the acoustic survey. NARWs are acoustically detected in CCB as early as mid-February (17 February) and vacate CCB by mid-May (15 May). Fig. 2B shows examples of upcalls detected by the automated classifier. Note that while all upcalls are characterized by an upsweep structure in the 50–350 Hz range, there is variability in shape and signal-to-noise ratio across different samples.

3.2. Upcall association

After association, we obtained 249 539 unique upcall instances between 17 February and 1 June 2019 (inclusive); all subsequent calculations consider this duration due to near-zero upcall detections in June. Of these, 10 913 (4%) were associated across all 5 channels, 15 360 (6%) across 4 channels, 41 822 (18%) across 3 channels, 60 811 (24%) across 2 channels, and the remaining 120 633 (48%) were unassociated. To visually verify the association procedure, Fig. 2B demonstrates detections associated across all 5 channels. As expected, the first arrival is higher in amplitude than its associates on different channels, though the relative differences in amplitude may vary depending on the location of the vocalizing whale. Additionally, some signal distortion due to propagation is visible across the examples. While relative clock drift between MARUs could affect association, we verified that association across all 5 units could be performed accurately throughout the acoustic survey. As an example, the 4 upcall instances shown were recorded on 20 February, 5 March, 7 April, and 9 May (in order from top to bottom). Overall, the total number of daily associated calls is highly correlated with the total number of daily detections, with a Pearson's correlation coefficient of 0.992 between these metrics across the duration of the acoustic survey. Additionally, the daily number of detections on a single unit was also highly correlated with the daily number of overall detections and with the number of daily associated calls; for MARU 5, the Pearson correlation coefficients were 0.988 and 0.977, respectively.

After completing the association step, we assessed vocal activity patterns. Fig. 3 displays hourly upcall occurrence throughout the acoustic survey. We observed diel variability in NARW vocal activity in CCB, with upcall rates higher at night (between sunset and sunrise) than during the day (between sunrise and sunset), consistent with prior studies (Mellinger et al. 2007, Mussoline et al. 2012) reporting on upcall rates measured in foraging grounds. Notably, throughout the acoustic survey, 62% of first arrivals occur at night, and consequently only 38% during daylight hours; furthermore, only 14% of the upcalls occur during the median flight interval.

3.3. Abundance estimation

Table 2 displays the calculated parameters (α and β) and quality-of-fit metrics (Poisson AIC, RMSE, and CV RMSE) for the 9 fitted models. Lower AIC and RMSE values indicate a better fit. Overall, using associated upcalls yielded an improvement over using total detections across the array, and a greater improvement over using detections from a single unit. The monotonically transformed models gave a significantly better fit than the standard linear models. While we observed a slightly lower CV RMSE for the square-root-transformed model with associated calls, the log-transformed models exhibited higher performance based on AIC and RMSE, and more realistic estimates at high upcall counts. Altogether, we applied the log-transformed model with associated calls for further inference and analysis, which implied the relationship:

$$\log([\text{whales}]) = -1.78 + 0.972 \log([\text{upcalls}])$$

Equivalently:

$$[\text{whales}] = 0.169\ [\text{upcalls}]^{0.972}.$$

Notably, this model suggests that the expected value of NARW individuals varies nearly linearly with the number of upcalls. Fig. 4 displays the expected values of the model along with the 95% confidence

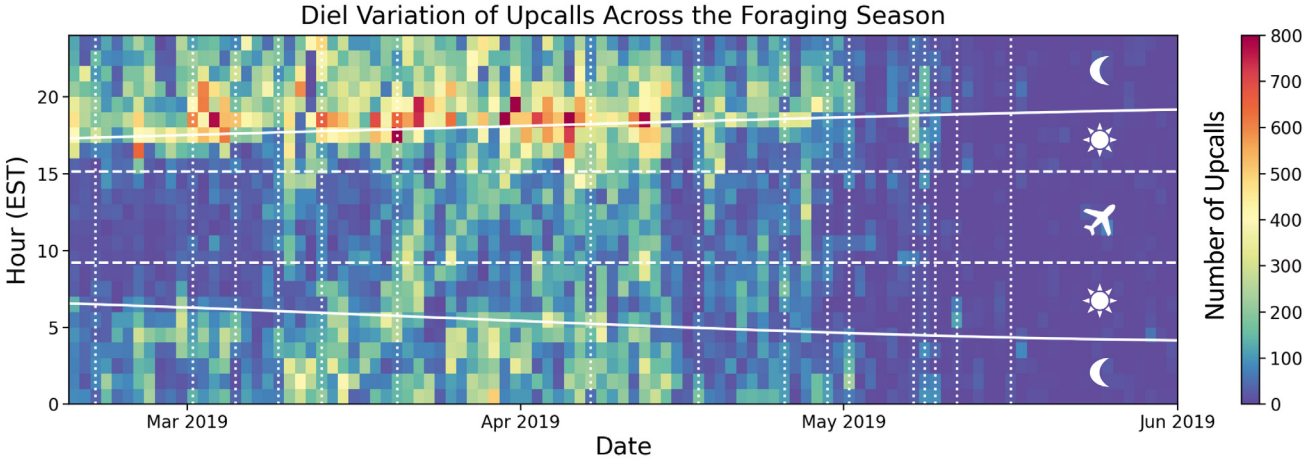

Fig. 3. After completing automated detection and association, we examined the number of distinct upcalls at hourly and daily scales throughout the acoustic survey. The color indicates the number of upcalls (varying from 0 in dark blue to 800 in dark red), horizontal solid white lines indicate daily sunrise and sunset, horizontal dashed white lines indicate the median start/end times of the aerial surveys, and vertical dotted white lines indicate aerial survey days. Days after 1 June are not shown due to very low upcall counts. The observed diel and seasonal patterns in vocalization activity are consistent with prior studies

interval for the parameters and the 95% prediction interval for new data points. For reference, at the median upcall count for the model (212), the 95% confidence interval width is 24–39 individuals, and the prediction interval width is 11–59 individuals. We discuss the contributing sources of uncertainty, as well as implications for ecological inference, in Section 4.3. Finally, Fig. 5 presents daily abundance estimates of NARWs in CCB throughout the acoustic survey. For each day, the figure shows the predicted number of individuals, the 95% prediction interval, and — on days with aerial surveys — the actual number of sightings. Days on which we extrapolate to upcall counts higher than observed in the 16 samples are indicated with transparent bars, and consequently have very high uncertainty. Considering half-month intervals, the average predicted NARW abundance is 20 whales over 17–28 February, 46 whales over 1–15 March, 89 whales over 16–31 March, 92 whales over 1–15 April, 52 whales over 16–30 April, 15 whales over 1–15 May, and 6 whales over 16–31 May.

Despite the high degree of correlation in detections across a single unit, detections across the array, and associated calls throughout the acoustic survey, there were notable differences in the models, with substantially lower quality-of-fit for models incorporating only MARU 5. To more clearly quantify these differences, we calculated the change in predicted daily abun-

Table 2. We considered different variations of models relating acoustic cue counts (upcalls) to abundance (the number of identified NARWs during aerial surveys). Three processing pathways of varying complexity were considered: using a 5-unit array and associating calls through time-difference-of-arrival (top 3 rows), counting all calls on the 5-unit array without association (central 3 rows), and counting only calls on the central recording unit (bottom 3 rows). For each of these variations, we compared quasi-Poisson models with 3 different monotonic transformations (linear, square-root, and logarithmic), with and without the association procedure. For each model, we report the regression parameters α and β along with statistical significance (*p < 0.05, **p < 0.01, ***p < 0.001). Quality-of-fit was evaluated with Akaike's information criterion (AIC) of the corresponding Poisson model, weighted root-mean-square-error (RMSE), and weighted root-mean-square-error of leave-one-out cross-validation (CV RMSE). Ultimately, we selected the logarithmic model with associated upcalls (in bold)

	Model formulation	α	β	AIC (Poisson)	RMSE	CV RMSE
Associated upcalls during median flight interval	Quasi-Poisson (linear)	2.69***	0.00275***	209	13.1	16.4
	Quasi-Poisson (square-root)	1.71***	0.111***	175	11.7	13.4
	Quasi-Poisson (logarithmic)	**−1.78***	**0.972*****	**149**	**11.5**	**14.3**
Detections during median flight interval	Quasi-Poisson (linear)	2.77***	0.00156***	222	14.9	17.9
	Quasi-Poisson (square-root)	1.90***	0.0794***	184	13.4	17.0
	Quasi-Poisson (logarithmic)	−1.52	0.861***	156	12.9	16.8
Detections on central unit during median flight interval	Quasi-Poisson (linear)	2.83***	0.00691***	244	18.8	30.8
	Quasi-Poisson (square-root)	2.01***	0.163***	197	16.5	25.6
	Quasi-Poisson (logarithmic)	0.110	0.801***	160	14.8	20.3

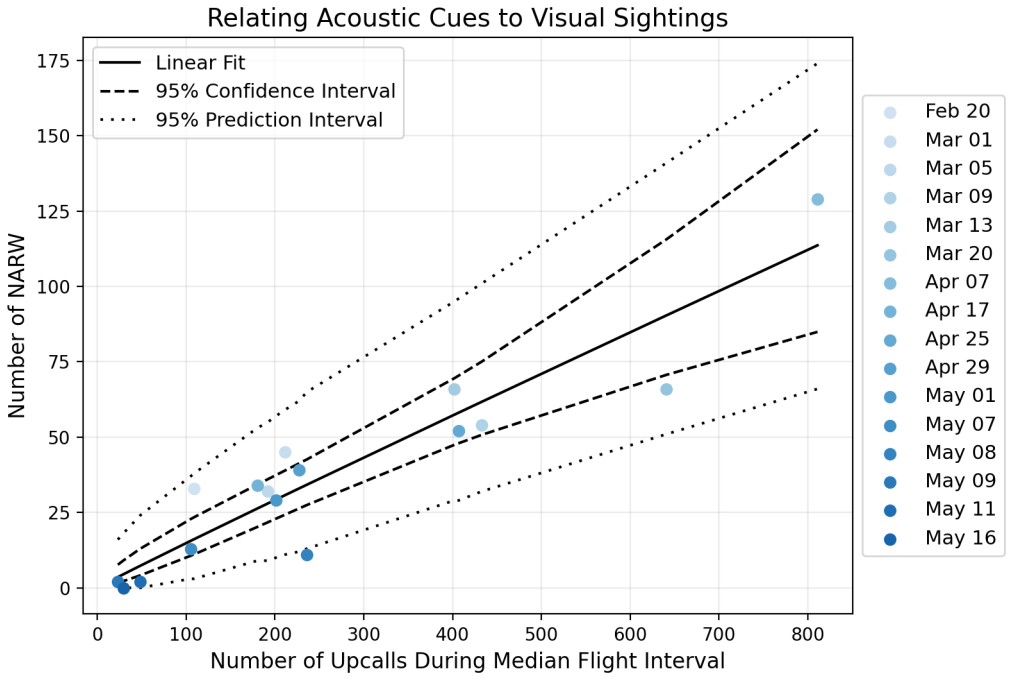

Fig. 4. To perform acoustic abundance estimation, we related the number of distinct upcalls to the number of sightings of North Atlantic right whales (NARWs) with quasi-Poisson regression with a logarithmic transformation. Each of the 16 points represents a day with concurrent acoustic and aerial surveys. The solid black line indicates the resulting fit, the dashed black line indicates the 95% confidence interval for the sample mean, and the dotted black line indicates the 95% prediction interval

dance between models as a percentage of the best-fitting model's prediction. In absolute value, the predictions resulting from the logarithmic model with associated calls differ from the predictions of the logarithmic model with detections by 11 ± 12% d^{-1} (mean ± SD), with a maximum difference of 24 individuals. Similarly, the predictions resulting from the logarithmic model with associated calls differ from the predictions of the logarithmic model with detections on the central unit by 20 ± 8% d^{-1}, with a maxi-

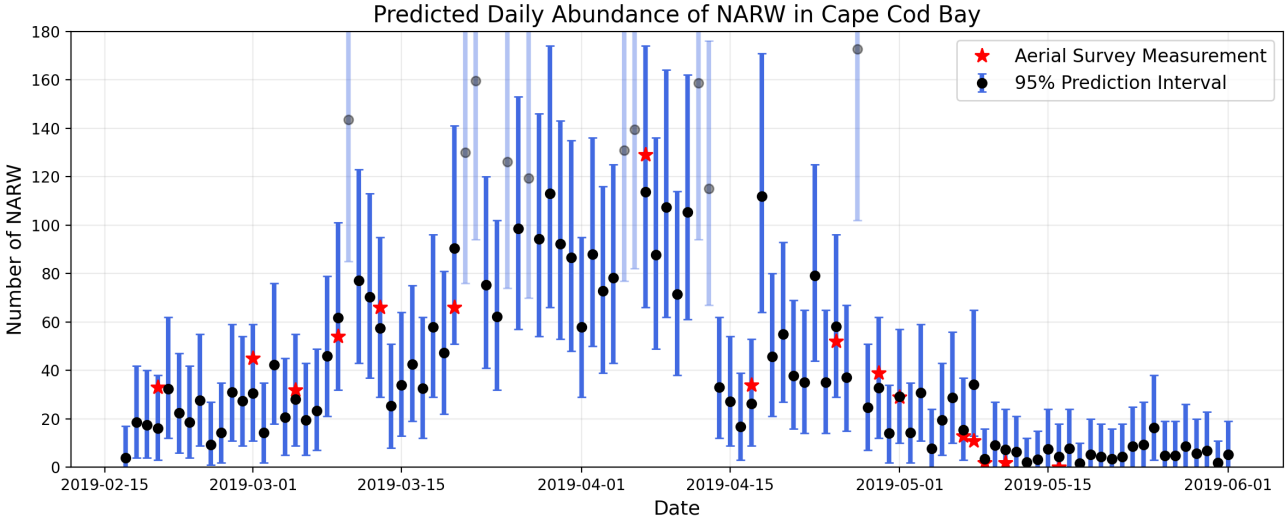

Fig. 5. Using quasi-Poisson regression to model whale observations as a function of the number of upcalls, we estimated NARW abundance in CCB throughout the acoustic survey. For each day, the red stars indicate aerial counts that were used to fit the model, the black points indicate the abundance prediction, and the blue bars indicate the 95% prediction interval. Days on which the model extrapolates to higher values than present in the calibration dataset are indicated with transparent points and bars

mum difference of 27 individuals. Lastly, the predictions resulting from the logarithmic model with detections across the array differ from the predictions of the logarithmic model with detections on the central unit by 16 ± 13% d^{-1}, with a maximum difference of 23 individuals.

4. DISCUSSION

4.1. NARW phenology

Broadly, our analysis of acoustic activity and abundance agrees with previous studies reporting diel patterns of upcall production in foraging grounds as well as seasonal presence in CCB. Throughout the season, NARWs exhibited a consistent diel pattern of upcall production, with higher call rates during light-limited hours than during daylight hours. Our results are consistent with previous PAM studies that document a rise in upcall rates at night in multiple other foraging grounds: from north to south, Roseway West (Nova Scotia, Canada), Jeffreys Ledge (MA, USA), and Stellwagen Bank National Marine Sanctuary (MA, USA; Mellinger et al. 2007, Mussoline et al. 2012). Additionally, Mayo et al. (2018) found that NARW seasonal occurrence in CCB varied minimally across 16 yr of aerial survey data, with consistent presence between January and mid-May. Our results uphold this trend for the months of this period for which we have data (early February through mid-May). Furthermore, assessing nearly 6 yr of PAM data, Charif et al. (2020) showed that NARW presence in Massachusetts Bay (MA, USA) is highest in late winter to spring, occurring as early as the first weeks of March and as late as the final week of April. Since Massachusetts Bay and CCB are adjacent habitats, NARW peak presence in CCB could be expected to correspond to peaks observed in Massachusetts Bay. Our results show that peak presence in CCB occurs in mid-March to mid-April, in agreement with those documented in Charif et al. (2020).

4.2. Analysis of findings

4.2.1. Detection

Passive acoustic data has the potential to estimate abundance at a higher time resolution than can be achieved through aerial surveys alone, but the manual tasks of the analytical workflow can delay these outputs significantly. While aerial surveys can provide abundance estimates on the same day as collection, PAM requires that the recording units are recovered and the data is pre-processed, analyzed, and summarized before abundance estimates are attained. Manually logging and even validating detector output can be unwieldy for long-term datasets. Consequently, using passive acoustic data to estimate abundance could take, at earliest, several months after initial data collection, a pace not always compatible with management timelines for NARWs or other threatened species. In this work, training a CNN was vital for rapid automated detection of hundreds of thousands of upcalls and the subsequent analysis of NARW phenology and abundance.

4.2.2. Association

The TDoA-based association procedure was effective at matching duplicate detections of the same upcall, and improved downstream abundance estimation based on model performance metrics. However, the algorithm may over- or under-associate under certain conditions. In particular, this approach associates opportunistically, and therefore may group together separate upcalls if they happen to occur at compatible times. In consequence, it may over-associate when calling density is high, and therefore may have caused a slight underestimation of the total number of upcalls in late March and April. In contrast, the algorithm assumes that a detection at a farther unit must imply detection at a closer unit, which may not always occur due to differing noise levels at different units and may therefore cause under-association. Incorporating signal similarity metrics within the association algorithm could help to improve the performance of this method in future work. The association statistics carry some implications for the upcall detection range and for the spatial distribution of vocalizing whales. Notably, as MARUs 1 and 3 are on the eastern and western corners of the array and are separated by a distance of 22 km, the repeated association of upcalls across 4 or 5 MARUs suggests that upcalls are frequently detected at distances further than the 9 km estimate from Clark et al. (2010). Furthermore, we find that association statistics for a particular day imply spatial distributions of upcalls that agree closely with the spatial distribution of aerial survey sightings. For instance, on 7 April, the visual survey revealed a large number of individuals clustered in the northeast corner, while the acoustic associations during the median flight interval indicated that out of 811 upcalls, the greatest contributions

were from unassociated calls on MARU 1 (136), unassociated calls on MARU 4 (131), and calls heard across MARUs 1, 4, and 5 (104). Similarly, on 1 May, sightings were concentrated in the southeast corner of CCB, and about 12% of all upcalls during the median flight interval were calls associated across MARUs 3, 4, and 5, compared to only 4% associated between MARUs 1 and 2. In future work, this association procedure could thereby permit a natural extension to acoustic spatially explicit capture–recapture (aSECR), an emerging abundance estimation technique (Marques et al. 2013, Franklin et al. 2022).

4.2.3. Abundance estimation

NARWs have been a historically difficult species for which to conduct acoustic abundance estimation because of the high individual-level variance in upcall rate. However, in contrast to abundance estimation using individual call rates, calibration with daily sightings data focuses on population-level call rate and allows for the representation of nonlinearity in upcall production. Overall, this study was fortunate to have access to independent abundance estimates within the same site and time frame as the acoustic survey, minimizing error and uncertainty associated with transferring call rates across populations and environments, as is often unavoidable in abundance estimation studies. Based on quality-of-fit metrics, association slightly improved the performance of models for abundance estimation, as did the usage of a 5-unit array over a single unit. Notably, on average, the differences in abundance estimates were smaller in magnitude than the uncertainty intervals; however, on individual days, these differences could be substantial. The small sample size available for this study limited our evaluation of abundance estimation accuracy and was insufficient for characterizing the effects of environmental factors, such as seasonality or temperature, on the calculated call rates. Further acoustic data collection would improve our understanding of these factors on abundance estimation and assess the variability of NARW vocal behavior across multiple years.

4.3. Limitations and sources of uncertainty

The abundance estimates themselves still bear a high degree of uncertainty that can complicate how to interpret these values for endangered species. Our estimates are concordant with well-understood NARW phenological patterns, yet we acknowledge that these estimates are not yet precise. On an average day, a 95% prediction interval can span approximately 50 whales. Equivalent to about 15% of the remaining population, this variance may be too great for confident integration within management planning at this time.

Multiple sources of uncertainty may be contributing to this variance. One source of uncertainty is the contrasting attributes of the 2 survey modalities, which impart differing observation and availability biases onto the visual and acoustic measurements. An aerial survey may cover a greater spatial extent across CCB overall, but each stretch of track line is observed briefly, with only whales near or at the surface being counted. On the other hand, acoustic methods are constrained by vocal activity and upcall detection range. Consequently, acoustic surveys may have smaller spatial coverage but can record NARWs at various depths and maintain simultaneous monitoring of the detectable area for each aerial survey. To temporally match the aerial and acoustic data, we only consider upcalls detected during the median flight interval, which is logistically limited to daylight hours. Per the diel variation patterns we observed in Fig. 3, NARWs are less vocally active during the day, which means we calibrated the acoustic–visual measurements at a time of day when the call rate is lower. While this should not affect the fidelity of our analysis, as the population-level call rates never fell silent when whales were present in CCB, the abundance relationship calculated in this work should not be extrapolated to nighttime hours. Furthermore, vessel traffic is highest during the day, causing heightened background noise levels that may reduce the quality of some recorded upcalls. A lower-quality upcall may fall below our selected threshold (as determined via the precision–recall curve) and thus would be excluded from the upcall count during its respective survey, possibly contributing more uncertainty to the model. Accumulation of error throughout the analysis pipeline, including false positive detections and incorrect associations, also can contribute to uncertainty. Both false positive detections and incorrect associations can arise from variance in environmental conditions, which we did not characterize in this study. Finally, the model's goodness-of-fit is ultimately limited by the availability and quality of 'ground-truth' abundance measurements. In this study, we benefitted from ground-truth aerial survey data that removed duplicate sightings, a level of accuracy that can only be attained for well-studied populations (such as NARWs) with identifiable individuals and

thus is a rarity at scale. In future applications of such a pipeline, a practitioner would likely only have access to raw sightings data, which would further exacerbate the uncertainty in final abundance estimates. While we see a statistically significant relationship in the model with 16 survey samples, if resources allow in future studies, a larger number of ground-truth abundance measurements would help to further characterize sources of uncertainty.

Previous studies have reached mixed conclusions about the efficacy of PAM for NARW abundance estimation due to the uncertainty introduced by behavior (i.e. group size, individual call rate). For instance, NARW foraging behavior may subject sightings reported by the aerial surveys to seasonal observation bias. Ganley et al. (2019) suggested that seasonal shifts in prey could influence the time that NARWs are observable at the surface and thus detectable by an aerial survey. Winter prey *Pseudocalanus* spp. aggregate in layers on the basin floor, driving down surface time to 34% in February (Ganley et al. 2019). In the spring, NARWs transition to surface-dwelling prey (e.g. *Calanus finmarchicus*), raising surface time to 55% by April (Ganley et al. 2019). Prey species availability and composition also undergo interannual variation, making seasonal observation bias difficult to predetermine for any one season. We acknowledge that the seasonal observation bias may affect the aerial survey sightings used in this study and thus may contribute uncertainty to the abundance estimates. Furthermore, NARW behavior may also introduce uncertainty via the acoustic measurements. NARW average call rates, particularly during the day, are very low: the calculated model implies an average call rate of about 6–8 associated upcalls per whale per day. Consequently, relative to an equivalent study conducted for a species with a higher call rate, these abundance estimates are more sensitive to a single upcall, yielding higher uncertainty. Additionally, while Durette-Morin et al. (2019) found a significant relationship between upcall production and abundance, Clark et al. (2010) and Franklin et al. (2022) did not. Of the latter, Franklin et al. (2022) suggested that this discrepancy may be attributable in part to higher variability in upcall rates observed in social aggregations. When group size increases, NARWs integrate upcalls within acoustic displays associated with surface active groups, groups of 2 or more NARWs engaging in close physical contact at the surface for putative reproductive purpose (Kraus & Hatch 2001, Parks & Tyack 2005). The quasi-Poisson regression we used to fit acoustic and aerial measurements reflects this ecological pattern, as the variance grows with the mean over time; consequently, as upcall counts and NARW sightings increase, the confidence interval widens and uncertainty compounds. This aligns with the growing uncertainty associated with the acoustic–visual calibration (Fig. 4) when NARW group size grows. Therefore, we posit that our acoustic–visual model may only hold up in smaller group sizes, which is consistent with previous work. The majority of observations within the aerial surveys included in our study was of individuals, with occasional groups of 2–3 whales, aligned with the smaller group focus reported in Durette-Morin et al. (2019) that also found a significant relationship between upcall production and abundance. In circumstances where group size is higher, all 3 NARW call types (upcalls, gunshots, tonal calls) may yield improved abundance estimates, as suggested in Franklin et al. (2022).

In line with similar studies (e.g. Franklin et al. 2022), we did not quantify the error and uncertainty associated with ground-truth aerial surveys. For instance, while the 3 partial aerial surveys only excluded 2 or 3 track lines, it is possible that a group of NARW was missed on those days, biasing the overall model towards underestimation. Ultimately, both acoustic and aerial surveys have limited detectability. Schliep et al. (2024) considered both aerial and acoustic surveys as providing only partial information on marine mammal abundance, and proposed a probabilistic methodology for fusing these 2 data sources, demonstrating its application for estimating NARW abundance in CCB. Without ground-truth abundance measurements, evaluation of acoustic abundance estimation is limited to simulation and consequently constrained by modeling assumptions. By integrating both aerial and acoustic surveys, we can get closer to demystifying what a detected upcall means in the context of overall species abundance.

5. CONCLUSION

Ultimately, expediting abundance estimation analysis via passive acoustic data could be a valuable addition to the suite of tools necessary for effective conservation. However, the trade-off between methodological limitations and precision remains a challenge. In combining very different survey modalities — aerial and acoustic — our abundance estimates have high uncertainty, which is especially consequential for a Critically Endangered species. Future studies should consider what degree of precision is realistic to

achieve with passive acoustic data alone, as well as how much uncertainty is acceptable for conservation applications. Further tests of this pipeline could reveal its adaptability for other endangered cetacean species, altogether increasing our collective capacity for monitoring in the service of conservation.

Acknowledgements. We thank Chris Tessaglia-Hymes and Dr. Christopher Clark at the K. Lisa Yang Center for Conservation Bioacoustics for instrument deployment and recovery. We also thank Amy James and Captain Marc Costa at the Center for Coastal Studies for facilitating aerial survey data collection and piloting the R/V *Shearwater*, respectively. We are also grateful to Kenneth Tyler Wilcox of the Cornell Statistical Consulting Unit for advice on the data analysis conducted for this project. This research was predominantly funded by the US Navy's Office of Naval Research (award no. N000014-18-1- 2808; program manager: Dr. Michael J. Weise) led by Dr. Scott Baker, Oregon State University. Further funding sources included the National Science Foundation Graduate Research Fellowship, Cornell Sloan Graduate Diversity Fellowship via the Alfred P. Sloan Foundation, Edward W. Rose Postdoctoral Fellowship, Herchel Smith Undergraduate Science Research Program, Harvard College Research Program, and the Museum of Comparative Zoology Grants-in-Aid of Undergraduate Research. Right whale research was carried out under NOAA Scientific Permit No. 19315. Finally, we thank three anonymous reviewers for their thoughtful and thorough feedback on this manuscript. The present work was part of the co-lead author's undergraduate honors thesis.

LITERATURE CITED

Abadi M, Agarwal A, Barham P, Brevdo E and others (2016) TensorFlow: large-scale machine learning on heterogeneous distributed systems. www.arxiv.org/abs/1603.04467v2 (Accessed 7 Sep 2024)

Barlow J, Taylor BL (2005) Estimates of sperm whale abundance in the Northeastern temperate Pacific from a combined acoustic and visual survey. Mar Mamm Sci 21: 429–445

Barlow J, Fregosi S, Thomas L, Harris D, Griffiths ET (2021) Acoustic detection range and population density of Cuvier's beaked whales estimated from near-surface hydrophones. J Acoust Soc Am 149:111–125

Barlow J, Moore JE, McCullough JLK, Griffiths ET (2022) Acoustic-based estimates of Cuvier's beaked whale (*Ziphius cavirostris*) density and abundance along the U.S. West Coast from drifting hydrophone recorders. Mar Mamm Sci 38:517–538

Brown MW, Marx MK (1998) Surveillance, Monitoring and management of North Atlantic right whales, *Eubalaena glacialis*, in Cape Cod Bay, Massachusetts: January to mid-May, 1998. Contract No. SCFWE3000-8365027, Commonwealth of Massachusetts, Boston, MA

Calupca TA, Fristrup KM, Clark CW (2000) A compact digital recording system for autonomous bioacoustic monitoring. J Acoust Soc Am 108:2582

Charif RA, Shiu Y, Muirhead CA, Clark CW, Parks SE, Rice AN (2020) Phenological changes in North Atlantic right whale habitat use in Massachusetts Bay. Glob Change Biol 26:734–745

Clark CW, Borsani JF, Notarbartolo-Di-sciara G (2002) Vocal activity of fin whales, *Balaenoptera physalus*, in the Ligurian Sea. Mar Mamm Sci 18:286–295

Clark CW, Gillespie DM, Nowacek DP, Parks SE (2007) Listening to their world: Acoustics for monitoring and protecting right whales in an urbanized ocean. In: Kraus S, Rolland RM (eds) The urban whale. Harvard University Press, Cambridge, MA, p 333–357

Clark CW, Brown MW, Corkeron P (2010) Visual and acoustic surveys for North Atlantic right whales, *Eubalaena glacialis*, in Cape Cod Bay, Massachusetts, 2001–2005: management implications. Mar Mamm Sci 26:837–854

Cooke JG (2020) *Eubalaena glacialis* (errata version published in 2020). The IUCN Red List of Threatened Species 2020:e.T41712A178589687. https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T41712A178589687.en

Corkeron P, Hamilton P, Bannister J, Best P and others (2018) The recovery of North Atlantic right whales, *Eubalaena glacialis*, has been constrained by human-caused mortality. R Soc Open Sci 5:180892

Davis GE, Baumgartner MF, Bonnell JM, Bell J and others (2017) Long-term passive acoustic recordings track the changing distribution of North Atlantic right whales (*Eubalaena glacialis*) from 2004 to 2014. Sci Rep 7:13460

Durette-Morin D, Davies KTA, Johnson HD, Brown MW, Moors-Murphy H, Martin B, Taggart CT (2019) Passive acoustic monitoring predicts daily variation in North Atlantic right whale presence and relative abundance in Roseway Basin, Canada. Mar Mamm Sci 35:1280–1303

Durette-Morin D, Evers C, Johnson HD, Kowarski K and others (2022) The distribution of North Atlantic right whales in Canadian waters from 2015–2017 revealed by passive acoustic monitoring. Front Mar Sci 9:976044

Estabrook B, Tielens J, Rahaman A, Ponirakis D, Clark C, Rice A (2022) Dynamic spatiotemporal acoustic occurrence of North Atlantic right whales in the offshore Rhode Island and Massachusetts Wind Energy Areas. Endang Species Res 49:115–133

Franklin K, Cole T, Cholewiak D, Duley P and others (2022) Using sonobuoys and visual surveys to characterize North Atlantic right whale (*Eubalaena glacialis*) calling behavior in the Gulf of St. Lawrence. Endang Species Res 49:159–174

Ganley LC, Brault S, Mayo CA (2019) What we see is not what there is: estimating North Atlantic right whale *Eubalaena glacialis* local abundance. Endang Species Res 38:101–113

Garrison LP, Adams J, Patterson EM, Good CP (2022) Assessing the risk of vessel strike mortality in North Atlantic right whales along the US East Coast. NOAA Tech Memo NMFS-SEFSC-757

Gillespie D (2019) DCLDE 2013 Workshop Dataset. University of St. Andrews. https://research-portal.st-andrews.ac.uk/en/datasets/dclde-2013-workshop-dataset

Helble TA, Ierley GR, D'Spain GL, Martin SW (2015) Automated acoustic localization and call association for vocalizing humpback whales on the Navy's Pacific Missile Range Facility. J Acoust Soc Am 137:11–21

Hudak C, Stamieszkin K, Mayo C (2023) North Atlantic right whale *Eubalaena glacialis* prey selection in Cape Cod Bay. Endang Species Res 51:15–29

Ibrahim AK, Zhuang H, Chérubin LM, Erdol N, O'Corry-Crowe G, Ali AM (2021) A multimodel deep learning

algorithm to detect North Atlantic right whale up-calls. J Acoust Soc Am 150:1264–1272
K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology (2024) Raven Pro: interactive sound analysis software. https://www.ravensoundsoftware.com/software/raven-pro/
Kenney RD, Winn HE, Macaulay MC (1995) Cetaceans in the Great South Channel, 1979–1989: right whale (*Eubalaena glacialis*). Cont Shelf Res 15:385–414 p
Kirsebom OS, Frazao F, Simard Y, Roy N, Matwin S, Giard S (2020) Performance of a deep neural network at detecting North Atlantic right whale upcalls. J Acoust Soc Am 147: 2636–2646
Knowlton AR, Clark JS, Hamilton PK, Kraus SD, Pettis HM, Rolland RM, Schick RS (2022) Fishing gear entanglement threatens recovery of critically endangered North Atlantic right whales. Conserv Sci Pract 4:e12736
Kraus SD, Hatch JJ (2001) Mating strategies in the North Atlantic right whale (*Eubalaena glacialis*). J Cetacean Res Manag 2:237–244
Madhusudhana S, Shiu Y, Klinck H, Fleishman E and others (2021) Improve automatic detection of animal call sequences with temporal context. J R Soc Interface 18: 20210297
Marchetto PM (2015) Addressing engineering challenges in bioacoustic recording. PhD dissertation, Cornell University, Ithaca, New York, NY
Marques TA, Thomas L, Ward J, DiMarzio N, Tyack PL (2009) Estimating cetacean population density using fixed passive acoustic sensors: an example with Blainville's 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. Endang Species Res 13:163–172
Marques TA, Thomas L, Martin SW, Mellinger DK and others (2013) Estimating animal population density using passive acoustics. Biol Rev Camb Philos Soc 88: 287–309
Marr D, Hildreth E (1980) Theory of edge detection. Proc R Soc B 207:187–217
Mayo CA, Ganley L, Hudak CA, Brault S, Marx MK, Burke E, Brown MW (2018) Distribution, demography, and behavior of North Atlantic right whales (*Eubalaena glacialis*) in Cape Cod Bay, Massachusetts, 1998–2013. Mar Mamm Sci 34:979–996
Mellinger DK, Nieukirk SL, Matsumoto H, Heimlich SL and others (2007) Seasonal occurrence of North Atlantic right whale (*Eubalaena glacialis*) vocalizations at two sites on the Scotian Shelf. Mar Mamm Sci 23:856–867
Meyer-Gutbrod EL, Greene CH, Sullivan PJ, Pershing AJ (2015) Climate-associated changes in prey availability drive reproductive dynamics of the North Atlantic right whale population. Mar Ecol Prog Ser 535:243–258
Miller BS, Madhusudhana S, Aulich MG, Kelly N (2023) Deep learning algorithm outperforms experienced human observer at detection of blue whale D-calls: a double-observer analysis. Remote Sens Ecol Conserv 9: 104–116
Morano JL, Rice AN, Tielens JT, Estabrook BJ, Murray A, Roberts BL, Clark CW (2012) Acoustically detected year-round presence of right whales in an urbanized migration corridor. Conserv Biol 26:698–707
Mussoline SE, Risch D, Hatch LT, Weinrich MT and others (2012) Seasonal and diel variation in North Atlantic right whale up-calls: implications for management and conservation in the northwestern Atlantic Ocean. Endang Species Res 17:17–26
New England Aquarium (2024) As the North Atlantic right whale population slowly increases, human activity remains a serious threat. New England Aquarium. https://www.neaq.org/as-the-north-atlantic-right-whale-population-slowly-increases-human-activity-remains-a-serious-threat/
Nichols OC, Kenney RD, Brown MW (2008) Spatial and temporal distribution of North Atlantic right whales (*Eubalaena glacialis*) in Cape Cod Bay, and implications for management. Fish Bull 106:270–280
Nosal EM (2007) Tracking marine mammals using passive acoustics. PhD dissertation, University of Hawai'i at Mānoa, Honolulu, HI
Pace RM, Corkeron PJ, Kraus SD (2017) State–space mark–recapture estimates reveal a recent decline in abundance of North Atlantic right whales. Ecol Evol 7:8730–8741
Parks SE, Tyack PL (2005) Sound production by North Atlantic right whales (*Eubalaena glacialis*) in surface active groups. J Acoust Soc Am 117:3297–3306
Parks S, Searby A, Célérier A, Johnson M, Nowacek D, Tyack P (2011) Sound production behavior of individual North Atlantic right whales: implications for passive acoustic monitoring. Endang Species Res 15:63–76
Roberts J, Yack T, Fujioka E, Halpin P and others (2024) North Atlantic right whale density surface model for the US Atlantic evaluated with passive acoustic monitoring. Mar Ecol Prog Ser 732:167–192
Rolland RM, Parks SE, Hunt KE, Castellote M and others (2012) Evidence that ship noise increases stress in right whales. Proc R Soc B 279:2363–2368
Sasse DB (2003) Job-related mortality of wildlife workers in the United States, 1937–2000. Wildl Soc Bull 31:1015–1020
Schliep EM, Gelfand AE, Clark CW, Mayo CA and others (2024) Assessing marine mammal abundance: a novel data fusion. Ann Appl Stat 18:3071–3090
Shiu Y, Palmer KJ, Roch MA, Fleishman E and others (2020) Deep neural networks for automated detection of marine mammal species. Sci Rep 10:607
Taylor BL, Rojas-Bracho L, Moore J, Jaramillo-Legorreta A and others (2017) Extinction is imminent for Mexico's endemic porpoise unless fishery bycatch is eliminated. Conserv Lett 10:588–595
Ver Hoef JM, Boveng PL (2007) Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data? Ecology 88:2766–2772
Watkins WA, Schevill WE (1976) Right whale feeding and baleen rattle. J Mammal 57:58–66
Zimmer WMX, Johnson MP, Madsen PT, Tyack PL (2005) Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*). J Acoust Soc Am 117:3919–3927

Editorial responsibility: Michael Noad, Gatton, Queensland, Australia
Reviewed by: G.E. Davis, T.A. Marques and 1 anonymous referee
Submitted: March 8, 2024; Accepted: December 16, 2024
Proofs received from author(s): February 17, 2025