POWER ANALYSIS FOR OPTIMAL DESIGN OF A PASSIVE ACOUSTIC MONITORING NETWORK IN THE VIRGINIA OFFSHORE WIND AREA

Authors:	Magda Chudzinska, Len Thomas
Report Code:	SMRUC-2023-010
Date:	Thursday, 10 August 2023

THIS REPORT IS TO BE CITED AS: CHUDZINSKA, THOMAS, ET AL. (2023) POWER ANALYSIS FOR OPTIMAL DESIGN OF PASSIVE ACOUSTIC MONITORING NETWORK IN THE VIRGINIA OFFSHORE WIND AREA. REPORT NUMBER SMRUC-RWS-2023-010 (UNPUBLISHED).

Document Control

Please consider this document as uncontrolled copy when printed.

Rev.	Date.	Reason for Issue.	Prep.	Chk.	Apr.	Client
1	05 May 2023	Version sent to Sue Barco	MC, LT			RWSC
2	12 May 2023	Version 2 sent to Sue Barco	MC, LT			RWSC
3	15 May 2023	Final Draft sent to RWSC	MC, LT	JW		RWSC
4	31 May 2023	Final sent to RWSC	MC, LT	JW, MR		RWSC

This project, Task #94.03, was funded by the Virginia Coastal Zone Management Program led by the Virginia Department of Environmental Quality through Grant #NA20NOS4190207 of the U.S. Department of Commerce, National Oceanic and Atmospheric Administration, under the Coastal Zone Management Act of 1972, as amended. The views expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Department of Commerce, NOAA, or any of its subagencies.

Contents

Table of Figures

1 Executive Summary

The goal of this project was to provide analyses and recommendations to support the optimal design of a baleen whale passive acoustic monitoring (PAM) network in the Virginia Offshore Area (VOWA). We used computer simulation to perform a statistical analysis, to determine the power to detect biologically realistic changes in whale distribution and behavior associated with construction and operation of the wind farm using three candidate designs. The three designs were: I) a regional PAM monitoring design provided by Van Parijs et al. (2021, doi:10.3389/fmars.2021.760840) consisting of a 20 x 20 km grid of sensors ("small PAM grid") located around the wind energy area (VOWA) and a 40 x 40 km grid ("large PAM grid"); II) a modified design with a 10 x 10 km grid replacing the small PAM grid; and III) a second modified design with a linear array of PAM stations in a T-configuration (i.e., with 3 "arms") centered on VOWA and denser sensor placement towards the center of the T, again replacing the small PAM grid. In all cases, each PAM station was assumed to consist of a bottom-mounted archival recorder and subsequent acoustic processing used to derive counts of detected vocalizations per unit time, which was assumed to be done independently at each station (so, for example, the possibility of localizing calls was not considered).

Statistical power is the probability of obtaining a statistically significant result in a hypothesis test, given that some specified change exists. We set a target power of 80% and a nominal false positive rate (i.e., probability of detecting a change if none exists) of 5%. We chose four study species: fin, sei, minke and North American right whale (NARW). We determined power using just the sensors within 20 km of VOWA ("small monitoring area") and within a larger circle around VOWA about as far as the continental shelf ("large monitoring area").

We simulated acoustic detection rate data by sampling from spatially- and temporally referenced whale density surfaces provided by Duke University at the spatial scale of 5 x 5 km and temporal scale of a month, applying the hypothesized changes due to wind farm construction or operation, and converting the simulated whale numbers to acoustic detection numbers using assumed average animal vocalization rates and effective detection ranges. We applied a statistical test to the simulated data, looking to see whether the relationship between acoustic detection rate and distance from wind farm changed between baseline data (collected over 1 year) and data collected during construction (1 year, only construction months). We repeated the simulation 500 times for each hypothesis, and species and calculated power as the proportion of simulations that yielded a statistically significant result.

Very little is known on how construction and operation of wind farms may affect baleen whales. We specified eight hypotheses on how simulated whales of the study species may respond. These could be broadly grouped into four categories: I) Construction and operation of wind farms had no effect on whale distribution and

acoustic behavior (H1[1] and H8). In H1 there was no systematic change in whale density or distribution over time; in H8 there was a region-wide decline, but unrelated to wind farms. II) Construction of wind farms caused a change in acoustic behavior (cue production rate) (H2). The change was modelled to be strongest close to the construction location and declined with increasing range. III) Construction of wind farms caused temporary displacement of whales (H3-H5). Again, the effect was modelled to be strongest close to the construction location; the hypotheses differ in whether the direction of displacement is dependent on habitat preference or other construction-related activities (such as shipping). IV) Operation of wind farms caused long-term change in whale distribution (H7). For hypotheses involving response to construction, we used two assumed dose-response functions, taken from a separate expert elicitation on the response of NARW to pile driving combined with a simple sound propagation model. In the less sensitive function (DR1[2]), there was a sharp decline in probability of response within the first 2 km from the source and reaching 1% at 18 km. In the more sensitive function (DR2), the probability of response declined gradually, reching 1% at 30 km.

Hypothesis H1 specifies no effect of wind farms and hence any significant result in the power analysis under this hypothesis is a false positive. We found that false positive rates were higher than the nominal level of 5%, ranging from 5-14%, but that false positive rate was lower when larger monitoring areas were used, and was lower under the T-design than the other designs.

Hypotheses H2-H5 involved effects of construction, using DR1 and DR2. We found that power under the Van Parijs et al. design was low (i.e., below the target of 80%) for all species and hypotheses. This was largely because the effect size, i.e., the proportion of animals monitored that responded, was small. Power was higher, but still below 80% threshold, if only sensors within 20 km of each VOWA were used (the “small monitoring area”), because this is focused on the area where the effect is strongest. Power was particularly low for fin whales because they can be detected over large distances, and so even sensors close to the sound source detect a mixture of responding and non-responding whales. Minke whales had the highest power, likely because they were only detectable over short ranges and had a higher acoustic encounter rate than other species (except fin whale). Sei whales and NARW typically had power values between fin and minke whales.

Power was also low under the Van Parijs et al. design to detect changes due to operation (H7).

The alternative monitoring designs both resulted in substantially higher power to detect effects of construction and operation but only for minke whales, because sampling effort was concentrated closer to the wind farm footprint where change was greatest.

[1] H indicated a hypothesis. These are numbered from 1 to 8.

[2] DR indicates a dose-response.

Power to detect the region-wide decline (H8) was high for minke whales, but not for the remaing species at VOWA under all designs. We did not explicitly examine power to detect long-term trends, but this result shows the potential for PAM as a population monitoring tool (although this would require additional information about spatial and temporal patterns of vocalization rate and acoustic detectability).

Our results are strongly contingent on the assumptions made and we suggest future studies that could be undertaken to improve the reliability and scope of the analysis. These include looking at finer temporal scales (which would require example acoustic data on which to base simulations) and examining alternative analysis methods.

Based on our results, we recommend replacing the 20 x 20 km small monitoring grid of sensors around VOWA with an alternative array that concentrates sensors where a response is expected and distributes sensors relatively evenly across the VOWA. Of the designs we tested, the T-design appears better than a 10 x 10 km grid of sensors, but other linear designs with closer sensor spacing nearer the centre of each wind farm are possible.

There is in addition a need for sensors at distances from the VOWA where no response is expected, and this role could be fulfilled by the 40 x 40 km grid. Monitoring over a larger area reduced the false positive rate.

To maximize the sample size of acoustic sensors we recommend pooling resources across stakeholders who are deploying sensors. Power will also likely be higher if analyses of construction effect were combined across VOWA and the neighbouring wind areas, and this is one of the additional investigations we recommend.

For species like fin whales with large acoustic detection distances, consideration should be given to localizing calls and undertaking effects analysis using the localizations.

One method to improve power is to accept a higher false-positive detection rate. We used a nominal false-positive rate of 5% and a target power of 80%, but these values are conventions and consideration could be given to using other values.

2 Introduction

2.1 Aims of the project

The overall aim of the project is to provide the RWSC Marine Mammal Subcommittee with analyses and recommendations to support the optimal design of a baleen whale Passive Acoustic Monitoring (PAM) Network in the Virginia Offshore Area (VOWA). A recommended regional PAM monitoring design for the Northeast U.S. offshore wind energy areas was provided by Van Parijs et al. (2021) (see Figure SI-2 in that paper, referred to here as the "Van Parijs et al. grid"). The design is based on a network of PAM stations at two spatial resolutions: a 20 x 20 km grid around the VOWA (referred to here as the "small PAM grid"), and a 40 x 40 km grid between the VOWA and neighboring wind farms (referred to as "large PAM grid"). Each PAM station comprises a single hydrophone bottom-mounted archival recorder. Here we present results, based on computer simulation, of the statistical power of this proposed design to detect biologically realistic changes in whale distribution and behavior associated with construction and operation of wind farms within the VOWA. We additionally evaluate two alternative designs: (1) a 10 x 10 km small PAM grid, and (2) a linear array of PAM stations in a T-configuration (i.e., with 3 "arms") centered on each development area ("T-design"). We follow convention in setting a target power (i.e., probability of detecting a change if it exists) of 80% or greater and a false positive rate (i.e., probability of detecting a change if none exists) of 5%.

The remainder of this document is laid out as follows. In the Introduction, we provide a brief overview of the spatial ecology of baleen whale species along the U.S. East Coast and a summary of the observed effects of disturbance related to offshore wind farm development. We use this information to develop hypotheses for what changes might occur, including also possible changes not related to offshore wind farms. These hypotheses of change provide the basis for our simulation study, focusing on the acoustic signal generated by the change. We give an overview of issues to be considered in using PAM to detect such changes. In the Methods section, we provide an overview of the power analysis and give details on each step of the analysis. We then present Results and a Discussion of the results, the caveats accompanying them, and our recommendations.

This report is part of larger, regional analysis for optimization of PAM monitoring along the East Coast of U.S. described in detail in Chudzinska and Thomas (2023).

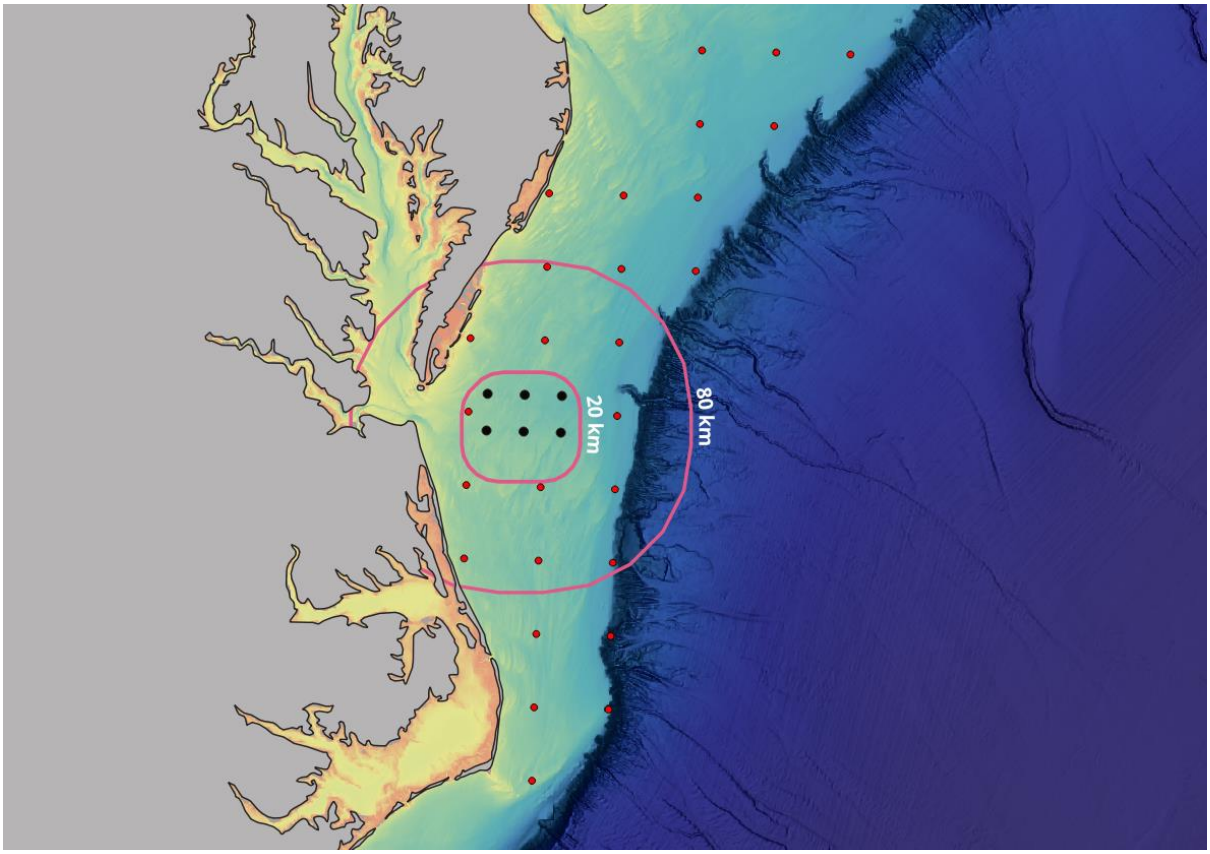


Figure 1. The Virginia Offshore Area sites, with monitoring areas shown around it and the PAM grids. The study site has two sizes of monitoring areas: 20 km buffer and larger buffer spanning over the continental shelf. The two PAM grids (Van Parijs et al. 2021) have a 40 km grid spacing (red dots) and a 20 km spacing (black dots).

2.2 Baleen whales along the East Coast of the U.S. and in VOWA area

Six baleen whale species are found in the western North Atlantic along the U.S. East Coast: minke (*Balaenoptera acutorostrata*), North Atlantic right (NARW; *Eubalaena glacialis*), humpback (*Megaptera novaeangliae*), sei (*Balaenoptera borealis*), fin (*Balaenoptera physalus*), and blue (*Balaenoptera musculus*) whales. These are mainly migratory species moving annually between summer breeding grounds and winter foraging areas.

The coastal waters off Virginia, U.S., contain a migratory corridor for several baleen whale species, such as fin, minke, humpback and North Atlantic right whales (NARW). For the latter species, the entire U.S. Atlantic coast represents a Biologically Important Area, and Virginia waters are an important part of their migratory corridor as they move between summer feeding grounds in the north and calving grounds in the south (LaBrecque et al. 2015). In terms of conservation status, fin whales are listed as vulnerable on the IUCN red list, sei and blue whales as endangered, and NARW as critically endangered. VOWA is located at the north end of the Currituck Sound Protraction block (see Salisbury et al. 2018 for detailed information on size, location, and physical characteristics of the VOWA).

Previous monitoring of density and distribution of baleen whales off Virginia show pronounced inter- and intra-annual variability in timing of presence of the whales at the study site (Salisbury et al. 2018, Burt and Thomas, 2021). For example, both acoustic and aerial monitoring show a peak in presence of NARW in winter months, but depending on year, this peak may occur in January, February, or March (Salisbury et al. 2018,). Similar inter annual variability in timing of presence has been reported along other parts of the Atlantic coast migratory route of whales (Tetra Tech and LGL, 2020). Baseline data collection is, therefore, essential to build an understanding of inter- and intra- year variability of species presence in an area. It will allow for inference to be drawn from any of the observed changes and whether they are a result of oceanographic, ecological, or climatological factors, or due to anthropogenic effects such as offshore wind farm construction.

There is a large variation in reported measured or modelled PAM detection ranges (see Section 2.5 for details) and acoustic behavior (cue rate) of the six baleen whale species. Fin whales have the largest reported detection range (> 120km (Salisbury et al. 2018, Estabrook et al. 2021)) and minke whales and NARW the smallest, although there is large variation for these species reported between different areas of the study site (Salisbury et al. 2018). The detection ranges for the remaining species have been reported to be ~4-25 km (Salisbury et al. 2018, Estabrook et al. 2021, Palmer et al. 2022). As the detection range for each species may vary over space, we use estimations of detection range closest to or within the study area; we also require more information than just mean or maximum detection range, and use only reports that provide at least three percentiles for detection (e.g., ranges at which 5, 50 and 95% of calls are estimated to be detected, see details in Methods section) (Salisbury et al. 2018, Estabrook et al. 2021). Studies estimating detection ranges from other areas or not providing percentiles (e.g. Gervaise et al. 2019, Kowarski et al. 2020, Palmer et al. 2022) were therefore not used here. There is also a large variation in cue production rate between species, rates at which individuals produce acoustic signals (Fregosi et al. 2022). For cue production rate, fin whales have been estimated to produce on average 45 cues/h (Stimpert et al. 2015) and minke whales and NARW only ~ 6 cues/h (Parks et al. 2011, Martin et al. 2013) however information for fin, sei and minke whales come from the Pacific region. Apart from NARW, no information on call rate from East Coast of the U.S. is available for minke, sei, fin and blue whales. No information on rates from the study site is available for any of the modelled species

After consulting with BOEM and RWSC, four baleen species were chosen for analysis: NARW, fin, minke and sei whales. Fin and minke whales are the most abundant species and are at the two extremes in terms of detection ranges and cue production rates. NARW and sei whales are less abundant species, but NARW are of highest conservation concern.

2.3 Potential effects of offshore wind development on behavior of cetaceans

The majority of studies describing the effects of construction and operation of wind farms on cetaceans come from Europe, where small odontocetes (dolphins and porpoise mainly) have been the main focus. Displacement effects, at ranges up to 10-26 km from the footprint of offshore wind farms during construction have been reported (Dähne et al. 2013, Brandt et al. 2016, Dähne et al. 2017, Brandt et al. 2018, Graham et al. 2019, Benhemma-Le Gall et al. 2021, Graham et al. 2023). Together, these studies indicate that the distance and duration of avoidance may be related to habitat quality, received noise level (which in turn is a function of source level and sound propagation conditions), hearing characteristics of the studied species, distance to the noise source, duration of exposure, level and type of mitigation and presence of other noise sources like construction vessels. There is variation among studies in the reported duration of effect, from hours (Dähne et al. 2017) and days (Brandt et al. 2018) to years (Teimann and Carstensen 2012), which also suggest that operation of offshore wind farms may affect cetacean behavior, including distribution. These large differences are hypothesized to be mainly driven by the habitat quality and the noise characteristics at the area during construction and operation.

The acoustic frequency range used by baleen whales to communicate, and presumed hearing sensitivity, overlaps more with frequencies produced by pile driving (Pyć et al. (2018), see Fig. 3 in Van Parijs et al. (2021)) than in the case of odontocetes. Little is known about the effect of pile-driving noise on baleen whales, and we are not aware of any empirical data on behavioral response to pile driving in this group. While the effects of pile driving on odontocetes may not be directly transferable to baleen whales some informed estimates can be made based on understanding of their hearing, response to other low-frequency noise sources (e.g., seismic or vessel) and models or auditory masking (Hatch et al. 2012, Erbe et al. 2016).

We suggest seven hypotheses, listed in Table 2, on how whales of the study species may respond to construction and operation of wind farms, and how this may affect acoustic detections We also outline what additional data (in addition to PAM) or analyses maybe required to distinguish between the hypotheses, although this is not the focus of the current report. For consistency with regional analysis (Chudzinska and Thomas, 2023), the numbering of hypotheses is kept the same between these two reports. H6 did not apply to VOWA in the regional study, hence it is omitted from this report, but the numbering of the remaining hypotheses remains the same. Broadly, the hypotheses can be grouped into four categories, as follows. I) Construction and operation of wind farms has no effect on whale distribution and acoustic behavior (H1 and H8). In H1 there is no systematic change in whale density or distribution over time; in H8 there is a region-wide decline, but unrelated to wind farms. II) Construction of wind farms causes a change in acoustic behavior (cue production rate) (H2). The change will be strongest close to the construction location and decline with increasing range. III) Construction of wind farms causes temporary displacement of whales (H3-H5). Again, the effect will be strongest close to the construction location; the hypotheses differ in whether the direction of displacement is dependent on habitat preference or other construction-related activities (such as shipping). IV) Operation of wind farms causes long-term change in whale distribution (H7).

In cases where we hypothesize that construction of wind farms results in behavioral response (either displacement or change in cue rate), to calculate the number of animals responding, we multiply the number of animals at each distance from the source (here derived from animal density maps such as Roberts et al. (2022)) by the probability that each animals will respond, obtained from an assumed dose-response function (Tyack and Thomas 2019). If the dose-response function uses received noise level rather than distance as the dose metric then the range-specific received level must be estimated (e.g., using a sound propagation model). To our knowledge, no empirical dose-response function has been derived for a baleen whale species' response to construction of offshore wind farms. However, an interim function for NARW was obtained using expert elicitation in December 2022 by a team from the University of St Andrews as part of a BOEM-funded project "Assessing population effects of offshore wind development on North Atlantic right whales". A distribution of functions was elicited, representing scientific uncertainty in the dose-response function; we use the first and third quartiles from this distribution (less and more sensitive, respectively, see Methods for details) and assume these apply to all species.

We note that baleen whale dose-response functions have been developed for other noise sources (e.g. Sivle et al. 2015, Dunlop et al. 2017, Dunlop et al. 2018, Dunlop et al. 2020) and this information was included in the input to the expert elicitation. However, in many cases the data have been summarized in the form of a received level at which probability of response is a given value, typically 50%. For example, 50% probability of inducing behavioral responses may be expected at a received sound pressure level (SPL) of 140-160 dB re 1 µPa based on the responses of migrating grey whales to airguns (Malme et al. 1984, Wood et al. 2012, see also summary in Pyć et al. 2018). The use of 50% thresholds to estimate zones of impact has been criticized by Tyack and Thomas

(2019), who showed that it can lead to a large underestimation of the number of animals responding. They advocate the use of a dose-response function or, if a single threshold value is preferred, then it should be based on the concept of an effective response range (see paper for details). Nevertheless, 50% thresholds may provide a helpful frame of reference. Based on a literature review and sound propagation modeling, Pyć et al. (2018) predicted 50% probability of response for baleen whale species in the Vineyard Wind 1 Area VOWA to be at 2-7 km from turbine installation, depending on exposure scenario, foundation type and attenuation level, and assuming a step function with certain response at a frequency-weighted sound pressure level of 140 dB (see Tables 10 and 13 in Pyć et al. 2018). Depending on the scenario (number of pilings per day, attenuation level and number of piles), no more than 1 to 40 individuals (depending on the baleen whale species) would be estimated to experience sound levels above the threshold criteria, which is no more than 4% of the population of a given species (Pyć et al. 2018). The recommended distance for monitoring and mitigation during Vineyard Wind 1 construction for North Atlantic right whales (NARW) was 10 km (Pyć et al. 2018).

For hypotheses where the response is displacement only, once the number of displaced animals is estimated, we will also consider the location the displaced animals move to. Direct evidence of this is (to our knowledge) lacking, so we considered a range of hypotheses. We set an upper limit on the displacement distance and within that distributed responding animals uniformly ("symmetrical displacement", H3), or according to the underlying habitat model (H4) or according to the habitat model but also accounting for patterns of vessel traffic (H5).

Construction of wind farms is not limited only to foundation installation: a range of activities happen before and after this installation, such as cable or turbine installation. In this project we focus only on periods when foundations are installed (referred to as 'construction' throughout the text, Table 1) as we are not aware of any data on the effect of other activities on baleen whales. We also consider hypotheses based around long-term operation. Piles can be installed in a variety of ways: impact hammering, vibrating piling or drilling and a mix of the methods can be used for a given wind farm. The presented study focuses on impact pile driving as this is the loudest of the methods.

Table 1. Timing of foundation installation for the studied wind energy areas and assumptions on corresponding baseline monitoring.

Wind energy area	Foundation installation	Baseline monitoring	Size of the large monitoring area (km)
VOWA	May – October 2024 May – October 2025 May - September 2026	May - October	80

2.4 Challenges of using PAMs to detect changes in behavior and distribution of baleen whales and the effect of these challenges on quantifying power.

Responses to anthropogenic disturbance, including noise, have been quantified using Behavioral Response Studies (BRSs). Such studies either involve controlled exposure experiments, where animals are exposed to a controlled level of a potential stressor, or in opportunistic contexts where exposure and concurrent activities are monitored in a strategic manner (see Harris et al. 2018 for a review on the context of disturbance from naval sonar). A variety of animal observation techniques have been employed, some focused on measuring the response of individual animals (e.g., animal-borne tags, ship-based or aerial focal follows, acoustic tracking arrays) and others on changes in occurrence or density of animals at the population level (e.g., visual surveys, PAM). Often, the approaches are complementary—for example, Tyack (2011) used results from both a controlled exposure experiment on a small number of individuals and a large-scale opportunistic population-level PAM study to infer an avoidance response by Blainville's beaked whales to naval sonar exercises.

We here focus on the use of a fixed array of archival PAM sensors to perform population-level inference on behavioral response. We note that other types of PAM system exist (e.g., glider-mounted, towed or buoyed; real-time vs archival) and can be used for mitigation and monitoring in the context of wind energy development—see, e.g., review by Van Parijs et al. (2021). Archival data is post-processed once the recordings are recovered, and detections of vocalizations from the target species are identified. If each sensor is analyzed independently then the resulting data are a count of detections per unit of monitoring time. If sensors are close enough together that the same vocalizations can be detected on multiple sensors, it may in addition be possible to localize animals making the sounds, allowing a finer level of inference about any possible behavioral change. Further, under some circumstances it may be possible to track individual vocalizing animals, allowing for individual level responses to be studied (see e.g. Durbach et al. 2021).

Returning to the independent sensor analysis, changes in detection rate associated with wind farm construction or operation could be caused by several factors: changes in detectability of vocalizations (e.g., via masking of the sounds), change in the frequency of false positive detections, change in animal vocalization behavior or changes in local animal distribution. Additional studies may be required to exclude the first two factors, although any masking should only occur while anthropogenic sounds are being made while biological changes may be more long term. Distinguishing between changes in vocalization behavior and changes in animal distribution also requires additional data collection—but both constitute a behavioral response that are of interest to detect. However, detecting such a change on a single sensor concurrent with wind farm construction or operation is not enough to infer the wind farm has caused the change because any change may be part of some larger-scale process that is independent of the wind farm activities. This is why it is crucial to monitor at a range of distances from the wind farm site: a change in detection rates close to wind farm activities that does not occur at further distances is much stronger evidence that the change is caused by the activities. Detecting an interaction between change in detection rate and distance from wind farm, while accounting for other factors such as monthly changes in detection rate in the baseline data, is the core of the statistical test used in the power analysis.

The density of the four baleen species chosen for this analysis is generally low. In particular, for NARW, construction of some of the wind farms is scheduled to happen in the months when they are at seasonally low density (Pyć et al. 2018). As a result, the number of whales responding to construction (or operation) of wind farms is also going to be low. For a given statistical test and chosen αlevel, statistical power increases with increasing effect size (i.e., true magnitude of change) and sample size, and decreases with increasing variability. For the statistical test evaluated here, effect size is related to the proportion of whales responding within the area monitored by the PAM devices, rather than the absolute number responding. Hence the low number responding will not necessarily cause low power if the proportion responding is consistent. However, low numbers may lead to high variability over time and space, which may result in low power. In addition, in some cases there may be no whales at all in the area, giving an effect size of zero.

Variability in baseline density in space and time will tend to decrease power unless it is dealt with as part of the analysis. The highest density of the studied species generally is along the continental shelf and there is therefore a natural gradient in whale density from VOWA towards the continental shelf. This gradient is accounted for in the analysis method used (which models acoustic encounter rate under baseline conditions as a function of distance from each wind farm), and so should not cause either decreased power or elevated false positive rate. Variability in time is partly accounted for by including month as a factor in the model.

Another issue that could potentially affect power is that some species (fin whale in particular) can be detected at a considerable distance from the acoustic sensors. In this case, if response to wind farms only occurs relatively close to the wind farm footprint, then sensors placed near the wind farm boundary will detect a mix of responding and non-responding animals. This will dilute the measured effect size and reduce power. We return to this in the Discussion.

As well as quantifying statistical power given a specified change, it is important to quantify the false- positive rate—i.e., the probability of detecting a statistically significant effect when none exists. We followed convention in using a threshold for statistical significance of p = 0.05, which should result in a 5% false positive rate. However, for complex statistical tests like those used here, the false positive rate can be different from the nominal value. In addition, it may be that random variation in animal density over time produces a change in distribution with respect to distance from a wind farm and so triggers a positive significance test. To evaluate false positive rate, we include a hypothesis of no change (H1) in our test suite.

Table 2. Summary of potential drivers, acoustic effect of these drivers, additional to PAM data required to study the driver and additional reading supporting the driver related to the construction and operation of offshore wind farms. For consistency with regional analysis (Chudzinska and Thomas, 2023), the numbering of hypotheses is kept the same between these two reports. H6 did apply to VOWA in the regional study, hence it is omitted from this report, but the numbering of the remaining hypotheses remains the same.

#	Drivers and hypothesis	Effect observed in PAM data	Additional data or analysis required to distinguish from other hypotheses	Additional reading	Tested scenarios and methods
1	Construction and operation activities have no effect on baleen whale distribution or behavior.	There are no changes in acoustic encounter rates with distance from wind farm between construction/operation versus the same area before construction (referred to hereafter as "over time").	Visual surveys and tagging to confirm no changes in behavior such as foraging or group behavior. Studies could additionally undertake analysis of stress hormones to understand whether there is a physiological response even if no behavioral response is observed.	Bailey et al. (2010)	Simulation: no re-distribution of whales or changes in cue rate will be generated for all the wind farms. Analysis: examine the relationship between acoustic encounter rate and distance to the wind farm under baseline and under construction/operation. Under this hypothesis, the relationship should not differ. The proportion of times a significant difference is found is an estimate of the false positive rate, i.e., the probability of detecting a change if the change is not present.
2	Construction activities have no effect on baleen whale distribution but have an effect	We would observe a change in acoustic encounter rate over time with distance from wind farm (compared with the baseline pattern), where the	Additional data, e.g., a visual survey or tagging study, would be needed to confirm that there is no displacement.	Benhemma-Le Gall et al. (2021)	Simulation: the number of responding whales will be calculated based on a dose-response function. We then simulate I) 100% decrease in

	on their behavior related to cue production.	change is driven by change in cue production or cue detection rate (see next column) and not by changes in distribution. As pile driving mainly occurs during the day (Heaney et al. 2020), an increase in cue rate may be observed at night as a function of time since last pile-driving event (although diurnal patterns in baseline would need to be considered). A decrease in encounter rate of acoustic detections could also result from decrease in cue detection rate due to masking (Erbe et al. 2016, Cholewiak et al. 2018). Comparing data collected during actual piling with in-between piling is needed to distinguish between these two options.			cue rate of the responding whales (referred to as "H2_100"); II) 50% decrease in the cue rate of the responding whales simulating either masking or changes in cue production occurring only during the day or a partial response (referred to as "H2_50"). No whale re-distribution is assumed for this hypothesis. Simulation to be done for all wind farms. Analysis: examine the relationship between acoustic encounter rate and distance to the wind farm under baseline and under construction. The proportion of times a significant difference is found is an estimate of power.
3	Construction activities cause displacement of whales away from the construction locations, with displacement occurring equally in all directions. There is no change in cue production of individual whales.	We would observe a decrease in acoustic encounter rate over time with distance from wind farm, compared to the baseline pattern.	Visual survey or tagging study to confirm that this is animal displacement and not a change in cue detection/production rate. Animals may increase their foraging effort when displaced from the wind farm to	Kraus et al. (2019) Pyć et al. (2018) Benhemma-Le Gall et al. (2021) Sivle et al. (2016)	Simulation: the number of responding whales will be calculated based on dose-response functions. We then simulate displacement of responding whales from the footprint of the wind farm with

			compensate for the lost foraging time when moving away from the site (Benhemma-Le Gall et al. 2021). In such a case we may expect an increase in foraging activity further away from the construction which could be confirmed by tagging. Note that this hypothesis is not considered very likely to be correct, as whales can be displaced to areas where they were previously not known to occur.		equal probability of displacement in all directions. Analysis: as for H2.
4	Construction activities cause displacement of whales away from activity locations preferentially towards higher density areas outside wind farm.	We would observe a decrease in acoustic encounter rate over time with distance from wind farm (compared to baseline pattern) with a corresponding increase being proportional to the observed baseline densities in the areas around the wind farm.	Same as H3.	Davis et al. (2020) Salisbury et al. (2018) Roberts et al. (2016) BOEM and NOAA (2022) Rolland et al. (2016) Ellison et al. (2012)	Simulation: the number of responding whales calculated as for H3. Displacement locations will be proportional to baseline density. Analysis: as for H2. Note this means we will test for displacement as a function of distance. In this report, we do not additionally seek to distinguish between H3 and H4 but this could be done by fitting an acoustic encounter density surface to the data and looking for an increase in detections proportional to baseline density (or

					proportional to baseline animal density as estimated by, e.g., Roberts et al. 2016).
5	Construction activities cause displacement of whales away from activity locations preferentially towards higher density areas outside of the wind farm, but this preference is lessened by additional anthropogenic activities such as shipping that are associated with construction but take place away from the piling locations.	We would observe a decrease in acoustic encounter rate over time with distance from wind farm, compared to baseline pattern. The corresponding increase in density at nearby sites would be a function of baseline density and additional anthropogenic activity.	Same as H3, plus data on other anthropogenic activity: vessel traffic, etc.	www.northeastoceandata.org https://globalfishingwatch.org https://marinecadastre.gov/accessais/	Simulation: the number of responding whales calculated as for H3. Displacement locations will be proportional to baseline density, modified according to an anthropogenic pressure map. This map will be generated based on AIS data scaled from 0 to 1 with 0 meaning high anthropogenic pressure. Analysis: as for H2. Note, as for H4, this means we do not seek to distinguish between H3, H4 or this hypothesis.
7	Displacement and alteration of whale behavior during construction leads to long-term displacement of whales. Such long-term displacement may be related to the noise of the operational turbines, anthropogenic activity related to maintenance or changes in	The displacement would not only occur during construction but also several months after construction stops and operation starts.	Aerial surveys to confirm displacement and distinguish from changes in cue production rate. Monitoring of prey composition and distribution within and outside wind farm	Though operational turbine noise levels are generally low/close to ambient, larger turbines may lead to behavioral response in low-frequency specialists such as baleen whales within ~1.4 km of turbines (Teilmann and	Simulation: to calculate the number of responding whales, density within the wind farm footprint was set to zero; displaced whales were redistributed according to baseline density (as for H4).

	prey distribution/behavior at the wind farm.		could determine if prey changes are driving effect.	Carstensen 2012, Stöber and Thomsen 2021).	Displacement took place in all months. Analysis: as for H4.
8	Displacement or decline of whales is not linked to activating related to construction or operation of wind farms but it is a large-scale phenomenon related to global changes in environment.	The decrease or displacement of whales would occur over the entire study region.	If PAM is to be used to detect a long-term decline in density/abundance within the region, then additional data collection is needed to determine if detectability and/or cue rate changes over time. Alternative population monitoring such as visual surveys may be used.	Global shift or decline of whales has been observed along the U.S. East Coast especially related to NARW and changes in distribution of their primary food (Ramp et al. 2015, Davis et al. 2017, Meyer-Gutbrod et al. 2018).	Simulation: a region-wide decrease in acoustic encounter rate was simulated that was equal in magnitude to the average decline under H3 within 20 km of a wind farm during construction. Analysis: unlike previous hypotheses, the focus here is on whether the effect of year is statistically significant (and whether the interaction between distance and phase is not).

3 Methods

3.1 Overview

For the selected study species (NARW, fin, minke and sei whales), the power analysis was conducted in two steps.

I) Local, at the level of VOWA, based on the PAM design of Van Parijs et al. and examining all hypotheses (H1-H8).

II) Local, as with step I, but using alternative PAM design(s) if step I showed low power to detect change.

An overview of these steps is given below. All analyses were undertaken using the R statistical software (R Core Team 2022).

For step I, to evaluate hypothesized changes within and around each wind farm, we did not use the entire regional PAM network. Rather, we sought to understand how power to detect change is affected by the size of the monitoring area and by the PAM density. We defined two sizes of monitoring area: I) the wind farm footprint plus a 20 km buffer (the "small monitoring area"), and II) the wind farm footprint plus a buffer large enough to cover the continental shelf ("large monitoring area") (Figure 1). To examine PAM density, we calculated power using I) both the large (40 x 40 km spacing) PAM grid and the small (20 x 20 km spacing) PAM grid suggested by Van Parijs et al. and II) just the small PAM grid. If power is high and false positive detection rate is low with just the small PAM grid, the monitoring scheme will be considerably less expensive. The grids are shown in Figure 1.

We evaluated all seven hypotheses as shown in Table 2. For each species and hypothesis, the power analysis proceeded through a series of sub-steps, shown in Figure 2. We first simulated a set of 500 random replicate density surfaces on a 5 x 5 km grid for relevant months and years from habitat-based density models (sub-step 1). For hypotheses involving construction (H2-H5), we defined the sound source location for each month as a subset of the VOWA (sub-step 2). We then calculated the number of responding whales during construction or operation according to the hypothesis (sub-step 3) and, for hypotheses that involved displacement (H3-H8) we generated the appropriate spatial change in whale density (sub-step 4). We next transformed the whale densities per grid cell into a number of vocalizations ("cues") per month (sub-step 5) and, for the given PAM design, to the number of cues detected on each PAM sensor (sub-step 6). These detection numbers formed the input data for a statistical analysis to determine whether the hypothesized effect was detected or not. Analysis involves comparing one year of baseline data (just the months of construction for the construction-related hypotheses) with the same months of data from a construction or operation period. This process was repeated 500 times, once for each replicate simulated dataset, and the proportion of statistically significant results was used as the estimate of power (or false positive rate for H1 since this hypothesis is that there is no change related to construction or operation) (sub-step 7). Note that while sub-step 1 involves stochasticity, all other sub-steps are deterministic.

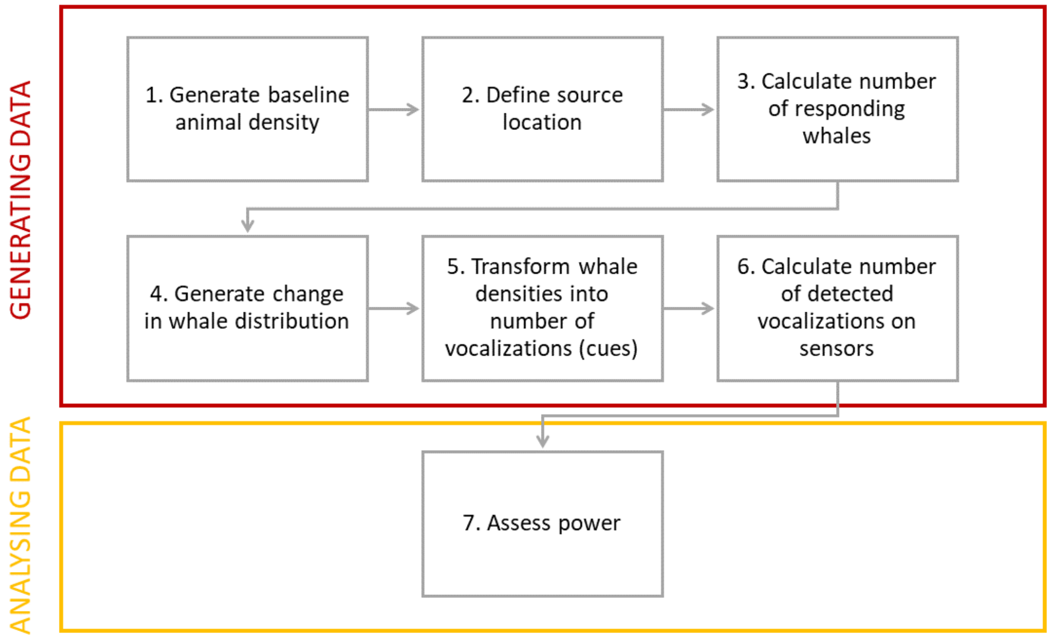


Figure 2. Flow chart of seven sub-steps of the analysis.

For construction-related hypotheses, we assumed 1 year of baseline monitoring and 1 year of construction, just in the months where construction activities were planned (Table 1). For operation-related hypotheses (H7) , we assumed 1 year of baseline monitoring and 1 full year of operation.

After completing step I (i.e., power analysis based on the Van Parijs et al. design) we found that power to detect hypotheses relating to construction and operation was not above the target level of 80% for any scenario. Therefore, in step II, we created two alternative PAM designs to replace the 20 x 20 km small grid of Van Parijs et al. in the vicinity of each wind farm: I) a 10 x 10 km grid and II) a linear array of PAM stations in a T-configuration. We repeated the same sub-steps to evaluate power.

3.2 Generating baseline animal density

Baseline animal density surfaces were generated using habitat-based marine mammal density models for the U.S. Atlantic (https://seamap.env.duke.edu/models/Duke/EC/) produced by the Duke University in collaboration with Marine Geospatial Ecology Laboratory (MGEL) of Duke University. These models were based on analysis of aerial and shipboard visual line transect survey data from multiple sources—see Roberts et al. (2016) and the above web site for details. For each species, we used the most recent version of the models (see

Table 3) available at the time we undertook our analysis; for minke whale there was a separate model for summer (April-October) and winter (November-March). The models predict density of the studied species with one-month temporal resolution and 5 x 5 km spatial resolution. Two types of models were produced by MGEL, depending on the species (and, for minke whale, the season): for fin whale and minke whale (winter) the models used climatological covariates whose value changed by month but were averaged across years, while for fin whale, NARW and minke whale (summer) the models used contemporaneous dynamic covariates whose value changed by month and year. In the latter case, we generated density surfaces using the most recent 5 years in the model (see

Table 3). The spatial coverage of the MGEL models varied between species (Figure 3A): models for sei whales and minke whales in winter did not extend much further southwest than VOWA.

The density models include two sources of uncertainty: I) uncertainty in the predicted average density surface given covariate values, represented by variances and covariances on model parameter values, and II) variability in the number of animals present in a given grid cell on any day, represented by a statistical distribution on predicted per-cell numbers (the MGEL models used a Tweedie distribution). To incorporate the first source of uncertainty, we used parametric bootstrap resampling (sampling model parameters from a multivariate normal distribution) to generate 500 realizations of the density surfaces for each species and month. For models that use contemporaneous dynamic covariates, we sampled 100 realizations from each year, making 500 in total. To incorporate the second source of uncertainty, for each realization and 5 x 5 grid cell, we generated a random value from the Tweedie distribution with mean equal to the parametric bootstrap value and scale equal to the value from the MGEL model divided by 30. The reason to divide by 30 was to account for the fact that we want an average density per grid cell per month while the MGEL models use day as the sampling unit.

The above procedure occasionally generated unrealistically high-density values, and we hence implemented an outlier removal procedure. We removed densities greater than an outlier threshold defined as $Q3 + 1.5 * IQR$ where *Q3* is the third quartile and *IQR* the interquartile range of the densities. The removed predictions were then replaced by new predictions generated by bootstrapping and generating values from Tweedie distribution, so the number of realization maps was always equal to 500. For all four species, no more than 20 realizations per species were above the defined threshold.

An example generated density map is shown in Figure 3B, and further examples for each species, wind farm and month are given in Appendix B.

Table 3. Version of the density surface distribution models used for each of the studied species, response variable distribution of each model, prediction years used in the generation of model realizations ("-" indicates the model did not contain year-referenced covariates) and references to the most recent model description.

Species	Model version	Response variable distribution and its parameters	Prediction years	Reference for the most recent model description
Fin whale	v12	Tweedie (p = 1.14, scale = 7.20)	-	Roberts et al. (2022)
NARW	v12	Tweedie (p = 1.23, scale = 13.45)	2016-2020	Roberts and Yack (2022b)
Minke whale	v10	Summer: Tweedie (p = 1.13, scale = 6.97) Winter: Tweedie (p = 1.06, scale = 5.96)	2015-2019 -	Roberts and Yack (2022a)
Sei whale	v10	Tweedie (p = 1.22, scale = 14.36)	2016-2020	Roberts and Yack (2022c)

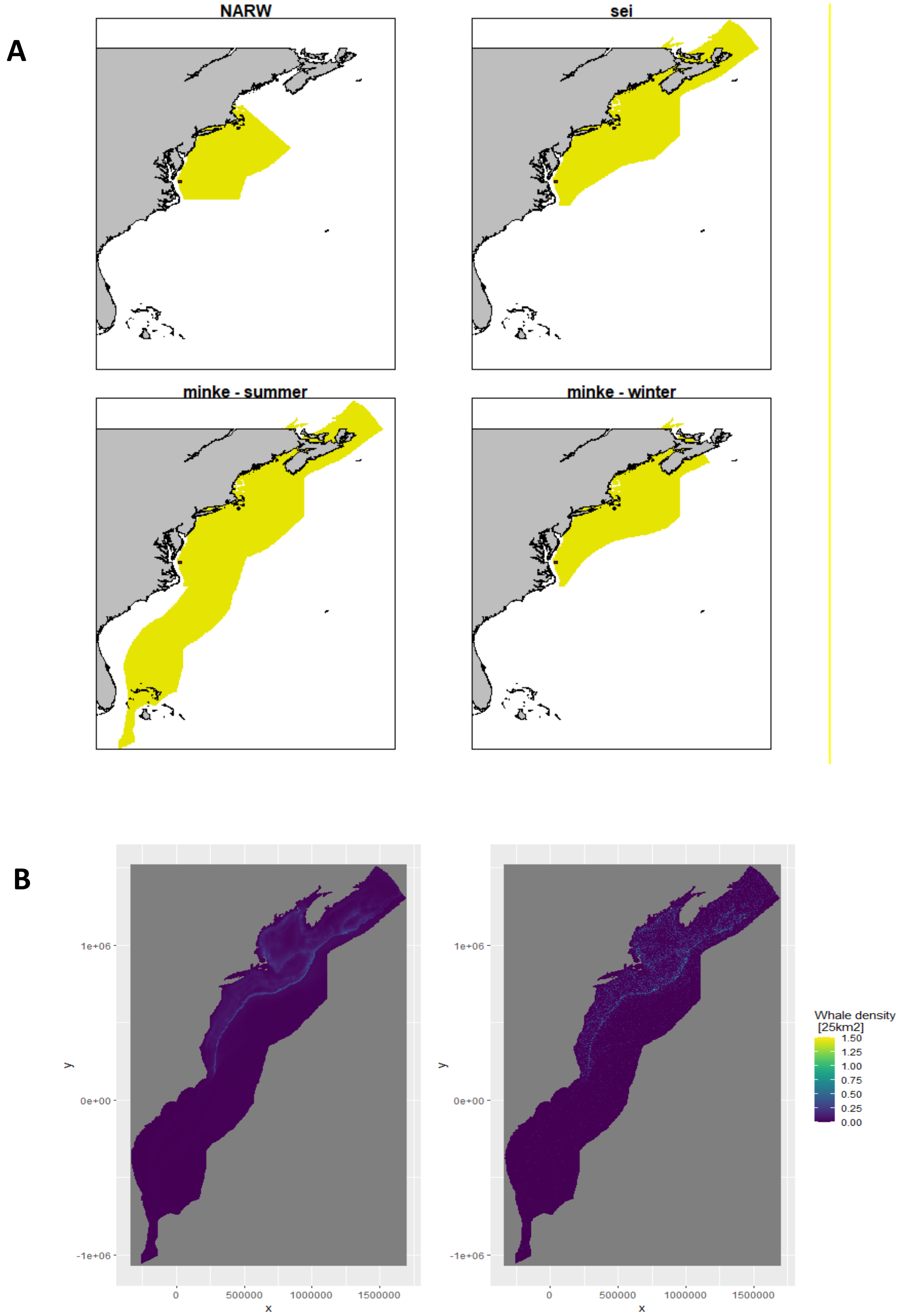


Figure 3. A: model extent for NARW, sei whale and minke whale. The black squares show location of the northernmost wind farm (VYWA) and the southernmost wind farm (VOWA) for reference. B: Density of fin whales in January an example of one out of 500 realization maps generated by parametric bootstrap (left) and then in addition accounting for daily variability (right).

3.3 Defining source location

This sub-step is relevant only for hypotheses involving wind farm construction (H2-H5, Table 2). As detailed location and order of construction of each turbine was not available for each wind farm, we assumed that at each month of the construction, piling took place within one section of the wind farm at a time. To define sound source location, the wind farm was, therefore, divided into as many approximately similarly sized polygons as months of piling (Table 1, see, e.g., Figure 4). The temporal ordering of the sections was chosen at random.

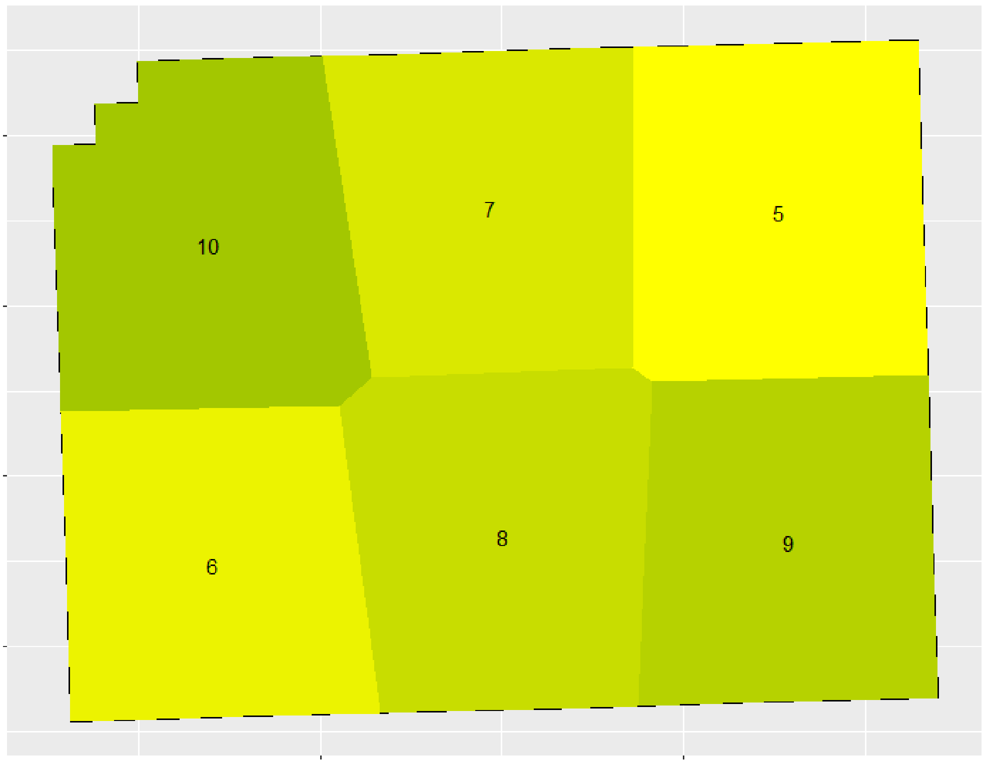


Figure 4. Division of the VOWA footprint based on location of piling. The piling takes place over 6 months and the numbers in each polygon indicate the piling month (e.g., 5 is May, 6 is June, etc.).

3.4 Calculating number of simulated responding whales

The number of responding whales for the hypothesis related to construction (H2-H5, Table 2) was calculated based on dose-response functions and uncertainty around them established for NARW during an expert elicitation conducted in December 2022 (Booth et al. 2023). The functions elicited related to the received level of sound from pile driving at which a foraging NARW would switch from a foraging to non-foraging state for at least the rest of the pile driving on that day. We chose two functions, one below the elicited mean dose-response function equal to 1st quartile (i.e., less sensitive) and one above the elicited mean equal to 3rd quartile (i.e., more sensitive). We refer to these as DR1 and DR2 respectively. The same two functions were used for all calculations. The dose response functions established in expert elicitation process were based on daily responses. Due to lack of information on cumulative response, we assumed that dose-response by month was the same as that based on daily responses. This would be the case, for example, if whales return to the baseline distribution at the end of each day after displacement.

We assumed impulsive source level of 200 dB re 1 µPa, including 10 dB broadband sound attenuation (following Pyć et al. (2018)), and transmission loss (TL) of:

$$TL = 15\log_{10}(R \times 1000) + \alpha R$$

where R is distance (km) from the source and α attenuation coefficient (dB km^{-1}). We chose α = 1.2, which was the mean value for areas of ~40 m depth as presented in Heaney et al. (2020) (see Table 3-1). We also assumed that at the distance where probability of response <= 1%, this probability was 0.

These dose-response functions resulted in a very sharp decline in probability of response within the first 2 km from the source and reaching 1% at 18 km for DR1 (green line, Figure 5) and a more gradual decline with probability of response 1% at 30 km for DR2 (orange line, Figure 5).

To calculate the number of responding whales for each 5 x 5 km density grid cell, the number of animals in that grid cell was multiplied by the probability of response from the dose-response function. Distances were taken from the center of each density grid cell to the closest edge of the sound source polygon. For grid cells with centers inside the sound source polygon, all whales were assumed to respond (see Discussion for effect of this assumption). An example showing probability of response under DR1 and DR2 for five months of piling at VYWA is shown in Figure 6.

For hypothesis related to operation (H7), we assumed that all whales within the footprint of the wind farm respond. In practice that meant altering the density of 5 x 5 km density grid cells whose center fell within the wind farm footprint, and not altering density for cells with center outside the footprint. An example for is shown in Figure 7.

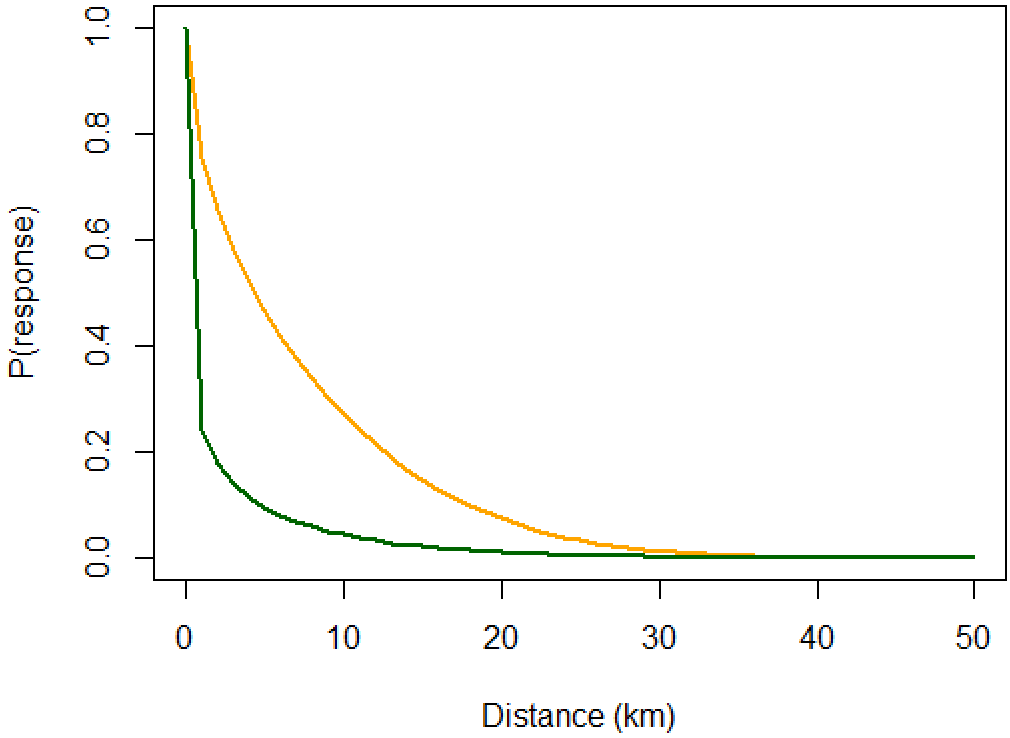


Figure 5. Probability of response with distance from the source for the two chosen dose-response functions assuming 200 dB source level. Green line is DR1, orange line is DR2.

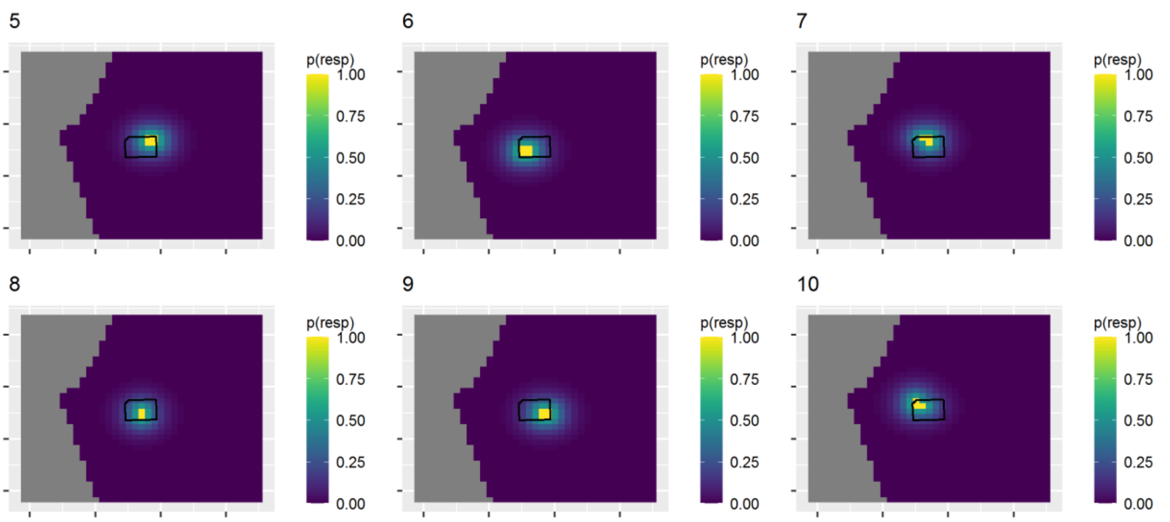


Figure 6. First six panels show the probability of response (p(resp)) for each of the piling months denoted by the title of each panel for based on DR1 for VOWA. The lower six panels were based on DR2. The difference between months was the location of the sound source.

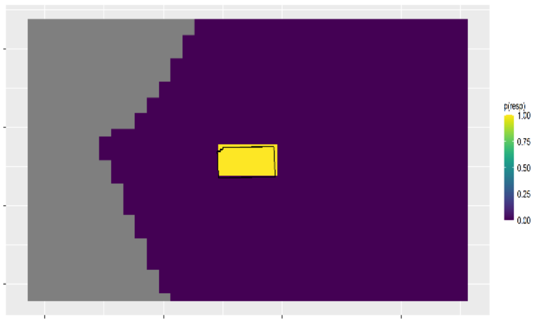


Figure 7. The probability of response (p(resp)) for one month of operation of VYWA. Probability of response was assumed to be 1 inside the wind farm footprint and 0 outside.

3.5 Generating change in simulated whale distribution and density

Hypotheses H3-H5 relate to redistribution due to construction, while H7 to redistribution due to operation. The locations of redistributed whales differed between hypotheses (Table 2). For each of the hypotheses, change was generated separately for each construction/operation month. We therefore assumed that changes in behavior of whales in one month was independent from changes in the previous or following months. Animals were assumed to redistribute within 100 km of the wind farm (but only within habitat covered by the density models – e.g., not onto land).

Under H3 (displacement due to construction, equally in all directions), responding animals were redistributed uniformly within the 100 km radius, excluding the wind farm footprint. Under H4 (displacement due to construction, with preference for higher density locations), animals were redistributed in proportion to the underlying simulated density surface, so areas of high density received proportionally more responding animals than those with low density, under the assumption that they will be more attractive to displaced animals.

Hypothesis H5 is that higher density locations receive more displaced animals, but that this is tempered by increased anthropogenic activities that are associated with construction, but outside of the wind farm area. As an approximation of this, we assumed that current shipping traffic is an indication of the locations of possible future increased traffic. We calculated mean monthly AIS per 5 x 5 km grid cell for each monitoring month based on mean of four days for each month (1, 10, 15 and 20th day of each month) from 2019 (source https://marinecadastre.gov/ais/). We then log transformed the number of vessels per grid cell and scaled to values between 0 and 1. We redistributed animals in proportion to the underlying simulated density surface multiplied by the scaled metric of vessel traffic.

Under H7 (operating wind farms displace whales) we allocated displaced whales according to the underlying simulated density surface (as with H4).

Hypothesis H8 involves a global decline in whale density, not related to wind farm construction or operation. To keep the size of the decline comparable to that in the vicinity of the wind farms under other hypotheses, we divided density by the ratio of the number of whales remaining within 20 km of each wind farm under H3 (symmetric displacement) DR2 divided by the number under H1 (no response). By including H8 in the analysis, it may be possible to understand whether a 20 km monitoring size allows for a distinction between a larger-scale, global decline and decline in detected cues due to activities at the WEA.

3.6 Transforming simulated whale densities into number of vocalizations (cues)

The above sub-steps generated animal density per grid cell and month under baseline conditions and under the hypotheses of change. These were transformed into number of vocalizations (cues) produced per grid cell and month by multiplying by an assumed cue production rate (Table 4), which was (for lack of better information) assumed to be constant over space and time per species.

Hypothesis H2 involved a decrease in cue production rate for responding animals. Two alternatives were used: I) 100% decrease in cue rate of responding whales (referred to as "H2_100") and II) 50% decrease in cue rate of responding whales ("H2_50"). The second alternative could be seen as a partial decrease in cue rate or a lessened ability to detect cues during piling activities. (Detectability is considered in the next sub-step, but the two explanations were not analyzed separately because their effect on the data at the temporal level of a month would be similar.)

Table 4. Individual cue production rate, type of cues and effective detection range for the four whale species.

Species	**Cue rate (cues/hour/individual), *c***	**Type of cues**	**Reference**	**Effective detection range (km), *v***
Minke whale	6.04	'boing' calls	Martin et al. (2013)	All wind farms: 8.6[3]
Fin whale	45.08	20 Hz pulse	Stimpert et al. (2015)	MAWA: 99.9 EOWA, VYWA: 93.8
Sei whale	10	30–87-Hz upsweeps and downsweeps combined	Baumgartner and Fratantoni (2008)[1] Calderan et al. (2014)[2]	All wind farms: 21.1[4]
NARW	6.2	Upcalls, variable tonal calls, gunshot sounds, and exhalations combined	Parks et al. (2011)	MAWA: 24.9 EOWA, VYWA: 9.2

[1] No individual cue production rate is given in this study. We used the minimum, but no zero values, from Fig. 6

[2] Back calculated based on call duration and inter-call intervals (Table 1 in the cited study).

[3] Percentiles of detection range for minke whale were not estimated in Estabrook et al. (2021) and we used values estimated by Salisbury et al. (2018) for all wind farms.

[4] Percentiles detection range for sei whale was not estimated in Salisbury et al. (2018) and we used values estimated by Estabrook et al. (2021) for all wind farms.

3.7 Calculating number of detected vocalizations on sensors

In this sub-step the cue densities were used, in conjunction with an assumption about cue detectability, to determine the number of detections per sensor. We first describe how sensor locations were determined—i.e., the PAM designs—and then how number of detections was determined given a sensor location.

4.7.1. PAM designs

We calculated the number of detected cues for three PAM designs: Van Parijs et al., and two alternative designs. The two alternative PAM designs have sensors placed closer to the footprints of each wind farm, where the effect of whale distribution and behavior was expected to be largest (Figure 5). The first design, referred to as "10 x 10 km grid", is an equally spaced grid of PAM sensors spaced 10 km from each other (Figure 9A). The original design for small PAM grid by Van Parijs et al. used 20 km spacing. The second design, referred to as "T-design", consists of PAM sensors in three rows with each 'arm' of T situated across the expected gradient of whale density (Figure 9B). Although such gradient differs between the four species both spatially and between months, there is a general trend for these species to have higher densities along the continental shelf. One 'arm' for the T-design is therefore, always pointing towards the continental shelf. The distance between sensors in the T-design is 2, 3, 4, 5, 6 and 8 km with sensors closer to each other in the center of the footprints of each wind farm. For both alternative designs, the numbers and locations of large PAM grid remains the same (Figure 8). The number of PAM sensors suggested by Van Parijs et al. in small PAM grid was 6 (Figure 1). The number of sensors in two alternative designs is 19 for T-design and 21 for 10x10 km grid (Figure 8).

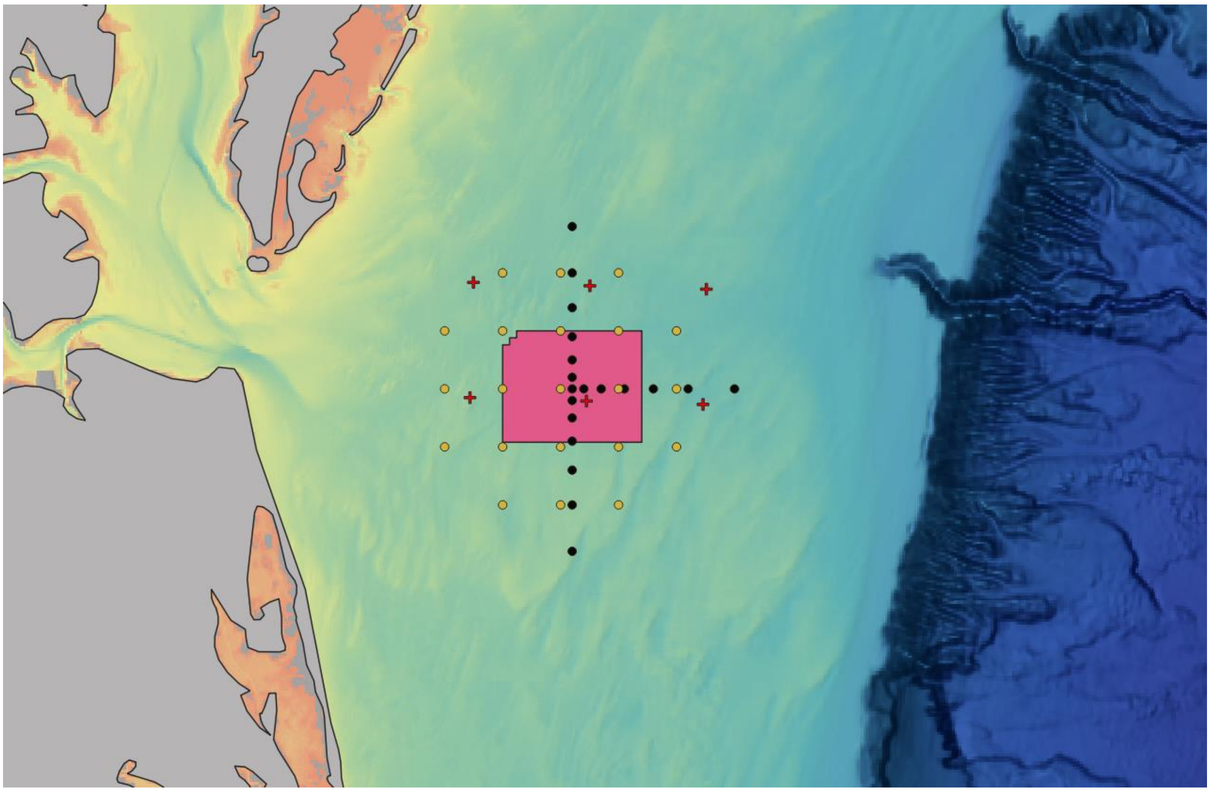


Figure 8. Two alternative modification of small PAM grid suggested by Van Parijs et al. for the studied site: A) 10x10 km grid (yellow dots), B) T-design (black dots). The large PAM grid is shown as red crosses.

4.7.2. Accounting for detectability

To determine the number of cues detected on each sensor given the sensor location one must account for the detection range of the cues. In general, the probability of detecting vocalizations (cues) of a given species decreases with increasing distance between animal and sensor (although many other factors are also involved). We used a concept from the distance sampling literature (see, e.g., Buckland et al. (2001) (Section 3.1.3) and Marques et al. 2009) of the effective detection area (EDA). The EDA is the circular area around a sensor within which as many vocalizations are missed as are detected outside it; hence the EDA can be thought of as a measure of the area monitored by a sensor. The radius of this circle is called the effective detection radius (EDR). EDA or EDR values have not been published for any of the four species for the study site, but empirical measurements of detectability have been collected at some sites and summary statistics published that allow them to be estimated. We used the reported detection ranges at which 5, 50 and 95 percent of vocalizations were estimated to have been detected by Salisbury et al. (2018) (see Table 5.2) We fitted a three parameter detection function (a two-part mixture of half-normal functions, Miller and Thomas (2015), to these three summary statistics using a least-squares algorithm, and given the fitted detection function we estimated EDA as

$$\hat{v} = 2\pi \int_{r=0}^{w} r\hat{g}(r)dr$$

where r is range [km], $\hat{g}(r)$ is the estimated detection probability and w is some suitably large truncation distance so that $\hat{g}(r) \approx 0$ at this range.

To calculate the number of detected vocalizations at a sensor given the EDA and the 5 x 5 km density grid, we summed the number of vocalizations produced in all grid cells within the EDA. For grid cells partially within the EDA, we pro-rated the number of vocalizations by the proportion of the grid cell within the EDA.

3.8 Assessing power to detect response.

We used analysis methods based on the concept a phase-gradient (PG, see Mackenzie et al. (2013), Methratta (2021) for an overview of the methods) design to detect effect of wind farm construction/operation. This involves estimating the relationship between the number of detected cues and distance from wind farm under baseline conditions (which thereby takes account of any pre-existing gradients) and determining whether this relationship is different during construction or operation. In statistical terms this means determining whether there is a significant interaction in acoustic detection rate between distance from wind farm and phase (baseline or construction/operation).

The specific analysis used was a generalized additive model (GAM), implemented using the *gam* function from the *mgcv* package (Wood 2017) in R. We used number of detected cues per each PAM station as the response, main effects for month, distance from wind farm, phase (baseline vs construction or operation) and (for the 5-year operational monitoring) year and an interaction between distance and phase[3]. The terms involving distance were specified as smooths (thin plate regression splines), allowing for flexibility in the relationship between number of detected cues and distance. The response was modelled with an overdispersed Poisson error structure with log-link function; estimation was via restricted maximum likelihood (REML). We recorded whether the interaction between distance and phase was statistically significant at an α-level of 0.05, and also for H8 (global changes over time not related to wind farm construction) where the interaction was not significant, but the main effect of phase was significant.

Power for each species, hypothesis, dose-response curve, and monitoring area and grid was estimated as the proportion of the 500 simulations cases when interaction between phase and distance was significant. If this percentage >= 80% we refer to it as 'high power'. For the remaining cases we refer to as 'low power'.

[3] The interaction term was specified in *gam* as s(distance, by=phase), with phase being an ordered factor with baseline the first level. This meant that the main effect of distance corresponded to the effect of distance under baseline and the interaction term to the effect under construction/operation.

4 Results

All the results are presented in Appendices C, D and E.

4.1 Simulated whale density, acoustic encounter rates and number of responding animals.

Mean number of responding whales under two dose-response functions: DR1 and DR2 (Figure 5) for each construction month (Table 1). Number of responding whales was proportional to baseline densities and was highest for the minke whales, which are the most abundant species and lowest for sei and NARW, which are least abundant species. In most months, number of responding whales was < 1 for all species and maximum of 3.23 for minke whales in May. Numbers of responding whales was lower under DR2 than DR1.

4.2 False positive rates

The simulations under H1 and H8 include no effect of wind farm construction or operation on acoustic detections, and hence any case with a statistically significant interaction between distance from wind farm and phase (before/during construction or operation) is a false positive. Given a threshold for significance of 0.05, we expect a 5% false positive rate, but since underlying density surfaces were sampled at random for each year (representing underlying variation in animal density over time), a higher false positive rate is possible.

The false positive rate under H1 ranged from 5.1-14.9% (Tables D1, E1 and E2 – see rows with H1 under "Hypothesis"). The rate was generally lower during construction (4.8-13.8%) than operation (5.1-14.9%). It was lower when all PAM sensors within the large monitoring area were used in the analysis, and highest when only the small PAM grid was used in the small monitoring area. There was no clear difference between the Van Paris et al. Design and the 10x10 grid, but false positive rate was generally lower for the T-design. There was no clear pattern among species.

4.3 Statistical power to detect change for Van Parijs et al. PAM design.

The power to detect change in number of detected cues for all hypotheses (Table 2) and each species is given in Table D-1 in Appendix D.

There was low statistical power (< 80%) for all hypotheses related to construction and operation (H2 – H7) regardless of the studied species and the chosen dose-response function. There was high statistical power to detect global decline (H8) for minke and fin whales but low for sei and NARW regardless of the monitoring size and PAMs grid used.

4.4 Statistical power to detect change with alternative PAM designs.

The statistical power to detect change in generated cue rates under the two alternative PAM designs is given in Tables E1 and E2 in Appendix E.

Neither the T-design nor 10x10 km grid resulted in high power to detect change in generated cue rates for fin, sei and NARW for all hypotheses related to construction and operation (H2-H7) of the wind farm, regardless of the number of PAM sensors used. The 10x10 design resulted in high power to detect change for H4 for minke whales and T-design also resulted in high power for the other two hypotheses assuming redistribution of whales during construction (H3 and H5). The power to detect change remained low for all hypotheses related to operation (H7) for minke whales under the two alternative designs. The power to detect change remained high for detecting global decline (H8) for minke but low for the remaining species for the two alternative PAM designs.

For cases where power to detect change is high, this power was higher for DR2 than DR1 by 1-6%. The power to detect change was also higher by 2.3-7% if only modified small PAM grid was used in comparison to use all grids.

5 Discussion

5.1 Discussion of results

The statistical power to detect biologically plausible changes in acoustic encounter rate due to construction and operation was generally low (i.e., <80%) under the Van Parijs et al. design regardless of the studied species. The main determinant of these results is the small effect size, which is a result of three factors.

First, the distance over which the response was assumed to occur is small. For hypotheses involving construction, where a dose-response function for response was assumed, p(response) was almost 0 by 35 km for both DRs and below 0.1 by 18 km for less sensitive DR. If a more sensitive dose-response function were assumed, where probability of response remained higher at greater distances, power to detect change would also be greater. For hypotheses related to operation of the wind farms, the effect size was even smaller than for hypotheses related to construction as assumed p(response) = 0 outside the wind farm footprints. This explains why power was higher for hypotheses related to construction than operation.

Second, under the Van Parijs et al. design, relatively few sensors are placed within or close to the wind farm footprint, in the area where p(response) is high and only six PAM sensors from small PAM grid were designed for VOWA. The design assumed a minimum 20 km spacing of the PAM sensors, which covers almost the entire dose-response function. The alternative designs both had considerably more sensors within and close to the wind farm footprints and the result was substantially increasing power. Of the two, the T-design gave higher power because the smaller between-sensor spacing at the middle of the wind farms resulted in more sensors where the response was greatest.

Third, baseline density of all four species was low, in some months equal or very close to zero. This may have produced a zero-effect size in some simulation replicates. For the T-design, sensors were placed along gradients of highest baseline density, and this may have contributed somewhat to the higher power. Note that none of the PAM designs resulted in high power to detect change for NARW and sei. This is not surprising as these two species are rarely present at the VOWA area.

Comparing the species, two factors seemed to drive variation in power between species. The major factor was variation in detection ranges, summarized in the EDR measure. Fin whales, in particular, had a large EDR and low power. As described earlier, the large detection distances meant that PAM sensors detected a mix of responding and non-responding whales, effectively "diluting" the effect size and resulting in low power. There are two potential solutions to this. The first is to increase (or impose) the received level (or signal to noise ratio) detection threshold for this species in the PAM detector, thereby decreasing the range over which detections are made. This would result in a decrease in the sample size of detections, but since detections per month were estimated to be over 100,000 in many months, this is unlikely to be problematic. A second solution is to use the fact that multiple sensors can detect the same vocalization for this species (given the sensor spacing) and attempt to localize vocalizations. This then would enable a more refined analysis of effect, which presumably would have higher power.

Power was higher under the displacement hypotheses H3-H5 than the acoustic behavior changes hypothesis H2. This is because the displacement hypotheses not only decreased acoustic detections in the vicinity of the VOWA, but also increased it at further distances from the VOWA (up to 100 km), by redistributing the displaced animals. Note that H3 (symmetrical displacement) is unlikely to be realistic, as it places animals in locations where they are not recorded at all under baseline conditions. H5 was based on AIS data, which is likely not an accurate representation of how disturbance could increase around VOWA during construction as we did not distinguish between AIS of the vessels from regular shipping lanes and vessels related to the construction of the wind farms—but in any case, the difference in power between H3, H4 and H5 was small, and no consistent pattern was detected.

One seeming paradox in the results is that power was somewhat higher when just the small PAM grid within the small monitoring area was considered and was lower when both the small and large PAM grids were considered within the large monitoring area. This finding is explained by again considering effect size. The large monitoring area contains mostly grid cells that are far from the VOWA and hence where no effect takes place. Hence the average effect size over the large monitoring area is smaller than that over the small monitoring area. Similarly, the large grid contains PAM sensors that are farther from the VOWA on average than the small grid, which is clustered around the VOWA. Hence the average effect size when large and small sensors are combined is less than for the small grid alone.

A conclusion from the above might be that there is no need for the large monitoring grid. However, the results from the false positive tests showed that false positive rate is lower with a larger monitoring area. This is a good reason to monitor over a larger area. Another is that the design will then be more robust to a misspecification of the dose-response function should animals be displaced over larger distances than were used here. False positive rates were found to be lower for the T-design than other regularly spaced PAM designs. This way the chance to falsely assign change in whale density with distance from the farms is lower. Power was good to detect a global decline but only for minke whales. Although this was only tested at the VOWA level, power is likely to be higher still when VOWA are combined in a regional analysis (see Chudzinska and Thomas 2023). We did not examine the potential of the PAM designs to monitor long-term population changes, but based on our initial study with H8 it seems that this might be a powerful approach compared with the current standard of visual line transect surveys, where power is often low due to high estimated variance. However, visual line transects surveys produce estimates of absolute density and abundance (notwithstanding issues associated with trackline detection bias and availability bias), while the PAM network tested here is designed only to produce acoustic cue detection rate. Separate studies of detection ranges and cue production rates would be required to enable production of absolute density estimates from the PAM network (Marques et al. 2009, Marques et al. 2013)

5.2 Limitations of analysis and potential future studies

In undertaking the power analysis, we made a number of assumptions. Here we review these and their potential effect on estimated power. We suggest future studies that might be undertaken to improve reliability of the estimated power values.

- The presented analysis is based on simulated data where we estimated the number of potentially detected cues using animal density models derived from visual data and literature-reported cue production rates for each species. For all four studied species, these cue rates come from very different habitats than the studied site. Additionally, the cue production rate was assumed constant over time and space for each species. If cue production rates are higher than our assumed values, then power will be higher; on the other hand, variation in cue production over time and space will reduce power. Targeted studies of cue rates in the studied species and within the study area (for example studies of factors affecting the calling behavior of NARW) would be required to produce better-supported inputs.
- One method for checking the realism of the simulated acoustic data would be to compare baseline simulated acoustic encounter rates with those measured on PAM sensors previously deployed within the study area. At the time of this study, only data on acoustic presence-absence were available, so further acoustic processing may be required to obtain measured acoustic encounter rates.
- The simulations assumed a consistent effect of each hypothesized change in each grid cell, month and VOWA. For each simulation realization, after randomly sampling from the animal density surface, the

other steps were deterministic. For example, it is infeasible that cue production rate is constant over space and time. This likely caused power to be over-estimated given that in the real world a consistent, deterministic effect cannot be expected.

- The assumption that all animals within the wind area footprint responded is almost certainly an overestimate of response and therefore results in an overestimate of power. However, the assumption for operation that no animals respond outside the wind area footprint likely results in an underestimate of power. The dose-response functions used were assumed to apply to all species and may be too sensitive or not sensitive enough. A simple propagation model was used in converting from a received level-based dose-response function to a distance-based one, and a single assumption about sound source level was used. The direction of any resulting bias in estimated power is unknown.
- Assuming the monthly variation in density between grid cells in the density surface had a scale parameter that was 1/30th the daily visual survey-derived estimates may over or under-estimate the variability (and hence power). Truncating outliers in the generated density surface likely had a minimal effect on power.
- The use of effective detection area in lieu of a full detection function when determining acoustic detections is an approximation and had an unknown effect on power (although likely not large). We assumed that effective detection area was constant in space and time. Variation in detection area, like other sources of variability, will decrease power unless it can be accounted for by collecting additional data alongside the acoustic encounter rates that will allow detectability to be estimated.
- We assume no change in response (sensitization or desensitization) with repeated exposure. Either of these would change the effect size and hence power.
- In the analysis some models fitted to simulated data failed to fully converge. The effect of this is likely minor but could be investigated further. For the small (20 km) monitoring size, number of PAM sensors used in the simulations was low (six in case on Van Parijs et al. design), which was the frequent reasons of models not converging. Increasing number of sensors, as suggested in the two alternative PAM designs, greatly reduced this problem.

In evaluating the power of hypotheses relating to construction we undertook only local analyses, analyzing each wind farm separately. Our experience from the regional analysis of operational scenarios was that combining wind farms into a regional analysis increased power (Chudzinska and Thomas 2023). Therefore, we expect higher power to detect construction effects could be gained by combining wind farms, and this would make a useful future study.

We analyzed data at the temporal scale of one month because this was the resolution available to us from the habitat-based density models. However, there is likely a strong signal of any construction effect in hourly or daily acoustic records (see, e.g., the diurnal pattern suggested under H2 in Table 2). Hence power is likely greater at this temporal level to detect a construction effect. We suggest that further efforts be made to obtain acoustic data that could be used as the basis for a simulation study at the daily level.

In creating and testing alternative designs, we tried two alternatives, the 10 x 10 km grid and the T-design, but many others are possible. In particular, the T-design could be replaced by a cross (x or +), although the branch running towards the shoreline would likely need to be truncated, and also would likely enter very low whale density habitat. In the T-design different inter-sensor spacings could be investigated, and the robustness of the results to mis-specification of the dose-response function checked (i.e., checking power if a very different dose-response function is used than the one used to optimize the sensor spacing). If an updated dose-response function were available, the T-design could be adjusted to account for the new function. Additionally, the T-design allows for adjusting spacing between sensors without necessarily increasing the total number of sensors used in the design. Such flexibility is lessened in a regularly spaced grid – firstly because of the regular spacing, and secondly because the number of sensors required goes up with the square of the grid spacing (e.g., a 5 x 5 km grid requires 4 times as many sensors as a 10 x 10 km grid to cover the same area). Overall, the optimal spacing of the sensors should be directly related to expected probability of response with distance from the wind farms.

The power analysis we undertook looked only at power to detect an effect. After an effect of wind farm construction or operation is detected then natural follow-up questions are: I) what is the dose-response function, and II) how many animals are affected? It is not certain that the design best suited to detecting an effect will also produce the most precise dose-response function or estimate of number of affected animals. For example, the most important part of a dose-response function for estimating number of affected animals is the part at larger ranges, because many more animals are exposed at these ranges than at smaller ranges (Tyack and Thomas 2019). To obtain a precise dose-response function at larger ranges may require more sensors at these ranges than in the T-design (although there are also the sensors of the large grid) or alternative data collection such as through tagging individuals. Optimizing the sensor array to estimate these quantities could be the subject of a future study.

- The statistical model we fitted was a one-dimensional model in the sense that the only spatial metric was distance from sound source. It is possible, and potentially more powerful given sufficient sensors, to estimate effects using a two-dimensional spatial model. This is the basis for the MRSea package (Scott-Hayward et al. 2017), which is routinely used in Europe to undertake phase-gradient analyses. Even if not used as part of the power analysis, such an MRSea-type analysis may be worth considering if the PAM monitoring grid is commissioned.
- We have not considered the power of the proposed grid for detecting long-term global trends in cetacean abundance. This would make a potentially useful future study. One factor that would need to be considered is that the sensor placement to detect wind farm effects may be non-random and so the resulting data may

not be suitable for analysis using typical "design-based" methods, but instead as part of a spatial modeling exercise. Such models are already a standard part of the analysis of long-term wildlife monitoring schemes.

- The Van Parijs. et al. design covers the area from the coast to the continental shelf (Figure 1). It is worth noting that all of the four chosen species frequently occur further off the shelf (e.g. Roberts et al. 2016a), beyond the area covered by Van Parijs et al., as well as alternative designs suggested in this report. Expanding the design further offshore was not tested in this study.

5.3 Recommendations

- Based on the results of this power analysis, we recommend replacing the proposed 20 x 20 km small grid of sensors around VOWA with an alternative array that concentrates sensors where a response is expected and distributes sensors relatively evenly across VOWA that are to be used as study sites.
- However, there is also a need for sensors at distances from the VOWA where no response is expected, and hence there is a role for the 40 x 40 km grid, at least out to the distances we tested (out to the continental shelf). Including both the local sensors and the 40 x 40 km grid appeared also to reduce the false positive detection rate.
- Of the designs we tested, the T-design appears better than a dense grid of sensors in terms of the amount of statistical power generated for a fixed investment of sensors.
- To maximize the sample size of acoustic sensors we recommend pooling resources across stakeholders who are deploying sensors.
- For species like fin whales with large acoustic detection distances, consideration should be given to localizing calls and undertaking effects analysis using the localizations.
- The power analysis we have undertaken could be improved, and we have made some suggestions for future studies in the previous section. One particular suggestion is to undertake analysis of existing PAM data to provide a cross-check that our simulated acoustic encounter rates are realistic.
- One method to improve power is to accept a higher false-positive detection rate. We used a nominal false-positive rate of 5% and a target power of 80%, but these values are merely conventions and consideration could be given to using other values.

6 Acknowledgements

We thank members of the Marine Mammal Subcommittee of the RWSC and BOEM for providing feedback on our planned analyses. We thank Jason Roberts of Duke University for providing the marine mammal density model outputs we used to simulate baseline animal density, Megan Ryder of SMRU Consulting for extracting AIS data, Jason Wood from SMRU Consulting and Susan Barco from Barco Marine Consulting for providing feedback and Cornelia Oedekoven from CREEM for discussions on the analytical methods. Funding for this project was provided by the Virginia Coastal Zone Management Program and the National Oceanic and Atmospheric Administration.

7 Literature Cited

Bailey, H., B. Senior, D. Simmons, J. Rusin, G. Picken, and P. M. Thompson. 2010. Assessing underwater noise levels during pile-driving at an offshore windfarm and its potential effects on marine mammals. Marine Pollution Bulletin **60**:888-897.

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

Benhemma-Le Gall, A., I. M. Graham, N. D. Merchant, and P. M. Thompson. 2021. Broad-scale responses of harbor porpoises to pile-driving and vessel activities during offshore windfarm construction. Frontiers in Marine Science **8**:664724.

BOEM, and NOAA. 2022. DRAFT BOEM and NOAA Fisheries North Atlantic Right Whale and Offshore Wind Strategy.

Booth, C., A. K. D. O' Hara, and L. Thomas. 2023. North Atlantic Right Whale Expert Elicitation. Assessing population effects of offshore wind development on North Atlantic Right Whales (Eubalaena glacials). BOEM Contract 140M0121C008 report, version 1.0.

Brandt, M. J., A.-C. Dragon, A. Diederichs, M. A. Bellmann, V. Wahl, W. Piper, J. Nabe-Nielsen, and G. Nehls. 2018. Disturbance of harbour porpoises during construction of the first seven offshore wind farms in Germany. Marine Ecology Progress Series **596**:213-232.

Brandt, M. J., A. Dragon, A. Diederichs, A. Schubert, V. Kosarev, G. Nehls, V. Wahl, A. Michalik, A. Braasch, C. Hinz, C. Katzer, D. Todeskino, M. Gauger, M. Laczny, and W. Piper. 2016. Effects of offshore pile driving on harbour porpoise abundance in the German Bight. Report prepared for Offshore Forum Windenergie.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. Introduction to Distance Sampling. Estimating abundance of biological populations. Oxford University Press, Oxford.

Calderan, S., B. Miller, K. Collins, P. Ensor, M. Double, R. Leaper, and J. Barlow. 2014. Low-frequency vocalizations of sei whales (Balaenoptera borealis) in the Southern Ocean. The Journal of the Acoustical Society of America **136**:EL418-EL423.

CETAP. 1982. A characterization of marine mammals and turtles in the mid- and North Atlantic areas of the U.S. outer continental shelf, final report. Washington, DC: Bureau of Land Management.

Cholewiak, D., C. W. Clark, D. Ponirakis, A. Frankel, L. T. Hatch, D. Risch, J. E. Stanistreet, M. Thompson, E. Vu, and S. M. Van Parijs. 2018. Communicating amidst the noise: modeling the aggregate influence of ambient and vessel noise on baleen whale communication space in a national marine sanctuary. Endangered Species Research **36**:59-75.

Chudzinska, M., and L. Thomas. 2023. Power analysis for optimal design of a passive acoustic monitoring network in the Atlantic Ocean. U.S. Department of the Interior, Bureau of Ocean Energy Management. xx p. Report No.: OCS Study BOEM 20xx-xxx. Contract No.: xxxxx.

Dähne, M., A. Gilles, K. Lucke, V. Peschko, S. Adler, K. Krugel, J. Sundermeyer, and U. Siebert. 2013. Effects of pile-driving on harbour porpoises (*Phocoena phocoena*) at the first offshore wind farm in Germany. Environmental Research Letters **8**:025002.

Dähne, M., J. Tougaard, J. Carstensen, A. Rose, and J. Nabe-Nielsen. 2017. Bubble curtains attenuate noise from offshore wind farm construction and reduce temporary habitat loss for harbour porpoises. Marine Ecology Progress Series **580**:221-237.

Davis, G. E., M. F. Baumgartner, J. M. Bonnell, J. Bell, C. Berchok, J. Bort Thornton, S. Brault, G. Buchanan, R. A. Charif, and D. Cholewiak. 2017. Long-term passive acoustic recordings track the changing distribution of North Atlantic right whales (Eubalaena glacialis) from 2004 to 2014. Scientific Reports **7**:1-12.

Davis, G. E., M. F. Baumgartner, P. J. Corkeron, J. Bell, C. Berchok, J. M. Bonnell, J. Bort Thornton, S. Brault, G. A. Buchanan, and D. M. Cholewiak. 2020. Exploring movement patterns and changing distributions of baleen whales in the western North Atlantic using a decade of passive acoustic data. Global change biology **26**:4812-4840.

Dunlop, R. A., R. D. McCauley, and M. J. Noad. 2020. Ships and air guns reduce social interactions in humpback whales at greater ranges than other behavioral impacts. Marine Pollution Bulletin **154**:111072.

Dunlop, R. A., M. J. Noad, R. D. McCauley, E. Kniest, R. Slade, D. Paton, and D. H. Cato. 2018. A behavioural dose-response model for migrating humpback whales and seismic air gun noise. Marine Pollution Bulletin **133**:506-516.

Dunlop, R. A., M. J. Noad, R. D. McCauley, L. Scott-Hayward, E. Kniest, R. Slade, D. Paton, and D. H. Cato. 2017. Determining the behavioural dose–response relationship of marine mammals to air gun noise and source proximity. Journal of Experimental Biology **220**:2878-2886.

Durbach, I. N., C. M. Harris, C. Martin, T. A. Helble, E. E. Henderson, G. Ierley, L. Thomas, and S. W. Martin. 2021. Changes in the movement and calling behavior of minke whales (Balaenoptera acutorostrata) in response to navy training. Frontiers in Marine Science **8**:660122.

Ellison, W., B. Southall, C. Clark, and A. Frankel. 2012. A new context-based approach to assess marine mammal behavioral responses to anthropogenic sounds. Conservation Biology **26**:21-28.

Erbe, C., C. Reichmuth, K. Cunningham, K. Lucke, and R. Dooling. 2016. Communication masking in marine mammals: A review and research strategy. Marine Pollution Bulletin.

Estabrook, B. J., K. Hodge, D. P. Salisbury, A. Rahaman, D. Ponirakis, D. Harris, J. Zeh, A. N. Park, and A. Rice. 2021. Final Report for New York Bight Whale Monitoring Passive Acoustic Surveys October 2017-October 2020. Contract C009925. New York State Department of Environmental Conservation. East Setauket, NY.

Fregosi, S., D. V. Harris, H. Matsumoto, D. K. Mellinger, S. W. Martin, B. Matsuyama, J. Barlow, and H. Klinck. 2022. Detection probability and density estimation of fin whales by a Seaglider. The Journal of the Acoustical Society of America **152**:2277-2291.

Gervaise, C., Y. Simard, F. Aulanier, and N. Roy. 2019. Optimal passive acoustic systems for real-time detection and localization of North Atlantic right whales in their feeding ground off Gaspé in the Gulf of St. Lawrence. Department of Fisheries and Oceans Ottawa, Canada, Can. Tech. Rep. Fish. Aquat. Sci. 3345: ix + 58 p.

Graham, I. M., D. Gillespie, K. C. Gkikopoulou, G. D. Hastie, and P. M. Thompson. 2023. Directional hydrophone clusters reveal evasive responses of small cetaceans to disturbance during construction at offshore windfarms. Biol Lett **19**:20220101.

Graham, I. M., N. D. Merchant, A. Farcas, T. R. C. Barton, B. Cheney, S. Bono, and P. M. Thompson. 2019. Harbour porpoise responses to pile-driving diminish over time. Royal Society Open Science **6**:190335.

Harris, C. M., L. Thomas, E. A. Falcone, J. Hildebrand, D. Houser, P. H. Kvadsheim, F. P. A. Lam, P. J. O. Miller, D. J. Moretti, A. J. Read, H. Slabbekoorn, B. L. Southall, P. L. Tyack, D. Wartzok, and V. M. Janik. 2018.

Marine mammals and sonar: dose-response studies, the risk-disturbance hypothesis and the role of exposure context. Journal of Applied Ecology.

Hatch, L. T., C. W. Clark, S. M. Van Parijs, A. S. Frankel, and D. W. Ponirakis. 2012. Quantifying loss of acoustic communication space for right whales in and around a US National Marine Sanctuary. Conservation Biology **26**:983-994.

Heaney, K. D., M. A. Ainslie, M. B. Halvorsen, K. D. Seger, R. Muller, M. Nijhof, and T. Lippert. 2020. A Parametric Analysis and Sensitivity Study of the Acoustic Propagation for Renewable Energy

Sources. Sterling (VA): U.S. Department of the Interior, Bureau of Ocean Energy Management. Prepared by CSA Ocean Sciences Inc. OCS Study BOEM 2020-011. 165 p.

Katona, S. K., and J. A. Beard. 1990. Population size, migrations and feeding aggregations of the humpback whale (Megaptera novaeangliae) in the western North Atlantic Ocean. Pages 295-306 Report of the International Whaling Commission (Special Issue 12).

Kenney, R. D., C. A. Mayo, and H. E. Winn. 2020. Migration and foraging strategies at varying spatial scales in western North Atlantic right whales: a review of hypotheses. J. Cetacean Res. Manage.:251-260.

Kowarski, K. A., B. J. Gaudet, A. J. Cole, E. E. Maxner, S. P. Turner, S. B. Martin, H. D. Johnson, and J. E. Moloney. 2020. Near real-time marine mammal monitoring from gliders: Practical challenges, system development, and management implications. J Acoust Soc Am **148**:1215.

Kraus, S., R. D. Kenney, and L. Thomas. 2019. A Framework for Studying the Effects of Offshore Wind Development on Marine Mammals and Turtles. Report prepared for the Massachusetts Clean Energy Center, Boston, MA 02110, and the Bureau of Ocean Energy Management.

Mackenzie, M., L. Scott-Hayward, C. S. Oedekoven, H. Skov, E. Humphreys, and E. Rexstad. 2013. Statistical Modelling of Seabird and Cetacean data: Guidance Document. University of St. Andrews contract for Marine Scotland; SB9 (CR/2012/05).

Malme, C., P. Miles, C. Clark, P. Tyack, and J. Bird. 1984. Investigations of the Potential Effects of Underwater Noise from Petroleum Industry Activities on Migrating Gray Whale Behavior—Phase II. U-S. Department of the Interior Minerals Management Service.

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

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 Blainville's beaked whales. The Journal of the Acoustical Society of America **125**:1982-1994.

Martin, S. W., T. A. Marques, L. Thomas, R. P. Morrissey, S. Jarvis, N. DiMarzio, D. Moretti, and D. K. Mellinger. 2013. Estimating minke whale (*Balaenoptera acutorostrata*) boing sound density using passive acoustic sensors. Marine Mammal Science **29**:142-158.

Methratta, E. T. 2021. Distance-Based Sampling Methods for Assessing the Ecological Effects of Offshore Wind Farms: Synthesis and Application to Fisheries Resource Studies. Frontiers in Marine Science **8**.

Meyer-Gutbrod, E. L., C. H. Greene, and K. T. Davies. 2018. Marine species range shifts necessitate advanced policy planning: The case of the North Atlantic right whale. Oceanography **31**:19-23.

Miller, D. L., and L. Thomas. 2015. Mixture models for distance sampling detection functions. PLoS ONE **10**:e0118726.

Mitchell, B. 1975. Preliminary report on Nova Scotia fishery for sei whales (Balaenoptera borealis). Report of the International Whaling Commission, 25, 218–225.

Morano, J. L., D. P. Salisbury, A. N. Rice, K. L. Conklin, K. L. Falk, and C. W. Clark. 2012. Seasonal and geographical patterns of fin whale song in the western North Atlantic Ocean. The Journal of the Acoustical Society of America **132**:1207-1212.

Muirhead, C. A., A. M. Warde, I. S. Biedron, A. Nicole Mihnovets, C. W. Clark, and A. N. Rice. 2018. Seasonal acoustic occurrence of blue, fin, and North Atlantic right whales in the New York Bight. Aquatic Conservation: Marine and Freshwater Ecosystems **28**:744-753.

Palmer, K. J., S. Tabbutt, D. Gillespie, J. Turner, P. King, D. Tollit, J. Thompson, and J. Wood. 2022. Evaluation of a coastal acoustic buoy for cetacean detections, bearing accuracy and exclusion zone monitoring. Methods in Ecology and Evolution.

Parks, S. E., A. Searby, A. Celerier, M. P. Johnson, D. P. Nowacek, and P. L. Tyack. 2011. Sound production behavior of individual North Atlantic right whales: implications for passive acoustic monitoring. Endanger Species Res **15**:63-76.

Pyć, C., D. Zeddies, S. L. Denes, and M. Weirathmueller. 2018. Appendix III-M: REVISED DRAFT - Supplemental Information for the Assessment of Potential Acoustic and Non-acoustic Impact Producing Factors on Marine Fauna during Construction of the Vineyard Wind Project. Document 001639, Version 3.1. Technical report by JASCO Applied Sciences (USA) Inc. for Vineyard Wind.

R Core Team. 2022. R: A language to environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Risch, D., M. Castellote, C. W. Clark, G. E. Davis, P. J. Dugan, L. E. Hodge, A. Kumar, K. Lucke, D. K. Mellinger, and S. L. Nieukirk. 2014. Seasonal migrations of North Atlantic minke whales: novel insights from large-scale passive acoustic monitoring networks. Movement Ecology **2**:24.

Risch, D., C. W. Clark, P. J. Dugan, M. Popescu, U. Siebert, and S. M. Van Parijs. 2013. Minke whale acoustic behavior and multi-year seasonal and diel vocalization patterns in Massachusetts Bay, USA. Marine Ecology Progress Series **489**:279-295.

Roberts, J., and T. Yack. 2022a. Density Model for Common Minke Whale (Balaenoptera acutorostrata) for the U.S. East Coast, Version 10, 2022-06-20, and Supplementary Report. Marine Geospatial Ecology Laboratory, Duke University, Durham, North Carolina. https://seamap.env.duke.edu/models/Duke/EC/EC_Common_minke_whale_history.html.

Roberts, J., and T. Yack. 2022b. Density Model for North Atlantic Right Whale (Eubalaena glacialis) for the U.S. East Coast, Version 12, 2022-02-14, and Supplementary Report. Marine Geospatial Ecology Laboratory, Duke University, Durham, North Carolina. https://seamap.env.duke.edu/models/Duke/EC/EC_North_Atlantic_right_whale_history.html.

Roberts, J., and T. Yack. 2022c. Density Model for Sei Whale (Balaenoptera borealis) for the U.S. East Coast, Version 10, 2022-06-20, and Supplementary Report. Marine Geospatial Ecology Laboratory, Duke University, Durham, North Carolina. https://seamap.env.duke.edu/models/Duke/EC/EC_Sei_whale_history.html.

Roberts, J., T. Yack, and P. N. Halpin. 2022. Density Model for Fin whale (Balaenoptera physalus) for the U.S. East Coast, Version 12, 2022-06-20. 2022-06-20, and Supplementary Report. Marine Geospatial Ecology Laboratory, Duke University, Durham, North Carolina.Prepared for Naval Facilities Engineering Command, Atlantic by the Duke University Marine Geospatial Ecology Lab, Durham, NC. https://seamap.env.duke.edu/models.

Roberts, J. J., B. D. Best, L. Mannocci, E. Fujioka, P. N. Halpin, D. L. Palka, L. P. Garrison, K. D. Mullin, T. V. Cole, and C. B. Khan. 2016. Habitat-based cetacean density models for the US Atlantic and Gulf of Mexico. Scientific Reports **6**:1-12.

Rolland, R. M., R. S. Schick, H. M. Pettis, A. R. Knowlton, P. K. Hamilton, J. S. Clark, and S. D. Kraus. 2016. Health of North Atlantic right whales *Eubalaena glacialis* over three decades: From individual health to demographic and population health trends. Marine Ecology Progress Series **542**:265-282.

Salisbury, D. P., B. J. Estabrook, H. Klinck, and A. Rice. 2018. Understanding marine mammal presence in the Virginia Offshore Wind Energy Area. Sterling (VA): US Department of the Interior, Bureau of Ocean Energy Management. OCS Study BOEM 2019-007. 103 p.

Scott-Hayward, L., C. S. Oedekoven, M. L. Mackenzie, C. Walker, and E. Rexstad. 2017. MRSea package (version 1.0-beta): Statistical Modelling of bird and cetacean distributions in offshore renewables development areas. Retrieved from http://creem2.st-and.ac.uk/software.aspx.

Sivle, L. D., P. H. Kvadsheim, C. Curé, S. Isojunno, P. J. Wensveen, F.-P. A. Lam, F. Visser, L. Kleivane, P. L. Tyack, and C. M. Harris. 2015. Severity of expert-identified behavioural responses of humpback whale, minke whale, and northern bottlenose whale to naval sonar. Aquatic Mammals **41**:469.

Sivle, L. D., P. J. Wensveen, P. Kvadsheim, F.-P. A. Lam, F. Visser, C. Curé, C. M. Harris, P. L. Tyack, and P. Miller. 2016. Naval sonar disrupts foraging behaviour in humpback whales. Marine Ecology Progress Series.

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

Stöber, U., and F. Thomsen. 2021. How could operational underwater sound from future offshore wind turbines impact marine life? The Journal of the Acoustical Society of America **149**:1791-1795.

Teilmann, J., and J. Carstensen. 2012. Negative long term effects on harbour porpoises from a large scale offshore wind farm in the Baltic — evidence of slow recovery. **045101**.

Tyack, P., and L. Thomas. 2019. Using dose–response functions to improve calculations of the impact of anthropogenic noise. Aquatic Conservation Marine and Freshwater Ecosystems. **29(S1)**:242-253.

Tyack, P. L. 2011. Using digital acoustic recording tags to detect marine mammals on Navy ranges and study their responses to Naval sonar.

Van Parijs, S. M., K. Baker, J. Carduner, J. Daly, G. E. Davis, C. Esch, S. Guan, A. Scholik-Schlomer, N. B. Sisson, and E. Staaterman. 2021. NOAA and BOEM Minimum Recommendations for Use of Passive Acoustic Listening Systems in Offshore Wind Energy Development Monitoring and Mitigation Programs. Frontiers in Marine Science **8**:760840.

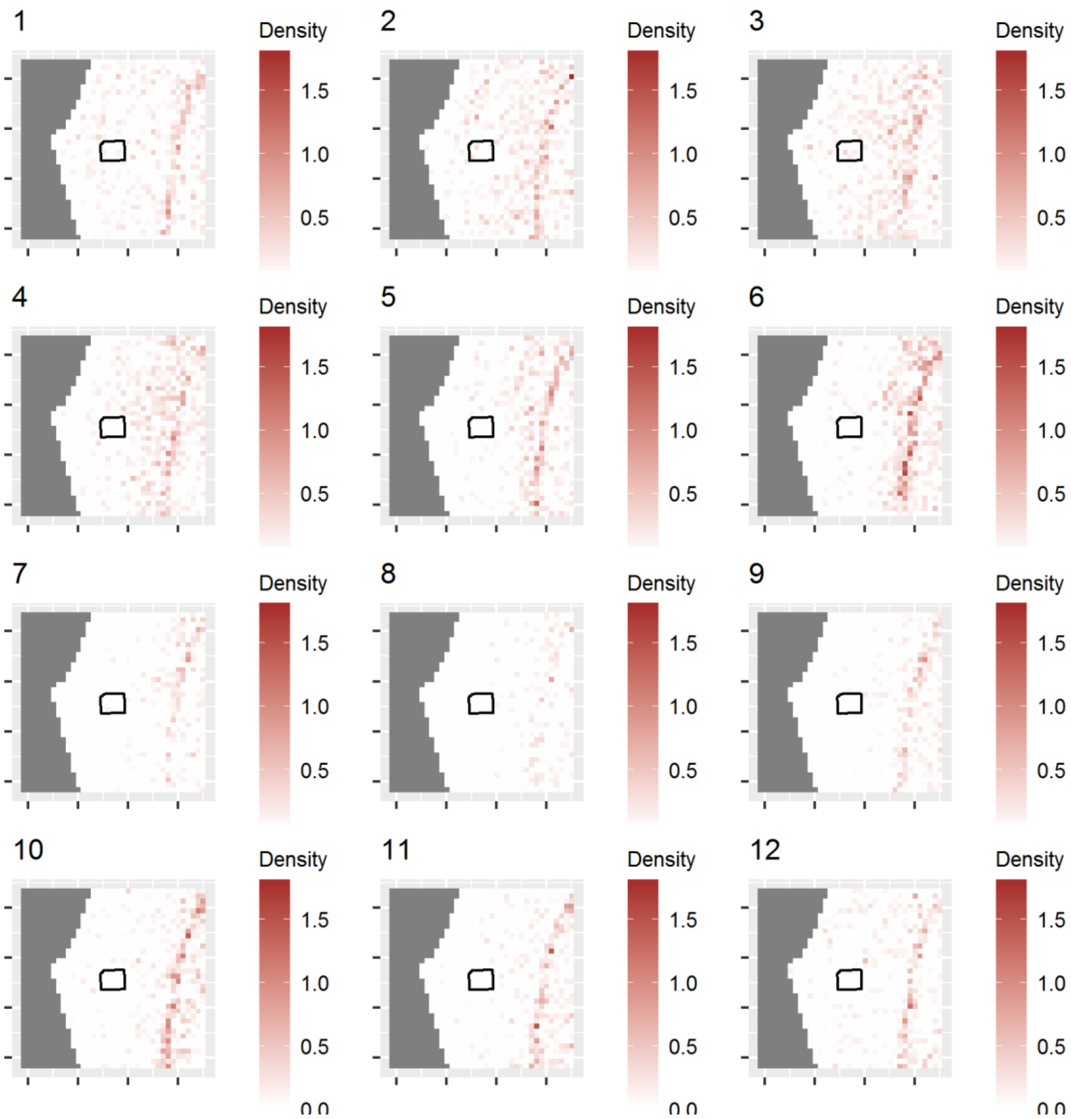
Vu, E. T., D. Risch, C. W. Clark, S. Gaylord, L. T. Hatch, M. A. Thompson, D. N. Wiley, and S. M. Van Parijs. 2012. Humpback whale song occurs extensively on feeding grounds in the western North Atlantic Ocean. Aquatic Biology **14**:175-183.

Wood, J., B. Southall, and D. Tollit. 2012. PG&E offshore 3-D seismic survey project EIR – Marine Mammal Technical Report., SMRU Ltd Report.

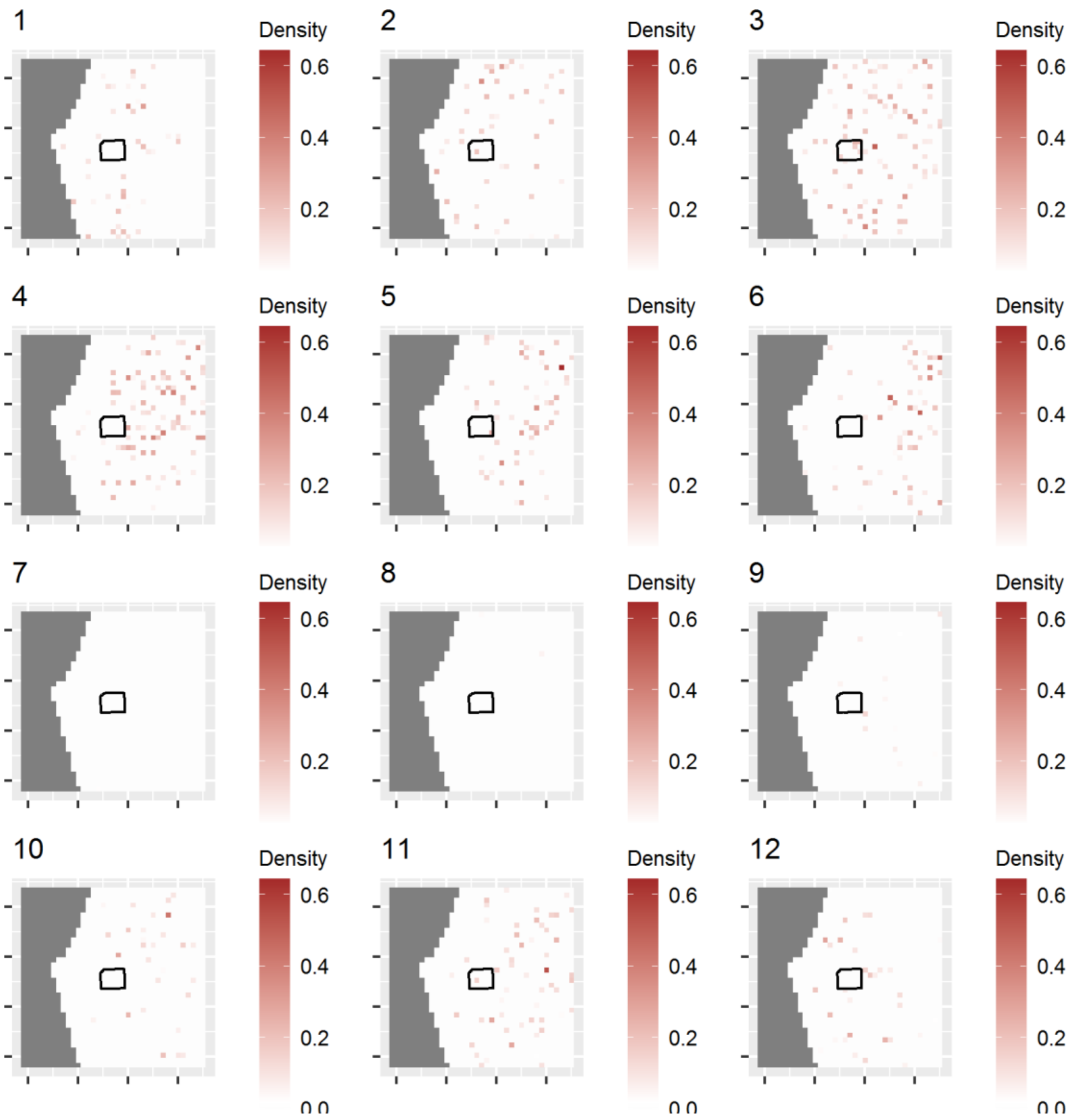
Wood, S. N. 2017. Generalized Additive Models: An Introduction with R (2nd edition). Chapman and Hall/CRC.

8 Appendix A Baseline density of the studied species at all studied wind farm areas for each month of the year.

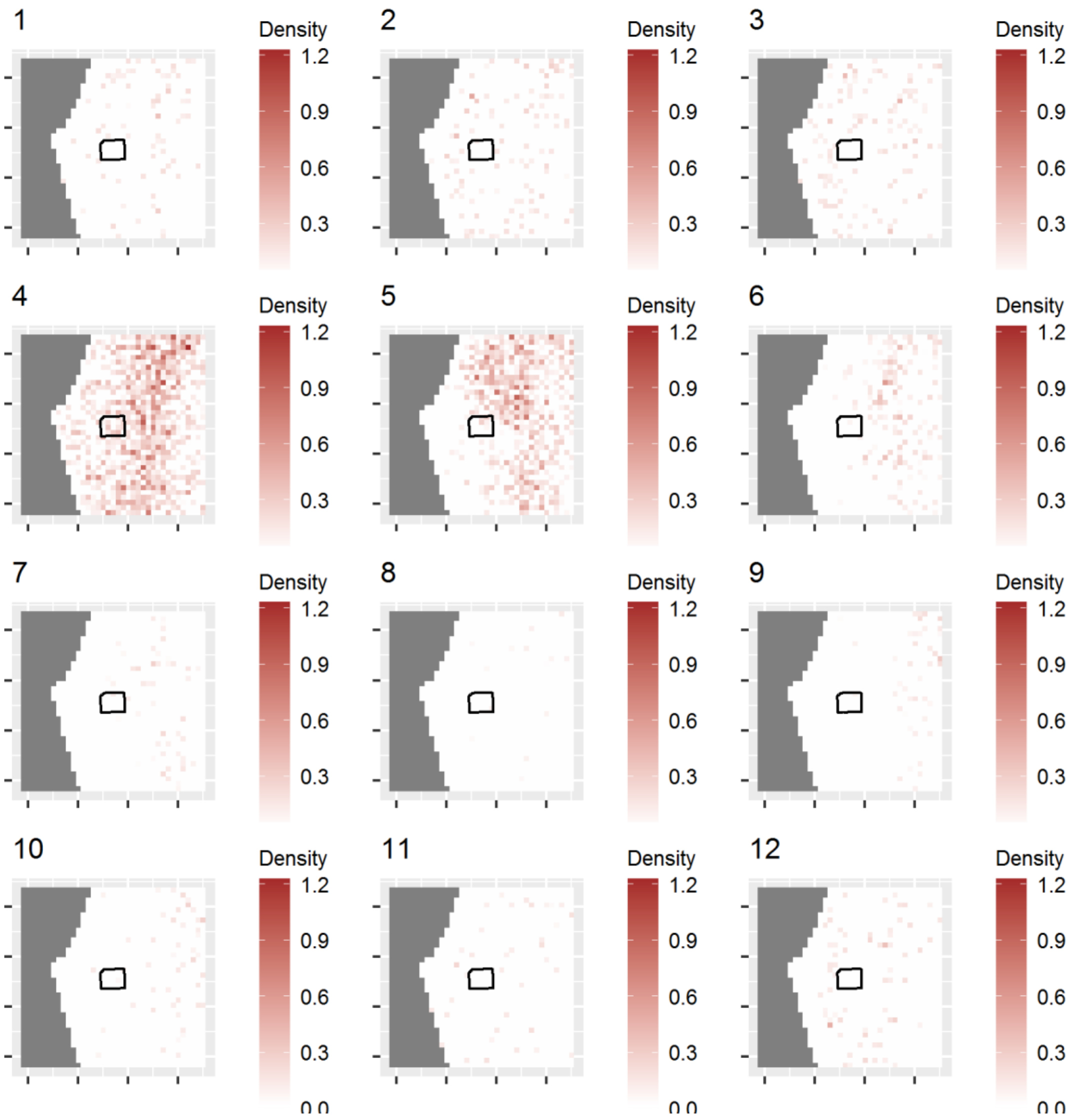
Fin whale



Sei whale


1
Density
0.6
0.4
0.2
2
Density
0.6
0.4
0.2
3
Density
0.6
0.4
0.2
4
Density
0.6
0.4
0.2
5
Density
0.6
0.4
0.2
6
Density
0.6
0.4
0.2
7
Density
0.6
0.4
0.2
8
Density
0.6
0.4
0.2
9
Density
0.6
0.4
0.2
10
Density
0.6
0.4
0.2
0.0
11
Density
0.6
0.4
0.2
0.0
12
Density
0.6
0.4
0.2
0.0

Minke whale



North Atlantic right whale (NARW)

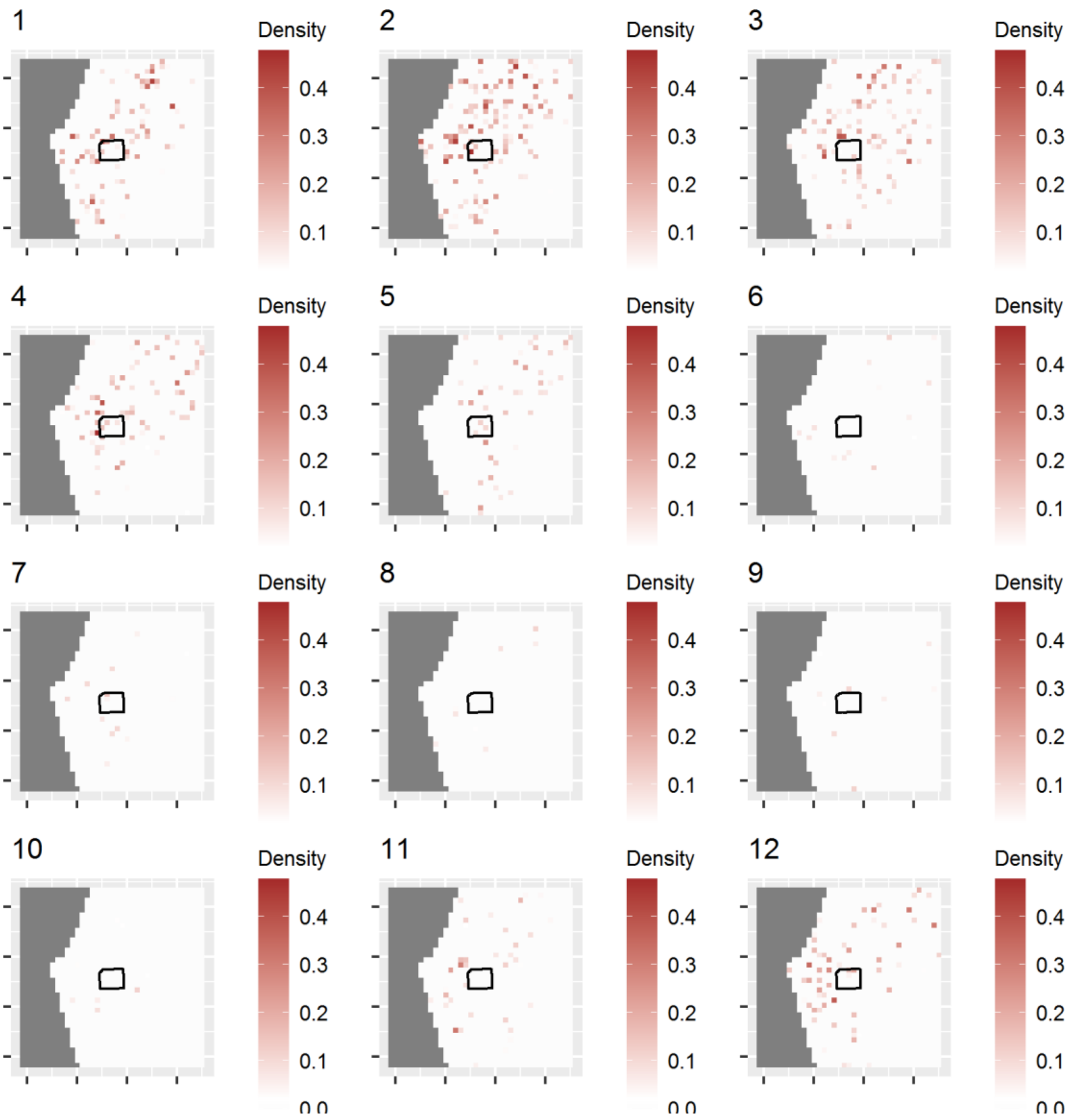


Figure A1. Densities are shown as individuals/25 km², and are displayed on a 5x5 km grid. Numbers over each sub-panel indicates month of the year (1 = January, etc.).

9 Appendix B: PAM design suggested by van Parijs for each monitoring area and effective detection ranges for the studied species.

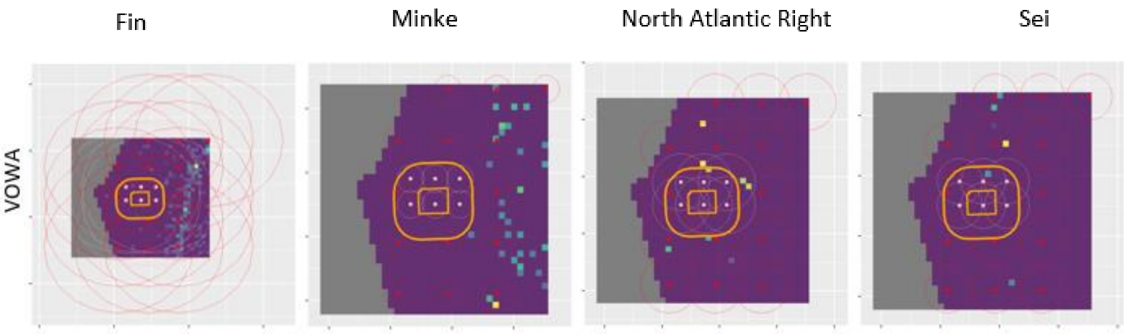


Figure B1. Area covered by all the PAM stations as suggested by Van Parijs et al. for the large monitoring site. Red circles show effective detection area for all four species (Error! Reference source not found.**. Pink dots and corresponding buffers show the small PAM grid and red dots show the large PAM grid. Orange lines show the small monitoring area.**

10 Appendix C. Number of responding whales.

Table C1. Mean number of responding whales over 500 realizations of the density surfaces for each month of construction at each wind farms. Two dose-response functions (DR1, and DR2), defining the probability of response as a function of distance from construction, were used. Effect size for each DR was calculated as proportion of all whales at the small monitoring area which responded. Densities are in number of whales per 25 km^2. The mean number of detected cues is calculated for baseline density for all PAM grids at the large monitoring area and is calculated as mean number of cues (cues per month) at the grids overlapping with EDR of each species.

Species	Wind farm	Month of construction	Mean responding (DR2)	Mean responding (DR1)	Effect size (DR2)	Effect size (DR1)	Average density at large monitoring area	Average density at small monitoring area	Mean number of detected cues
Fin	VOWA	5	0.52	0.21	0.057	0.139	0.07	0.02	6240000
Fin	VOWA	6	0.42	0.17	0.043	0.103	0.09	0.03	8550000
Fin	VOWA	7	0.09	0.04	0.045	0.11	0.03	0.01	2481000
Fin	VOWA	8	0.03	0.01	0.023	0.072	0.02	0.005	1377000
Fin	VOWA	9	0.09	0.04	0.044	0.113	0.04	0.005	3330000

Fin	VOWA	10	0.13	0.05	0.028	0.078	0.07	0.01	5880000
Minke	VOWA	5	3.23	1.39	0.073	0.169	0.11	0.12	64200
Minke	VOWA	6	0.28	0.11	0.024	0.061	0.03	0.03	19500
Minke	VOWA	7	0.15	0.06	0.044	0.115	0.01	0.01	5580
Minke	VOWA	8	0.02	0.01	0.028	0.072	0.005	0.005	2085
Minke	VOWA	9	0.03	0.01	0.06	0.133	0.01	0.005	2745
Minke	VOWA	10	0.06	0.02	0.033	0.088	0.01	0.005	5010
Sei	VOWA	5	0.15	0.06	0.062	0.151	0.01	0.01	19800
Sei	VOWA	6	0.02	0.01	0.027	0.063	0.005	0.005	8850
Sei	VOWA	7	0	0	0.076	0.147	0.005	0.005	330
Sei	VOWA	8	0	0	0.028	0.096	0.005	0.005	270
Sei	VOWA	9	0.01	0.01	0.078	0.156	0.005	0.005	1080
Sei	VOWA	10	0.03	0.01	0.036	0.095	0.005	0.005	5190
NARW	VOWA	5	0.12	0.05	0.073	0.166	0.005	0.005	11760
NARW	VOWA	6	0.03	0.01	0.05	0.133	0.005	0.005	9420
NARW	VOWA	7	0.02	0.01	0.045	0.115	0.005	0.005	4650
NARW	VOWA	8	0.02	0.01	0.043	0.111	0.005	0.005	4800
NARW	VOWA	9	0.03	0.01	0.067	0.146	0.005	0.005	6360
NARW	VOWA	10	0.03	0.01	0.054	0.123	0.005	0.005	10650

11 Appendix D. Power to detect change in distribution or behavior of whales using van Parijs design.

Table D1. Statistical power (%) to detect change in cue rates under seven hypotheses (see Table 2 Error! Reference source not found.**) for each studied species, and combination of monitoring size and PAM grid. H1 (no effect) was calculated both for construction phase (H1_co) and operation (H1_op). For hypotheses related to construction of the windfarms (H2-H5), power to detect change was calculated for two dose-response functions: DR1 and DR2. Scenarios with power >= 80% are marked in bold.**

Design	Hypothesis	Species	Wind farm	Large monitoring area & all PAM grids	Large monitoring area & small PAM grid only
Van Parijs et al.	H1_co	Minke	VOWA	6.7	8.7
Van Parijs et al.	H1_op	Minke	VOWA	7.8	11.8
Van Parijs et al.	H1_co	Sei	VOWA	9.8	13.8
Van Parijs et al.	H1_op	Sei	VOWA	9.9	14.9
Van Parijs et al.	H1_co	Fin	VOWA	4.8	6
Van Parijs et al.	H1_op	Fin	VOWA	5.1	8.1
Van Parijs et al.	H1_co	NARW	VOWA	7.8	11.8
Van Parijs et al.	H1_op	NARW	VOWA	8.4	12.4
Van Parijs et al.	H2_50_DR1	Minke	VOWA	14.8	21.8
Van Parijs et al.	H2_50_DR2	Minke	VOWA	15.8	21.8
Van Parijs et al.	H2_100_DR1	Minke	VOWA	18.2	25.2
Van Parijs et al.	H2_100_DR2	Minke	VOWA	21.6	25.6
Van Parijs et al.	H2_50_DR1	Sei	VOWA	2.9	7.9
Van Parijs et al.	H2_50_DR2	Sei	VOWA	3.2	5.2
Van Parijs et al.	H2_100_DR1	Sei	VOWA	9.2	14.2
Van Parijs et al.	H2_100_DR2	Sei	VOWA	7.6	13.6
Van Parijs et al.	H2_50_DR1	Fin	VOWA	0.5	4.5
Van Parijs et al.	H2_50_DR2	Fin	VOWA	1.3	4.3
Van Parijs et al.	H2_100_DR1	Fin	VOWA	3.2	5.2
Van Parijs et al.	H2_100_DR2	Fin	VOWA	4	10
Van Parijs et al.	H2_50_DR1	NARW	VOWA	8.6	15.6
Van Parijs et al.	H2_50_DR2	NARW	VOWA	7.6	9.6
Van Parijs et al.	H2_100_DR1	NARW	VOWA	9.2	15.2
Van Parijs et al.	H2_100_DR2	NARW	VOWA	9.6	13.6

Van Parijs et al.	H3_DR1	Minke	VOWA	14.2	19.2
Van Parijs et al.	H3_DR2	Minke	VOWA	16.2	22.2
Van Parijs et al.	H3_DR1	Sei	VOWA	8.2	15.2
Van Parijs et al.	H3_DR2	Sei	VOWA	9.6	13.6
Van Parijs et al.	H3_DR1	Fin	VOWA	0.4	5.4
Van Parijs et al.	H3_DR2	Fin	VOWA	2.3	9.3
Van Parijs et al.	H3_DR1	NARW	VOWA	12.3	16.3
Van Parijs et al.	H3_DR2	NARW	VOWA	15.6	19.6
Van Parijs et al.	H4_DR1	Minke	VOWA	24.2	27.2
Van Parijs et al.	H4_DR2	Minke	VOWA	27.3	30.3
Van Parijs et al.	H4_DR1	Sei	VOWA	9.5	13.5
Van Parijs et al.	H4_DR2	Sei	VOWA	9.3	14.3
Van Parijs et al.	H4_DR1	Fin	VOWA	4.5	7.5
Van Parijs et al.	H4_DR2	Fin	VOWA	3.8	6.8
Van Parijs et al.	H4_DR1	NARW	VOWA	15.8	21.8
Van Parijs et al.	H4_DR2	NARW	VOWA	16.4	22.4
Van Parijs et al.	H5_DR1	Minke	VOWA	24.2	30.2
Van Parijs et al.	H5_DR2	Minke	VOWA	26.3	33.3
Van Parijs et al.	H5_DR1	Sei	VOWA	8.8	11.8
Van Parijs et al.	H5_DR2	Sei	VOWA	8.5	10.5
Van Parijs et al.	H5_DR1	Fin	VOWA	2	5
Van Parijs et al.	H5_DR2	Fin	VOWA	4	7
Van Parijs et al.	H5_DR1	NARW	VOWA	7.8	13.8
Van Parijs et al.	H5_DR2	NARW	VOWA	7.9	11.9
Van Parijs et al.	H7	Minke	VOWA	6	9
Van Parijs et al.	H7	Sei	VOWA	2	7
Van Parijs et al.	H7	Fin	VOWA	2.3	7.3
Van Parijs et al.	H7	NARW	VOWA	0.4	2.4

Van Parijs et al.	H8	Minke	VOWA	**87**	**82**
Van Parijs et al.	H8	Sei	VOWA	48.4	51.4
Van Parijs et al.	H8	Fin	VOWA	11.8	13.4
Van Parijs et al.	H8	NARW	VOWA	62	63

12 Appendix E. Power to detect change in distribution or behavior of whales assuming two alternative PAM designs.

12.1 10 x 10 km grid

Table E1. Statistical power (%) to detect change in cue rates under seven hypotheses (see Error! Reference source not found.**) for each studied species, and combination of monitoring size and PAM grid assuming modified small PAM grid using 10x10 km grid. For hypotheses related to construction of the windfarms (H2-H5), power to detect change was calculated for two dose-response functions: DR1 and DR2. For alternative designs estimating power for small monitoring area and modified small PAM only was not conducted.**

Design	Hypothesis	Species	Wind farm	Large monitoring area & all PAM grid	Large monitoring area & modified small PAM grid only
10x10	H1_co	Minke	VOWA	5.7	6.7
10x10	H1_op	Minke	VOWA	5.0	9.5
10x10	H1_co	Sei	VOWA	9.8	13.0
10x10	H1_op	Sei	VOWA	9.5	12.3
10x10	H1_co	Fin	VOWA	5.2	3.5
10x10	H1_op	Fin	VOWA	5.4	7.9
10x10	H1_co	NARW	VOWA	5.8	11.3
10x10	H1_op	NARW	VOWA	8.2	10.2
10x10	H2_50_DR1	Minke	VOWA	56.8	64.5
10x10	H2_50_DR2	Minke	VOWA	60.0	66.0
10x10	H2_100_DR1	Minke	VOWA	60.0	63.2
10x10	H2_100_DR2	Minke	VOWA	56.8	60.8

10x10	H2_50_DR1	Sei	VOWA	23.2	28.4
10x10	H2_50_DR2	Sei	VOWA	25.0	27.7
10x10	H2_100_DR1	Sei	VOWA	18.7	22.6
10x10	H2_100_DR2	Sei	VOWA	18.2	22.5
10x10	H2_50_DR1	Fin	VOWA	7.5	13.3
10x10	H2_50_DR2	Fin	VOWA	8.8	11.3
10x10	H2_100_DR1	Fin	VOWA	7.4	11.8
10x10	H2_100_DR2	Fin	VOWA	9.2	11.9
10x10	H2_50_DR1	NARW	VOWA	50.0	57.2
10x10	H2_50_DR2	NARW	VOWA	46.0	52.1
10x10	H2_100_DR1	NARW	VOWA	35.2	31.6
10x10	H2_100_DR2	NARW	VOWA	35.0	37.4
10x10	H3_DR1	Minke	VOWA	69.8	72.1
10x10	H3_DR2	Minke	VOWA	69.8	72.2
10x10	H3_DR1	Sei	VOWA	21.0	24.6
10x10	H3_DR2	Sei	VOWA	21.8	26.6
10x10	H3_DR1	Fin	VOWA	11.0	17.7
10x10	H3_DR2	Fin	VOWA	11.5	15.5
10x10	H3_DR1	NARW	VOWA	51.0	55.9
10x10	H3_DR2	NARW	VOWA	55.3	57.5
10x10	H4_DR1	Minke	VOWA	**87.0**	**91.0**
10x10	H4_DR2	Minke	VOWA	**89.8**	**96.8**
10x10	H4_DR1	Sei	VOWA	21.5	25.2
10x10	H4_DR2	Sei	VOWA	25.1	32.2
10x10	H4_DR1	Fin	VOWA	9.5	15.5
10x10	H4_DR2	Fin	VOWA	9.7	12.6
10x10	H4_DR1	NARW	VOWA	38.2	33.4
10x10	H4_DR2	NARW	VOWA	35.8	30.2
10x10	H5_DR1	Minke	VOWA	67.0	69.2

10x10	H5_DR2	Minke	VOWA	79.1	79.0
10x10	H5_DR1	Sei	VOWA	22.0	26.4
10x10	H5_DR2	Sei	VOWA	22.3	26.5
10x10	H5_DR1	Fin	VOWA	12.0	17.3
10x10	H5_DR2	Fin	VOWA	8.9	13.2
10x10	H5_DR1	NARW	VOWA	38.2	30.2
10x10	H5_DR2	NARW	VOWA	35.8	39.1
10x10	H7	Minke	VOWA	73.8	76.8
10x10	H7	Sei	VOWA	23.6	30.6
10x10	H7	Fin	VOWA	5.6	8.7
10x10	H7	NARW	VOWA	31.4	33.7
10x10	H8	Minke	VOWA	**94.3**	**95.8**
10x10	H8	Sei	VOWA	44.1	47.3
10x10	H8	Fin	VOWA	14.6	17.9
10x10	H8	NARW	VOWA	45.2	48.6

12.2 T-design

Table E2. Statistical power (%) to detect change in cue rates under seven hypotheses (see Error! Reference source not found.**for each studied species and combination of monitoring size and PAM grid assuming modified small PAM grid using T-design grid. For hypotheses related to construction of the windfarms (H2-H5), power to detect change was calculated for two dose-response functions: DR1 and DR2.**

Design	Hypothesis	Species	Wind farm	Large monitoring area & all PAM grids	Large monitoring area & modified small PAM grid only
T-design	H1_co	Minke	VOWA	4.2	6.5
T-design	H1_op	Minke	VOWA	5.6	11.2
T-design	H1_co	Sei	VOWA	9.2	11.5
T-design	H1_op	Sei	VOWA	11.7	18.5
T-design	H1_co	Fin	VOWA	8	12.3
T-design	H1_op	Fin	VOWA	12.5	16.5
T-design	H1_co	NARW	VOWA	5.8	10.1
T-design	H1_op	NARW	VOWA	8.2	15.5
T-design	H2_50_DR1	Minke	VOWA	68.4	74.8

T-design	H2_50_DR2	Minke	VOWA	78.8	82.3
T-design	H2_100_DR1	Minke	VOWA	81	83.8
T-design	H2_100_DR2	Minke	VOWA	83	89.8
T-design	H2_50_DR1	Sei	VOWA	58.2	65.8
T-design	H2_50_DR2	Sei	VOWA	58.2	61.0
T-design	H2_100_DR1	Sei	VOWA	57.8	62.5
T-design	H2_100_DR2	Sei	VOWA	57.8	61.5
T-design	H2_50_DR1	Fin	VOWA	9	15.6
T-design	H2_50_DR2	Fin	VOWA	11.8	17.4
T-design	H2_100_DR1	Fin	VOWA	15	18.0
T-design	H2_100_DR2	Fin	VOWA	15	19.8
T-design	H2_50_DR1	NARW	VOWA	40.6	43.3
T-design	H2_50_DR2	NARW	VOWA	47.4	53.9
T-design	H2_100_DR1	NARW	VOWA	47.8	53.2
T-design	H2_100_DR2	NARW	VOWA	46.6	53.9
T-design	H3_DR1	Minke	VOWA	**85**	**88.7**
T-design	H3_DR2	Minke	VOWA	**89**	**94.9**
T-design	H3_DR1	Sei	VOWA	58.4	64.6
T-design	H3_DR2	Sei	VOWA	58.4	60.5
T-design	H3_DR1	Fin	VOWA	10	12.2
T-design	H3_DR2	Fin	VOWA	12	18.2
T-design	H3_DR1	NARW	VOWA	41.2	47.8
T-design	H3_DR2	NARW	VOWA	47.4	42.9
T-design	H4_DR1	Minke	VOWA	**85**	**89.7**
T-design	H4_DR2	Minke	VOWA	**91**	**93.3**
T-design	H4_DR1	Sei	VOWA	58.2	64.6
T-design	H4_DR2	Sei	VOWA	58.2	61.1
T-design	H4_DR1	Fin	VOWA	11	15.1
T-design	H4_DR2	Fin	VOWA	13	16.7
T-design	H4_DR1	NARW	VOWA	39.4	46.9
T-design	H4_DR2	NARW	VOWA	37.6	43.6
T-design	H5_DR1	Minke	VOWA	**89**	**91.6**
T-design	H5_DR2	Minke	VOWA	**90**	**93.1**
T-design	H5_DR1	Sei	VOWA	58.2	60.7
T-design	H5_DR2	Sei	VOWA	58.2	63.3
T-design	H5_DR1	Fin	VOWA	10	13.1
T-design	H5_DR2	Fin	VOWA	12	16.4
T-design	H5_DR1	NARW	VOWA	39.8	45.6
T-design	H5_DR2	NARW	VOWA	37.4	44.6
T-design	H7	Minke	VOWA	78.4	85.2
T-design	H7	Sei	VOWA	58	63.0
T-design	H7	Fin	VOWA	8.4	8.2
T-design	H7	NARW	VOWA	61.6	65.1

T-design	H8	Minke	VOWA	**82.4**	**89.0**
T-design	H8	Sei	VOWA	44.6	45.6
T-design	H8	Fin	VOWA	12.6	13.4
T-design	H8	NARW	VOWA	41.8	47.0

13 Addendum: Recommendations for Developing a Baleen Whale Monitoring Plan for Virginia's Wind Energy Area

Submitted By Susan G. Barco

Barco Marine Consulting, contracted through the Regional Wildlife Science Collaborative for Offshore Wind and Coastal States Stewardship Foundation

13.1 Executive Summary

The project, funded by the Virginia Coastal Zone Management Program (VA CZM), was conducted to begin the process of developing an environmental monitoring plan to study the impact of offshore wind (OSW) development and operation on baleen whales in Virginia. For this project, we determined the efficacy of using passive acoustic monitoring (PAM) for monitoring the effect of OSW construction on baleen whales, particularly the North Atlantic right whale (*Eubalaena glacialis*-NARW) which is a critically endangered species that migrates offshore between summer feeding grounds to the north and winter breeding grounds to the south of Virginia. At its inception, the core work for this project was planned to be conducted by researchers at the Centre for Research into Ecological and Environmental Modelling (CREEM; https://www.creem.st-andrews.ac.uk/). The experts at CREEM specialize in developing and applying innovative mathematical and statistical methods to solve problems in biology, ecology and geography. Much of the CREEM scientists' recent work has focused on studying the effect of offshore wind in the UK and Europe on marine mammals and the environment. The funding for this project was contracted to the Regional Wildlife Science Collaborative for Offshore Wind (RWSC) which also sought funding for a similar coastwide analysis. As such, this project became part of a larger, more complex regional analysis supported by BOEM and described in Chudzinska and Thomas (in review).

CREEM researchers were contracted to develop a statistical power analysis to determine the ability to detect changes to baleen whale occurrence using PAM in the waters surrounding the Coastal Virginia Offshore Wind (CVOW) development area. During the time between project proposal and funding, scientists at the National Marine Fisheries Service and BOEM published minimum recommendations for use of passive acoustic listening systems in offshore wind energy development monitoring and mitigation programs for the US Atlantic continental shelf (Van Parijs et al. 2021). The CREEM scientists used the Van Parijs et al. (2021) recommendations along with existing distance sampling survey data on baleen whale species in Virginia to develop data sets that were used to test several hypotheses about how whales may respond to offshore wind construction, particularly pile driving. Pile driving was chosen because studies in other areas suggested it is likely to evoke the greatest response from cetaceans (whales, dolphins and porpoises). If this study shows that impacts of pile driving cannot be detected, then other sound related impacts associated with wind turbine operation are unlikely to be detected as well. The hypotheses ranged from no response to displacement from the region. Because critically endangered North Atlantic right whales may not occur in Virginia in sufficient numbers to detect changes in their occurrence, the project focused on four baleen whale species each with different vocal characteristics and predicted density: North Atlantic right whales (*Eubalaena glacialis*-NARW), fin (*Baleanoptera physalus*), minke (*Baleanoptera acurostrata*) and sei whales (*Baleanoptera borealis*).

Results of the analyses showed that the combination of 1) relatively low and variable baleen whale density off Virginia, 2) expected limits on timing of construction to avoid times when NARW are in the region, and 3) uncertainty associated with sole use of whale vocalization via PAM (rates of vocal cues, detection distance, behavioral changes in vocalization, etc.) as a measurement of occurrence combined to make the probability of detecting changes in whale occurrence extremely low even when a significant number of PAM units are deployed. These results suggest that use of PAM as the sole method of monitoring baleen whales is not recommended for impact assessment of these species in Virginia. PAM, especially units that provide real time detection notification may, however, be useful for a broad scope environmental monitoring and as mitigation tools during the construction phase of the project. A significant PAM array of 20 units or more, combined with other research methods such as surveys, focal follows, and tagging are required both for developing baseline data on abundance and distribution and to detect impacts of OSW construction on baleen whales off Virginia.

13.2 Acknowledgements

Funding for this project was provided by the Virginia Coastal Zone Management Program and the National Oceanic and Atmospheric Administration.

13.3 Introduction & Background

Understanding the impacts of offshore wind (OSW) development on marine and avian wildlife requires prior knowledge of species occurrence. Ideally, multiple years of consistently collected baseline data on species abundance and environmental correlates should be collected in the region of OSW development in order to predict possible impacts, detect changes if they occur, and assign causality of detected changes to different variables (Kraus et al. 2019). Several baleen whale species occur in the waters off Virginia, and the area primarily serves as a migratory corridor between high latitude summer feeding areas to the north and lower latitude breeding areas to the south of the region.

Critically endangered North Atlantic right whales (*Eubalaena glacialis*-NARW) are thought to migrate southward in November and December and northward in March and April, but individual NARW have been detected in all months of the year (Salisbury et al. 2016, 2018). Other baleen whale species present in Virginia continental shelf waters are humpback (*Megaptera novaeangliae*), fin (*Baleanoptera physalus*), minke (*Baleanoptera acurostrata*) and sei whales (*Baleanoptera borealis*). Blue whales (*Baleanoptera musculus*) likely occur offshore of the continental shelf. Non-breeding (primarily juvenile) humpback, fin and minke whales may remain in Mid-Atlantic waters in winter months to feed when prey is available.

Of all the marine species that occur off the Virginia coast which may be impacted by OSW development, the NARW is the most critically endangered, and, therefore of the greatest concern for research and mitigation. Of the baleen whale species found in the region, however, estimated density is lowest for NARW (Roberts et al. 2016). Because the estimated number and timing of all whales migrating off Virginia each year also varies naturally, the ability to detect changes caused by human activities on these species is exceedingly difficult. Knowing this, the project was designed to include multiple baleen whale species that could, hopefully, collectively increase the statistical power to detect and assign cause for observed changes.

Passive acoustic monitoring (PAM) is a relatively low cost means of detecting animals that vocalize underwater. PAM units can be moored in one place or be attached to autonomous underwater vehicles such as gliders that can be programmed to follow survey lines. Both moored and autonomous can be designed to archive data that are later downloaded or transmit detected vocalizations in near real time. Artificial intelligence is used to identify vocalizations and other sounds recorded by PAM units and different units are designed to detect different frequency ranges, thus making them more or less likely to capture sounds made by certain species or species groups. Baleen whale species such as right whales make a variety of sounds, and PAM recordings can be searched for a variety of species-specific sounds. Cost of sound analyses increase with each type of vocalization added to analyses and real time units, where sound processing occurs onboard the PAM unit instead after being downloaded may have processing limitations that affect battery life, cost and other variables.

Generally different types of PAM units are deployed for baleen whales (mysticetes) which produce lower frequency sounds and toothed whales (odontocetes) which produce higher frequency sounds including biosonar or echolocation. One challenge of using PAM alone as a methodology is that animals produce sounds (vocal cues) at different rates and times, and a lack of cues does not necessarily indicate a lack of animals in an area. Likewise, a high number of detected cues could indicate many animals each with a relatively low cue rate or a smaller number that have high cue rates. Finally sounds produced at different frequency and volume travel different distances underwater and environmental parameters (water temperature, salinity, thermocline, bottom substrate, depth, etc.) also affect how sound travels. Fin whale vocalizations, for example, can be detected from much greater distances (~100km) than those from minke whales (~9km - Table 4 p.30 in Chudzinska et al. in review report).

It is important to understand the complexities and limitations of conducting PAM research on baleen whale presence in order to develop a comprehensive wildlife monitoring plan for any area. The technology, though relatively inexpensive compared to survey and tagging efforts, requires thoughtful experimental design and analyses prior to deployment in order to understand what meaningful information can be provided. The justification for this project was to understand what level of effort and experimental design would be required use PAM to monitor baleen whales in the vicinity of the Virginia wind energy area. With funding from BOEM, additional analyses were conducted through the Regional Wildlife Science Collaborative for Offshore Wind

(RWSC) for other areas along the Atlantic coast which have baleen whale presence including other migratory aeras as well as summer feeding areas (Chudzinski and Thomas 2023).

13.4 Brief Summary of Methods and Results

A detailed discussion of methods and results are available in the SMRUC report and Figure 2 (p. 22) summarizes the methods used (Chudzinski et al. 2in review). In general, the analysis required several complex steps that are summarized below:

1) Generate data to be used in the modeling effort: Collection and analysis of existing data on whale presence in the region by month as well as information on whale vocalization by species and timing of construction activities that are most likely to impact baleen whales were all required in a multistep process to generate a model data test to test hypotheses. These steps included:
    a. Monthly baleen whale density from surveys
    b. PAM data for vocal cues detected in the region
    c. Existing and estimated construction timing restrictions
    d. Information on cue rates and detection distance of various vocalizations made by different whale species

    From these data, the researchers calculated the number of whales likely to be in the area during the months of construction.
2) Develop hypotheses and data for testing hypotheses. Develop PAM array designs to test against various hypotheses.
3) Transform estimated number of whales into number of vocalizations (cues) detected by each PAM unit in each PAM array and determine statistical power to detect change based on experimental scenario and PAM array (described below).

The researchers used two different experimental scenarios to assess the ability to detect changes in whale occurrence. The first assumed that the data were collected before-during and/or before-after (BD/BA) construction and compared for differences between the time periods, and the second assumed a phase-gradient design (PG) where PAM units were located at known distances from the area of interest and data were collected simultaneously at increasing distances from the area where construction noise would be located. The complexities of each approach were described in detail in the SMRU report.

The PAM array designs that were tested included a minimum distance PAM design where units were placed at equal distances across the region (20km x 20km) and closer in specific wind energy areas (10km x 10km; Van Parjs et al. 2021). When this design failed to detect changes, the researchers assumed that approximately 20 units were deployed in 10km x 10km across a broader area that included but extended beyond the wind energy area. An alternative design, expected to be more effective for PG experimental approaches, a 'T-shaped' design with approximately 20 units placed closer together at the intersection of the 'T' within the wind energy area and farther apart at the ends of the three 'arms' outside the wind energy area (see Figure 8 p.32 in Chudzinski and Thomas 2023) was also tested.

Based on, among other things, construction limitations (expected to be limited to May-Oct), whale density, cue rates, and detection distance, there was only month (May) where only one species (minke whale), was predicted to be detected more than one time (3.23 detections). All other species were predicted to be detected >1 time per month with several estimated to be zero (see Appendix C p.50-1 in Chudzinski and Thomas 2023). Few expected detections meant that there was no power to detect changes for any one species at any specific time regardless of the experimental design or approach. Collectively, assuming 12 months of construction over 2 years (May-Oct for two years) ability to detect change in whale cues in the vicinity of the Virginia wind energy area was relatively low, and the possibility of incorrectly predicting that a change occurred when it did not occur (false positive error) was relatively high (see Appendices D and E in Chudzinski and Thomas 2023). When applied across a larger region, such as the mid-Atlantic, New York Bight and Southern New England regions, power to detect global decline in whale presence was relatively high using all of the PAM designs.

13.5 Discussion

The results were disappointing but not unexpected. The low density of baleen whales off Virginia at any one time combined with expected construction restrictions limiting pile driving to months when NARW are less likely to be in the region mean that testing the impacts of construction on baleen whales using PAM units, even in high numbers, is not feasible. For the same reasons, PAM used in conjunction with other methods would be unlikely to allow for impacts on baleen whales to be measured statistically in Virginia waters. If such a project were to be developed, a phase gradient approach using a high level of effort with multiple methodologies would be needed to have any power to detect changes in whale presence. Use of PAM with other methods such as surveys and tagging would, however, contribute significantly to continued environmental monitoring for baleen whales in this important migratory corridor. Multiple years of consistently collected data that include both distance sampling surveys, tag tracks, and vocal cues collected using PAM would help scientist to understand the migratory paths and timing of baleen whale movement off Virginia's coast.

If the purpose of using PAM is solely to detect baleen whales for environmental monitoring/mitigation and not to detect impacts of wind energy development, deployment of PAM units, especially units capable of transmitting near real time information, is an appropriate method to assist in determining GO/NO GO status for construction activities. PAM alone, however, cannot determine whale presence in the regional since whales must be vocalizing in order to be detected. PAM can act as one of several tools to provide information on whale presence, but, due to its limitations, PAM should not be the only method of determining whale presence in a sensitive area.

Finally, if researchers are interested in studying impacts of wind energy development on acoustically sensitive cetaceans in Virginia waters, studying more common species that are likely to be in the region when construction activity occurs such as bottlenose (*Tursiops truncatus*) or common (*Delphinus delphis*) dolphins provide greater likelihood of success. Developing a robust experimental design with multiple methodologies and testing the power to detect change prior to start would still be strongly recommended.

13.6 Literature Cited

Chudzinska M, Thomas L (in review) Power analysis for optimal design of a passive acoustic monitoring network in the Atlantic Ocean. U.S. Department of the Interior, Bureau of Ocean Energy Management. Xx p. Report No.: OCS Study BOEM 20xx-xxx. Contract No.: xxxxx.

Chudzinska M, Thomas, L, Wood, J (2023) Power analysis for optimal design of passive acoustic monitoring network in the Virginia offshore wind area. Report number SMRUC- RWS-2023-010 (unpublished).

Kraus SD, Kenney RD, Thomas L (2019) A Framework for Studying the Effects of Offshore Wind Development on Marine Mammals and Turtles. Report prepared for the Massachusetts Clean Energy Center, Boston, MA 02110, and the Bureau of Ocean Energy Management.

Roberts JJ, Best BD, Mannocci L, Fujioka E, Halpin PN, Palka DL, Garrison LP, Mullin KD, Cole TV, Khan CB (2016) Habitat-based cetacean density models for the US Atlantic and Gulf of Mexico. *Scientific Reports* 6:1-12.

Salisbury DP, Estabrook BJ, Klinck H, Rice A (2018) Understanding marine mammal presence in the Virginia Offshore Wind Energy Area. Sterling (VA): US Department of the Interior, Bureau of Ocean Energy Management. OCS Study BOEM 2019-007. 103 p.

Salisbury DP, Clark CW, Rice A. N (2016) Right whale occurrence in the coastal waters of Virginia, USA: Endangered species presence in a rapidly developing energy market. *Marine Mammal Science*, *32*(2), 508-519.

Van Parijs SM, Baker K, Carduner J, Daly J, Davis GE, Esch C, Guan S, Scholik-Schlomer A, Sisson NB, Staaterman E (2021) NOAA and BOEM Minimum Recommendations for Use of Passive Acoustic Listening Systems in Offshore Wind Energy Development Monitoring and Mitigation Programs. *Frontiers in Marine Science* 8:760840