Assessing the Viability of Density Estimation for Cetaceans from Passive Acoustic Fixed Sensors throughout the Life Cycle of an Offshore E&P Field Development

Authors:	Booth, C.G., Oedekoven, C.S., Gillespie, D., Macaulay, J., Plunkett, R, Joy, R., Harris, D., Wood, J., Marques, T. A., Marshall, L., Verfuss, U.K., Tyack, P. Johnson, M., & Thomas, L.
Report Code:	SMRUC-OGP-2017-001
Date:	Tuesday, 04 July 2017

THIS REPORT IS TO BE CITED AS:

Booth, C.G., Oedekoven, C.S., Gillespie, D., Macaulay, J., Plunkett, R, Joy, R., Harris, D., Wood, J., Marques, T. A., Marshall, L., Verfuss, U.K., Tyack, P. Johnson, M., & Thomas, L. 2017. Assessing the Viability of Density Estimation for Cetaceans from Passive Acoustic Fixed Sensors throughout the Life Cycle of an Offshore E&P Field Development. Report number: SMRUC-OGP-2017-001. Submitted to IOGP Sound and Marine Life Joint Industry Programme (Unpublished).

Contents

Figures

Tables

Glossary of Terms, Acronyms and Abbreviations

This report covers a number of specific subject areas and as such we use a number of distinct terms. In addition to the Glossary below and outlined throughout the document, here we draw attention to a key element of the use of PAM for density estimation. Critically, not all PAM systems have the same functionality; however they can be grouped into two distinct classes relevant to implementing DECAF (via Distance Sampling or Spatially Explicit Capture Recapture). These are nodes that can only detect (throughout this study termed **n-nodes**, because they generate *n, i.e. counts of the object of interest, be it animals, groups or cues,* in the density estimation equation) and nodes that can localize (i.e. provide range estimates to vocalizing animals) (termed **p-nodes**, because the data can be used to estimate the detection probability, or equivalently, the effective detection area of sensors). In the description of the model elements, please note: we use italic font for naming parameters (e.g. *c.cv*) – see section 4 for more details.

A number of other terms that are important for this report are outlined in the table below.

Term	Description
Ambient noise	That part of the total noise background observed with a non-directional hydrophone which is not due to the hydrophone and its manner of mounting or to some identifiable localized source of noise (Urick 1984).
Aperture	The average distance between hydrophones in a localizing cluster.
AVADECAF	Assessing the Viability of Density Estimation of Cetaceans from Acoustic Fixed sensors.
Availability bias	In a distance sampling context, missing animals at the point or on the line is generally attributed to availability bias, i.e. animals that are not available to be detected. In the context of PAM systems, the equivalent of this would be if animals were not producing cues when directly on the point. This availability bias is accounted for via the cue production rate which takes into account the proportion of time that the animal does not produce any cues.
Background noise	All acoustic sound detected in the environment at a time, including all sound in the ocean, and excluding the signal of interest, system noise, electrical noise and self-noise.
Beamwidth	The width of the beam in which a sound is generated by a marine mammal. Classically from studies of bat and odontocete echolocation clicks. Beamwidth is commonly expressed as the angle within which the level is within 3 dB of that at the center of the beam.
Bit depth	The precision with which a digitizer can measure voltage changes.
Cluster	A group of >1 hydrophones set up to work together on a single node.
Coherent detection	Detecting a single vocalization on all or a significant subset of hydrophones within an array.
Cue	A marine mammal vocalization. Cues per time unit form a cue rate.
CV	Coefficient of Variation.
DCL	Detection, Classification and Localization.
DECAF	Density Estimation for Cetaceans from passive Acoustic Fixed sensors. The original project (2007-2011) where the PAM density estimation methods were developed, funded

Term	**Description**
	by NOPP, looking into density estimation of cetaceans from passive acoustic data.
Detection probability	The average probability of detecting an object of interest in the surveyed area, given that it was available to be detected in the surveyed area.
Detection range	The range at which an animal is detected on a PAM system.
Distance sampling	A method for assessing wildlife abundance that accounts for imperfect detectability of the animals using information about the distances to the observed objects – perpendicular distances from the line for line transects or radial distances from the point for point transects.
DTAG	Digital Acoustic Recording Tag.
E&P	Exploration and Production.
Effective survey area	The effective survey area is the area where there are, on average, as many animals undetected within the surveyed area, as those detected outside the surveyed area. This corresponds mathematically to the surveyed area (as defined in a density estimator) multiplied by the average probability of detection.
Electrical noise	Any electrical interference resulting from thermal noise and semiconductor junction noise as well as sources such as ground loops, which create a humming sound in electrical systems, or radio interference.
Flow noise	Component of self-noise that results from turbulence as water flows around a hydrophone.
FN	False negative (or missed) detection, sometimes also referred to as a “Type 2” Error.
FP	False positive detection - A signal misidentified during a detection and classification process as the signal of interest - sometimes also referred to as a “Type 1” Error.
HF	High-frequency, ranging from 15 kHz to 150 kHz.
JIP	Joint Industry Programme.
LF	Low-frequency, ranging from 10 Hz to 1 kHz.
Localization	The process of determining some positional information based on received acoustic data.
Lossy Compression Algorithm	A compression algorithm for acoustic data which does not allow raw data to be recovered. For example, converting from .wav files (raw) to X3 (a non-lossy algorithm) and then back to .wav will return exactly the same .wav file. It is not possible to do this with lossy algorithms such as .mp3. Essentially lossy algorithms degrade and lose some raw data.
Match uncertainty	The inability to determine which detections on different hydrophones come from the same source.
MF	Mid-frequency, ranging from 1 kHz to 15 kHz.
Multiplier	Any factor required to convert the estimated “object of interest” density into animal density. Group size and call production rate are examples of multipliers.

Term	Description
NBHF	Narrow Band High Frequency.
Narrow beam profile	A vocalization may be received with higher level if an animal is on-axis i.e. directly facing a receiver. The signal level relative to the on-axis level as a function of off-axis angle is termed the beam profile. A narrow beam profile occurs when the received level reduces rapidly as off-axis angle increases.
Node	A type of PAM system - either for localization/range estimation or 'detection only' (see details above)
Noise	Any energy which is not signal and can potentially interfere with the detection, classification, and localization of signals. Any sound that is not of interest / should not be detected. Noise can be mechanical, environmental or biological in nature. See section 3.2.3.1 for details
n-node	PAM nodes that can detect but cannot localize cues.
Omnidirectional	Receiving signals from or transmitting in all directions.
PAM	Passive Acoustic Monitoring.
p-node	PAM nodes that are capable of providing range estimates to vocalizing animals.
Perception bias	In the context of PAM systems, missed detections at the point or on the line are often the result of perception bias, i.e., when the animals are available for detection but they are still missed for some other reason due to perception of the signal. For example, due to the directionality of an animal's beam profile, which is captured through the parameter described by a_{perc}, referring to perception bias at the point.
Range estimation	The process by which range to an animal is obtained in Distance Sampling.
RMS	Root Mean Square.
SECR	Spatially Explicit Capture Re-capture.
Self-noise	Energy originating from the recording system itself.
Sensor	Another term for 'PAM system' or 'node (see n-node and p-node)'.
Spatial aliasing	Unreliable localizations due to incorrect temporal matches of a narrowband signal received at multiple receivers.
SNR	Signal-to-noise ratio.
System noise	The electrical noise which is an inherent part of the properly working system. A component of self-noise.
Target species	Species for which the monitoring needs to be conducted.
Target Motion Analysis	A localization method. This method is used to localize data collected on a hydrophone array which is moving at significantly faster than the target species. Multiple detections over a time are used to localize the position of an animal. The analysis algorithms assume the animal is stationary during the detection period which introduces significant

Term	Description
	localization errors. This is therefore a generally low accuracy localization method but widely used in towed array surveys and sufficient for distance estimations.
Total noise	The sum of all kinds of noise that can be sensed by a system, excluding the signal.
Transmission loss	Attenuation of the amplitude of a signal or cue passing between two points of a transmission path (here: animal to receiver for passive systems, and sender to reflector to receiver for active systems).
Vocalization	A specific type of sound produced by an animal (in the context of DECAF, see 'Cue' above).

Executive Summary

The objective of this JIP-funded project was to investigate how an array of fixed PAM sensors could be used to estimate cetacean density/abundance across the life cycle (30-50 years) and spatial extent of a 'typical' oil and gas fields. As part of such an exercise, we built off existing JIP (and other funded) reviews (e.g. Moretti, et al 2009 & Norris et al. 2010) before we reviewed the available knowledge on marine mammal species to determine their (current) suitability of PAM for density estimation (based on the state of the key building blocks needed to employ DECAF methods). This included key reviews of detection, classification and crucially localization (DCL) capabilities which have direct implications for employing DECAF methods.

A fundamental element of the project was a set of power analyses and simulation exercises to explore the feasibility and utility of implementing DECAF methods considering different PAM survey designs, species vocal characteristics, DCL capabilities, variable environments (and the role of error), using Marques et al. (2009) as the foundation for this study. This analysis also considered the effect on the bias and precision of density/abundance estimates when integrating auxiliary data sources (such as DTAG or other acoustic tag data) into a DECAF analysis and we provide recommendations for further development to improve effectiveness and accuracy of estimating marine mammal abundance using PAM methods. As part of this report and in support of the sensitivity analyses, we also explored the availability of suitable datasets for use in parameterizing (i.e. providing inputs to) the simulation tool.

We have explored the wide range of biological, practical and statistical elements that are comprised within a long-term fixed PAM deployment to estimate density. It is crucial to understand the vocalization behavior of the species of interest and we know enough to demonstrate that the signals produced by marine mammals vary a great deal between species (ranging in frequency, duration, source level, directionality etc.). There is a need to deepen our understanding of vocalization behavior (including how cue production rate varies seasonally, or between individual animals due to differences in sex, age or motivation) – and this is likely to continue to be improved with site-specific studies. Using multipliers from the scientific literature or which have been estimated from datasets collected from another study population (or even the same study population but at a different location or time) may result in biased and ultimately, incorrect results. In addition, there is a pressing need to understand (for a given species of interest) the effects of anthropogenic activity on cue production rate, as this will impact the entire density estimation process (cue rate is a key multiplier and incorrect multipliers result in incorrect estimates of density). To collect these data, the use of acoustic tags (e.g. DTAG, Acousonde) that can record the vocalizations of focal animals, can be extremely valuable but will require significant investment over the course of an E&P site life span.

Given that the vocalizations of some species are also known to vary over time and between regions, and this can dramatically affect the efficiency of some automatic detectors in any DCL effort. Misclassification of species that produce similar sounds remains a major limiting factor and requires further development to be dealt with, and the assessment of detector performance is often hampered by a lack of ground-truthed test data (i.e. recordings for which the species and circumstances of vocalization are sufficiently well-understood). Different call types and different background noise conditions require different detection algorithms, so multiple detectors and classifiers may be required to process a single dataset. A detector / classifier is characterized by how efficiently it correctly identifies calls and by its false positive rate. While it is generally not necessary to have a perfectly efficient detector, or to reduce the false positive rate to zero, it is important to be able to quantify them and minimize their variance. Recordings from static or towed PAM systems, cross-referenced with visual identification of species have also proved invaluable, as these inform DCL efforts by enabling the development of call classification algorithms.

To localize vocal animals, the frequency, source level, the directionality of a vocalization, and diving behavior of a species, will dictate the methodological approach. The optimum design of a *localizing hydrophone array* is therefore species-specific. In addition, even for the optimal localizing array, it is unlikely that the probability of localization is the same as the probability of detection, especially for directional vocalizations (it may not be possible to localize all animals that are detected). Thus, in a distance sampling framework, to apply the probability of localization to 'n-nodes' requires an understanding of how to convert a probability of localization to a probability of detection. This process is, at present, poorly understood. There are two potential pitfalls in localization which directly affect DECAF: 1) Error in localization increases significantly with range from a hydrophone array and is roughly proportional to the spacing between hydrophones (the array aperture size). Therefore, although an animal may be detectable on all or some hydrophones within an array, the aperture size of the array may mean that the animal cannot be localized accurately. 2) Some vocalizations are highly directional (i.e. the echolocation clicks of odontocetes). This can lead to a situation where an animal may be within the effective localization range of an array but is orientated in such a way that it does not ensonify enough of the hydrophones within the array to allow its position to be calculated.

As noted above, the development of the (**AVADECAF** ('Assessing the ViAbility of DECAF')) simulation tool is the first time that all the elements of DECAF-based approaches have been brought together in a simulation. This is the first time that we have been able to assess which components of a long-term fixed PAM survey, which aims to detect significant changes in animal density, are the most crucial to a monitoring effort. From this first application, it is clear that certain elements of the tool could be developed to become more complex (and ideas for future development are discussed below). However, these adjustments will be made much easier now that the simulation framework is in place. Furthermore, the use of sensitivity analysis alongside this framework can provide a roadmap to guide the direction in which research effort should now proceed. This project therefore represents a significant step forward in making long-term PAM monitoring efforts more focused and cost-effective.

Using the AVADECAF tool, more than 1,000 simulation scenarios, with many thousands of iterations and hundreds of millions of simulated animals in total were run to inform a large sensitivity analysis, enabling us to explore which elements of a PAM survey are or greatest importance. In advance of any simulations, an understanding of vocal behavior and particularly, of cue production rate was identified as a clear knowledge gap, which may be dealt with either by the use of acoustic tags or by an extensive program of visual-acoustic focal follow observations. The results of the sensitivity analysis showed that a higher ratio of p-nodes (localizing nodes) to n-nodes (detection-only nodes) improves the power of a survey. This is because the use of more p-node systems generates a larger number of detection functions across the study site (and updated to deal with temporal changes in the environment that might affect detection functions) meaning we can adequately capture the spatial and temporal heterogeneity that is likely to exist across a given area. If we do not adequately control for this heterogeneity then the power of a survey will be low, and this is a key trade-off in designing a fixed PAM survey.

We also identified the importance of using detectors and classifiers that have a moderate-to-low false positive rate – i.e. some false positives (i.e. detections when no animal is present) can be accommodated, but the power of the model decreases when using poor detectors / classifiers. For all variables, keeping the variances low resulted in increased survey power. Conversely, where the spread of data around mean estimates was large, the power of all surveys was decreased (irrespective of cue rate, perception bias or FP rate). Therefore, a premium must be put on robust data collection and analysis of these data. It should also be acknowledged that where variances are likely to be high (e.g. where cue rate is assessed for a range of individuals and produces a wide range of cue rates) then this can drive down the power of a survey to produce meaningful results. Research efforts must be directed to advance this area in order to improve the viability of DECAF approaches (see: Recommendations).

This document provides a comprehensive review of a wide range of factors. In order to summarize some of the key challenges of each of the components of DECAF, we've provide boxed summaries outlining the pitfalls associated with each element. The relevant boxed summaries can be found below:

Box Details	Page
Box A: Species vocalization behavior	Page 22
Box B: Detection and Classification	Page 39
Box C: Localization	Page 56
Box D: Flowchart - The possibilities and pitfalls of localization	Page 77
Box E: PAM systems for DECAF	Page 78
Box F: Using PAM for estimating density	Page 95
Box G: Sensitivity analyses – a cautionary note	Page 108
Box H: How many whales to attach DTAGs to: a separate study in power.	Page 115

We conclude the report by outlining practical recommendations for how the field of density estimation using PAM methods can be developed further. Increased engagement between offshore industries and scientists would ensure that suitable data collection and analytical systems are developed, and that the direction of DCL approaches addresses industry needs. In light of the results and the findings of the reviews, it seems that there are two key dependencies for any PAM survey (in terms of hardware) and those are the need for systems that have range estimation capacity (from which robust detection functions can be generated) and the *in situ* collection of auxiliary data to inform cue rate (a key element that was challenging to adequately capture in the simulation tool). We outline a series of steps to develop suitable and cost-effective PAM systems capable of providing range estimates to vocal animals, which is a critical requirement for monitoring changes in animal density. Whilst we provide recommendations for next steps in such an effort – it's important to highlight that collecting the suitable information to deliver density estimation on an E&P site will require significant effort, even for a single species. Importantly, we outline the 'before, during and after' research approach (including the collection of auxiliary data to inform cue rate, a key parameter) that should be adopted to advance this field and to fill in data gaps in key parameters that limit the utility of DECAF for long-term E&P developments.

1 Project framework and approach

This report was funded by the International Association of Oil & Gas Producers following the Request for Proposals (RFP) Number JIP III-15-03 "*Long-term Fixed Acoustic Monitoring of Marine Mammals throughout the Life Cycle of an Offshore exploration and production (E&P) Field Development*" from the Joint Industry Programme on E&P Sound and Marine Life - Phase III (henceforth 'JIP'), released on 25th March 2015.

The objectives outlined in the RFP were to critically assess the viability and value of designing and implementing a 'life of field' type 'demonstration' project (30-50 years) that uses fixed passive acoustic monitoring (**PAM**) to determine animal density based on methods developed by the **Density Estimation for Cetaceans from passive Acoustic Fixed sensors (DECAF)** project, an international collaborative effort that ran between 2007 – 2011 and was funded by the US National Oceanographic Projects Partnership**.**

The objective of this current JIP-funded project was to investigate how an array of fixed PAM sensors could be used to estimate cetacean density/abundance across the temporal life cycle and spatial extent of 'typical' oil and gas fields. As part of such an exercise we reviewed the available knowledge on marine mammal species to determine their (current) suitability of PAM for density estimation (based on the state of the key building blocks needed to employ DECAF methods). This included key reviews of **detection**, **classification** and **crucially localization** (**DCL**) capabilities which have direct implications for employing DECAF methods. A crucial element of the project was a set of power analyses and simulation exercises to explore the feasibility and utility of implementing DECAF methods considering different **PAM survey designs**, **species vocal characteristics**, **DCL capabilities, variable environments** (and the **role of error**) using Marques et al. (2009) as the foundation for this study. This analysis also considered the effect on the bias and precision of density/abundance estimates when integrating auxiliary data sources (such as DTAG data) into a DECAF analysis and we provide recommendations for further development to improve effectiveness and accuracy of estimating marine mammal abundance using PAM methods. **The development of this simulation tool represents the first time that each of the elements of DECAF have been integrated into a simulation setting** and – this novel ability to explore the sensitivities of DECAF modelling and the value of different elements of a PAM programme represents – **a significant step forward in making fixed PAM surveys more cost effective**. As part of this report and in support of the sensitivity analyses, we also explored the availability of suitable datasets for use in parameterizing (i.e. providing inputs to) the simulation tool.

In the course of delivering the above elements, it became clear that the bulk of DECAF research (and certainly PAM in general) has focused primarily on cetacean species and less so on pinnipeds. Whilst here we are focused more broadly on marine mammals, and thus considering if or how DECAF approaches might be applied to pinnipeds, it should be stated at the outset that the focus of research to date means this review is heavily-weighted towards cetaceans. However, much of what we present in subsequent sections can be applied to pinniped species for situations where there is sufficient knowledge and other more established methods of studying pinniped density/spatial usage are unavailable (e.g. haul-out counts, telemetry etc. – which we do not consider here).

In assessing the viability of fixed PAM for density estimation, we have explored many of the complexities of implementing DECAF approaches. In order to present those complexities here, we have focused primarily on single-species case studies – particularly when developing sensitivity analyses. We acknowledge that in most (if not all) real-world PAM scenarios, there will be a need to develop more complex 'nested arrays' which are suitable for collecting the required data to estimate density for multiple marine mammal species (i.e. each with different vocal characteristics, dive behaviors and temporal variations in occurrence etc.). However, given the infancy and complexities of this research area, we have focused on a simplistic array conditions as our starting point.

In order to deliver the above, SMRU Consulting teamed up with the Sea Mammal Research Unit (SMRU), the Centre for Research into Ecological and Environmental Modelling (CREEM) to form a team of highly experienced experts in the fields of acoustics, marine mammal biology and statistics (with two additional high level scientific advisors) to undertake a comprehensive review and assessment on the viability and value of the use of fixed PAM for determining marine mammal density over the life cycle of offshore field developments. This review of the capabilities, value and viabilities of existing and developing fixed PAM systems and the current state of biological knowledge required for implementing DECAF methodology led to recommendations on how to best set up a field study to advance the feasibility of DECAF approaches and identify knowledge gaps and areas for further development and research.

2 Introduction

2.1.1 Background

The exploration and production of energy resources such as oil and gas from offshore sites involves operating in marine ecosystems, which have varying levels of sensitivity as they are home to a range of marine organisms. The goal of the oil and gas industry and of industry regulators is to produce oil and gas efficiently while avoiding impacts and minimizing effects on marine life. Industries that operate in the marine environment are under increasing pressure to explore and assess the potential for their activities to affect marine life. Though the length of 'life cycle' and spatial extent of E&P sites varies between regions, these typically range between 25-40 (and sometimes 50) years and over 10-30 km^2 respectively. Despite significant interest and effort into researching the potential effects of oil and gas E&P on the marine environment, data gaps still exist – all of which add to uncertainty in the assessment of the potential risks of industry activities. One critical knowledge gap area is to identify and quantify changes in marine mammal density over life time (before, during and after) of offshore activities, such as, oil and gas field developments. Information, such as animal distribution, movement, habitat use, and any potential disturbance by E&P operations over time and will help to informing long-term risk assessments and focus management and mitigation efforts.

Marine mammals have traditionally been surveyed using a suite of visual survey methods. Of these, the most established for offshore populations uses distance sampling, specifically line-transect survey methods (Buckland et al. 2001a), although cliff top observations and mark-recapture (photo-identification) have been used for inshore and small populations. Jewell et al. (2012) and Kaschner et al. (2012) demonstrate that these surveys generally do not have the power to detect anything but the most catastrophic declines for most populations. One of the main drivers of the large uncertainties in abundance estimates from line transect surveys, which limit our ability to detect trends, is often the limited sample sizes of encountered animals during the survey.

PAM relies on detecting the sounds produced by animals. In the marine environment, the species group for which PAM is best developed is cetaceans (whales, dolphins and porpoises). Pinnipeds (seals, fur seals and sea lions) have also been studied using these approaches, though less is known about the underwater vocal behavior (with the possible exception of some polar species). Many marine mammal species are known to make sounds of some type which in general is considered to make them a strong candidate for PAM approaches (though there are additional complexities to be considered). Sound propagates extremely effectively underwater and many marine mammals are highly vocal. Significant amount of research coupled with recent advances in computing, sound capture hardware and specialist software (e.g. PAMGuard - (Gillespie et al. 2008); www.pamguard.org) have greatly enhanced the efficacy and cost effectiveness of PAM in contributing to monitoring objectives. For some species, passive acoustic line transect surveys can be conducted using towed hydrophones deployed from vessels (e.g., Barlow and Taylor (2005); Gillespie et al. (2005); Lewis et al. (2007)).

While such surveys can often be conducted from smaller and more cost effective vessels than visual line transect surveys, the sample size of encountered animals is still often limited by available vessel time. Fixed PAM installations allow for cost-effective long-term monitoring over a limited spatial area for extended intervals (e.g. months - years – decades). Most importantly, persistent deployments of multiple PAM sensors has the ability to dramatically increase sample size, and thereby increase the ability to detect trends in a given population over time (Gerrodette et al. 2011).

The ability to monitor marine mammal density across the life cycle of an E&P field has the potential to contribute to industry's ability to robustly assess the potential effects to animals of offshore operations. There is also the potential that fixed PAM methods can provide cost-effective monitoring approach before, during and after an industry activity and therefore contribute to assessments of environmental risk and management/mitigation plans, and help to minimize the risk of negative interactions with marine mammals. Therefore an important element, when considering the wider application of density estimation using PAM (or any other data) is what the regulatory or operational requirements for such data are. With respect to regulatory elements, these vary around the world and will be dictated by the region in which an E&P field is sited. Some regulatory systems are extremely species focused and site-specific and, for example, will require that a particularly vulnerable species is protected either throughout the year and its range, or for just a specific time-area window (for example, linked to the breeding season). Other regulatory approaches are protective of many different species to the same degree and therefore certain components discussed in this report will be less important. For example, identifying to a species level may be less important (than under a more stringent or focused regulations), and in such cases, some of the requirements of DCL for a given PAM survey can be relaxed.

It should be noted, however, that as with visual survey methods PAM as a tool is not without constraints: marine mammals vary enormously both in the characteristics of their vocalizations and in their vocal behavior (e.g. frequency range, source level, directionality of sounds, and especially temporal or seasonal variations in vocalization etc.). Such factors have the potential to affect the performance of PAM systems. The attribution of sounds to a target species also continues to be a challenging area and can severely limit our ability to assess rare species if their sounds are confused with more common species (Caillat et al. 2013). In addition, detection will be hampered by sound occurring in the same frequency band as marine mammal vocalizations. A wide variety of PAM systems exist with different deployment life spans and functional capabilities - therefore not all PAM systems are created equal. The data collected by PAM systems are either archival or delivered in real-time, the latter of which providing an opportunity for up-to-date feedback for E&P operations and for potential risk to be minimized. Fortunately, there are a number of useful systems.

Coupled with the advances in hardware and software for PAM, there have been a series of recent developments of statistical methods to design suitable surveys and analyze the collected acoustic detection data (Marques et al. 2013a). These methods provide robust density and/or abundance estimates, which are fundamental to understanding cetacean basic biology, the number of animals around an industry operation, and therefore may contribute to the assessment of the potential effects of marine industries on marine mammals. However many PAM systems are capable of only detecting species, with no further information regarding the source of the sound. In order to implement some DECAF approaches, being able to estimate the range to the vocalizing animal is fundamental and the absence of 'off-the-shelf' systems to deliver this functionality remains a key obstacle. Fortunately, this is also an area where some development is occurring, and we explore this further here.

The use of PAM for estimating density is underpinned by a number of key elements including the hardware capabilities of systems (and how they are configured and deployed), the species behavior (vocal and dive), current software available for DCL. This report is focused on assessing the viability of using DECAF and thus such subjects are viewed through that lens.

2.1.2 Density estimation using PAM

Most marine mammals routinely produce sounds for a variety of reasons, from feeding to social interaction. The notion that sounds produced by animals contain information to estimate animal density has become prevalent and is reviewed in Marques et al. (2013a).

In a fixed passive acoustic data density estimation exercise we need to convert a number of detected sounds into an estimate of animal density. To do so a number of fundamental questions must be answered. These can be separated in two sets, those relating to (1) survey design / field methods and those related to (2) statistical analysis.

Regarding survey design, we need to define whether fixed or moving platforms will be used, how to distribute sensors in the area of interest, and what will be the objects of interest to “count”, the animals themselves (e.g. Ward et al. (2012)), groups of animals (e.g. Moretti et al. (2010)) or individual sounds (e.g., Marques et al. (2009)). An individual sound is an example of an indirect “cue” that indicates animal presence (Buckland et al. 2001a). All these choices can have profound implications on the analysis process. Efforts should not be spared at this stage since, more often than not, data collected using deficient survey designs/field methods will not be able to be salvaged even with optimal analysis methods. We therefore advocate that appropriate consideration should be given to survey design and field methods before any data collection takes place and offer advice in this regard.

To estimate animal density from acoustic data, the data analysis stage involves quantifying the efficacy and reliability of the detection and classification system being used, particularly regarding false positive detections and the effective area surveyed (see Glossary & Sections 3 and 4 for definition of false positives and effective survey area, respectively). The effective area surveyed, a concept directly linked to the probability that a sensor will detect an object of interest (whether an animal, group of animals or cue), in a given area around it, is typically estimated by either distance sampling (Buckland et al. 2001a) or spatially explicit capture recapture (SECR) methods (e.g., Borchers (2012)), though other “non-standard” methods have also been developed (Marques et al. (2013a), and see Section 2.1.2.1 for an overview of all methods). Distance sampling and SECR require that additional information be collected about the objects of interest. Distance sampling requires horizontal ranges to the objects of interest to estimate the detection probability (Buckland et al. 2001a). The decrease in the number of detections as a function of distance can be related to the probability of detecting the objects of interest. SECR requires that the same acoustic event is detected (often the term “captured” is used) and can be identified and associated across multiple instruments. The spatial pattern in these detections will convey information about both detectability and density. SECR can also be extended by incorporating additional information such as signal arrival times or the bearings of the objects of interest to improve the precision of results (Kidney et al. (2016), Borchers et al. (2015), Stevenson et al. (2015), Stevenson (2016)). Further, in the case that animals themselves are not being counted directly, we need the required multipliers to convert density of detected objects, those being either groups or cues, to animal density. This will include either an understanding of individual sound production rate or the average group size and the group sound production rate. Since these multipliers are likely to vary over space and/or time, ideally these should be collected at the same time and place as the survey being conducted. If collecting concurrent multiplier and survey data is impossible, a potential future approach may be to use a model to relate the multipliers to available covariates and then predict the multiplier values for the conditions present while surveying. However, the development of such models is still an emerging research area and so collection of multiplier data during the survey is currently the preferred and recommended option. If any multiplier depends on animal density, then there is no alternative but to collect data about the multipliers alongside the survey data.

2.1.2.1 *PAM approaches to estimate detection probability*

There are a number of critical elements that affect the viability of a fixed PAM survey. In this section we discuss the process of estimating the probability of an animal being detected.

Fixed PAM surveys take place from static monitoring points. As animals will be missed within surveyed areas, there are several methods that can be used to estimate the probability of detection. As noted above, distance sampling, for example, is a widely-used, versatile method that has been widely applied to marine surveys (Buckland et al. 2001a, Buckland et al. 2015). Distance sampling has been applied to a wide range of visual and passive acoustic data to estimate the abundance of marine mammals (and many other taxa). Distance sampling data have been collected from a variety of marine surveying platforms including vessels and aircraft that follow survey lines (line transect sampling) and stationary monitoring points (point transect sampling). Mark-recapture is another standard abundance estimation method (Borchers et al. 2002) but has an alternative estimation framework to distance sampling, where the re-identification of detected objects enables probability of detection to be estimated. Mark recapture has been used with visual marine mammal survey data to estimate abundances (e.g. Cheney et al. (2014)) and for other marine taxa too (e.g. Dutton et al. (2005); Bradshaw et al. (2007)). A challenge of the utility of mark-recapture methods is that, although abundance can be estimated, it is non-trivial to then determine the size of the surveyed area, which is required to estimate animal density. However, a recent extension to the typical mark-recapture methodologies is that of SECR, which allows density, as well as abundance, to be inferred from the collected data. Despite being a relatively recent addition to the statistical toolbox, SECR is considered to be a “standard” abundance and density estimation approach given its genesis from mark-recapture and also distance sampling (Borchers 2012, Borchers et al. 2015).

SECR has been successfully applied to both visual (Pirotta et al. 2014) and passive acoustic (Marques et al. 2012) (Martin et al. 2013) data to estimate the abundance and density of marine mammals. In this project, we will only focus on standard analysis methods that can readily estimate density as well as abundance i.e., distance sampling and SECR.

The two approaches require different data to be collected. Distance sampling requires that horizontal range information is available for each object of interest. That is, the horizontal distance between the PAM sensor and the detected animal can be determined. In general, SECR analyses requires that individuals are re-identified, i.e. “re-captured” across surveying occasions and that the location of each “capture” event is recorded (note that this is not necessarily the same as needing to record the location of each detected animal). When performed in PAM surveys, the hydrophones are considered to be “traps” of known location, and the same acoustically encounter object of interest is re-captured simultaneously across multiple hydrophones. Therefore, a data requirement of SECR when used in a PAM context is that the object of interest can be matched across the PAM systems in an array, if it has been recorded by more than one sensor.

There are also a number of alternative non-standard approaches specifically developed for estimating density from PAM data, often utilizing auxiliary data. For example, a “trial-based” method implemented by Marques et al. (2009) used passive acoustic tag data (specifically DTAG data: Johnson and Tyack (2003)) to assess if clicks produced by the tagged animal were detected on fixed hydrophones in the area. The probability of detecting a beaked whale on a fixed hydrophone could then be modelled as a function of range (Marques et al. 2009). Another non-standard method uses a simulation that incorporates sound propagation models, data about the source level of the target signal, as well as information about the efficiency of the detection and classification system to estimate the probability of detection (Kusel et al. 2011). While these non-standard methods do not require as much empirical information about the detections i.e., horizontal ranges or matching signals across systems are not required, the disadvantage of these methods is that the reliance on auxiliary datasets often increases the number of assumptions made throughout the analysis. It is for this reason that we focus on distance sampling and SECR for the remainder of the report.

2.1.3 Considerations in how fixed PAM and DECAF methods could contribute to offshore industry needs

While fixed PAM has clearly been demonstrated to be a viable method of monitoring some species like beaked whales and sperm whales (*Physeter macrocephalus*) (e.g. Marques et al. (2009); Ward et al. (2012)), the wide variety of survey options outlined in the review by Marques et al. (2013a) informs us that there is no PAM monitoring panacea and indeed, that in some circumstances, there may not be a viable solution. In particular, our ability to derive useful information from PAM surveys will depend on:

- The species (or groups of species) of interest vocalizing in the area of interest at the time of interest
- Our ability to detect those vocalizations and to distinguish them from other sound sources, including vocalizations of other non-target species.
- Our ability to deploy sufficient numbers of acoustic sensors with adequate bandwidth & detection capabilities to detect and localize sounds from target species.
- Knowledge of the species vocal behavior in the area of interest at the time of interest
- An ability to estimate the probability of detecting those vocalizations (generally as a function of distance from the PAM system).
- The power of any approach used to detect trends in abundance will also be heavily dependent on the population density and the amplitude of natural fluctuations, with small sample sizes and wide natural fluctuations making it increasingly hard to detect trends in abundance. This will be a harder problem the rarer the species.

Given the complexities of the DECAF approach in designing suitable PAM surveys and analyzing fixed PAM data, there are a number of practical considerations for their employment across an offshore area of development. With the ultimate goal of robust density estimates to feed into assessments of possible long-term risk, the following elements are considered:

- Our understanding of the vocal behavior of species of interest (acoustic properties, behavioral context and seasonality) (section 3.1);
- Information on the current state of detection, classification and localization (DCL) capabilities (section 3.2 and 3.3)
- The currently available autonomous PAM systems and their capabilities and constraints (section 3.4)
- The development of a Power Analysis Tool to simulate the real PAM surveys (section 4.3)
- A review of the currently available datasets / studies / methods to parameterize DECAF approaches (in section 4)
- An assessment (via sensitivity analysis) of the power of fixed PAM surveys to detect changes in animal density through time (section 5)

Assessing the above elements allows for a review and assessment of the viability and value of fixed PAM deployments for density estimation (with consideration of the need for integrating auxiliary data collected via other platforms/methods).

3 Review of PAM components of DECAF

3.1 Species vocalization behavior

Box A: Why is this topic challenging?

Having an understanding of the sounds that marine mammals produce is an essential starting point for density estimation using PAM methods. In terms of PAM surveys, it is crucial to understand the vocalization behavior of the species of interest. This can start with the frequency range of the calls (to help choose a PAM system that is capable of detecting the frequency of interest). The signals produced by marine mammal vary a great deal between species (ranging in frequency, duration, source level, directionality etc.).

The current state of knowledge of vocalization behavior could be described as broad but not very deep, as we know a little about a wide range of species but only ~10-15% of marine mammal species have been studied sufficiently to provide the inputs that we require for density estimation.

Those inputs are (for PAM methods) detail knowledge of cue production rate and how that varies among individuals of different age/sex classes, as well as seasonally and under different conditions (e.g. different depths, foraging / breeding ground etc.). In order to collect these data, typically animals are tagged with acoustic tags (e.g. DTAGs, Acousonde) which can record vocalizations of the focal animal, or they are observed using a focal follow method whilst PAM systems are recording to cross reference visual sightings with acoustic detections to derive vocalization rate. Currently there are a limited number of species for which this has been achieved. Therefore data is needed to be collected on site for target species. Using multipliers from the literature or that have been estimated from datasets collected from another study population (or even the same study population but at a different location and/or time of year) may result in the application of the wrong multiplier and, ultimately, biased results. However, it should be noted that it may be the best available way forward for initial use until such times that additional information specific to the area and species of interest may be available.

A lot of data from deployed recording PAM systems, cross referenced with visual identification of species have proved invaluable as these can feed into DCL efforts to develop classifiers and/or algorithms.

Marine mammals spend most, or all, of their lives at sea, and for the majority of that time they are submerged. Sound propagates efficiently through water and marine mammals rely on the use of sound to communicate with conspecifics, for predator avoidance, to locate and capture prey, mate selection and social interactions (e.g. Akamatsu et al. (1994); Au et al. (2004); Goodson and Sturtivant (1996); Hastie et al. (2006); Janik (2000, 2009); Madsen et al. (2005a), Madsen et al. (2005b); Rendell and Whitehead (2004); Schulz et al. (2008)).

In terms of PAM surveys with the objective of estimating density, it is crucial to understand the vocalization behavior of the species of interest. The following paragraphs attempt to summarize some of the key considerations. Key features of vocalizations that may influence the probability of detection of a signal are the directionality, duration, bandwidth and source level (SL) (if known) of vocalizations. In addition, many species vary their vocalization behavior with season, time of day and with dive behavior (discussed further below). There may also be confounding variables to consider, like other species that co-occur with the species of interest whose vocalizations can affect our ability to accurately classify the species of interest and/or changes in vocalization behavior driven by responses to anthropogenic activities.

We also need to consider frequency range of the call of interest (to help choose a PAM system that is capable of detecting the frequency of interest). The signals produced by marine mammals vary a great deal between species. In terms of the vocalization frequency alone, they range from the infrasonic moans of blue whales (*Balaenoptera musculus*) around 10 Hz to the ultrasonic clicks produced by harbor porpoises (*Phocoena phocoena*) over >130 kHz.

There are very few marine mammal species for which source levels have been comprehensively characterized (for some species there are one or two estimates which are largely context specific), and the available measurements demonstrate wide variability within a single species. For example, Au et al. (1999) measured a peak-to-peak source level (SL) for harbor porpoise of 172 dB re 1 µPa @1 m SPL (noting most clicks were produced with SL between 155 and 170 dB re 1 µPa @1 m $SPL_{p\text{-}p}$, whereas Mohl and Andersen (1973) reported a peak value of 140 dB re 1 µPa @1 m SPL. It is expected that animals actively vary their SL, most likely in an adaptive manner. Like most birds and mammals tested, marine mammals often increase the SL of their calls as the backgrounds sound level increases (Tyack and Janik 2013). Recently, Linnenschmidt et al. (2012) reported a porpoise varying its SL between 145 and 175 dB re 1 µPa @1 m peak-to-peak SPL. At the other end of the size scale and frequency spectrum, the SL of blue whale calls have been reported to vary between 174 and 189 dB re 1 µP @1 m SPL (Samaran et al. (2010); Sirovic et al. (2007)). The most intense sounds recorded to date are the on-beam clicks of sperm whales with an on-axis SL of up to 236 dB re 1 µPa @1 m SPL (Mohl et al. 2003), but far lower off-axis levels of about 190 dB re 1 µPa @1 m SPL (Zimmer et al. 2005b).

As it relates to the estimation of density it is *theoretically* possible to estimate range to a vocalizing animal by back-calculating from the measured received level (RL) at a hydrophone (i.e. inferring range from RL and assumed SL). However, this is dependent on a large number of assumptions about the SL of the vocalization, the orientation of the animal with respect to the receiver and the acoustic propagation conditions between the animal and the receiver. As demonstrated above, there is a very broad variation in SL and this factor alone can lead to significant errors in range estimation. We cannot therefore, advocate this approach.

Another crucial factor in the use of PAM for estimating density is an understanding of the potential for perception bias to affect the density estimation process. Specifically, perception bias in the context of PAM surveys occurs when animals are present in the region and vocalizing (and are thereby available to be detected) – but are not detected. This could occur due to the beamwidth with which a species' vocalizations are produced. Beamwidth is a measure of the radiation pattern of a sound source and is indicative of the degree of directionality with which the signal is produced. Many marine mammal vocalizations, especially those high frequency clicks produced by odontocetes for echolocation, are produced 'directionally' in a narrow, forward-facing beam outside of which there is relatively little energy (e.g., Au et al. (2006a); Goodson and Sturtivant (1996); Zimmer et al. (2005a), Zimmer et al. (2005b)). This means that the detection probability will depend on the location of the PAM sensor in relation to the axis of the vocalizing animal. This has the potential to affect detection probability during a dive and therefore will also depend on the movements of the animal and the extent to which it moves its acoustic beam (usually by changing body orientation) and 'scans' the wider environment. Baleen whale calls are largely omnidirectional. Sperm whale clicks are not but the off-axis level is still high enough to enable detection at several kilometers. For such species we can assume that we are detecting (close to) all the signals produced by an individual when they are present at 'zero horizontal distance' from the node.

In addition, marine mammals vocalizations vary greatly in duration with some very short signals (a few microseconds) such as the echolocation clicks of toothed whales (odontocetes), and longer, tonal sounds – including the low frequency moans of baleen whales (>30 seconds). Many signals are frequency modulated within the duration of the vocalizations, which can provide important information for distinguishing between species and in recognizing and excluding noise from various anthropogenic sources. Therefore for designing long-term PAM deployments (e.g., years), the temporal patterns in cue production for different species of interest must also be considered. These temporal patterns can be driven by a large number of factors, such as behavior, age, sex, motivation, and thus vary with time of day, time of year, group dynamics, availability of prey etc. Variability in calling behavior has been documented for some species and some regions – for example fin whale (*Balaenoptera physalus*) and humpback whale (*Megaptera novaeangliae*) songs (Watkins et al. 1987) and some blue whale calls (Leroy et al. 2016) are known to be produced variably with the year (Vu et al. 2012). In contrast to the relative predictability of some of the species noted above, Matthews et al. (2001) indicated that vocalization rates from North Atlantic Right whales (*Eubalaena glacialis*) are highly variable and appear to depend on the behavioral state of the animal with periods of silence often lasting for long periods. There are also diel variations in production (e.g. bottlenose dolphins *Tursiops truncatus*; Branstetter et al. (2012)). Diel variations have been observed in oceanic dolphins which are commonly detected acoustically around dawn and dusk (di Sciara and Gordon 1997), which could be indicative of increased feeding effort at night or changes in social or group cohesion signals.

Whereas the vocalizations and vocal production rate of some baleen whale species varies seasonally, with some animals calling predominantly during migration and on the breeding grounds, small odontocete species are habitually vocal. Harbor porpoises have been demonstrated to vocalize nearly constantly – indeed Akamatsu et al. (2005) deployed an acoustic recording (i.e. sound detecting) tag on a free-ranging harbor porpoise and observed that breaks between bouts of vocalizations were rarely >20 seconds. Studies conducting focal follows of sperm whales have also shown highly regular vocalization patterns whereby the animals vocalize near continuously while undertaking long foraging dives and are then silent while resting at the surface for ~15 – 20 minutes. Tag attachments to beaked whales show similar dive/surface vocalization behavior to that of sperm whales, but silent intervals between dives may last for well over an hour (Tyack et al. 2006). Some beaked whale species are more likely than others to vocalize during the descent phase of their dives, to only begin clicking regularly at great depth, and remain largely silent during their ascent (e.g. Cuvier's (*Ziphius cavirostris*) and Blainville's (*Mesoplodon densirostris*) beaked whales; Tyack et al. (2006)). Other beaked whale species however, begin clicking from the start of their dives – (e.g. Baird's (*Berardius bairdii*) beaked whales; Stimpert et al. (2014)). In addition animals increase their click production rate when foraging at the bottom of dives whilst homing in ('buzzing') on prey during foraging dives (e.g. Miller et al. (2004)). Because these echolocation clicks are so tied to foraging behavior, it also means they vary depending on the vertical distribution of their prey (which can also vary seasonally and spatially - Baird et al. (2005), Teloni et al. (2008), Benoit-Bird et al. (2016)).

Large seasonal variations in vocal behaviors are usually associated with seasonal breeding. In many (possibly all) baleen whales, males are much more vocal during the breeding season than at other times of year (in the context of DECAF, this means the acoustic cue production rate may only be indicative the density of a single sex). In some species, such as the humpback whale, long complex vocalization patterns or songs are only produced by males most commonly during the breeding season (but also on feeding grounds and during migrations). In addition, for some migratory species there are also likely to be geographic variations in vocalizations, both in terms of structure (i.e. how the call sounds and how identifiable it is) and call production rate. In theory as animals move towards breeding areas, one would expect that the rate of production of calls associated with reproduction would increase.

Given the scope of this study, it is important to consider how acoustic behavior may be affected by human activities. Animals may become less vocal and/or animals producing directional signals may be more likely to be oriented away from the disturbance sources. These behavioral changes would reduce detection probability. Alternatively, detection rates could increase if, for example, echolocating animals investigate this new presence or animals vocalize in response to it (e.g. Rendell and Gordon (1999)). Blackwell et al. (2015) measured bowhead whale (*Balaena mysticetus*) call rates using bottom-mounted recorders and observed that vocalization rates increased when pulses from seismic surveys were just detectable (though above another RL threshold animals went silent and therefore became undetectable). In a similar study, Di Iorio and Clark (2010) monitored call rates of blue whales using bottom mounted recorders. They found that call rates were elevated on days when airgun and/or sparker sound sources were also detected.

The tendency of animals to increase the SL of their calls in increased background sound conditions (known as 'the Lombard effect') (e.g. Tyack and Janik (2013), Holt et al. (2009)) may moderate the lower probability of detecting the calls if human activities increase the background sound level and therefore lower the signal-to-noise ratio (SNR) of animal vocalizations. However, this theory would need further investigation in order to fully understand the potential for this to occur (and the direction in which it could bias the density estimation processes).

There remains a paucity of information about the vocal repertoire of a number of marine mammal species. In many cases it is practical (and necessary) to group species together by their morphology and life history. The grouping of marine mammal species according to their vocal behavior was explored by Verfuss et al. (2016). We make use of those same groupings here (shown in Table 1).

Table 1 - Categorization of marine mammal species and species groups for the consideration of PAM surveys (and vocalization behavior). Note this table differs from that presented by Verfuss et al (2016) as non-marine mammals and River dolphin species have been removed. N.B. This table is not designed to be complete but provides some useful example groupings.

Category	Vocalization characteristics
Blue and Fin whales	Males produce powerful stereotyped low frequency calls (<30 Hz) in the breeding season. Females much less vocal or mute. Low call rates.
Humpback, Right and Bowhead Whales	Vocalize in the mid to low frequencies. Males more vocal in the breeding season. Both sexes also produce other vocalizations year-round. Some are extremely loud and characteristic, such as Pacific humpback feeding screams or right whale gunshots. Cue rate is seasonally variable and gender specific and overall moderate to low.
Other balaenopterids	Minke whale "boing" vocalizations with most energy between 1 and 5 kHz. Vocalization probably seasonally and sex specific.
Sperm whales	Powerful signals in the mid to high kHz band. Almost continuously vocal for most of their dives. Sporadic social calls produced when resting/socializing at the sea surface.
Beaked whales	Characteristic narrow band high frequency clicks with a distinctive frequency upsweep in the clicks (when observed on axis). Moderately powerful at source but highly directional and mostly produced at considerable depths.
Blackfish / Other Dolphins including pilot whales	Powerful click signals over a broad frequency range extending to ultrasonic frequencies. Directionality depends on the signal – with directionality of clicks likely similar to beaked whales. Highly variable call rates including silent periods.
Kogia (Pygmy and Dwarf Sperm Whales)	Narrow band high frequency clicks, similar to those of porpoises. Vocalizations usually made at substantial depths, likely to be highly directional.
Porpoises / *Cephalorhynchus*	High vocalization rates. Highly directional, narrow band very high frequency clicks centered at around 130 kHz. Source levels relatively low.
Pinnipeds	Some highly vocal, especially in breeding season, other species rarely vocalize. Mid to high frequency range calls.

3.1.1 Summary of the current state of knowledge of species vocalization behavior

The following section attempts to provide a high-level overview of marine mammal vocalization behavior that is relevant to implementing DECAF methods. The information presented here is by no means an exhaustive summary of each species vocalization behaviors. This topic is covered extensively in the literature and has previous reviews have comprehensively summarized the available literature at the time (e.g.: Richardson (1995), Au and Hastings (2008), Zimmer (2011), Todd et al. (2015) etc.). However, the intention is to highlight the current state of knowledge for the species, covering key questions of whether the species group vocal behavior is well understood, whether it varies significantly between age/sex classes, whether or not the species has been acoustically tagged etc. These are the knowledge areas which are of particular interest given our objective of assessing the viability of DECAF for different marine mammal species. A summary table is shown below (Table 2).

Table 2 - Summary of the vocalization parameter data available in the literature for various species of marine mammal. Source level is a measure of how loud the sound produced by the animal is. Directionality is a measure of how directional the sound is - sounds can be highly directional with most power emitted from the animal in one direction or omnidirectional and emitted more broadly in directions around the animal. Cue rate is the number of vocalizations produced per unit of time. Seasonal refers to where there is evidence that vocalization types are produced seasonally (e.g. during the breeding season). Sex Specific refers to whether or not there is evidence that the vocalizations of this species/group are produced by one specific sex (e.g.: males produce vocalizations during courtship).

	Frequency range	**Source level**	**Direction-ality**	**Cue rate**	**Acoustic Tag data**	**Seasonal**	**Sex Specific**
Blue & Fin Whales	Y	Y		Y	Y	Y	Y
Humpback, Right & Bowhead Whales	Y	Y		Y	Y	Y	Y
Other balaenopterids	Y	Y		Y	Y	Y	Y
Sperm Whales	Y	Y	Y	Y	Y		Y
Beaked Whales	Y	Y	Y	Y	Y		
Blackfish & Other Dolphins	Y	Y		Y	Y		
***Kogia* spp.**	Y	Y	Y				
Porpoise & *Cephalorhynchus* spp. Dolphins	Y	Y	Y		Y		
Pinnipeds	Y	Y			Y	Y	Y

It is important to consider how information on vocalization behavior is collected. Below we summarize studies using acoustic tags in several places in the following pages but they (and their pros and cons) have not been discussed as this is out of scope for this report. But tag data provide cue production rates from the head of the animal. Other results discussed come from tended and untended recorders (i.e. with or without supporting visual observations) or from towed arrays. Such studies provide information on rates of detected calls or vocal activity. If such studies are combined with focal follows then they might be able to generate cue rates, but otherwise their utility is limited in DECAF (in terms of providing estimates of cue rate).

3.1.1.1 *Blue and Fin whales*

Both blue and fin whales have been successfully tagged with acoustic tags in several studies, and have been well studied in fixed PAM studies over the past few decades. As a result, there is a strong knowledge base about the vocalization behavior for these species, making them a good candidate species for attempting to estimate density using PAM data. PAM data have previously been used to detect, classify and localize fin whales using the 20-Hz call; however a lack of cue production rate information has prevented previous efforts to estimate density from PAM data using distance sampling (e.g. Matias and Harris (2015); Harris et al. (2013)).

Blue and fin whales produce powerful stereotyped low frequency (<30 Hz), high intensity (>180 dB re: 1 µPa at 1m SPL) and lengthy calls (up to 20 sec) and many studies report females to be much less vocal than males in both species (e.g. McDonald et al. (2001); Croll et al. (2002); Oleson et al. (2007)). These low frequency calls propagate well underwater so the potential detection range is considerable (tens to hundreds of km). Both blue and fin whale calls are omnidirectional, with a directivity index score of 0 dB (e.g. Zimmer (2011) which means that PAM systems should be able to detect their calls irrespective of the orientation of the animal in relation to the PAM system, therefore avoiding the perception bias issue when applying distance sampling methods to estimate density from PAM data.

Studies have shown that unlike some other baleen whale species, blue whales vocalize year round, though seasonal peaks in calling activity has been demonstrated in some studies (e.g. Stafford et al. (2001); Širović et al. (2004)). It has also been shown that blue whales calling activity is higher at night which is thought to be linked to higher prey densities near the surface at night (at least in the eastern tropical Pacific, Stafford et al. (2005)). Blue whales produce several different vocalizations but studies have focused primarily on three call types: low-frequency calls (AB calls), calls with decreasing frequency (downsweep D calls) and highly variable amplitude or frequency-modulated calls. While these call types are typical of blue whales worldwide, the frequency, duration and inter-call interval has been shown to differ between geographic regions (e.g. Mellinger and Clark (2003); McDonald et al. (2006)), which has important implications for any DCL efforts.

Blue whales have successfully been tagged with acoustic tags which has provided detailed information on call types, call frequencies, cue production rates, seasonality and sex differences in vocalizations (e.g. Oleson et al. (2007); Burtenshaw et al. (2004); Stimpert et al. (2015)). For example, Oleson et al. (2007) report on data from 38 blue whales tagged with acoustic tags. In this study only one third of the tagged animals vocalized and there were clear differences between the sexes and with behavioral state. Overall males were more vocal than females and some call types, such as the AB calls, were produced exclusively by males and usually by solitary travelling individuals. Downsweep D calls were made by both sexes and were associated with feeding behaviors. Call production rates were quite variable across the study region (off California) ranging from ~4 to 43 per hour and most calls were produced whilst animals were at shallower depths in the water column and more often when travelling than during foraging bouts. Fin whale calls typically consist of short, repetitive downsweeps, with geographic variation in the frequency range and the inter-call interval between fin whale calls in the Atlantic and Pacific (Thompson et al. (1992); Širović et al. (2004)). While fin whales produce a variety of vocalizations, their 20-Hz call is the most studied and is described as being around 1 second in duration, sweeping downwards between 30 and 15 Hz (e.g. Sirovic et al. (2007); Stafford et al. (2009); Nieukirk et al. (2012)). Seasonal variations in fin whale calling has been documented; for example, fin whale calls in the Antarctic were found to be seasonal with calls detected between February and June (Širović et al. 2004) where as in the Gulf of California they are detected from late summer to late spring (Thompson et al. 1992) though these seasonal differences may be driven by insufficient sampling effort year round.

3.1.1.2 *Humpback, Right and Bowhead whales*

Both humpback and right whales have been successfully tagged with acoustic tags in several studies, and so there are good data on the vocalization behavior for these species, which should make them suitable candidates to calculate density estimation from PAM data. By contrast, while bowhead whales have been tagged with DTAGs, there is no data available in the literature on their vocalization behavior from these tag deployments and in general there is a lack of sufficient data with which one could use to estimate density from PAM data for this species.

Humpback, Right and Bowhead Whales vocalize in the mid to low frequencies (30 Hz to 10 kHz). Males are more vocal in the breeding season when many produce complex songs or mating calls. Both sexes also produce other vocalizations at all times of the year. Some of these calls are higher amplitude and characteristic, such as Pacific humpback feeding calls (e.g. D'Vincent et al. (1985)) or right whale gunshots (e.g. Laurinolli et al. (2003); Parks et al. (2005), Parks et al. (2012)). The cue rate has been shown to be seasonally variable, gender specific and varies with group size.

Male humpback whales produce highly stereotyped, repetitive songs as a breeding display. These songs are long and complex and consist of various song elements. Payne and McVay (1971) first described the structure of humpback whale song as an ascending series of units (notes), phrases, themes and songs with units varying in frequency between 30 Hz to over 10 kHz. It has also been shown that there are high frequency components to the songs too with harmonics reaching >24 kHz (Au et al. 2006b). Humpback whale songs have been shown to vary temporally and geographically (e.g. Cerchio et al. (2001); Garland et al. (2011)) and in response to anthropogenic noise (e.g. Miller et al. (2000); Risch et al. (2012)). While humpback whale songs are traditionally heard at their breeding grounds, they have also been recorded singing along their migration routes and at feeding grounds (Vu et al. 2012). Humpback whale songs are long and continuous (a song session can last for several hours) and, given the large distances over which they can be heard, if multiple animals are singing then a chorus is created which makes it difficult to localize separate singing individuals (Marques et al., 2009). Other non-song humpback whale vocalizations include social sounds such as grunts, pulses, moans and shrieks which occur when humpback whales are present in a group; these social vocalizations increase overall with group size while the detection rate per individual remains constant (e.g. Silber (1986); Dunlop et al. (2010); Dunlop et al. (2013)).

Right whales produce a variety of vocalization types including screams, gunshots, blows, upcalls, warbles, and downcalls. The predominant vocalization type varies spatially and temporally; for example the most common call type in the Bay of Fundy, Canada was the scream call (Parks and Tyack 2005) with gunshot calls being predominantly detected in the mid-summer and early autumn in the North Atlantic which is when the right whale breeding season is presumed to occur (Van Parijs et al. 2009). Bouts of gunshot vocalizations can extend to several hours in duration (Parks et al. 2012) and are produced exclusively at or near the surface (Parks and Tyack 2005). Call type varies by sex, with scream calls being produced predominantly by females while gunshots and warbles are produced predominantly by males (Parks and Tyack 2005). Cue rates are highly variable; for example, Parks et al. (2011) tagged 35 right whales with acoustic tags (DTAGs) and found that over half of the tags recorded no calls at all. Those that did record calls produced an average of 6.4 calls per hour but were highly variable, ranging as high as 200 per hour. Call rates also vary with group size, with single whales having the lowest call rates (0-10 per hour), the call rate for small groups (2-10 individuals) was ~60 per hour while the largest groups with more than 10 individuals, had detection rates up to 700 per hour (Matthews et al. 2001). From larger groups and at night, when vocalizations were produced, overall silent periods were typically ~<10 mins, but individuals might remain silent for 120-150 minutes (e.g. Matthews et al. (2001); Parks et al. (2011)).

Bowhead whales produce a large variety of sounds (including songs, moans, pulses, high frequency calls and gargles) during their spring migration, fall migration, and summer feeding, most of which are below 300 Hz and made up of simple frequency-modulated structures (Clark and Johnson 1984). Source levels for calls and songs during the spring migration are estimated to be up to 189 dB re 1 µPa-m (Cummings and Holliday, 1985), and bowhead calls can be detected out to over 10 km away (Cummings and Holliday, 1985). Call rates fluctuate, but can be very high for a baleen whale with over 2,000 calls detected a day (up to 600 detections/h) at locations along their migration route (e.g. the Alaskan Beaufort Sea, Blackwell et al. (2007)); however it was not known how many whales were present to produce these high call rates. The calling behavior of bowhead whales has been shown to change in response to airgun sounds, where animals near to the airgun sound source cease calling (Blackwell et al. 2013).

3.1.1.3 *Other balaenopterids*

There are almost no data available from acoustic tags deployed on other balaenopterid species such as minke (*Balaenoptera acutorostrata*), Bryde's (*Balaenoptera brydei*), Omura's (*Balaenoptera omurai*) and sei whales (*Balaenoptera borealis*) and, where there are data, reports are limited to a couple of individuals or studies and descriptions of vocalizations are limited to specific vocalization types (e.g. minke whale boings and bio-duck sounds – see details below). Data obtained from other methods such as static PAM devices have shown that minke whale vocalizations vary greatly with region and so caution should be taken when attempting to employ DECAF-based methods for estimating animal density; it should be ensured that the data are appropriate to the specific region of interest. Despite a general lack of vocalization data for most of the species in this group, it has been shown that using the SECR method is a suitable approach for estimating minke whale call density from their boing sounds (e.g. Marques et al. (2012); Martin et al. (2013)) and discussed later in more detail (section 4.1.2).

Minke whales have been shown to produce a range of vocalization types including: low-frequency downsweeps, pulse trains, grunts, ratchets, single pulses, boings, bio-duck and higher frequency clicks (e.g. Rankin and Barlow (2005)). The minke whale "boing" vocalization is one of the most distinctive and stereotyped vocalization. It is characterized by a brief pulse followed by a longer, frequency- and amplitude-modulated segment with most energy between 1 and 5 kHz (e.g. Rankin and Barlow (2005); Oswald et al. (2011); Ou et al. (2012)). Studies have shown that vocalization rates vary seasonally and geographically. For example, minke whale pulse trains in Massachusetts Bay, USA were only detected between August to November and were predominantly detected at night (Risch et al. 2013). Minke whale boing vocalizations differ in detection rate and duration between the eastern and central North Pacific regions (e.g. Rankin and Barlow (2005); Oswald et al. (2008)). Interestingly, it has been shown that boings increase with increasing number of animals present in the area Thompson and Friedl (1982). Martin et al. (2013) report a cue rate of 6 boings/hour based on a small acoustic sample with a focal follow of a single animal, though the authors note that this individual whale may not be representative of the population and that it may be that only male whales vocalize in the breeding season, which would mean that there is a proportion of the population which cannot be detected using PAM.

Minke whales are notoriously difficult to tag; however one study (Risch et al. 2014) reports on the successful tagging on two minke whales which provided data on two of their vocalization types. These two tags reported low vocalization rates, with most calls produced when the animal was close to the surface. The 'bio-duck' sound was produced just before a feeding dive and consisted of a series of pulses at regular intervals with frequencies between 146-165 Hz and low frequency downsweeps were recorded with lower peak frequencies of 83.1 Hz (Risch et al. 2014). To our knowledge, this is the only vocalization data available from acoustic tags deployed on minke whales. Another group have been successful in tagging them off Iceland (Miller pers. comm.) – but the data are so far unavailable.

It has been noted previously that some minke whale boing classifiers can be sensitive to humpback whale songs which may cause problems for species identification and false positive rates in regions where these two species overlap spatially and temporally (i.e. Hawaii in the winter months, Ou et al. (2012)).

Bryde's whales produce a variety of vocalizations including calls, moans, growls and pulses. Their calls are low-frequency and vary in frequency and duration and are similar to the calls of other balaenopterid species, with the key difference that they tend to produce shorter vocalizations than many other balaenopterids (Heimlich et al.

2005) – though a recent paper indicates they may also produce longer duration calls (Rice et al. 2014). Calls vary slightly with region but frequencies are similar around 60 Hz or less at three Bryde's whale habitats: the Eastern Tropical Pacific, the southern Caribbean, and the northwest Pacific near the coast of Japan (Oleson et al. 2003) and studies have concluded that while Bryde's whale vocalizations vary slightly with geographic region, there are some universal characteristics that allows them to be easily identified to species (Heimlich et al. 2005). Bryde's whales have been tagged in the Hauraki Gulf where down-sweep calls were the most identifiable call type recorded. This study highlighted that down-sweep call types are produced rarely and occurs in bouts when the whales are at shallow depths (Constantine et al. 2012).

Until recently, Omura's whales were classified as Bryde's whales and have only recently been identified as a separate species. Therefore, very little information on them is available in the published literature, with most of the data being provided from carcasses obtained from whaling or strandings. Recently, the first acoustic measurements of confirmed Omura's whales have been documented (Cerchio et al. 2015), and aside from these data, there are no other data available in the literature for this species. Acoustic recordings have shown that Omura's whales produce stereotyped low frequency calls which are described as long duration calls (9.2 seconds) made up of rhythmically repetitive sequences that are broadband in nature (14.9-52.9 Hz) with a peak frequency of 36.1 Hz (Cerchio et al. 2015). These calls are thought to be associated with breeding behavior based on concurrent visual sightings of behavior, however, it is not known if the sounds are sex specific to males as with some other balaenopterid songs (e.g. humpback whales), nor is it known if these calls are produced out with the breeding season or region.

There have been few studies where sei whale vocalizations have been verified with visual sightings and no acoustic tags have been deployed on this species, therefore, data on their acoustic behavior is limited. Data obtained from combined visual and acoustic data near the Auckland Islands has shown that sei whales produce low frequency (34-87 Hz) up- and down-sweep calls when at their summer feeding grounds, that are short in duration (up to 1.7 seconds) with inter-call-intervals of up to 13.5 seconds (e.g. Calderan et al. (2014)). Similar data has been shown for sei whales east of Cape Cod where calls swept from 82 to 34 Hz over 1.4 seconds (Baumgartner and Fratantoni 2008). This study cautions that a lack of sei whale call detections does not mean that no sei whales are present as they found that in 17 (29%) of the 59 hourly periods with no acoustic detections, sei whales were confirmed as present from visual observations.

3.1.1.4 *Sperm whales*

Sperm whales have been successfully tagged with acoustic tags in several studies and so their vocalization behavior is relatively well documented. In particular, their click vocalizations are well suited to allow for density estimation as they are easily identified and classified and data on the necessary vocalization parameters such as frequency and cue rate are available in the literature. Studies have previously used sperm whale echolocation clicks recorded on fixed instruments to estimate the density of animals (e.g. Ward et al. (2012), which confirms their suitability for DECAF work.

Sperm whales make long dives lasting ~45 mins. During these dives they produce extremely powerful signals (235 dB re 1 µPa p-p, Madsen et al. (2002a), Madsen et al. (2002b), Mohl et al. (2003)) that are highly directional (directivity index of 26.7 dB, Zimmer et al. (2005b)) and are in the mid to high frequency band (2-25 kHz, Madsen et al. (2002a), Madsen et al. (2002b)). The highly directional characteristics of sperm whale clicks (e.g. Møhl et al. (2000)) would normally present an issue with perception bias for DECAF based methods to estimate density from PAM data, however, their clicks also contain less-directional components (broad backward-directed beam with a directivity index of 7.4 dB, Zimmer et al. (2005b)) and, given that sperm whale clicks have such a high source level (some of the highest biologically produced source levels recorded), they propagate well underwater, which allows for the large, omnidirectional detection ranges of several to tens of kilometers for deep stationary hydrophones in good conditions. Good data are available on sperm whales' acoustic behavior, in particular recently from acoustic tags (DTAGs) (e.g. Fais (2014); Miller et al. (2004); Watwood et al. (2006)). Different click types (echolocation click, buzzes, codas and the slow click) are produced by sperm whales and these serve different functions. The slow click type produced primarily by males at the feeding grounds and when in proximity to female groups (occurring at lower latitudes) is likely to be linked to long range communication between males, while echolocation clicks and buzzes are produced by both sexes and are associated with foraging and prey capture (Oliveira et al. 2013). The cue rate for sperm whale echolocation buzzes has been shown to differ with dive depth, where inter-buzz-intervals were longer during shallow dives compared to medium and long dives and they click almost constantly throughout the descent, foraging and, in some areas, ascent phases of their dives (Fais et al. 2015).

The social sounds produced by sperm whales include 'coda clicks' which are composed of a series of pulses where the delay between pulses is linked to animal size, and are produced by females and their offspring possibly to communicate identity and activity state to maintain social cohesion (Oliveira et al. 2016). Sperm whales have been shown to respond to anthropogenic sounds by modifying their vocalizations. For example, in response to naval sonar, sperm whales increase their codas and slow click social sound production rates (Curé et al. 2016).

3.1.1.5 *Beaked whales*

While certain species of beaked whale have been well studied (i.e.: Cuvier's, Blainville's beaked whales and Northern bottlenose whales), there is very little information on several other species such as Deraniyagala's (*Mesoplodon hotaula*), Stejner's (*Mesoplodon stejnegeri*) and Gervais' (*Mesoplodon europaeus*) beaked whales (Todd et al. 2015). For this reason, PAM data can be used to estimate animal densities for the most studied species for which vocalization parameter data are available, but caution should be taken when applying this approach to other, less well studied beaked whale species. Previous studies have shown the applicability of Blainville's beaked whale click vocalizations for density estimation (e.g. Ward et al. (2008); DiMarzio et al. (2008); Marques et al. (2009); Moretti et al. (2010); Ward et al. (2011)).

Beaked whales have been shown to produce a species-dependent wide range of vocalizations including clicks, calls, pulses and whistles. Beaked whales produce characteristic high frequency clicks (up to 48 roughly 25 – 70 kHz) with a distinctive frequency upsweep in the clicks (Shaffer et al. 2013). Clicks are moderately powerful at source (200-220 dB re 1 µPa at 1 m p-p), are highly directional and are mostly produced at considerable depths (Johnson et al. 2004). The clicks produced by different species of beaked whales such as Blainville's beaked whales, Cuvier's beaked whales, Gervais' beaked whales and the Northern bottlenose whales have similar characteristics but differ primarily in center frequency and duration (e.g. Gillespie et al. (2009); Stimpert et al. (2014)). For Cuvier's and Blainville's beaked whales, there are good data available from acoustic tags: these show long silent periods of shallow diving between deep foraging dives where echolocation takes place (Tyack et al. 2006), with vocal periods of around 28% for Cuvier's beaked whales and 17 to 19% for Blainville's beaked whales (Barlow et al. (2013); Arranz et al. (2011)). Blainville's beaked whales have been observed as being mostly silent during the descent and ascent phases of their dives and they rarely vocalize during shallow dives; for example, during their deep foraging dives they produce foraging clicks and buzzes, which accounts for approximately 18% of the time (Aguilar de Soto et al. 2012). Tagged whales produce approximately 3,500 echolocation clicks per dive (Madsen et al. 2014).

Detection ranges of several kilometers are possible for bottom-mounted hydrophones (e.g., Marques et al., 2009). Extensive work has also been done to model and measure detection probability for static bottom-mounted hydrophones and to estimate density from PAM surveys (e.g. Hildebrand et al. (2016); Gassmann et al. (2015); Kusel et al. (2011); Marques et al. (2009); Ward et al. (2011)).

Beaked whales have been shown to alter the clicking behavior in response to anthropogenic sounds. For example, Cuvier's beaked whales stop producing echolocation clicks when exposed to mid-frequency sonar (DeRuiter et al., 2013). A similar response was also obtained from an animal incidentally exposed to high level vessel noise (Aguilar et al. 2006).

3.1.1.6 *Blackfish and Other dolphins*

'Blackfish' is a colloquial grouping encapsulating the following species: beluga (*Delphinapterus leucas*), false killer (*Pseudorca crassidens*), killer (*Orcinus orca*), pilot (*Globicephala melas* & *G. macrorhynchus*), melon-headed (*Peponocephala electra*) and pygmy killer (*Feresa attenuata*) whales while 'other dolphins' refers all other delphinid species (with the exception of *Cephalorhynchus spp.* dolphins). Of these, there are some species for which vocalization parameter data has been well studied and that would be suitable species to attempt to estimate animal density with PAM data; these include killer whales and bottlenose dolphins. There is considerably less data available on many of the other blackfish and oceanic dolphin species, where most vocalization parameter data required for density estimation are lacking. Acoustic (DTAG) data are available for pilot whales, killer whales, bottlenose dolphins and Risso's dolphins (*Grampus griseus*).

All species within this group produce powerful signals over a broad frequency range with echolocation clicks extending to ultrasonic frequencies of >100 kHz. The lower frequency whistles are mainly omnidirectional and propagate well with ranges in good conditions of several kilometers. By contrast, the higher frequency clicks produced by this species group are generally highly directional (e.g. Au et al. (1995); Kyhn et al. (2010); Rasmussen et al. (2004); Wahlberg et al. (2011) and many more – see references within Todd et al. (2015)) which can present a challenge in terms of applying DECAF methods as an additional perception bias parameter may be required, to account for the fact that highly directional clicks may be not be detected by the instruments (even at close ranges). This problem may be mitigated somewhat by the typically large group sizes of these animals that result in cues being produced frequently in all directions.

Bottlenose dolphins are one of the most intensely studied species in this group and vocalization data are available from various regions globally although mostly for coastal dwelling animals: oceanic bottlenose dolphins could behave differently. Vocalization parameters are available for echolocation clicks as well as for tonal sounds such as whistles and squeals. These data have highlighted that bottlenose dolphin clicks and whistles vary geographically in terms of source levels, central frequencies, durations and sound production rate (e.g. Jones and Sayigh (2002); Hawkins (2010); May-Collado and Wartzok (2008)).

Risso's dolphins have been successfully tagged with acoustic tags and so information on cue types, frequency, cue rate and associated behavioral states are available (e.g. Arranz et al. (2016)). Both species of pilot whales have also been successfully tagged and the data obtained from these studies have shown that their vocal behavior is typical of biosonar foraging, with echolocation clicks interspersed with buzzes (Aguilar Soto et al. 2008). The data for short-finned pilot whales have also shown little vocalization during the initial descent phase of daytime deep dives, with regular clicks starting to be produced at depths around 300-400m during prey search and capture with buzzes mainly recorded at the deepest parts of the dives. In contrast, foraging occurs in deep and shallow dives at night-time and clicking is continuous throughout these dives.

This species group typically has high click production rates, though the rate of production of tonal communication signals is highly variable and often dependent on behavioral state. For example, mammal eating killer whales are known to employ "stealth" strategies while hunting, producing few and sporadic vocalizations. At these times their cue production rate would be very low compared to after a marine mammal kill or during surface active behaviors (e.g. Deecke et al. (2005)). Many dolphin species form large groups which will have much higher collective cue rates than individuals, making them easier to detect (but group size is all but impossible to accurately estimate). However, as a further confounding factor, many oceanic dolphins form multi species groups and differentiating between species is difficult. Classification of oceanic dolphin whistles and clicks to species level (if that is desired/required by regulators) can be challenging and would likely require classifiers optimized for different regions (Gillespie et al. (2013); Oswald et al. (2007))(see section 3.2). For example, the whistles of nine odontocete species in the eastern tropical Pacific overlap in characteristics and studies have found that some species, such as the bottlenose dolphin, false killer whale and pilot whale are the most distinctive while striped dolphins were hard to separate and classify to species level (Oswald et al. 2003). This means that collecting PAM data using automated detectors/classifiers for certain species may result in high false positive rates, particularly where they overlap spatially and temporally with other similarly vocalizing species.

3.1.1.7 Kogia *spp.*

There is almost no data available in the literature on the vocalization behavior of *Kogia* species. As such, there is little or no data available to be able to estimate animal density from PAM data. These two species have not been tagged with acoustic tags and information on their vocal behavior is generally lacking. For example, data on pygmy sperm whale (*Kogia breviceps*) vocalizations comes mainly from a study comprising of one stranded animal that was placed in a holding tank (Marten 2000). It is currently not possible to differentiate between the two species acoustically (Merkens et al. 2016).

Kogia (Pygmy, and Dwarf Sperm Whales *Kogia sima*) are considered to produce narrow band high frequency (NBHF) clicks, similar to those of porpoises. Pygmy and dwarf sperm whale clicks frequencies range between 60->200 kHz with peak frequencies of around 125 kHz with varying click duration and repetition rate (Marten 2000). In the wild these vocalizations are presumed to be made at substantial depths, and are likely to be highly directional.

3.1.1.8 *Porpoises and Cephalorhynchus spp. dolphins*

Porpoise vocalization parameters have been very well studied both in captivity and in the wild, and as such, they are well suited to attempt to estimate animal density from their vocalizations. In comparison, there is a lack of data for most species of *Cephalorhynchus* spp. dolphins and given the similarities in the narrow-band-high-frequency click behavior PAM systems may not be able to discriminate between different species without training of classifiers with local data. The main limitation of PAM surveys for this species group is that the narrow band high frequency clicks are only detectable out to a short distance, and then, due to the directionality of the clicks, only when the animals is oriented towards the PAM device.

Porpoises (and *Cephalorhynchus* spp. dolphins) have consistently high vocalization rates, producing highly directional, very high frequency clicks in a narrow frequency band, centered at around 130 kHz (Akamatsu et al. (2005); Villadsgaard et al. (2007)). Free ranging wild harbor porpoise have been shown to vocalize almost constantly (Akamatsu et al. 2005, Akamatsu et al. 2007) while in other regions it has been shown that click activity varies cyclically with higher click detections at night (e.g. Keenan et al. (2011)). Source levels are relatively low, and this combined with high rates of absorption at these frequencies limits detection range to several hundred meters (e.g. Leaper and Gordon (2012)). Porpoise clicks are highly distinctive and can be readily distinguished from the wideband clicks of most dolphins, blackfish and beaked whales.

Cephalorhynchus dolphins produce mainly NBHF clicks with some species also producing buzzes, broad-band clicks and whistles. The NBHF clicks of *Cephalorhynchus* dolphins show very similar characteristics between species (e.g. Kyhn et al. (2009); Götz et al. (2010); Leeney et al. (2011)). The whistles produced by Commerson's dolphins (*Cephalorhynchus commersonii*) have a frequency between 16 and 48 kHz and have been recorded in the presence of mothers and calves which suggests the whistles are produced for parental behaviors (Reyes Reyes et al. 2016). The broad-band clicks of Commerson's dolphins have a peak frequency of 44 kHz (well below the NBHF click frequencies) and are thought to play a role in adult communication as they are primarily produced during activity states such as resting, socializing and travelling (Reyes Reyes et al. 2016). Hector's dolphins (*Cephalorhynchus hectori*) produce ultrasonic clicks (peak frequency 129 kHz) with a high click rate which is used in the context of for foraging and communication; this species has not been shown to produce whistles or any other audible sounds (e.g. Dawson (1991); Kyhn et al. (2009)).

3.1.1.9 *Pinnipeds*

With the exception of males of certain species vocalizing during the breeding season, pinnipeds in general do not reliably vocalize. Because of this, data on vocalization parameters are limited to a few species and not all information necessary for DECAF-based methods is available for most species.

Males of some species of pinniped are highly vocal, especially in the breeding season, though other species rarely vocalize. It is the species that breed in water that tend to produce the majority of the vocalizations underwater as part of courtship and territorial signals (e.g. Van Parijs et al. (2003)). Therefore it is these species might be suitable for employing DECAF approaches – but of course such species are available to be studied in detail (e.g. census methods) whilst hauled-out on land, and so such methods would likely be favored over PAM. Pinniped vocalizations tend to be in the low-to-mid frequency range, with vocalization types ranging between lower frequency growls and grunts to mid frequency roars (overall frequency range approximately <1 kHz to 4 kHz). Harbor seals (*Phoca vitulina*) have been recorded producing a variety of underwater sounds including grunts, groans, creaks, bubbly growls and roars, which are all produced primarily by males and are thought to be produced in male-male competition and displays to females (Hanggi and Schusterman 1994). Ribbon seals (*Histriophoca fasciata*) produce a variety of sounds including roars, yells, hisses and grunts with frequencies between 0.07 and 17.5 kHz, though most sounds produced were at the lower frequency range between 0.2-1 kHz (Mizuguchi et al. 2016a). Since ribbon seals breed on ice, the function of their underwater vocalizations is unknown but may be linked to social interactions and territorial behavior (Mizuguchi et al. 2016a). Ringed seals (*Pusa hispida*) produce a variety of sounds that are linked to different behaviors; knocks are produced during aggressive encounters, yelps and barks are submissive signals and snorts are associated with reproductive displays (Mizuguchi et al. 2016b). Bearded seals (*Erignathus barbatus*) call mainly during the mating period where they produce long trills, thought to indicate male quality, with higher levels of vocal activity at night (Frouin-Mouy et al. 2016). For some species of pinniped, most vocalization data available come from captive studies which may not be representative of the vocalization behavior of wild animals.

3.2 Detection and Classification of Cetacean Vocalizations

Box B: Why is this topic challenging?

The volumes of data collected using modern acoustic monitoring equipment are too large for humans to listen to or browse by hand (for instance, viewing spectrogram displays). We are therefore increasingly reliant on automatic computerized systems which search for and extract calls of interest from acoustic data sets. Different call types and different background sound level conditions require different detection algorithms, so multiple detectors and classifiers may be required to process a single dataset. A detector / classifier is characterized by how efficiently it correctly identifies calls and by its false positive rate. While it is generally not necessary to have a perfectly efficient detector, or to reduce false positive rate to zero, it is important to be able to quantify them. Both vary with local background sound level conditions, so detector performance can vary with both time and location as conditions vary. Some types of background sound (e.g. uniform white noise) are unlikely to create false positive detections, but will reduce detection efficiency. Other (intermittent of transient) types of sound will both reduce efficiency and create false positives. Calls from some species are also known to vary over time or between regions which can also dramatically affect the efficiency of some types of detector. Mis-classification of species that produce similar sounds remains a major limiting factor. Characterization of detector performance is often hampered by a lack of ground-truthed test data.

3.2.1 Introduction

Hardware for collecting acoustic data is described in section 3.4. In this section we discuss how those data can be processed to extract the calls of particular species and also to extract other parameters such as detection (horizontal) range which may be required for density estimation.

The JIP has previously funded two projects that have reviewed the state of the fields of automatic detection, classification and localization (DCL) of marine mammal vocalizations (Moretti et al. 2009, Norris et al. 2010). These subject areas have direct relevance to long-term fixed acoustic monitoring and play a key role affecting our ability to estimate marine mammal density using PAM approaches. The areas of DCL have developed hand in hand with technology. As underwater acoustic equipment has developed, the need for automatic DCL techniques has increased, as we now are capable of readily collecting many terabytes of data, covering wider frequency ranges via improved hardware capabilities. In addition, as computer processing power has increased, more sophisticated DCL techniques have become possible. Without automation of the DCL process, the acoustic data volumes would make long-term monitoring resource, cost and time-prohibitive. However, it is important to point out that automation does not mean a complete lack of human involvement in the DCL process. At the very least, a human will commonly be needed to quantify the performance of DCL algorithms. More realistically, they will be needed at several steps in the DCL process to make decisions, guide the process and improve the accuracy of the algorithms, especially at the beginning of a new monitoring project.

Computer algorithms have the advantages of being fast and consistent in their decision processes but can become unstable in the presence of unexpected signals and usually require retraining for different datasets with different background sound characteristics and animal vocalization signal. Computers can also make accurate measurements from the data (such as of received signal levels or time of arrival differences). Humans on the other hand can perform inconsistently but are capable of dealing with unexpected signals and the human brain is considered to be by far the most adaptable pattern recognition instrument available. Indeed, all of us are familiar with the CAPTCHA, or “Completely Automated Public Turing test to tell Computers and Humans Apart” – those twisted letters with lines through them we have to type in to web sites to show that we are a human and not a computer (Von Ahn et al. 2008). A child can generally decode a CAPTCHA with ease whereas computer hackers struggle to decode them. In many PAM studies, a combination of human browsing and computer algorithm are used together with the computer performing an initial scan of data to find detections and a human checking either all or a subset of the detections to either exclude false detections or to provide a false positive error rate for use in the later stages of analysis.

While often discussed separately, Detection and Classification are inextricably linked. Indeed, some algorithms effectively perform detection and classification as a single step, other algorithms may run a highly efficient detector that has a high false positive rate, but then use a classifier to reject false calls and separate others by species (Moretti et al. 2009).

The objective of the following sections is to provide an overview of what Detectors and Classifiers attempt to achieve and which Detection and Classification algorithms are suitable for which types of sound. Our discussion is within the framework of the requirements for density estimation, possibly of multiple species within a single dataset. We include an overview of the key points of the two previous JIP funded reports, and provide an update on the fields of automatic detection and classification. Localization is covered separately in section 3.3.

3.2.2 Performance Metrics

In any detection system, be it human or computer based, two types of error can occur. False positives occur when the system reports a sound of interest as being present when in reality it was not. False negatives occur when animal sounds which were made are missed by the detection system. The concepts of False Negatives and Efficiency (described below) are inextricably linked, Efficiency being '1 – the rate of False Negatives' and the rate of False Negatives being '1 – Efficiency'. Generally, there is a direct trade-off between the rate of false positives (hereafter referred to as FPs) and false negatives (FNs): A low detection threshold or an overly generous human operator will tend to have a low FN rate and a high rate of FPs. A high detection threshold or a strict human will be more likely to have fewer FPs but will also have a higher fate of FNs. A classical way to describe a detection system is a 'Receiver Operator Characteristic' or ROC curve on which the detection efficiency is plotted against the probability of a false alarm. For a simple binary hypothesis test, (such as "Was a target detected in that sonar or radar ping?") the ROC curve provides an easily visualized view of detector performance. On an ROC curve, the perfect detector would have a detection efficiency of 1.0 and a false positive probability of 0, while random guesses would correspond to the 1:1 line on a ROC chart. The area under the curve, AUC, is an overall measure of performance. An example is shown in Figure 1. A binary classification based on a single numeric variable can be used to distinguish two populations. An example threshold is shown in the left panel. Varying that threshold results in a ROC curve. The threshold from the left panel is highlighted in the right panel, making the link between the two plots explicit.

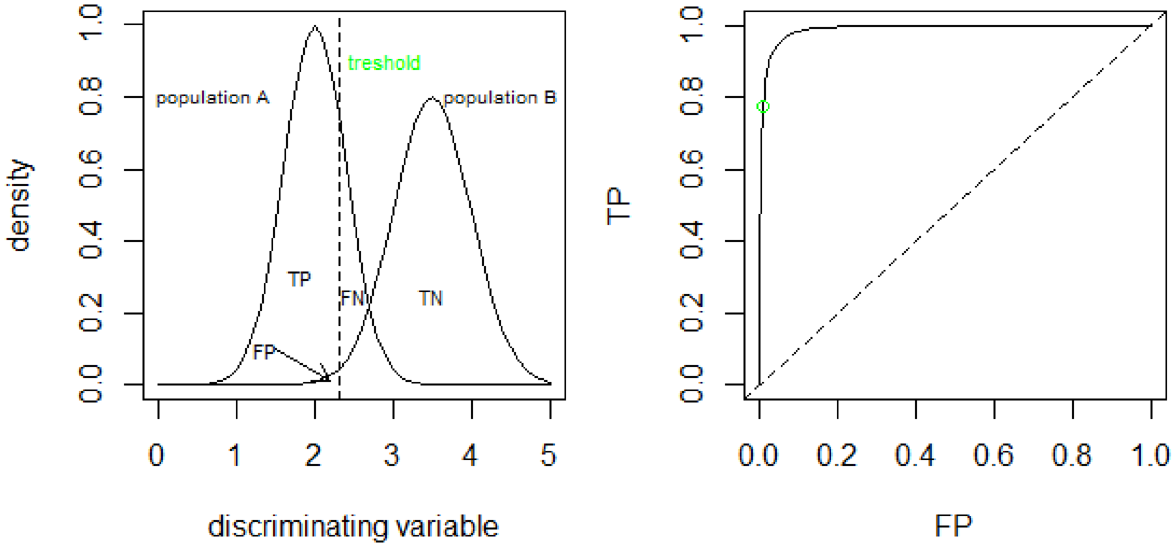


Figure 1 - A classification example, with two equally abundant populations A and B, and a single discriminating variable. The mean of A is 2 and the mean of B is 3.5. Left plot: Given a specific threshold for the discriminating variable (example = 2.3) for which observations with lower values are classed as A and those above as B, we obtain corresponding true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). Right panel: Varying the threshold and representing the true positive rate, TP/(TP+FN) as a function of FP/(FP+TN) is known as the ROC curve. A random classifier would correspond to the dashed line in the right plot.

For marine mammal call detection, where the rate of possible FPs cannot be easily defined (i.e. determining 100% would be challenging without a complete knowledge of animals present (and those vocalizing) in the region) it is often more practical to use a detector performance curve (Roch et al., 2011) in which "Recall" (proportion of genuine calls that are detected, i.e. the Detector Efficiency) is plotted against "Precision" (number of true detections divided by the total number of detections, both true and false). On a detector performance curve, the perfect detector would have both a Recall of 100% and a Precision of 100%. An example detector performance curve is shown in Figure 2. It can be seen from the figure that if a low (5 dB) detection threshold is selected, then a Recall of over 80% is achieved, but at the cost of a poor (60%) precision. This means that for every ten calls detected, four of them would be FP's. If the detection threshold is raised to 8 dB, then the Recall is still over 75% and the precision rises to over 90%, i.e. out of ten detected calls, only one will be an FP.

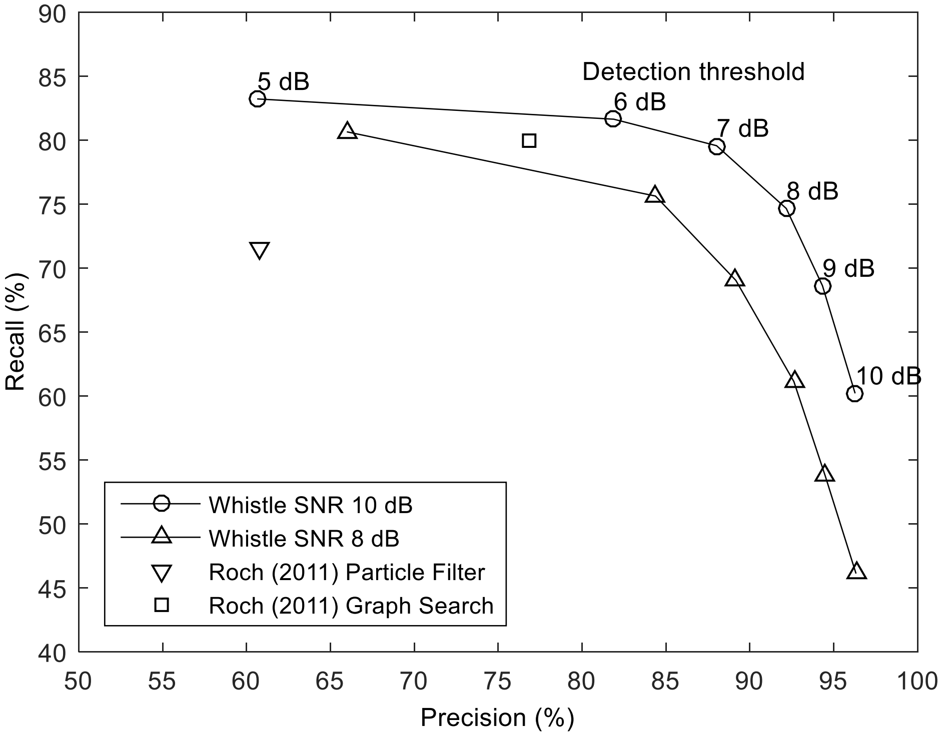


Figure 2, Example detector performance curve taken from (Gillespie et al. 2013)**. "Whistle detection precision and recall for minimum whistle signal to noise ratios of 8 and 10 dB and varying detection thresholds. Precision is the percentage of detected calls which matched a human detection. Recall is the percentage of human detected calls which were automatically detected."**

As with the detection process, no classifier is perfect and when using classifiers to separate calls by species, errors will occur. Classifier performance is often described in terms of a "Confusion Matrix" (Example in

Table 3). A confusion matrix shows how each individual species is correctly and incorrectly classified. Classification is discussed in more detail in section 3.2.5. When using the outputs of multistage detectors and classifiers, it will be necessary to understand the combined effects of the errors in all stages of the system. For example, if the detection efficiency for a given call type is 90%, but the correct classification rate is only 80%, then the true probability of correctly "Detecting", i.e. being able to make use of, one of those calls is 90% x 80% = 72%. Care must always be taken to ensure that what a detector developer describes as "Efficiency" is correctly translated into the detection probability values required for abundance estimation.

Table 3: Example Confusion Matrix taken from (Gillespie et al. 2013)**: Classifier Confusion Matrix for four species of odontocete found in the Polar Atlantic Region. Numbers on the diagonals represent the percentage of correctly classified sections of data. Off diagonal entries represent mis-classifications. For example, 92.3% of Beluga (BEL) sections were correctly classified, with 6.5% being mis-classified as pilot whales (LPLT), 1.2% as White Beaked dolphin (WBD) and none as White-sided dolphin (WSD). Numbers in parentheses are the standard error on the estimates of correct classification rates.**

		% Classified As			
		BEL	LPLT	WBD	WSD
Actual Species	BEL	92.3 (19.4)	6.5	1.2	0.0
	LPLT	2.6	97.4 (4.2)	0.0	0.0
	WBD	0.0	0.0	93.3 (19.7)	6.7
	WSD	0.3	0.0	4.7	95.0 (3.3)

Generally, detector efficiency and classifier accuracy will reduce for quieter calls and from more distant animals. If a distance sampling approach is used for data analysis (see section 4) then detector efficiency is largely encompassed within the detection probability term $\hat{p}$ in equation (4) and the analysis methods can provide a robust estimate irrespective of how $\hat{p}(r)$ changes with distance. However, it is still generally important that $\hat{p}$ at horizontal distance 0 from the receiver is close to 1, i.e. a call from an animal directly above or below the receiver is certain to be detected. In the presence of misclassification, it may be necessary to estimate the number of calls correctly classified close to the receiver (i.e. close to distance 0) to obtain the correct value for $\hat{p}$ at zero distance. This is explored in much more detail throughout section 4.

A significant effect of misclassification is on the abundance estimation of rare species. This topic was explored in detail by (Caillat et al. 2013). To summarize, even if misclassification rates have been measured precisely, since each classification of each detection is an independent random event which may or may not create a misclassification, the actual number of misclassifications in a given dataset will be subject to random fluctuations. For example, if 10% of calls from species B are misclassified as species A, then from a sample of 1000 calls of species B, on average we would expect 100 of these calls (10%) to be misclassified. Since the misclassifications occur at random and independently, we will probably get somewhere between 90 and 110 extra species A calls rather than precisely 100. This means, that although we know the 10% misclassification rate precisely, we will always subtract off exactly 100 calls from the count of species A, introducing an additional error of +/- 10 calls. If the total number of calls of species A and species B are similar, this will not have a significant effect of the count of species A (10 / 1000 = 1%). If however species A were very rare and we only had 10 calls from species A to start with, then the additional error in the count of species A from misclassified species B calls would be around 100%. It is therefore essential, when working with rare species to ensure that the incorrect classification rate of a species as the 'rare species of interest' is extremely low. This may be achievable either by training detectors with the express intent of reducing misclassifications from the more common species or by being prepared to check (if possible) every call of the rare species. An example of the problem of detecting rare species comes from North Atlantic right whales. Gillespie (2004) described a situation whereby tens of thousands of interference sounds, mostly from humpback whales, were received for every true right whale calls, so a detector was developed which only had an efficiency of 60% for right whale calls, but could reduce the false alarm rate from humpbacks down to < 2 false detections per day. However for the detector to be useful, all calls still need to be manually checked.

3.2.3 Sources and Causes of Error

3.2.3.1 *Signal vs Noise*

Both FP and FN are caused by noise. 'Noise', which is discussed in more detail in the following paragraphs, is by definition is any sound that is not of interest (i.e. the 'signal'). High noise levels will mask calls of interest causing FN's and will create FP's. Several types of noise are typically present in PAM data sets in varying degrees. Noise can either be **Persistent** or **Intermittent** and different analytical processes may be needed to deal with each type. Persistent noises are those which occur throughout the entire dataset or at least occur for extended periods of time (weeks – months). For instance, snapping shrimp causing false click detections in coastal regions. Intermittent noise sources are things such as passing ships which create high levels of interference for a relatively short period of time. Persistent noise sources are best dealt with automatically. If they are always there, then it may be possible to retrain detection and classification algorithms to recognize those sounds and to reject them from the analysis. Where it is not possible to entirely remove a persistent noise source, it may be possible to measure a false positive rate and to incorporate this into the analysis (see section 4.1 of this report or - the canonical density estimator correction factor in *equation 3* of Marques et al. (2013a). Intermittent noise sources can be harder to deal with analytically and by their very nature are often hard to train an algorithm to recognize. Dealing with intermittent interference is often best dealt with by a human checking results and rejecting those caused by events such as passing vessels.

System Noise is the electrical noise inherent in any electronic system. Even in "Perfect" electronics, random thermal movement of electrons generates a voltage. The noise generated is spectrally flat (or white noise). On a home stereo system this is apparent as a "hiss" when the volume is turned up high. Active components such as FETs (Field Effect Transistors) and Op-Amps (Operational Amplifiers) used in the input stages of preamplifiers also add broadband noise but this tends to be strongest at low frequencies, and is spectrally flat above a 'knee' frequency determined by the design of the device. Many systems also have other sources of electrical noise, often caused by high frequency switching power supplies or digital components of the recording system which create spikes on the systems ground and power supplies. This can sometimes be heard as a buzz, but is often at higher frequencies, beyond the limits of human hearing, but well within the frequency range of the PAM system and overlapping with the frequency range of some cetacean vocalizations. Autonomous recording systems based on spinning hard drives also suffer from mechanical noise which couples from the hard drives into the recording system's body and then the water where it is readily picked up by the PAM hydrophones. It is out of the scope of this assessment to review PAM systems by their system noise. Such variables are likely to vary across systems but also be dependent on the age of the unit, its internal components but also how the devices are configured or arranged for deployments. Depending on the configuration, there is potential for significant self-noise to be generated. This should be assessed at the design stage for PAM arrays.

Ambient Ocean Noise is always present. Standard curves for mean ocean noise levels are well established (Wenz 1962) which generally show higher levels at lower frequencies. Distant shipping and seismic disturbance (earthquakes rather than E&P) dominate at low frequencies, with surface turbulence, wind, rain and sediment movements all affecting noise levels at higher frequencies. At very high frequencies (above 150 kHz) is it likely that noise is dominated by thermal movement of the water molecules.

Flow-noise can also be present in audio data. It is caused by turbulent water flow past a pressure sensitive hydrophone and is analogous to wind flowing past a microphone. It can create significant (up to 50 dB) noise at frequencies below 500 Hz (Bassett et al. 2012, Bassett et al. 2014) which could interfere with the detection and classification of baleen whale calls. For static receivers, this would only be a concern in areas where there are strong (i.e. > 0.5 m/s) tidal or ocean currents

Biological Activity can also create significant amounts of noise. In some areas noise is dominated by shrimp or schooling fish. Variation in biological activity can lead to different habitats having different sound signatures (Radford et al. 2010) and these sound signatures in turn can be used by larval fish and invertebrates to locate appropriate habitat for settlement (e.g. Stanley et al. (2012); Mann et al. (2007)). Changes in noise driven by biological activity can change by location as described above and by time (seasonally and daily and on lunar cycles) which can further complicate acoustic detection of cetacean species.

Local Human Activity often dominates local soundscapes, especially in coastal areas. Ship traffic near ports and marine industrial activity at ports can add large amounts of noise to the local soundscape (Bassett et al. (2012); Merchant et al. (2014); Veirs et al. (2016). Anthropogenic activities can generate noises that are continuous in nature, but which are experienced intermittently in a local area. These would include vessel noise (from large commercial ships to small recreational boats) which is continuous in nature, but, because the vessel is moving, is experienced intermittently in specific locations. Anthropogenic activities such as pile driving and seismic exploration generate impulsive sound which are intermittent in nature, but these activities can last from weeks to months in specific areas. Bathymetric sonars and/or fish-finders operate continuously on many vessels generating another type of intermittent noise, in this case regular narrow-band pulses at high frequencies.

Other Cetacean Species which are not of primary interest in PAM surveys can also be considered as a type of noise. Passive acoustic density estimation for cetaceans requires that the calls and clicks of different species be differentiated, if the goal is to estimate density at the species level (as opposed to a higher taxonomic grouping; e.g. beaked whales). For related species, this may prove more challenging. Techniques have been developed to classify cetacean calls and clicks (e.g. Gillespie et al. (2013); Roch et al. (2011a), Roch et al. (2011b)), but the ability to classify calls and clicks correctly to the species level will change depending on the species assemblage present (or calling), which will change between locations and may change across the year. More details on classification are provided in section 3.2.5 below.

3.2.3.2 *Call Variability*

Another reason that calls may be missed by a detector is **Call Variability**. This is a particular problem if detectors have been tuned to a very particular sound type and that sound type then changes either through differences between different populations of the same species, or temporal drift in the vocalizations of a population over time, or even due to a change in orientation of the calling animal with respect to the receiver. This is potentially most problematic when using correlation detectors as discussed below.

3.2.4 Detection Algorithms

In the field of DCL, there are many, many different detection algorithms available. Marine mammals make a great variety of sounds that could be detected ranging from short duration, relatively high frequency echolocation clicks to longer duration but broadband sounds, and highly tonal whistles and calls (which also may or may not be stereotypical). Moreover, these sounds are made in a variety of ambient noise environments requiring varying degrees of filtering. Because of this large variety of sound types, no one algorithm will perform well on all these sounds. The choice of algorithm will depend both on the type of sound (or more often the mixture of sounds) to be detected and also the types of noise in the dataset. It is often the case that multiple algorithms are required to extract different sound types from a single dataset.

A signal can be detected if the Received Level of the signal, after processing, rises above some threshold. Many algorithms work by dynamically measuring noise and adjusting the detection threshold so that it is set to some predetermined factor (e.g. 10 dB) above the measured background level. As the noise level rises, so too does the detection threshold. These are known as Constant False Alarm Rate or CFAR detectors (Kay 1993). An important consequence of using a CFAR detector is that the quietest sound that can be detected (and therefore the maximum detection range of the system) will vary over time and / or location if the background noise level is changing over time and / or location.

The purpose of a detection algorithm is to somehow make the signal bigger and the noise smaller. Therefore maximizing the signal to noise ratio (SNR). The ability of an algorithm to increase the SNR in the data is known as Processing Gain. The signal and noise characteristics determine the maximum processing gain that can be achieved by any detector. A good detector will perform close to this maximum while and a poor detector will not realize all of the SNR improvement possible.

Some algorithms achieve a useful result in just one step (e.g. detection) while others may perform more than one (e.g. detection then classification) (Moretti et al. 2009). Because noise can cause false positive detections, the first step in signal detection is usually a conditioning of the signal. This can take the form of a **gain control** to average noise levels over time, **equalization or filtering within specific frequency bands**, or **filtering of a spectrogram image** (Moretti et al. 2009). Some algorithms can be run in the time domain and are often used for detecting clicks, whereas algorithms searching for signals on a spectrogram image of the data are more often used for long duration and/or tonal sounds (Moretti et al. (2009), Norris et al. (2010)).It is important to consider that there are trade—offs in selecting which detector/classifier to use for a species or sound of interest. For example because matched filtering and image cross-correlation use a template (or 'kernel') that represents the signal of interest, they work well (i.e. they have a high Processor Gain) on well-characterized and highly stereotyped calls, but poorly (low Processor Gain) on variable calls. A benefit of correlation detectors is that because they are so specific to the call of interest, they are a classifier as well as a detector (Moretti et al. 2009). Some types of common detection algorithm and their application are discussed below in sections 3.2.4.1 to 3.2.4.3

3.2.4.1 *Correlation Detectors*

Correlation detectors work by comparing a template of a sound with acoustic data. Where the template matches the acoustic signal the algorithm generates a high output and when there is no match the algorithm output is low. Correlation detectors can work in both the time domain and the frequency domain (discussed further below). Templates can either be a real call, ideally one having a high signal to noise ratio, or can be generated synthetically. Synthetic templates have the advantage of being noise free and can be generated to represent an average or typical sound from the species of interest. On the other hand, selecting a high SNR call is often straight forward and is guaranteed to be accurate for at least that one sound and therefore might be considered preferable.

A time domain correlation detector (commonly known as a Matched Filter) is the optimal detector for a known signal in white (i.e. spectrally flat) noise (Kay 1993). A Matched Filter works by cross- correlating a template of the known signal with the incoming time-series data. When the signal is present, and the template matches the signal, the output is at its maximum and will be lower at other times. For highly stereotyped sounds, Matched Filters have a Processor Gain which is dependent on the product of the duration and the bandwidth of the signal; i.e. they maximize the signal to noise ratio in the detector output to the extent possible given the signal characteristics. Some signals, in particular the short duration clicks made by some delphinids, have little processing gain making a matched filter detector unnecessary for these calls.

Matched Filters have been used to detect a number of whale species. Harris et al. (2013) used a Matched Filter to detect fin whales using data from bottom seismometers. The three component accelerometers in the seismometers were used to reject false detections which did not originate from the sea surface and 10% of remaining detections checked manually resulting in a 95% correct detection rate. Ward et al. (2008) used a Matched Filter to detect Blainville's beaked whale clicks. FP rates were estimated using synthetic white noise data so their results cannot be used to assess how well the Matched Filter would have been able to reject other types of noise. Both of these studies used a real recording of a high SNR call from the same study site as the filter template. Stafford et al. (1998) used a synthetic filter kernel based on mean values for frequency and time characteristics of blue whale calls to detect animals using data from a Navy bottom mounted array.

In practice, few marine mammal calls are consistent enough for a Matched Filter to remain optimal as even a small shift in the frequency of a call away from the frequency used in the template will cause them to perform poorly. A useful variant on the time domain Matched Filter is spectrogram correlation. Instead of cross-correlating a template in time, a two dimensional template of the time–frequency contour of the signal matches the spectrogram of audio data (Mellinger and Clark 2000) . For many types of call, particularly stereotyped baleen whale calls, spectrogram correlation provides high Processor Gain along with the flexibility to deal with some variation between calls. This method was used by (Širović et al. 2004) to assess the seasonality of both blue and fin whales in the Antarctic. The detection threshold was set such that the FP rate was around 1% of all detected calls and the calls from days with low call counts were checked by an operator. Baumgartner and Fratantoni (2008) used a modified version of this detector to detect Sei whales using data collected from an ocean glider, though later replaced this with a pitch tracking algorithm (Baumgartner and Mussoline 2011).

For humpback whales, which produce a wide variety of tonal call types, Abbot et al. (2010) developed a modified version of Mellinger's spectrogram correlation method (Mellinger and Clark 2000) which employed multiple correlation kernels to match different song units and only declared a detection when multiple units were detected in order to avoid false alarms from single kernel outputs. For multi-year and multi-location studies, Abbot et al. (2010) state that templates would need to be updated to keep pace with changes in Humpback song patterns in order to maintain detector efficiency over multiple years.

3.2.4.2 *Energy Type Detectors*

Many types of marine mammal calls, such as dolphin whistles and the calls from some baleen whales are highly variable meaning that correlation type detectors such as matched filters or spectrogram correlation will be of limited use. Energy-in-band detectors can be used with these types of calls, but may have lower Processing Gain than a correlation detector. As with correlation detectors, energy detectors can work either on time series data or on a spectrogram. In the time-domain, data are generally filtered so as to only include the frequency band of interest and the power of the filtered signal is integrated over a time period matching the duration of the calls in order to improve Processor Gain. Energy detectors are suitable for many types of transient signal such as odontocete echolocation clicks and variable types of baleen whale calls. However, energy detectors are very non-discriminatory and would detect a down sweeping or constant signal as readily as an upsweep. Therefore they are less useful in situations where distinguishing between two species that produce sounds in the same energy band. They are also very likely to be triggered by transient noise sources, for instance, a simple energy detector for sperm whale clicks is likely to also trigger on impulsive sounds from a cavitating propeller or on signals from other odontocetes.

Energy-based detectors are widely used for the detection of odontocete echolocation clicks (e.g. Leaper et al. (2000), Gillespie et al. (2005), Lewis et al. (2007)). Gillespie and Leaper (1996) used a simple energy based detector to detect porpoise and sperm whale clicks. High rates of false alarm were dealt with by manually identifying sequences of clicks (click trains) on consistent bearings. In these studies, the regularity of the clicks in the click trains and the human's ability to pick them out from the display was possibly of more importance then information extracted from the clicks themselves. Gillespie and Caillat (2008) used the same detector output and combined it with a statistical classifier which conducted a more detailed frequency analysis of the initial detector output to select and separate beaked whale, pilot whale and Risso's dolphin clicks.

Several species of whale have been detected using a Teager-Kaiser energy operator (Glotin et al. (2008), Baumann-Pickering et al. (2010)), though in its basic form it offers no species identification and will trigger on any transient signal within the systems frequency range, it being most sensitive to frequencies at one quarter of the system's sample rate. Klinck and Mellinger (2011) developed a slightly more sophisticated algorithm, which provides a degree of species specificity, by comparing the energy ratio output from two band pass filters, triggering when that ratio exceeds a predetermined threshold. This was successfully used to distinguish beaked whale clicks from Risso's dolphin and Pilot whale clicks, Helble et al. (2012) used a power law detector, which emphasizes spectral peaks rising above background noise, to successfully detect humpback whales with detection likelihood of 95% for a 5% false alarm rate when compared to a human, significantly outperforming an energy detector.

3.2.4.3 *Contour Tracking Detectors*

Energy Detectors are generally more suited to the detection of discrete signal with significant signal-free periods (i.e. periods of silence) between sounds during which time the background noise levels are measured. They will not therefore perform well with multiple overlapping sounds such as are often received from groups of dolphin as they won't distinguish well between overlapping sounds. For dolphin whistle detection, a number of researchers have developed contour tracking algorithms which analyze a spectrogram of sound data for sweeping contours rising above background. These have been used successfully both with baleen whales (Baumgartner & Mussoline 2011; Gillespie 2004) and with odontocete whistles (e.g. Gillespie (2004), Johansson and White (2011), Mellinger et al. (2011), Roch et al. (2011b), Gillespie et al. (2013)) and with baleen whale calls (Gillespie (2004), Baumgartner et al. (2013)). The detection parts of these algorithms are generally not species specific since they will detect any tonal noise present in the data. They are therefore usually combined with a classifier, either running in conjunction with the detector or as a separate processing stage. For instance, Baumgartner et al. (2013) deployed a pitch tracking algorithm on a low power processor installed on an ocean glider. Detected pitch tracks were transmitted to shore for classification by comparing with a library of contour types and checking by a human operator. Gillespie et al. (2013) describe a system in which detection and classification of odontocete whistles are two distinct steps in the analysis process. Within the PAMGuard implementation of this detector, the output of the detection part of the algorithm can also be fed to an alternative classifier described in Oswald et al. (2007).

3.2.5 Classification

While detectors aim to identify as many signals of interest as possible, they often result in large numbers of false positives, especially if the detector is set to be sensitive (which, for example, would be important if the species of interest is rare). Just as signal conditioning helps to minimize false positives, so do classifiers. They can help to differentiate between a single call or click type of interest (a true positive) and a false positive (FP), or differentiate between multiple call or click types and FP. A classifier generally consists of two separate components, a Feature Extractor and a Classification Algorithm which uses the extracted features to differentiate between call or click types. Classification algorithms can include **clustering**, **decision trees**, **multivariate discriminant function analyses**, **artificial neural networks**, **Gaussian mixture models**, **support vector machines**, etc. (reviewed in Norris et al. (2010)). These vary on how much training data they need and have strengths and weaknesses depending on the kinds of acoustic classification that is being done.

While detection theory is well understood for simple continuous noise types (distinct from 'signal types'), many of the types of noise described above in section 3.2.3.1 are highly variable and unpredictable, meaning that many types of detectors will incur a significant number of false alarms. Furthermore, as highlighted above, many detectors are also non-species specific, for example, a contour tracking whistle detector will detect the whistles of many species of dolphin and a simple energy based click detector will detect the clicks of many different species of odontocete. It is therefore common to follow the detection stage of the processor with a **Classification** process which attempts to distinguish between cetacean sounds and "other" sources of sound (anthropogenic or natural), which are considered as noise from the cetacean signal detection perspective and also attempts to distinguish between cetacean species. Classification may also look at multiple single sounds as part of its decision making process and may even include spatial information, for instance in the separation of odontocete click trains from noise by looking for the regularity of clicks detected from a consistent bearing to a cluster of hydrophones.

In the case of dealing with unwanted noise sources, as with detection, the ideal is that this process can occur automatically using well defined computer algorithms, but human operators can also play a significant role, particularly in the rejection of unexpected anthropogenic sound sources. With well-developed software tools, a small amount of human input can go a long way. For instance, when classifying small odontocetes, several classifiers have been developed which can be used to classify clicks and whistles to a species level (Oswald et al. (2007), Gillespie et al. (2013)). However, if false detections are incurred, for instance as a result of the sound produced by a passing vessel, then a classifier which has been trained on multiple species of dolphin may not be able to correctly classify the unexpected sounds from the vessel which may generate FP's. A human on the other hand, needn't look at every whistle or click detection but easily examine and reject the entire group of false detections.

The error rate of a classifier will depend on the mix of species/signal present and the types and levels of noise. For instance, Gillespie et al. (2013) showed that they could achieve a 94% accurate classification rate for whistles from four species of odontocete, but that the mean correct classification rate dropped to 58.5% when 12 species were included. When monitoring animal populations at the species level, a call from species A which is classified as species B will have the same effect on any estimates of the abundance of species A as a missed detection. In addition, any call from species B classified as species A must be treated as a False Positive. As discussed above (section 3.2.2), significant problems arise when one species is more common than the other making it particularly hard to monitor rare species acoustically.

3.2.5.1 *Feature Extraction*

In order for most classifiers of tonal signals to operate, features must be extracted from the detections. Norris et al. (2010) suggest feature extraction should be considered a separate step between detection and classification because of its importance to the success of classifiers and point out that this step has been largely ignored in the DCL literature and that software tools to do this were not readily available when that review was written. Typically, features that are extracted include the start, end and other frequency related measures of the call, call duration, and inter-call interval (Norris et al. 2010). The accuracy of these measurements is affected by the spectrogram settings (due to the frequency/time resolution trade off inherent in Fourier transforms). Common feature extractions include **peak frequency tracing** and **curve fitting** which both try to extract the frequency contours of tonal calls (Norris et al. 2010). They both struggle to accurately trace call contours when there are overlapping calls, which is common for groups of dolphins. ROCCA software (http://biowaves.net/services-technologies/tech-rocca), which is now a PAMGuard module, was developed to allow human operators to manually fix frequency contours for accurate feature extraction. Norris et al. (2010) also review other feature extraction algorithms that may be of value, including **Capon's minimal variance – spectral estimator / watershed method**, **warping -based signal decomposition**, **particle filters**, **cepstral features**, and **generalized perceptual linear prediction models**. The choice of which feature extraction tool to use will be influenced by the availability / ease of use of appropriate software to implement it as well as the number of signal types that need to be classified. The choice of which features are used for classification needs to be considered carefully to maximize the difference between signal types while minimizing the number and robustness of features used, as the more features that are utilized or the more unreliable the feature, the larger the training dataset that is needed for many classifiers (Norris et al. 2010).

3.2.6 Advances in Detection and Classification since the last SML-JIP reviews

In their discussion, Norris et al. (2010) concluded that the detection of stereotyped calls and echolocation clicks were well developed (although the latter generates large numbers of false positives) and that future detection efforts should focus on variable calls and calls with low signal to noise ratio. They also concluded that feature extraction and classification needed much work. In addition, there needed to be a change from DCL efforts on single species to detectors and classifiers that worked on datasets with many different species and many different noise sources and levels. Here we examine if the field of DCL has progressed in the way Norris et al. (2010) suggested. This section will review progress in the fields of DCL by focusing on the progress made at the Detection, Classification, Localization, and Density Estimation (DCLDE) of Marine Mammals using Passive Acoustics Workshops which are held every two years. These are international workshops that bring together the leading practitioners in these fields around specific themes and datasets. The first three workshops focused on individual species with stereotypical calls (Halifax, 2003, North Atlantic right whales) and clicks (Monaco, 2005, sperm whales; Boston, 2009, beaked whales), but have expanded since then to multiple species and multiple call types (Oregon, 2011, odontocete calls and clicks; St. Andrews, 2013, right whales but with multiple calls and a much larger dataset; San Diego, 2015, odontocete and mysticete calls/clicks at multiple locations and seasons). There is a noteworthy trend in workshop foci moving towards more complicated DCL approaches.

In the decade from the first workshop in 2005 to the latest in 2015, the number of participants has increased, suggesting more practitioners are actively involved in these fields. There has also been a great deal of progress in that decade as well as an increase in the complexity of the DCL approaches that are being used. DCL has seen a continued application of methods first developed or used in other disciplines. Hardware that was deployed and is used for geophysics research work has been actively used for DCL on low frequency baleen whale species such as fin and blue whales (e.g. Wilcock et al. 2015). **Mel-frequency cepstral coefficients** and **formant analysis**, both developed for human speech processing, have been used for both right whale DC and blue whale DC, respectively (Lara et al. 2013, Lockhart 2015). DCL practitioners have borrowed **chirplet transforms** and **Weiner filters** from signal processing and applied them to right whale detection and classification and to improve background noise reduction for beaked whale detection, respectively (Roy et al. 2013). The field of statistics has provided much inspiration for a variety of DCL methods, including **Hidden Markov Models** to classify humpback calls (Pace et al. 2011), **Bayesian** approaches to localize walrus knocks (Rideout et al. 2013) and **Gaussian mixture models** to classify delphinid echolocation clicks (Trinh et al. 2015). Computer science has also provided significant inspiration, with the application of **deep learning** to the classification of blue whale calls (Karnowski and Movshovitz-Attias 2015) and **neural networks** to classify bowhead calls (Thode et al. 2011).

These are but a few examples that demonstrate the interdisciplinary nature of DCL approaches and the increasing maturity of these fields. This maturity does not however mean that the rate of progress of these fields has become stagnant. These fields continue to see active development. It is now more common that precision and recall rates exceed 70%, even for detection and classification of multiple species (e.g. Martin et al. (2015)). Less than a quarter of oral presentations at the last three DCLDE workshops have conducted DCL on more than two species (22, 15 and 16% of presentations in 2011, 2013, and 2015 workshops). However, there has been a trend towards including most, if not all vocal marine mammal species present in a given study area in DCL projects and using multiple DCL approaches (e.g. Muoy et al. (2015)). Perhaps the region that has seen the most concerted DCL efforts on multiple species has been the California Current due to the large number of delphinid species and the relatively high efforts that have gone into visual and PAM density estimates there. This has generated a large challenge for those working on DCL in that region, and led to the focus on the 2015 DCLDE Workshop. DCL efforts in the California Current may well show the direction that will be needed for long-term acoustic monitoring of offshore E&P fields. For example, Rankin et al. (2015) used compound classification of whistles, echolocation clicks and burst pulse calls on recorded delphinid species. This approach combines results of classifiers for multiple vocalization types from each species to improve the classification results for a single vocalization type.

To give an example of how such approaches could be applied to density estimation (but has not yet been to date), let us consider a scenario where echolocation clicks are being counted as cues to estimate density in an area of interest, but there are three delphinid species in the area and it is difficult to accurately classify the echolocation clicks of these species. Using classifiers on delphinid whistles and burst pulse calls, in addition to click classifiers, should allow for an improved accuracy of classification of the three species at the encounter level, assuming that all call types should occur within the analysis interval. Once an encounter is classified to species, all clicks can be assumed to be of that species and cue counting may proceed using the echolocation clicks. This can however only work if a single species is present at any one time.

3.2.7 Summary of Detection and Classification Progress

While the sections above demonstrate some progress in the field of DCL for many types of marine mammal calls, many methods fail to make the transition to practical application. This is perhaps due to the limited nature of many training datasets, which results in detectors being suitable for only a limited set of circumstances, particularly with regard to the number of individual animals in the training set, the nature of confounding noises and the range of species to be monitored. For any monitoring program the final choice of detector can only be made based on knowledge of the mix of species to be detected and the types of confounding noise.

While demonstrated to work well for a number of species, correlation detectors have the disadvantage that they tend to only detect a single call type. Therefore multiple templates are required in studies searching for more than one species, or for a species which makes more than one type of sound. This can make them impractical to deploy and additional classifiers may still be needed since more complex sounds may be detected by more than one correlator. The most practical and effective methods for multi species detection are therefore to deploy a relatively small set of quite general energy and contour tracking detectors for an initial scan of the data and then to attempt to classify sounds to species at a later stage.

A detect-then-classify approach can also be important for practical reasons. Long-term monitoring programs can rapidly generate many Terabytes of raw audio data. Even relatively simple detectors can be used to significantly reduce data volumes which need to be managed on a day to day basis. Not only does this reduce cost, but also makes it much easier to both check and to further develop species classifiers since the volume of data that need to be reprocessed is considerably reduced.

Obviously, researchers and marine users will always want to use the best available detector / classifier to process their acoustic data. The choice of which is "best" is however often impossible to make due to a lack of suitable ground-truthed test data. While a limited number of human annotated datasets do now exist, each of those datasets contains only a small mix of species and often even more limited types of noise so may not therefore be directly applicable to a specific study site. While better detectors / classifiers will improve the quality of an acoustic dataset and the accuracy of any inferences derived from the dataset, 100% correct detection and classification rates will never be achieved, and indeed, are generally unnecessary for density estimation.

More important than selecting the "best" detector is understanding detector performance both in terms of efficiency and false positive rate for the study site. False positives can sometimes be dealt with relatively simply by examining a subset of detections by hand and estimating the fraction of calls which are either from a different species of from some other noise source. This approach was used by Kusel et al. (2011) to assess the false alarm rate in a set of beaked whale detections and could be applied to any species with a consistently distinctive call type. However, this approach cannot be used for many species where humans are themselves unable to identify calls to species. While a human can easily tell if a dolphin whistle was indeed a dolphin or some other type of noise, it is often not possible for a human to tell different dolphin species apart. Where this is the case, we are reliant on classifiers which have been trained with recordings of known species (from visual observation). It is quite likely that there are regional variations in call types for many species, so a classifier trained with data from one region may perform differently with data from a different region. Clearly therefore, if classifiers of this type are to be used, then recordings of the local population should be obtained in order to correctly train classifiers. This is something that can be readily achieved via the acoustic tagging of animals (e.g. using DTAGs) and later in this report we discuss recommendations how to implement a field study (section 6.2) and the temporal and spatial sampling required to achieve this.

3.3 Localization

Box C: Why is this topic challenging?

Localization is the process by which the location of an animal can be determined by detecting vocalizations from the animal on a group of or single hydrophone(s). With DECAF, when we refer to localization we are interested in the ability the estimate horizontal range to an animal in order to estimate a detection function.

A common method of localization involves detecting the same vocalization on multiple dispersed hydrophones. The time-of-arrival of the vocalization at each hydrophone is slightly different and allows for the location of the vocalizing animal to be calculated. However, errors in localization are inherent. They occur from the point an animal vocalizes, to the digitizing of the received signal. For example, sound can be distorted as it travels to the different hydrophones introducing errors in timing measurements, the position of hydrophones may not be known precisely and/or there may be slight errors in synchronizing the clocks between different hydrophones. If measured properly, most of these errors can be compensated for or propagated accurately to errors in the location of animals.

There are two potential pitfalls in localization which directly affects DECAF. 1) Error in localization increases significantly with range from a hydrophone array and is roughly proportional to the spacing between hydrophones (the array aperture size). Therefore, although an animal may be detectible on all or some hydrophones within an array, the aperture size of the array may mean it cannot be accurately localized. 2) Some vocalizations are highly directional (i.e. clicks from odontocetes). This can lead to a situation where an animal may be within the effective localization range of an array but is orientated in such a way that it does not ensonify enough hydrophones within the array to allow a position to be calculated.

The optimum spacing of hydrophones within an array is therefore dependent on a variety of factors. The frequency and source level dictate how far a signal can travel before it is no longer detectible above the background noise. The aperture of an array must be significantly less than this range. If vocalizations are directional, then the aperture size must also balance localization accuracy (i.e. a larger aperture) with the chance of coherently detecting an animal on enough elements within the array to make localization possible (i.e. a smaller aperture). As frequency, source level, the directionality of a vocalization and diving behavior are all species specific, the optimum design of a localizing hydrophone array is species specific.

In addition, even for the most optimal localizing array, it is unlikely that the probability of localization is the same as the probability of detection, especially for directional vocalizations: thus, in a distance sampling framework, to apply the probability of localization to 'n-nodes' requires an understanding of how to convert a probability of localization to a probability of detection.

3.3.1 Introduction

Density estimation methods using PAM often require at least some positional information (e.g. bearing, range, depth/any combination of the three) to estimate animal density. PAM sensors in different configurations can be used to obtain such positional information by detecting and classifying the vocalizations of different species and then *localizing* (i.e. calculating a direction to, range to or position of) the source of the vocalization (i.e. the animal). Localization has been applied over the last four decades (Watkins and Schevill 1972) to study the behavior and density of many species of cetaceans and remains an important and actively used methodology in marine mammal science. Although the literature contains many examples of studies which have used some form of localization, localization remains an often difficult and somewhat nuanced set of methods which tend to be highly species and application specific.

There are many factors which must be considered before deploying PAM sensors to localize a single or group of species. The type of hardware and software required to effectively localize is highly dependent on a target species' typical vocalizations, what type of localization information is required (e.g. bearing and/or range information) and environmental conditions. Even the most effective localization system is unlikely to accurately localize all detected vocalizations and any system used to calculate a range to an animal will be subject to substantial increases in localization error the further the animal is from the PAM sensors. For example, the general rule of thumb when using an array of distributed PAM sensors to localize an animal's range, is that the maximum effective localization range is 2-10 times the maximum distance between sensors in the array (bearings to animals are more robust and essentially limited by the detection range). Deep water environments also pose a challenge as horizontal range accuracy falls off substantially with increasing depth difference between a vocalizing animal and the PAM array as the slant range increases substantially. Shallow water environments present a different set of challenges especially for low frequency vocalizations due to the complex propagation of sound in shallow water.

However, although there are limitations, passive acoustic localization is important for density estimation as it provides ranges to animals (to generate a detection function in distance sampling). There is no requirement to tag animals (specifically for localization, but as noted in other sections this will be extremely valuable for other DECAF components), the position of multiple vocalizing animals within a study area can be determined and the data-to-cost ratio is often favorable. Therefore, localization using arrays of PAM sensors represents a powerful methodology which can be very useful in DECAF context, for example to calculate a probability of detection or enhance SECR methods. However, the many errors and pitfalls which can be introduced means the type and placement of hardware and subsequent software used in analysis of the data, requires careful consideration. Although many of the challenges in localization are generic across different types of hardware, this section is written in the context of a *seabed-mounted DECAF system which consists of a set of 'p nodes' dispersed within a larger group of n nodes*. The focus here is on the nodes which are localizing hydrophone arrays capable of resolving the range of a vocalizing animal and thus in a distance sampling context, can be used to calculate a probability of detection. The next section explores different localization methods which can be applied to vocalizing cetaceans and details the various possibilities and pitfalls in using an array of PAM sensors to localize different species.

3.3.2 General Localization Methods

PAM localization refers to a broad set of methods to obtain one, two or three-dimensional information on the location of vocalizing animals. DECAF is concerned with the estimation of horizontal range, but there is valuable information in deriving the depth of the animal relative to the sensor. Determining this is a non-trivial exercise (both in physical terms for PAM deployments and statistically). For example an estimate of horizontal range typically requires the depth of acoustically-detected animals to be estimated. Later we discuss statistical approaches that advance this area – for example, see Cox et al. (2011), Harris et al. (2013) for alternative approaches to extending standard distance sampling to account for animals at depth. Three localization methods are discussed here: a) time delay based localization; b) hydrophones combined with particle velocity sensors; and c) matched field processing.

The most common localization method is to use an array of receivers (in this case single hydrophones) to detect a sound source; the time delays between the sound arriving at the different receivers are used to calculate the bearing to or position of the source (Watkins and Schevill 1972). Time delay measurements however, do not necessarily have to be between two individual receivers. Echoes from the seabed and surface can be used to create 'virtual' time delays which can also be used to provide additional positional information or improve existing localizations (Aubauer et al. 2000). Beamforming is also often cited in cetacean literature, particularly when using towed arrays; beamforming is a method of utilizing time delays from multiple closely spaced hydrophones to determine a bearing to a source; it can be very effective in high background noise environments or in a situation which requires an array to target detections from certain directions (e.g. Miller and Tyack (1998), Thode et al. (2000a)).

As well as arrays of hydrophones, some single receivers can calculate bearings to animals based on particle velocity (e.g. DIFAR buoys). The DIFAR system uses a single hydrophone to record the pressure wave and a 2-axis accelerometer measures incoming particle acceleration from which particle velocity is estimated. Phase and amplitude differences between the signals on the two accelerometers and hydrophone provide a horizontal bearing to a sound source. DIFAR systems can also be deployed in an array to obtain ranges or precise location of animals by crossing bearings from multiple receivers (Miller 2012). Similarly, seabed-mounted seismometers with particle velocity sensors have been opportunistically used to provide bearings to fin whales (Harris et al. 2013). Particle velocity sensors are generally only useful for LF vocalizations due to the typically limited bandwidth of accelerometers.

Matched field processing (MFP) is a set of techniques which compare received acoustic data to a set of modelled acoustic fields. If the typical vocalizations of the study animal can be accurately described and realistic propagation models calculated, then the acoustic signals which might be received on different hydrophones at different locations within an array can be predicted. Then by comparing the model to the received data it is possible to estimate the location of the animal. The advantage of MFP is that it compensates for (and indeed uses) the substantial warping of acoustic signals which can occur when acoustic signals interact with the surface, seabed or thermocline over *long* ranges, something which can otherwise introduce substantial errors into time delay measurements. However this requires intimate knowledge of the surrounding environment, including bathymetry, bottom type and sound speed profiles which can be costly to gather (Baggeroer et al. 1988, Thode et al. 2000b, Thode et al. 2006).

For the remainder of this report we focus on hardware and software which is designed to localize based on time delay measurements since this is the most applicable methodology to multiple different species and the most developed and generic method in the literature.

3.3.3 Array Design

Time delay based localization requires an array of hydrophones Time delay based localization requires an array of hydrophones (or a surface/bottom reflection which can create a 'virtual' hydrophone allowing a single hydrophone to act as an array (Au & Hastings 2008). The number and distribution of hydrophones determines the *type of localization information* which can be determined *e.g.* a bearing, range or 3D fix. The *accuracy of localization for a given sound* is highly dependent on two important factors; the number of hydrophones within the array and the distribution and distance between those hydrophones.

The absolute minimum number of hydrophones is one + the number of dimensions to be resolved, i.e. 3D tracking requires an absolute minimum of four hydrophones distributed in three dimensions. Increasing the number of hydrophones within an array increases the number of time delay measurements between receivers. Thus, more observed data results in a greater accuracy and also allows the estimation of localization accuracy.

The choice of the number of hydrophones and how to distribute them throughout the array to provide the best localization performance requires careful consideration. The distribution of hydrophones*, i.e.* how they are arranged within an array, determines what localization information is possible. For example, a linear hydrophone array can only restrict the location of a vocalizing animal to a circle perpendicular to and centered on the axis of the array irrespective of the number of hydrophones in the array. For a vertical linear array this would give you a range and depth but no bearing information. The greater the number of hydrophones, the greater the accuracy in depth and range but a bearing to the animal is impossible to recover with this array geometry. A planar array can determine two possible equally probable 3D locations, of which one will be the correct location of the vocalizing animal. If a planar array is deployed close to the sea floor or sea surface, then there is of course only one viable solution since the other will be either below the sea floor or above the surface. An array distributed in three dimensions (i.e., with four or more non-co-planar hydrophones) can provide a unique 3D location provided all of the hydrophones detect the same sound (Møhl et al. 2001).

The distance between hydrophones is perhaps the most complex and least understood design consideration. Generally, the probability distribution of time delay error, usually caused by errors in the location of the hydrophones, sound speed or cross correlation of the signal between hydrophones is fixed for a given localization. As aperture size (the distance between hydrophones) increases, the magnitude of the time delay between two hydrophones increases and the time delay error becomes proportionally smaller. Therefore, the average aperture of the array should be made as large as possible. However, to calculate a time delay, the signal must be detected on a minimum number of sensors. The higher the frequency and the more directional a signal, the less likely that it will be detected on widely separated hydrophones. Thus, the distribution and distance of hydrophones is a trade-off. Hydrophones must be spaced closed enough that individual vocalizations are consistently detected on a minimum number of receivers (this is from now on referred as *coherently detected*) but the array spacing must also be large enough to provide accurate localization information at useful ranges.

There are various methods that can be used to determine this optimal spacing of receivers. A simple calculation might suffice for some baleen whales, however odontocetes with more narrow beam clicks, which also may or may not produce less directional tonal vocalizations, require careful consideration. A rule of thumb often used within the marine mammal bioacoustics community is that an array can only localize accurately to range of between 2-10 times the maximum spacing between hydrophones within the array. The next three sections cover different factors which must be taken into account when considering the optimum aperture size of an array (with assumed omnidirectional response – as is likely given the currently available PAM technology).

3.3.4 Vocalization Type

Cetacean vocalizations vary hugely in frequency, from a few tens of Hz for some baleen species to porpoise and dolphin clicks which contain frequency components >200 kHz. Frequency, along with SL, are major considerations in array spacing. The higher the frequency the greater the associated attenuation in seawater - *e.g.* a 200 dB re 1 µPa at 1 m porpoise click is detectable at hundreds of meters whilst a blue whale call made at the same amplitude can be detectable for tens or hundreds of kilometers depending on propagation conditions. Obviously, source level is important as the louder a vocalization the greater its detection range. Therefore, when considering if vocalizations will be coherently detected[1] on widely--spaced hydrophones, frequency and source levels are the first parameters to consider. Vocalization frequency and bandwidth also affects how accurately it is possible to make timing measurements between the signals arrival on different hydrophones. While modern signal processing techniques can measure time delays to a fraction of a wavelength, is a fundamental limit to

---

[1] Coherently detected refers to a minimum number of hydrophones within an array detecting a vocalization. The minimum number of hydrophones is determined by the type of localization information which is required. For example, for a range to be calculated, then three hydrophones would need to detect a vocalization. Note that this can also change depending on the design of an array. For example, if an array consisted of multiple dispersed units, each with an array of three closely-spaced hydrophones, then at least two units would need to be ensonfied for a range estimation to be possible. The minimum number of hydrophones would therefore be six, rather than three.

timing accuracy (the Cramer-Rao lower bound) that is dependent on the SNR and bandwidth of the signal. Therefore, it is unlikely that for a tonal baleen whale call at a frequency of 10's of Hz, timing accuracy will be better than a few milliseconds, whereas for high frequency species such as the harbor porpoise, timing accuracy of a few microseconds can be readily achieved.

An extra complication is added by the beam profile of animals. Odontocetes, either entirely or in addition to a tonal repertoire, utilize short duration clicks to sense surroundings, hunt for prey and/or sometimes communicate. All the echolocation clicks from odontocetes are directional, instead of a uniform spherically propagating pulse they should be considered more like a flashlight varying in intensity with animal behavior. Figure 3 shows the example beam profile of a bottlenose dolphin: note that to the sides and behind the animal there is significantly reduced acoustic energy.

Therefore, when considering the optimal hydrophone spacing for this group of species, especially for clicking species only, it is vital to have additional knowledge of the beam profile and pattern of scanning with the beam. For some species, *e.g.* beaked whales and harbor porpoises, an unfortunate combination of higher frequency vocalizations, a lack of tonal repertoire and narrow beam profile, means that the detection probability is particularly correlated with both the orientation and location of the animal. Thus, the spacing between hydrophones has to be significantly reduced to stand a chance of coherently detecting clicks and it must also be assumed that an animal will only be detected if it is at least partially oriented towards the hydrophone array unless it is very close. Therefore, localization information cannot be calculated for all clicks (and therefore only portions of a dive are usually recorded), even if the animal is located well within the detection range of hydrophones. The consequences of this for density estimation are discussed in section 4.

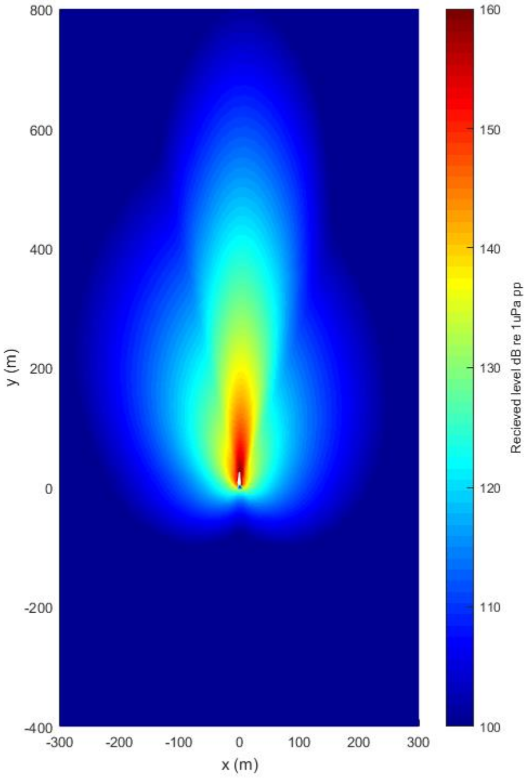


Figure 3. An example of a directional vocalization. This is an estimated harbor porpoise beam profile for a 191 dB re 1 uPa on axis source level. The porpoise is pointing in the y direction and located at x,y = (0,0). The beam profile was measured by (Koblitz et al. 2012, Finneran et al. 2014). (Note: Finneran beam profile measurements are from a bottlenose dolphin; these were used for some parts of the beam profile as a full 360° beam profile of a harbor porpoise has not been published.)

3.3.5 *Spatial Aliasing and Match Uncertainty*

The last section discussed how increasing aperture size reduces the chances of coherently detecting vocalizations, especially for narrow beam profiles. Larger apertures can also be problematic if there are multiple vocalizing animals and/or very high vocalization rates

Match uncertainty is a form of aliasing where a sequence of vocalizations of one or more animals introduces a source of error. Many marine mammals vary their vocalization rate and there is often the possibility that animals will be in groups with a high overall vocalization rates. Direct path signals are also often mixed with surface and bottom echoes. It is therefore generally the case that there is some ambiguity as to which vocalizations match between different hydrophones. Match uncertainty is defined as the difficulty in determining which detections on which hydrophones belong to a single vocalization from an animal. Consider a detection (e.g. a dolphin click) on a single hydrophone, named here the primary hydrophone. The detection was probably detected on at least some other hydrophones within the array. To find those vocalizations a time window before and after the time of the detection is searched on each hydrophone within the array. The time window size is equal to the distance from the primary hydrophone divided by the speed of sound. In an ideal situation, only one detection is observed on each of the searched hydrophones. When more than one detection is observed there may be spectral or temporal properties of the vocalization which can be used to match to the primary detection *e.g.* the frequency modulation of a whistle or patterns of inter click intervals (e.g. Ward et al. (2008)). However, for many species it is impossible to distinguish the vocalizations from different individuals, leading to the issue of match uncertainty.

There have been multiple attempts at solving match uncertainty in the literature. For example Baggenstoss (2011a) developed algorithms to remove echoes and obtain click trains of individual sperm whales recorded on different hydrophones within an array. Baggenstoss (2011b) then correlated these click trains and used the resulting time delay measurements to create multiple localisation solutions, applied weights to each and began a pruning process to remove unlikely solutions, thus ending up with correct positions of one or more animals. Baggenstoss (2013a) also proposed an alternate method to track beaked whale clicks by using a ‘correlogram’ of time delay measurements to create time delay contours for pairs of hydrophones. These contours were then associated by comparing different pairs of hydrophones which share the same hydrophone. Miller (2010) discussed two possible solutions to click matching in his thesis on tracking sperm whales. The first involved initially separating click trains using angle of arrival and then searching for clicks on other hydrophones based on the current inter-click interval (ICI) of individual sperm whales. The second involved using a method proposed by Giraudet and Glotin (2006) which requires at least four hydrophones to be ensonified and then use a set of basic arithmetic relationships to determine ‘true’ time delay measurements. Finally, Nosal (2013) proposed a method wereby a likelihood surface is created for every possible time delay measurement between hydrophone pairs. The surfaces between all pairs are then multiplied together and persisitant peaks over multiple calls above a threshold are clustered and extracted, allowing the positions of multiple whales to be determined. These solutions would need to be tested for different species and it may well be that different algorithms are far better suited to different species.

A hardware based approach to match unceratinty might be the use of clusterred hydrophones. For example an eight hydrophone array could be designed as an evenly dsitributed array of hydrophones or two units of four closely spaced (orders of magnitude below the maximum detection range of an animal) hydrophones. Within each unit the clusterred hydrophones are spaced so closely that the time window is a small fraction of the vocalization rate and therefore match uncertainity is unlikely to occur. Thus the only match uncertainty occurs between the two units. This is a lot easier to solve than match uncertainty between eight evenly distributed hydrophones, however because clusterred systems are less disperssed, localisation would be less accurate and therefore the effective loclaisation range reduced.

3.3.6 Signal Distortion

Acoustic signals which have travelled large distances can become distorted and in addition, waveforms at different angles from an odontocete's beam profile can have very different amplitude modulations and varying spectra (Au et al. 2012). This may result in a reduction of timing accuracy, and possibly also exacerbate difficulties in matching vocalizations on different receivers as discussed above. For very narrowband signals, e.g. pygmy blue whales and NBHF species, additional noise or distortion can make accurate timing measurements difficult. A common method to determine time delays is to cross correlate signals and pick the highest peak of the resulting cross correlation function as the time delay. This works well with frequency modulated signals (e.g. broadband dolphin clicks), however if a signal is narrowband, multiple similar sized peaks in the cross-correlation function are present and so picking the wrong peak and therefore wrong time delay measurement is likely.

The only way to ensure that timing measurements are accurate is to make the spacing between hydrophones less than half the wavelength of the target signal, however this is often not practical for an array targeted at multiple species (wavelengths range between 15 m and less than 11 mm) and certainly could not be implemented for localizing arrays which require a wide aperture between at least some hydrophones. However, there are also analytical approaches which can help reduce error; e.g. correlograms can be constructed from click trains and time delay contours tracked. This can 'smooth out' some of the error in individual time delay measurements (Baggenstoss 2011a, b; Baggenstoss 2013a). Alternatively, rather than cross correlating waveforms of a narrowband signal, the waveform envelope might be more accurate, or using the rising edge of a signal.

3.3.7 Potential Array designs

The optimum array aperture for a given species is therefore a balance between multiple competing factors. Ideally aperture sizes should be as large as possible to make any errors in time delay measurements a smaller proportion of the overall time delay value, thus increasing localization accuracy. The source level and frequency dictate how far a vocalization travels before it becomes undetectable and therefore put an initial limit on the aperture size; the array aperture should be at least an order of magnitude below this. Narrow beam profiles require a further reduction the aperture size to ensure that vocalizations are consistently coherently detected. Matching detections on different hydrophones and compensating for signal distortion can be significantly easier if hydrophones are clustered; however this will reduce localization accuracy and may not be necessary as sophisticated matching algorithms are available for some species.

Designing an array which balances these factors is difficult. Frequency, source level, beam profiles and vocalizations rates are all species specific and therefore the optimum aperture size will vary, depending on species. The challenge is therefore designing an array which *can effectively localize multiple species.*

There are a number of potential solutions for a multi-species localizing array. An initial design might be to select the smallest required aperture and create a uniform density of hydrophones over the proposed study area. However, this quickly becomes impractical. Imagine a 20 m aperture for harbor porpoises over multiple square kilometers – this would result in thousands of hydrophones. Therefore, a uniform density of sensors with aperture sizes that can effectively localize the vocalizations of baleen whales, dolphins, beaked whales and NBHF species is, in most cases going to be prohibitively expensive, both in hardware and analysis costs.

A more pragmatic solution would be the use of two or more different typical aperture sizes targeted at different species. For example, a nested hexagonal configuration with 4,000 m radius for baleen whales and 500 m radius for beaked whales and dolphins is shown in Figure 4. Both array types would be capable of providing ranges to animals within their typical performance limitations. The potential area within which a beaked whale can be localized is significantly reduced compared to a baleen whale; however, the number of sensors required is potentially manageable compared to a uniform distribution of hydrophones at the smallest aperture size.

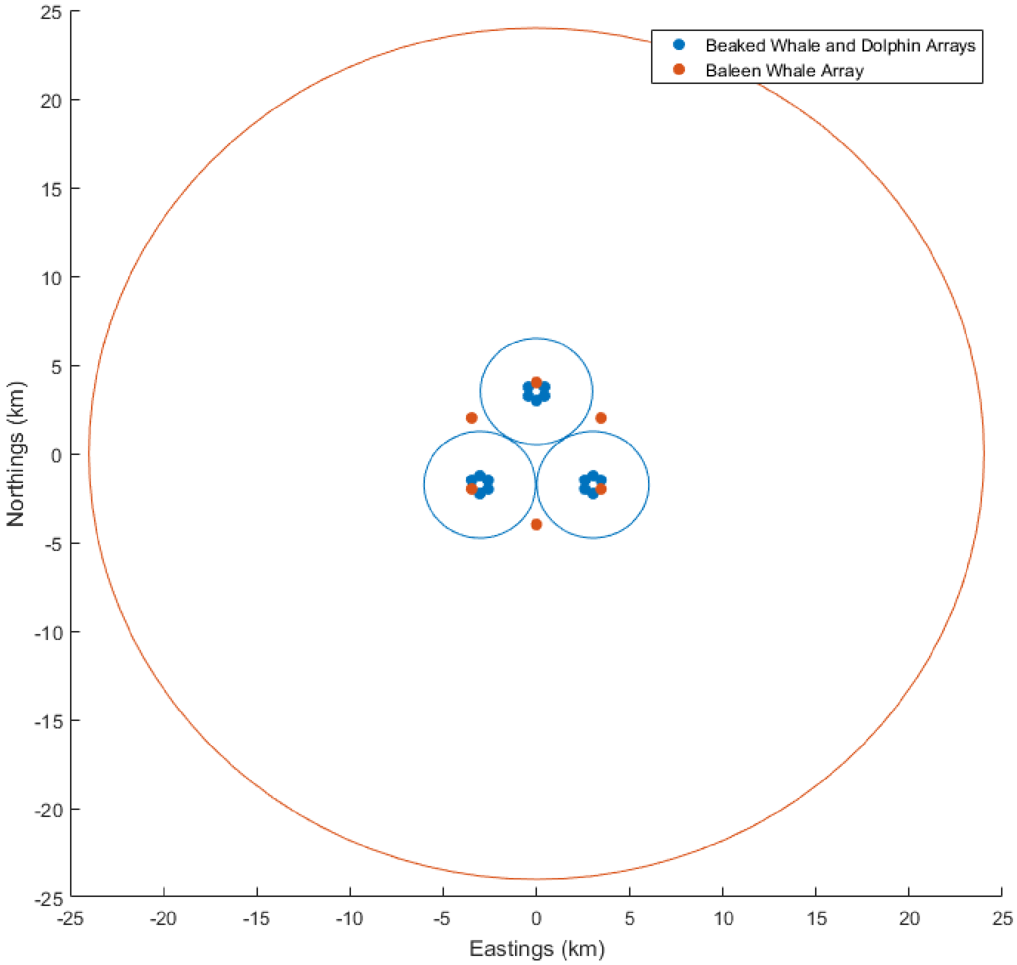


Figure 4. Example of a 21-element nested hydrophone array. The individual blue hexagonal arrays are capable of coherently detecting and localizing ranges to directional beaked whale and dolphin clicks. The orange array is larger and therefore can coherently detect and localize low frequency omnidirectional vocalizations of baleen whales to much greater ranges than the blue hexagonal arrays (baleen whale vocalizations would be detected on all hydrophones, blue and orange). The blue and orange circles represent an 'effective localization' range of 3 times the array aperture. This is for illustration purposes only; in reality, localization error increases with range from the array. The magnitude of the increase is dependent on a multitude of factors discussed below. The 'effective localization range' is therefore simply a defined threshold of acceptable error (as defined by requirements of the detection function).

Another potential solution is to use *clustered* hydrophones (Wiggins et al. 2012). These are generally single deployable units which have multiple closely-spaced hydrophones (orders of magnitude less than the maximum detection range). The close spacing of the hydrophones means that each unit is only capable of calculating a horizontal and vertical bearing to an animal. The advantage of this approach is mainly pragmatic, if a fixed seabed system consists of many widely-spaced hydrophones (e.g. the baleen whale array in Figure 4), adding some clustered stations, *e.g.* a multichannel HARP device, may be a cost-effective means to provide additional localization information on higher frequency species. Instead of using smaller nested hexagonal arrays as in Figure 4, a few clustered hydrophone systems could be added between the baleen whale hydrophones. If the clustered system was ensonified by a beaked whale, then it would be likely that at least one hydrophone in the baleen whale array would also be ensonified. The combination of bearing from the cluster and single time delay between the cluster and baleen whale array would be enough to calculate a range to the beaked whale. Thus, rather than deploying 15 additional hydrophones in nested hexagonal arrays, two or three clustered systems could suffice instead; a potentially much more cost effective strategy, however localization accuracy would be significantly reduced compared to the nested hexagonal configuration. Clustered hydrophone units could also be used as a standalone system if HF species are the primary target of a study. For example, multiple 4-channel HARPs have been deployed on the seabed to localize beaked whales (see section 3.3.11).

3.3.8 Localization Errors

The last section focused on the distribution of hydrophones and optimal aperture size. For any design of hydrophone array there are multiple sources of error which must also be considered pre-deployment.

3.3.8.1 *Clock Synchronization*

A key challenge when using separate bottom mounted recorders for localization is clock synchronization. Localization algorithms based on time of arrival measurements require that time differences can be measured highly accurately between devices, however, individual clocks on different devices will all be running at slightly different speeds, with differences of between 1 and 10 parts per million (ppm). This sounds minor; however, it should be considered that a 10-ppm clock drift equates to nearly 2 s per day and a 1 ppm clock drift will give an error of 1 second after 5 days. Clock drift is also temperature dependent and therefore, in a variable temperature environment, likely to vary. If multiple autonomous units are deployed, even if clocks are synced upon deployment it is likely that the error in timing, due to drift, will rapidly make localization impossible unless actions are taken to accurately measure clock drift during the course of a deployment.

By far the best solution is for all signals to be digitized together using a single clock which requires signals from all hydrophones in an array to be connected via a cable or radio signal. If hydrophones are to be distributed over a wide area, then running all signals to the same digitizing chassis, or running clock signals to separate chassis may be impractical in which case alternative methods must be used.

If recording units have a surface presence, then it is possible to add a GPS receiver which can detect a highly accurate (one nanosecond accuracy) pulse per second included in the GPS signal. This can be used to measure clock drift to an accuracy of a single sample. This is only possible if there is a surface presence. A surface component on a mooring also provides the opportunity to telemeter data to a remote collecting point (e.g. on shore or on a platform) but can greatly increase the potential for failure or loss due to bad weather, ship collision and biofouling.

For arrays with large aperture it might be possible to estimate clock drift after a deployment, *e.g.* by pinging the sensors from set locations. The clock drift could then be extrapolated and any errors in the extrapolation (for example due to varying temperature) would be a small enough proportion of the time delays to make localization possible (Wiggins et al. 2013).

Another option is to have pingers integrated into the PAM units themselves. If each unit sends out a ping at regular intervals, then it is possible to determine the relative clock drift between units. If positions are well known, then a single unit pinging will be sufficient to clock align all units. If positions of units are not well known, then if all units ping, both position and clock drift can be determined. When studying marine mammals care would of course need to be taken to ensure that the ping signals did not affect animals distribution or behavior. However, if hydrophone units are unlikely to move (i.e. by being bottom-mounted) an initial set of pinging would suffice to determine position and then clock synchronization could be performed by recording sounds at known locations e.g. a boat passing over the array.

3.3.8.2 *The Position of Sensors*

One of the most important factors in localization is accurate knowledge of the locations of each hydrophone within an array. Errors in hydrophone position propagate to errors in the localized position of animals and therefore range estimations. For static arrays, the position of hydrophones can be recovered post-deployment as mentioned above, by broadcasting sounds from known locations or by having pingers on hydrophones within the array itself (Watkins and Schevill 1972). However, some error in hydrophone positions will inevitably still be present in any system. For example, even for closely-spaced hydrophones on a rigid frame there is still some uncertainty precisely where the piezo electric element is located and its exact effective center for aperture measurements. And for much larger aperture systems, calculating the positon of hydrophones from pinging sources will unavoidably lead to errors (for example slight distortion in sound speed refracting the ping).

3.3.9 Environmental Influences

Temperature, salinity and depth can change the speed of sound in water and thus in most environments there exists a variable sound speed profile. Localizing in such an environment requires algorithms which take the sound speed profile into account and compensate appropriately and there are many examples in the literature the literature on how to do this (e.g. Thode (2005)). However, this requires accurate information on the sound speed profile.

In addition to introducing errors in the position of animals, the sound speed profile may have influence on the optimum location of the array. For example, a downward refracting environment might mean an array on the seabed will detect more animals whilst an upward refracting environment would favor a system on the sea surface. The depth at which animals vocalize would also be a consideration here.

Noise is another potential source of error in localization. One of the core rules of PAM is that a signal power, after processing, must be greater than the background noise (in its frequency band) in order to be detected. Noise has implications for all aspects of PAM, detection, classification and localization. Detection and classification are covered elsewhere in this document; however, for localization, noise may prevent clicks being coherently detected on different hydrophones, which will reduce the ranges at which localization is possible. It will also introduce more timing error in the cross-correlation of signals, although for most species this is likely to be a very small proportion of the overall timing error.

Environmental influences therefore have the potential to introduce significant performance constraints for a localizing array and can be broadly categorized into factors for which there is no solution and factors which need to be measured. For example, ambient noise is an environmental factor which can, at worst, be impossible to solve and must be initially measured when a site is being considered for PAM and throughout a deployment. Sources of noise such as wind, waves and rain are reduced in deep water and so it is likely that, in most locations, noise conditions will be appropriate for bottom-mounted PAM arrays.

The sound speed profile through the water column is important also, and can be compensated for if there is adequate data. For example, sound speed profiles can be measured pre-deployment to check that animals in different parts of the water columns will be detected; and monitored post-deployment so that localization accuracy can be recovered. In many countries, data on speeds is readily available from regular CTD sampling and can be interpolated with physical oceanographic models (this is expected to be available as part of the development of an E&P field). The ARGOS network (Roemmich et al. 2009) also provides worldwide coverage. Such information should therefore be readily accessible and sound propagation can be modelled using various algorithms and open access programs e.g. ACTUP (Duncan and Maggi 2006). It would be advisable to make in situ CTD measurements during monitoring to test the accuracy of any sound propagation models used.

3.3.10 *Simulation of Errors*

In the above discussion localization errors were split into three categories: a) the ability to detect an animal on multiple receivers and match detections; b) errors in time delay measurements (e.g. from synchronization issues, noise and sound propagation); and c) errors in the position of receivers.

Accurate knowledge of an animal's diving behavior/movements, acoustic beam profile and acoustic behavior along with environmental conditions is needed (and for some species is available or can be collected) to simulate the likelihood that a group of receivers will detect an individual and what match uncertainty issues may occur. Building such a simulation that incorporates detailed animal movements in 4D is beyond the scope of this report (as it is extremely complex to parameterize and implement). However, if it is assumed that an animal can be detected on multiple receivers, the contribution that errors in time delay measurements and position of receivers make to the final localization can be simulated. Timing errors (e.g. due to synchronization, sound propagation and signal distortion) and errors in position can be combined into one overall timing error measurement, simply by calculating the root mean square value. The error in position is converted into a timing error by *timing error = position error/ sound speed.*

Error surfaces were generated to simulate the localization accuracy of a 6-element, 20 m diameter hexagonal array at different ranges and depths and for different overall timing errors. This array was chosen as a simple design which has been in discussed in a recent review (Moretti et al. 2009). The PAMGuard Sound Acquisition and Click Detector modules were used to generate simulated clicks in a 200 m x 200 m grid around the array, 0-100 m above the array in 5 m steps (Figure 5). A simulated click on each hydrophone was generated assuming a source was located at every grid point and then localized using a Markov chain Monte Carlo (MCMC) algorithm assuming a % standard deviation overall time error. The errors predicted by the Markov chain Monte Carlo algorithm were then used to generate error surfaces and this was repeated for different % overall timing error values. The 20 m aperture was simply for simulation purposes. Time delay errors were converted to a % of the overall time delay value (which is a function of aperture size) and the array the error surface scaled so that they could be used for multiple aperture sizes.

3.3.10.1 *Propagation of Time Delay and Receiver Errors.*

Figure 5 shows the mean horizontal range error versus range from the hydrophone array for different average overall timing error values. Range is quoted in units of maximum array aperture size so that these errors can be scaled for different sizes of array and thus different target species. Time errors are quoted as a percentage of the maximum time delay possible. For example, for a 1.5 km array the maximum time delay is 1500 m/1500 ms^{-1} so 1 second. A 1% timing error for a 1.5 km aperture would therefore be 0.01 seconds and likewise for a 150 m array would be 0.001 s. Figure 6 shows an example using the data from Figure 5 for a 1 km aperture array, about the size that might be used to localize sperm whales. The errors are represented as both milliseconds (representing timing error) and the distance error between hydrophones. Both Figures 5 and 6 assume the animal is at a similar depth to the array. All simulations assumed a 5 ms^{-1} error in sound speed (equivalent to an additional 0.3% error in timing).

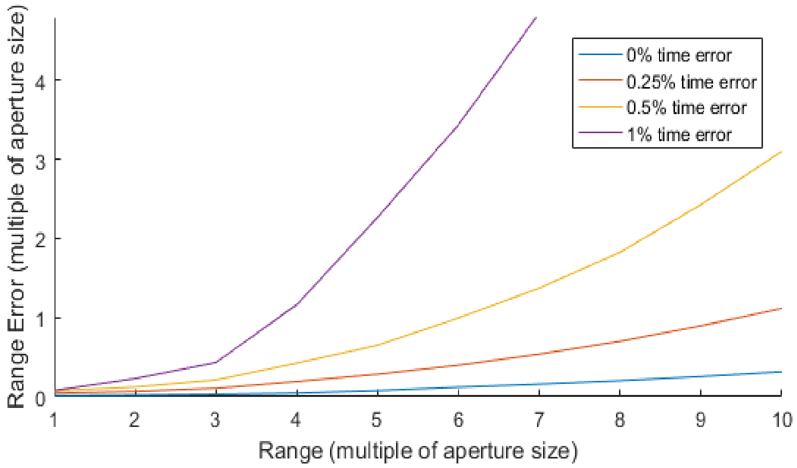


Figure 5. The horizontal range error in localized position versus the range from a 6-element hexagonal array. The ranges are quoted in multiples of the aperture size of the array and time delay errors are quoted as a percentage of the largest possible time delay measurement. This can be scaled to any aperture of hexagonal array. The animal is located at a depth similar to that of the array.

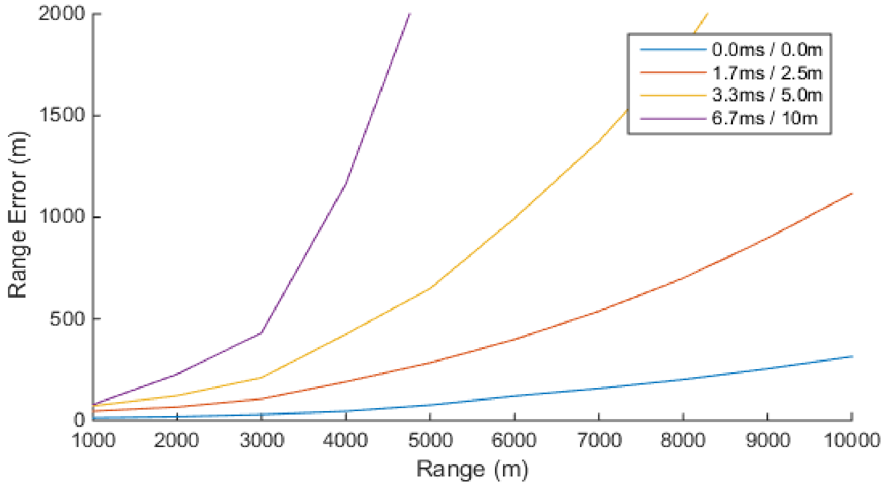


Figure 6. An example of the horizontal range error in localized position versus the range from a 6-element hexagonal array with an aperture size of 1000 m. This is simply scaled data from Figure 6. As is obvious, even relatively small errors in the positions of hydrophones and/or timing errors can result in very large errors in the localized positions of animals.

3.3.10.2 *Localization accuracy and depth*

It is important to note that, when considering localization accuracy, we are dealing with a 3D environment. Figure 7 shows error contours when the localization error (standard deviation) in horizontal range error becomes greater than 10 m at different heights above an array. The timing error used was 0.25% (equivalent to 5 mm error in hydrophone positions) with 5 ms^{-1} error in sound speed and the array aperture was 20 m.

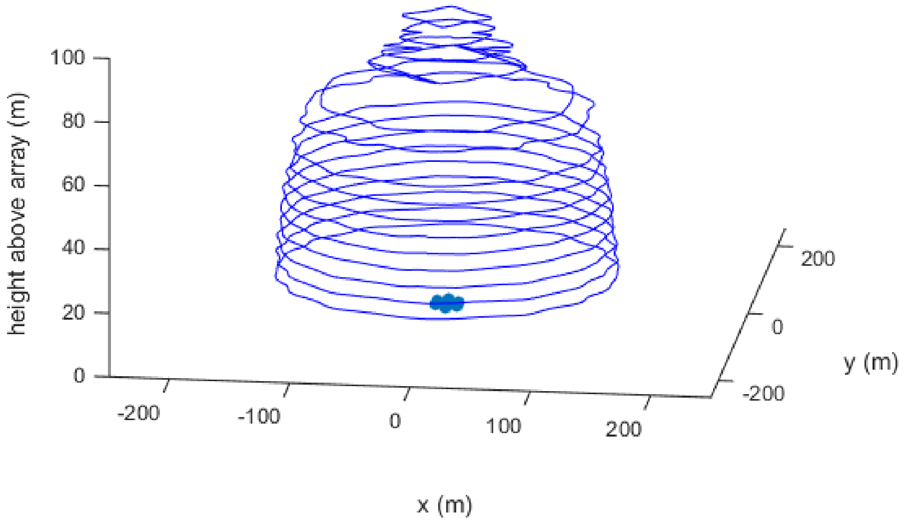


Figure 7. Contours showing the point at which horizontal range error is 10 m at different depths. An array with an aperture size of 20 m is used.

The simulations show that reasonably small errors in time delay measurements and receiver positions can propagate to large errors in localization accuracy. As shown in Figure 7 even a <1/100 of a second error in timing can cause substantial errors on localization. As such, completely[2] unsynchronized systems are will require extremely complex processing to yield reliable localizations and any array must have receiver positions which are known accurately. In addition, Figures 5 and 6 represent a best-case scenario, when an animal is roughly at the same depth as the array. As the difference in depth between a vocalizing animal and the array increases, localization errors also increase in the horizontal plane, as shown in Figure 7. Note that Figure 7 is the depth above the array, so if this hexagonal array was placed in 1,000 m depth it would only be able to accurately localize animals which were diving at >900 m. This example is with a relatively small aperture targeted towards only the highest frequency vocalizations; a 500 m aperture with similar error values would likely cover the entire water column with similar accuracy in 1,000 m of water.

These simulations therefore, demonstrate that timing and/or position errors must be a major consideration in designing a localizing hydrophone array. If an array cannot be synchronized properly and/or receiver locations cannot be surveyed accurately and/or the water depth is significantly larger than the effective range of the array, then the ability to accurately localize animals in the horizontal plane is going to be limited, which will have potential follow on consequences for density estimation as discussed later in section 4.1.

3.3.11 Examples of Localization

There are multiple factors which can be adversely influence localization and simulations show the potential for relatively small measurement errors to introduce large uncertainties in the position localized positions of animals. However, it is important to note that none of these are insurmountable problems. There are many studies in the literature which have carefully designed and successfully utilized localizing arrays to study marine mammals. The next section covers two such studies which have utilized two different seabed PAM systems to localize beaked whales and humpback whales.

---

[2] 'Completely' is used here because synchronization is not necessarily needed between all hydrophones in an array e.g. clusters of closely-spaced hydrophones can be unsynchronized and still localize accurately. Synchronization could also be performed via some other means, e.g. using ambient noise, however this would have to be thoroughly evaluated and errors known.

3.3.11.1 *Beaked Whale*

Beaked whales are a group of deep diving whales. The vocal repertoire of beaked whale species studied to date is mostly stereotyped by characteristic frequency modulated clicks of which there are five distinct types that are likely species specific. The beam profiles which have been measured so far (Shaffer et al. 2013) are all narrow with significant differences between on axis and off axis waveforms; combined with high attenuation in sea water, this makes beaked whales are an exemplar species for the problem of detection matching on widely separated hydrophone arrays. However due in part to their vulnerability to naval sonar (Tyack et al. 2011) they are also a highly studied group of whales and there have been many attempts at PAM localization.

The difficulty in detecting and localizing beaked whales is perhaps summarized best by DTAG studies in the AUTEC naval range, were 82 hydrophones are deployed at 2 km depth with a 4 km hydrophone aperture. Only 11% of clicks detected on a tagged Blainville's beaked whale were detected on 3 hydrophones, the minimum to determine a range, and only 1% were detected on 4 hydrophones, the minimum required for 3D localization (Ward et al. 2008). Clearly hydrophone arrays with 4 km apertures are not optimal for beaked whale localization, not a surprising finding as measured probability of detection for Cuvier's beaked whales is estimated to be near zero at 4 km (Zimmer et al. 2005a) and ~20% for Blainville's beaked whales (Marques et al. 2009).

Gassmann et al. (2015) developed a more practical hydrophone array to study beaked whales. This consisted of two 4-channel and three single-channel autonomous high-frequency acoustic recording packages (HARPs), deployed from a vessel. The hydrophones within each HARP were synchronized; however, there was no synchronization between different HARP units. The typical aperture size between HARP units was 500 to 1,000 m and positions of each HARP unit on the seabed was determined by localizing a vessel at a known location. The 4-channel HARPS consisted of 4 hydrophones separated by ~1 m and so were capable of calculating 3D vectors of vocalizing beaked whales, but not ranges (Wiggins et al. 2012). However, if the beaked whale was detected on both 4-channel HARPs a single 3D location could be resolved with any coherent detections on the single channel HARP units further enhanced the accuracy of the localization.

Results show that the array was capable of tracking beaked whales over an area of 20 km^2 and several convincing tracks are presented suggesting that clicks were consistently coherently detected at hydrophone separations of 500 m. However, little in the way of localization accuracy is presented. Gassmann et al. (2015) note that the system could not resolve movements less than 20-30 m; further evaluation of localization accuracy could be a focus of a feasibility study if this type of system was to be employed. Gassmann et al. (2015) also use a semi-manual click matching technique when multiple animal are present, which is not ideal for automated analysis of data collected over significant time periods, however alternative sophisticated click matching algorithms have been developed for beaked whales and could be applied to these data (Baggenstoss 2013b).

The approach by Gassmann et al. (2015) is advantageous for number of reasons. Autonomous units, such as HARPs (or cheaper alternatives once developed) provide flexibility both in the length of time it takes to set up the array and the potential areas which can be studied (and for incorporating technological improvements as they become available) (Figure 8). Species-specific detection, classification and localization algorithms are well developed and so although beaked whales are a less than ideal candidate species for localization, a consistent research effort over the last few decades has tackled many of the most intractable problems.

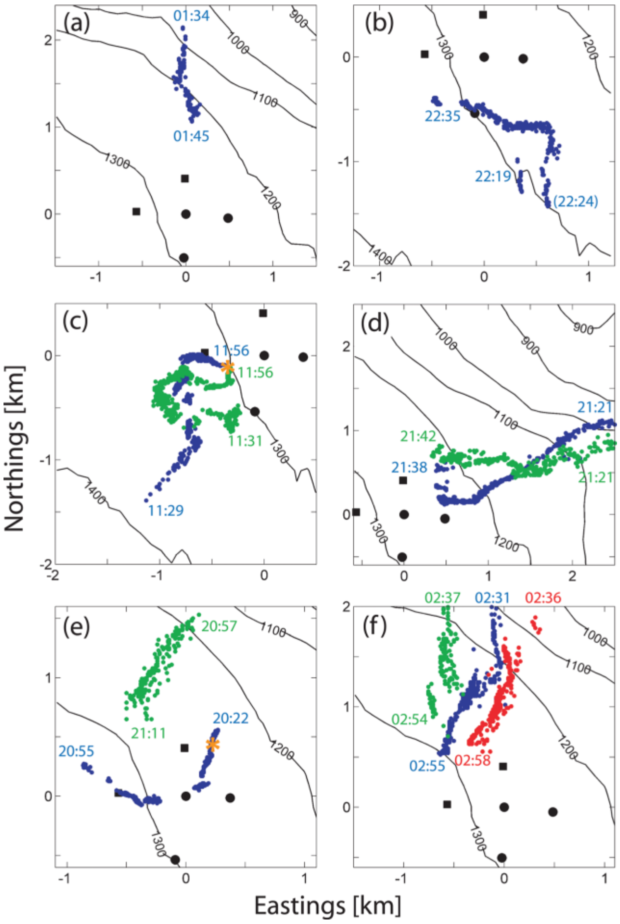


Figure 8 - Reproduced from Gassmann et al. (2015). This shows various examples of beaked whale localizations from an autonomous seabed-mounted HARP array (encounter periods lasted 11-33 minutes). Different colors represent different animals. The square dots are 4-channel HARP units and the circular dots are single channel HARP units.

3.3.11.2 *Humpback Whale*

Baleen whales are generally more likely to be detected coherently on widely-spaced hydrophones. However, the disadvantage is that they generally vocalize less frequently and more unpredictably than odontocetes and hence the efficacy of PAM in general is limited.

However, there have been many studies which have used localizing arrays to track baleen whales; these include opportunistic studies with seismometers, military hydrophone ranges and hydrophones used to enforce the nuclear test ban (e.g. McDonald et al. (1995), Dunn and Hernandez (2009), Wilcock (2012)) to the targeted deployment of floating DIFAR buoys for real time tracking of blue whales in the Antarctic (Miller 2012).

Here, we focus on a case study by Helble et al. (2015) on humpback whales, which have amongst the highest frequency vocalizations amongst baleen whales. The study is based in Hawaii within a 14-element bottom-mounted hydrophone array. These types of dispersed bottom-mounted arrays have been used in many previous studies to track baleen whales (Tiemann et al. 2004, Martin et al. 2013). The focus in these papers is on tackling detection matching problems and ensuring that the data analysis methods are highly automated, something which will be vital for a pragmatic approach to DECAF.

Humpback whale songs consist of individual units. These units can be similar amongst different individuals and units from an individual can be repeated (Payne and McVay 1971). Thus, the detection matching problem discussed in above is as relevant to humpback whales as it is to beaked whales. To address this Helble et al.

(2015) used an a generalized power law detector (Helble et al. 2012) to automatically extract humpback whale tonal vocalizations from raw sound files on all hydrophones within the array. A combined detection matching and localization algorithm developed by Nosal (2013) was then used. As discussed in above this algorithm works by generating a probability surfaces between different pairs of hydrophones in a wide baseline array. No attempt is made to match detections; a surface is generated for every possible pair of detections which can generate a realistic time delay measurement (i.e. a time delay which is no greater than the time of travel between the two hydrophones). These surfaces are then multiplied together and consistent peaks are clustered; persistent peaks represent tracks of one or more animals - as such, this method is capable of automatically tracking multiple animals through any large baseline array.

Helble et al. (2015) demonstrated that using a bottom-mounted array it is possible to localize humpback whales in real-time using highly automated methods which could potentially be applied to a wide variety of species. They were further able to demonstrate that such methods are robust to a variety of environmental and noise conditions, however a drawback of the study is that localizations are only performed in 2D and although localization errors are simulated, there is no true ground truth on localization accuracy. 2D localization may be appropriate for longer ranges, where the water depth is a much smaller percentage of the overall range or where the depth distribution of animals is well known, however this also means these methods may not be applicable to all species without further testing.

3.3.12 Implications of Localization for Density Estimation

Two methods of density estimation are explored in this document: distance sampling and SECR. Localization errors affect both methods, though distance sampling is most affected, as SECR does not rely on localization.

In a DECAF context a distance sampling framework requires monitoring points, or nodes, from which horizontal ranges to vocalizing animals can be estimated. In general, a hydrophone array is required at each node that, at a minimum, must be capable of determining a horizontal range to a vocalizing animal. If it is not possible for arrays to be placed at all nodes, then a mix of localizing arrays and sensors with single hydrophones may be possible.

The major weakness of this approach is that, for some species, the probability of localization will not be the same as the probability of detection. In fact, it may the case that the probability of localization is both different in magnitude (i.e. less likely) and does not scale linearly with the probability of detection at different ranges, i.e. the two functions are different shapes. This is an area that should be addressed in any future development of this tool as it will have implications for the tool simulations.

Many of the issues above contribute to this problem; however the range dependence of localization errors and detection matching are the two key factors most likely to exacerbate a dichotomy between probability of detection and the probability of localization. As shown in section 3.3.8 the error in localization increases significantly with range. Thus, for many situations, the effective operating localization range of a hydrophone array is significantly less than the detection range. Detection matching is another major contributing factor. Errors in detection matching mean that, even when a detection has occurred, a localization may not be possible. Additionally a minimum number of hydrophones may not be triggered. For example, consider an animal with a very narrow beam profile next to the hydrophone array. It may be very close to a hydrophone element but facing such a way that only one other hydrophone in the array is ensonified, making localization impossible. Thus the probability of detection at the individual hydrophone is very high but the probability of localization for the array is very low. In such instances it might be possible to identify an unbiased subset of detections in order to estimate the detection function. Alternatively, it might be possible to use SNR to establish a relationship between those calls that are localizable and those that are only detectable. However this is a very novel research area and requires further investigation.

SECR is less impacted by the challenges of localizations as, unlike distance sampling, the method does not require localizations of, or even ranges to, vocalizing animals. This makes SECR an attractive approach; the effects of localization error do not need to be addressed. However, if localization is possible, then care must be taken in an SECR analysis not to use localization algorithms to facilitate the detection matching process across instruments. If this was done, and non-localizable detections were omitted from the analysis, then the same issue as distance sampling is created: the estimated probability of detection would be relevant for localizable calls only.

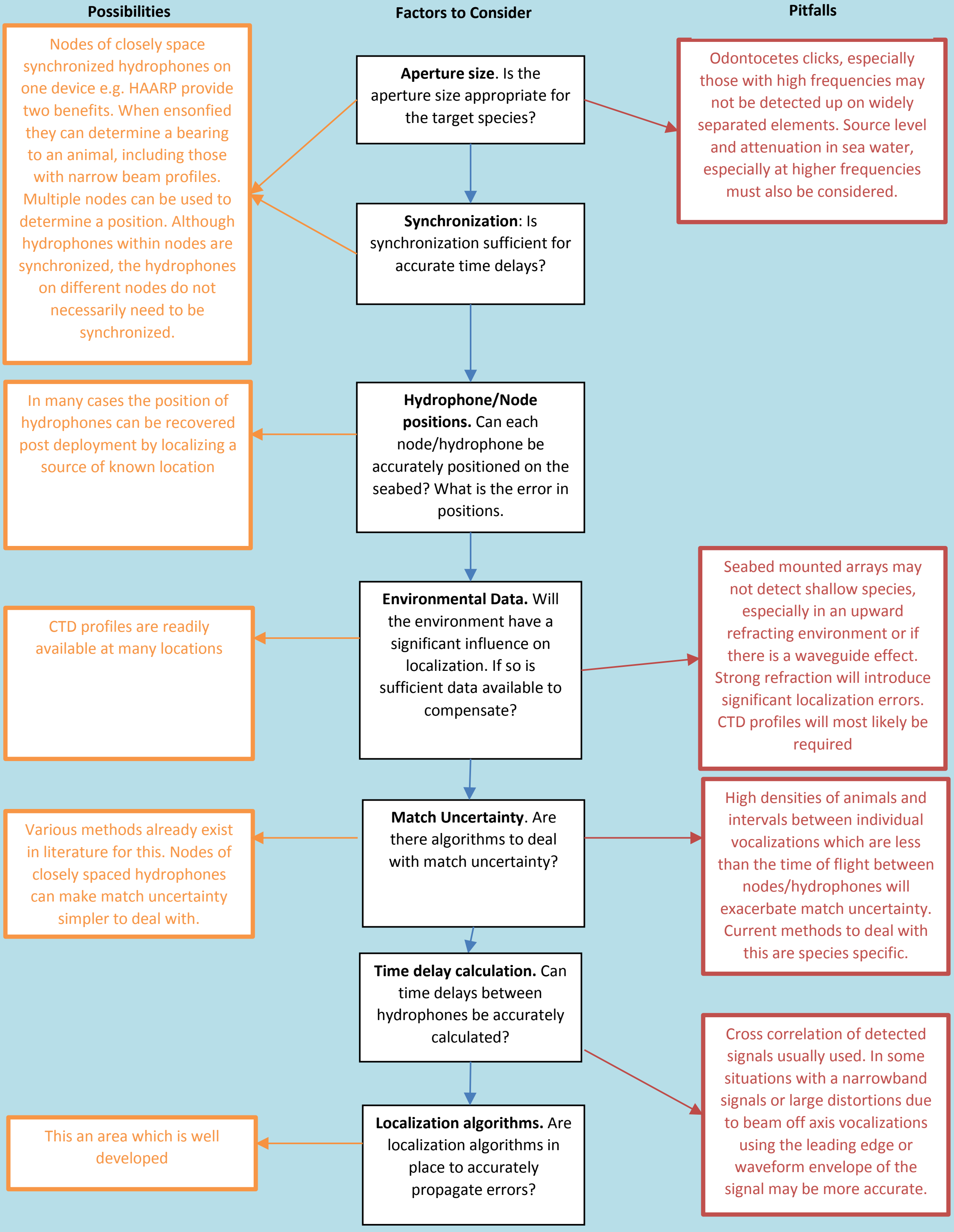
3.3.13 Localization Summary

Localization is extremely challenging and labor intensive in practice, but is important to estimate as precisely as possible (as errors will affect the power of a PAM survey) and these challenges are summarized below. Hardware and analysis algorithms must be appropriate for the environment and study species. A multitude of errors must be accounted for, sophisticated signal processing techniques are needed to match vocalizations on different hydrophones and probabilistic localization algorithms are needed to accurately propagate errors. Although numerous, these are not insurmountable problems, there are plenty of examples where these challenges have been overcome on a species by species basis, as the two above case studies demonstrate.

The main difficulty in localization is therefore, designing an array which can be used to study *multiple* species with highly variable vocal repertoires. However as discussed above many of the challenges in localization, for example determining an optimum aperture size, are dependent on vocal properties and therefore *species specific*. Whilst a single broad band hydrophone may be able to detect the vocalizations of all marine mammal species, from large baleen whales to small odontocetes, it is challenging to design a single localizing array which can work over the multitude of scales required to localize these different species groups. Instead a targeted approach is required, were perhaps different types of arrays are integrated and different localization algorithms and even density estimation techniques are used depending on the type of detected vocalization.

A second issue arises when using localizing arrays to determine the probability of detection. It is unlikely that the probability of localization is the same as the probability of detection. For some species, such as baleen whales it may be very close. However, for other species, especially those with directional vocalizations, it is likely the probability of localization is neither the same magnitude as, or linearly scales with, the probability of detection. The difference between the probability of localization and detection will therefore also be species specific. Further research is required to evaluate the magnitude of this problem and effects it may have on density estimation. The pitfalls and possibilities of localization are outlined in the Box D below.

Box D: Flowchart - The possibilities and pitfalls of localization



Possibilities

Nodes of closely space synchronized hydrophones on one device e.g. HAARP provide two benefits. When ensonfied they can determine a bearing to an animal, including those with narrow beam profiles. Multiple nodes can be used to determine a position. Although hydrophones within nodes are synchronized, the hydrophones on different nodes do not necessarily need to be synchronized.

In many cases the position of hydrophones can be recovered post deployment by localizing a source of known location

CTD profiles are readily available at many locations

Various methods already exist in literature for this. Nodes of closely spaced hydrophones can make match uncertainty simpler to deal with.

This an area which is well developed

Factors to Consider

Aperture size. Is the aperture size appropriate for the target species?

Synchronization: Is synchronization sufficient for accurate time delays?

Hydrophone/Node positions. Can each node/hydrophone be accurately positioned on the seabed? What is the error in positions.

Environmental Data. Will the environment have a significant influence on localization. If so is sufficient data available to compensate?

Match Uncertainty. Are there algorithms to deal with match uncertainty?

Time delay calculation. Can time delays between hydrophones be accurately calculated?

Localization algorithms. Are localization algorithms in place to accurately propagate errors?

Pitfalls

Odontocetes clicks, especially those with high frequencies may not be detected up on widely separated elements. Source level and attenuation in sea water, especially at higher frequencies must also be considered.

Seabed mounted arrays may not detect shallow species, especially in an upward refracting environment or if there is a waveguide effect. Strong refraction will introduce significant localization errors. CTD profiles will most likely be required

High densities of animals and intervals between individual vocalizations which are less than the time of flight between nodes/hydrophones will exacerbate match uncertainty. Current methods to deal with this are species specific.

Cross correlation of detected signals usually used. In some situations with a narrowband signals or large distortions due to beam off axis vocalizations using the leading edge or waveform envelope of the signal may be more accurate.

3.4 PAM Systems for DECAF

Box E: Why is this topic challenging?

There are a range of n-node (detection only) systems available and many of them are extremely cost effective, portable systems. Most of these are autonomous detectors or recorders that require retrieval. At the other end of the spectrum there are fixed cabled arrays (these typically exist for other purposes (e.g. military) but provide an opportunity for marine research.

Sousa-Lima et al (2013) and previous IOGP reviews provide a comprehensive update. Here we only present new or updated systems/technology. For DECAF approaches, the choice of system will depend heavily on species of interest and site-conditions (e.g. depth, distance to shore/infrastructure).

In terms of hardware to be deployed for density estimation, (notwithstanding the need for auxiliary data to inform cue rates for the target species) range estimation is the key. However, there are no off-the-shelf options, fixed cabled arrays offer 'easy' power and data transfer solutions, but may be restrictive and expensive for short duration studies. However cabled systems may offer cost effective solutions over decadal time scales compared to using multiple individual recorders. Some innovative systems (e.g. HARP) allow for a more cost-effective and flexible monitoring approach (cf. fixed, cabled systems developed for other purposes) – but still proprietary and of limited availability to mainstream users.

For all systems, data volumes are an issue (to a greater or lesser extent dependent on species vocal behavior). Therefore hybrid systems and streaming systems have added value – they lower data volumes and analysis times required.

The need on an oil and gas E&P field to estimate density for a range of species or species groups will likely require some form of bespoke specific 'nested array' which has different spacings for different species (to make an array multifunctional).

3.4.1 Introduction to fixed PAM surveys

For any fixed PAM survey, the selection of equipment will be one of the main factors which determine whether or not required information can be extracted from the data. Purchase, deployment and subsequent servicing of equipment are also likely to be significant cost elements. Selecting suitable equipment which is capable of collecting the required data in a reliable way over the required lifetime, without incurring unnecessary cost is therefore one of the key challenges in planning a PAM study.

The design of PAM survey to support any risk assessment must start with some indication of the likely species (or species groups) of interest for the survey (**the target species**). In the context of fixed PAM surveys providing density estimates to contribute to risk assessments, it is likely that some form of literature or data review for the site of interest has been completed, and therefore a list of target species will have been identified in advance of the planning of a survey. In addition, such reviews will have likely identified whether the species is common to the region or rare and also what sensitivities/protections are in place (further guiding the selection of species in the scope of a PAM survey).

Such prior planning will be crucial as the selection of equipment needs to consider the vocal repertoire of the species. That is, the frequency range of the vocalizations produced by the species of interest (to ensure that the PAM systems chosen have sufficient **sampling frequency** and suitable **hydrophone elements** capable of detecting and recording the calls). Prior knowledge of likely **detection ranges** for each species is also essential for array design. In addition, understanding the temporal occurrence of species – are they in the region year-round or do they migrate to/through the site at particular times of year (which can bound the period in which is crucial to have PAM survey effort). This also applies to species which tend to vary their calling behavior with particular times of year. In addition, there may be age/sex class variation in vocalization rate. Understanding the factors that cause the species vocalization rate to vary within the survey region, is important as all can dramatically affect density estimates. Understanding this temporal component provides guidance on the selection of PAM equipment as data collection can be limited by the **battery life** and/or **data storage capacity** of the PAM system, i.e. some systems require periodic at-sea servicing. Other PAM systems can be powered from land or an infrastructure hub, to which data can be streamed. The utility of autonomous and cabled systems, along with some of the trade-offs of their use, is discussed later.

The selection of suitable PAM equipment must consider the site of interest. This involves considering elements like **distance from land or infrastructure** hub (if cabling systems and/or streaming data), the **water depth**, the **bottom topography** and **substrate**, the **exposure of the site** (e.g. whether it is in a very sheltered region, subject to significant adverse weather etc.), the **background sound** sources on the site and what **other marine users** or **stakeholders use the same area** (e.g. high shipping, active fishing grounds etc.) which can affect how deployments are designed or how systems are installed.

A final fundamental consideration in the choice of PAM system is the objective of the monitoring effort. In this study we are principally interested in the use of PAM for estimate the density of marine mammals and to detect changes in density) during the life cycle of an industrial activity.

Therefore choice of system is dependent on both the technical elements mentioned above and the requirements for density estimation. The main requirements are:

- Ability to detect a wide range of vocalizations of the target species. Hydrophones and recording systems must have broadband frequency ranges and appropriate sensitivity.
- A system designed which can calculate a range to animals and therefore a probability of detection. This requires that at least some nodes are localizing hydrophone arrays introducing a degree of extra complexity in both hardware and data analysis.
- A system which can localize different species effectively. This likely requires significantly different sizes of hydrophone array for different groups of species as discussed in section 3.3.7.

The JIP has previously funded two key projects that have comprehensively reviewed the state of technology available to support PAM efforts for many of the key elements outlined above (Moretti et al. (2009), Norris et al. (2010)). One of these reports led to an even more focused review of PAM systems (Sousa-Lima et al. 2013). However, technology for PAM is in a constant state of development requiring periodic re-evaluation of capabilities. **Therefore the objective of this report is not to repeat those efforts, but to provide a complementary update on the development and availability of PAM technologies.** In addition, this report follows two other JIP reports, one exploring methods for monitoring marine mammals in low visibility environments (PAM was covered extensively) and a second focused on data collection via autonomous platforms (where a number of new systems were covered) (Verfuss et al. 2015, Verfuss et al. 2016). Here we provide a summary of those new systems that can be deployed in a fixed PAM effort. Crucially, this overview will be focused on discussing the components of a PAM survey which can limit or enhance any PAM effort specifically to generate density estimates. Coupled with the development of DCL efforts, the development of PAM hardware has advanced the field of such monitoring significantly over the past decade and more (as technology advances, more sophisticated analyses have become possible and vice versa). **Therefore, understanding these topics is integral to a successful long-term fixed acoustic monitoring of potential**

changes in density of marine mammals throughout the life cycle of an offshore E&P field. Given the volumes of data that can be generated now with the advances in PAM technology, DCL plays a crucial role in keeping analysis manageable. This is discussed extensively in section 3.2. With respect to PAM technology, each of the elements noted above are discussed and described in further detail below (3.4.2 – 3.4.6).

3.4.2 Sampling rate and data volumes

Most PAM systems include a human operator in any decision-making process, whether that be real-time decision making or post-analysis of data. By their very nature, remotely-deployed platforms do not accommodate a human operator, so data must be stored or transmitted in a form that retains sufficient information for detection verification purposes.

PAM systems for autonomous data collection can broadly be divided into three categories:

1. Raw data systems
2. Systems with onboard data processing.
3. Systems that can collect both raw data and processed data (hybrid systems).

The increased capacity of modern hard drives, and more recently of flash memory cards, has led to the development of a wide variety of systems that simply stream raw data to storage media for later analysis. As devices increase in acoustic bandwidth, so the pressures on data storage, data processing and power requirements become more acute. It is important to consider this element further before exploring the systems that are available. Addressing the question of how one deals with the volumes of data produced should be done before selecting a particular model of PAM system as the cost implications of processing data are significant and by the end of any monitoring effort will far exceed the cost of purchasing or leasing PAM hardware. Therefore, a key consideration in the design of a long-term PAM survey is that of the volume of data to be generated and the subsequent resources required to process and analyze the data over the duration of the monitoring effort.

Unless the levels and types of background noise in the study area are well understood it is impossible to predict in advance of a deployment, which detection and classification algorithms are likely to be most effective, or exactly how algorithms should be configured. Wherever possible therefore, raw acoustic data should be archived in order that algorithms can be tested and adjusted if required and there is a strong case for carrying out preliminary studies, obtaining data from the site (probably using simplified PAM methodology) against which DCL algorithms can be tested. This is particularly important for life of field deployments. Storing raw data can however require high volume storage which may not be available on small, low cost devices. Data volume per unit time increases with the number of channels recorded, the sampling rate, and the sample size (bit depth). Storage volumes for different sample rates and deployment durations are shown in Table 4, indicating the storage capacity required for deployments of a single hydrophone sampling with 16-bit precision (up to 96dB dynamic range). If multiple hydrophones are used, then the data volume should be multiplied accordingly. If 24-bit rather than 16-bit digitization is used, then the volumes will increase by a factor 1.5. A 24-bit system allows for a larger range (dynamic range) of amplitudes to be measured (assuming the hydrophone and preamplifier are also capable of the same range). This may be important for some areas where there is an especially large difference between the smallest and highest sound levels. For scientific data collection, it is not possible to use lossy compression (see Glossary) algorithms such as MP3 since these are likely to throw away important data - they are designed to keep components of sound pleasing to the human ear, and may discard components of sound used by marine mammals. There are however, a number of data compression algorithms that are lossless and can be used to reduce data volume. All readers are familiar with the zip file compression system; this is not however, efficient for compressing sound data. A widely used compression system for music is FLAC (https://xiph.org/flac/), which can achieve compression ratios of around 3:1. Similar levels of compression can be achieved with lower processing power, using a bespoke algorithm (known as X3) developed by Johnson et al. (2013). X3 exploits the fact that most PAM data is dominated by low frequency background sound interspersed by loud transients, which is quite different than the musical applications for which FLAC was designed. If X3 is used, for planning purposes it is generally possible to assume that data volumes will be approximately 33% of those given in Table 4. The table also indicates the kinds of species groups and vocalizations that are possible to record at different sampling rates.

Table 4. Data volume for deployments of a single channel PAM recording system digitizing with 16-bit accuracy.

	Deployment Duration				
Sample Rate	**1 day**	**30 days**	**90 days**	**365 days**	**Species Accessible**
1 kHz	0.16 GB	4.8 GB	14.5 GB	58.7 GB	Blue and fin whales
2 kHz	0.33 GB	9.7 GB	29.0 GB	117.5 GB	Baleen whales (and those above)
48 kHz	7.7 GB	231.7 GB	695.2 GB	2,800 GB (2.73 TB)	Sperm whales, most dolphin whistles and some clicks (and those above)
192 kHz	30.9 GB	927.0 GB	2,700 GB (2.63 TB)	11,000 GB (10.74 TB)	Sperm whales, most dolphin whistles and clicks and most beaked whale clicks (and those above)
500 kHz	80.5 GB	2.4 TB	7,100 GB (6.93 TB)	28,700 GB (28.03 TB)	Dolphin, beaked whale and porpoise clicks. (and those above)

For baleen whales whose vocalizations can be recorded at low sampling rates, a year of continuous recording might generate only a modest amount of data. Even sampling at a standard human audio rate of 48 kHz for a year would generate no more data than can be fitted on a single portable hard disk drive (<3 Tb). However, for the majority of dolphin, beaked whale and porpoise echolocation clicks, higher bandwidths are required, which generate a significantly greater volume of data. The majority of systems reviewed by Sousa-Lima et al. (2013) had data storage capacities of 1-64 GB; only a few had greater capacity – most notably the HARP which has 1.92 TB onboard storage. Therefore, all of these systems will be limited by their data capacity for any species group other than baleen whales. However, it is reasonable to predict that in ten years the technology will have developed further, such that longer duration deployments with fewer maintenance visits to recover data will be possible.

A common technique applied to extend deployment duration is the duty-cycling so that a PAM system will periodically collect data for a short period before becoming dormant. There are savings for battery power and data storage since no data are collected whilst dormant. Duty-cycling can cause limitations however, as important signals may be missed – particularly for rare species or species that vocalize only infrequently. For any study limited by the sample size of encountered animals, duty-cycling will mean that either more devices are required or that deployments need to last longer, therefore achieving little overall improvement in data collection coverage or duration.

Cabled systems are not governed by these limitations of power and data-storage; they are powered from land or from local marine infrastructure meaning that raw data can be collected continuously. Also, any issues with cabled or telemetry-linked PAM units can be identified immediately and addressed accordingly – problems with autonomous units may not become evident until their data have been retrieved. This is a clear benefit of fixed, cabled systems compared to autonomous, battery-powered recorders. The cost implications of installing a cabled, fixed array of multiple sensors may be high (Moretti et al. 2009)(section 3.4.6), but not prohibitively so, as considering the duration of a life-of-field monitoring project this may be a cost effective option. Data loss due to malfunction or damage to autonomous units may be mitigated against to some extent by deploying back-up units, and by servicing the units at shorter intervals so as to detect any data collection problems sooner.

3.4.3 Previous Sound and Marine Life-JIP PAM system reviews

In the following sections, we explore the two main types of system which can form a larger PAM array for density estimation. These include n-node systems that can only detect (and potentially classify) marine mammal vocalizations (3.4.4) and p-node systems that are capable of providing a range to the animal (section 3.4.5). Moretti et al. (2009) provides a comprehensive review of the potential of cabled PAM systems in density estimation and therefore we build on their recommendations and primarily focus here on novel developments and alternative approaches to collecting the input data required for density estimation In 2013, the Sousa-Lima et al. (2013) review focused on the capabilities of fixed, autonomous acoustic recording devices:

> *"Fixed autonomous acoustic recording devices (autonomous recorders [ARs]) are defined as any electronic recording system that acquires and stores acoustic data internally (i.e., without a cable or radio link to transmit data to a receiving station), is deployed semi-permanently underwater (via a mooring, buoy, or attached to the sea floor), and must be retrieved to access the data."* – Sousa-Lima et al. (2013).

3.4.4 Advances in PAM technology since the last SML-JIP reviews

In their discussion, Moretti et al. (2009), Norris et al. (2010) and Sousa-Lima et al. (2013) concluded that there were numerous technical areas in need of development. Some research areas of PAM technology recommended to be further developed by those reviews were:

- increased power capacity and low-power systems (section 3.4.4.1)
- systems capable of wireless transmission of data (section 3.4.4.2)
- installation of an expansive fixed cabled array (see section 3.4.6)

Has the field of PAM developed as these reviews suggested? This section reviews progress in PAM hardware with particular focus to how advancements aid density estimation efforts using PAM.

3.4.4.1 *Increased power capacity and low power PAM systems*

The development of such systems will be important for E&P life cycle developments spanning 30-50 years as it will increase deployment durations (over higher power systems) and thus is an area for further investigation and development. However, based on the current state of knowledge and because DECAF application is the focus here, and highlights a paucity of information in a number of research areas required to estimate density. For example, given the wide range of 'off-the-shelf' detection-only systems, but the crucial role that DCL plays in generating density estimates in a cost-effective manner - our focus is on how systems have developed to improve and streamline capabilities to support PAM deployments for density estimation.

Hence, one such area of growth, aligned with the recommendations of Norris et al. (2010), is the development of 'hybrid' systems that are capable of monitoring for a wide range of species but use onboard detection and classification algorithms to reduce onboard data storage requirements (particularly for high frequency data) and to streamline analysis. Hybrid systems can transmit data in real time which provides data security in the event of equipment loss and speeds up the analysis process however these advantages and the likely cost savings from not having to physically recover data from devices must be weighed against the extra contextual information that raw data can provide. The key consideration therefore, when deploying an automatic or hybrid system, is how much data analysis should be done on board, and how much back on shore. As discussed above, storage of raw data allows for more detailed analysis, but manually analyzing all raw data can be time consuming - most data will only ever be processed using automatic algorithms in any case. While this is often the case, the key thing is that we often do not know in advance is which parts of the data can reliably be left to automatic algorithms and which will require more intervention. In designing a hybrid system, it is generally safer therefore, to retain as much data as possible, using automated algorithms to discard the minimum amount of raw data that will prevent the device storage from filling up rather than trying to deploy algorithms which do the maximum amount of processing and retain a minimum amount of data.

3.4.4.1.1 *SoundTrap [Ocean Instruments NZ]*

An example of a hybrid PAM system is the SoundTrap, which is a flexible system capable of operating in a range of modes. With lossless compression, the 128 GB SD card will provide 65 days of continuous recording at 36 kHz (and 48 days at 48 kHz) but can only record for 4.5 days at its maximum sample rate of 576 kHz. However, on-board processing allows the device to sample at its full rate of 576 kHz, but decimate data to 48 kHz and store the decimated data in X3 format. At the same time an automatic click detector processes the full rate, high frequency data with a simple energy-based click detector. Waveforms of detected clicks are also stored (also in X3 format), but since these comprise only a very small fraction of the raw data, the clicks do not contribute significantly to overall storage volumes. This means that the device can operate for nearly as long as if it were only recording at one of the lower sampling rates. The simple click detector runs at a high false positive rate, so detected clicks must be further processed (either automatically, manually or both) on shore to assign detections to species or discard them as background sound. In our experience, click data alone is sufficient for detecting and classifying odontocete echolocation clicks and importantly, the single simple detector can detect clicks from many species. The many types of sound (baleen whale calls, dolphin whistles, sperm whale clicks) occurring at lower frequencies can all be detected through offline processing of the lower frequency raw data.

3.4.4.1.2 *Decimus [SA Instrumentation Ltd]*

The Decimus system is another hybrid system which combines recording with automatic detection. It contains a processor running embedded versions of the click and whistle detection algorithms from the PAMGuard software and can be configured to detect a wide range of different sounds. It consists of a low power embedded processor and a high speed (500 kHz) four channel acquisition system. The processor has been programmed to detect both echolocation clicks and whistles in real-time and to either store or transmit detection data only. Enough raw data (short sound clips varying in length from less than a millisecond for an echolocation click to up to a second for a dolphin whistle) for each detection are stored so that an operator can still make a final decision and reject false detections. The exact amount of stored data depends on the frequency of encounters and the false detection rate from background noise – but Decimus can reduce a daily raw data volume by scaling 10,000:1. This volume of data can be stored for years before filling even a moderate sized SD card and can also be transmitted to shore using either 3G or Wireless modem technology. Where transmission of even these volumes of data becomes impractical (e.g. when using a satellite link) a limited summary of system status can be transmitted to at least inform operators that the system is running correctly.

A disadvantage of this system is that although it is acquiring data on four channels, the detectors are only processing a single channel of data and are only using the other channels to measure phase (time delay) information. This means that the hydrophones need to be closely-spaced and a single system is only suitable for measuring bearings. In addition, the system lacks a marine housing and requires integration into a platform. This system has been successfully used both on moored buoys and on a Liquid Robotics Waveglider for the detection of harbor porpoise and other odontocetes.

Since the 2013 review, other standard, 'detection-only' systems have been further developed – many of which are low power, helping extend battery life (e.g. microMARS). These are described below.

3.4.4.1.3 *SM2M / SM3M [Wildlife Acoustics]*

Sousa-Lima et al. (2013) covered this system's predecessor, the SM2M. This is a well-established, recording only unit which can sample at up to 384 kHz (Recording bandwidth 2-192 kHz). They are usually used in moored systems but have also been used in drifting buoy systems. Their longevity is heavily dependent on target species (and so the chosen sampling rate) and like most systems can be augmented by duty-cycling.

3.4.4.1.4 *AMAR G3 [JASCO]*

This system has been updated since 2013. Now the unit can be configured with up to eight hydrophones (a range of different hydrophones can be used) and up to 1792 GB (1.75 TB) solid state hard drive (a single drive for all data collected). It also can be housed in a range of depth-rated housings, spanning 250 m, 2,500 m and 6700 m. Lifetime can be extended over one year, but this depends on the number of hydrophones used, the target species (sampling rate) and the duty-cycling configuration.

3.4.4.1.5 *microMARS (Desert Start Systems LLC)*

The microMARS system did not appear in the Sousa-Lima et al. (2013) review, but like many of the newer systems, is small, portable and has potential for extensive hard drive space (4096 GB (4 TB) maximum for 'eight card version'). It is depth rated to 300 m and samples at up to 250 kHz.

3.4.4.1.6 *RS-ORCA [Seiche & RS AQUA]*

The RS-Orca supports multiple input channels with sample rates up to 768 kHz and up to 4 TB of internal solid state storage and rated down between 200 – 3,500 m. The RS-HILO is a real time, dual channel, underwater acoustic data acquisition instrument designed for underwater noise measurements for 'long term deployments' – though the 128 GB SD card requires duty cycling to extend deployments up to 1 year.

3.4.4.2 *Data transmission (wireless) systems*

All of the systems reviewed by Sousa-Lima et al. (2013) collect and store data internally, i.e. without a cable or radio link to transmit data. Certainly new systems that require less 'hands-on' maintenance have been developed in recent years. Much has focused on the wireless transmission and remote control of PAM systems. A recent SML-JIP funded report (Verfuss et al. 2015) provides extensive information on systems that can be used for real time data collection and transmission from autonomous vehicles. In principle data collection for DECAF could be done using many of those vehicles and we refer the reader to that report. However here we are principally focused on those new developments which can be deployed in a fixed PAM array and can improve our capabilities to robustly estimate density using PAM data. The following systems have been developed, which have the potential to advance this field.

3.4.4.2.1 *Seiche Wireless Detection System (W-PAM or WDS) [Seiche Ltd.]*

The Seiche WDS is a highly configurable system which can be utilized for true real-time data transmission, with a wireless transmission range of up to a theoretical maximum of 10 km, typically using 2.4 GHz/5 GHz bands. Other Seiche wireless systems have incorporated a satellite data-link to enable truly remote monitoring. A typical system configuration would acquire data from 1-4 hydrophones, at a base sampling rate of 500 kHz per channel. It has two modes to enable real-time monitoring: a) In Mode 1, an analogue-to-digital sampling device is installed within the unit on the transmitting platform (e.g. a network of moored buoys, a raft or unmanned surface vehicle USV) and the full dataset is transferred to a PC at the receiver station (e.g. a support vessel or infra-structure hub) for processing in PAMGuard. The analyst can then view and utilize the full required frequency range for monitoring in true real-time. This configuration requires a wide bandwidth to allow receipt of full dataset at the receiver station. This mode has been utilized during the construction phase of offshore wind farms (e.g. Höschle et al. 2013; Nehls et al. 2015).

b) In Mode 2, an electronic processing unit is installed on the transmitting platform within the unit, where the audio signal is processed in PAMGuard. The analyst at the receiving station has full control and viewing access of incoming LF and HF data within the PAMGuard user interface via a remote software link. This configuration requires less bandwidth and offers greater telemetry range. Additionally, no data loss is suffered should the link be lost as it is possible to record data at the base unit. This mode has been used, for example, for marine mammal surveys using USVs (Pierpoint et al. 2016).

Autonomous versions of the Seiche system (record only, no transmission) have been utilized for sound source verification using moored and drifting buoys, soundscape monitoring in shallow habitats using rafts, marine mammal detection from a remote semi-submersible drilling rig, and with the AutoNaut wave-driven USV (C. Pierpoint, Seiche Ltd., pers. comm.).

3.4.4.2.2 *Coastal Acoustic Buoy (CAB) [SMRU Consulting]*

This system is a small buoy that can be deployed from a small vessel with two people (weight is ~100 lbs). It houses a Decimus system (section 3.4.4.1.2) and therefore allows a hybrid approach. The CAB can record raw acoustic data (from up to four hydrophones) on the unit which can be analyzed at a later date, and transmit PAMGuard binary files (kilobytes in size) via wireless transmission (900 MHz) in real time up to 10 km from the unit.

3.4.4.2.3 *SDA14 and SDA416 [rtSYS]*

Developed by rtSYS, these devices can sample at up to 1 MHz. Four channels are available per board and boards can be connected for more channels. The rtSYS systems have been trialed on wind farm construction sites. The systems can measure noise in 1/3 octave bands and perform click event detection and similar to the Seiche systems, data can be streamed back to a control station over short ranges. They have been deployed on autonomous underwater vehicles (AUVs) and buoys.

rtSYS have recently been experimenting with underwater data transmission. However, this development is still very much in early trial stages.

3.4.4.2.4 *Auto Detection Buoys [Cornell University]*

Developed for the detection of North Atlantic Right Whales (for which only a low sampling rate is required) from buoys in the approaches to Boston Harbor, the systems monitor in real time, while software searches for specific sound types and sends a sample of candidate detections to shore for human verification. To date, this has been used extensively on bespoke moored buoys with a relatively low sampling rate of ~4 kHz).

3.4.5 *Systems for range estimation*

As noted above, there are a multitude of systems available that are capable of detecting marine mammal species. However, more crucial to any effort to estimate density from PAM data using Distance Sampling section 4.1.1) is the integration of systems capable of localization into a PAM array. This is because the localization systems measure range to detected calls and can therefore provide us with the data to generate detection functions (i.e. how detection probability of cues vary with range from the PAM device).

The hardware options for PAM, as discussed in the previous sections, are numerous, not prohibitively expensive and generally easy for researchers to purchase or access. These could be called 'off the shelf' solutions.

However, using PAM systems to localize introduces a new level of complexity and there are currently, as of the writing of this report, no 'off the shelf' solutions. Neither is there a comprehensive software package which comes close to being able to handle the multitude of different algorithms required to localize different species on different types of hydrophone array effectively (discussed in section 3.3). The most comprehensive software packages currently available are PAMGuard, Raven and Ishmael.

Generally, the hardware that has been used to further the field of localizing cetaceans has either comprised of a hydrophone array deployed from a vessel / surface buoy or a pre-existing seabed-mounted fixed array system (usually military) from which marine mammal acoustic data has been appropriated opportunistically.

3.4.5.1 *Vessel-mounted Systems – towed arrays*

Many surveys, mitigation strategies and research projects use towed arrays to collect acoustic data. Towed arrays generally consist of a linear array of two or more hydrophones towed a few hundred meters behind a ship. The array can estimate range to a vocalizing animal using Target Motion Analysis (Leaper et al. 2000), as long as the animal remains at approximately the same 2D location whilst it is being tracked (i.e. it is slow moving compared to the towed array). Since towed arrays can calculate a range to an animal they are suitable for Distance Sampling; thus when used as part of well-designed line transect survey, and usually combined with a concurrent visual study, towed hydrophone arrays can be used to calculate the absolute density of cetaceans in a surveyed area (Thompson et al. 1994, Hammond et al. 2006, Leaper and Gordon 2012, Marques et al. 2013a).

More complex tetrahedral towed arrays can determine the depth and absolute location of animals using similar target motion techniques (Zimmer 2013). These are slowly becoming more common; however their use remains rare in the literature. Beamforming arrays are often used in a towed configuration and can be particularly useful for behavioral studies (e.g. Miller and Tyack (1998)).

3.4.5.2 *Drifting Systems*

Hydrophone arrays deployed from a drifting vessel (i.e. not moving through the water) are widely used in studies where a wide aperture array (generally one order of magnitude less than maximum detection range e.g. a 1 km detection range would mean a 100 m aperture) is required to calculate more accurate information on animal movements. There are numerous examples of drifting systems which can vary hugely in size depending on the study species vocalization type. For example vertically orientated linear arrays on the order of hundreds of meters in length have been used to localize the clicks (~10-20 kHz) of sperm whales (*Physeter macrocephalus*) (Møhl et al. 2000, Wahlberg 2002, Heerfordt et al. 2007) whilst similar but much smaller systems a few tens of meters or less have been used to localize the very high frequency (~130 kHz) and therefore short range, vocalizations of harbor porpoises (Villadsgaard et al. 2007, Kyhn et al. 2013, Macaulay et al. 2017). Another potential configuration is to use autonomous drifting units rather than vessel-based systems. For example, (Miller and Dawson 2009) deployed four free floating autonomous drifting hydrophones and boat-based stereo array to estimate dive profiles tracks of sperm whales in 3D. Matsumoto et al. (2013) and Griffiths and Barlow (2015) have also been developing a low cost drifting buoy system to survey beaked whales. Multiple floating DIFAR buoys (radio-linked hydrophones capable of determining a bearing) have been used to localize blue whales in the arctic (Miller 2012, Miller et al. 2014).

3.4.5.3 *Bottom-mounted systems*

Seabed-mounted systems have the great advantage over drifting PAM systems of being able to monitor the same area over long time periods, unaffected by weather and without posing a potential collision hazard in busy shipping areas. They are therefore probably the only practical option for DECAF.

However, primarily due to cost of development, deployment and maintenance, seabed-mounted systems capable of localization and dedicated to marine mammal science are very rare. The majority of cetacean studies using seabed-mounted systems are either military installations such as SOSUS (Stafford et al 1998) and AUTEC (Ward et al. 2008) or equipment used for other primary purposes, e.g. detecting neutrinos (Riccobene, et al 2009), seismic monitoring (Soule, et al 2013), hydrophones used to monitor the nuclear test ban treaty (Miksis-Olds et al. 2014) or piggybacking on tidal turbines (Malinka, et al, in review). However, there are a few dedicated systems that have been deployed to study marine mammals. Cabled systems have been used to study 3D positions of beluga whales (*Delphinapterus leucas*) (Roy et al. 2010). A variety of other bespoke cabled systems are discussed in Norris et al. (2010).

An innovative example of a dedicated seabed-mounted system is the use of multiple autonomous HARP units (Wiggins et al. 2012) to localize the relatively high frequency vocalizations of beaked whales (Gassmann et al. 2015). The HARP units, which are configured for localizing, are autonomous recorders with four closely-spaced hydrophones (~30 cm) in a tetrahedral configuration. They have the great advantage of not requiring any cabled connection between different units and can be deployed and recovered from different areas. However, to resolve the exact position on the seabed and maintain some synchronization an external sound source at a known location or pingers located on the units themselves are required. The lack of ultra-accurate synchronization and the fact that hydrophones are clustered on individual autonomous units, rather than spread evenly, also means that localization errors are generally greater however, despite this, they can resolve realistic tracks of beaked whales, potentially one of the worst candidate species for localization (this study is discussed in detail in section 3.3.11). Discrete autonomous units for localization have also been applied to much lower frequency signals over greater ranges. The Cornell pop-up network has been used successfully to localize right whale calls and has been instrumental in many studies (e.g. Laurinolli et al. 2003) and Directional Autonomous Seafloor Acoustic Recorder (DASARs) have been used to localize bowhead whale calls (Blackwell et al. 2015).

Given the wide range of possible species and locations, a bespoke solution will be necessary, therefore attempting to specify one here will be misleading; therefore we have outlined some possible solutions of some of the issues raised in section 3.3 that could be practically implemented to reduce errors. For density estimation, where detection range is one the primary measures of interest, a practical solution might be to construct vertical arrays, with all hydrophones within the array cabled to a common digitizing unit mounted somewhere between the hydrophones and the bottom anchor / release mechanism. This solves all problems associated with synchronizing clocks between multiple units at different bottom locations. It also means that, when working in deep water, moorings can be designed so that all arrays are at the same water depth, which will reduce difference in horizontal detection range and other performance metrics between units (e.g. the fraction of calls localized). So long as each mooring is designed with sufficient buoyancy directly above the array and weight beneath it so that the array itself remains vertical, overall sideways drift will have little bearing on density estimation. Problems of array spacing discussed in section 8.1.8 would still apply, so arrays would have to be designed with a specific subset of species in mind. Altitude of the array would also be a key consideration. For any species group, best results will be obtained when the array is closest to the depth the animals typically vocalize at, so for beaked whales, the array should be at a depth of several 100s of meters. For small cetaceans, a deep array might be beyond their detection range, so the array would need to be at a much shallower depth. Consideration could be given to mounting two or more arrays on a single mooring in order to cover a wide range of species groups. If the devices were briefly tested with pingers (upon deployment and briefly) so that spacing between vertical array units and inter unit timing could be determined, then species with large detection ranged could also be localized by using multiple vertical arrays. Fitting pingers for longer term testing might adversely affect cetaceans (and thus bias their density estimates)(e.g. Watkins and Schevill (1975)).

3.4.6 PAM surveys for Density Estimation in practice

Moretti et al. (2009) outlines a simple but comprehensive list of PAM survey elements to be implemented in order to develop a fit-for-purpose PAM fixed array installation. These include site selection, suitable design/selection of PAM systems (comprising in-water hardware; array design, shore or top-side electronics (including signal processing/tracking/display hardware and software), the installation and maintenance of an array, and processing and analysis of the data generated. As noted in section 3.4.5 the majority of bottom-mounted arrays suitable for density estimation are US Navy cabled hydrophone systems or cabled ocean observatories (this is also covered in detail in Norris et al. (2010)), where the driver for these systems are operational needs; any environmental information gleaned from them is secondary. Moretti et al. (2009) attempt to estimate the costs associated with different PAM array survey approaches in different site conditions (see Table 5 from Moretti et al. (2009) below). We have not attempted to repeat this exercise as the tables below likely to be broadly indicative of the costs associated with such a long-term fixed installation (i.e. either utilizing cabled hydrophones or bespoke portable systems).

Table 5 – Amended Table 2 from Moretti et al. (2009) exploring the elements and associated costs of different PAM arrays. It is unclear from that report, exactly which costs are in $k (USD) increments and which are $USD values and a number of terms are presented without further explanation. Please note final costs are definitely $k. See Moretti, et al. (2009) for descriptions of systems.

	Fixed/Mux Array	Single Cabled Hydrophone Array	Portable Mux Array	Recording Buoy	Moored Surface Buoy
Water Depth (feet)	100 - 20,000	101 - 20,000	< 1000	100 - 20000	< 1000
Area of Coverage (nm^2)	50 - 1,000	50 - 1,000	< 100	50-500	< 50
Maximum Number of Sensors	100's	100's	50	~50	20
Track Accuracy	<10m	<10m	<10m	<15m	<20m
Usable deployment period	20 years	20 years	2-3 years	1 year max	5 years*
#Sensors for Analysis Case	49	49	49	49	49
Node Cost ($K)	£150	£40	£75	£40	£100

Installation (days) Best / Worst Case	12/24	20/30	12/24	5/10	10/20
Survey (days) Best / Worst Case	7/14	7/15	7/14	7/14	0.5/2
Retrieval (days) Best / Worst Case	N/A	N/A	8/13	2/5	3/8
Installation Costs -Best Case ($k)	$500	$780	$500	$120	$220
Retrieval Costs	--	--	$500	$300	--
Refurb Costs (per install)	--	--	$50	$100	--
Yearly Maintenance	--	--	$50	--	$300
Operating Costs (per day)	--	--	--	--	--
ship/personnel/hardware	--	--	--	--	--
Cable Cost ($k)	$3,360	$11,566	$3,780	--	--
Node Costs	$7,350	$1,960	$3,675	$1,960	$4,900
NRE	$300	$300	$300	$50	$100
Acquisition Cost ($K) w/ Installation	$11,510	$14,606	$8,255	$2,130	$5,220
M&O per year	$50	$50	$50	$200	$250
20 year w/ M&O costs ($K)	$12,510	$15,606	$14,255	$6,130	$7,720
Systems Engineering	$1,200	$1,200	$1,200	$500	$1,200
Program Management	$500	$500	$500	$500	$500
Program Cost ($K)	**$14,210**	**$17,306**	**$15,955**	**$7,130**	**$9,420**

*assumes yearly required maintenance with refurb on 5-year basis
** assumes portable mux array recovery every 2 years
*** for tests: 2 tests x ship ($15k/day x 10days) + staff (5 * 2.4K/day*10)

A clear conclusion from the assessment made by Moretti et al. (2009) is that the installation and maintenance of *any* of the considered fixed PAM array designs will be a costly endeavor. Depending on the exact specifications, capabilities and design of the array, the cost for a 20 year PAM array survey would range between $7.1 – $17.3m (USD). It is important to note that this does not cover the analysis costs (e.g. DCL analyses and density estimation efforts). The majority of such installations in operation now are operated by military or governments and their primary function is not as a marine mammal monitoring system – therefore there are a number of roles such arrays could fulfil. It is unclear if PAM arrays for this purpose would also be utilized to address operational issues for E&P installations (e.g. leak monitoring or 4D surveys). Most PAM systems are not designed to last for extended periods and therefore whether it is portable PAM systems or cabled hydrophones that are installed it is likely the systems will require maintenance to maintain optimum performance (for DCL). It is likely (particularly in shallow, temperate regions) that units are prone to biofouling, which has the potential to affect the detection capabilities of acoustic systems (Heupel et al. 2008). Maintenance of a fixed cable system would likely be more complex and expensive than servicing autonomous recorders (which in theory could be accessed using acoustic releases) – though the user of submersible vehicles, such as Remotely Operated Vehicles in industry is so prevalent that this might represent a cost effective approach for maintenance. It is unclear how fixed installations would be decommissioned at the end of lifetime.

Despite this cost, cabled systems do offer a number of benefits in that they allow for an adaptive monitoring framework and for changes to be made to the sampling regime and/or (and in the case of portable PAM systems) changes to the array layout to best fit the site. Furthermore, by having recent access to the data, it will be possible to refine and update DCL efforts to optimize their performance over time. The main advantages and disadvantages of both classes of system are presented below (Table 6).

Table 6 - Summary of advantages and disadvantages of fixed cabled arrays and portable autonomous units.

System Type	Advantages	Disadvantages
Cabled fixed array	• Synchronized hydrophones reduce localization error • Data is obtained in real time. • Long lifespan and consistent dataset. Potential cost-saving for decadal monitoring	• Very expensive initial capital outlay • Still requires maintenance • Costly decommissioning of cables and infrastructure and the end of the project.
Portable autonomous units	• Initial capital pay an order of magnitude less than that of a cabled system • Flexible. Can be deployed in different locations and rearranged • The technology will improve and become more cost effective over the next 40 years • It is possible that deployments will be cheap as the oil industry is expert at this type of thing. • Much more useful tool as a whole for the research community, especially if software is developed and methods to build these devices for little cost is shared.	• Few off the shelf systems available • Synchronization requires additional analysis and likely not as accurate as a cabled systems • Position of hydrophones has to be determined on every deployment • Hydrophones likely clustered so for some species. These will not be as accurate as widely dispersed arrays.

3.4.6.1 *Personnel for analysis*

An important but often overlooked element of conducting PAM surveys is the analysis personnel required. This is a particularly crucial element for PAM surveys where density estimation may be used to inform project planning and environment risk assessments throughout a project life cycle. This is because such efforts differ from exploratory research where timing for delivery of results is typically less critical. While all systems incorporate a degree of automation, a substantial degree of user intervention and interpretation is required and it is recognized by many that systems that combine automated detectors in conjunction with experienced operators will probably always underpin the most effective systems (Sheldon-Robert et al., 2008). In this respect, the human analysts are a crucial part of any PAM system and effectiveness and performance will be substantially influenced by the experience and skill of PAM analysts. An approximate estimate of the cost of analysts over a lifetime of duration would likely be $100,000 per year (based on a post-doc level analyst conducting PAM and DECAF analysis) – although this would potentially need to be up-scaled according to the number and type of PAM systems deployed and the number of species of interest.

4 Using PAM data for estimating density

Box F: Why is this topic challenging?

The method best suited for a given survey will vary depending on the target species, and the logistics and resources of a proposed survey. Thorough survey planning is essential to assess the options.

In particular, the object of interest (to be acoustically detected and counted) needs to be defined – this may be individual vocalizations, animals or groups.

Another main challenge of density estimation is estimating all the required multipliers for the chosen density estimation method.

Multipliers to be estimated include detection probability (and the associated effective survey area), cue production rate and group size. Data, auxiliary to the main survey, may be required to estimate multipliers. Auxiliary data should be collected at the same time and place as the main survey, to ensure that the estimated parameters are accurate.

4.1 Detailed overview of DECAF and input components

In this section, we introduce methods for analyzing acoustic data from fixed hydrophone systems. These DECAF methods are described in more detail in Marques et al. (2013a). As mentioned in previous sections of this report, the two main methods for estimating animal density from passive acoustic data are distance sampling and SECR. Both rely on different assumptions and require different data to be collected. We briefly review the concepts of density estimation here.

If we were able to count all animals *N* in the study area of size *A*, we could simply obtain the density of animals as

$$D_{animals} = \frac{N}{A}. \tag{1}$$

Note that we use italic font for naming parameters. For most wildlife surveys it is impossible to survey the whole area of interest, often because the study area is too large. Hence, we only sample a randomly located fraction *a* of the area *A*, which we refer to as the covered area, in which we detect *n* animals from a population of size *N*. Assuming that we detected all animals the density in the covered area is now:

$$D_{animals} = \frac{n}{a}. \tag{2}$$

We further acknowledge that we do not detect all individuals within the search area with certainty. Hence, we estimate the average detection probability *p* within the search area *a* and an estimator of density becomes

$$\hat{D}_{animals} = \frac{n}{a\hat{p}} \tag{3}$$

Note the use of a hat to show that the metric of interest is being estimated.

4.1.1 Animal density estimation using point transect sampling and cue counting

Point transect sampling is a form of distance sampling relevant to DECAF methods. In typical visual surveys with point transects, an observer remains at a point for a predetermined period of time and records all the detected animals and their radial distances to the point. In this case the search area is defined as $a = k\pi w^2$ where *k* is the number of points and *w* is a truncation distance, i.e. the furthest distance at which animals were detected. A statistical model, a detection function, is fitted to the observed distances, which describes the decrease in detection probabilities with increasing distance from the point and allows us to estimate *p,* the average detection probabilities in the search area. It is important to note that distance sampling requires not the slant distance but the horizontal distance between a vocalizing animal and a hydrophone array, i.e. as if animals were projected into the ocean floor. Therefore, depth of the animal might need to be estimated to estimate the horizontal or projected distance between the animal and the instruments.

Typically, in the context of a PAM density estimation survey designed for distance sampling, each node represents a point transect and we observe cues produced by the animals of interest, e.g. echolocation clicks or vocalizations, rather than animals themselves. Hence, from the acoustic recording of length T (i.e. PAM survey effort) at each of k nodes (i.e. PAM systems), we observe a total number of cues n_c for a species of interest. As for conventional distance sampling methods, distances from the node to these cues are required to estimate the decay in detection probabilities with increasing distance from the point. These distances can be obtained using fixed hydrophone systems with localization capacity, however, obtaining accurate and precise distances can be extremely complex (see Sections 3.3 for details). We further acknowledge that a known proportion of the ‘cue’ detections, n_c, might be false positives, denoted with f (see Section 8.1.8.2), such that (*1-f*) is the proportion of correctly detected cues. Rather than each cue indicating a single animal as with surveys that count individuals directly, animals produces cues at an average ‘cue production rate’, *c* (e.g. number of cues per second). Now the estimator for density can account for *f* and *c* (Marques et al. 2009, Marques et al. 2013a):

$$D_{animals} = \frac{n_c(1-f)}{k\pi w^2 pTc} \tag{4}$$

Note that $\frac{n_c(1-f)}{Tc}$ is equivalent to *n* from equation 3. To obtain an estimate of cue production rate generally requires the collection of additional auxiliary data. It cannot be estimated using only the acoustic data from the nodes as not all cues produced by animals within the search area are detected with certainty, nor can we include silent animals in any straightforward way. Note that we use different parameter names compared to Marques et al. (2009, 2013). These authors use *r* to denote cue rate while we use *c*.

The effective survey area can be derived from the detection probability and is equivalent to $k\pi w^2 p$ in equation 4. The effective survey is defined as the area where there are, on average, as many animals undetected within the area, as those detected outside the area (within the bounds of the original surveyed region).

As mentioned above, detection probabilities generally decrease with increasing distance from the sensor – a feature usually captured by fitting a detection function to the observed distances and estimating the average detection probability, *p*. To reliably estimate cue production rate, we would need to be able to detect *all* cues that were made by an individual regardless of distance from the sensor. Hence, this is often done by attaching DTAGs (Johnson and Tyack 2003) to a selected number of individuals. The acoustic data from a DTAG allow counting the number of cues produced by the individual over the time period that the tag was attached. It is recommended that DTAG data are collected from multiple animals from the same study area throughout the same time period as the acoustic recordings from the sensors (see Box H in section 5.2.3). This is because there are likely differences in cue production rate due to various reasons, including but not limited to differences between individuals, differences in the behavior of the animals due seasonality or to presence of other species (predatory or prey). Cue production rates can also be estimated from combined visual and acoustic focal follow data (e.g. (Marques et al. 2011)).

Conventional distance sampling methods generally rely on the assumption that all animals (or cues in this case) on the point are detected with certainty. However, some cues produced by cetaceans are highly directional, such as the echolocation clicks of harbor porpoises (see Section 3.3.4 for an estimated beam profile of harbor porpoise clicks) or beaked whales. Furthermore, it is possible that an animal may have a very small horizontal distance from a hydrophone but the vertical distance between a bottom-mounted hydrophone and an animal vocalizing at the surface may be large. As a result, the probability that a sensor detects a given cue not only depends on horizontal distance from the sensor but also on the depth and orientation of the animal producing the cue in relation to the sensor. Hence, not all cues produced directly over the sensor are necessarily detected with certainty and we need to estimate an additional parameter defining this perception bias: parameter a_{perc} represents the proportion of cues that are missed directly over the sensor. We note that generally, in a distance sampling context, missing animals at the point or on the line is attributed to availability bias, i.e. animals that are not available to be detected. In our context, however, the equivalent of this would be if animals were not producing cues when directly on the point. However, this availability bias is already accounted for via the cue production rate, which takes into account the proportion of time that the animal does not produce any cues. In the context of PAM systems, missed detections at the point or on the line are often the result of perception bias, for example, due to the directionality of an animal's beam profile, which is captured through the parameter described by a_{perc}, referring to perception bias at the point. The density estimator incorporates bias by including the a_{perc} term as follows:

$$D_{animals} = \frac{n_c(1-f)}{k\pi w^2 pTca_{perc}} \tag{5}$$

4.1.2 Animal density estimation using spatially explicit capture recapture methods

Spatially explicit capture recapture (SECR) methods are applicable in the context of PAM and DECAF methods when the same object of interest can be detected and associated across multiple nodes but is not necessarily localized (e.g., Borchers and Efford 2008). From that perspective, it is a method that can cope with all detections of interest without the need for explicitly localizing or ranging to any of them. As with distance sampling methods, the concept of a detection function takes a central role. SECR was developed as a natural extension of conventional capture recapture (CR) methods, aiming to solve two of the major problems with CR: (1) unmodelled heterogeneity in detection/capture probabilities and (2) the inability to estimate density due to an undefined sampling area. Within SECR (1) part of the unmodelled heterogeneity in detection probabilities that plagues standard capture recapture approaches is accounted for the spatial component, and (2) density itself becomes a parameter in the SECR likelihood, so we have a direct estimate of density and a description of how it changes over space (typically assumed uniform, which is computationally easier to implement, but not a requirement of the methods). We can therefore estimate the population abundance over any given area of interest by integrating under an estimated density surface. We note upfront explicitly that therefore, under SECR, rather than a derived parameter of the detection probability, as in distance sampling methods, density can be estimated directly as one of the parameters of interest.

SECR was developed in the context of effective captures of animals, hence requiring multiple capture sessions, and basing inferences in the capture histories obtained across these. But it soon became apparent that in the case of acoustic data, SECR might be used to estimate density of sounds based on a single capture occasion. Under this setting the hydrophones are called "proximity" detectors, to convey the notion that the animal can be captured when in the vicinity of the detector. The "recaptures" of a sound in multiple sensors are analogous to the capture of animals in different camera traps, and the resulting capture histories can be used to estimate a detection function as a function of distance from any instrument in the SECR array (this is estimated as part of the SECR method). From this information on detectability we can then estimate the effective surveying area of each sensor. From this the effective survey area of a field of sensors follows naturally. Due to being based on CR methodology, SECR cannot be formulated using the estimator in Eqn 5, though many of the underlying concepts regarding detection probability and effective survey area are the same. There is also still a need for multipliers such as cue production rate or group size, depending on the object of interest.

Optimal sensor placement for a SECR survey is an object of current research that has not yet received much attention, perhaps Solmann et al 2012, in a non-acoustic setting, being the exception. It is desirable that there is a good range in the number of sensors that objects of interest are detected on. If sensors are so close together that most objects of interest are detected on a large number of sensors, the design is inefficient, as we could have spaced the sensors better to obtain better spatial coverage. On the contrary, if sensors are so far apart that no recaptures are recorded, with calls detected on single sensors, there is little information about detectability in the data. A middle ground is therefore desired, leading to a large proportion of cues detected at two or more nodes, providing strong information on detectability, and hence allowing the estimation of both detectability and density from the recorded capture histories.

4.1.3 Comparing distance sampling and SECR: method assumptions, advantages and disadvantages

We are exploring two different methods to estimate density, each with its pros and cons. In this section, the two methods are compared, using the assumptions that must be met to reliably use each method as a basis for discussion.

Each density estimation method has associated assumptions of varying severity. Violation of key assumptions can lead to bias in estimated detection probabilities and, ultimately, abundance or density estimates. There are four main assumptions in distance sampling (Buckland et al. 2015), in the “standard” situation where the object of detection is an individual animal. (We mention cue-based distance sampling below.) These are that: (1) transect lines or points have been placed randomly with respect to the distribution of the study species (this is important not only for the design-based abundance inference, but also for the detection probability estimation), (2) any animal on the transect line (i.e., at zero distance), or at the center of the monitoring point, is detected with certainty, (3) detected animals are detected at their initial location, i.e., there is no animal movement and (4) measured distances are accurate.

Distance sampling survey design has been well documented (Buckland et al. 2001b, 2004, Buckland et al. 2015) so meeting assumption (1) should be possible at the survey design stage. Auxiliary analyses using acoustic tag data and/or propagation modelling may be required to estimate the probability of detecting an animal at zero horizontal range from the hydrophones (assumption 2) and animal movement can be dealt with by using a cue-based approach (because cues are instantaneous and so do not move by definition) or by estimating density over time intervals that are short enough to assume that individual animals or groups are stationary. This is known as a “snapshot” approach (Buckland 2006). The incorporation of animal movement models is an area of current research focus (Glennie et al. 2015). Measurement error, if quantifiable, can be accounted for (e.g. Marques (2004)).

The main assumption of SECR is that captures can be accurately associated (i.e., re-identified) across capture occasions. When using SECR with passive acoustic data, it is also currently assumed that (1) detection probability is only a function of animal-sensor distance (which implies, for example, that all calls are emitted at the same source level and that there is no effect of signal directionality on detection probability), (2) the underlying animal density surface is constant in the effective survey area. However, each of these assumptions can be addressed through extensions to the existing SECR methods and software (Borchers et al. 2015, Stevenson et al. 2015).

Distance sampling is a well-established method with many years of research into all aspects of the approach. By contrast, SECR is a relatively new method and optimal survey design is still being investigated (e.g. Kidney et al. (2016)). The effect of mis-matching detections across sensors and creating incorrect capture histories has not been investigated for SECR (though this has been investigated in mark recapture e.g. Lukacs and Burnham (2005)); method development is required to address some of the key assumptions relating to acoustic data, which are currently unrealistic for many species of interest. However, the fact that SECR does not rely on localizations is a large advantage of the method over distance sampling, which is why it has been considered in this report, despite there being a need for key method developments. Further, it has already been shown that adding information relevant to animal location, such as received level, bearing or arrival time, improves the precision of SECR estimates (Borchers et al. 2015, Stevenson et al. 2015, Kidney et al. 2016, Stevenson 2016).

It is important to note, however, that both methods will rely on multipliers (such as cue rate, average group size, probability of detection at zero distance, etc.) to obtain estimates of absolute abundance or density. It is key that (a) multipliers are estimated as precisely and accurately as possible, and also (b) any spatio-temporal variability in the multipliers should be quantified and the density estimator updated with new parameter values over space and time, as needed.

4.1.4 Challenges for the simulation tool

The goal of the simulation tool is to assess the power to detect a trend in animal densities over a number of survey years using data collected with fixed PAM devices, given specific scenarios defined as a set of assumptions. When using the simulation tool, the user specifies and, hence, knows, the true animal densities and the true decline of densities over the survey period. The tool can be configured to simulate a range of scenarios of what might be observed with a PAM system and uses these data in a regression model to test for a significant trend in animal densities over time. A scenario is comprised of three elements and is defined by:

- the study area size and shape,
- the survey design properties,
- the characteristics of the species of interest.

Each of these elements needs to be defined by the user by parameterizing input variables while the data generating and observing mechanisms that replicate scenarios are built into the simulation tool and described in detail below (Section 4.3). Making these scenarios realistic is challenging due to the reasons we summarize in this section. Here we describe the different components of the simulation tool and the potential sources of variability encountered in nature if applicable.

Defining the study region is done by specifying four coordinates that delineate the outermost points of a rectangular study region. Challenges here include defining the study area such that it is the appropriate size for capturing the effect of an impact. If the selected study region is too large, then the effect might be difficult to detect when taking averages over a large area with natural variability; if the study region specified is too small, then the dataset might be too small to allow reliable estimation of animal densities. Naturally this could be extended for non-rectangular survey areas, but given the conceptual nature of the exercise, that was not implemented at this.

The survey design is defined by specifying the spacing between nodes that can simply detect (n-nodes, because they generate *n* in the density estimator) (in km) and between nodes that can localize (p-nodes, because the data are used to estimate detection probability). Note that p-nodes also have the capability to collect information on number of detections, so are also n-nodes by default. In addition, all nodes in SECR are used to estimate detection probability so there is no need to distinguish between n- and p-nodes in SECR. Determining node spacing will generate the total number of nodes that are placed into the survey area and how the p-nodes are spaced within the region. Heterogeneity in the detection properties most likely exists between individual nodes. These heterogeneities can be caused by different sensitivities of the individual nodes themselves and/or varying amounts of background noise surrounding the individual nodes against which the cues are detected, or by varying propagation conditions around the nodes. In addition, heterogeneities may be caused by variations in animal behavior across the study areas (e.g. different dive ecology or vocalization behavior). Background noise might be strongest in the immediate vicinity of a given project or human activity and during the time of construction but may also have natural causes. This heterogeneity can lead to strong biases in the estimated animal densities.

As stated above, p-nodes are used to obtain distances to the observed cues, in addition to the counts of observed cues. From n-nodes we only obtain counts of detected cues. Observed distances from the p-node nodes are used to estimate the average detection probabilities *p* from equation (5) which is then applied to all nodes. From a statistical point of view, in an ideal scenario all nodes would be p-nodes and all p-nodes would be capable of observing distances with high precision and accuracy. While the former is generally too expensive and logistically too challenging, the latter is simply unrealistic at the current stage given the practical limitations of localization marine mammal vocalizations (Section 3.1.1). However, we still recommend including a relatively high number of p-nodes, eight or more, in any given survey design. In locations with expected large heterogeneity between nodes, using a large number of p-nodes may protect against the potential bias in estimating *p* while still capturing a good estimate of the variability in fitting a detection function, and hence estimating *p*. Bias in estimating *p* may give rise to bias in estimating animal densities and hence, in estimating the trend in animal density, the true parameter of interest.

Distances to the cues are generally observed from the p-nodes with poor accuracy and precision (see Sections 3.3.10). Here we distinguish between random and systematic error in the distances measured by the p-nodes. Systematic distance bias is error that occurs when the estimated distances to cues are always overestimated (or always underestimated) as can happen with factors related to, but not limited to, temperature and salinity affecting sound speed profiles, sound refracting properties associated with a particular location, background and/or sensor-specific errors (these should be collected on site). Systematic bias is implemented in the simulation by adding or subtracting a distance to each cue that is a percentage of the 'true' localization distances. The magnitude of the systematic bias can be specified by the user as a percentage multiplier on the cue distance estimate. Random error occurs when the distances are measured with low precision and can be over-or underestimated. This source of variability was also added to the simulation, where the user can then specify the uncertainty as a percentage, where the percentage is a function of the aperture size (more details are given in Section 3.3.7). Systematic and random error will introduce variability and bias to the system which will decreases accuracy in estimating animal densities and hence, decrease power in estimating trends in densities.

Defining the species of interest is challenging as little is known about most cetacean species and the user of the tool needs to define a relatively large number of parameters. To populate these parameters, the user needs to research published literature. The following description of the individual components needed for the species of interest aims to assist in interpreting the parameters correctly.

As described above in this section, density of animals refers to the expected number of animals within a defined study area. It generally represents an average over a certain time period, namely the period over which data were collected, or over smaller defined time periods within the overall monitoring period, if changes over time are to be investigated. Marques et al. (2009), for example, estimated that the density of beaked whales (*Mesopolodon densirostris)* was on average 25.3 animals per 1000 km^2 over the recording period of six days with a coefficient of variation (CV) of 20.8%. Hence, this inference is limited to the time period of the recording as animal densities fluctuate naturally within the study area on varying time scales. The CV of 20.8% represents an estimate of the uncertainty in average animal density over the 6 day survey period – i.e., it includes only the observation error.

There are typically two sources of “error” (i.e., uncertainty) in estimates of trend in animal density (see, e.g., Buckland et al. (2004), Ahrestani et al. (2013)). The first is observation error, which is uncertainty in the trend estimates caused by sampling (e.g., how precisely the detection probability could be estimated will contribute to observation error). The second is process error, which is random variation in the underlying animal density over time that is not captured by the trend model.

To be realistic, a simulation study of the power to detect trends should include both observation and process error. Here, estimates of observation error for a particular scenario are generated by simulating a survey under particular conditions and recording the resulting uncertainty in the simulated detections. Process error can be included by making an assumption about year-to-year variability in animal density in addition to a smooth trend (called a “random decline” scenario later). Realistic values for this parameter were not available for the Blainville’s beaked whale study on which the simulation was based, and so this source of uncertainty was set to zero. Estimating it requires a time series of data for which a trend line has been fit and so overall uncertainty is available, and also for which observation error has been estimated – the result when the observation error has been subtracted from the total error is the process error (see, e.g. Newman and Lindley (2006)).

Animal densities also vary on different spatial scales; some of these spatial scales are likely smaller than the study area of interest for this project, for example due to density hot spots or gradients arising from geographical features such as the continental slope or underwater canyons. While acknowledging these natural fluctuations of animal densities in space and time, we are interested in assessing whether the average density of animals within a defined study area declined over the course of the study period. Such a decline could manifest itself as an overall decline in animal densities throughout the study area or with varying degrees within the study area, e.g. with the strongest decline in the area of highest disturbance from a construction site. We set up a simulation that considers variability across all the input parameters, assessing the effect of changing the mean and coefficient of variation (C.V.) on our ability to detect a significant decline in density given that the decline exists, i.e., our statistical power.

The cue production rate, c is the average rate at which the individual animals of the species of interest produce cues and is often estimated using DTAG data (see Section 3.1). These are acoustic tags that are attached directly to animals, generally via a suction cup mechanism, and remain attached for some time while continuously recording several parameters as well as sound. The tag data allows counting all cues produced by the animal over the period that the tag was attached to it, though additional signal processing analyses may be required to distinguish between cues from the tagged animal and conspecifics (e.g. Goldbogen et al. (2014)). . The rate at which animals produce cues in real life varies between individual animals and depends on a range of factors including, for example, the presence of other species, the behavior of the animals, temporal fluctuations (e.g. seasonal or annual).

Sound tag studies are also useful in estimating perception bias, a_{perc} when the acoustic record from a tagged animal can be compared to the acoustic records from a fixed PAM array, e.g., when tagged animals are within the fixed PAM array study area, or are within range of a drifting receiver with similar capabilities. The proportion of cues missed at or close to the instruments then allows the perception bias parameter, a_{perc} to be estimated. Hence, it is important that tag studies are conducted at the same time that the nodes are recording cues, so that both perception bias and cue production rate are estimated accurately at the time of the monitoring effort. Similar to estimating cue rates, perception bias may vary depending on a range of factors including, for example, the behavior of the animals, the presence of other species, temporal fluctuations (e.g. seasonal). We recommend that enough DTAG data are collected to allow seasonal and annual variation in cue production rate to be investigated.

For estimating the false positive rate in a real life situation, sections of the acoustic recordings are proofed for false positives assuming a gold-standard. As this is generally done by a human observer (see Sections 3.2), these sections often only represent a small fraction of the total length of acoustic recordings. False positive rates are generally not constant throughout a recording but may depend on the sensitivity and the settings of the detectors, the presence of other species and/or the level of background noise during the recording. Hence, in an attempt to reduce bias, the total effort of proofing is generally divided into sections systematically spaced throughout the recordings.

The utility of relative animal density, or indexes of abundance, compared to absolute density or abundance, is often the subject of heated debate (e.g. Anderson 2001). Relative density estimates may be the result of not being able to estimate all parameters in equation (5), e.g. due to the lack of data. . Often ignored are those which pertain to adjusting observed number of cues for imperfect detection. These components are the parameters p and a_{perc} from equation (5) above. Alternatively, cue production rate may be omitted, leaving call density as a measure of relative density. Not including these parameters in estimating animal densities is a source of potentially large bias in the resulting relative density measures. If any of the parameters are not constant through time or space (which is quite feasible), then the observed spatiotemporal trends in the relative density estimates will not reflect the unobserved underlying spatiotemporal trend in absolute animal density, but a confounded effect of the two phenomena. Therefore, incorrect conclusions might be drawn from a power analysis using relative density measures.

4.2 Available datasets to inform density estimation approaches

The concept that acoustic data from fixed instruments can be used to derive estimates of marine mammal density has been promoted in the literature for many years (e.g. McDonald and Fox (1999), Mellinger and Barlow (2003), Mellinger et al. (2007)). In more recent years, the DECAF project demonstrated several case studies where fixed acoustic data were used to estimate cetacean density. All of the datasets used in DECAF were from two US Navy ranges – the Atlantic Undersea Test and Evaluation Center (AUTEC) and the Pacific Missile Range Facility (PMRF). Blainville's beaked whale density was estimated at AUTEC using two different methods (Marques et al. 2009, Kusel et al. 2011). Sperm whale density was also estimated using AUTEC data (Ward et al., 2011). PMRF data were used to estimate minke whale density using spatially explicit capture recapture (Marques et al. 2012, Martin et al. 2013).

Marine mammal monitoring data from naval ranges can be considered as opportunistic, given that navy ranges have other primary purposes. Other opportunistic datasets have been used to implement DECAF-based methods. The Comprehensive Nuclear Test Ban Treaty Organization has deployed hydrophone arrays as part of its International Monitoring System (www.ctbto.org) and these have been used to apply DECAF-based methods , specifically those described in Kusel et al. (2011) to blue whale calls (Harris, 2012). Ocean Bottom Seismometers have also been used to demonstrate that distance sampling can be applied to fin whale recordings (Harris et al., 2013).

More recently, DECAF-based methods have been applied to dedicated marine mammal monitoring efforts. Marques et al. (2011b) estimated North Pacific right whale density using acoustic distance sampling. The method described in Kusel et al. (2011), which uses a simulation incorporating sound propagation models to estimate detection probability, has been extended and applied to recordings of Cuvier's and Gervais' beaked whale (Hildebrand et al. 2016), delphinid species (Frasier et al. 2016) and humpback whales (Helbe et al. 2013, Helble et al. 2013). Harbor porpoise density has also been estimated using an approach based on Marques et al. (2009), though a combined visual-acoustic trial was used to estimate the detection function, rather than data from acoustic tags (Kyhn et al. 2012).

4.3 Description of simulation tool

The objective of the simulation tool is to allow the user to assess the power to detect a change in animal densities for a given scenario. A scenario is defined by a range of parameters pertaining to the survey design using a fixed PAM system, the study area and the species of interest. By replicating realistic survey conditions, it can be assessed how likely it is that fixed PAM surveys over time will be able to detect trends in the population of interest. Various scenarios can be tested using the simulation tool.

As an example, the tool can be used to assess the effect of reducing the number of sensors while other parameters of the scenario remain unchanged. This would allow the determination of the minimum number of sensors needed to retrieve enough power for detecting potential effects.

Uncertainty in the individual parameter estimates necessary for estimating animal density from the detected cues arise from several reasons as described in this report which, in turn, can lead to inability to detect a trend in animal densities. If the parameters that contribute the largest amount of uncertainty can be identified, these aspects of the survey can be targeted with the aim of reducing the variance of these components. One of the returns of the tool is the contributions of the individual components to the overall variance of the density estimates, i.e., an error budget, and hence, allows the user to investigate the effects of uncertainty of the individual components (see Section 5.1). This can be used to reduce the uncertainty of the overall density estimates, which, in turn, will likely increase the power to detect trends in the estimated animal density.

There are four main steps involved in the simulation tool (Figure 9) and four different levels at which these steps are performed. The levels pertain to the simulation, the iteration, survey year and season where each simulation consists of 1000 iterations and a number of years and seasons over which the surveys were conducted, the latter two specified by the user.

The first step consists of creating a data set that attempts to replicate a certain scenario (Figure 9). This step is divided into generating the study area and survey design as well as the species of interest which are all used in combination to generate cue data across the study region from an assumed population process model, or scenario. The second step consists of estimating animal abundance using the “observed” PAM data. The first and second steps are conducted for each season and each survey year for a given iteration of the simulation. The third step consists of fitting a generalized linear model (GLM) to the estimated animal densities from all the seasons and years and determine if a change can be detected. For a given simulation, steps 1–3 are repeated during 1000 iterations (although some of the elements such as the study area and parameters for the survey design remain fixed). From step 3, the percentage of 1000 iterations with p-values <0.05, denoting a significant change parameter in the fitted GLM, represents the power to detect the trend given the scenario. We note that only changes in the direction (decline or increase in abundance) that the user specifies are included in this percentage; if, for example, the interest is to detect a decline in animal abundance, only those iterations with a significantly negative change parameter are included. In Appendix 8.1, we describe individual steps and components for the simulation in more detail.

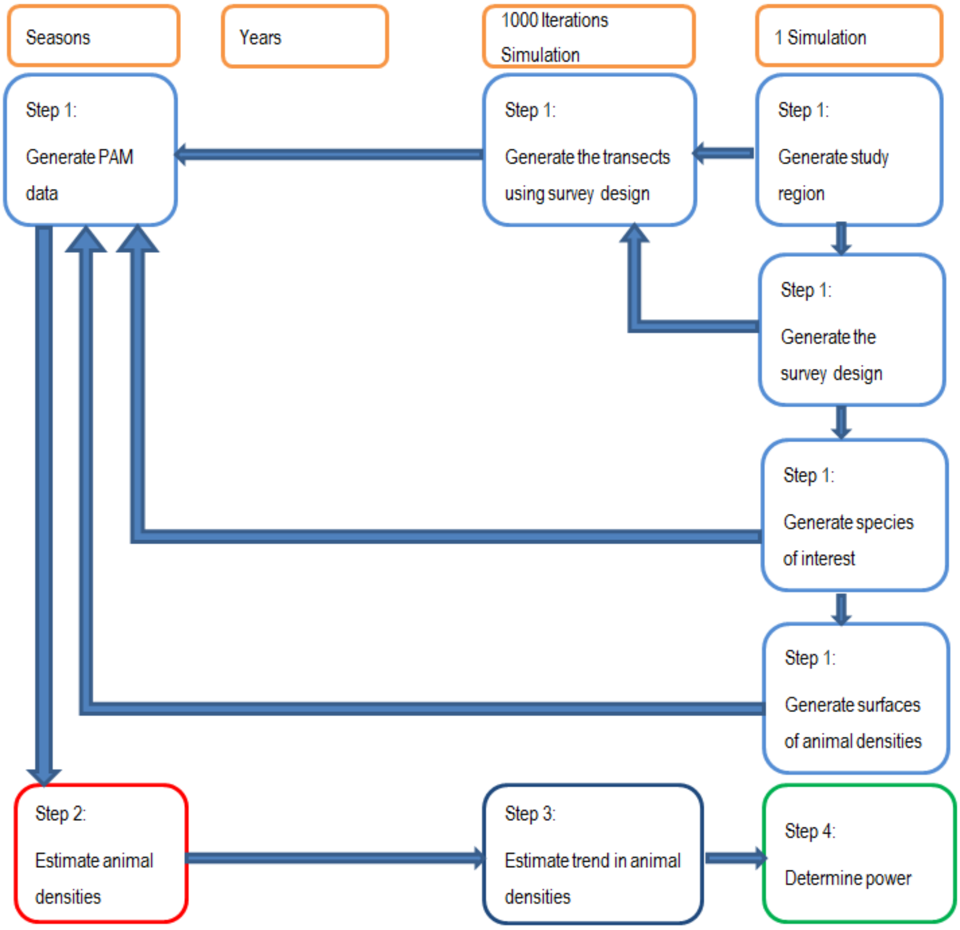


Figure 9 - Diagram depicting the different levels (top row) of the simulation tool and which level the four steps (color-coded) of the simulation belong to. Note that step 1, can itself be divided into smaller steps which span across the lowest (season) and the two highest levels (iteration and simulation). Transects are the Distance sampling term for PAM systems in this instance.

5 Application: Exploring sensitivities of PAM survey elements

5.1 Introduction to sensitivity analyses

Following the development of the simulation tool, the next step was to explore, under controlled settings, the different elements of the PAM survey and the analysis components and their overall ability to affect the power of the survey. The project started with the development of a Distance Sampling version of the AVADECAF tool and latterly redeveloped the core tool infrastructure to use also SECR methods. The bulk of sensitivity analyses have been run using the DS method (see section 5.2 and 5.3) and it is challenging to understand how the sensitivities of the SECR method would differ from those of DS approaches.

Box G: Sensitivity analyses – a cautionary note

There are a number of components that are important to consider when conducting sensitivity analyses. One particularly crucial one is that we can only assess the sensitivity of a limited number of parameters. By doing this we can explore the change in power of a given survey caused by the change in the parameter of interest. However it is to be expected that a more realistic scenario (with multiple variables altered) would be extremely challenging to interpret.

Furthermore in controlling patterns within simulations we are forced to make assumptions about variables (e.g. such as cue production rate) being constant (or varying in a consistent manner) year to year. This is likely not to reflect reality. With the specific example of cue rate, we have already identified that there are a wide range of elements that fluctuate between individuals that will and there will be variation among some age/sex classes of certain species (e.g. not producing any vocalizations) and/or, production changing seasonally and by exposure etc. Where possible we need to monitor these specific cue rate elements on site – to ensure these issues can be managed.

Here, the effect of interest is likely to be that caused by exposure to industrial activity. So this needs to be framed not just as years but as intensity and duration of exposure to activity – therefore this type of assessment is challenging here as we are not focusing specifically on the exposure, but the general power to detect a trend.

5.2 Sensitivity analyses – Distance sampling method

To assess the relative contribution of PAM elements to the overall power to detect a specific annual change in animal densities, it was necessary to first determine a suitable baseline against which to compare scenarios. We chose a scenario based on a published study of beaked whales from Marques et al. (2009) to define the main parameters of the survey. These values come from a series of research studies and analyses conducted on AUTEC in the Bahamas. Whilst we concede that the choice of a beaked whale species on a PAM instrumented US Naval range is not ideal (as it is unlikely be replicated on an E&P site) – it provides the most complete assessment using DECAF methods to date – and therefore a solid foundation from which to assess sensitivities. Please note that we did not restrict survey scenarios to the actual number of hydrophones deployed on AUTEC. Furthermore, as our focus here is exploring how varying individual parameters affects the power of a specific PAM survey, specifying exact parameter values is less important than the change in power as conditions, i.e., their interaction and variability, vary. The baseline against which power to detect a 5% annual change was assessed is outlined in Appendix 8.2.

There are several approaches to analyzing time series data to investigate (1) whether trends are occurring through time, and (2) whether there is enough statistical power to detect a trend in the data (Thomas *et al*., 2004). Assessing statistical power is a way of determining whether the coefficients of variation associated with density estimates in a time series dataset are precise enough to be able to detect a trend, if one was present in the time series. In this project, at each iteration of the simulation, we fitted a generalized linear model to the density estimates through time and assessed whether the estimated model trend coefficient showed a significant decline (using a p-value of 0.05). We could then use the many simulation iterations to assess the power of the survey (i.e., the proportion of simulations where the trend was statistically significant). For example, with the baseline settings and the values from Marques et al. (2009), we found this generated a high power for the baseline surveys of ~98.2% (that is, in 982 out of 1,000 of simulations, a significant annual decline was detected when the beaked whale densities declined by 5%).

The following sections outline a series of PAM survey specific questions that we explored via sensitivity analyses. Each analysis is explained in more detail before presenting the results in comparison to the baseline scenario outlined above.

5.2.1 What is the effect of the number and type of nodes on the power of analysis to detect change in density?

A fundamental question in the design of any PAM survey is deciding on the number of units to be deployed. When the objective of the survey is to generate estimates of density for marine mammals, it is important to understand the number of PAM systems required and the ratio of 'detection only' vs 'localization' systems, i.e. n-node vs p-nodes. As described above, localizing nodes provide range estimates to vocalizing animals. This was assessed by changing the spacing of PAM units (within the 60 x 30 km study area) compared to the baseline scenario. For this purpose, we explored how the power to detect a change in animal densities changed with changing PAM array designs. The results from the sensitivity analysis and different tested array configurations are shown below (Figure 10 and Table 7). We explored three different scenarios with varying values for parameter *sigma.cv* (pertaining to the random variability in detection functions between detectors). As a reminder, in order to introduce heterogeneity between the different PAM nodes for a given season and year, values other than the default *Null* can be specified for *sigma.cv*. This parameter indicates the amount of variability in the scale parameter of the detection function *sigma* and hence in the average detection probabilities *p* between nodes for in each season and year. The three scenarios we explored in Figure 10 include one where with almost no heterogeneity in detection functions between nodes (Figure 10A), one scenario with a small amount of heterogeneity across the study region (10B) and one scenario where there is a reasonable amount of heterogeneity (10C).

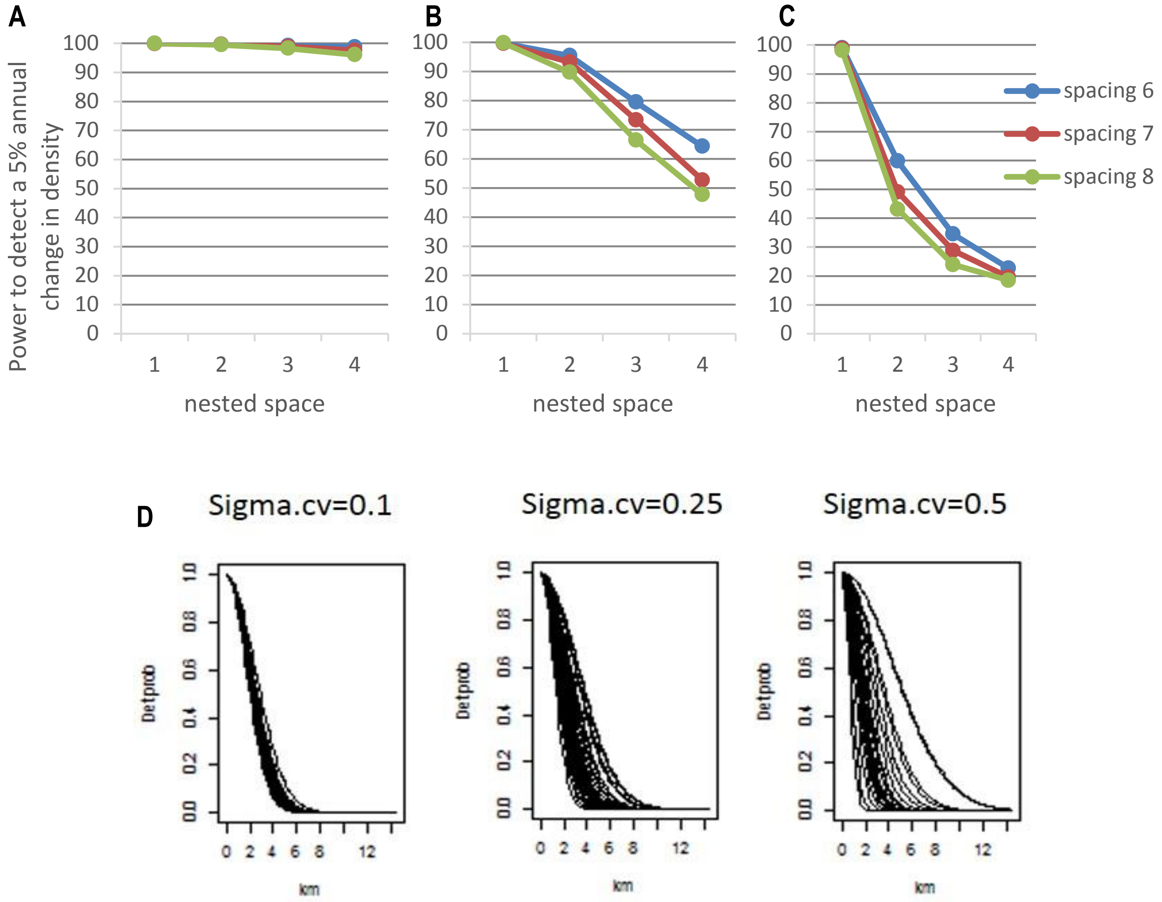


Figure 10 - Effects of changing the number of p-nodes (localizing PAM units) via parameter *nested space* and the total number of nodes (PAM units) via parameter *spacing* on the power to detect a 5% annual change in animal density for varying values for the heterogeneity in the detection function (parameter *sigma.cv*) – A (sigma.cv = 0.1), B (sigma.cv = 0.25), C (sigma.cv = 0.5), illustrated for the corresponding values of *sigma.cv* in the bottom panel (D). Top panel: different *spacing* values in km are shown in color key.

Table 7 shows the corresponding number of nodes in the various spacing / nested.space combinations tested for this study region.

Table 7 - The combinations of spacings and nested.space values tested in simulations and the number of p-nodes ('localising') for each combination. Along the top row the spacings (and resulting number of n-nodes) are shown. On the left side, the different nested.space values drive how many p-nodes are required (see values in table). This is for a survey area of 60 x 30 km.

		spacing		
		6 km / 50 nodes	**7 km / 37 nodes**	**8 km / 28 nodes**
nested.space	**1**	50	37	28
	2	12.5	9.2	7
	3	5.6	4.1	3
	4	3.1	2.3	1.9

What we observe is that when there is very little heterogeneity in detection functions between nodes, the number of nodes (and consequently the number of localization nodes) has little bearing on the power of the survey.

However it is very unlikely that such an environment exists in a real PAM survey for marine mammals. In those simulations with low to moderate heterogeneity (middle and right panels), we found that the total number of nodes had less effect on the power than the number of p-nodes within the array. As the number of p-nodes decreased (i.e. as a result of increasing values for *nested.space*), the power of the survey array to detect a 5% annual change reduced steeply (Figure 10). If all of the nodes deployed were capable of range estimation, pooling robustness (Buckland et al. 2001) would allow us to estimate the average detection function across all nodes and therefore develop an unbiased picture of this detection function. However when we set *nested.space* to 2 (meaning each p-node is surrounded by one n-node in each direction before another p-node), the power drops by 10% in scenarios with minor heterogeneity but by 40-60% where the study site is quite variable in terms of detection probabilities. This issue becomes more pronounced as the number of p-nodes decreases further. This is because with fewer localization nodes, we leave larger gaps across the study region, over which we must assume the same detection function applies (when in reality – where there is variability – it will be a poor assessment of detection functions). Bias can occur, for example, if detection probabilities were particularly low at the p-nodes compared to the n-nodes. Then, the average detection probabilities would be underestimated and, as a result of applying this same detection function to all the n-nodes, densities would be overestimated. This is a crucial point because in this controlled sensitivity analysis we can assess what biases occur and the direction in which they push the estimates. But in a PAM survey you will only have the data you collected and will not know if these have been subject to biases (which can lead to over- or under-estimates of animal density).

5.2.2 What is the effect of different detection functions on power?

In Section 5.2.1 we explored the effect of changing the survey design (number of nodes) under different conditions (by varying *sigma.cv* in simulations – the heterogeneity in detection functions). Here, we explore how that variance around the scale parameter *sigma* (resulting in different detection functions) affects the power to detect an annual decline for different values of *sigma.* Figure 11 provides an example of how the different *sigma.cv* values affect the detection function in simulations for the default *sigma* value in the beaked whale scenario *sigma* = 2).

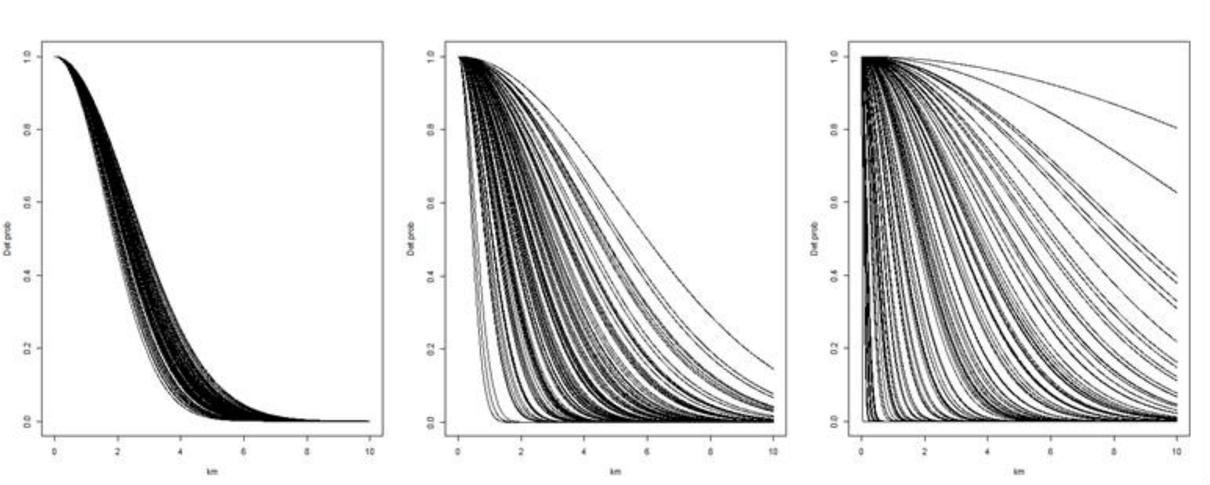


Figure 11 – Examples of how an increase in *sigma.cv* affects the variability in detection functions in simulations for a default value of *sigma* = 2. For each plot, *sigma* =2 but the *sigma.cv* varies between 10% (left), 50% (middle) and 90% (right) (i.e. 0.1, 0.5 and 0.9). y-axis is Detection Probability; x-axis is km.

Figure 23 (Appendix 8.1) shows how the shape of the detection function changes as the scale parameter is changed. These different detection functions can be the result of different environments in which the PAM survey is taking place. For example, for a detection function with a value of *sigma* of 1 detection probabilities drop off more steeply at close distances from the PAM sensor increases compared to the default value we use in the beaked whale example of *sigma* = 2. This could be indicative of an environment in which detection is compromised (e.g. by a noisy environment or a faulty detector) – and therefore only cues produced by animals close to the sensor are detected. Higher scale parameter (*sigma*) values present scenarios in which the probability of detecting a signal stays high for a longer distance from the sensor (see Figure 23 in Appendix 8.1).

Here, we explored how different combinations in the shape of the detection function (governed by *sigma*) and variance around the detection function (governed by *sigma.cv*) affected the power of a PAM survey (with all other parameters remaining constant across the simulations. Figure 12 shows that it is the variance around the detection function (i.e. the variability in detection function across the study region) that drives the power in the survey as opposed to the values for *sigma*. As *sigma* changes, the curves in Figure 12 remain close to one another, but increasing *sigma.cv* (increasing variability) results in a decrease in power for the simulated PAM survey. Under the baseline scenario where *sigma.cv* = 0.1 the power is still high, but a *sigma.cv* value of 0.5 results in a decrease in power of 70% from the baseline scenario (with all other parameters constant).

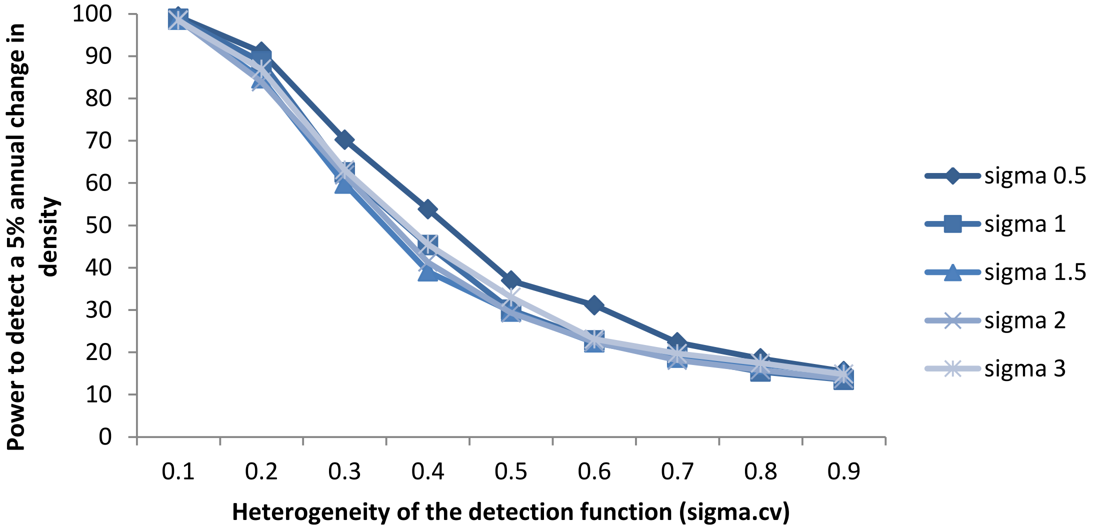


Figure 12 Effects of varying the detection function (via *sigma*) and the precision of the detection function estimate (via *sigma.cv*) on the power to detect a 5% annual decline in density.

In Figure 12 you can see that the curves for larger values for *sigma* are typically below curves for lower values of *sigma* for ranges of sigma.cv 0.2 – 0.8. Meaning, for a given *sigma.cv*, higher sigma values will result in slightly lower power.

5.2.3 What is the effect of adjusting c, a and f (and variances) on power?

In the sections above we explored the effect that increased variance (in *sigma.cv*) has on the power of simulated PAM surveys when changing the number of p-nodes or the value for *sigma*. Here we examine how elements that we (in theory) have greater control over via auxiliary data we collect and/or the analysis practices we undertake can affect the power of the survey. Specifically we want to look at how changes in estimates of cue rate, perception bias (both largely dictated by vocalization behavior) and false positive rate (dictated by DC approaches employed and the effort invested in exploring the efficacy of the detector/classifiers). We also want to study how the variances around these estimates affect power.

Figures 13 shows the results for species with different cue rates and variance (*c.cv*) values. What we observe is that it is the variance that drives the power of the survey. The power did not change significantly with different cue rate values. The simulations with lower variance values (i.e. *c.cv* = 0.01 – 0.2) had little effect on the power, but as variance increased > 0.2 (i.e. 20%), power dropped steeply by 30% (*c.cv* = 0.4; 40%) and ~70% (*c.cv* = 0.6, 60%).

What this means in practical terms, if we can obtain a precise estimate of cue rate, then the power of the survey will remain high, but as variance around the cue rate increases, the power declines. This places a premium on obtaining precise cue rate information. This is further explored in Box H – which looks at how the collection DTAG data can improve power.

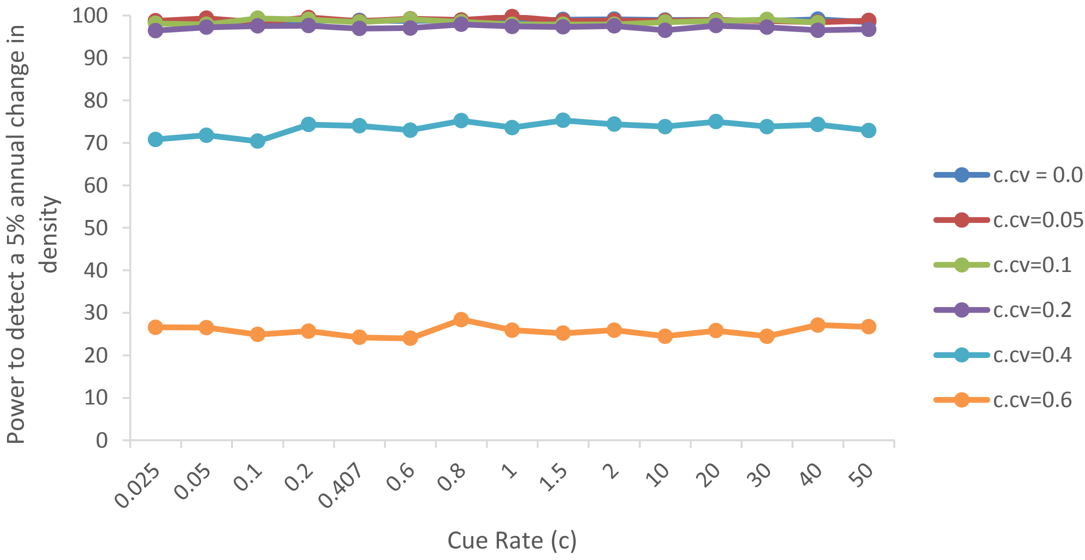
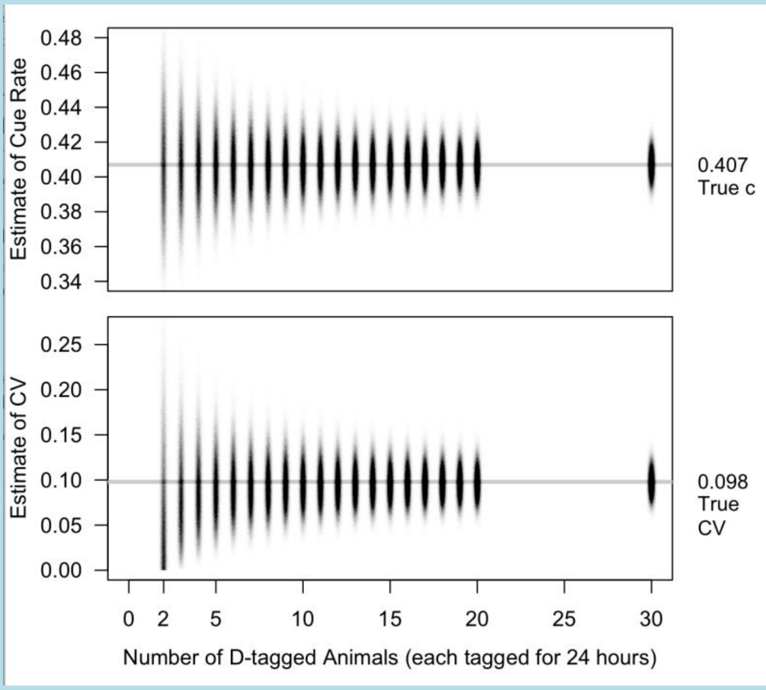


Figure 13 - Effects of varying cue rates (*c*) and the precision of cue rate (*c.cv*) on the power to detect a 5% annual change in density.

BOX H: How many whales to attach DTAGs to: a separate study in power.

We ran a secondary power analysis to see the effect of the number of tags on the precision and accuracy of estimating the cue rate and its CV, from tagging from 2 to 30 whales. The results shown in the figure below suggest there is little advantage gained in precision after 7 or 8 whales are tagged. The figure also identifies the sample CV as a biased estimate for the 'true' CV at small sample sizes.



To examine the effect of tagging whales on a study's ability to estimate the cue rate and its variability accurately, we set the following simulation up. We took the observed cue rate and its standard deviation from the 5 tagged beaked whales in (Marques et al. *2009) as the population parameters c and σ_c. These parameters are used to set the shape and scale parameters of a gamma distributed data-generating function and generate a random realization for each year-season sampling period[3]. We use this random c realization, along with a CV_c generated from the following X^2-distribution to generate the cue rate for each individual tagged whale as follows (i.e., the observed cue rates estimated for each tagged whale).*

We acknowledge, however, that this analysis assumes that the initially tagged whales were representative of the population of beaked whales of interest in their vocal behavior. This assumption could easily be violated if a proportion of the population were more likely to be tagged, and this behavior was correlated to their vocal behavior. Furthermore, there are many factors that could influence the level of variability of a given species' vocal behavior, so the amount of required tagging may differ spatially and temporally for the same species. Cue production rate, and how best to estimate it for the purposes of density estimation, are active research topics and there is still a lot to investigate. Nonetheless, this preliminary power analysis demonstrates how future tagging studies could be facilitated by estimating an optimal sample size of tagged animals.

3 - The one realization of this cue rate from a gamma distribution is the underlying cue rate for that season-year combination, and is therefore randomly different from other season-years due to factors affecting cue rates such as for example, the behavior of the animals, the presence of other species, temporal fluctuations, etc.

Figure 14 shows the effect of variable perception bias on survey power. As with cue rate, when the variance is low ($a.cv \leq 0.1$), there is little effect on survey power (irrespective of the perception bias value, a_{perc}). However as variance increases > 0.2 the power drops steeply by ~35% (a.cv = 0.2), ~80% (a.cv = 0.4) and ~85% (*a.cv* = 0.6). When variance is high, the power drops slightly as perception bias increased – but the main effect is driven by variance.

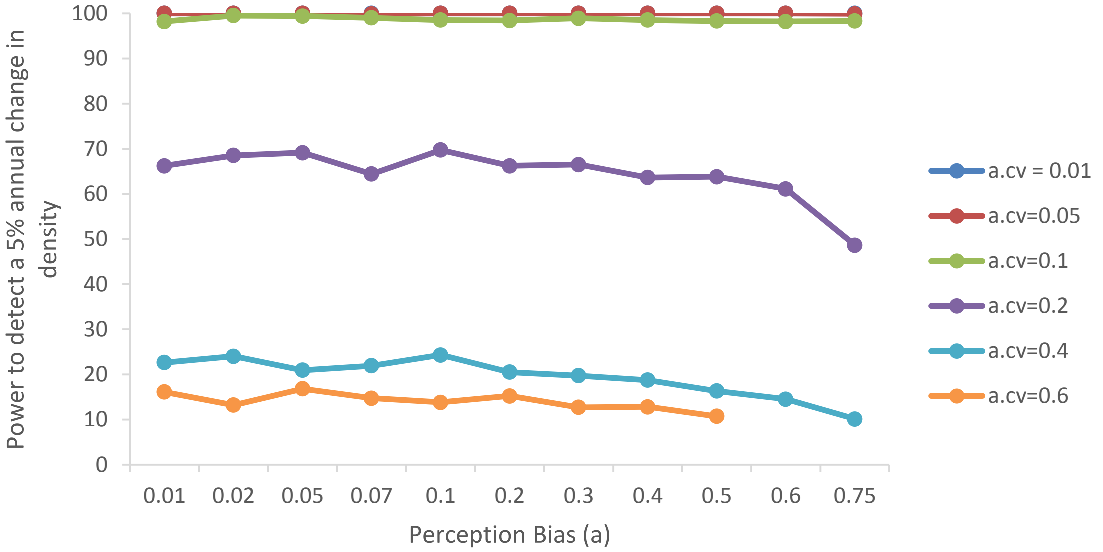


Figure 14 - Effect of varying the perception bias parameter (a_{perc}) and its precision ($a.cv$) on the power to detect a 5% annual change in animal density.

Figure 15 shows the effect of variable false positive rate on survey power (e.g. driven by the quality of detector/classifier). As with cue rate, when the variance is low (*f.cv* ≤ 0.1), there is little effect on survey power (irrespective of the false positive rate, *f*). However as variance increases > 0.2 the power drops steeply as false positive rate increases (i.e. the detector/classifier performance gets worse). When variance is high, the power drops further as false positive rate increased – but the main effect is driven by variance / high false positive rates. This highlights the value of developing robust detectors and classifiers for a species, but also being able to minimize the variance associated with the false positive rate.

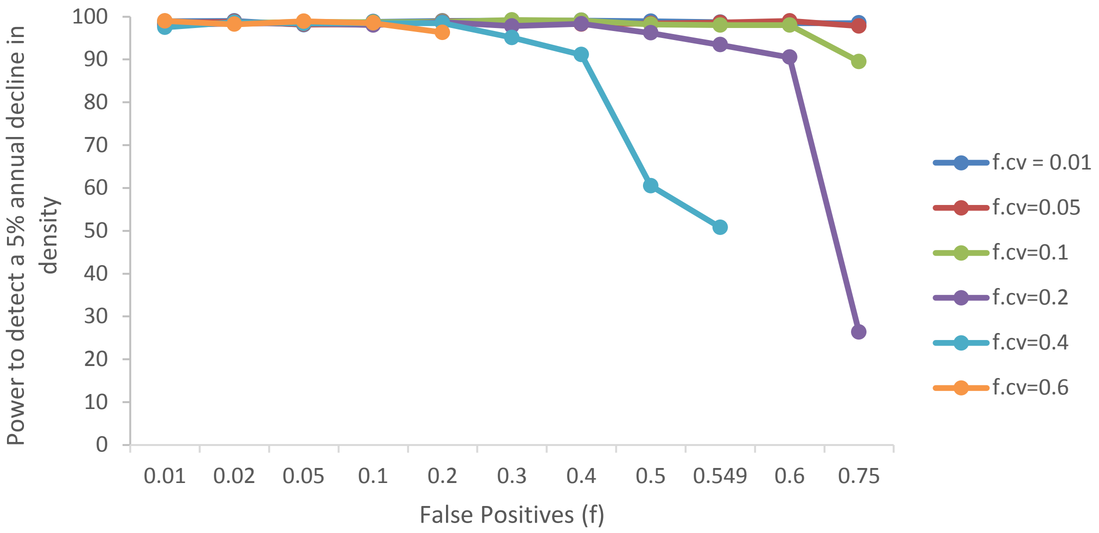


Figure 15 Effect of varying the false positive rate precision (*f.cv*) on the power to detect a 5% annual change in animal density.

For each of the three parameters investigated in this section, cue rate, perception bias and false positive rate, increasing the variance of the parameter had a negative effect on the power to detect an annual decline. The values of the parameters themselves did generally not influence the power.

5.2.4 What is the effect of localization performance (how well it works) on power?

As discussed in section 3.3 the role of localization (for range estimation) is fundamental to density estimation using distance sampling approaches. The power analysis tool has been developed with the ability to explore different levels of localization performance (i.e. how well the PAM system is capable of generating range estimates to the vocalizing animal (see section 3.3 and 8.1.9)). At one extreme, the tool can assess the power of a given survey where we assume that localization is possible at all distances and therefore horizontal range estimates are attained for all cues out to the truncation distance. This level of localization is unlikely to be achievable in the marine environment with wild animals – particularly given the level of understanding in vocalization behavior (section 3.1) and the challenges inherent in range estimation with PAM surveys (section 3.3). Therefore we also explored scenarios where range estimates are only achievable close to the cluster of sensors, but beyond a maximum horizontal range, animals are 'detected only'. We explored how the power of the survey declines when that maximum horizontal distance where localization is possible is 250 m, 500 m and 1,000m (compared with 12 km in the baseline). Figure 16 indicates that as we move away from this unrealistic best case (range estimates attained for all cues) to the more realistic scenarios where localization is only possible close to the sensor, the overall power of the survey declines but only marginally (e.g. 20%). However this is not insignificant and as we've acknowledged previously, the tool is likely over-simplistically modelling localization and errors and as such these results should be interpreted with caution.

One other important note here is because we are assessing how error in localization (and localization performance) affects power, is that here we are assuming that we have a constant error rate. Under such circumstances where measurement errors stay the same throughout and are controlled for (as they are here in this simulation) then the power to detect a trend will not change (because range measurements are made poorly, but consistently poorly throughout the deployment). This is a further reason to interpret these results with caution

– as any factor during a long-term deployment – that affects the error (e.g. an improvement in range estimation or PAM technology) could affect the survey results.

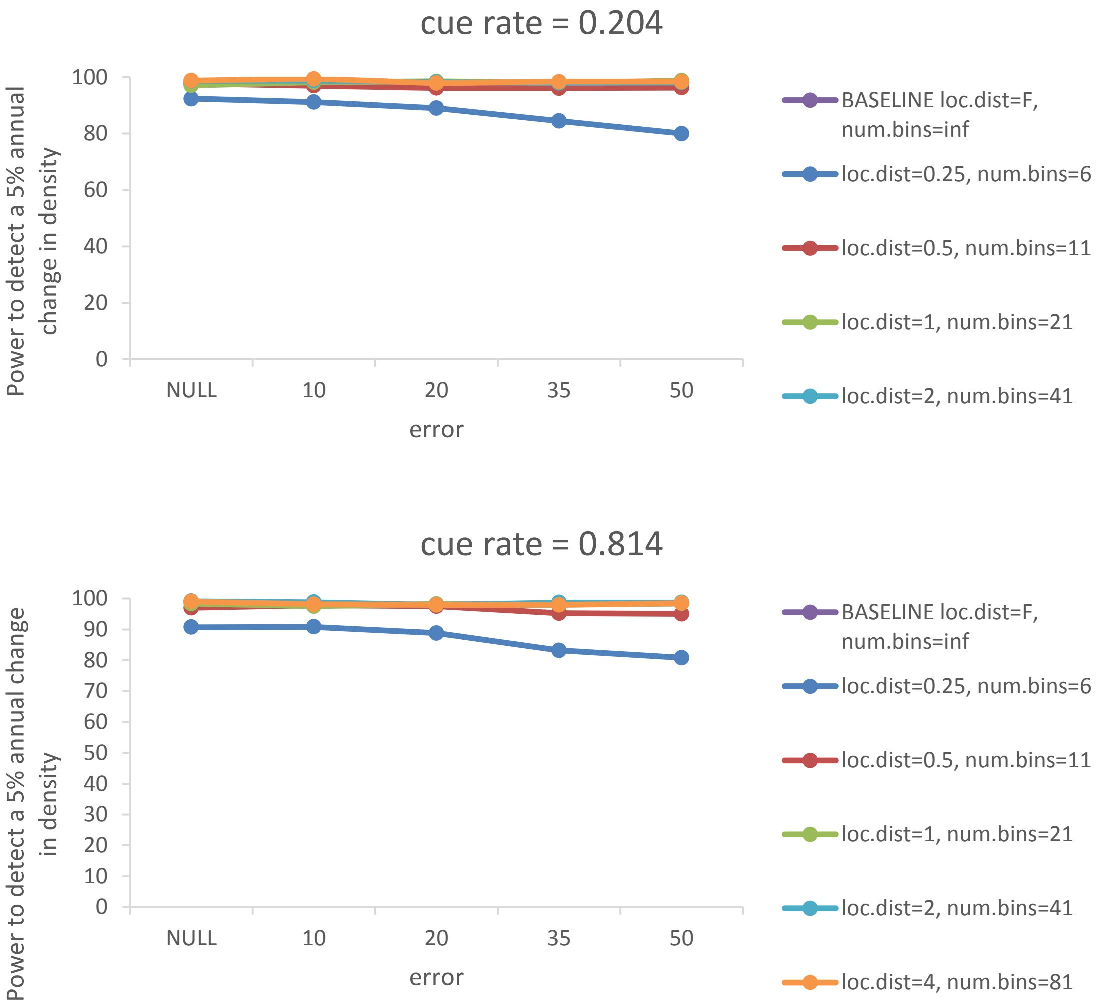


Figure 16 Effect of varying the systematic error in localization (*error*) on the power to detect a 5% annual change in density when the localization distances are truncated at varying distances from the p-node (*loc.dist*) and when the detection function data are fit to varying number of bins (*num.bins*).

5.2.5 How error affects power and biases impact PAM surveys

In this section we explore the effect of varying amounts of heterogeneity in the detection functions (defined by parameter *sigma.cv*) in the presence of systematic and random errors in the distances measured by the localization nodes (i.e. how precisely or not the distances to animals are measured), defined by parameters *error* and *error.spread*, respectively (sources of error are discussed extensively in section 3.3 – therefore we know this a real concern for such systems in PAM surveys for density estimation). For a given value of *sigma.cv*, the value for *error* or *error.spread* did not have a strong effect on the power (Figure 17). However, as before, increasing the value for *sigma.cv* reduced the power to detect the annual decline.

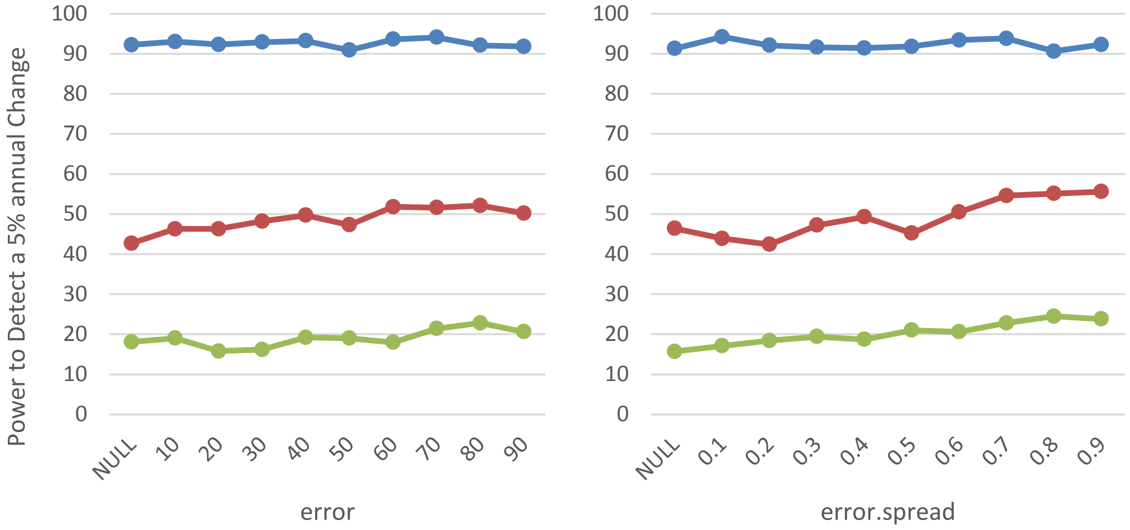


Figure 17- Effects of changing parameters *error* (left, with *error.spread* = 0.5) and *error.spread* (right, with error = 50) on the power to detect an annual decline of 5% in animal densities for three different values of *sigma.cv*, blue: 0.1, red: 0.25, green: 0.5.

A by-product of introducing error (via *error* or *error.spread*) is that this will result effectively result in a larger *sigma* value; hence, the seemingly increasing power with increasing *sigma* but only when sigma.cv is 0.25 or 0.5 (red and green respectively). Although power is constant across changes in *error* and *error.spread*, accuracy on the estimate of density decreases or the estimate of population density decreases with an increase in both *error* and *error.spread* localization errors.

As hydrophone performance, or sound propagation qualities vary by unit and by location, the detection function is then specific to the hydrophone and its location. This is what we aimed to capture with the parameter *sigma.cv* which describes the amount of variability between detection functions for a given year and season during the part of the simulation where we generated the PAM data. In the analyses, however, we assumed that all hydrophones were detecting cues similarly, and fitted the same detection function to all cue data collected in the study area. This between-node variability propagated error into the precision and accuracy of the fitted detection function. And, as the differences between nodes increased (*sigma.cv* gets larger), there was more bias introduced into the estimated detection function. This bias in the detection function then again propagated into the estimate of animal density, with the consequence of dropping the power to detect a decline.

Understanding the implications of bias being introduced to PAM surveys for density estimation is crucial. In this controlled sensitivity analysis we can assess what biases occur and the direction in which they push the estimates. However, in a PAM survey only the data collected at the PAM stations will be available and it is unknown if these have been subject to biases (which can lead to over- or under-estimates of animal density) (see Figure 18 and 19).

We therefore have underestimated whale density if the systematic error applied is positive, and this is seen in this figure where *error* increases as you move to the right across the panels. Likewise, with random localization errors, *error.spread*, we have assumed a symmetric, Gaussian error form with the mean and variance a function of the true distance. However the expected number of cues per unit distance should increase at a rate that is the square of the distance. Therefore we are adding random error, but putting more cues farther from the hydrophone than would be expected, with the consequence of moving the detection function shoulder to the right (hence, increasing the average detection probabilities *p*), and again as *p* is in the denominator of the density estimator (see eqn 4 or Marques et al. 2009), we underestimate whale density.

The risk here is that with under-estimation of the population size across the length of the study, can have dangerous interpretations if not recognized. The conclusion from the right most panel in the series below is that we have estimated the population to be 25% of the actual population size.

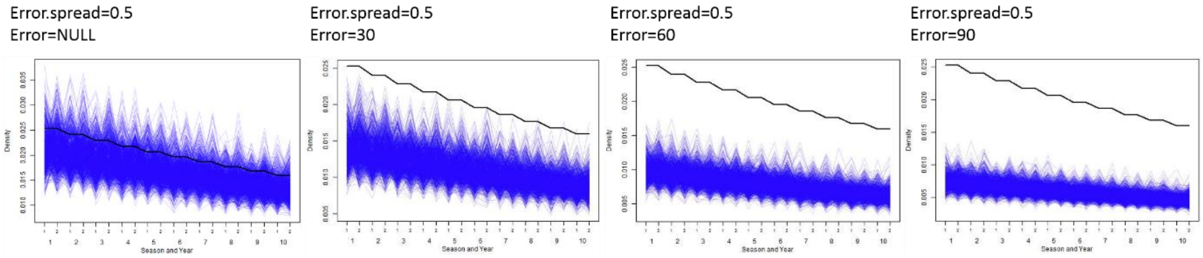


Figure 18 - How varying the random localization error (error.spread) changes the estimated density (blue lines) in relation to the true density (black line). Error = 50% for each plot.

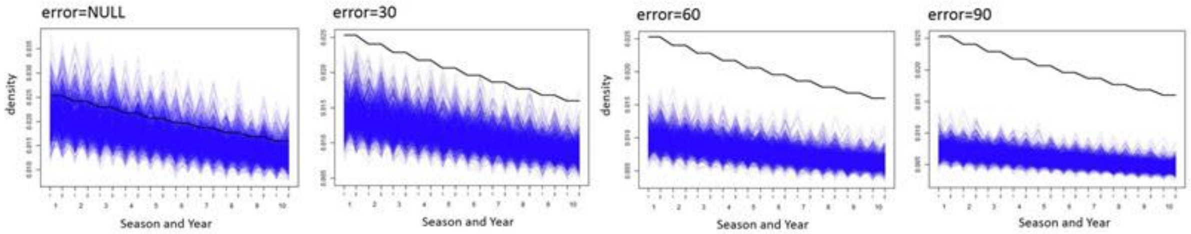


Figure 19 How varying the systematic localization error (*error*) changes the estimated density (blue lines) in relation to the true density (black line). *Error.spread* = 0.5 for each plot.

This means that although we may have detected a significant decline in power, the absolute number of animals impacted by the project would have been underestimated. By exploring the relationships presented above (Figure 18 & 19) we highlight the importance of keeping systematic errors in localization small (which has important consequences for the accuracy in the estimation of whale density, even if the precision around the estimate of density, and therefore the power to detect a decline, remains high). Critically, understanding the implications of error being introduced to PAM surveys for density estimation is crucial. Because in a real PAM survey, one only has the data collected and it was not be possible to understand the extent to which these have been subject to biases (which can lead to over- or under-estimates of animal density) – hence, here we stress the emphasis must be put on minimizing errors in localization for range estimation in any real-life PAM deployment where the objective is density estimation.

5.2.6 What is the effect on how often you 'analyze' your data (i.e. get density estimates) on power?

The simulation tool is configured such that it works as though a 10 year survey have been conducted (though the number of years can be changed if the user wishes) and at the end of the survey, a PAM technician or statistician is analyzing the data collected. We explored how the power of the baseline PAM survey varies if we adjust the number of years the survey is run for and explored the power of the survey at the end of each year. Simply put, having more survey data results in increased power (Figure 20) (though once the survey length exceeds 10 years (for this scenario) the power will always be close to 100%). This provides information about how the amount of survey effort affects the power of any given survey. Given the baseline scenario defined for the beaked whale example. We note, that this assessment of how many years are required to obtain a power of, e.g., 95% to detect an annual decline will depend on each of the parameters specified.

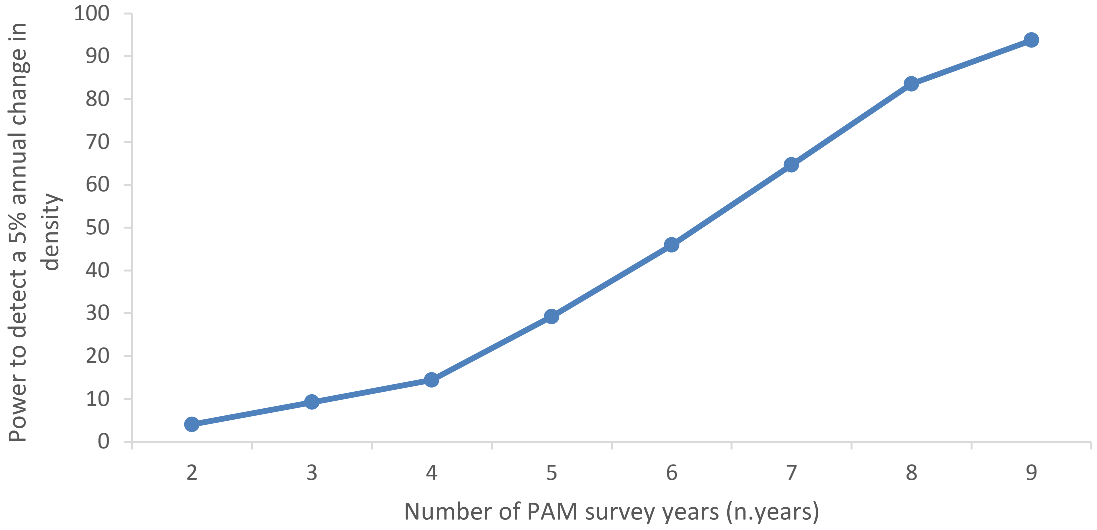


Figure 20 - How varying the number of years the PAM survey is conducted for (*n.years*) changes the power to detect a 5% annual decline in density.

Another part of the analysis following a PAM survey is considering the number of density estimates calculated for each year (i.e. the temporal period of analysis). We therefore explored how power changed depending on how many density estimates we generated from our (simulated) PAM data. We find that the more density estimates we retrospectively generate via analysis within each year, the higher the power (Figure 21).

This suggests that, if the collected data support it, as fine a temporal unit as is sensible should be selected when maximizing power is the focus (this assumes a PAM user has sufficient data following the survey). It is important to note that different survey set-ups will have different 'baseline' power and therefore adjusting this parameter (*n.seas*) will affect the power of the survey program dependent on the survey parameters selected.

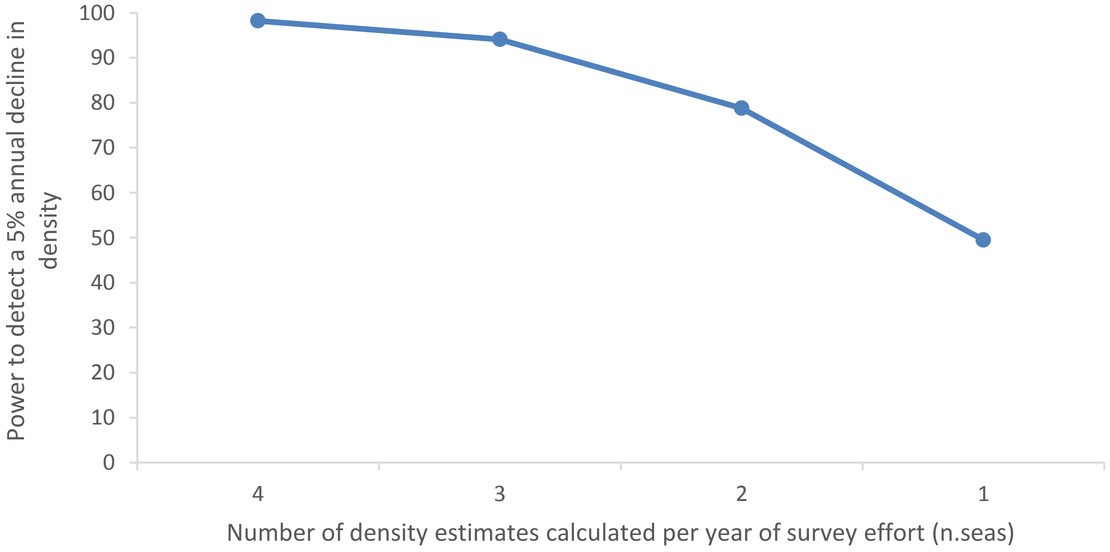


Figure 21 - Shows how the survey power varies with the number of density estimates calculated for each year of the study (*n.seas*).

5.3 Exploring SECR sensitivities

The tool has the capability to explore PAM scenarios utilizing SECR approaches to density estimation. As noted in section 4.1.2, SECR is a relatively novel approach, however, some of the information required to simulate data here is different from the Distance Sampling tool. This introduces a practical limitation in what can be assessed using the SECR part of the tool. Whilst the DS version of the tool runs relatively quickly (e.g. in minutes) and as such we were able to run t hundreds of simulations for the sensitivity analysis above, the SECR version of the tool is more cumbersome and takes 40-50 hours to run on a standard laptop. This has severely limited our ability to assess the sensitivities of the SECR methods. It is, however, something that could be further developed and explored in future studies using high power supercomputers to run simulations very quickly.

We do not have sufficient SECR simulations completed to facilitate a sensitivity analysis. To conduct an equivalent analysis to that of the Distance sampling version of the tool, would likely require on the order of ~1,500 days of continuous simulation time – therefore any further simulations need to be carefully selected given the structure of this method and the limitations this imposes on any SECR tool. This is an area for further development and analysis.

6 Conclusions & Recommendations

6.1 Conclusions

Before this project started the intention here was to deliver a comparative critical assessment of PAM systems in their ability to be used for density estimation using DECAF methods. However in light of the results and the findings of the reviews, it seems that there are two key dependencies for any PAM survey (in terms of hardware) and those are the need for systems that have range estimation capacity (from which robust detection functions can be generated) and the *in situ* collection of auxiliary data to inform cue rate (a key element that was challenging to adequately capture in the simulation tool). The needs for these is discussed elsewhere in this report (section 3.4 and section 7.4) and summarized by Johnson et al. (2009).

In the previous sections we have explored the wide range of biological, practical and statistical elements that are comprised within a long-term fixed PAM deployment to estimate density. In the following sub-sections we outline some of the key conclusions and recommendations to be considered for future developments.

Having an understanding of the sounds that marine mammals produce is an essential starting point for density estimation using PAM methods. In terms of PAM surveys, it is crucial to understand the vocalization behavior of the species of interest and we know enough to demonstrate that the signals produced by marine mammal vary a great deal between species (ranging in frequency, duration, source level, directionality etc.). There is a need to deepen our understanding of vocalization behavior – and this is likely to continue to be improved in site-specific studies.

One of the key inputs for PAM methods are detailed knowledge of cue production rate and how that varies among individuals of different age/sex classes, as well as seasonally and under different conditions (e.g. different depths, foraging / breeding ground etc.). In order to collect these data, tagging animals with acoustic tags (e.g. DTAGs, Acousonde) which can record vocalizations of the focal animal will be extremely valuable. A lot of data from deployed recording PAM systems, cross-referenced with visual identification of species have proved invaluable as these can feed into DCL efforts to develop classifiers and/or algorithms.

Calls from some species are also known to vary over time or between regions which will dramatically affect the efficiency of some types of detector. Mis-classification of species that produce similar sounds remains a major limiting factor and requires further development but characterization of detector performance is often hampered by a lack of ground-truthed test data. Different call types and different background noise conditions require different detection algorithms, so multiple detectors and classifiers may be required to process a single dataset. A detector / classifier is characterized by how efficiently it correctly identifies calls and by its false positive rate. While it is generally not necessary to have a perfectly efficient detector, or to reduce false positive rate to zero, it is important to be able to quantify them (and minimize variances).

For localization - frequency, source level, the directionality of a vocalization and diving behavior of a species will dictate how this approached; therefore the optimum design of a *localizing hydrophone array* is *species specific*.

In addition, even for the most optimal localizing array, it is unlikely that the probability of localization is the same as the probability of detection, especially for directional vocalizations: thus, in a distance sampling framework, to apply the probability of localization to 'n-nodes' requires an understanding of how to convert a probability of localization to a probability of detection. But this is poorly understood.

There are two potential pitfalls in localization which directly affects DECAF. 1) Error in localization increases significantly with range from a hydrophone array and is roughly proportional to the spacing between hydrophones (the array aperture size). Therefore, although an animal may be detectible on all or some hydrophones within an array, the aperture size of the array may mean it cannot be accurately localized. 2) Some vocalizations are highly directional (*i.e.* clicks from odontocetes). This can lead to a situation where an animal may be within the effective localization range of an array but is orientated in such a way that it does not ensonify and therefore be detected by enough hydrophones within the array to allow a position to be calculated.

Finally it is crucial to consider one of the key advances made by this report. **The development of the AVADECAF simulation tool is the first time that all the elements of DECAF-based approaches have been brought together in a simulation capability.** Therefore, it is the first time that we have been able to assess which components of a PAM survey design, which aims to detect changes in density, are the most crucial to a monitoring effort. It is clear that each element of the tool could be developed to become more complex (and ideas for future development are discussed below). However, these adjustments will be made much easier now that the simulation framework is in place. Furthermore, the use of sensitivity analyses and this framework can provide a roadmap to guide where research effort needs to go next. Therefore this represents a significant step forward in making fixed PAM monitoring efforts more focused and cost effective for identifying long-term changes in cetacean density estimations.

6.2 Recommendations

In this section we outline the recommendations for improvement to the tool, before discussing how data / technology gaps can be filled for long-term PAM monitoring during the life cycle of E&P activities.

6.2.1 Distance sampling, SECR and the power analysis tool

6.2.1.1 *Complex region boundaries*

Currently, the simulation tool only allows for using rectangular regions with vertical north-south and horizontal east-west orientation. This way, the creation of a buffer zone of width *w* (the truncation distance) surrounding the region is very straightforward. This *region.withbuffer* is required for both the distance sampling and the SECR approaches as it is assumed that, while nodes are only placed within the region, the cues can be detected from outside the region but not from outside the region with the buffer. Furthermore, the function for the SECR approach *transects.to.traps* relies on this orientation of the region for allocating positions to the nodes. However, complex regions would make the simulation more realistic and flexible. Creating a complex region is already implemented in the R package DSsim. Hence, a method for creating the *region.withbuffer* would need to be implemented and the function *transects.to.traps* would need to be updated.

6.2.1.2 *Spatial models for generating a density surface*

Currently the simulation tool uses the functions *make.density* and *add.hotspots* from the R package DSsim to create an animal density surface which is used for both the distance sampling and the SECR approaches. These functions do not rely on an existing model of animal densities. The user has to define a density value which is used to create a flat density surface to start with. Then hotspots can be added where the user has to define by how much the density is higher in a specified location as well as how steep this increased density decays back to the flat density surface. In some study areas, however, spatial density models for cetacean species exist already and may be used to create a more realistic scenario. Spatial density models are more flexible and realistic than our current approach as they describe the relationship the animal densities and covariates (environmental and/or spatial) and predict animal abundances throughout the study area based on this relationship. For example, some species might be more abundant along the 200m depth contour (the shelf break), a pattern that would be difficult to implement with the current approach. Furthermore, if the creation of the cue data was based on a spatial density surface, we could use the spatially explicit power analyses methods from the MRSea package (Scott-Hayward et al. 2013) for more detailed analyses.

6.2.1.3 *Expanding detection functions for generating distances*

In this version of the simulation tool the detection function for both creating the detections as well as for estimating the detection function is limited to the half-normal. In some cases, however, the actual shape of the detection function might be better described as a hazard-rate. It would be easy to update the simulation tool to use a hazard-rate detection function for the estimation. It would also be easy to update the tool for creating detections using a hazard-rate function for the SECR method in general and for distance sampling method when using the grid-based approach described above in Section 8.1.7.2. However, the creation of the data using the default method of point densities (Section 8.1.7.1) uses the inversed cumulative distribution function where the cumulative distribution function is based on a half-normal (Buckland et al. 2001b). This method would need to be developed for the hazard-rate function. In addition, detection function models should be expanded (in both distance sampling and SECR approaches) to be able to incorporate variables that may affect detectability (e.g., ambient noise) at both the data generation and analysis stages.

6.2.1.4 *Spatial models for detecting a change*

The current version of the tool allows investigating whether average animal densities within the study region changed. This method is based on obtaining a single estimate of density throughout the study region for each year and season and applying testing for a change (either decline or increase) in a generalized linear model fitted to these density values. Oftentimes, the change may be localized around the immediate area of a given human activity and occur as a redistribution into areas surrounding this area. This is one example which can result in overall densities in the study area remaining the same while they changed in the area with human activity. We implemented such an effect for the creation of the cue data with the methods described in Section 8.1.5.3 (Impact decline). However, our current estimation methods do not allow estimating a localized change. The benefit of using a spatial model for creating animal densities as described in Section 7.5.2 is that it would make it possible to implement the spatially explicit power analyses methods developed for the MRSea package (Scott-Hayward et al. 2013). These methods allow detecting localized changes and their statistical significance at each point of the density surface.

6.2.1.5 *Model selection for the detection function*

Currently only the half-normal detection function is used to estimate one global detection function for all nodes for a given year and season. However, the sensitivity analyses above have shown that heterogeneity in detection probabilities is a major cause for losing power in detecting a change. Hence, it is important to allow more flexibility in the detection function and use the best fitting model to estimate average detection probabilities. Future development should therefore include model selection between the half-normal and hazard-rate key functions and potentially adjustment terms or covariates for the detection model.

6.2.1.6 *Simulating the estimation of the perception bias*

For SECR methods, it is important to first better understand the sensitivities of the tool. One area that arose during development is the parameter pertaining to perception bias on and near the node is estimated using the *secr.fit* function. For distance sampling methods, however, this presents a problem as estimation of this parameter generally requires auxiliary data and complex modelling approaches. The current version of the tool lacks a method for simulating the estimation of this parameter. Instead we introduce stochasticity by drawing a new random sample from a beta distribution (with mean and variance defined by the user).

6.2.1.7 *The variance of cue production rate and false positive rate*

One of the downfalls of the current simulation tool is that we use the same CV value for generating the cue data as for estimating the parameters as well as for estimating the percent contributions to the overall variance. This is the case for cue production rate and false positive rate for both distance sampling and SECR methods and for the perception bias parameter for distance sampling methods. Ideally, the variances of these parameters (and how they may also change over time) would be estimated from either the primary data (if possible) or auxiliary data. Future work should include methods that simulate this estimation.

6.2.1.8 *Probability of detection vs probability of localizations*

A limitation of density estimate using PAM efforts is that for some species, the probability of localization will not be the same as the probability of detection. In fact, it may be the case that the probability of localization is both different in magnitude (i.e. less likely) and does not scale linearly with the probability of detection at different ranges, i.e. the two functions are different shapes. This is an area that should be addressed in any future development of this tool as it will have implications for the tool simulations. This might be addressed with the development of a separate simulation module that could explore the key elements (and run independently of the AVADECAF tool, but inform parameter selection and drive new developments to the model algorithms).

Specifically, such a tool module could simulate probability and accuracy of localization and operate as a standalone java based tool with a simple user interface. The key inputs would be information on: animal vocalization beam profile, acoustic behavior (source levels and inter-click-intervals for clicks sounds), dive behavior metrics (e.g. distribution of depth, descent angle, ascent angle etc.), environmental noise conditions, sound propagation conditions (along specific information on DCL performance and errors). By generating a number of simulated dives of vocalizing animals which can be localized and provide outputs on the 'probability of detection', the 'probability of localization (including the issue of detection matching) and estimate the error surfaces at different depths (receiver and animal). Whilst such a development would be less precise as measuring these parameters it remains an extremely cost-effective way to improve the AVADECAF tool.

6.2.2 Technology development, data needs and fieldwork

Localization is a challenging area and localization performance is species specific. In addition, the probability of localization is unlikely to be the same as the probability of detection for some species. It is necessary that localization performance and the potential dichotomy between detection and localization functions is evaluated before a full system is deployed.

The following tasks would be beneficial to advance the use of fixed PAM systems for density estimation of cetaceans:

6.2.2.1 *Engagement of industry and scientific communities*

Industry and scientists should engage further to assess what (from an industry perspective) is going to be a more cost effective and pragmatic option to advance this field - a cabled array (perhaps similar to that used at AUTEC), or autonomous units that are deployed and recovered, or some combination of the two types of PAM technologies – in order to determine the most suitable technology to be applied or developed to aid monitoring of cetaceans throughout the life cycle of an E&P site.

There is also a pressing need for the scientific community to engage in developing DCL methods in order to advance the field of density estimation. Currently there is a disconnect between the development of DCL approaches and the use of PAM for density estimation. This is because many efforts in advancing DCL approaches are done independently of efforts to improve density estimation using PAM. Via increased engagement and collaboration between experts in fields of PAM, DCL and density estimation, the key elements of PAM and DCL to improve our ability to using PAM for estimating density. Without this direct collaboration, these fields will develop independently and in parallel – which may hamper the value of such research efforts. The viability of DECAF can be improved as DC and L improvements are made. Therefore focusing effort on improvements to DCL specifically for DE (with an industry focus) is a priority area.

6.2.2.2 *A small-scale test deployment*

Following engagement between industry and scientists, and a decision made on whether to use an autonomous or cabled system, a small test array of sensors should be deployed (see 6.2.4). This would provide an assessment of localization performance for different target species, which is vital. The array would include p-nodes (alongside d-nodes) and could provide data to assess the following:

- The localization performance of the array, *i.e.* horizontal range error versus range could be tested with pingers of different frequencies and at different depths.
- The proportion of detections detected on different hydrophones, and the proportion of detections successfully localized could be easily measured.
- Algorithms for detection, classification and localization of different species could be evaluated.
- The simulation results on probability of detection versus the probability of localization possibly could be tested with experiments using DTAGs or similar tag sensors in the area in which the PAM array is to be deployed (with the intention of detected vocalizations of the tagged animal on the array).

The results from this small-scale, trial deployment could feed into a more realistic simulation to answer how many p-nodes are required to achieve an acceptable power to detect change for different species, and how this could be applied for nested arrays (for surveying multiple target species).

6.2.2.3 *PAM system assessment for use with DECAF*

Relating to the AVADECAF tool results, there would also be value in exploring a detailed analysis of the hardware, software, performance requirement and error budget for three localization cases: (i) harbor porpoises, (ii) beaked whales, and (iii) blue whales, with the objective of assessing the situations in which there are/are not ready-to-go solutions that achieve satisfactory localization performance. These test cases are chosen due the wide range in vocalization frequencies and beam patterns. From the 100Hz tones of blue whale to the highly directional 40-80 kHz beaked whale clicks and finally the ultra-high frequency 130 kHz clicks of harbor porpoises.

6.2.2.4 *Cost-effective localization*

Long-term fixed location PAM would benefit from the development of cheap portable systems capable of collecting both long-term acoustic data and data that can be used to estimate detectability (e.g. distance to detections). Portable systems specifically designed for DECAF-based methods would be an extremely useful hardware advance and may provide an option in areas where a fixed cabled installation is not suitable. Ideally all survey data from these devices would be retrievable during the course of the survey itself rather than only after the study is concluded. Such capabilities allows for quality assessment of the data and fault detection throughout the deployment (and not only following retrieval of equipment and analysis).

Currently, there are no ‘off the shelf’ seabed-mounted and cabled, wide baseline or vertical arrays. These would need to be developed and constructed if DECAF methods are to be effectively applied in a robust manner. Moretti et al. (2009) estimated a substantial cost for this (>$10 million per array). However, it is our view that there are multiple systems available which could be adapted for use as autonomous seabed-mounted systems and which would satisfy the data needs for applying DECAF methods. HARPs are perhaps an initial obvious choice as they have been used before. However, they are not commercially available and potentially, they could be prohibitively expensive to rent over the duration of a long-term deployment. Another potential solution is to use off the shelf, 4-channel, autonomous recorders (e.g. rtSYS, SoundTrap, JASCO AMAR, Decimus), develop a frame on which to mount them, and apparatus to deploy and recover in deep water; or a system to incorporate them into vertical arrays. Griffiths and Barlow (2015) describe a drifting buoy system constructed for approximately $5,000 which was based on a Wildlife Acoustics sound recorder. Based on these pre-existing recorders, an autonomous unit with the capabilities of a HARP and either a compact tetrahedral or a vertical array could conceivably be built for ~$20-30k per unit after development costs. This figure does not include deployment, servicing and analysis costs which would be substantial for any type of equipment (even fixed cabled arrays for which cleaning would need to be conducted periodically).

A detailed cost analysis is therefore required. Can autonomous systems continuously deployed and recovered com in substantially cheaper than a fixed system? This must consider both the likely continuous capability improvement and cost reduction of autonomous devices over the lifetime of a cabled system but also the disadvantages of potential equipment loss, damage and reduction in performance and abilities compared to a fully synchronized cabled array.

Whether using cabled or autonomous hardware, to make a localizing system efficient, there must be a high level of automation in the analysis of data. Employing multiple manual analysts and/or highly trained academics would be prohibitively expensive. Therefore, a full suite of appropriate and easy software should be developed and made openly available for all analysis tasks. Localization, as discussed in section 3.3, is species specific. Therefore different DCL algorithms are required, depending on the target species. Although there are many localization algorithms in the literature, often these have either only been tested on one species or not effectively tested at all other than in a few brief examples.

Therefore, an initial stage would be the development and deployment of a test array. The localization accuracy of the array could be tested by pinging from a known location and the effectiveness of different algorithms evaluated when applied to different marine mammal species. This would form an important initial assessment of the efficacy of autonomous units to localize and help iron out problems when a final system is decided upon.

In summary, we recommend a thorough exploration of the potential cost savings and performance of multi hydrophone autonomous systems to form the localizing and simple nodes of a large baseline array. The potential advantages of such as system as opposed to a fixed cabled array are mainly pragmatic. Autonomous arrays appear to be more cost-effective and it has been shown they can also be used to accurately localize vocalizing cetaceans. In addition they provide a more targeted, fault-tolerant and generic localization system, which, if developed with appropriate software tools could feed into the wider research community as a useful tool. By being available to the wider research community, this will serve to increase engagement in this research area and speed-up the development of systems and monitoring approaches specifically to address issues of using PAM for density estimation. However, there are existing sub-surface infrastructure (e.g. cabled power , transmission, command/control systems) in place for various marine industries. The existence and location of such infrastructure should be considered in planning any long-term PAM effort as it may provide a cost-effective foundation from which to build an effective monitoring program and address some of the issues raised above. Therefore, a logical next step is to explore in partnership with industry, whether autonomous systems can be deployed and recovered in a cost-effective manner and how this compares to the deployment and lifetime maintenance costs of a fixed cabled system.

6.2.2.5 *Planning a life of field DECAF effort*

Here we outline a best case research program to provide the foundation from which to develop an array for marine mammal density estimation. This serves to outline the complexity of the process and therefore the value in addressing the main research questions that remain obstacles to the viable implementation of DECAF on an E&P field. It is important, therefore, to acknowledge that implementing DECAF approaches for E&P life of field is not viable at present, without additional auxiliary efforts. The developments outlined here could make it viable.

At the initial planning phase, we recommend that the marine mammal species likely to be present at the E&P site are identified. Their suitability for PAM and for using acoustic tags to measure cue rates from individuals (based on past experience and consultation with experts) should be discussed. A key step would be decide the taxonomic level needed for analyses (*i.e.* can species be grouped together or is there a specific species which is considered a priority?). With this in mind, an assessment of available information (which might include previous abundance estimates, species distributions, previous acoustic tag data and habitat preferences etc.) should be made and this information used to design a series of field studies.

Such a research program could be developed to assess the following as part of year-long preliminary study:

- Measure environmental variables required to model acoustic propagation, and test the models with in-situ measurements.
- Conduct during each critical season, visual and acoustic vessel-based surveys, ocean sampling of features relevant for habitat modelling, and with acoustic tagging of appropriate species.
- Deploy a small number of bottom-mounted recorders (n-nodes (and potentially p-nodes)) for duration of the studies (or longer). This might include the use of portable (temporary) vertical line arrays to measure the detection rate of calls by tagged animals at different depths.

- Analysis of the data collected could estimate site-specific detection functions (to help plan the number of nodes and design of array), and cue rates by age/sex class and season.
- Use data from field study to design array for estimating changes in density for that site and for relevant species. This could also be integrated with the localization solutions presented in 6.3 and 6.4.

In addition, in order to explore potential effects of offshore activities (and better fit PAM arrays for density estimation to meeting E&P regulatory needs), the following could be investigated throughout the life time of an E&P development to determine a suitable PAM design to test for changes in density relative to offshore activities:

- Conduct sound data collection efforts before offshore activities are implemented to get a sense of seasonal and annual variability in density surface, along with relevant ocean parameters that may affect distribution.
- Record not just detections of animal calls but also measures of ambient sound levels from natural and anthropogenic sources. Include potential control sites with similar oceanography and marine mammals but where development is not planned (*i.e.* reference sites).
- Continue to record sound levels and animal detections during offshore activities. Make sure that the anthropogenic sound itself does not affect probability of detecting animal calls, and measure how cue rate varies when animals are exposed to noise.
- Continue to monitor sound levels and marine mammal vocalizations after operations have ceased to evaluate return to baseline (or densities predicted by ocean parameters).

7 Acknowledgements

We would like to thank IOGP for funding this work and in particular David Hedgeland for all of his guidance and feedback throughout the project. We'd also like to that thank Chris Pierpoint at Seiche Ltd. for providing a review of the report and the Project Steering Group from the Sound and Marine Life Joint Industry Programme for reviewing the report and helping improve it.

8 Appendices

8.1 Description of AVADECAF simulation tool

8.1.1 Generating the study area

Currently the tool only allows specifying rectangular study regions defined by four sets of x-y coordinates with the axes of the rectangle in a north-south and east-west orientation (example in Figure 22).). The x-y coordinates should be expressed in km (same unit as the truncation distance and distances to the nodes defined below). This region represents the area of interest over which a change in average densities shall be inferred. Then, a buffer distance is applied to the study region to define an area of analyses that includes the former as well as its surrounding buffer. This buffer distance is the width of the user-defined truncation distance (which represents the furthest distance at which the cues can be detected, see more details below).

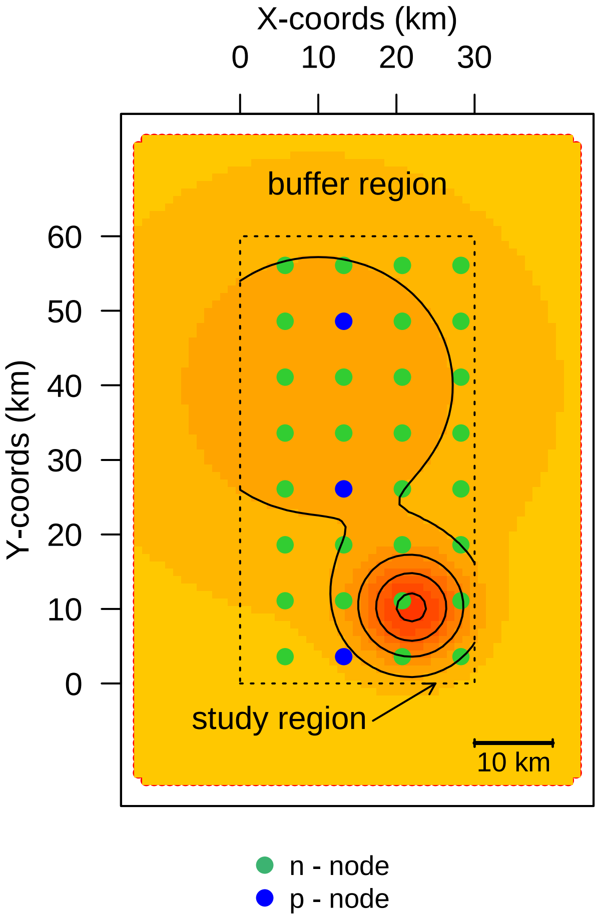


Figure 22 - Figure depicting possible user-selected inputs for Step 1 of the simulation. This figure represents a rectangular study area, including a 30 km x 60 km study region, a buffer region with truncation distance, w=12 km, spacing of n-node and p-nodes every 7.5 km, as well as the nested spacing of p-nodes set to every 3rd p-node in both X- and Y-coordinate directions. The generated PAM data shown here as a continuous heat map has x-space=0.25 km and y-space=0.25 km grid resolution. In the figure, regions of higher density are colored red, and regions of lower density in orange, contours depict the shape of the hotspot in cue density.

8.1.2 Generating the survey design

The survey design is defined by the spacing between all nodes in km and the number of n-nodes between the p-nodes in integers (e.g. nested space = 3 ensures that p-nodes are placed every 3rd node, Figure 22). The spacing between the nodes determines the total number of nodes placed into the region. The spacing between p-nodes determines how many of the nodes are capable of localization. This survey design is then used to generate the transects (i.e. nodes).

The user also needs to specify the scale parameter (*sigma*) of the half-normal detection function which, together with the truncation distance (*w*), defines the average detection probabilities *p*. Examples for a half-normal detection function are shown in Figure 23. The truncation distance is the furthest distance at which the cues produced by the animals of interest can be detected. For smaller values of *sigma* detection probabilities decrease more rapidly at closer distances. Several options are available for introducing heterogeneity in detection probabilities. The user can specify the same value for all seasons and years. However, heterogeneity between seasons and years can be introduced by specifying different values for the seasons and/or years which may reflect seasonal and/or annual differences in species presence, behavior and/or background noise. The user can also introduce heterogeneity between the different nodes by specifying a value for *sigma.cv* which indicates the amount of variability in the *sigma* and hence in the average detection probabilities *p* between nodes for a given season and year.

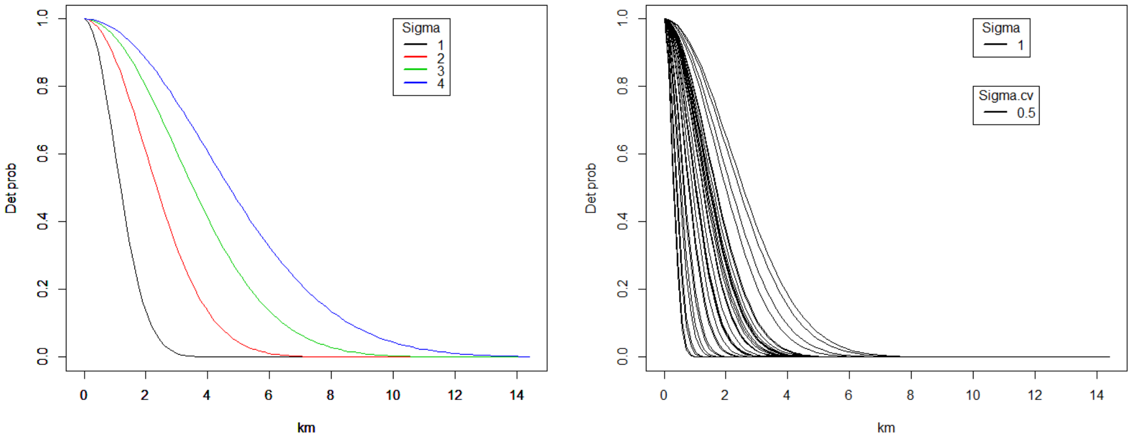


Figure 23 Left: four examples of a half-normal detection function, each with a different scale (*sigma*) parameter. The truncation distance is set within the simulation, and should be set at the furthest detection distance for a hydrophone. Right: 32 different half-normal detection functions with *sigma* = 1 and *sigma.cv* = 0.5.

As described in previous sections distances are measured with error by the p-nodes, and by incorporating error in PAM estimates of distances will introduce variability and bias to the system causing decreased accuracy in estimating animal densities and decreased statistical power to detect trends in densities. To incorporate this in the simulation tool the user can specify a systematic bias (*error*) and/or random error (*error.spread*) which is applied in the following manner: This source of variability was added as Gaussian error around the ‘true’ distance. The mean of the random error was the observed distance and mean squared error was the square of the observed distance

$$newDistance = |trueDistance \times (1 + b_{ran})|, b_{ran} \sim N(error, error.spread^2) \qquad (6)$$

In the case of no systematic or random error, the arguments *error* or *error.spread* are set to 0 in equation (6), respectively (or to *NULL* in the tool).

Other parameters pertaining to the survey are the number of survey years, the number of seasons surveyed each year, the number of days surveyed in each season (*days*) and the time (*time*) surveyed during each day (specified in seconds). The latter generally equals 24 * 60 * 60 but allows specifying smaller values e.g. due to duty cycling.

8.1.3 Generating the species of interest

The user is required to define several parameters including the average density (*d1*) in the first year and first season of the study, as well as information pertaining to the spatial (location and gradient of hotspots) and the temporal variability in densities (the seasonal fluctuations (*season*) in average densities and the annual change (referred to in the code as *decline* as we are interested in detecting a decline in densities) in densities over the study period). The annual change is the parameter of interest which we aim to retrieve by estimating the trend in densities over all seasons and years in step 2. The change can be specified as a fixed or random annual change and/or via a target change (e.g. 50%) arrived at after the duration of the study. Counts of cues at each node are generated from a negative binomial distribution to allow for extra-Poisson variability, typical of biological datasets. With the negative binomial distribution (unlike with the Poisson), the variance can be larger than the mean, and thus the variance inflation factor *n.cv* needs to be defined. This parameter reflects the proportion of overdispersion in the cue data, e.g. 1.2 represents a 20% over-dispersion in cue count. Note that when *n.cv* = 1, the negative binomial distribution reduces to a Poisson distribution when using distance sampling methods, i.e. for method = “DS”). Furthermore, we need to specify values for the parameters *c*, f and a_{perc} from equation (5), i.e. for the cue production rate, false positive rate and perception bias. For *c*, f and a_{perc}, CVs need to be specified which reflect the variability in the parameters appropriately. All these parameters are needed for generating the n_c from equation (5), respectively, for step 1. See below for generating distances to the cues which allows estimating parameter *p* in step 2.

8.1.4 Generating surfaces of animal densities

We begin by using the parameters *decline* (annual change in animal densities) and *season* to create the matrix *decline.mat* – a matrix containing the proportional changes in average densities for each year and season in relation to *d1*. We then use the average density from year 1, season 1 (*d1*) to construct a flat density surface for the region within the buffer region with the `DSsim` function `make.density()` which consists of a grid of density points, each denoted by an x-y-coordinate, and with a density value equal to *d1*. The spacing between density points in the surface is specified using *x.space* and *y.space* and should be relatively small. We then add hotspots to this surface until the spatial variability is a good representation of the real study area. For adding hotspots, we use `DSsim:make.hotspot()` where for each hotspot the user has to define the center location, the increase in density at this location as well as the steepness of the decay in densities towards the flat surface – smaller values result in steeper gradients. After adding the hotspots, we readjust the values of the density surface such that the mean of all density values in the region without the buffer equals *d1*. We use these adjusted density values in combination with the matrix *decline.mat* to generate animal density surfaces for all years and seasons at the simulation level. (We use these animal densities for generating cue counts for all nodes and distances for the p-nodes at the season level). For generating animal density surfaces for all years and seasons, the user has two options for what type of impact will be applied to the densities which are described in the following two subsections 8.1.5.1 - 8.1.5.2. Both options result in a data array (*all.densities*) which contains the density surfaces for each year and each season and remains constant for all iterations of the simulation. The locations of the density points in each of the density surfaces are the same and obtained from the density object created with the `DSsim` functions.

8.1.4.1 *Equal proportion change*

There are two methods to define an overall annual change in density (parameter *decline*). This is the simpler of the two options. Here, we assume an equal proportion change throughout the study region. Note, that as the interest is generally to detect a decline, the change parameter is called *decline* in the R code. Implied in this notation is that, e.g., a 5% decline in animal densities is defined as *decline* = 0.05, whereas a 5% increase in animal densities would be defined as a negative value, i.e. *decline* = -0.05. We multiply the matrix *decline.mat* with the vector of adjusted animal densities from year 1, season 1 and obtain *all.densities*. This ensures an exponential change in population size over time.

8.1.4.2 *Random change*

The alternate way to incorporate an annual change in density is through a random change – we referred to this earlier as "process error". Here, we generate from a lognormal distribution and adjust the mean of that distribution such that it equals some overall rate of change set by the user. We multiply the matrix *decline.mat* with the vector of random change rates, and then multiplied by the adjusted animal densities from year 1, season 1 to obtain *all.densities*. This ensures a random exponentially-distributed change in population size over time.

8.1.4.3 *Impact change*

This option allows manipulating the way the annual change affects the density surface. For implementing an impact change we assume that the change in densities is the strongest in the center of the study region with a decay of the impact towards the outer areas of the study region. This impact increases each year while densities are readjusted so that average densities across the entire region without the buffer are equal to those from the equal proportion change (Section 8.1.5.1). The impact change pattern can be added to an equal proportion change or a random change.

8.1.5 Generating node placement using the survey design

Generating the placement of the n-node and p-nodes in the study region, occurs anew at each iteration using `DSsim:generate.transects`; hence, the location, the exact number and the ratio of p-nodes vs n-nodes may vary between iterations. We use the region without the buffer for this purpose which ensures that each node will be located inside the study region while the search radius around each node is fully covered by the density surface.

8.1.6 Generating the PAM_{dist} data using distance sampling methods.

As we aim to produce an estimate of animal density for each season and year, we generate a new PAM_{dist} data set at the lowest level, the season, using the parameters and objects created in the previous Sections. The PAM_{dist} data set consists of the detected number of cues for each node as well as the observed distances from each p-node. To generate these, we need to use the array *all.densities* generated in Section 8.3.4 as well as the detection function and detection probability *p*, cue production rate *c* (and CV *c.cv*), false positive rate f (and CV *f.cv*), perception bias a_{perc} (and CV *a.cv*) and variance inflation factor *n.cv*. To incorporate the variability described for cue production rate, false positive rate and perception bias we draw new random samples for each season and year for these parameters using a gamma or beta distribution: *c.sim ~ Gamma(c, c.cv)*, *f.sim ~ Beta(f, f.cv), a.sim ~ Beta(a, a.cv)*. To generate the number of detected cues, n_c, and distances using these, the user has two options as described below.

8.1.6.1 *Using point densities*

This is the simpler, however, less realistic approach for generating numbers of cues and distances. For each node we identify the closest point in the density surface and multiply the respective density values with the search area (defined by truncation distance *w* around the node) to obtain the expected number of animals within the search area. After multiplication with the survey time (*T* = *days* x *time*), the detection probability *p* and parameters *c.sim*, *a.sim* and division by (1-*f.sim*) we obtain the expected number of observed cues for each node. We then draw random samples from a negative binomial distribution with mean equal to the expected number of observed cues and the variance inflation factor defined above (Section 8.3.3). This is done separately for each node. The random samples are the number of detected cues n_c.

In the next step, we generate n_c distances assuming that we can localize (estimate distance to) all cues out to a truncation distance. In Section 5.2.4 we discuss a simulation scenario in which the p-node can only estimate distance out to a specified distance beyond which a p-node becomes a n-node (still counting cues, but without the ability to estimate distance for that cue). This latter scenario aims to replicate some of the challenges associated with localization as discussed in Section 3.3. Here we sample from the probability density function of observed distances for point transects using a half-normal detection function (Buckland et al. 2001b). Although this method does not allow specifying a truncation distance *w*, we sample until we have generated n_c distances with distance equal to or less than *w*.

In the case that a systematic or random distance measurement error was specified, the generated (true) distances are converted into observed distances with error using equation (6).

8.1.6.2 *Using grid-based distances*

Compared to using point densities, using grid-based distances is more complex and more time-consuming when running the simulation but more realistic as it captures non-uniform densities within the search radius. This approach uses a x-space by y-space resolution grid containing all the density points from the surface within a circle around the nodes with radius equal to the truncation distance, w. Which density points fall within the search areas as well as their distances to the nodes is determined using the x-y-coordinates from the density surface and the nodes. The area for each grid cell is determined by the *x*- and *y-space* that the user defined when creating the density surface. The expected number of animals in each grid cell is obtained by multiplying the area of the grid cell with the density value. After multiplication with the survey time (*days* and *time*), the detection probability *p* and parameters *c.sim*, *a.sim* and division by (1-*f.sim*) we obtain the expected number of observed cues for each grid cell within the search radius around each node. We use the expected number of observed cues as the mean and the variance inflation factor to draw random samples from a negative binomial for each grid cell which represent the observed number of cues. The distances for these cues are given by the distance of the density points to the nodes. Hence, to avoid rounding issues, we recommend specifying relatively small values for *x.space* and *y.space*. (In Figure 10, we have selected *x-space* and *y-space* values of 0.25 km).

In the case that a systematic or random distance measurement error was specified, the generated (true) distances are converted into observed distances with error using equation (6).

8.1.7 Generating the $\mathrm{PAM}_{\mathrm{SECR}}$ data using SECR methods.

As for distance sampling methods, we aim to produce an estimate of animal density for each season and year and, hence, generate a new $\mathrm{PAM}_{\mathrm{SECR}}$ data set at the lowest level, the season, using the parameters and objects created in the previous Sections. The $\mathrm{PAM}_{\mathrm{SECR}}$ data set consists of a capture history of detected cues and at which nodes each of them was detected at. To generate these, we need to use the array *all.densities* generated in Section 8.3.4 as well as the detection function cue production rate *c* (and CV *c.cv*), false positive rate f (and CV *f.cv*), perception bias a_{perc} (and CV *a.cv*) and variance inflation factor *n.cv* which were defined in Section 8.3.3. To incorporate the variability for cue production rate, false positive rate and perception bias we draw new random samples for each season and year for these parameters using a gamma or beta distribution: *c.sim ~ Gamma(c, c.cv)*, *f.sim ~ Beta(f, f.cv), a.sim ~ Beta(a, a.cv)*. However, we note that when using SECR methods, the default values for a_{perc} and *a.cv* are fixed at 1 and 0, respectively; these default values were used in the simulations (though other values could be specified, if required by the user). To generate the detections, we make use of various simulation functions from the *secr* package (Efford and Fewster 2013). The current version of the tool does not implement extensions to SECR methods (i.e., making additional use of bearing or time difference of arrival information). Furthermore, introducing variability between each node's detection function using covariates is also not possible when using SECR (in contrast to using a distance sampling approach). This is an area that should be considered in future development of the tool.

8.1.8 Estimate animal density using distance sampling methods

Estimating animal density from PAM_{dist} data requires estimating the various contributing components and their variances including the encounter rate, the probability of detection, the cue production rate, the false positive rate and the perception bias (see equation (6)). We use the detected cues to estimate encounter rate (*exp.e.n*) and its variance. The encounter rate of cues is given by the total number of observed cues n_c (generated during the simulation, see above) divided by the total search area ($k\pi w^2$) and the time surveyed (*T* = *days* x *time*). Encounter rate variance is estimated using the P1-estimator from Fewster et al. (2009). For estimating the average detection probability within the search area (p, is coded as *average.p*) and its variance we fit a half-normal detection function to the observed distances using the function `ddf` of the `mrds` package.

Perception bias $\mathrm{a_{perc}}$ represents the proportion of cues that were produced by the animals that were detected directly over the node. While this parameter can be set to 1 for a range of species (e.g. when species produce vocalizations which are omnidirectional such as humpback whales and when the depth difference between the animal and the receiver is relatively small), it can be particularly difficult to estimate in other cases; see for example Marques et al. (2009) who model the probability of detecting a click at a sensor as a function of the orientation of the animal in relation to the sensor using horizontal and vertical off-angle (in addition to the sensor). However, in any real-world study, estimating this parameter requires additional data to be collected and complex analyses which are outside the scope of this simulation tool. Hence, to reflect that there is uncertainty in estimating this parameter in a real-study – as opposed to knowing the truth – we take a sample *a.x* from a beta distribution bound by 0 and 1 with mean equal to *a.sim* and CV equal to *a.cv*.

Estimating parameters c and f is described in more detail in the following sections.

8.1.8.1 *Simulating the estimation of cue production rate c*

Cue production rate is the average rate at which animals of the species of interest produce cues and is often estimated using data from sound recording tags in a real-world study (see Section 3.1). These are tags that are attached directly to animals, generally via a suction cup mechanism, and remain attached for some time while they continuously record several parameters as well as sound. The sound data allows counting up all cues produced by the animal that the tag was attached to. To estimate cue production rate from sound tag data, the total number of cues produced by the individual are divided by the length of the recording. To obtain an unbiased estimate of cue production rate for the primary acoustic survey, tags should be attached to multiple individuals and at different times throughout the survey. Due to the potential biases described in Section 3.1, more tags applied will improve the estimate of cue production rate, as well as the estimate of its CV.

To simulate the process of estimating the cue production rate from DTAG data, the user of the tool specifies the number of DTAGs (*trials*) used for a particular year and season. We then generate an observation of *c.sim* (a randomly drawn gamma-distributed sample of cue production rate for a given year and season), as this represents 'truth' for this parameter during the respective year and season, as well as a randomly generated observation of *c.sim.cv* (a randomly drawn chi-square distributed sample of the CV for cue production). These samples are then used to generate the gamma distributed random realizations for cue production rates in each of the different trials (e.g., if there were 8 DTAGs on 8 whales, 8 gamma distributed random realizations were generated from a distribution with an underlying mean and CV of c.sim, and c.sim.cv). The mean of the samples represents the estimated cue production rate for that year and season. With this process we acknowledge that while there is a true underlying average cue production rate for this year and season, there will be variability between different animals and different recording times within the season. The generating process is unbiased for the mean estimation, however, the more DTAGs deployed (i.e. the higher the number of *trials*), the closer the estimate is likely to be to the truth.

8.1.8.2 *Simulating the estimation of false positive rate f*

For estimating the false positive rate in a real life situation (Section 3.2 above), sections of the acoustic recordings are proofed for false positives. As this is generally done by a human observer, these sections often only represent a small fraction of the total length of acoustic recordings. To reduce bias, the total effort of proofing is divided into sections throughout the recordings.

To simulate this process, we generate numbers of observed cues from trials of acoustic recordings of certain length (representing the sections that the human observer would proof). The user specifies the number of trials as well as the length of the trials which, in combination, represent the total effort spent to estimate the false positive rate from existing cue count data (Figure 24). The more effort spent, i.e. using an increasing number of trials and/or length of trials, the closer the estimate for f will be to the truth. The number of true positives *csim.gamma* for each trial are generated using *c.sim* (the randomly drawn sample of cue production rate for a given year and season) and *c.cv* (a predefined CV for cue production rate). Then we create the total number of observed cues *obs.c* for each trial which consists of both the true and false positives, where *obs.c* = *csim.gamma* + *f.sim*obs.c* or *obs.c* = *csim.gamma/(1-f.sim)*. We use *f.sim*, the randomly drawn sample of false positive rate for a given year and season, as for this year and season it represents truth – as opposed to f (the predefined parameter value for false positive rate).

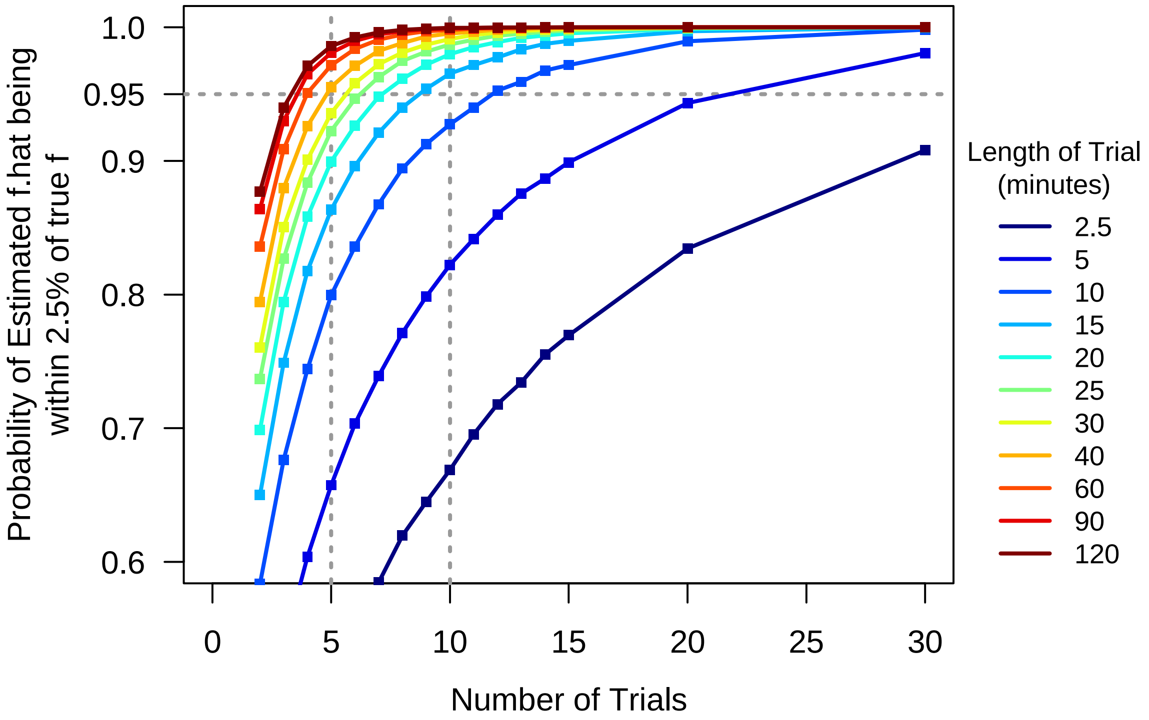


Figure 24 - This plots the number of trials (x-axis) vs the probability of estimating f.hat being within 2.5% of true false positive rate, f. Each line represents the length of trial in minutes that a human observer would check a cue series for. A horizontal grey line is drawn at 95%, and two vertical lines are drawn at 5 trials and 10 trials. The figure shows that a minimum of 5-trials of 40 minutes long would be required to be 95% sure the estimate of f, is within 2.5% of the 'true' f, 19 times out of 20 (95% of the time). It shows that a minimum of 10 trials of 15 minutes long would be required to be 95% sure of being within 2.5% of the 'true' f, 19 times out of 20. (etc.).

We generate false positive observations *f.hat* using a beta-binomial distribution where the probability of success is a random variable with a gamma distribution defined by mean equal to *f.sim* (and CV = *f.cv*, the predefined CV for false positive rate). The mean of the ratios *f.hat/obs.c* over all trials is the estimate of f for a given year and season.

8.1.9 Inclusion of fixed PAM localization errors in the simulation.

Localization is typically only possible over a proportion of the range that cues can be detected. Therefore, we set up a simple sensitivity scenario, to look at the effect of systematic bias and random errors, over a subset of cues for which distances could be estimated, beyond which the simulation records only presence. The localization errors are associated with fitting the detection function at the p-nodes, and this sensitivity is only a quick attempt at implementing a simple approach of many other more complicated ways that these errors with localization might affect the estimation of density. By assuming that localization is possible at all distances, we set "adapted.for.localisation.range=F" and distance estimates are attained for all cues out to the truncation distance (w=12 km), and if error and error.spread are also both set to NULL, then those distances are estimated without any error. This sensitivity then sets a maximum localization distance (max.loc.distance), beyond which the p-nodes detect vocalizations but they can only be lumped into one distance bin -- from max.loc.distance to truncation distance "w"). To implement this sensitivity, there is a "num.bins" object, and this is required because the detection function has to be discretised into a step function over localization distances, e.g., num.bins=6 means there are 5 bins within max.loc.distance, plus 1 for the bin that includes all detections greater than max.loc.distance. Random error and systematic bias to distance estimation is added as already discussed.

8.1.10 Estimate animal density using SECR methods

Estimating animal density from PAM_{SECR} data requires estimating the various contributing components including density of cues, the probability of detection, the cue production rate, the false positive rate and the perception bias. We use the `secr.fit` function to estimate density of cues in the region from the capture history. This function also estimates the average detection probability p and the perception bias a_{perc}. Currently, the tool does not allow calculating the percent contribution of each component to the overall variance estimate, though a CV for the density estimate can be estimated from the output of the tool. However this is critical for strategies to reduce error and should be considered in future developments.

Estimating parameters, c and f is done as outlined above. These estimates are used to convert the estimated cue density into density of animals.

8.2 Simulations

8.2.1 Baseline description

For PAM survey specific parameters we decided to use a study area of 60 x 30 km. While we acknowledge that this is likely to be larger than a typical area of E&P development, with respect to the effects of noise, the potential impact footprints are much larger than the E&P area itself. Under the presumption that some E&P activities have the potential to displace animals, monitoring an area that is too small may result in a total reduction of animal densities across the area but no information about over what ranges an activity is causing an impact. Therefore we have erred on the side of caution by running simulations for a larger study area (which is likely to capture any 'holes' or hotspots in density surface created naturally or by an activity. To ensure this does not bias the tool and its outputs, the size of the study area is one of the variables we explored relative to changing power of PAM surveys in section 5.

We ran the baseline survey with sensors spaced 7.5 km apart (resulting in a total of 32 sensors spread across our study area) and a *nested.space* value of 3 (which defines from the grid of sensors, how far apart those with localization capabilities are, see section 8.1.3). This resulted in a typical survey with 32 detection-only 'n-nodes' and ~3 'p-nodes' capable of localization (this is approximate as it varied across each loop of the simulation). We had each sensor collecting data throughout each year with a duty cycle of 100%.

In analyzing the simulated PAM survey data, we decided to generate four density estimates in each year of a 10-year survey. This appeared to be a reasonable number of years to run a survey over; any increase in survey length would only increase the power of the survey.

The Marques et al. (2009) study summarizes the analytical methodology and auxiliary data sources used in their analyses. The baseline parameters that were varied across scenarios (section 5) are presented in Table 8 below. The species of interest in the Marques et al. (2009) study was Blainville's beaked whale (*Mesoplodon densirostris*) and the animal density estimated in their DECAF analysis was 25.3 animals / km^2.This value was used here as the starting value for density of animals for the simulations, i.e. animal densities in the first year of the ten year survey period. In Marques et al., cue (vocalization) rate was estimated from five animals equipped with DTAGs. Their estimated cue rate was 0.407 clicks per second and a variability value (CV) of 9.8%. Via a comprehensive analysis of the data collected on the AUTEC hydrophones the false positive rate was estimated to be 0.549 (CV = 1.99%). In order to incorporate the effects of perception bias into the simulations, following discussion among the project team, a beaked whale perception bias value of 0.07 (CV = 10%) was chosen. These estimates were influenced by the narrow beamwidth of the highly directional beaked whale clicks meaning that there was a high probability that animals could be in the study area, within the immediate vicinity, and hence, well within the detection distance of a sensor, without being detected while producing clicks (e.g. because they are not directly pointing towards the sensor in the three-dimensional space, e.g. by facing away from the sensor or and their narrow beam of clicks was not detected).

The other parameters selected in the baseline scenarios are shown in Table 8 – and their importance explored in section 5.2.

Table 8 - Baseline parameters explored in subsequent analyses.

Parameter	Values	Parameter	Values
n.seas	4	error	10
season	c(1,1,1,1)	error.spread	0.1
days	c(rep(1,90),rep(2,92), rep(3,92),rep(4,91))	n.cv	1.2
spacing	7.5	nested.space	3
c	0.407	impact.grid	F
c.cv	0.098	grid.based distances	F
c.bias	0	mean density year1 seas1	0.0253
a	0.07	sub.number.of.cues	5000
a.cv	0.1	adapted.for.localisation.range	F
a.bias	0	w	12
f	0.549	max.loc.distance	w
f.cv	0.0199	num.bins	Inf
f.bias	0	time	24*60*60
sigma	2		
sigma.cv	0.1		

8.2.2 Assessing use of ‘grid-based-distances’ and ’impact grid’ functions

We assessed what effect the use of *grid.based.distances* and *impact.grid* had on power in simulations. We observed that there was no material change in simulation power (as all changes were driven by changes in sigma.cv and nested.space (as observed in section 5)(Figure 25). However setting these elements to be TRUE (i.e. so that they were used in the simulations) allows for different levels of realism to be simulated. For example, using the gridbased method or centering the impact change in the middle of the region. Using the grid based approach did result in longer run times than when it was set to FALSE.

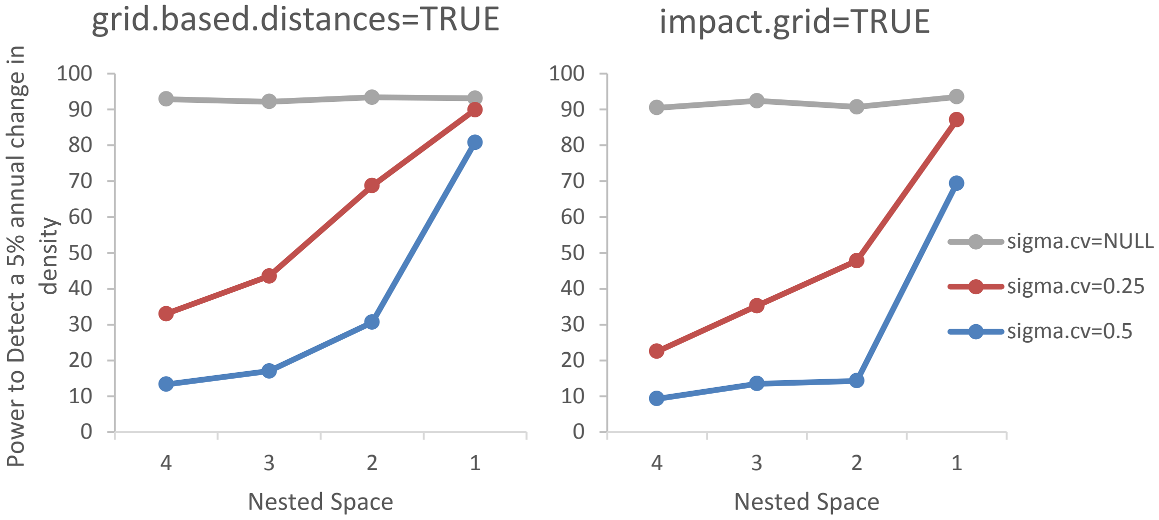


Figure 25 - How varying the number of localizing PAM units (the nested space) changes the power to detect a 5% annual change in density for varying values of precision for the detection function estimate (*sigma.cv*), when distances and counts were created using the gridbased method (*grid.based.distances*=TRUE) and when the change was centered in the middle of region (*impact.grid*=TRUE).

9 References

Abbot, T. A., V. E. Premus, and P. A. Abbot. 2010. A real-time method for autonomous passive acoustic detection-classification of humpback whales. Journal of the Acoustical Society of America **127**:2894-2903.

Anderson, D. R. 2001. The need to get the basics right in wildlife field studies. Wildlife Society Bulletin, 29, 1294-1297.

Aguilar de Soto, N., P. T. Madsen, P. Tyack, P. Arranz, J. Marrero, A. Fais, E. Revelli, and M. Johnson. 2012. No shallow talk: Cryptic strategy in the vocal communication of Blainville's beaked whales. Marine Mammal Science **28**:E75-E92.

Aguilar Soto, N., M. P. Johnson, P. T. Madsen, F. Díaz, I. Domínguez, A. Brito, and P. Tyack. 2008. Cheetahs of the deep sea: deep foraging sprints in short-finned pilot whales off Tenerife (Canary Islands). Journal of Animal Ecology **77**:936-947.

Ahrestani, F. S., M. Hebblewhite, and E. Post. 2013. The importance of observation versus process error in analyses of global ungulate populations. Scientific Reports **3**:3125.

Akamatsu, T., Y. Hatakeyama, T. Kojima, and H. Soeda. 1994. Echolocation rates of two harbor porpoises (Phocoena phocoena). Marine Mammal Science **10**:401-411.

Akamatsu, T., J. Teilmann, L. A. Miller, J. Tougaard, R. Dietz, D. Wang, K. Wang, U. Siebert, and Y. Naito. 2007. Comparison of echolocation behaviour between coastal and riverine porpoises. Deep Sea Research Part II: Topical Studies in Oceanography **54**:290-297.

Akamatsu, T., D. Wang, K. Wang, and Y. Naito. 2005. Biosonar behaviour of free-ranging porpoises. Proceedings of the Royal Society of London B: Biological Sciences **272**:797-801.

Arranz, P., N. A. de Soto, P. T. Madsen, A. Brito, F. Bordes, and M. P. Johnson. 2011. Following a Foraging Fish-Finder: Diel Habitat Use of Blainville's Beaked Whales Revealed by Echolocation. PLoS One **6**.

Arranz, P., S. DeRuiter, A. Stimpert, S. Neves, A. Friedlaender, J. Goldbogen, F. Visser, J. Calambokidis, B. Southall, and P. Tyack. 2016. Discrimination of fast click-series produced by tagged Risso's dolphins (Grampus griseus) for echolocation or communication. Journal of Experimental Biology **219**:2898-2907.

Au, W. W. 2012. The sonar of dolphins. Springer Science & Business Media.

Au, W. W., B. Branstetter, P. W. Moore, and J. J. Finneran. 2012. Dolphin biosonar signals measured at extreme off-axis angles: Insights to sound propagation in the head. Journal of the Acoustical Society of America **132**:1199-1206.

Au, W. W., J. K. Ford, J. K. Horne, and K. A. N. Allman. 2004. Echolocation signals of free-ranging killer whales (Orcinus orca) and modeling of foraging for chinook salmon (Oncorhynchus tshawytscha). The Journal of the Acoustical Society of America **115**:901-909.

Au, W. W., and M. C. Hastings. 2008. Principles of marine bioacoustics. Springer.

Au, W. W., R. A. Kastelein, K. J. Benoit-Bird, T. W. Cranford, and M. F. McKenna. 2006a. Acoustic radiation from the head of echolocating harbor porpoises (Phocoena phocoena). Journal of Experimental Biology **209**:2726-2733.

Au, W. W., R. A. Kastelein, T. Rippe, and N. M. Schooneman. 1999. Transmission beam pattern and echolocation signals of a harbor porpoise (*Phocoena phocoena*). The Journal of the Acoustical Society of America **106**:3699-3705.

Au, W. W., A. A. Pack, M. O. Lammers, L. M. Herman, M. H. Deakos, and K. Andrews. 2006b. Acoustic properties of humpback whale songs. The Journal of the Acoustical Society of America **120**:1103-1110.

Au, W. W., J. L. Pawloski, P. E. Nachtigall, M. Blonz, and R. C. Gisner. 1995. Echolocation signals and transmission beam pattern of a false killer whale (Pseudorca crassidens). The Journal of the Acoustical Society of America **98**:51-59.

Aubauer, R., M. O. Lammers, and W. W. Au. 2000. One-hydrophone method of estimating distance and depth of phonating dolphins in shallow water. The Journal of the Acoustical Society of America **107**:2744-2749.

Baggenstoss, P. M. 2011a. An algorithm for the localization of multiple interfering sperm whales using multi-sensor time difference of arrival. The Journal of the Acoustical Society of America **130**:102-112.

Baggenstoss, P. M. 2011b. Separation of sperm whale click-trains for multipath rejection. The Journal of the Acoustical Society of America **129**:3598-3609.

Baggenstoss, P. M. 2013a. Processing advances for localization of beaked whales using time difference of arrival. Journal of the Acoustical Society of America **133**:4065-4076.

Baggenstoss, P. M. 2013b. Processing advances for localization of beaked whales using time difference of arrival. The Journal of the Acoustical Society of America **133**:4065-4076.

Baggeroer, A. B., W. Kuperman, and H. Schmidt. 1988. Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem. The Journal of the Acoustical Society of America **83**:571-587.

Baird, R. W., M. B. Hanson, and L. M. Dill. 2005. Factors influencing the diving behaviour of fish-eating killer whales: sex differences and diel and interannual variation in diving rates. Canadian journal of zoology **83**:257-267.

Barlow, J., and B. L. Taylor. 2005. Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey. Marine Mammal Science **21**:429-445.

Barlow, J., P. L. Tyack, M. P. Johnson, R. W. Baird, G. S. Schorr, R. D. Andrews, and N. Aguilar de Soto. 2013. Trackline and point detection probabilities for acoustic surveys of Cuvier's and Blainville's beaked whales. Journal of the Acoustical Society of America **134**:2486-2496.

Bassett, C., B. Polagye, M. Holt, and J. Thomson. 2012. A vessel noise budget for Admiralty Inlet, Puget Sound, Washington (USA). The Journal of the Acoustical Society of America **132**:3706-3719.

Bassett, C., J. Thomson, P. H. Dahl, and B. Polagye. 2014. Flow-noise and turbulence in two tidal channels. The Journal of the Acoustical Society of America **135**:1764-1774.

Baumann-Pickering, S., S. M. Wiggins, E. H. Roth, M. A. Roch, H. U. Schnitzler, and J. A. Hildebrand. 2010. Echolocation signals of a beaked whale at Palmyra Atoll. Journal of the Acoustical Society of America **127**:3790-3799.

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.

Baumgartner, M. F., D. M. Fratantoni, T. P. Hurst, M. W. Brown, T. V. Cole, S. M. Van Parijs, and M. Johnson. 2013. Real-time reporting of baleen whale passive acoustic detections from ocean gliders. Journal of the Acoustical Society of America **134**:1814-1823.

Baumgartner, M. F., and S. E. Mussoline. 2011. A generalized baleen whale call detection and classification system. The Journal of the Acoustical Society of America **129**:2889-2902.

Benoit-Bird, K. J., B. L. Southall, and M. A. Moline. 2016. Predator-guided sampling reveals biotic structure in the bathypelagic. Page 20152457 *in* Proc. R. Soc. B. The Royal Society.

Blackwell, S. B., C. S. Nations, T. L. McDonald, C. R. Greene, A. M. Thode, M. Guerra, and A. Michael Macrander. 2013. Effects of airgun sounds on bowhead whale calling rates in the Alaskan Beaufort Sea. Marine Mammal Science **29**:E342-E365.

Blackwell, S. B., C. S. Nations, T. L. McDonald, A. M. Thode, D. Mathias, K. H. Kim, C. R. Greene Jr, and A. M. Macrander. 2015. Effects of Airgun Sounds on Bowhead Whale Calling Rates: Evidence for Two Behavioral Thresholds. PLoS One **10**:e0125720.

Blackwell, S. B., W. Richardson, C. Greene Jr, and B. Streever. 2007. Bowhead whale (Balaena mysticetus) migration and calling behaviour in the Alaskan Beaufort Sea, Autumn 2001-04: An acoustic localization study. Arctic:255-270.

Borchers, D. 2012. A non-technical overview of spatially explicit capture–recapture models. Journal of Ornithology **152**:435-444.

Borchers, D., B. Stevenson, D. Kidney, L. Thomas, and T. Marques. 2015. A unifying model for capture–recapture and distance sampling surveys of wildlife populations. Journal of the American Statistical Association **110**:195-204.

Borchers, D. L., S. T. Buckland, and W. Zucchini. 2002. Estimating animal abundance: closed populations. Springer Science & Business Media.

Bradshaw, C. J., H. F. Mollet, and M. G. Meekan. 2007. Inferring population trends for the world's largest fish from mark–recapture estimates of survival. Journal of Animal Ecology **76**:480-489.

Branstetter, B. K., J. J. Finneran, E. A. Fletcher, B. C. Weisman, and S. H. Ridgway. 2012. Dolphins can maintain vigilant behavior through echolocation for 15 days without interruption or cognitive impairment. PLoS One **7**:e47478.

Buckland, S. T. 2006. Point-transect surveys for songbirds: Robust methodologies. Auk **123**:345-357.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001a. Introduction to distance sampling estimating abundance of biological populations.

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

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2004. Advanced Distance Sampling. Estimating abundance of biological populations. Oxford University Press, Oxford.

Buckland, S. T., E. A. Rexstad, T. A. Marques, and C. S. Oedekoven. 2015. Distance sampling: methods and applications. Springer.

Burtenshaw, J. C., E. M. Oleson, J. A. Hildebrand, M. A. McDonald, R. K. Andrew, B. M. Howe, and J. A. Mercer. 2004. Acoustic and satellite remote sensing of blue whale seasonality and habitat in the Northeast Pacific. Deep Sea Research Part II: Topical Studies in Oceanography **51**:967-986.

Caillat, M., L. Thomas, and D. Gillespie. 2013. The effects of acoustic miscassification on cetacean species abundance estimation. Journal of the Acoustical Society of America **134**:2469-2476.

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.

Cerchio, S., B. Andrianantenaina, A. Lindsay, M. Rekdahl, N. Andrianarivelo, and T. Rasoloarijao. 2015. Omura's whales (Balaenoptera omurai) off northwest Madagascar: ecology, behaviour and conservation needs. Royal Society Open Science **2**:150301.

Cerchio, S., J. K. Jacobsen, and T. F. Norris. 2001. Temporal and geographical variation in songs of humpback whales, Megaptera novaeangliae: synchronous change in Hawaiian and Mexican breeding assemblages. Animal Behaviour **62**:313-329.

Cheney, B., R. Corkrey, J. W. Durban, K. Grellier, P. S. Hammond, V. Islas-Villanueva, V. M. Janik, S. M. Lusseau, K. M. Parsons, and N. J. Quick. 2014. Long-term trends in the use of a protected area by small cetaceans in relation to changes in population status. Global Ecology and Conservation.

Clark, C. W., and J. H. Johnson. 1984. The sounds of the bowhead whale, Balaena mysticetus, during the spring migrations of 1979 and 1980. Canadian journal of zoology **62**:1436-1441.

Constantine, R., N. A. Soto, and D. M. Johnson. 2012. Sharing the Waters: Minimising Ship Collisions with Bryde's Whales in the Hauraki Gulf. publisher not identified.

Cox, M. J., D. L. Borchers, D. A. Demer, G. R. Cutter, and A. S. Brierley. 2011. Estimating the density of Antarctic krill (Euphausia superba) from multi-beam echo-sounder observations using distance sampling methods. Journal of the Royal Statistical Society: Series C (Applied Statistics) **60**:301-316.

Croll, D. A., C. W. Clark, A. Acevedo, B. Tershy, S. Flores, J. Gedamke, and J. Urban. 2002. Bioacoustics: Only male fin whales sing loud songs. Nature **417**:809-809.

Curé, C., S. Isojunno, F. Visser, P. J. Wensveen, L. D. Sivle, P. H. Kvadsheim, F.-P. A. Lam, and P. J. Miller. 2016. Biological significance of sperm whale responses to sonar: comparison with anti-predator responses. Endangered Species Research **31**:89-102.

D'Vincent, C. G., R. M. Nilson, and R. E. Hanna. 1985. Vocalization and coordinated feeding behavior of the humpback whale in southeastern Alaska. Scientific Reports of the Whales Research Institute **36**:41-47.

Dawson, S. M. 1991. Clicks and communication: the behavioural and social contexts of Hector's dolphin vocalizations. Ethology **88**:265-276.

Deecke, V. B., J. K. Ford, and P. J. Slater. 2005. The vocal behaviour of mammal-eating killer whales: communicating with costly calls. Animal Behaviour **69**:395-405.

Di Iorio, L., and C. W. Clark. 2010. Exposure to seismic survey alters blue whale acoustic communication. Biology letters **6**:51-54.

di Sciara, G. N., and J. Gordon. 1997. Bioacoustics: a tool for the conservation of cetaceans in the Mediterranean Sea. Marine & Freshwater Behaviour & Phy **30**:125-146.

DiMarzio, N., D. Moretti, J. Ward, R. Morrissey, S. Jarvis, A. M. Izzi, M. Johnson, P. Tyack, and A. Hansen. 2008. Passive acoustic measurement of dive vocal behavior and group size of Blainville's beaked whale (Mesoplodon densirostris) in the Tongue of the Ocean (TOTO). Canadian Acoustics **36**:166-172.

Duncan, A. J., and A. L. Maggi. 2006. A consistent, user friendly interface for running a variety of underwater acoustic propagation codes. Pages 471-477 *in* Proceedings of ACOUSTICS.

Dunlop, R. A., D. H. Cato, and M. J. Noad. 2010. Your attention please: increasing ambient noise levels elicits a change in communication behaviour in humpback whales (*Megaptera novaeangliae*). Proceedings of the Royal Society B-Biological Sciences **277**:2521-2529.

Dunlop, R. A., D. H. Cato, M. J. Noad, and D. M. Stokes. 2013. Source levels of social sounds in migrating humpback whales (Megaptera novaeangliae). The Journal of the Acoustical Society of America **134**:706-714.

Dunn, R. A., and O. Hernandez. 2009. Tracking blue whales in the eastern tropical Pacific with an ocean-bottom seismometer and hydrophone array. The Journal of the Acoustical Society of America **126**:1084-1094.

Dutton, D. L., P. H. Dutton, M. Chaloupka, and R. H. Boulon. 2005. Increase of a Caribbean leatherback turtle Dermochelys coriacea nesting population linked to long-term nest protection. Biological Conservation **126**:186-194.

Efford, M. G., and R. M. Fewster. 2013. Estimating population size by spatially explicit capture–recapture. Oikos **122**:918-928.

Fais, A. 2014. Acoustic behaviour of sperm whales (*Physeter macrocephalus*): applications in ecology and population studies. La Laguna, Tenerife, Spain, La Laguna.

Fais, A., T. Lewis, D. Zitterbart, O. Alvarez, A. Tejedor, and N. Aguilar Soto. 2015. Abundance and distribution of sperm whales in the Canary Islands: are ship-strikes sustainable? PLoS One **Submitted**.

Fewster, R. M., S. T. Buckland, K. P. Burnham, D. L. Borchers, P. E. Jupp, J. L. Laake, and L. Thomas. 2009. Estimating the encounter rate of variance in distance sampling. Biometrics **65**:225-236.

Finneran, J. J., B. K. Branstetter, D. S. Houser, P. W. Moore, J. Mulsow, C. Martin, and S. Perisho. 2014. High-resolution measurement of a bottlenose dolphin's (*Tursiops truncatus*) biosonar transmission beam pattern in the horizontal plane. The Journal of the Acoustical Society of America **136**:2025-2038.

Frasier, K. E., S. M. Wiggins, D. Harris, T. A. Marques, L. Thomas, and J. A. Hildebrand. 2016. Delphinid echolocation click detection probability on near-seafloor sensors. The Journal of the Acoustical Society of America **140**:1918-1930.

Frouin-Mouy, H., X. Mouy, B. Martin, and D. Hannay. 2016. Underwater acoustic behavior of bearded seals (Erignathus barbatus) in the northeastern Chukchi Sea, 2007–2010. Marine Mammal Science **32**:141-160.

Garland, E. C., A. W. Goldizen, M. L. Rekdahl, R. Constantine, C. Garrigue, N. D. Hauser, M. M. Poole, J. Robbins, and M. J. Noad. 2011. Dynamic Horizontal Cultural Transmission of Humpback Whale Song at the Ocean Basin Scale. Current Biology **21**:687-691.

Gassmann, M., S. M. Wiggins, and J. A. Hildebrand. 2015. Three-dimensional tracking of Cuvier's beaked whales' echolocation sounds using nested hydrophone arrays. The Journal of the Acoustical Society of America **138**:2483-2494.

Gerrodette, T., B. L. Taylor, R. Swift, S. Rankin, A. M. Jaramillo-Legorreta, and L. Rojas-Bracho. 2011. A combined visual and acoustic estimate of 2008 abundance, and change in abundance since 1997, for the vaquita, Phocoena sinus. Marine Mammal Science **27**:E79-E100.

Gillespie, D. 2004. Detection and classification of right whale calls using an'edge'detector operating on a smoothed spectrogram. Canadian Acoustics **32**:39-47.

Gillespie, D., P. Berggren, S. Brown, I. Kuklik, C. Lacey, T. Lewis, J. Matthews, R. McLanaghan, A. Moscrop, and N. Tregenza. 2005. Relative abundance of harbour porpoises (*Phocoena phocoena*) from acoustic and visual surveys of the Baltic Sea and adjacent waters during 2001 and 2002. J. Cetacean. Res. Manage **7**:51-57.

Gillespie, D., and M. Caillat. 2008. Statistical classification of odontocete clicks. Canadian Acoustics **36**:20-26.

Gillespie, D., M. Caillat, J. Gordon, and P. White. 2013. Automatic detection and classification of odontocete whistles. The Journal of the Acoustical Society of America **134**:2427-2437.

Gillespie, D., C. Dunn, J. Gordon, D. Claridge, C. Embling, and I. Boyd. 2009. Field recordings of Gervais' beaked whales Mesoplodon europaeus from the Bahamas. The Journal of the Acoustical Society of America **125**:3428-3433.

Gillespie, D., and R. Leaper. 1996. Detection of sperm whale (Physeter macrocephalus) clicks and discrimination of individual vocalisations. Eur. Res. Cetaceans **10**:87-91.

Gillespie, D., D. Mellinger, J. Gordon, D. McLaren, P. Redmond, R. McHugh, P. Trinder, X. Deng, and A. Thode. 2008. PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localisation of cetaceans. Journal of the Acoustical Society of America **30**:54-62.

Giraudet, P., and H. Glotin. 2006. Real-time 3D tracking of whales by echo-robust precise TDOA estimates with a widely-spaced hydrophone array. Applied Acoustics **67**:1106-1117.

Glennie, R., S. T. Buckland, and L. Thomas. 2015. The effect of animal movement on line transect estimates of abundance. PLoS One **10**:e0121333.

Glotin, H., F. Caudal, and P. Giraudet. 2008. Whale cocktail party: real-time multiple tracking and signal analyses. Canadian Acoustics **36**:139-145.

Goldbogen, J. A., A. K. Stimpert, S. L. DeRuiter, J. Calambokidis, A. S. Friedlaender, G. S. Schorr, D. J. Moretti, P. L. Tyack, and B. L. Southall. 2014. Using accelerometers to determine the calling behavior of tagged baleen whales. The Journal of Experimental Biology:jeb-103259.

Goodson, A., and C. Sturtivant. 1996. Sonar characteristics of the harbour porpoise (Phocoena phocoena): source levels and spectrum. ICES Journal of Marine Science: Journal du Conseil **53**:465-472.

Götz, T., R. Antunes, and S. Heinrich. 2010. Echolocation clicks of free-ranging Chilean dolphins (Cephalorhynchus eutropia). The Journal of the Acoustical Society of America **128**:563-566.

Griffiths, E. T., and J. Barlow. 2015. Equipment performance report for the Drifting Acoustic Spar Buoy Recorder (DASBR). NOAA Tech. Memo. NMFS-SWFSC **543**:36.

Hammond, P., K. Mc Leod, and M. Scheidat. 2006. Small Cetaceans in the European Atlantic and North Sea (SCANS-II). Final Report. Saint Andrews.

Hanggi, E. B., and R. J. Schusterman. 1994. Underwater acoustic displays and individual variation in male harbour seals, Phoca vitulina. Animal Behaviour **48**:1275-1283.

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

Hastie, G., B. Wilson, and P. Thompson. 2006. Diving deep in a foraging hotspot: acoustic insights into bottlenose dolphin dive depths and feeding behaviour. Marine Biology **148**:1181-1188.

Hawkins, E. R. 2010. Geographic variations in the whistles of bottlenose dolphins *(Tursiops aduncus*) along the east and west coasts of Australia. Journal of the Acoustical Society of America **128**:924-935.

Heerfordt, A., B. Møhl, and M. Wahlberg. 2007. A wideband connection to sperm whales: A fiber-optic, deep-sea hydrophone array. Deep Sea Research Part I: Oceanographic Research Papers **54**:428-436.

Heimlich, S. L., D. K. Mellinger, S. L. Nieukirk, and C. G. Fox. 2005. Types, distribution, and seasonal occurrence of sounds attributed to Bryde's whales (Balaenoptera edeni) recorded in the eastern tropical Pacific, 1999–2001. The Journal of the Acoustical Society of America **118**:1830-1837.

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

Helble, T. A., G. L. D'Spain, G. S. Campbell, and J. A. Hildebrand. 2013. Calibrating passive acoustic monitoring: Correcting humpback whale call detections for site-specific and time-dependent environmental characteristics. Journal of the Acoustical Society of America **134**:EL400-EL406.

Helble, T. A., G. R. Ierley, G. L. D'Spain, and S. W. Martin. 2015. Automated acoustic localization and call association for vocalizing humpback whales on the Navy's Pacific Missile Range Facility. The Journal of the Acoustical Society of America **137**:11-21.

Helble, T. A., G. R. Ierley, G. L. D'Spain, M. A. Roch, and J. A. Hildebrand. 2012. A generalized power-law detection algorithm for humpback whale vocalizations. The Journal of the Acoustical Society of America **131**:2682-2699.

Heupel, M. R., K. L. Reiss, B. G. Yeiser, and C. A. Simpfendorfer. 2008. Effects of biofouling on performance of moored data logging acoustic receivers. Limnology and Oceanography: Methods **6**:327-335.

Hildebrand, J. A., S. Baumann-Pickering, A. Giddings, A. Brewer, E. R. Jacobs, J. S. Trickey, S. M. Wiggins, and M. A. McDonald. 2016. Progress Report on the Application of Passive Acoustic Monitoring to Density Estimation of Cuvier's Beaked Whales.

Holt, M. M., D. P. Noren, V. Veirs, C. K. Emmons, and S. Veirs. 2009. Speaking up: Killer whales (Orcinus orca) increase their call amplitude in response to vessel noise. The Journal of the Acoustical Society of America **125**:EL27-EL32.

Höschle, C., Pierpoint, C., Diederichs, A., Betke, K., Matuschek, R., Karnatz, W., Nielsen, S. & Nehls, G. (2013). SAVE from harmful noise – real-time detection system of harbour porpoises before pile driving. European Cetacean Society Conference, Setubal, Portugal.

Janik, V. M. 2000. Source levels and the estimated active space of bottlenose dolphin (Tursiops truncatus) whistles in the Moray Firth, Scotland. Journal of Comparative Physiology A **186**:673-680.

Janik, V. M. 2009. Acoustic communication in delphinids. Advances in the Study of Behavior **40**:123-157.

Jewell, R., L. Thomas, C. Harris, K. Kaschner, R. Wiff, P. Hammond, and N. Quick. 2012. Global analysis of cetacean line-transect surveys: detecting trends in cetacean density. Marine Ecology Progress Series **453**:227-240.

Johansson, A. T., and P. R. White. 2011. An adaptive filter-based method for robust, automatic detection and frequency estimation of whistles. Journal of the Acoustical Society of America **130**:893-903.

Johnson, M., P. T. Madsen, W. M. Zimmer, N. A. De Soto, and P. L. Tyack. 2004. Beaked whales echolocate on prey. Proceedings of the Royal Society of London B: Biological Sciences **271**:S383-S386.

Johnson, M., J. Partan, and T. Hurst. 2013. Low complexity lossless compression of underwater sound recordings. The Journal of the Acoustical Society of America **133**:1387-1398.

Johnson, M. P., and P. L. Tyack. 2003. A digital acoustic recording tag for measuring the response of wild marine mammals to sound. IEEE Journal of Oceanic Engineering **28**:3-12.

Jones, G. J., and L. S. Sayigh. 2002. Geographic variation in rates of vocal production of free-ranging bottlenose dolphins. Marine Mammal Science **18**:374-393.

Karnowski, J., and Y. Movshovitz-Attias. 2015. Classification of blue whale D calls and fin whale 40-Hz calls using deep learning.*in* 7th International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics, La Jolla, CA.

Kaschner, K., N. J. Quick, R. Jewell, R. Williams, and C. M. Harris. 2012. Global Coverage of Cetacean Line-Transect Surveys: Status Quo, Data Gaps and Future Challenges. PLoS One **7**.

Kay, S. M. 1993. Fundamentals of statistical signal processing, volume I: estimation theory.

Keenan, G., C. Sparling, H. Williams, and F. Fortune. 2011. SeaGen Environmental Monitoring Programme. 9S8562/R/303719/Edin.

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

Klinck, H., and D. K. Mellinger. 2011. The energy ratio mapping algorithm: A tool to improve the energy-based detection of odontocete echolocation clicks. Journal of the Acoustical Society of America **129**:1807-1812.

Koblitz, J. C., M. Wahlberg, P. Stilz, P. T. Madsen, K. Beedholm, and H. U. Schnitzler. 2012. Asymmetry and dynamics of a narrow sonar beam in an echolocating harbor porpoise. Journal of the Acoustical Society of America **131**:2315-2324.

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

Kyhn, L. A., F. H. Jensen, K. Beedholm, J. Tougaard, M. Hansen, and P. T. Madsen. 2010. Echolocation in sympatric Peale's dolphins (Lagenorhynchus australis) and Commerson's dolphins (Cephalorhynchus commersonii) producing narrow-band high-frequency clicks. Journal of Experimental Biology **213**:1940-1949.

Kyhn, L. A., J. Tougaard, K. Beedholm, F. H. Jensen, E. Ashe, R. Williams, and P. T. Madsen. 2013. Clicking in a Killer Whale Habitat: Narrow-Band, High-Frequency Biosonar Clicks of Harbour Porpoise (Phocoena phocoena) and Dall's Porpoise (Phocoenoides dalli). PLoS One **8**.

Kyhn, L. A., J. Tougaard, F. Jensen, M. Wahlberg, G. Stone, A. Yoshinaga, K. Beedholm, and P. T. Madsen. 2009. Feeding at a high pitch: Source parameters of narrow band, high-frequency clicks from echolocating off-shore hourglass dolphins and coastal Hector’s dolphins. The Journal of the Acoustical Society of America **125**:1783-1791.

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

Lara, G., R. Miralles, and A. Carrion. 2013. Right Whale activity detector and sound classificator using Mel-frequency Cepstral Coefficients.*in* 6th International Workshop on Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics, St Andrews, Scotland.

Laurinolli, M. H., A. E. Hay, F. Desharnais, and C. T. Taggart. 2003. Localization of North Atlantic right whale sounds in the Bay of Fundy using a sonobuoy array. Marine Mammal Science **19**:708-723.

Leaper, R., D. Gillespie, J. Gordon, and J. Matthews. 2003. Abundance assessment of sperm whales.

Leaper, R., D. Gillespie, and V. Papastavrou. 2000. Results of passive acoustic surveys for odontocetes in the Southern Ocean. Journal of Cetacean Research and Management **2**:187-196.

Leaper, R., and J. Gordon. 2012. Marine Mammal Acoustic and Visual Surveys - Analysis of Neart Na Gaoithe data.

Leeney, R. H., D. Carslake, and S. H. Elwen. 2011. Using Static Acoustic Monitoring to Describe Echolocation Behaviour of Heaviside's Dolphins (*Cephalorhynchus heavisidii*) in Namibia. Aquatic Mammals **37**:151-160.

Leroy, E. C., F. Samaran, J. Bonnel, and J.-Y. Royer. 2016. Seasonal and Diel Vocalization Patterns of Antarctic Blue Whale (Balaenoptera musculus intermedia) in the Southern Indian Ocean: A Multi-Year and Multi-Site Study. PLoS One **11**:e0163587.

Lewis, T., D. Gillespie, C. Lacey, J. Matthews, M. Danbolt, R. Leaper, R. McLanaghan, and A. Moscrop. 2007. Sperm whale abundance estimates from acoustic surveys of the Ionian Sea and Straits of Sicily in 2003. Journal of the Marine Biological Association of the United Kingdom **87**:353-357.

Linnenschmidt, M., L. N. Kloepper, M. Wahlberg, and P. E. Nachtigall. 2012. Stereotypical rapid source level regulation in the harbour porpoise biosonar. Naturwissenschaften **99**:767-771.

Lockhart, S. 2015. Detecting Blue Whale Calls By Extending a Speech Processing 81 Algorithm: A Proof of Concept.*in* 7th International Workshop on Detection, Classification, Localization, and Density Estimation (DCLDE) of Marine Mammals using Passive Acoustics, La Jolla, CA.

Lukacs, P. M., and K. P. Burnham. 2005. Research notes: Estimating population size from DNA-based closed capture–recapture data incorporating genotyping error. Journal of Wildlife Management **69**:396-403.

MacAulay, J., Gordon, J., Gillespie, D., Malinka, C., and Northridge, S. (2017). “Passive acoustic methods for fine-scale tracking of harbour porpoises in tidal rapids,” J. Acoust. Soc. Am., 141, 1120–1132.

Madsen, P., M. Johnson, N. A. De Soto, W. Zimmer, and P. Tyack. 2005a. Biosonar performance of foraging beaked whales (Mesoplodon densirostris). Journal of Experimental Biology **208**:181-194.

Madsen, P., R. Payne, N. Kristiansen, M. Wahlberg, I. Kerr, and B. Møhl. 2002a. Sperm whale sound production studied with ultrasound time/depth-recording tags. Journal of Experimental Biology **205**:1899-1906.

Madsen, P., M. Wahlberg, and B. M+©hl. 2002b. Male sperm whale (Physeter macrocephalus) acoustics in a high-latitude habitat: implications for echolocation and communication. Behavioral Ecology and Sociobiology **53**:31-41.

Madsen, P. T., D. Carder, K. Bedholm, and S. Ridgway. 2005b. Porpoise clicks from a sperm whale nose—Convergent evolution of 130 kHz pulses in toothed whale sonars? Bioacoustics **15**:195-206.

Madsen, P. T., N. A. de Soto, P. L. Tyack, and M. Johnson. 2014. Beaked whales. Current Biology **24**:R728-R730.

Malinka C. E., Gillespie, D. M., MacAulay, J. D. J., Joy, R., Sparling, C. 2017. First in-situ passive acoustic monitoring for marine mammals during operation of a tidal turbine in Ramsey Sound, Wales. Marine Ecology Progress Series (submitted for review 30/05/2017).

Mann, D. A., B. M. Casper, K. S. Boyle, and T. C. Tricas. 2007. On the attraction of larval fishes to reef sounds. Marine Ecology-Progress Series **338**:307.

Marques, T. A. 2004. Predicting and correcting bias caused by measurement error in line transect sampling using multiplicative error models. Biometrics **60**:757-763.

Marques, T. A., S. T. Buckland, R. Bispo, and B. Howland. 2013a. Accounting for animal density gradients using independent information in distance sampling surveys. Statistical Methods and Applications **22**:67-80.

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

Marques, T. A., L. Thomas, S. W. Martin, D. K. Mellinger, S. Jarvis, R. P. Morrissey, C. A. Ciminello, and N. DiMarzio. 2012. Spatially explicit capture-recapture methods to estimate minke whale density from data collected at bottom-mounted hydrophones. Journal of Ornithology **152**:S445-S455.

Marques, T. A., L. Thomas, S. W. Martin, D. K. Mellinger, J. A. Ward, D. J. Moretti, D. Harris, and P. L. Tyack. 2013b. 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.

Marten, K. 2000. Ultrasonic analysis of pygmy sperm whale (Kogia breviceps) and Hubbs' beaked whale (Mesoplodon carlhubbsi) clicks. Aquatic Mammals **26**:45-48.

Martin, B., K. Kowarski, H. Murphy-Moors, and C. Evers. 2015. Differentiating marine mammal clicks using time-series properties.*in* 7th International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics, La Jolla, CA.

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.

Matias, L., and D. Harris. 2015. A single-station method for the detection, classification and location of fin whale calls using ocean-bottom seismic stations. The Journal of the Acoustical Society of America **138**:504-520.

Matthews, J., S. Brown, D. Gillespie, M. Johnson, R. McLanaghan, A. Moscrop, D. Nowacek, R. Leaper, T. Lewis, and P. Tyack. 2001. Vocalisation rates of the North Atlantic right whale (Eubalaena glacialis). Journal of Cetacean Research and Management **3**:271-282.

May-Collado, L. J., and D. Wartzok. 2008. A comparison of bottlenose dolphin whistles in the Atlantic Ocean: factors promoting whistle variation. Journal of mammalogy **89**:1229-1240.

McDonald, M. A., J. Calambokidis, A. M. Teranishi, and J. A. Hildebrand. 2001. The acoustic calls of blue whales off California with gender data. The Journal of the Acoustical Society of America **109**:1728-1735.

McDonald, M. A., and C. G. Fox. 1999. Passive acoustic methods applied to fin whale population density estimation. The Journal of the Acoustical Society of America **105**:2643-2651.

McDonald, M. A., J. A. Hildebrand, and S. C. Webb. 1995. Blue and fin whales observed on a seafloor array in the Northeast Pacific. The Journal of the Acoustical Society of America **98**:712-721.

McDonald, M. A., S. L. Mesnick, and J. A. Hildebrand. 2006. Biogeographic characterization of blue whale song worldwide: using song to identify populations. Journal of Cetacean Research and Management **8**.

Matsumoto, H., Jones, C., Klinck, H., Mellinger, D. K., Dziak, R. P., & Meinig, C. (2013). Tracking beaked whales with a passive acoustic profiler float. The Journal of the Acoustical Society of America, 133(2), 731–40. https://doi.org/10.1121/1.4773260.

Mellinger, D. K., and J. Barlow. 2003. Future Directions for Acoustic Marine Mammal Surveys: Stock Assessment and Habitat Use: Report of a Workshop Held in La Jolla, California, 20-22 November 2002. US Department of Commerce, National Oceanic and Atmospheric Administration, Pacific Marine Environmental Laboratory.

Mellinger, D. K., and C. W. Clark. 2000. Recognizing transient low-frequency whale sounds by spectrogram correlation. The Journal of the Acoustical Society of America **107**:3518-3529.

Mellinger, D. K., and C. W. Clark. 2003. Blue whale (Balaenoptera musculus) sounds from the North Atlantic. The Journal of the Acoustical Society of America **114**:1108-1119.

Mellinger, D. K., S. W. Martin, R. P. Morrissey, L. Thomas, and J. J. Yosco. 2011. A method for detecting whistles, moans, and other frequency contour sounds. Journal of the Acoustical Society of America **129**:4055-4061.

Mellinger, D. K., K. M. Stafford, S. Moore, R. P. Dziak, and H. Matsumoto. 2007. Fixed passive acoustic observation methods for cetaceans. Oceanography **20**:36.

Merchant, N. D., E. Pirotta, T. R. Barton, and P. M. Thompson. 2014. Monitoring ship noise to assess the impact of coastal developments on marine mammals. Marine pollution bulletin **78**:85-95.

Merkens, K., Y. Barkley, M. Hill, and E. Oleson. 2016. Dwarf sperm whale (Kogia sima) echolocation clicks from Guam (Western North Pacific Ocean). The Journal of the Acoustical Society of America **140**:3415-3415.

Miksis-Olds, J. L., Vernon, J. A., & Heaney, K. D. (2015). ESOMM-2014: The impact of ocean sound dynamics on estimates of signal detection range. Aquatic Mammals, 41(4), 444-453. http://dx.doi.org/10.1578/AM.41.4.2015.444

Miller, B., and S. Dawson. 2009. A large-aperture low-cost hydrophone array for tracking whales from small boats. The Journal of the Acoustical Society of America **126**:2248-2256.

Miller, B. S. 2010. Acoustically derived growth rates and three-dimensional localisation of sperm whales (Physeter macrocephalus) in Kaikoura, New Zealand. University of Otago.

Miller, B. S. 2012. Real-time tracking of blue whales using DIFAR sonobuoys.*in* Proceedings of acoustics.

Miller, B. S., D. Gillespie, G. Weatherup, S. Calderan, and M. C. Double. 2014. Software for the localisation of baleen whale calls using DIFAR sonobuoys: PAMGuard DIFAR. Submitt to Sci Comm 65b Int Whal Comm Bled, Slov SC/65b/SH **7**.

Miller, P. J., N. Biassoni, A. Samuels, and P. L. Tyack. 2000. Whale songs lengthen in response to sonar. Nature **405**:903-903.

Miller, P. J., M. P. Johnson, and P. L. Tyack. 2004. Sperm whale behaviour indicates the use of echolocation click buzzes 'creaks' in prey capture. Proceedings of the Royal Society of London B: Biological Sciences **271**:2239-2247.

Miller, P. J., and P. L. Tyack. 1998. A small towed beamforming array to identify vocalizing resident killer whales (Orcinus orca) concurrent with focal behavioral observations. Deep Sea Research Part II: Topical Studies in Oceanography **45**:1389-1405.

Mizuguchi, D., Y. Mitani, and S. Kohshima. 2016a. Geographically specific underwater vocalizations of ribbon seals (Histriophoca fasciata) in the Okhotsk Sea suggest a discrete population. Marine Mammal Science **32**:1138-1151.

Mizuguchi, D., M. Tsunokawa, M. Kawamoto, and S. Kohshima. 2016b. Underwater vocalizations and associated behavior in captive ringed seals (Pusa hispida). Polar biology **39**:659-669.

Mohl, B., and S. Andersen. 1973. Echolocation: high frequency component in the click of the Harbour Porpoise (*Phocoena ph. L.*). The Journal of the Acoustical Society of America **54**:1368-1372.

Møhl, B., M. Wahlberg, and A. Heerfordt. 2001. A large-aperture array of nonlinked receivers for acoustic positioning of biological sound sources. The Journal of the Acoustical Society of America **109**:434-437.

Mohl, B., M. Wahlberg, P. T. Madsen, A. Heerfordt, and A. Lund. 2003. The monopulsed nature of sperm whale clicks. The Journal of the Acoustical Society of America **114**:1143-1154.

Møhl, B., M. Wahlberg, P. T. Madsen, L. A. Miller, and A. Surlykke. 2000. Sperm whale clicks: Directionality and source level revisited. The Journal of the Acoustical Society of America **107**:638-648.

Moore, P. W., L. A. Dankiewicz, and D. S. Houser. 2008. Beamwidth control and angular target detection in an echolocating bottlenose dolphin (Tursiops truncatus). The Journal of the Acoustical Society of America **124**:3324-3332.

Moretti, D., T. Casey, and D. K. Mellinger. 2009. Fixed passive acoustic marine mammal monitoring for estimating species abundance and mitigating the effect of operations on the marine environment.

Moretti, D., T. Marques, L. Thomas, N. DiMarzio, A. Dilley, R. Morrissey, E. McCarthy, J. Ward, and S. Jarvis. 2010. A dive counting density estimation method for Blainville's beaked whale (Mesoplodon densirostris) using a bottom-mounted hydrophone field as applied to a Mid-Frequency Active (MFA) sonar operation. Applied Acoustics **71**:1036-1042.

Muoy, X., J. Moloney, T. Dakin, J. Pilkington, A. Riera, J. Ford, K. Kanes, and P. Mouy. 2015. Automatic Marine Mammal Monitoring off British-Columbia, Canada.*in* 7th International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics, La Jolla, CA.

Nehls, G., Höschle, C., Kosarev, V., Brandt, M.J., Diederichs, A. 2015. Applying the wireless detection system (WDS) – a realtime monitoring tool for harbour porpoise activity around construction sites. The 21st Biennial Society of Marine Mammalogy Conference. San Francisco, USA.

Newman, K. B., and S. T. Lindley. 2006. Accounting for demographic and environmental stochasticity, observation error, and parameter uncertainty in fish population dynamics models. North American Journal of Fisheries Management **26**:685-701.

Nieukirk, S. L., D. K. Mellinger, S. E. Moore, K. Klinck, R. P. Dziak, and J. Goslin. 2012. Sounds from airguns and fin whales recorded in the mid-Atlantic Ocean, 1999-2009. Journal of the Acoustical Society of America **131**:1102-1112.

Norris, T., J. Oswald, and R. Sousa-Lima. 2010. A review and inventory of fixed installation passive acoustic monitoring methods and technologies. Final report to the International Association of Oil and Gas Producers Joint Industry Programme on E&P Sound and Marine Life.

Nosal, E.-M. 2013. Methods for tracking multiple marine mammals with wide-baseline passive acoustic arrays. The Journal of the Acoustical Society of America **134**:2383-2392.

Oleson, E. M., J. Barlow, J. Gordon, S. Rankin, and J. A. Hildebrand. 2003. Low frequency calls of Bryde's whales.

Oleson, E. M., J. Calambokidis, W. C. Burgess, M. A. McDonald, C. A. LeDuc, and J. A. Hildebrand. 2007. Behavioral context of call production by eastern North Pacific blue whales.

Oliveira, C., M. Wahlberg, M. Johnson, P. J. Miller, and P. T. Madsen. 2013. The function of male sperm whale slow clicks in a high latitude habitat: Communication, echolocation, or prey debilitation? The Journal of the Acoustical Society of America **133**:3135-3144.

Oliveira, C., M. Wahlberg, M. A. Silva, M. Johnson, R. Antunes, D. M. Wisniewska, A. Fais, J. Gonçalves, and P. T. Madsen. 2016. Sperm whale codas may encode individuality as well as clan identity. The Journal of the Acoustical Society of America **139**:2860-2869.

Oswald, J., T. Norris, W. Au, and F. Duennebier. 2008. Temporal and geographic patterns in the occurrence and distribution of minke whale (Balaenoptera acutorostrata) boings in the central and western North Pacific. The Journal of the Acoustical Society of America **124**:2484-2484.

Oswald, J. N., W. W. Au, and F. Duennebier. 2011. Minke whale (Balaenoptera acutorostrata) boings detected at the Station ALOHA Cabled Observatory. The Journal of the Acoustical Society of America **129**:3353-3360.

Oswald, J. N., J. Barlow, and T. F. Norris. 2003. Acoustic identification of nine delphinid species in the eastern tropical Pacific Ocean. Marine Mammal Science **19**:20-037.

Oswald, J. N., S. Rankin, J. Barlow, and M. O. Lammers. 2007. A tool for real-time acoustic species identification of delphinid whistles. The Journal of the Acoustical Society of America **122**:587-595.

Ou, H., W. W. Au, and J. N. Oswald. 2012. A non-spectrogram-correlation method of automatically detecting minke whale boings. Journal of the Acoustical Society of America **132**:EL317-EL322.

Pace, F., P. White, and W. Adam. 2011. Classification of humpback whale (Megaptera novaeangliae) calls using Hidden Markov Models.*in* Fifth International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics Oregon, USA.

Parks, S. E., P. K. Hamilton, S. D. Kraus, and P. L. Tyack. 2005. The gunshot sound produced by male North Atlantic right whales (Eubalaena glacialis) and its potential function in reproductive advertisement. Marine Mammal Science **21**:458-475.

Parks, S. E., C. F. Hotchkin, K. A. Cortopassi, and C. W. Clark. 2012. Characteristics of gunshot sound displays by North Atlantic right whales in the Bay of Fundy. The Journal of the Acoustical Society of America **131**:3173-3179.

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.

Parks, S. E., and P. L. Tyack. 2005. Sound production by North Atlantic right whales (Eubalaena glacialis) in surface active groups. The Journal of the Acoustical Society of America **117**:3297-3306.

Payne, R. S., and S. McVay. 1971. Songs of humpback whales. Science **173**:585-597.

Pierpoint, C., Sidorovskaia, N., Heath, B., Marks, K., Scala, L., Morrison, B., Rutter, D., Royston, J., Fuselier, J. (2016). LADC-GOMRI: Passive Acoustic Monitoring Of Marine Mammals Using Automated Surface Vehicles. Gulf Of Mexico Oil Spill Science Conference, Tampa, Florida, 1st-4th Feb. 2016.

Pirotta, E., K. L. Brookes, I. M. Graham, and P. M. Thompson. 2014. Variation in harbour porpoise activity in response to seismic survey noise. Biology letters **10**:20131090.

Radford, C., J. Stanley, C. Tindle, J. Montgomery, and A. Jeffs. 2010. Localised coastal habitats have distinct underwater sound signatures. Marine Ecology Progress Series **401**:21-29.

Rankin, S., and J. Barlow. 2005. Source of the North Pacific “boing” sound attributed to minke whales. The Journal of the Acoustical Society of America **118**:3346-3351.

Rankin, S., J. Barlow, J. L. Keating, and J. Oswald. 2015. Improving classification of dolphin groups: compound classification using whistles, echolocation, and burst pulses.*in* 7th International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics, La Jolla, CA.

Rasmussen, M. H., M. Wahlberg, and L. A. Miller. 2004. Estimated transmission beam pattern of clicks recorded from free-ranging white-beaked dolphins (Lagenorhynchus albirostris). The Journal of the Acoustical Society of America **116**:1826-1831.

Rendell, L., and H. Whitehead. 2004. Do sperm whales share coda vocalizations? Insights into coda usage from acoustic size measurement. Animal Behaviour **67**:865-874.

Rendell, L. E., and J. C. D. Gordon. 1999. Vocal response of long-finned pilot whales (*Globicephala melas*) to military sonar in the Ligurian Sea. Marine Mammal Science **15**:198-204.

Reyes Reyes, M. V., V. P. Tossenberger, M. A. Iñiguez, J. A. Hildebrand, and M. L. Melcón. 2016. Communication sounds of Commerson's dolphins (Cephalorhynchus commersonii) and contextual use of vocalizations. Marine Mammal Science **32**:1219-1233.

Rice A.N., Palmer K.J., Tielens J.T., Muirhead C.A., Clark C.W. 2014. Potential Bryde's whale (Balaenoptera edeni) calls recorded in the northern Gulf of Mexico. Journal of the Acoustical Society of America;**135**(5):3066-76.

Riccobene, G., Aiello, S., Aloisio, A., Ameli, F., Amore, I., Anghinolfi, M., … Wischnewski, R. (2009). Long-term measurements of acoustic back ground noise in very deep sea. Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 604(1–2 SUPPL.). https://doi.org/10.1016/j.nima.2009.03.195

Richardson, W. 1995. Marine mammals and noise., Toronto: Academic Press.

Rideout, B. P., S. E. Dosso, and D. E. Hannay. 2013. Underwater passive acoustic localization of Pacfiic walruses in the northeastern Chukchi Sea. Journal of the Acoustical Society of America **134**:2534-2545.

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.

Risch, D., P. J. Corkeron, W. T. Ellison, and S. M. Van Parijs. 2012. Changes in Humpback Whale Song Occurrence in Response to an Acoustic Source 200 km Away. PLoS One **7**.

Risch, D., N. J. Gales, J. Gedamke, L. Kindermann, D. P. Nowacek, A. J. Read, U. Siebert, I. C. Van Opzeeland, S. M. Van Parijs, and A. S. Friedlaender. 2014. Mysterious bio-duck sound attributed to the Antarctic minke whale (Balaenoptera bonaerensis). Biology letters **10**:20140175.

Roch, M. A., T. S. Brandes, B. Patel, Y. Barkley, S. Baumann-Pickering, and M. S. Soldevilla. 2011a. Automated extraction of odontocete whistle contours. Journal of the Acoustical Society of America **130**:2212-2223.

Roch, M. A., H. Klinck, S. Baumann-Pickering, D. K. Mellinger, S. Qui, M. S. Soldevilla, and J. A. Hildebrand. 2011b. Classification of echolocation clicks from odontocetes in the Southern California Bight. Journal of the Acoustical Society of America **129**:467-475.

Roy, N., Y. Simard, M. Bahoura, and S. Giard. 2013. DCLDE 2013 workshop data set: Detection of Right Whale contact calls with supervised spectrogram templates and chirplet approaches.*in* 6th International Workshop on Detection, Classification, Localization and Density Estimation of Marine Mammals using Passive Acoustics, St Andrews, Scotland.

Roy, N., Y. Simard, and C. Gervaise. 2010. 3D tracking of foraging belugas from their clicks: Experiment from a coastal hydrophone array. Applied Acoustics **71**:1050-1056.

Samaran, F., C. Guinet, O. Adam, J. F. o. Motsch, and Y. Cansi. 2010. Source level estimation of two blue whale subspecies in southwestern Indian Ocean. The Journal of the Acoustical Society of America **127**:3800-3808.

Schulz, T. M., H. Whitehead, S. Gero, and L. Rendell. 2008. Overlapping and matching of codas in vocal interactions between sperm whales: insights into communication function. Animal Behaviour **76**:1977-1988.

Scott-Hayward, L., C. Oedekoven, M. Mackenzie, C. Walker, and E. Rexstad. 2013. MRSea package (version 0.2.0): Statistical Modelling of bird and cetacean distributions in offshore renewables development areas. University of St. Andrews.

Shaffer, J. W., D. Moretti, S. Jarvis, P. Tyack, and M. Johnson. 2013. Effective beam pattern of the Blainville's beaked whale (*Mesoplodon densirostris*) and implications for passive acoustic monitoring. Journal of the Acoustical Society of America **133**:1770-1784.

Silber, G. K. 1986. The relationship of social vocalizations to surface behavior and aggression in the Hawaiian humpback whale (Megaptera novaeangliae). Canadian journal of zoology **64**:2075-2080.

Sirovic, A., J. A. Hildebrand, and S. M. Wiggins. 2007. Blue and fin whale call source levels and propagation range in the Southern Ocean. The Journal of the Acoustical Society of America **122**:1208-1215.

Širović, A., J. A. Hildebrand, S. M. Wiggins, M. A. McDonald, S. E. Moore, and D. Thiele. 2004. Seasonality of blue and fin whale calls and the influence of sea ice in the Western Antarctic Peninsula. Deep Sea Research Part II: Topical Studies in Oceanography **51**:2327-2344.

Soule, D. C., & Wilcock, W. S. D. (2013). Fin whale tracks recorded by a seismic network on the Juan de Fuca Ridge, Northeast Pacific Ocean. The Journal of the Acoustical Society of America, 133(3), 1751. https://doi.org/10.1121/1.4774275

Sousa-Lima, R. S., T. F. Norris, J. N. Oswald, and D. P. Fernandes. 2013. A Review and Inventory of Fixed Autonomous Recorders for Passive Acoustic Monitoring of Marine Mammals. Aquatic Mammals **39**:23-53.

Stafford, K. M., C. G. Fox, and D. S. Clark. 1998. Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean. The Journal of the Acoustical Society of America **104**:3616-3625.

Stafford, K. M., J. J. Citta, S. E. Moore, M. A. Daher, and J. E. George. 2009. Environmental correlates of blue and fin whale call detections in the North Pacific Ocean from 1997 to 2002.

Stafford, K. M., S. E. Moore, and C. G. Fox. 2005. Diel variation in blue whale calls recorded in the eastern tropical Pacific. Animal Behaviour **69**:951-958.

Stafford, K. M., S. L. Nieukirk, and C. G. Fox. 2001. Geographic and seasonal variation of blue whale calls in the North Pacific. Journal of Cetacean Research and Management **3**:65-76.

Stanley, J. A., C. A. Radford, and A. G. Jeffs. 2012. Location, location, location: finding a suitable home among the noise. Proceedings of the Royal Society of London B: Biological Sciences **279**:3622-3631.

Stevenson, B. 2016. Methods in spatially explicit capture-recapture. University of St Andrews.

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

Stimpert, A., S. L. DeRuiter, B. Southall, D. Moretti, E. Falcone, J. Goldbogen, A. Friedlaender, G. Schorr, and J. Calambokidis. 2014. Acoustic and foraging behavior of a Baird's beaked whale, Berardius bairdii, exposed to simulated sonar. Scientific Reports **4**:7031.

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.

Teloni, V., J. P. Mark, M. J. Patrick, and M. T. Peter. 2008. Shallow food for deep divers: Dynamic foraging behavior of male sperm whales in a high latitude habitat. Journal of Experimental Marine Biology and Ecology **354**:119-131.

Tervo, O. M., M. F. Christoffersen, S. E. Parks, R. Møbjerg Kristensen, and P. Teglberg Madsen. 2011. Evidence for simultaneous sound production in the bowhead whale (Balaena mysticetus). The Journal of the Acoustical Society of America **130**:2257-2262.

Thode, A. 2005. Three-dimensional passive acoustic tracking of sperm whales (Physeter macrocephalus) in ray-refracting environments. The Journal of the Acoustical Society of America **118**:3575-3584.

Thode, A., K. Kim, S. B. Blackwell, C. R. Greene Jr, C. S. Nations, T. McDonald, and A. Macrander. 2011. Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys, 2007-2010 *in* Fifth International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics Oregon, USA.

Thode, A., T. Norris, and J. Barlow. 2000a. Frequency beamforming of dolphin whistles using a sparse three-element towed array. The Journal of the Acoustical Society of America **107**:3581-3584.

Thode, A. M., G. L. D'Spain, and W. Kuperman. 2000b. Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations. The Journal of the Acoustical Society of America **107**:1286-1300.

Thode, A. M., P. Gerstoft, W. C. Burgess, K. G. Sabra, M. Guerra, M. D. Stokes, M. Noad, and D. H. Cato. 2006. A Portable Matched-Field Processing System Using Passive Acoustic Time Synchronization. IEEE Journal of Oceanic Engineering **31**:696-710.

Thompson, P. O., L. T. Findley, and O. Vidal. 1992. 20-Hz pulses and other vocalizations of fin whales, B alaenopteraphysalus, in the Gulf of California, Mexico. The Journal of the Acoustical Society of America **92**:3051-3057.

Thompson, P. O., and W. A. Friedl. 1982. A long-term study of low frequency sounds from several species of whales off Oahu, Hawaii. Biological Systems.

Thompson, S., S. T. Buckland, D. R. Anderson, K. P. Burnham, and J. L. Laake. 1994. Distance Sampling: Estimating Abundance of Biological Populations. Biometrics **50**:891.

Tiemann, C. O., M. B. Porter, and L. N. Frazer. 2004. Localization of marine mammals near Hawaii using an acoustic propagation model. The Journal of the Acoustical Society of America **115**:2834-2843.

Todd, V., I. Todd, J. Gardiner, and E. Morrin. 2015. Marine mammal observer and passive acoustic monitoring handbook. Pelagic Publishing Ltd.

Trinh, Y., V. Rao, S. Lindeneau, M. Ackerman, S. Baumann-Pickering, and M. A. Roch. 2015. Unsupervised Clustering of Delphinid Encounters.*in* 7th International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics, La Jolla, CA.

Tyack, P. L., and V. M. Janik. 2013. Effects of noise on acoustic signal production in marine mammals. Pages 251-271 Animal communication and noise. Springer.

Tyack, P. L., M. Johnson, N. A. Soto, A. Sturlese, and P. T. Madsen. 2006. Extreme diving of beaked whales. Journal of Experimental Biology **209**:4238-4253.

Tyack, P. L., W. M. X. Zimmer, D. Moretti, B. L. Southall, D. E. Claridge, J. W. Durban, C. W. Clark, A. D'Amico, N. DiMarzio, S. Jarvis, E. McCarthy, R. Morrissey, J. Ward, and I. L. Boyd. 2011. Beaked Whales Respond to Simulated and Actual Navy Sonar. PLoS One **6**.

Van Parijs, S. M., C. W. Clark, R. S. Sousa-Lima, S. E. Parks, S. Rankin, D. Risch, and I. van Opzeeland. 2009. Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales. Marine Ecology Progress Series **395**:21-36.

Van Parijs, S. M., C. Lydersen, and K. M. Kovacs. 2003. Vocalizations and movements suggest alternative mating tactics in male bearded seals. Animal Behaviour **65**:273-283.

Veirs, S., V. Veirs, and J. D. Wood. 2016. Ship noise extends to frequencies used for echolocation by endangered killer whales. PeerJ **4**:e1657.

Verfuss, U., D. Gillespie, J. Gordon, T. Marques, B. Miller, R. Plunkett, J. Theriault, D. Tollit, D. Zitterbart, P. Hubert, and L. Thomas. 2016. Low Visibility Real-Time Monitoring Techniques Review. Report Number SMRUM-OGP-2015-002 Provided To IOGP, March 2016 (Unpublished).

Verfuss, U. K., A. S. Aniceto, M. Biuw, S. Fielding, D. Gillespie, D. Harris, G. Jimenez, P. Johnston, R. Plunkett, A. Sivertsen, A. Solbø, R. Storvold, and R. Wyatt. 2015. Literature Review: Understanding The Current State Of Autonomous Technologies To Improve/Expand Observation And Detection Of Marine Species. Report Number SMRUC-OGP-2015-23015 Provided To IOGP, 24th February, 2016 (Unpublished).

Villadsgaard, A., M. Wahlberg, and J. Tougaard. 2007. Echolocation signals of wild harbour porpoises, *Phocoena phocoena*. Journal of Experimental Biology **210**:56-64.

Von Ahn, L., B. Maurer, C. McMillen, D. Abraham, and M. Blum. 2008. recaptcha: Human-based character recognition via web security measures. Science **321**:1465-1468.

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.

Wahlberg, M. 2002. The acoustic behaviour of diving sperm whales observed with a hydrophone array. Journal of Experimental Marine Biology and Ecology **281**:53-62.

Wahlberg, M., F. H. Jensen, N. Aguilar Soto, K. Beedholm, L. Bejder, C. Oliveira, M. Rasmussen, M. Simon, A. Villadsgaard, and P. T. Madsen. 2011. Source parameters of echolocation clicks from wild bottlenose dolphins (Tursiops aduncus and Tursiops truncatus). The Journal of the Acoustical Society of America **130**:2263-2274.

Ward, J., S. Jarvis, D. Moretti, R. Morrissey, N. DiMarzio, M. Johnson, P. Tyack, L. Thomas, and T. Marques. 2011. Beaked whale (*Mesoplodon densirostris*) passive acoustic detection in increasing ambient noise. Journal of the Acoustical Society of America **129**:662-669.

Ward, J., R. Morrissey, D. Moretti, N. DiMarzio, S. Jarvis, M. Johnson, P. Tyack, and C. White. 2008. Passive Acoustic Detection and Localization of Mesoplodon Densirostris (Blainville's Beaked Whale) Vocalization Using Distributed Bottom-Mounted Hydrophones in Conjunction With a Digital Tag (DTAG) Recording. DTIC Document.

Ward, J., L. Thomas, S. Jarvis, N. DiMarzio, D. Moretti, T. A. Marques, C. Dunn, E. Hartvig, and P. Tyack. 2012. Passive acoustic density estimation of sperm whales in the Tongue of the Ocean, Bahamas. Marine Mammal Science **28**:E444-E455.

Watkins, W., and W. Schevill. 1975. Sperm whales (Physeter catodon) react to pingers. Pages 123IN1125-1124IN2129 *in* Deep Sea Research and Oceanographic Abstracts. Elsevier.

Watkins, W. A., and W. E. Schevill. 1972. Sound source location by arrival-times on a non-rigid three-dimensional hydrophone array. Pages 691-706 *in* Deep Sea Research and Oceanographic Abstracts. Elsevier.

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

Watwood, S. L., P. J. Miller, M. Johnson, P. T. Madsen, and P. L. Tyack. 2006. Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*). Journal of Animal Ecology **75**:814-825.

Wenz, G. M. 1962. Acoustic ambient noise in the ocean: Spectra and sources. The Journal of the Acoustical Society of America **34**:1936-1956.

Wiggins, S. M., K. E. Frasier, E. Henderson, Elizabeth, and J. A. Hildebrand. 2013. Tracking dolphin whistles using an autonomous acoustic recorder array. Journal of the Acoustical Society of America **133**:3813-3818.

Wiggins, S. M., M. A. McDonald, and J. A. Hildebrand. 2012. Beaked whale and dolphin tracking using a multichannel autonomous acoustic recorder. Journal of the Acoustical Society of America **131**:156-163.

Wilcock, W. S. 2012. Tracking fin whales in the northeast Pacific Ocean with a seafloor seismic network. The Journal of the Acoustical Society of America **132**:2408-2419.

Wilcock, W. S., R. Wade, D. K. Mellinger, S. L. Nieukirk, M. Weirathmueller, and E. Frederickson. 2015. Tracking blue whales with the Cascadia Initiative ocean bottom 51 seismograph network.*in* The 7th International DCLDE [Detection, Classification, Localization, and Density Estimation] Workshop, La Jolla, CA.

Zimmer, W. M. 2011. Passive acoustic monitoring of cetaceans. Cambridge University Press.

Zimmer, W. M., M. P. Johnson, P. T. Madsen, and P. L. Tyack. 2005a. Echolocation clicks of free-ranging Cuviers beaked whales (*Ziphius cavirostris*). The Journal of the Acoustical Society of America **117**:3919-3927.

Zimmer, W. M., P. L. Tyack, M. P. Johnson, and P. T. Madsen. 2005b. Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis. The Journal of the Acoustical Society of America **117**:1473-1485.

Zimmer, W. M. X. 2013. Range estimation of cetaceans with compact volumetric arrays. Journal of the Acoustical Society of America **134**:2610-2618.