Effectiveness of Navy lookout teams in detecting marine mammals

Final report on HDR Task Order 002/1000300001856

Authors:	C. Oedekoven and L. Thomas
Report Code:	CREEM-24289-1
Date:	22 March 2022

Please cite this report as: Oedekoven, C. and L. Thomas (2022) Effectiveness of Navy lookout teams in detecting cetaceans. Report number CREEM-24289-1. Provided to HDR Inc, 22 March 2022 (unpublished).

Document Control

Please consider this document as uncontrolled copy when printed

Rev.	Date.	Reason for Issue.	Prepared by	Checked by
1	10/Dec/2021	Draft sent to HDR for review	CO/LT	LT/CO
2	22/March/2022	Revised version sent to HDR	CO/LT	LT/CO

Executive Summary

The United States (US) Navy uses lookouts (LOs) to detect objects in the water in the vicinity of ships. One class of object that LOs are trained to detect is marine mammals[1]; this forms an important component of the Navy's procedures for marine mammal mitigation during training activities[2]. As well as dedicated lookouts, detections of marine mammals may also be made by other members of the ship's crew such as officers on the bridge ("watchstanders") or sonar technicians, although acoustic detections require visual confirmation. We refer to these personnel together as the "lookout team" (LT). The primary goal of this project was to determine how effective LTs are at detecting marine mammals before they entered a defined set of mitigation ranges during mid-frequency active sonar training activities. These ranges were 200, 500 and 1,000 yards. A secondary goal was to compare this effectiveness with that of trained marine mammal observers (MMOs).

In collaboration with Navy environmental personnel, we developed a field protocol for at-sea experiments, where MMOs set up trials by locating marine mammals around Navy ships training with mid-frequency active sonar and determined whether these animals were detected by the LT. We also developed new analytical methods that allow estimation of the probability of animals approaching to within a specified mitigation range without being detected (probability of remaining undetected, PrU). These methods include a model for the surfacing pattern of animal pods[3], and for the range-dependent probability of detecting a pod each time it surfaces. Crucially, the methods allow us to account for the possibility that animal pods may remain undetected by both MMOs and the LT. The methods are flexible in allowing for various patterns of animal surfacing and various experimental configurations (in terms of communication between MMO and LT positions and whether repeat surfacings of the same pod are recorded). They are, however, simplistic in assuming that there is no measurement error (in surfacing location, taxon designation and whether duplicate detections are correctly assigned), that pods only move in the vertical plane (i.e., there is negligible horizontal movement during the period when the pod is within observation range), and that the ship moves at constant speed in a straight line. We tested the new analytical methods using computer simulation and found they generally produce unbiased estimates when the model assumptions are met, although in some circumstances (including those in our at-sea study) it is not possible to estimate both detectability and surfacing pattern; in this situation if the parameters governing surfacing pattern are known then unbiased estimation of detectability, and hence PrU, is possible.

A total of 27 embarks were conducted between 2010 and 2019, mostly on destroyer class ships. These generated 716 valid sightings of animal pods. Each sighting consisted of one or more detection of a marine mammal pod by the MMO and/or LT positions; to be valid there had to be enough information recorded to derive a taxonomic code at the level needed for analysis (see below), pod size and, for each detection, a location (relative to the ship) and an observer position. There were no valid acoustic detections, and so all LT detections were generated by the LOs or watchstanders. Some species of small cetacean are known to approach ships and "bowride"; after discussion with Navy environmental personnel it was decided to exclude detections of pods

[1] Seals and turtles are also included in mitigation for various activities; however, they were not included in this study.

[2] The Navy's required mitigations for each training activity are described in each Letter of Authorization (LOA), and lookout configurations are dependent on the type of ship and training activity (see AFTT and HSTT Training LOAs; Section 6(a)(2)) (NMFS 2019, 2020).

[3] We use the term "pod" to refer to a group of one or more marine mammals. This term is typically used only for cetaceans but, as we document lower down in the report, there were not enough pinniped detections to include them in the analysis.

observed during the sighting to engage in bowriding behavior. There were 46 such sightings, with first detections predominantly made at close ranges. After excluding these, 670 sightings remained.

Our data collection protocol asked MMOs to prioritize new sightings over repeated detections (resights) of an already-sighted school, and so resights were not recorded consistently. We therefore used analytical methods that require data only on the first detection of a pod by each position. Analysis at species level was not possible because of limited sample size, and because many sightings were not identified using a taxonomic code that refers to species, but instead to a higher taxonomic level such as "large whale" or "dolphin"[4]. We therefore divided the data into four groups according to similarity in surfacing pattern and detectability: rorquals (i.e., large baleen whales), sperm whales, small cetaceans in small pods (6 or less) (SCSP) and small cetaceans in large pods (more than 6) (SCLP). We assumed the parameters governing surfacing pattern for each group were known, and we used values derived from the literature. For the sperm whale group, for which there were only two sightings, we used the detectability parameters estimated for rorquals. There were not enough detections of pinnipeds for us to estimate range-dependent probability of detection from the detection data and, unlike sperm whales, we elected not to use the estimated detectability parameters from one of the other groups; hence, our results only cover cetaceans.

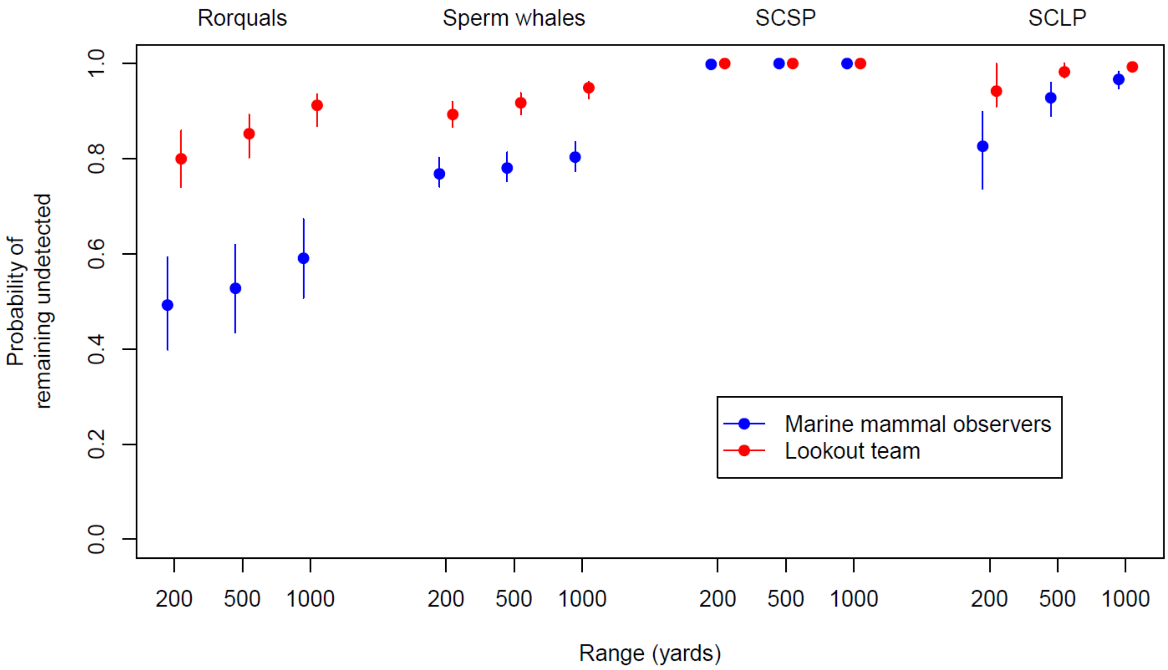
Before undertaking the modelling we performed some exploratory analyses, including calculating a simple distance-specific index of effectiveness at 200, 500 and 1,000 yards (yds) for rorquals, SCSP and SCLP. For this analysis, we quantified LT effectiveness as the number of pods detected by the LT *before* they enter within the mitigation range divided by the total number of pods thought to have entered within the mitigation range (as estimated by the number seen by the LT or MMOs within a given distance of the ship's track). We speculate that this provides an upper bound on absolute effectiveness, because it does not take account of pods that pass through the mitigation zones undetected by either position. Estimated effectiveness was highest for rorquals: 0.35, 0.21 and 0.13 at 200, 500 and 1,000 yds for the LT and 0.74, 0.70 and 0.54 respectively for MMOs. It was lowest for SCSP: 0.03, 0.03 and 0.02 respectively at 200, 500 and 1,000 yds for the LT and 0.25, 0.29 and 0.14 respectively for MMOs. The estimates for SCLP were similar to SCSP for the LT but higher than SCSP for MMOs.

Results from the modelling analysis to obtain PrU are summarized graphically in the figure on the next page. For each group, we estimated PrU at 200, 500 and 1,000 yards (yds). Please note that, although the results are quoted at these ranges, all of the data from each taxonomic group (including data beyond 1,000 yds) was used in deriving the results with these models. For rorquals the estimated PrU at 200 yds for the LT was 0.80 (95% confidence interval (CI) 0.74-0.86), rising to 0.91 (95% CI 0.87-0.94) at 1000 yds. PrU is the complement of effectiveness, so estimated absolute effectiveness was 1-0.80=0.20 at 200 yds and 1-0.91=0.09 at 1,000 yds. As expected, these values are slightly lower than the simple distance-specific index of effectiveness quoted in the previous paragraph (and this pattern held true for all such comparisons). MMOs were estimated to be considerably better, with PrU at 200 yds of 0.49 (95% CI 0.40-0.59) and at 1,000 yds of 0.59 (95% CI 0.51-0.67).

Taking the estimated detectability parameters and applying them to sperm whales, where time spent underwater is considerably higher, led to PrU for the LT of 0.89 (95% CI 0.87-0.92) at 200 yds and 0.95 (95% CI 0.93-0.96) at 1,000 yds. MMO PrU for sperm whales was 0.77 (95% CI 0.74-0.80) at

[4] A full list of taxonomic codes is given in Appendix A. One reason that identification to species level was sometimes not possible was that, unlike many research cruises, Navy ships did not approach pods in order to confirm species identification.

200 yds and 0.80 (95% CI 0.77-0.84) at 1,000 yds. Hence, in this case the difference between LT PrU and MMO PrU was smaller because the long dive times place an insurmountable constraint on any visual observation position, no matter how good.



Estimated probability that a pod of marine mammals of the taxonomic group shown along the top remains undetected by the Navy lookout team (blue) or marine mammal observers (red) at ranges of 200, 500 and 1000 yards from the ship. Dots show estimates and vertical lines give 95% confidence limits. Note that the sperm whale results assume their detectability while on the surface is the same as rorquals.

For small cetaceans, many of the first detections by both LT and MMO positions were at very close ranges, well within the smallest mitigation range of 200 yds, even after bowriding pods were removed. Because of this, for the SCSP group, the estimated PrU was close to 1 at all mitigation ranges tested and for both positions. We speculate that this result was caused by a combination of (a) genuinely low detectability combined with the surfacing pattern of this group, (b) fast and possibly responsive movement (attraction to the boat) by some pods, which violates a model assumption, (c) some rounding of detection distances and possibly angles, which violates another model assumption. For the SCLP group, which are assumed to have a surfacing pattern that makes them more available for detection, results improved slightly compared to the SLSP group. Estimated LT PrU for this group was 0.94 (95% CI 0.91-1.00) at 200 yds and 0.99 (95% CI 0.99-1.00) at 1,000 yds. The equivalent estimates for MMOs were 0.83 (95% CI 0.74-0.90) at 200yds and 0.97 (95% CI 0.95-0.98) at 1,000yds. Overall, for small cetaceans, we conclude that PrU is high (and hence effectiveness low) across pod sizes, caused by a combination of low detectability of small pods and possibly responsive movement of some taxa within the small cetacean groups.

We summarize our findings as follows:

1. Based on the data and analyses presented here, the ship's lookout team (LT) have approximately an 80% chance of failing to detect a pod of large baleen whales (rorquals) before they come closer than a mitigation range of 200 yards. This probability of a pod remaining undetected (PrU) rises to 85% at 500 yards and 91% at 1,000 yards.
2. The marine mammal observers (MMOs) performed better for this taxonomic group: for example, the PrU at 200 yards was lower at 49%. Note that the MMO team consisted of two dedicated observers while the LT consisted varying number of LOs depending on the type of ship and the training activity the ship was engaged in.
3. For species (sperm whales) with longer dive times but the assumed same detectability as rorquals, the PrU for both LT and MMOs was estimated to be higher (e.g., 89% for LT and 77% for MMOs at 200 yards), with less difference between the LT and MMOs.
4. For small cetaceans the majority of first detections of a pod (particularly those made by the LT) took place at very close range regardless of pod size. Estimated PrU for small pods (1-6 individuals) was close to 100% for any range, while for large pods this probability was lower for 200yds at 94% for the LT and at 83% for MMOs and for 500 yds at 98% for the LT and 93% for the MMOs. Small cetacean pods are genuinely difficult to detect, but in addition a limitation of our model was that it assumed no horizontal movement while some small cetaceans are attracted to ships and can move quickly (although we excluded pods where bowriding behavior was noted explicitly). Despite this it seems clear that PrU is high for small cetaceans.
5. We did not estimate PrU for beaked whales as none were recorded in the surveys. However, given they are not as detectable as sperm whales but have similar dive patterns, we would expect their PrU to be higher than sperm whales.
6. Our analyses assumed that the average surfacing pattern is known for each taxonomic group and used values taken from the literature. In reality, surfacing pattern varies by species and will likely differ from literature values. We undertook some sensitivity analyses and found that results were largely the same, except for sperm whales where assumptions about dive pattern made some difference to the predicted PrU. Overall our findings are unlikely to differ substantially if uncertainty and heterogeneity in surfacing could be included. Deviation of ship trajectory from the straight-line constant-speed assumption will also have some effect on results, but ship trajectory was unknown to us.
7. If further data collection were envisaged in the future, we would encourage further revision and tightening of the data recording procedures, in collaboration with the analysts.
8. Further analytical developments could include incorporation of responsive animal movement, changing ship trajectory and measurement error.

Table of Contents

List of abbreviations

AFTT	Atlantic Fleet Training and Testing
CREEM	Centre for Research into Ecological and Environmental Modelling
DMMO	Data recorder marine mammal observer
EDA	Effective detection area
ESHW	Effective strip half-width
HSTT	Hawaii-Southern California Training and Testing
g(0)	Probability of detection at zero perpendicular distance from the direction of travel
IE	Index of effectiveness of an observation position (LT or MMO)
LMMO	Liaison marine mammal observer
LO	Lookout
LOA	Letter of authorization
m	meter(s)
min	minute(s)
LT	Lookout team
MMO	Marine mammal observer
PrU	Probability of remaining undetected (at some specified mitigation range from the ship)
SCLP	Small cetaceans in large pods (more than 6 individuals)
SCSP	Small cetaceans in small pods (6 or fewer individuals)
SOCAL	Southern California (offshore range)
SMMO	Surveying marine mammal observer
US	United States
yds	yards

Introduction

The United States (US) Navy undertakes mitigation for marine mammals during training activities as part of mitigation procedures designed to minimize risk to these animals. One component of this mitigation is the shipboard lookouts (LOs), who are part of the standard operating procedure that ships use to detect objects (including marine mammals and other animals) around the ship during operations. The LOs are an element of monitoring requirements specified by the National Marine Fisheries Service in the Letters of Authorization (LOAs) issued pursuant to the Marine Mammal Protection Act for the US Navy's training and testing. As well as dedicated LOs, detections of marine mammals may also be made by other members of the ship's crew such as officers on the bridge (watchstanders) or sonar technicians (although in the latter case visual confirmation is required). We refer to all these personnel together as the "lookout team" (LT). The primary goal of this project was to determine how effective the LTs are at detecting marine mammals before they enter a defined set of mitigation ranges. This was achieved undertaking a set of at-sea trials where we could compare LT observations with those made by teams of highly trained civilian or contractor Marine Mammal Observers (MMOs) who were stationed on board Navy ships training with mid-frequency active sonar. This setup enabled a secondary aim of determining how LT effectiveness compared with that of MMO teams.

History of project

This project was initiated in 2010, when researchers from the Centre for Research into Ecological and Environmental Modelling (CREEM) collaborated with marine biologists from the Navy to design a field protocol for the at-sea trials. The protocol was revised slightly (Burt and Thomas 2010) after the first four cruises. In parallel, these initial four cruises (which generated 125 sightings) were used as the basis for initial development of the required modelling approach (Rexstad and Thomas 2010a, Thomas et al. 2011). The new analytical methods constitute a substantial extension of previous methods for analysis of line transect survey data, and in turn formed the basis for two peer-reviewed publications (Langrock et al. 2013, Borchers and Langrock 2015).

Data collection continued, with eight cruises having been completed by Feb 2012, yielding 182 sightings. An analysis of these data was undertaken (Thomas et al. 2012), with the data divided into two functional groups: "large animals" (i.e., large whale species) and "small animals" (mainly dolphins). The main conclusion was that more data was required for reliable results; it was recommended that, if possible, data collection was focused on a single ship type (destroyer) to minimize heterogeneity and location (SOCAL) to maximize detections per cruise. Notwithstanding the small sample size, the preliminary analysis indicated that probability of an animal pod (group of 1 or more animals) coming to within close range of the ship and yet remaining undetected (here denoted "PrU") by the LT may be quite high – for example estimated PrU at 100 meters (m) (109 yards, yds) ranged from 0.74 to 0.94 depending on assumptions about animal surfacing behavior.

Further data collection was undertaken between 2012 and 2019, with the data collection protocol receiving a further round of revisions by Navy personnel (Department of the Navy 2016) – this version is included as Appendix B. There are now 27 embarks complete, with 716 valid sightings (Table 1; a valid sighting is where one or more detection is made by the MMOs and/or LT, and all required information was recorded, such as taxon, distances, etc. – see Methods). The Navy has requested an analysis of observer effectiveness based on these data. After discussion with Navy personnel, the three mitigation ranges for calculating PrU used in this report are 1,000, 500 and 200 yds (914, 457 and 183 m). In other words, we seek to quantify the probability of a marine mammal school coming to within 1,000 yds of the ship without being detected, 500 yds without being detected, and 200 yds without being detected.

Report overview

In the Methods section, we briefly summarize the field survey protocol and give a summary of the modelling approach developed. We have further extended the approach used in Thomas et al. (2012) and Borchers and Langrock (2015) to account for the fact that, in the survey protocol, MMOs are aware of LT detections but not vice versa, and we describe the extension. Full technical details are given in Appendix C. We then summarize the simulation studies used to test these new methods, and how we applied these methods to the complete survey dataset.

In the Results section we present results from the simulation study, exploratory analyses of the survey dataset and our modelling results. Further results are given in Appendix C. We finish by discussing our results.

Methods

Field survey protocol

A detailed field protocol was developed and refined; this is reproduced in Appendix B. Here we give a summary.

During Navy mid-frequency active sonar training activities, dedicated LOs (in varying numbers depending on the type of ship) are positioned at the forward part of the ship, with additional lookouts serving as members of the bridge watch. On destroyers and cruisers, LOs may be located on the bridge wings. The Navy's required mitigations for each training activity are described in each LOA, and lookout configurations are dependent on the type of ship and training activity (see AFTT and HSTT Training LOAs; Section 6(a)(2)) (NMFS 2019, 2020). When sonar technicians are on duty during certain training activities, they report acoustic detections of marine mammals to the bridge. Visual detections are sometimes made by officers on the bridge (as well as other crew members). We refer to all of these observers together as the lookout team (LT). Officers on the bridge or in combat operations center were responsible for entering marine mammal records into a log, but we did not use this log in our experiments as it was not sufficiently detailed for our purpose. Instead, one of the MMO team acted as a liaison with the primary LO (see below) and kept detailed records of the detections made. During the experiments, every effort was made to allow the LT to operate as normal: MMOs were careful not to interfere with the work of the LT, or to cue them when the MMO team made sightings (see below). The only condition under which the MMO would inform the LT of a MMO sighting was when the animal was close aboard and a strike or close encounter might occur.

On each experimental embark, a team of four MMOs were embarked on a Navy ship (frigate, destroyer or cruiser) training with mid-frequency active sonar. The four rotated in duties and collected data during daylight hours and as much of the cruise as possible, including in all sea states and visibility conditions. At any one time during data collection, two MMOs were surveying for animals, one acted as liaison with the LT and one acted as data recorder or was resting. Where it is important to distinguish these duties, we refer to surveying marine mammal observers as SMMOs, the liaison as LMMO and the data recorder as DMMO. All MMOs were in radio contact with one another. MMOs kept standardized records of their survey effort, who was on which station, high-level ship activity (e.g., whether sonar was in use), weather and visibility (Appendix B, Section 4).

The SMMOs were generally stationed on the bridge wings of the ship. Their main functions were to detect and track marine mammals and determine whether detections made by the LT and reported by the LMMO were duplicates with sightings they had made. The SMMOs were careful to operate in such a way as not to cue the LT when SMMOs made detections. The SMMOs searched with naked eye and 7x50 binoculars; the port observer searched from about 5 degrees starboard to abeam on the port side while the starboard observer searched from about 5 degrees port to abeam on the starboard side. On detecting an animal (or pod of animals) a set of information was recorded (see Section 4 of Appendix B) including taxonomic code, range, bearing, pod size and behavior. The taxonomic code allowed for identification at a range of taxonomic levels, from high-level ("large whale", "dolphin", etc.) down to species – a full list is given in Appendix A. One reason that identification to species level was sometimes not possible was that, unlike many research cruises, Navy ships did not purposefully approach marine mammals to confirm species identification. As part of the protocol, SMMOs attempted to track the pod until past abeam and record each resight where possible; however, under the protocol they were asked to give higher priority to detecting new animals and so tracking was not done consistently. The reason for prioritizing sighting new pods is that each new sighting by the SMMOs sets up a new "trial" for the LT, which is either a success (LT detects school) or failure (LT does not detect school) giving direct information relevant to the primary goal of determining observer effectiveness. On the other hand, tracking already-sighted pods helps to identify whether a later LT detection is a match for an already-sighted SMMO pod and so is useful in determining duplicates; consistent tracking data, if gathered, can also be used in the estimation procedure (see below). Overall, prioritizing direct information on observer effectiveness was judged to be more important where a choice had to be made.

The LMMO was stationed on the bridge to observe the OT. Depending upon the configuration of the lookouts, the LMMO may be positioned either inside the pilot house, on one bridge wing, or moving between the bridge wings and the pilot house. Their main function was to record information on the first detections of pods by the LT, particularly the LO. Information on resights were recorded if possible. The LMMO passed relevant information on the detection to the rest of the MMO team as soon as possible – this included the LT's estimation of range, bearing and taxon. In some cases the LT made detections before the SMMOs.

The MMO team were responsible for determining whether a surfacing seen by either observation position (SMMO or LT) was a resight of a previous pod and whether a surfacing seen by the LT was a duplicate of a sighting by the SMMOs. Determinations were classified as either definite (at least 90% likely), probable (50-90% likely) or remote (less than 50% likely – although in practice this category was not used).

In the following we use the term "sighting" to refer to one or more detections of the same pod by either or both positions. We refer to an individual surfacing that is detected by either or both positions as a "detection". After the first detection of a pod by a position (MMO or LT), second and subsequent surfacings that are detected by the same position are referred to as "resights".

Exploratory data analysis

Initial analyses of the data were undertaken several times and are documented in previous reports (Rexstad and Thomas 2010a, b, Thomas et al. 2011, Thomas et al. 2012). Here we undertook the following analyses in advance of modelling to estimate PrU, which is described in subsequent sections.

Note, in this and subsequent sections, unless stated otherwise, we use the term "marine mammal observer" (MMO) to mean the surveying marine mammal observers (SMMOs).

Some data formatting and cleaning was required. Data files often had different structures and column headers. It was not clear how reticle readings were converted into distances (even though the formula was given in most files, the values could not be reproduced). Hence, we used the NOAA conversion to calculate distances from reticles (Kinzey et al. 2000) using the height of the ship given in the data files. Sightings and resights were often not clearly labeled and matches of sightings or individual detections between MMOs and LT were often described verbally in the comment section. This required manual inspection of each record and assigning corrected sighting and resight numbers. Data analysis required unique identifiers for each sighted pod and each resighted surfacing of the respective pod and which observer team had detected these. Information on who may have cued who also needed to be obtained from the comment section.

We removed sightings where it was not clear which position made each detection, where it was not possible to derive the perpendicular or horizontal distance, taxonomic code or pod size (the latter two were often derived from other detections within the same sighting). We removed a single detection where the observer was recorded as "sonar" for which the species was logged as "Biologics" and no corresponding visual confirmation in the same sighting existed which may have aided to improve the species identification. The only detection reported by the sonar technician, logged as an un-identified marine mammal, was also not included in the analysis as this taxonomic code could not be attributed to any of the four groups with similar surfacing behavior described above. We removed one detection (of a humpback whale in Hawaii) where the observer was recorded as "Aerial" because this detection came from a contracted monitoring aircraft conducting a simultaneous survey, not a Navy asset. We also removed any detections made after the pod passed abeam of the ship because the pod was then beyond its closest point of approach; however we did use these detections to help with taxonomic identification of any previous detections made during the same sighting. Remaining sightings and detections-within-sightings are referred to as "valid".

In matching detections to form a single sighting, we used all cases where the MMOs judged the match to be "definite" or "probable". We planned to discard all cases where the match probability was judged "remote", but in practice there were none of these in the dataset.

One piece of information noted by the observers relates to animal behavior, and this included cases where sightings were of pods observed to actively engage in "bowriding" behavior at some point during the sighting. Such pods were typically first observed at close range. After discussion with Navy environmental it was decided to exclude these detections (see Discussion). Other pods noted to be "closing" on the ship were not excluded because it was not known whether they were responding to the ship or would have been travelling in that direction without the ship's presence.

A simple index of the effectiveness (IE) of an observation position is the proportion of pods known to be present that were detected by that position. This is not an absolute measure of effectiveness as it does not account for pods undetected by either position, nor does it incorporate range-dependent detection. Denoting the number of marine mammal sightings by LTs, MMOs and both combined as n_{LT}, n_{MMO} and n respectively, we calculated the indices of effectiveness as

$$IE_{LT} = \frac{n_{LT}}{n}$$

$$IE_{MMO} = \frac{n_{MMO}}{n}$$

As an initial view of sighting location relative to the mitigation ranges (200, 500 and 1,000 yds), we plotted first detection location for each sighting separately for the LT and MMOs. If the majority of first detections are made within the mitigation ranges, this implies low effectiveness to detect pods before they enter these ranges. However, this does not account for pods that are unobserved, which provides motivation for the more sophisticated modeling described in the next section.

To provide a quantitative summary of the information in the plots, we extended the index of effectiveness to incorporate the spatial location of detections and hence derive a simple index of effectiveness in detecting pods before they enter the mitigation ranges. Under the assumption that pods do not change their perpendicular distance from the ship, then any pods first sighted within perpendicular distance x will pass within range x of the ship as the ship moves forward and past the pod. Hence all detections of pods made by either position within perpendicular distance x, denoted n_x, form a set of "trials". The number of these first detected by the LT at ranges r greater than x, denoted $n_{LT,r>x}$, is the number of "success" for the LT because each of these pods was detected before it entered the mitigation range. Similarly, the number of trials first detected at ranges greater than x by the MMOs, $n_{MMO,r>x}$ is the number of MMO "successes". We calculated the distance-specific indices of effectiveness as

$$IE_{LT}(x) = \frac{n_{LT,r>x}}{n_x}$$
$$IE_{MMO}(x) = \frac{n_{MMO,r>x}}{n_x}$$

for $x = 200$, 500 and 1,000 yds. This simple measure is likely an over-estimate of effectiveness because the n_x comprises only the pods known to be present because they were detected by one or both positions – it does not account for pods that were present but not detected by either position. (It does, however, assume that the perpendicular distance a pod is first detected at is their perpendicular distance when the ship passes abeam – see Discussion.) Again, this provides motivation for the methods described next.

Estimating range-dependent probability of remaining undetected (PrU)

Here we give an outline of the approach developed; full details are in Appendix C. Note that the methods are also of potential use in shipboard line-transect surveys for estimating density of marine fauna, where there is also the problem of estimating detectability in the presence of non-continuous availability.

Conceptually, the modeling approach used can be divided into two parts: a process model describing the horizontal and vertical movement of pods relative to the ship transporting the observers, and an observation model describing the way that observers sight and record pods given the pods' horizontal and vertical location. We describe each in turn.

For the vertical movement component of the process model, animals are assumed to be in one of two behavioral states: (1) diving relatively shallowly where they may on occasion be at the surface (e.g., to breathe), or (2) diving relatively deeply such that they will not surface. In state 1, they are assumed to surface at random intervals (drawn from an exponential distribution) with an average rate that is a model parameter, λ; while in state 2, they do not surface. Surfacings are assumed to be instantaneous. The probability of switching from one state to the other at any instant only depends upon the current state, and is governed by two model parameters, q_{12} and q_{21} (the state switching rates). This framework gives the flexibility to allow for clusters of surfacings followed by extended periods underwater. It also includes two special cases: intermittent and continuous availability. Intermittent availability is where one or more member of the pod is on the surface for extended periods between dives, and can be accommodated by assuming pods are at the surface while in state 1; continuous availability is where one or more member of the pod is always at the surface and can be accommodated by additionally removing the state 2 component from the model (or by setting the switching rate q_{21} to a suitably high value, so that negligible time is spent in state 2).

In theory, given sufficient data, it is possible to estimate the availability parameters q_{12}, q_{21} and λ as well as the observation model parameters described below. However, in practice, these parameters are only accurately identifiable in a limited range of circumstances – for example with large datasets with reliable resighting data and an amenable detection process (see Borchers and Langrock 2015). Borchers and Langrock (2015) distinguished between no assumed knowledge of the availability parameters (where all three parameters are estimated), partial knowledge (where q_{12} and q_{21} are assumed known and λ is estimated) and full knowledge (where all three parameters are assumed known). In the work carried out as part of this current contract we assumed full knowledge and used parameter values from the literature (see "Application to survey data", below).

The horizontal component of the process model is simplistic. Schools are assumed not to move horizontally, while the ship is assumed to move in a straight line at a constant speed of 12 knots. This means that pods do not change their location relative to the ship in the *x*-direction (i.e., perpendicular to the direction of travel of the ship), but their position in the *y*-direction (parallel to the direction of travel of the ship) gets smaller at a rate of 12 nautical miles per hour (with zero being abeam).

The observation model assumes that the LT and MMO positions each have separate two-dimensional hazard functions, describing the probability of detecting a surfacing given its (*x*,*y*) location. Animals beyond some specified distance from the observer are assumed to be undetectable, as are animals below the surface. The hazard functions have parameters that fix their shape; the function used in this report, the inverse power hazard function, has three parameters, α, β and γ, and is of the form:

$$h(x,y) = \frac{\alpha\beta^{\gamma}}{(\beta^2 + x^2 + y^2)^{\gamma/2}}$$

(see figures in Results for example fits). (An alternative hazard function, which also has three parameters, is the exponential power function – see Borchers and Langrock 2015.) Hence there are six parameters in the observation model: three-parameters for the LT hazard function and three parameters for the MMO hazard function. It is assumed that the location of sighted animals is recorded without error, that duplicate identification is certain (i.e., that there are no mistakes in determining whether a surfacing was detected by MMO alone, LT alone, or both), and that the object detected has been correctly identified to the taxonomic level used in the analysis. The model also assumes some maximum detection range beyond which there are no detections, and can optionally use a perpendicular truncation distance w (this latter is just like standard line transect analysis).

The model framework is capable of accommodating the first detection of a sighting by either or both observation position and also resights by either or both positions of later surfacings. It is likely that the detection function for an observation position changes significantly after that position becomes aware of a school, and this can be accommodated by allowing the detection function parameters to differ between initial detections and resights. As noted previously, our survey protocol meant that resights were not collected consistently, and so in the data analyses performed here we used a version of the model that utilizes information only from first detections by each position.

Initial development of this model was documented in Thomas et al. (2011, 2012) with more developed versions given in Langrock et al. (2013) (which considered only one observation position) and Borchers et al. (2015) (which included multiple independent observation positions). The latter assumed the two positions operated fully independently, so further development was required to accommodate the one-way dependence in our survey protocol: the MMO position could make detections independently of the LT, but when the LT made a detection the MMO position was informed, creating a dependence. The additional development to accommodate this one-way dependence is described in detail in Appendix C. When only first detection data are used, then each pod encounter can yield one of four kinds of sighting record: (1) MMOs detected the pod but LT did not; (2) MMOs detected the pod and LT detected it on the same surfacing; (3) MMOs detected the pod but LT missed that surfacing; LT detected the pod on a subsequent surfacing (before the pod went abeam); (4) LT detected a pod but MMOs missed it until the LT brought it to their attention.

The methods require specification of a forward truncation distance, beyond which detection probability is assumed to be 0, and a perpendicular distance, w, which functions like the truncation distance in line transect analysis. In the analysis of survey data reported here we used the furthest observed forward distance and furthest observed perpendicular distance as these two truncation points. Hence all valid detections were used in the analysis. Results reported here are insensitive to the choice of truncation distance in the sense that using much larger distances would have yielded the same results (but taken more computer time to run) because the estimated detection hazard at the truncation points was extremely small.

The model is fitted to data using maximum likelihood. This yields parameter estimates and estimates of uncertainty in the model parameters.

Given a fitted model, it is straightforward to estimate the probability of remaining undetected (PrU) to a given set of mitigation ranges. This involves integrating the surfacing and detection models in the forward distance direction from the maximum detection range up to the mitigation range, and in the perpendicular distance direction from the truncation distance w to zero.

The same technique can be used to calculate the probability of detection up to a given forward distance. An important application of this technique in the context of line-transect surveys is to calculate the probability of detecting an animal between the maximum detection distance and a forward distance of 0 (i.e., when the animal goes abeam of the ship, for a given perpendicular distance). This equates to the perpendicular detection probability, $g(x)$. Of particular interest is $g(0)$, the probability of detecting an animal at zero perpendicular distance. We use this as one metric to compare different detection functions in the analyses that follow. Another metric related to line-transect surveys is the effective strip half-width (ESHW, i.e., the perpendicular distance within which as many animals are missed as are detected outside that distance) (Buckland et al. 2001).

Uncertainty in the derived quantities PrU, ESHW and g(0) was obtained from a parametric bootstrap procedure. We sampled 1,000 random parameter estimates of the model, assuming estimates were distributed according to a multivariate normal distribution with mean equal to the maximum likelihood estimates and using the estimated covariance matrix. For each sample, we calculated PrU (at each mitigation range), ESHW and g(0). From these 1,000 estimates of each quantity, we estimated variance as the empirical variance in values and used the percentile method to obtain confidence intervals.

We evaluated the goodness-of-fit of the fitted models by comparing the observed number of detections in a set of distance intervals with those predicted under the model. We made visual plots for both MMOs and the LT showing observed and expected detections in the forward and perpendicular distance directions, and also in a set of range bins. For the latter we also undertook χ^2 goodness-of-fit tests.

Simulation studies

Simulation is a useful way to examine the performance of a new statistical method under conditions where the truth is known. Many random replicate datasets are generated from a model and then analyzed; results can then be used to compute various metrics such as percentage bias (i.e., the mean of the estimates minus the true value, all divided by the true value and multiplied by 100).

A small simulation study of the initial model was documented in Thomas et al. (2012), and more extensive studies given in Langrock et al. (2013) and Borchers and Langrock (2015). These established (among other things) that the analysis method is capable of estimating:

- both availability parameters and ESHW with low bias from single or independent-observer double platform (i.e., two independent observation position) data for a combination of parameters that yield a peak in detections away from $y = 0$;
- both availability parameters and ESHW with low bias from independent-observer double platform data where data are collected on resights, for a combination of parameters that yield a peak in detections close to $y = 0$;
- ESHW and potentially surfacing rate λ with low bias from independent-observer double platform data when only first detection data are collected, for a combination of parameters that yield a peak in detections close to $y = 0$, if the other two availability parameters are known.

Some other scenarios demonstrated considerable bias – for example independent-observer double platform data with only first detections recorded when there is a peak in detections close to y=0.

For this report, we extended the modeling approach to account for one-way independence between MMO and LT positions and using first detections by either observer only; hence we wished to test this approach through simulation, for the scenario where availability parameters are known. We also wished to examine any potential bias in estimation of PrU, which was the main focus of the observer effectiveness study.

To this end, we constructed two simulation scenarios. In both cases the pod availability pattern was taken from a preliminary analysis of the rorqual dataset (see below), yielding $q_{12} = 0.00104$ m^{-1}, $q_{21} = 0.00064$ m^{-1} and $\lambda = 0.0090$ m^{-1} (calculated as 1 over the distance travelled by the ship between events). With an assumed maximum detection range of 15,726 m and ship speed of 12 knots, this yielded an expected x surfacings per pod while the pod was in the visual field of the observers. The detection hazard parameters were also taken from the rorqual analysis, which yielded for the LT position $\theta = (0.05,\ 8103.08, 5.47)$ and for the MMO position $\theta = (0.10, 8103.08, 6.69)$. The resulting detection hazards are shown in Figure 1. In the first scenario, we assumed full independence between positions, and retained all detections. In the second scenario, which is closer to that in our survey data, we assumed one-way independence between the MMO and LT positions, and we retained only first detections for each sighting. The cumulative distribution of first detection distances is shown in Figure 2 (green dashed lines). The interpretation of this plot is that each point gives the probability of detecting a pod at that perpendicular distance *x* by the time it gets to forward distance *y* from the ship – hence the values at *y*=0 give the probability of detecting a pod at perpendicular distance x by the time it passes abeam of the ship. The value at *y*=0 and *x*=0 is what is referred to as *g(0)* in the line transect literature, and in this simulation is approximately 0.52 for the LT position and 0.71 for the MMO position.

To generate a simulated dataset, we repeated the following until we had 300 sightings. We first generated a random perpendicular distance between 0 and $w = 6{,}500$ m. We then generated a random availability pattern and for each surfacing used the detection hazard for each position to determine whether the surfacing was detected. For scenario 1 we retained all detections, while for scenario 2 we retained only the first detections by each position.

Each simulated dataset was analyzed using the independent observer all detections model (Scenario 1) and the one-way independence first detection model (Scenario 2), in both cases assuming the availability parameters were known. We repeated this exercise for 100 replicate simulations and calculated percentage bias in estimates of PrU at 200, 500 and 1,000 yards, as well as ESHW and g(0).

Application to survey data

Previous reports (Rexstad and Thomas 2010a, b, Thomas et al. 2011, Thomas et al. 2012) undertook preliminary analyses of data available at that time. The most recent (Thomas et al. 2012) divided the data taxonomically into large and small cetaceans. For the former, a preliminary version of the analysis methods used here were employed assuming three sets of availability parameters corresponding to long shallow dives, intermediate and long deep dives. Estimated PrU at 100 m (109 yards) for the LT was 0.74, 0.79 and 0.94 for the three scenarios, while for the MMOs the estimated PrU at 100m was 0.38, 0.48 and 0.81. PrU at larger range was lower. Small cetaceans were assumed, optimistically, to be continuously available, and this yielded LT PrU at 100m of 0.75. More realistically (for some species), an instantaneous availability model with assumed expected times in deep and shallow dive states of 2 minutes gave an LT PrU of 0.95.

Given the larger dataset available for this report, we undertook a re-evaluation of the taxonomic level at which observer effectiveness could be evaluated. To this end, we took guidance from the NOAA hierarchical classification system of sighting-categories (KInzey et al. 2000) to construct a similar hierarchy using the taxonomic codes used during the surveys (Appendix A). The majority of whales identified at lower levels than the level 5 category WHALE (unidentified whale) were rorquals (Table 2). The 277 sightings in this group represented 41% of all sightings. Most of these identified to species were humpback whales (84 sightings), but also including Balaenoptera sp. (29), Minke whales (3), blue whales (16) and fin whales (1). Nearly 50% in this group were not identified to species (94 sightings of unidentified whales and 50 sightings of unidentified large whales). As almost all sightings identified to species were rorquals, we grouped all these sightings (with the exception of two sperm whale sightings) into the rorqual group for analysis. Regarding availability pattern for this group, as the majority were classified as humpback whales, we assumed that humpback whale availability parameters would form a reasonable proxy for the rorquals group. We assumed that members of the rorquals group would spend on average 2.4 min near the surface (state 1) and 4.2 min diving (state 2) (Dolphin 1987). While in state 1, members of this group would surface on average once every 0.3 min. Hence, parameters q_{12} and q_{21} were calculated using: $q_{12} = \frac{1}{2.4*12*1852/60}$, $q_{21} = \frac{1}{4.2*12*1852/60}$ and $\lambda_1 = \frac{1}{0.3*12*1852/60}$, respectively.

In grouping rorqual species together, we excluded one sighting with another large whale species: the single sperm whale sighting. As an ESA-listed endangered species it was of interest to obtain an approximate estimate of PrU. On the assumption that the detection hazard for this species was the same as that of the rorquals group, we combined the estimated rorqual detection hazard with availability information derived from Drouot et al. (2004). We assumed that sperm whales on average spend 9.1 min near the surface (state 1) and 44.8 min diving (state 2); while near the surface we assumed they produce on average 4.6 blows per min.

For small cetacean species, one important factor affecting both availability and detectability is pod size. After examining the distribution of observed pod sizes, we divided the small cetacean sightings into those with an estimated pod size of 6 or fewer (“small cetacean small pod”, SCSP) and those with a pod size of more than 6 (“small cetacean large school”, SCLP). We analyzed these two groups separately, for the SCSP group assuming instantaneous availability and for SCLP assuming intermittent availability. We used the study by Scott and Chivers (2009) to populate the parameters for the availability model: for both groups we assumed that the pods on average spend 0.99 min near the surface (state 1) and 1.26 min diving (state 2). For the SCSP group we assumed that on average every 0.1 min at least one animal of the pod was at the surface while the pod is in state 1 and hence, making the pod available to be detected. For the SCLP group we assumed that at least one member of the pod was at the surface at all times while in state 1.

While the values described above represented the settings for the main analyses, we conducted a sensitivity analysis for each group. As part of this, we used other suitable values from the literature to populate the availability model (see Appendix C for details).

Results

Exploratory analysis of survey data

After data cleaning there were 716 valid sightings in the data set. Of these 46 were recorded as including bowriding behavior and were therefore excluded, giving 670 remaining sightings (Table 2). There were 5 sightings of animals recorded as “closing” but not recorded as bowriding during the sighting; these were retained. The remaining sightings were divided into the four groups described earlier (rorquals, sperm whales, SCSP and SCLP, 544 sightings) and other animal taxa outside these groups, including pinnipeds, turtles and fish (126 sightings) that were not analyzed further.

The rorquals group encompassed 277 sightings (Table 3), comprising 301 detected surfacings by either or both positions (including first detections and subsequent resights). Out of the 277 sightings, 212 were detected by the MMOs only, 21 by the LT only and 44 by both teams. This gives an index of effectiveness for the MMOs of $IE_{MMO} = 0.92$ and for the LT of $IE_{LT} = 0.23$, i.e., 4 times lower. Out of the 44 sightings detected by both positions, 19 were detected by the MMOs first, 5 by the LT first and 20 by both positions during the same surfacing. 104 encounters could be identified to species, 29 to genus and the remaining 144 to higher taxonomic levels.

The location of first detections of rorquals relative to the ship is shown in Figure 3. The percentage of first detections made by the LT within 200, 500 and 1,000 yds of the ship was 6.2%, 13.8% and 32.3%; the corresponding percentages for the MMOs were 0.8%, 3.1% and 13.3%. Comparing these values from the two positions, it is clear that the LT were focusing their search effort closer to the ship than the MMOs.

This finding is also reflected in the distance-specific indices of effectiveness (Table 4). Of all pods first detected within 200 yds perpendicular distance from the vessel’s track 74% of them were detected by the MMOs outside of the 200 yds mitigation range ($IE_{MMO}(200)$ = 0.74) while 35% were detected by the LT outside this range ($IE_{MMO}(200)$ = 0.35). Similarly, for the 500 yds distance, MMOs detected 70% of pods outside 500 yds range while the LT detected 21%; for the 1000 yds distance, MMOs detected 54% of pods outside 1000 yds range while the LT detected 13%.

As indicated earlier, there were only 2 sperm whale sightings so no exploratory analysis was performed.

The SCSP group included 178 sightings (Table 5), comprising 201 detections made by either or both positions (including first detections or subsequent resights). Out of the 178 sightings, 125 were detected by the MMO only, 20 by the LT only and 33 by both. This gives an index of effectiveness for the MMOs of $IE_{MMO} = 0.88$ and for the LT of $IE_{LT} = 0.30$, i.e., almost 3 times lower. Of the 33 encounters detected by both positions, five were detected first by the MMOs, four first by the LT and 24 by both positions during the same surfacing. 63 of the 178 sightings were identified to species, four to genus and 111 to higher taxonomic levels. We note that 20 sightings were excluded from the analysis as they were recorded as 'bowriding'; hence, the total number of sightings and detections included in the analyses were 178 and 201, respectively.

The location of first detections of SCSP relative to the ship is shown in Figure 4. The first detections were generally much closer to the ship than for rorquals. Indeed, many detections were made at distances that would indicate the pod was at or almost at the bow of the ship when first detected: 13.2% of LT first detections and 8.8% of MMO first detections were given recorded ranges of 10 yds or less. The percentage of first detections made within 200, 500 and 1000 yds of the ship was 52.8, 77.4, 88.7% for the LT and 32.3, 43.7, 74.1% for the MMOs.

The close ranges at which detections were first made meant that the distance-specific indices of effectiveness (Table 6) were lower for SCSP compared with rorquals. Of the pods first detected within 200 yds perpendicular distance from the vessel's track 25% of them were detected by the MMOs outside of the 200 yds mitigation range while only 3% were detected by the LT outside this range. For the 500 yds distance, MMOs detected 29% of pods outside 500 yds range while the LT detected 3%. For the 1000 yds distance, MMOs detected 14% of pods outside 1000 yds range while the LT detected 2%.

The SCLP group included 87 sightings (Table 7) and 136 detections by either or both positions. Out of the 87 sightings, 58 were by the MMOs only, nine by the LT only and 20 by both positions. This gives an index of effectiveness for the MMOs of $IE_{MMO} = 0.90$ and for the LT of $IE_{LT} = 0.33$, i.e., approximately 3 times lower. Of the 20 detected by both, six were detected first by the MMOs, one first by the LT and 13 by both simultaneously. Of the 87 sightings, 49 were identified to species, ten to genus and 28 to higher taxonomic levels. As for the SCSP group, pods recorded as bowriding were excluded from the analyses leaving 87 sightings and 94 detections in the analyzed data set.

The location of first detections of SCLP relative to the ship is shown in Figure 5. The percentage of pods first signed at very close range (10 yds or less) was again high for the LT at 17.2% for the LT but 0% for the MMOs. The percentage of first detections made within 200, 500 and 1000 yds of the ship was 48.3, 69.0 and 86.2% for the LT and 11.5, 24.4 and 46.2% for the MMOs. Combined with the very large proportion of detections made at close ranges, there was a significant tail of detections at larger ranges (Figure 5).

The distance-specific indices of effectiveness (Table 8) were generally somewhat higher for SCLP than the SCSP, although still very low for the LT. Of the pods first detected within 200 yds perpendicular distance from the vessel's track 41% of them were detected by the MMOs outside of the 200 yds mitigation range while only 6% were detected by the LT outside this range. For the 500 yds distance, MMOs detected 33% of pods outside 500 yds range while the LT detected 4%. For the 1000 yds distance, MMOs detected 31% of pods outside 1000 yds range while the LT detected 3%.

Simulation study

Visual comparison of observed vs expected estimates provides an informal means for checking the simulation. The true detection hazard from simulation scenario 2 (which is closer to our real data situation) for MMOs and the LT is shown on the top panel of Figure 1 and an example estimated hazard function from one simulated dataset is shown on the bottom panel. The two appear very similar. Likewise, Figure 2 shows the cumulative distribution of first detections in the 2-d plane as green dashed lines, and the estimated distribution from a single dataset as black lines and shading. They are again similar, although the black lines are within in the green dashed lines, indicating slight underestimation of detection probability at closer ranges. Further diagnostic plots are provided in Appendix C.

A more formal quantification of performance is the percentage bias calculated over the 100 simulated realizations. Bias in estimation of all quantities of interest was very low (<3%) for both scenario 1 (independent observers, all detections) and scenario 2 (one-way independence, first detection only) (Table 9).

Range-dependent probability of remaining undetected (PrU)

Full details of results, including additional diagnostic plots, goodness-of-fit test results and sensitivity analysis are given in Appendix C. An extended summary is given here.

The largest forward detection distance was 15,725 m and the largest perpendicular distance was 13,425 m; these were used as truncation distances in the analysis for all species groups.

The estimated detection hazard for the rorqual group was very flat and near zero in most areas for both positions, only exhibiting a sharp rise within small radii around zero radial distance from the ship (Figure 6). This increase in detection hazard was much more pronounced for MMOs compared to LT. The former also rose to much higher values at x=0, y=0 compared to the latter. The estimated cumulative distribution of first detections is shown in Figure 7. The value shown in Figure 7 at x=0, y=0 is the estimate of *g(0)* for this position, and this is also given in Table 10 together with the estimated effective strip half-widths (ESHWs) and the range-dependent probability of remaining undetected (PrU). For the LT, estimated PrU was 0.80 (95% confidence interval (CI) 0.74-0.86) at 200 yds, 0.85 (95% CI 0.80-0.89) at 500 yds, and 0.91 (95% CI 0.87-0.94) at 1000 yds. PrU for MMOs were lower (i.e., better): 0.49 (95% CI 0.40-0.59) at 200 yds, 0.53 (95%CI 0.43-0.62) at 500 yds and 0.59 (0.51-0.67) at 1000 yds. Overall these are around 1.6 times lower. Results from the χ^2 goodness-of-fit tests indicated that the fit of the rorquals model was poor for both MMO and LT data (Appendix C Table 12); visual inspection of diagnostic plots (Appendix C Figures 11 and 12) showed the model under-predicted MMO detections at close ranges in the perpendicular distance direction and over-predicted in the forward distance direction; for the LT the lack of fit appeared again to be under-prediction at close range in the perpendicular distance direction. The sensitivity analysis, which comprised a second run with alternative availability model parameters, produced almost identical results to the main analysis (Appendix C Table 11).

Taking the estimated detectability parameters for rorquals and applying them to sperm whales, where time spent underwater is considerably higher, led to the estimated cumulative distribution of distances shown in Figure 8, and detection statistics given in Table 10. As would be expected, predicted *g(0)* and ESHW are lower, and the PrUs are higher (Table 10). For the LT estimated PrU was 0.89 (95% CI 0.87-0.92) at 200 yds, 0.92 (0.89-0.94) at 500 yds and 0.95 (0.93-0.96) at 1,000 yds. MMO PrU for sperm whales was 0.77 (95% CI 0.74-0.80) at 200 yds, 0.78 (0.75-0.81) at 500 yds and 0.80 (95% CI 0.77-0.84) at 1,000 yds. Note that the difference between the LT and MMO PrUs was smaller in this case, because a major limitation on PrU comes from the dive behavior and that is the same for both positions. The sensitivity analysis, using two sets of divergent availability model parameters, produced somewhat different results (e.g., estimated PrU at 200 yds varying from 0.60 to 0.87 for MMOs and from 0.79 to 0.95 for the LT) (Appendix C Table 13).

For small cetaceans in small pods (1-6 individuals, SCSP), the estimated detection hazard was extremely "spiked" at small ranges (Figure 9). Most pods were estimated to remain undetected, and of those detected, the estimated cumulative distribution of first detections (Figure 10) showed that these detections are likely to be at very close ranges. The estimated PrUs were close to 1 for both observation positions and all mitigation ranges (Table 10). As with rorquals, the χ^2 goodness-of-fit tests indicated a poor fit for both MMO and LT data (Appendix C Table 17); however the diagnostic plots (Appendix C Figures 16 and 17) did not indicate any obvious problems, and showed that the large spike in detections in forward, perpendicular and radial distances appeared to be well captured by the model. The sensitivity analysis, which included 3 divergent sets of availability parameters, gave almost identical results to the main analysis (Appendix C Table 16).

For small cetaceans in large pods (7+ individuals, SCLP), the estimated detection hazard was also spiked at small ranges (Figure 11), although less so than for SCSP. The estimated cumulative distribution of first detections (Figure 12) indicated that most pods detected are first detected at distances within the mitigation ranges. For the LT estimated PrU was 0.94 (95% CI 0.91-1.00) at 200 yds, 0.98 (0.97-1.00) at 500 yds and 0.99 (0.99-1.00) at 1,000 yds. MMO PrU for SCLP was 0.83 (95% CI 0.74-0.90) at 200 yds, 0.93 (0.89-0.96) at 500 yds and 0.97 (95% CI 0.95-0.98) at 1,000 yds. Once again the χ^2 goodness-of-fit tests indicated a poor fit for both MMO and LT data (Appendix C Table 20); the diagnostic plots (Appendix C figures 20 and 21) showed that although the model did a good job of fitting the spike in detections at small forward and perpendicular distances, generally underfitted for ranges <6000 yds and overfitted the larger ranges. The sensitivity analysis, which comprised one set of alternative availability parameters, gave almost identical results to the main analysis (Appendix C Table 19).

Discussion

The main goal of this project was to quantify the effectiveness of the Navy LOs and other members of the LT in detecting marine mammals before they enter a set of specified mitigation ranges. To achieve this, the Navy expended considerable effort in deploying MMOs on board ships during training exercises, where they have generated data on their own detections and those of the LT. We developed the experimental protocol, and have developed new methods for analysis of the resulting data. One additional benefit of the new analytical methods is that they are applicable to estimating animal density from line transect surveys, which could potentially also aid in producing better estimates of marine mammal density for use in the Navy's Marine Species Density Database.

In analyzing the experimental data, we grouped species along taxonomic lines, aiming to put together taxa with similar sightability and diving behavior. This gave us four groups: rorquals, sperm whales, SCSP and SCLP.

Rorquals are large whales often with conspicuous blows, and can therefore be relatively easy to sight when surfacing compared with small cetaceans – although the mean pod sizes were small (means between 1-2 individuals, depending on the species, Table 3). Given the forward truncation distance used in our analyses (15,725m) and assumed ship speed (12 knots) a patch of water at perpendicular distance x=0 would be in view for 42 mins. During this time, using the assumed dive parameters used in the main analysis for rorquals (2.4 mins in state 1 with inter-surfacing interval 0.4 mins, 4.2 mins in state 2), the expected number of surfacings while in view is 37. We found that, for MMOs, the probability of detecting an animal at x=0, g(0), was 0.53 (95% CI 0.43-0.62). This is lower than the estimate of 0.92 derived by Barlow and Forney (2007) for "large whales (most baleen whales and killer whales)" on NOAA line transect surveys. However, Barlow and Forney acknowledged that their methods tended to produce over-estimates; also NOAA survey data is collected only in good survey conditions while the experimental trials reported here took place in all sea state and sightability conditions.

We found for rorquals in our exploratory analysis that the LT detected only 23% of the surfacings known to have taken place within visual range, while the MMOs detected 93% of them. This leads to a naïve conclusion that the LT is 4 times less effective than the MMOs at detecting roquals. However, this is not directly relevant to mitigation effectiveness as it ignores range (the LT are unlikely to be scanning for objects at large perpendicular distances, for example) and pods unobserved by both teams. A more refined exploratory analysis was undertaken, examining all detections made within 200, 500 and 1,000 yds perpendicular distance from the ship track (assuming the ship continued in a straight line) and determining what proportion of these were detected by the MMOs and the LT before entering the a 200, 500 and 1,000 yard radius from the ship. In every case, the proportion of detections by the MMOs was substantially higher than that for the LT – for example for 500 yds the MMOs detected 70% of pods before they entered this mitigation range while the LT detected 21% – 3 times worse. This appeared to be due to a combination of two factors: the LT detected far fewer pods overall and also tended to make a larger proportion of their detections within the mitigation ranges rather than outside of them, as would be required for effective mitigaton.

The relative performance of MMOs and LT was not of primary interest, partly because the MMO team was typically larger and also they were tasked only with detecting animals in the water, while the LT had other responsibilities. The primary interest was in estimating the absolute effectiveness of the LT, and the simple analyses discussed in the previous paragraph may provide an upper estimate of effectiveness because they do not account for animals missed by both positions. One caveat, however, is that the distance-specific indices do assume that the ship moves in a straight line and that animals do not move in the perpendicular direction – these assumptions are required for it to be true that pods seen at perpendicular distance x will pass within range x of the ship as it passes by. Some animals seen by the MMOs far ahead of the ship may have moved away before the ship passed, and so not entered within mitigation range.

The more sophisticated modelling exercise was intended to address the issue of animals undetected by both positions, and produce absolute estimates of PrU, which can be seen as the complement of effectiveness. For example, the estimate of PrU at 500 yds for of 0.53 for the MMOs and 0.85 for the LT mean that the MMOs are estimated to have had a (1-0.53)x100 = 47% chance of detecting a rorqual pod before it reached the 500 yard boundary while the LT had a (1-0.85)x100 = 15% chance. If the logic of the previous paragraph is correct that the previously-quoted simple distance-based indices of effectiveness provide an upper bound then we would expect these model-based estimates to be lower. Indeed they are in both cases (47% vs 70% for MMOs and 15% vs 21% for the LT), and also for all cases across all three sets of mitigation ranges. This provides some reassurance that the analyses are producing estimates that are, at least approximately, correct.

The analysis leading to estimation of PrU made the following assumptions:

1. the ship travels at 12 knots in a straight line;
2. pods do not move in the horizontal plane;
3. pods are uniformly distributed with respect to perpendicular distance from the ship;
4. time in near-surface and deep dive phases follow exponential distributions with known parameters, as does the inter-surfacing interval while in the near-surface state;
5. dive behavior is not affected by presence of the ship;
6. pod location and taxon are recorded accurately;

7. all first detections of a sighting are recorded, and MMOs do not cue the LT.

The effect of violation of these assumptions on estimation of PrU is typically not intuitively obvious. It could be studied via simulation. In the present study assumption 1 is likely broken. For rorquals assumption 2 may be mildly violated, although pods typically move slowly compared with ship speed; there is in some cases known to be avoidance behavior. There is no reason to suspect assumption 3 is violated for rorquals. For assumption 4, the grouping of species together means there is certainly variation in dive behavior within the group. The sensitivity analysis performed gave very similar results, so it may be that variation in dive behavior is not of primary importance for this group. Assumption 5 is likely violated, and warrants further study. For assumption 6, we have not attempted to look at measurement accuracy but one avenue here would be to compare LT and MMO records. We did encounter some difficulties in processing the data collected, and return to this below. In addition there is some rounding of distances, and possibly angles, evident in the data (see Figure 3). Assumption 7 appears to have been met.

Overall, we judge that our results for rorquals are broadly robust, but further investigations could be taken to assess this. One concern is the poor goodness-of-fit (as with the other two taxa analyzed), and we return to this below.

For the sperm whale taxon, we additionally assumed that the rorqual detection hazard parameters apply to this taxon. This enabled us to derive estimates of PrU, and illustrates an approach that might be applied to other taxa such as beaked whales if a suitable proxy taxon can be identified for which detectability of surfacings is measurable. Our findings that PrU for this taxon is higher than rorquals makes intuitive sense, as does the finding that PrU is more similar between the LT and MMOs because it is dominated for both observation positions by the lower surface availability. The sensitivity analysis showed that estimates of PrU and hence absolute effectiveness were somewhat influenced by the availability parameters, and hence more work is needed to determine what parameters are most reasonable for these long, deep divers.

For SCSP, the great majority of first detections were made within the mitigation ranges (nearly 90% of LT and 74% of MMO first detections within 1,000 yds of the ship). This clearly indicates that neither position were able to detect pods before they entered the mitigation ranges. The distance-specific indices of effectiveness for MMOs ranged from 25% at 200 yds to 14% at 1,000 yds and for the LT from 3% at 200 yds to 2% at 1,000 yds. It is important to note that these results are after excluding pods noted to have been bowriding, the majority of which were first detected well within 200 yds and so would have made these effectiveness figures even lower had they been included.

The model-based estimates of PrU for SCSP were not satisfactory, being unrealistically high (close to 1) for both positions and all mitigation ranges. This was likely caused by the significant proportion of first detections (around 10%) that were recorded at very close ranges (10 yds or less), leading to extremely “spiked” estimated detection hazard functions. The recorded very close ranges may be a combination of fast and responsive movement in some pods, violating assumptions 2 (no horizontal movement) and 3 (uniform distribution with respect to perpendicular distance). It may be that rounding of distances also contributed to the estimation difficulties. Despite these issues, a highly spiked detection hazard may not be unrealistic: for example Roberts et al. (2015 Figure 13) present a perpendicular distance detection function fitted to over 500 sightings of bottlenose dolphin from line transect surveys that exhibits a similarly spiked shape.

For SCLP, the great majority of first detections made by the LT (nearly 90%) were again within the mitigation ranges, while for the MMOs less than half (46.2%) were seen within 1,000 yds. This led to the LT having similar very low distance-specific indices of effectiveness to SCSP (e.g., 6% at 200 yds) but the MMOs having higher estimated effectiveness (e.g., 41% at 200 yds).

The model-based estimates of PrU for SCLP were very high for both platforms, although not much higher than would be indicated from the distance-specific indices for the LT. Like the SCSP, the fit was strongly influenced by the preponderance of first detections recorded at very close ranges, for the LT, where 17% of first detections were recorded as being at 10 yds or less from the ship. Similar problems of fast movement and rounding error are likely present in this group as for the SCSP.

Given the above, we conclude that effectiveness in detecting small cetaceans before they reach even the 200 yard mitigation range is very low for the LT, while for the MMOs it is likely to be low for small pods and somewhat higher for larger pods.

If more quantitative estimates were required, it may be possible to extend the analysis methods to account for animal movement and rounding error. Another factor potentially causing lack of fit is the assumption that detection hazard is the same in all directions. For rorquals at least, there is some evidence from the diagnostic plots in Appendix C that there are more detections at close perpendicular distances and fewer at close forward distances than predicted by the model – implying that observers may be searching further away in the forward distance direction than abeam. This may make sense for observers seeking to detect objects ahead of the ship, and so allowing different detection hazard parameters in the forward and perpendicular distance directions may improve model fit.

In preparing this report, we were required to expend considerable effort on data checking and cleaning due to inconsistencies in data recording. Some data had to be discarded. Thomas et al. (2012) recommended that, in advance of any further data collection, MMOs liaise with our group to discuss experimental data collection and recording protocols and that exploratory analyses are undertaken by way of data validation after every cruise. We again make this recommendation which we believe, if implemented, would reduce data loss and lead to better data quality in future.

Acknowledgements

We thank the dedicated MMOs who collected the field survey data, as well as the ships' commanders and crews who participated in the trials. Jay Barlow, Debbie Palka, and Lance Garrison of the US National Marine Fisheries Service provided helpful advice during construction of the experimental protocols. From within CREEM, David Borchers, Louise Burt, Roland Langrock and Eric Rexstad all provided valuable input into earlier parts of the project, as documented in previous reports.

Literature cited

Barlow J. and K.A. Forney. 2007. Abundance and density of cetaceans in the California Current ecosystem. Fish. Bull. 105: 509-526.

Borchers, D.L., and R. Langrock. 2015. Double-observer line transect surveys with Markov-modulated Poisson process models for animal availability. Biometrics 71 (4): 1060–9.

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

Burt, M.L. and L. Thomas. 2010. Calibrating US Navy lookout observer effectiveness: information for marine mammal observers, Version 2.1. Document submitted as part of Technical report IAW CDRL A003. University of St Andrews.

Department of the Navy. 2016. U.S. Navy Lookout Effectiveness Study Marine Mammal Observer Survey Protocol. August 2016. (Included as Appendix B of this report.)

Dolphin, W.F. 1987. Ventilation and dive patterns of humpback whales, *Megaptera novaeangliae*, on their alaskan feeding grounds. Canadian Journal of Zoology 65 (1): 83–90.

Drouot, V., A. Gannier, and J. Goold. 2004. Diving and feeding behaviour of sperm whales (*Physeter macrocephalus*) in the northwestern Mediterranean Sea." Aquatic Mammals 30: 419–26.

Kinzey, D., P. Olson and T. Gerrodette. 2000. Marine mammal data collection procedures on research ship line-transect surveys by the Southwest Fisheries Science Center. Southwest Fisheries Science Center, Administrative Report LJ-00-08, 32 p.

Langrock, R., D.L. Borchers, and H.J. Skaug. 2013. Markov-modulated nonhomogeneous Poisson processes for modeling detections in surveys of marine mammal abundance. Journal of the American Statistical Association 108 (503): 840–51.

NMFS. 2019. Letter of Authorization for US Navy training activities in the Atlantic Fleet Training and Testing (AFTT) Study Area.

NMFS. 2020. Letter of Authorization for US Navy training activities in the Hawaii-Southern California Training and Testing (HSTT) Study Area

Rexstad, E., and L. Thomas. 2010a. Preliminary analysis of observer effectiveness data gathered from FFG-A, DDG-A, DDG-B and DDG-C cruises. Technical report IAW CDRL A002. University of St Andrews.

Rexstad, E., and L. Thomas. 2010b. Observer effectiveness study – results from FFG-A, DDG-A, DDG-B and DDG-C cruises – Task C final survey protocol report. Technical report IAW CDRL A003. University of St Andrews.

Roberts J.J., Best B.D., Mannocci L., Fujioka E., Halpin P.N., Palka D.L., Garrison L.P., Mullin K.D., Cole T.V.N., Khan C.B., McLellan W.M., Pabst D.A. and Lockhart G.G. 2015. Density Model for Bottlenose Dolphin (*Tursiops truncatus*) for the U.S. Gulf of Mexico Version 3.3, 2015-10-07, and Supplementary Report. Marine Geospatial Ecology Lab, Duke University, Durham, North CarolinaScott, M., and S. Chivers. 2009. Movements and diving behavior of belagic spotted dolphins. Publications, Agencies; Staff of the US Department of Commerce. 46.

Thomas, L., E. Rexstad and D.L. Borchers. 2011. Calibrating US Navy lookout observer effectiveness – Task D: Final Data Analysis. Technical report IAW CDRL A004. University of St Andrews.

Thomas, L., R. Langrock, E. Rexstad and D.L. Borchers. 2012. Lookout effectiveness analysis development. Final Contract Report. HDR Task Order 002-0000105055. University of St Andrews.

Tables and Figures

Table 1. Cruises included in this study. Ship codes starting with CG are cruisers, DDG are destroyers and FFG are frigates. Study areas included Atlantic Fleet Training and Testing (AFTT), Hawaii Rance Complex (HRC) and the Southern California Range Complex (SOCAL).

Ship	Month	Year	# sightings	Study area
FFG-A	Feb	2010	21	HRC
DDG-A	Mar	2010	11	AFTT
DDG-B	Jun	2010	15	AFTT
DDG-C	Jul	2010	84	SOCAL
CG-A	Nov	2010	7	HRC
DDG-D	Feb	2011	29	HRC
DDG-E	Apr	2011	21	SOCAL
DDG-F	Nov	2011	5	HRC
DDG-G	Feb	2012	13	HRC
FFG-B	May/Jun	2012	24	AFTT
DDG-H	Jul	2012	62	SOCAL
DDG-I	Feb	2013	5	HRC
DDG-J	Aug	2013	2	HRC
DDG-K	Jan	2014	57	HRC
CG-B	Feb	2014	7	HRC
CG-C	Aug	2014	23	AFTT
DDG-L	Feb	2015	34	HRC
DDG-M	Apr	2015	3	AFTT
DDG-N	Feb	2016	12	HRC
DDG-O	Mar/Apr	2016	52	AFTT
DDG-P	Aug	2016	44	AFTT
DDG-Q	Aug	2017	56	AFTT
DDG-R	Feb	2018	22	HRC
DDG-S	Jun	2018	34	AFTT
DDG-T	Feb	2019	30	HRC
CG-D	Mar	2019	15	AFTT
DDG-U	Sep	2019	28	AFTT
Cruises	**27**	**Total**	**716**	

Table 2. Summary of sightings from lookout effectiveness field surveys. The hierarchy of taxon codes is given in Appendix A.

Taxon	Scientific name	Common name	# sightings	Mean pod size	SD pod size	# det by MMO only	# det by LO only	#det by both	# both: MMO first	# both: LO first	# both: same time
BAL--	Balaenoptera sp.	--	29	1.27	0.53	23	4	2.00	1	0	1
BALAC	Balaenoptera acutorostrata	Minke whale	3	1	0	1	0	2.00	1	0	1
BALMU	Balaenoptera musculus	Blue whale	16	2	0.91	12	0	4.00	3	0	1
BALPH	Balaenoptera physalus	Fin whale	1	1	NA	1	0	0.00	0	0	0
BLACK	--	Blackfish	2	2	0	2	0	0.00	0	0	0
CARCA	Caretta caretta	Loggerhead turtle	41	1	0	36	1	4.00	0	1	3
CET--	Unidentified cetacean		9	3.11	2.85	8	1	0.00	0	0	0
CHEMY	Chelonia mydas	Green turtle	12	1.08	0.29	10	0	2.00	0	0	2
DEL--	Delphinus sp.	Unid. common dolphin	11	92.32	109.03	8	0	3.00	2	0	1
DELCA	Delphinus capensis	Long-beaked common dolphin	2	42.5	38.89	1	0	1.00	1	0	0
DELDE	Delphinus delphis	Short-beaked common dolphin	11	27	38	8	0	3	1	0	2
DERCO	Dermochelys coriacea	Leatherback turtle	1	1	NA	0	0	1.00	0	0	1
DOLPH	--	Unid. dolphin	129	3.99	4.04	100	17	12.00	2	2	8
GLO--	Globicephala sp.	Unid. pilot whale	3	3.67	2.08	2	0	1.00	0	0	1
GLOMA	Globicephala macrorhynchus	Short-finned pilot whale	15	12.93	11.96	11	0	4.00	1	2	1
GRAGR	Grampus griseus	Risso’s dolphin	7	13.71	12.42	6	0	1.00	0	0	1
LAGAC	Lagenorhynchus acutus	Atlantic white-sided dolphin	1	8	NA	1	0	0.00	0	0	0
LEPKE	Lepidochelys kempii	Kemp's ridley turtle	1	1	NA	1	0	0.00	0	0	0
LGWHA	--	Unid. large whale	50	1.32	0.56	44	3	3.00	1	0	2
MEGNO	Megaptera novaeangliae	Humpback whale	84	1.98	1	58	3	23.00	10	3	10
MIXED	--	--	6	31.37	36.35	2	1	3.00	1	0	2
MOLMO	Mola mola	Pelagic sunfish	1	1	NA	1	0	0.00	0	0	0
ORCOR	Orcinus orca	Killer whale	1	1	NA	1	0	0.00	0	0	0
PHOPO	Phocoena phocoena	Harbor porpoise	1	1	NA	1	0	0.00	0	0	0
PHYMA	Physeter microcephalus	Sperm whale	2	1	NA	2	0	0.00	0	0	0
SMALL	--	Unid. small cetacean	8	6.69	4.17	5	0	3.00	0	0	3

STEAT	Stenella attenuata	Pantropical spotted dolphin	1	35	NA	1	0	0.00	0	0	0
STEBR	Steno bredanensis	Rough-toothed dolphin	1	NaN	NA	1	0	0.00	0	0	0
STECO	Stenella coeruleoalba	Striped dolphin	3	8.83	4.37	2	0	1.00	1	0	0
STEFR	Stenella frontalis	Atlantic spotted dolphin	30	6.77	8.04	20	0	10.00	0	0	10
STELO	Stenella longirostris	Spinner dolphin	1	45	NA	0	0	1.00	0	0	1
TURTL	--	Unid. turtle	33	14.61	43.51	29	3	1.00	0	0	1
TURTR	Tursiops truncatus	Common bottlenose dolphin	24	6.96	9.91	14	1	9.00	2	0	7
UN-MM	--	Unid. marine mammal	3	1.33	0.58	3	0	0.00	0	0	0
UNID	--	Unid. animal	1	1	NA	1	0	0.00	0	0	0
UNOTA	--	Unid. eared seal	5	1.2	0.45	2	1	2.00	0	0	2
UNPIN	--	Unid. pinniped	13	1.5	0.53	8	3	2.00	1	0	1
WHALE	--	Unid. whale	94	1.4	0.72	73	11	10.00	3	2	5
ZALCA	Zalophus californianus	California sea lion	14	1.23	0.6	8	0	6.00	0	0	6
Total			**670**			**507**	**49**	**114**	**31**	**10**	**73**

Table 3. Summary of sightings included in the rorquals group from lookout effectiveness field surveys. The hierarchy of taxon codes is given in Appendix A.

Taxon	# sightings	Mean pod size	SD pod size	# det by MMO only	# det by LO only	#det by both	# both: MMO first	# both: LO first	# both: same time
BAL--	29	1.27	0.53	23	4	2	1	0	1
BALAC	3	1	0	1	0	2	1	0	1
BALMU	16	2	0.91	12	0	4	3	0	1
BALPH	1	1	NA	1	0	0	0	0	0
LGWHA	50	1.32	0.56	44	3	3	1	0	2
MEGNO	84	1.98	1	58	3	23	10	3	10
WHALE	94	1.4	0.72	73	11	10	3	2	5
Total	**277**			**212**	**21**	**44**	**19**	**5**	**20**

Table 4. Distance-specific indices of effectiveness for the rorquals group. The index of effectiveness $IE_p(x)$ is the proportion of pods thought to have entered within mitigation range x, n_x, that were successfully detected by position p before they entered within that range, $n_{p,r>x}$. The calculation of n_x does not account for pods not detected by either position.

		MMO		LT	
	#trials n_x	#successes $n_{MMO,r>x}$	effectiveness $IE_{MMO}(x)$	#successes $n_{LT,r>x}$	effectiveness $IE_{LR}(x)$
x=200 yds	34	25	0.74	12	0.35
x=500 yds	58	41	0.70	12	0.21
x=1000 yds	104	57	0.54	14	0.13

Table 5. Summary of sightings included in the SCSP group from lookout effectiveness field surveys. The hierarchy of taxon codes is given in Appendix A.

Taxon	# sightings	Mean pod size	SD pod size	# det by MMO only	# det by LO only	#det by both	# both: MMO first	# both: LO first	# both: same time
BLACK	2	2	0	2	0	0	0	0	0
DEL--	1	1	NA	1	0	0	0	0	0
DELDE	4	2	1.41	4	0	0	0	0	0
DOLPH	103	2.59	1.57	79	13	11	2	2	7
GLO--	3	3.67	2.08	2	0	1	0	0	1
GLOMA	5	3.5	1.94	3	0	2	1	1	0
GRAGR	2	3.5	2.12	2	0	0	0	0	0
MIXED	2	4.17	2.59	0	1	1	0	0	1
ORCOR	1	1	NA	1	0	0	0	0	0
PHOPO	1	1	NA	1	0	0	0	0	0
SMALL	4	2.88	0.85	1	0	3	0	0	3
STECO	1	4	NA	1	0	0	0	0	0
STEFR	28	2.57	1.75	16	4	8	0	1	7
TURTR	21	3.48	1.78	12	2	7	2	0	5
Total	**178**			**125**	**20**	**33**	**5**	**4**	**24**

Table 6. Distance-specific indices of effectiveness for the SCSP group. The index of effectiveness $IE_p(x)$ is the proportion of pods thought to have entered within mitigation range x, n_x, that were successfully detected by position p before they entered within that range, $n_{p,r>x}$. The calculation of n_x does not account for pods not detected by either position.

		MMO		LT	
	#trials n_x	#successes $n_{MMO,r>x}$	effectiveness $IE_{MMO}(x)$	#successes $n_{LT,r>x}$	effectiveness $IE_{LR}(x)$
x=200 yds	102	25	0.25	3	0.03
x=500 yds	144	42	0.29	5	0.03
x=1000 yds	181	26	0.14	3	0.02

Table 7. Summary of sightings included in the SCLP group from lookout effectiveness field surveys. The hierarchy of taxon codes is given in Appendix A.

Taxon	# sightings	Mean pod size	SD pod size	# det by MMO only	# det by LO only	#det by both	# both: MMO first	# both: LO first	# both: same time
DEL--	10	101.45	110.41	7	0	3	2	0	1
DELCA	2	42.5	38.89	1	0	1	1	0	0
DELDE	7	40.64	42.79	4	0	3	1	0	2
DOLPH	21	11.69	5.15	18	2	1	0	0	1
GLOMA	10	17.64	12.12	8	0	2	0	1	1
GRAGR	5	17.8	12.54	4	0	1	0	0	1
LAGAC	1	8	NA	1	0	0	0	0	0
MIXED	3	49.5	37.49	2	0	1	1	0	0
SMALL	4	10.5	1	4	0	0	0	0	0
STEAT	1	35	NA	1	0	0	0	0	0
STECO	2	11.25	1.77	1	0	1	1	0	0
STEFR	13	15.85	8.36	5	5	3	0	0	3
STELO	2	26	26.87	0	1	1	0	0	1
TURTR	6	24	15.52	2	1	3	0	0	3
Total	**87**			**58**	**9**	**20**	**6**	**1**	**13**

Table 8. Distance-specific indices of effectiveness for the SCLP group. The index of effectiveness $IE_p(x)$ is the proportion of pods thought to have entered within mitigation range x, n_x, that were successfully detected by position p before they entered within that range, $n_{p,r>x}$. The calculation of n_x does not account for pods not detected by either position.

	MMO			LT	
	#trials n_x	#successes $n_{MMO,r>x}$	effectiveness $IE_{MMO}(x)$	#successes $n_{LT,r>x}$	effectiveness $IE_{LR}(x)$
x=200 yds	34	14	0.41	2	0.06
x=500 yds	49	16	0.33	2	0.04
x=1000 yds	74	23	0.31	2	0.03

Table 9. Simulation study results, showing mean percent bias in estimates of the effective strip half-width (ESHW), the trackline detection probability (g(0)) and the probability of remaining undetected (PrU) calculated for 200 yds, 500 yds and 1,000 yds for simulation scenarios 1 (two-way independence, all detections) and 2 (one-way independence, first detections only). MMO is marine mammal observers and LT is lookout team.

	Scenario 1: two-way independence all detections		Scenario 2: one-way independence first detection only	
	MMO	LT	MMO	LT
ESHW	0.1	-0.3	-1.5	-1.2
g(0)	-0.2	-1.9	-0.6	-2.5
PrU 200 yds	0.4	1.9	1.4	2.6
PrU 500 yds	0.3	1.7	1.4	2.4
PrU 1,000 yds	0.2	1.4	1.4	2.0

Table 10. *Survey data results, showing estimated effective strip half-width (ESHW), the trackline detection probability (g(0)) and the probability of remaining undetected (PrU) calculated at 200 yds, 500 yds and 1,000 yds for four cetacean taxa. MMO is marine mammal observers and LT is lookout team. Values in brackets are 95% confidence intervals. (Note, estimated PrUs for small cetaceans are 1.00 when rounded to 2 decimal places, but are denoted 0.99 to indicate that they are not exactly 1.)*

	Rorqual		Sperm whale		Small cetaceans in small pods (6 or fewer)		Small cetaceans in large pods (more than 6)	
	MMO	LT	MMO	LT	MMO	LT	MMO	LT
ESHW	1739 m (1396-2126)	408 m (310-579)	886 m (733-1060)	234 (178-318)	0.66 m (0.26-1.63)	0.192 m (0.069-0.481)	240 m (137-363)	70 m (0-109)
g(0)	0.53 (0.43-0.62)	0.24 (0.16-0.31)	0.24 (0.20-0.27)	0.12 (0.09-0.154)	0.0027 (0.0011-0.0064)	0.0011 (0.0004-0.0027)	0.49 (0.31-0.68)	0.25 (0.00-0.37)
PrU 200 yds	0.49 (0.40-0.59)	0.80 (0.74-0.86)	0.77 (0.74-0.80)	0.89 (0.87-0.92)	0.99 (0.99-0.99)	0.99 (0.99-1.00)	0.83 (0.74-0.90)	0.94 (0.91-1.00)
PrU 500 yds	0.53 (0.43-0.62)	0.85 (0.80-0.89)	0.78 (0.75-0.81)	0.92 (0.89-0.94)	0.99 (0.99-0.99)	1.00 (0.99-1.00)	0.93 (0.89-0.96)	0.98 (0.97-1.00)
PrU 1,000 yds	0.59 (0.51-0.67)	0.91 (0.87-0.94)	0.80 (0.77-0.84)	0.95 (0.93-0.96)	1.0000 (0.99-1.00)	1.00 (1.00-1.00)	0.97 (0.95-0.98)	0.99 (0.99-1.00)

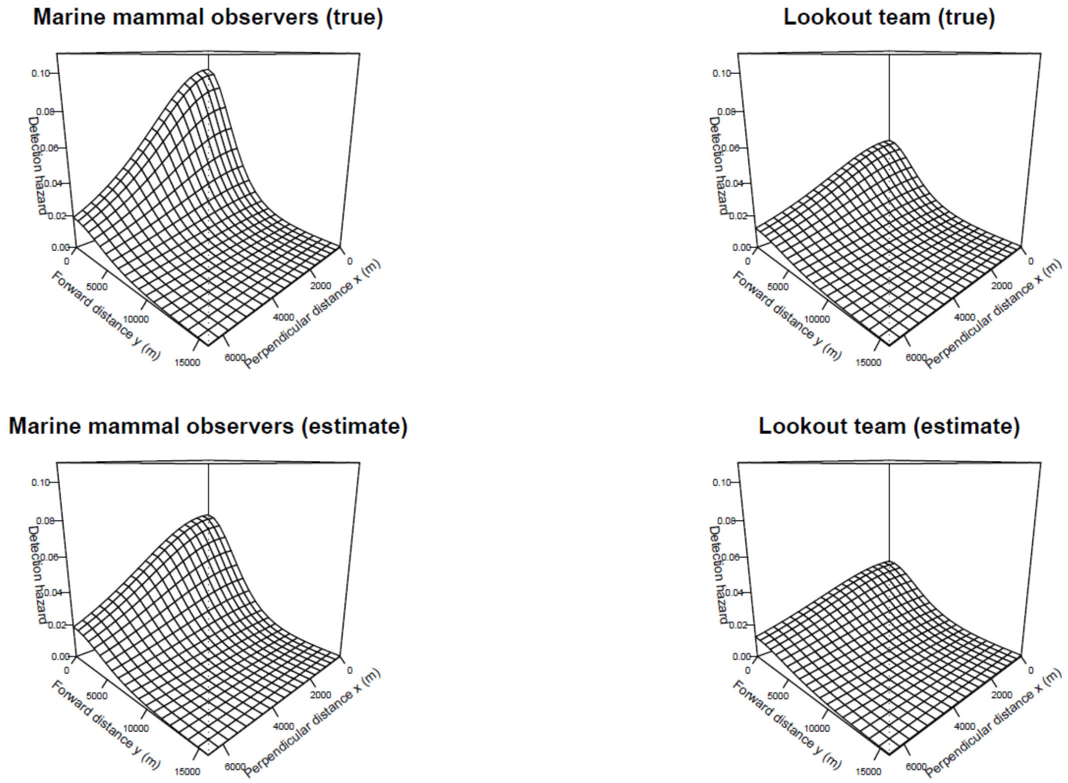


Figure 1. True detection hazard from the simulation study (top) and example estimated detection hazard from one example dataset generated under scenario 2 (bottom).

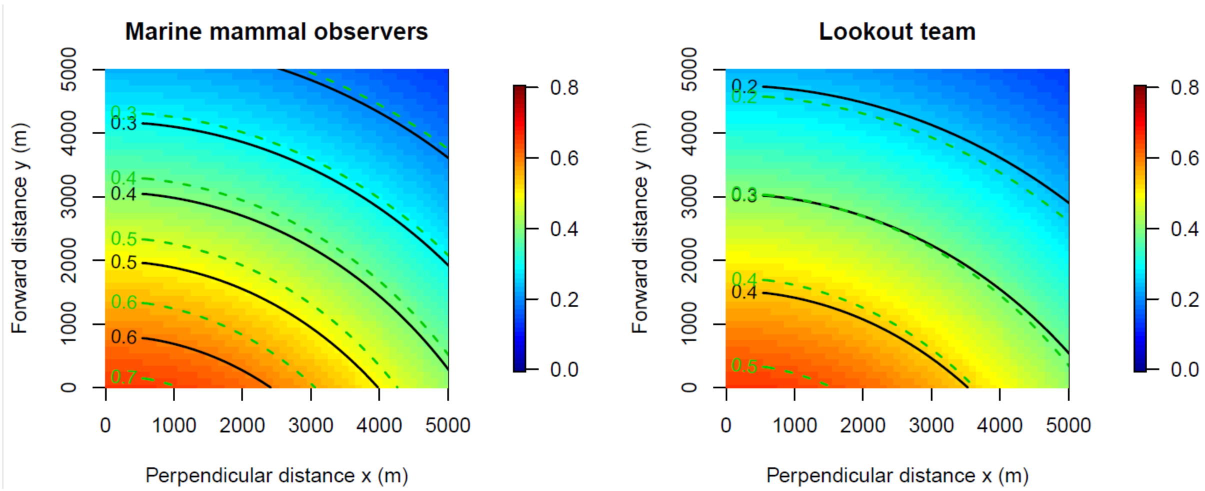


Figure 2. Cumulative distribution of distance to first detection from simulation scenario 2. Green dashed lines are the true distribution (i.e., calculated with the true simulation parameters); black lines and shading show estimated distribution from an analysis of one example dataset generated from the simulation.

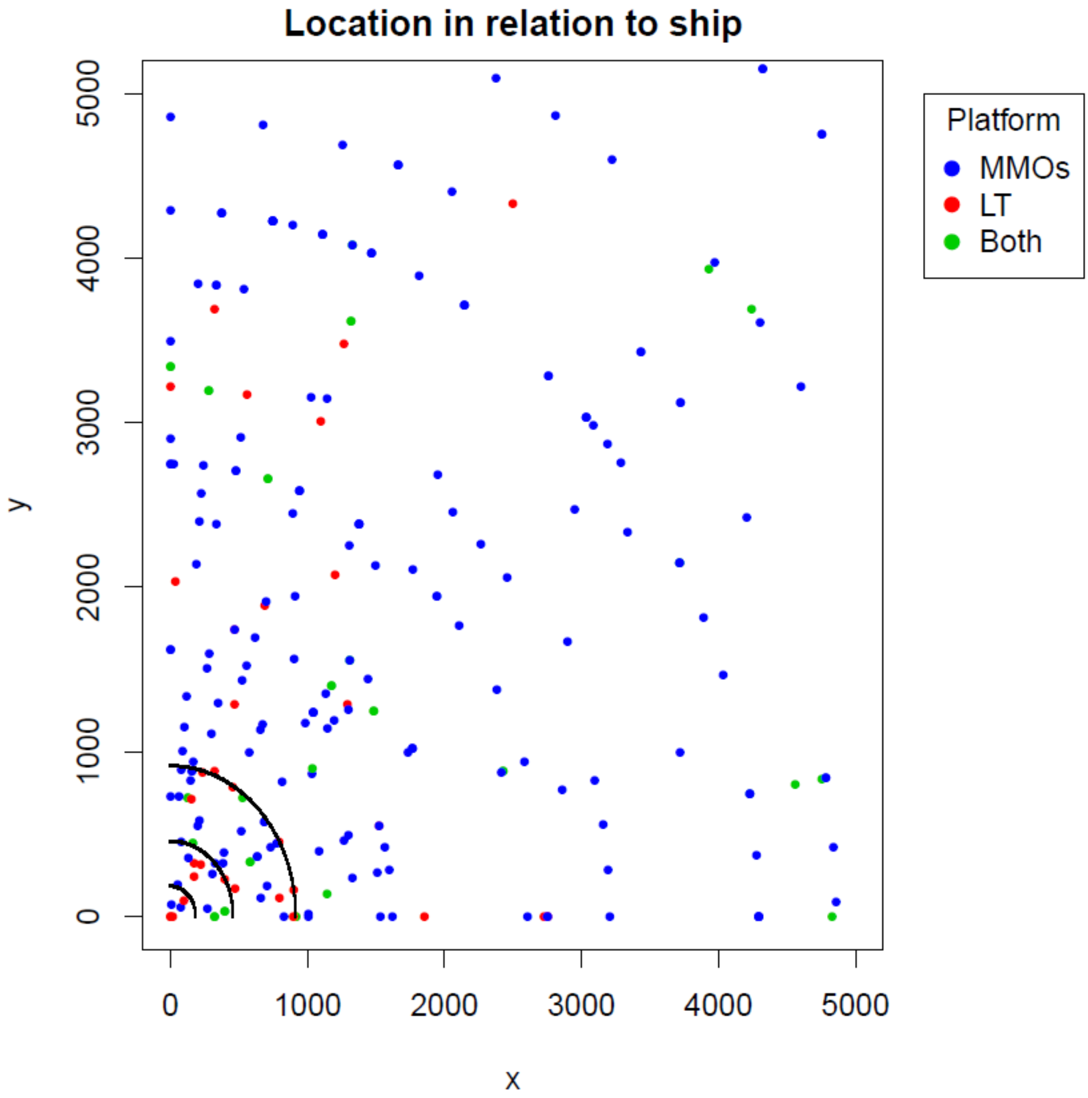


Figure 3. Location of rorqual first detections relative to ship location at (0, 0). Lines represent the three mitigation ranges of 200, 500 and 1,000 yards.

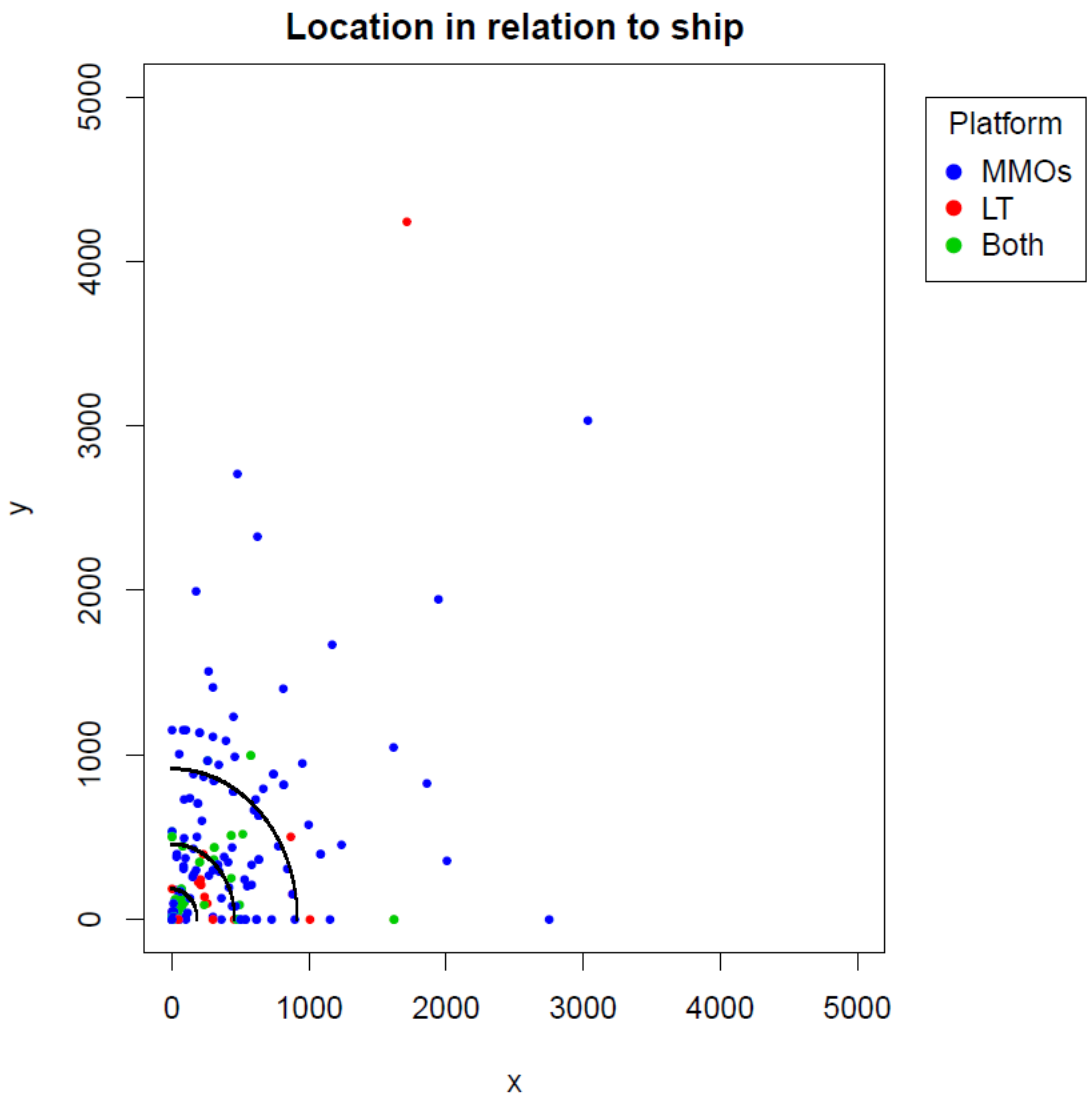


Figure 4. Location of small cetacean small pods (SCSP) first detections relative to ship location at (0, 0). Lines represent the three mitigation ranges of 200, 500 and 1,000 yards.

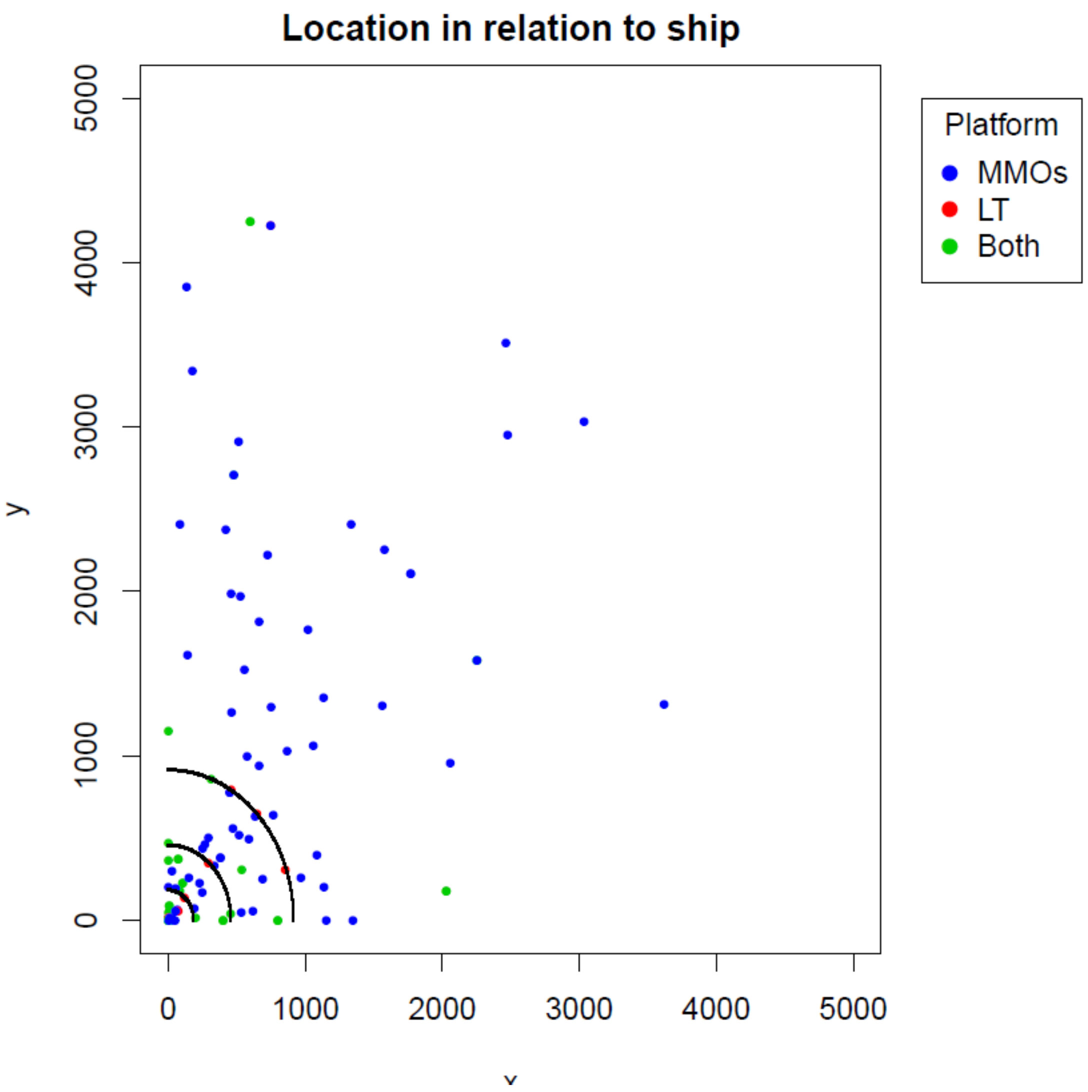


Figure 5. Location of small cetacean large pods (SCLP) first detections relative to ship location at (0, 0). Lines represent the three mitigation ranges of 200, 500 and 1,000 yards.

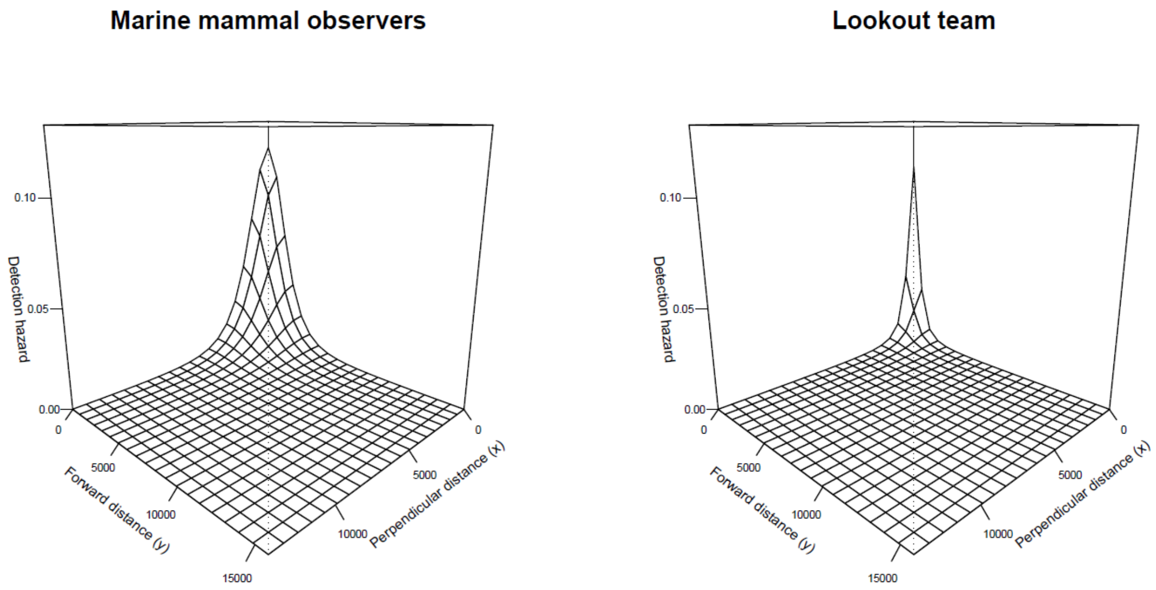


Figure 6. Estimated detection hazard of a surfacing from model fitted to rorqual group.

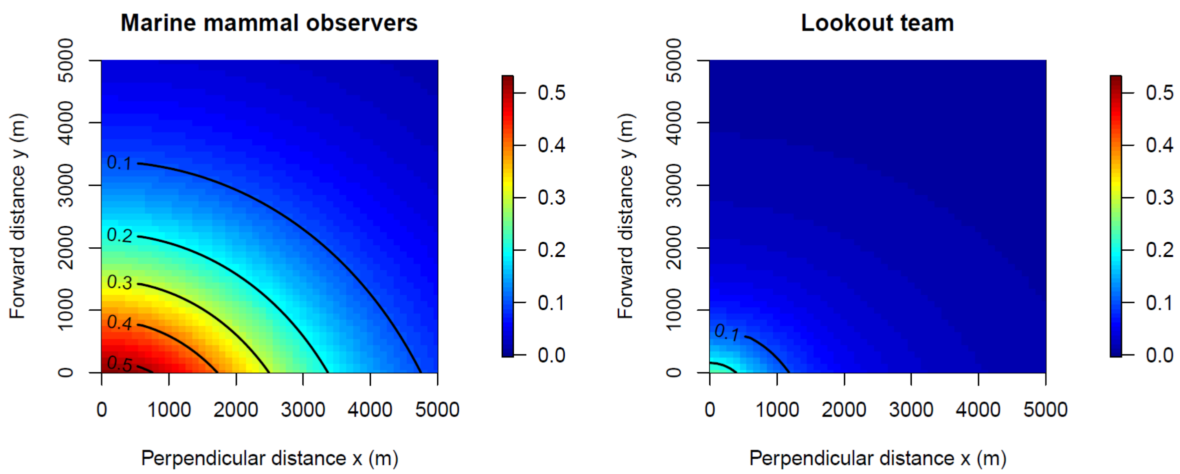


Figure 7. Estimated cumulative distribution of distances to first detection from model fitted to the rorqual group.

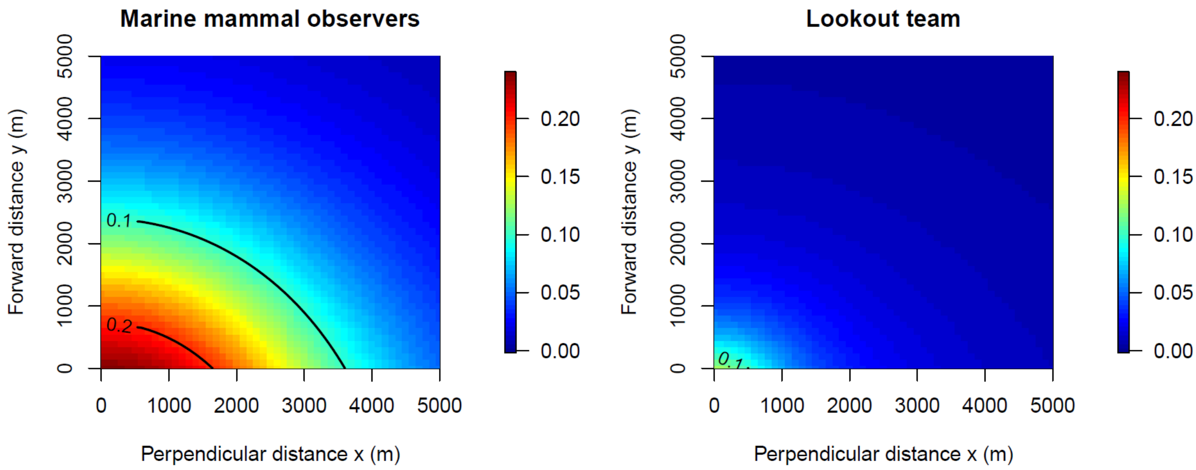


Figure 8. Estimated cumulative distribution of distances to first detection for sperm whales calculated using the estimated detection hazard from the model fitted to the rorqual group.

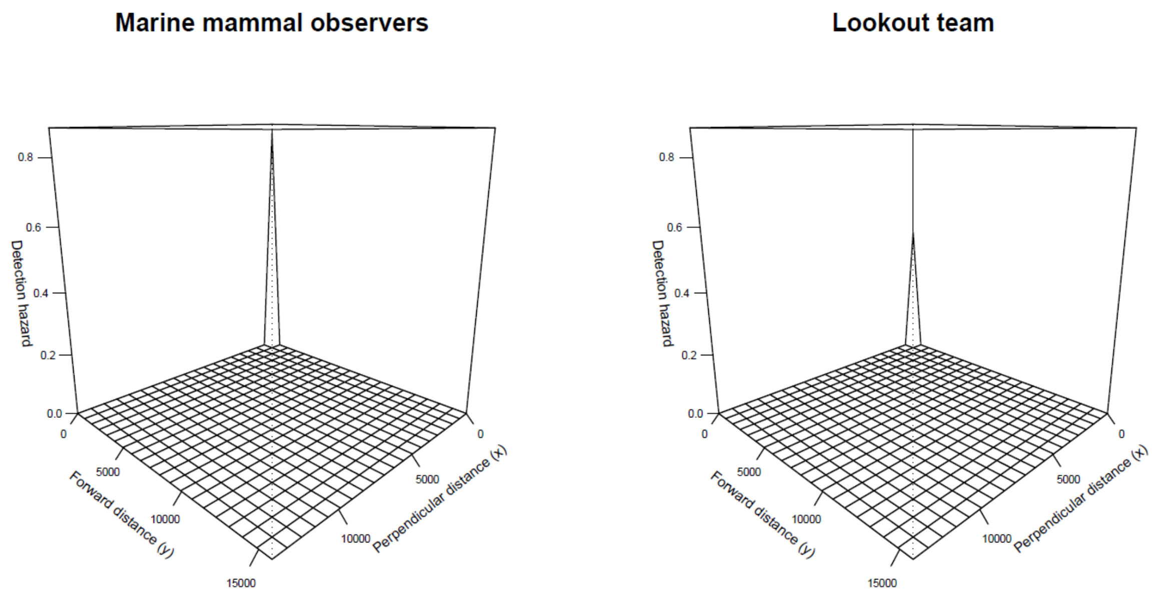


Figure 9. Estimated detection hazard of a surfacing from model fitted to the small cetaceans in small pods (SCSP) group.

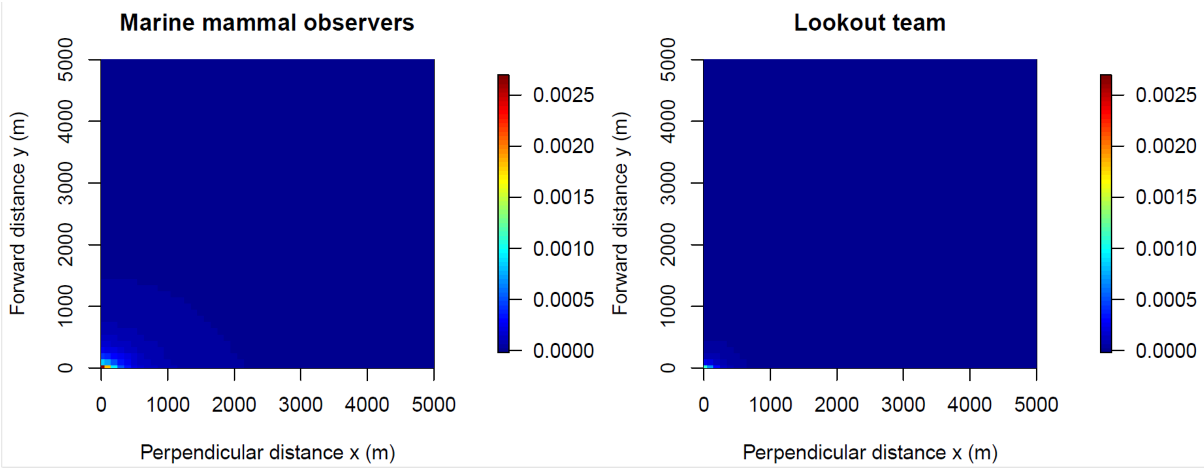


Figure 10. Estimated cumulative distribution of distances to first detection from model fitted to the small cetaceans in small pods (SCSP) group.

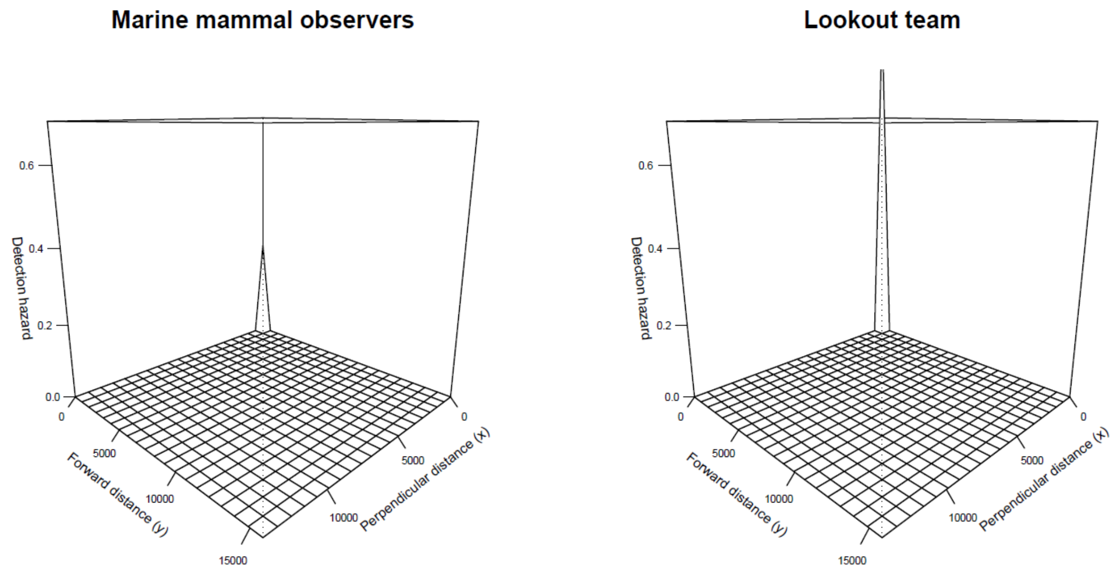


Figure 11. Estimated detection hazard of a surfacing from model fitted to the small cetaceans in small pods (SCLP) group.

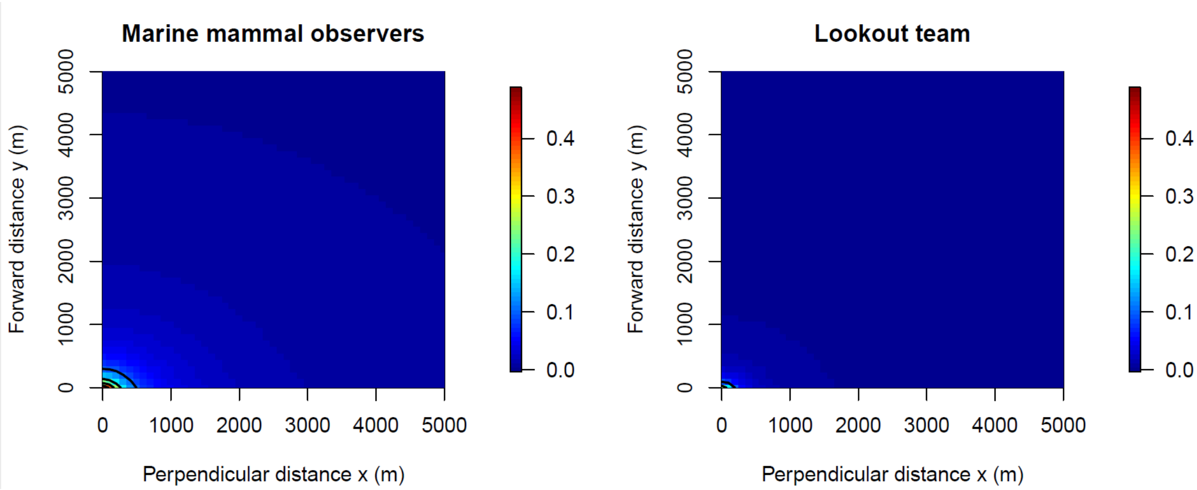


Figure 12. Estimated cumulative distribution of distances to first detection from model fitted to the small cetaceans in small pods (SCLP) group.

Appendix A: Taxonomic codes used in lookout effectiveness surveys.

Common and scientific names for each code is given in Table 4.4 of Appendix B.

Level1	Level2	Level3	Level4	Level5	Level6	Level7
BALAC	NA	BAL--	LGWHA	WHALE	CET--	UN-MM
BALED	NA	BAL--	LGWHA	WHALE	CET--	UN-MM
BALBO	NA	BAL--	LGWHA	WHALE	CET--	UN-MM
BALMU	NA	BAL--	LGWHA	WHALE	CET--	UN-MM
BALPH	NA	BAL--	LGWHA	WHALE	CET--	UN-MM
MEGNO	NA	BAL--	LGWHA	WHALE	CET--	UN-MM
ESCRO	NA	NA	LGWHA	WHALE	CET--	UN-MM
MESDE	MES--	ZIP--	LGWHA	WHALE	CET--	UN-MM
ZIPCA	NA	ZIP--	LGWHA	WHALE	CET--	UN-MM
INDPA	NA	ZIP--	LGWHA	WHALE	CET--	UN-MM
BERBA	NA	ZIP--	LGWHA	WHALE	CET--	UN-MM
PHYMA	NA	NA	LGWHA	WHALE	CET--	UN-MM
KOGBR	KOG--	NA	NA	SMALL	CET--	UN-MM
KOGSI	KOG--	NA	NA	SMALL	CET--	UN-MM
ORCOR	NA	BLACK	NA	SMALL	CET--	UN-MM
PSECR	NA	BLACK	NA	SMALL	CET--	UN-MM
FERAT	NA	BLACK	NA	SMALL	CET--	UN-MM
PEPEL	NA	BLACK	NA	SMALL	CET--	UN-MM
GLOMA	GLO--	BLACK	NA	SMALL	CET--	UN-MM
TURTR	NA	NA	DOLPH	SMALL	CET--	UN-MM
GRAGR	NA	NA	DOLPH	SMALL	CET--	UN-MM
STEFR	STE--	NA	DOLPH	SMALL	CET--	UN-MM
STEAT	STE--	NA	DOLPH	SMALL	CET--	UN-MM
STELO	STE--	NA	DOLPH	SMALL	CET--	UN-MM
STECO	STE--	NA	DOLPH	SMALL	CET--	UN-MM
STEBR	NA	NA	DOLPH	SMALL	CET--	UN-MM
DELDE	DEL--	NA	DOLPH	SMALL	CET--	UN-MM
DELCA	DEL--	NA	DOLPH	SMALL	CET--	UN-MM
LAGHO	NA	NA	DOLPH	SMALL	CET--	UN-MM
LAGOB	NA	NA	DOLPH	SMALL	CET--	UN-MM
LISBO	NA	NA	DOLPH	SMALL	CET--	UN-MM
CHEMY	NA	NA	NA	NA	TURTL	NA
EREIM	NA	NA	NA	NA	TURTL	NA
LEPKE	NA	NA	NA	NA	TURTL	NA
DERCO	NA	NA	NA	NA	TURTL	NA
CARCA	NA	NA	NA	NA	TURTL	NA
LEPOL	NA	NA	NA	NA	TURTL	NA
NEOSC	NA	NA	UNOTA	SEALS	UNPIN	UN-MM
ZALCA	NA	NA	UNOTA	SEALS	UNPIN	UN-MM
PHOVI	NA	NA	UNOTA	SEALS	UNPIN	UN-MM
MIXED	NA	NA	NA	NA	NA	UN-MM
MOLMO	NA	NA	NA	NA	FISH	NA

Appendix B: Calibrating US Navy lookout observer effectiveness. Information for Marine Mammal Observers, Version 2.1.

Appendix C: Markov-modulated Poisson process models for lookout effectiveness data