MARK RECAPTURE DISTANCE SAMPLING: USING ACOUSTICS TO ESTIMATE THE FRACTION OF DOLPHINS MISSED BY OBSERVERS DURING SHIPBOARD LINE-TRANSECT SURVEYS

Shannon Rankin*[1], Cornelia Oedekoven[2], and Frederick Archer[1]

[1] Southwest Fisheries Science Center, NMFS, NOAA, 8901 La Jolla Shores Drive, La Jolla, California 92037.

[2] Centre for Research into Ecological and Environmental Modeling, University of St. Andrews, The Observatory, Buchanan Gardens, St. Andrews, United Kingdom

*Corresponding Author: shannon.rankin@noaa.gov; ORCID ID: 0000-0002-8405-5169

ACKNOWLEDGEMENTS

This work benefited from the help of the officers and crew of the NOAA Ship *McArthur II*. We also thank the cruise leaders, marine mammal observers, Liz Ferguson (field data collection) and Jay Barlow (who made this project possible). Funding was provided by the U.S. Navy's N45 Program and NOAA Southwest Fisheries Science Center. We thank Jeff Laake (NOAA Fisheries) and David Borchers (University of St Andrews) for help with the mrds package.

Keywords: distance sampling, line-transect surveys, mark-recapture, passive acoustics, mark-recapture distance sampling, trackline detection probability, double observer methods, abundance estimation, dolphins

ABSTRACT

Cetacean abundance estimation often relies on distance sampling methods using shipboard visual line-transect surveys, which assumes that all animals on the trackline are detected and that the detection of animals decreases with increasing distance from the trackline. Mark-Recapture Distance Sampling (MRDS) typically employs a secondary visual observation team and may be used to identify the fraction of animals detected on the trackline when it is suspected that animals may have been missed. For species that are difficult to detect using visual observation methods, such as deep-diving species or those with cryptic surfacing behavior, this secondary team may be prone to the same limitations in detection as the primary observation team and alternative modes of detection may improve estimates. Here we examine the potential use of passive acoustic detection as a secondary platform for MRDS of rough-toothed dolphins (*Steno bredanensis)* during a combined visual and acoustic shipboard line-transect survey. The average trackline detection probability for rough-toothed dolphins was less than one for both the trial configuration (average $p(0)$ = 0.45 for the visual team) and independent observer configuration (average $p(0)$= 0.37 for the visual, $p(0)$= 0.77 for the acoustic and $p(0)$=0.84 for both teams combined). This study, while limited in scope, strongly suggests that passive acoustic methods may be an effective alternative for estimating $p(0)$ for cetaceans.

1. INTRODUCTION

Cetacean abundance estimation typically relies on shipboard line-transect surveys using visual detections. Here, a ship travels along predefined tracklines and the visual observers record the species identification, number of animals in the group, and the distances at which the group was detected. For line-transect sampling, it is frequently assumed that *all* animals on the trackline are detected and the probability of detecting a group of animals, $g(x)$, decreases with distance x from the transect line. Detection of animals on the trackline is often referred to as $g(0)$, however, to avoid confusion we follow the convention suggested by Laake and Borchers (2004) and refer to it as $p(0)$. Mark-Recapture Distance Sampling (MRDS - Borchers et al. 1998) is an alternative to traditional line-transect methods that allows estimation of trackline detection probability, $p(0)$, and thus, does not rely on perfect detection on the trackline. MRDS generally requires two independent or conditionally independent observation platforms, where all covariates that affect detectability are included in the model. This independence requires that, at a minimum, platforms do not cue each other (see below for further criteria). For shipboard surveys, these are typically two visual platforms on the same survey vessel (Borchers et al. 2006; Laake and Borchers 2004).

Visual detection of cetaceans may be adversely affected by variables such as weather or diving behavior (Barlow et al. 2001). If this results in a failure to detect animals on the trackline, then the assumption that $p(0)$ = 1 will be violated, resulting in negatively biased abundance estimates. There are two main biases that may lead to violation of this assumption: availability bias and perception bias. Availability bias occurs when animals are missed because they are not available for detection, a common problem with long-diving species. Perception bias occurs when animals are available for detection but are missed for other reasons, such as cryptic surfacing behavior, poor weather conditions, or observer fatigue.

Using independent observation platforms allows us to quantify these biases. However, it is crucial that detections made from the two platforms are truly independent (Burt et al. 2014). If both platforms are subject to the same limitations (as is often the case when using two visual platforms on the same vessel), MRDS can only address some aspects of perception bias, such as observer experience or fatigue. Perception bias due to behavior, weather conditions and availability biases remain. Consideration of a second independent platform using an alternative method of detection that does not have the same limitations as the primary platform may allow for consideration of both perception and availability bias.

The use of passive acoustic monitoring (PAM) methods during shipboard cetacean surveys has increased dramatically in recent years (vanParijs et al. 2009; Marques et al. 2013); however, one role that has received little attention is its use as an independent platform to evaluate the fraction of cetaceans missed during shipboard line-transect surveys. Passive acoustic detection of cetaceans is not limited to calm sea states (Rankin et al. 2008b) and has proven a valuable tool for detecting species with cryptic surface behavior (Rankin and Barlow 2005; Gerrodette et al. 2011). Many long-diving species are acoustically active during their dives (Barlow and Taylor 2005; Barlow et al. 2013; Pérez et al. 2017), when they are not available for visual observation. In these cases, passive acoustic methods may provide a more suitable independent detection method for MRDS than a second team of visual observers because acoustic methods can also detect submerged animals.

Here we present an example of how passive acoustic detections may be used to estimate the fraction of dolphins missed by the visual observer team during line-transect surveys. To this end, we used acoustic detections of rough-toothed dolphins (*Steno bredanensis*) to 'mark' detections and set up 'trials' for the visual observation team during a combined visual and acoustic line-transect survey conducted in the eastern tropical Pacific (ETP). This species was chosen for study because of their distinctive vocalizations that could be easily identified by an experienced acoustician in real-time during the survey (Rankin et al. 2015). We analyzed these detections using MRDS methods to estimate the number of groups missed by the visual team within the strip width. The emphasis of this paper is on future research needs and method development, rather than providing reliable results of $p(0)$ that can be used for population estimates of rough-toothed dolphins.

2. MATERIALS AND METHODS

2.1. Visual line transect data

The data were collected during the *Stenella* Abundance Research Line Transect and Ecosystem cruise (STARLITE) conducted by the Southwest Fisheries Science Center (SWFSC) in 2007 on board the NOAA Ship *McArthur II* (Archer et al. 2008). Visual observation followed standard SWFSC protocol (Kinzey et al. 2000), using two visual observers on 25×150 'big eye' binoculars scanning forward of the vessel and one visual observer recording data and scanning the near field with naked eye. When cetaceans were detected, observers obtained species identification (to the lowest taxonomic level possible) and group size estimates. Standard SWFSC surveys are conducted in "closing" mode, where the visual team usually suspends survey effort upon sighting a cetacean group and the ship approaches the group for accurate group size estimates and species identification. During STARLITE, the survey alternated between a day of closing mode, followed by a day of "passing" mode along the same segment covered in the previous day. In passing mode, search effort remained uninterrupted and sighting information (species identification and group size estimates) was obtained from a distance. Sighting information was relayed to the acoustics team, but the visual team was not aware of acoustic detections.

2.2. Acoustic line transect data

We towed a hydrophone array 300 m behind the ship during daylight hours on passing mode days, which allowed for a direct comparison of visual and acoustic detection of cetaceans. Hydrophones had internal pre-amplification and sensitivity from 1 kHz to 40 kHz (± 5 dB re 1 μPa at 1 m); a minimum of two hydrophones with 4 m separation provided localization capabilities using time difference of arrival between the hydrophones. High-frequency recordings were made on a computer hard disk (96 kS/sec) using a MOTU Traveler digital audio interface with Raven software (Center for Conservation Bioacoustics 2011) on a desktop computer.

The acoustician on watch monitored sounds in real-time aurally via headphones and visually using a spectrographic display. ISHMAEL software (Mellinger 2001) was used for real-time spectrographic monitoring and estimation of bearing angles to the sound source. Bearing angles were plotted with a custom mapping program (Whaltrak) and the location of the sound source was estimated by the convergence of successive bearing angles (Rankin et al. 2008a). The vessel traveled in an imperfect straight line, and this, combined with the slight delay between detecting and plotting a bearing angle would lead to variations in convergence on either side of the vessel. Therefore, we calculated beam distance (distance at which the sound source passed 90^o) as the average perpendicular distance from the trackline to the left and right localizations at the time the group passed the hydrophone array. For species that spend limited time at depth, such as rough-toothed dolphins, this perpendicular distance is assumed to equal the beam distance (slant angle = 0).

We identified visual and acoustic detections as matches (duplicate ID for MRDS) when visual and acoustic detections of the same (or similar) species occurred in close proximity to each other. Specifically, acoustic bearing angles and approximate distances as estimated from acoustic detections must match those recorded by the visual team, ideally for multiple sighting updates as well as when the animals passed the beam of the ship. In the case of potential ambiguity, the acoustician would specifically request an updated location for a group to confirm a match. For MRDS analyses, it is necessary to assign a single perpendicular distance to a group of animals, regardless of which observer detected it. Due to potential animal movement and a time delay between detections made by the visual and acoustic teams, perpendicular distances for a given group differed between these two observer platforms. It is generally recommended to use the perpendicular distance from the observer who first detected the group as it is assumed that the animals will have had less of a chance to move in response to the observers (in this case, the ship; Burt et al. 2014). For the same reason, we used the perpendicular distances from the visual team when possible as these were obtained at the time of first detection (perpendicular distances for acoustic detections were obtained when animals passed the beam). Perpendicular distances were truncated at w = 5 km following recommendations by Buckland et al. (2001).

An experienced acoustician (S. Rankin) made acoustic identification of rough-toothed dolphins in the field. This species frequently produces stepped whistles, which feature several distinct frequency jumps with no time gap (Rankin et al. 2015). The stepped whistles produced by rough-toothed dolphins are unique in that the entire whistle and/or the individual components are often downswept. The echolocation clicks and burst pulses of rough-toothed dolphins have low-frequency energy (extending below 15 kHz) and are often produced in short repeated 'packets'. These characteristics are both common to rough-toothed dolphins and rare to other species encountered in the study area. Detection of calls with these distinct features has been found to be indicative of the presence of rough-toothed dolphins (Rankin et al. 2015).

2.3. Mark-recapture distance sampling

In the following we speak of 'observers', instead of 'teams of observers' or 'observation platforms' for brevity. Laake and Borchers (2004) identified three observation configurations for MRDS which each have two observers: independent, trial, and removal configuration. Each configuration differs according to which observer sets up the trials for the other and which observer may be cued by the other. For the independent observer configuration, two observers, say observer 1 and observer 2, search independently of each other and their detections serve as trials for the other. Here it is essential that neither observer cues the other. For the trial configuration, only one of the observers (observer 2), who is usually looking further ahead, sets up trials for observer 1. In this trial configuration, observer 2 may be cued by observer 1, however, it is essential that observer 1 is not cued by observer 2. For the removal configuration, observer 2 is aware of what observer 1 detects and detects objects missed by observer 1 (here again, observer 1 must not be cued by observer 2).

Another concern for MRDS studies is the level of independence between the two observers (Borchers et al. 2006; Buckland et al. 2009; Burt et al. 2014). Even if observers are not directly cued by one another, there may be preferential detection of the 'most observable', which effectively violates the assumption of independence. This unmodelled heterogeneity may include body size (larger animals are easier to detect visually) or group size (larger group sizes may be easier to detect by both visual and acoustic methods). Ideally, all variables that affect the probability of detection would be included as covariates in the model; in reality, it may be difficult to observe or record all sources of heterogeneity. A test to identify unmodelled heterogeneity can be made by comparing the detection functions of the mark-recapture (MR) model and the distance-sampling (DS) model; the shape of the detection function for these models should be the same if the observers are independent (Burt et al. 2014). If the shape of the models is similar, full independence may be considered. The more limiting point independence can be used when there is potential dependence in detections (unmodelled heterogeneity). For point independence we only assume that observers make independent observations at distance zero. Here, we exploit the fact that the distribution of distances can generally be assumed known for DS data, e.g. uniform for line transects, as long as the survey followed a random design (Buckland et al. 2001). We can estimate the shape of the detection function with only the DS model fitted to the observed distances, i.e. without using the MR data (i.e. the trials) (Burt et al. 2014). However, the MR model is needed to estimate the detection probabilities at distance zero (which the DS model assumes to be 1). See below for further details.

For our study, we considered both the simpler trial configuration and the more complete independent configuration. For each configuration, we tested for both full and point independence between the visual and acoustic detections. The removal configuration assumes equal detection probability for each observer and is inappropriate for most applications considering acoustics as the secondary observer (Laake and Borchers 2004).

In the trial configuration, a trial consisted of a distinct group of rough-toothed dolphins detected by the acoustics team (observer 2). If the visual observers (observer 1) also detected the same group, the trial was a success, otherwise a failure. For the independent observer configuration, detections by each team (visual and acoustic) set up trials for the other. In this case, if detections were made by both teams they were considered successes (if not, they were considered failures). Visual observations were independent of acoustic detections and acoustic cues were unlikely to be biased by visual detection even though the acoustics team were informed of sightings.

MRDS analyses were done in R (vs 3.4.4, R core team 2018) using the mrds package (vs 2.1.18, Laake et al. 2018). Here, we must specify an MR model and, for point independence, a DS model, describing the conditional and relative detection functions, respectively, as a function of perpendicular distance y from the trackline. The MR model, $p(y)$, is a binomial generalized linear model fitted to the trial data which specifies the form of the conditional detection functions, i.e. the probability the group is seen by observer 1 given it was seen by observer 2, or vice versa. The DS model is the relative detection function, $g(y)$, which is fitted to the distances of all detected groups combined (the unique observations detected by either observer). The DS model assumes $g(0)$ = 1 and that groups are uniformly distributed with respect to the transect line.

The detection function, $d(y)$, is equal to the MR model for full independence, such that $d(y) = p(y)$, while for point independence it is a combination of the DS and MR models, $d(y) = p(0)g(y)$. The average detection probability, P_a, within the search strip a, is estimated using $P_a = \int_0^w d(y)\, dy/w$, where w is the truncation distance (i.e., the largest y included in the analyses).

For the DS model, we tested the fit of hazard-rate and half-normal key functions with or without covariates (Buckland et al. 2001). For either key function the scale parameter can be expressed as a function of covariates to model heterogeneity in detection probabilities (e.g. Marques et al. 2007).

We also tested Beaufort sea state both as a linear covariate and as a factor covariate in the DS and MR models. As a linear covariate, we assumed a linear relationship on the link scale between the response and the covariate and estimated one extra coefficient for the respective model. To test as a factor covariate, we fitted the a separate coefficient for each Beaufort sea state (where the coefficient for the lowest observed Beaufort sea state, 1, is absorbed in the intercept) and estimated four extra parameters for the respective model (one each for Beaufort sea states 2-5), allowing more flexibility in the relationship compared to the linear term.

Differences between observation platforms were tested under the independent observer configuration by including *observer* as a covariate in the MR model (note that *observer* cannot be included as a covariate in the DS model as each detected group is only included once in the analyses, regardless of which observer detected it).

Model selection for a given configuration was done using minimum AIC (Akaike 1973), including the level of independence (full or point), choice of key function and covariates for the MR and DS models. The Δ AIC values were calculated as the AIC value of the respective model minus the minimum AIC of all contending models; AIC weights were calculated using methods described in Wagenmakers and Farrell (2004). AIC cannot be used for choosing the configuration (i.e. independent observer, trial or removal) as the data included in the analyses differs between them.

3. RESULTS

A total of 19 days of combined visual and acoustic effort were conducted in passing mode during the 2007 STARLITE survey (Fig. 1). There were 46 detections of rough-toothed dolphins using visual and acoustic detection methods. Visual observation included 18 groups, of which four groups were detected exclusively using visual methods (these groups were not detected using passive acoustics). Acoustic methods detected 42 groups, of which 28 were detected exclusively using acoustic methods.

The frequency of detection decreased with increasing distance from the trackline for all detections (Fig. 2a) and for the visual and acoustic observers (Fig. 2 b, c). For any given distance, a higher proportion of groups were acoustic detections (Fig. 2b vs 2c) and the acoustics team did not miss any of the visual detections at distances greater than 2 km (Fig. 2c).

3.1. Trial configuration

In the trial configuration, where acoustic detections set up trials for visual detections, the best fitting model with the minimum AIC score considered point independence with a hazard-rate detection function without covariates for the DS model and Beaufort fitted as a linear covariate for the MR model (Table 1). An additional five models scored Δ AIC values < 2. The three models with full independence scored the highest AIC values out of the 15 models tested. According to the best model, average trackline detection probability $p(0)$ across all Beaufort sea states was 0.45 (SE=0.16) for the visual team (Table 1).

Table 1. MRDS models under the trial configuration including Δ AIC values (calculated as the AIC – minimum AIC) and AIC weights (Wagenmakers and Farrell 2004), estimates and standard errors (SE) of the trackline detection probability ($p(0)$) and average detection probability (P_a) of the visual team within the covered area. Only models with Δ AIC < 2 are shown; results for all models provided in supplementary table S1. Note that $p(0)$ values are averaged across all Beaufort sea states and that models which did not converge are not listed here. Key to abbreviations: DS = distance sampling, MR = mark recapture, AIC = Akaike information criterion, SE = standard error, HR = hazard-rate, HN = half-normal, Bft = linear *Beaufort*, fac(Bft) = *Beaufort* fitted as factor.

Model	**Level of independence**	**DS model**		**MR model**	**Δ AIC**	**AIC weight**	***p(0)* visual**	**SE**	P_a	**SE**
		Key function	**Covariates: distance+**	**Covariates: distance+**						
PI.DS(HR).MR(Bft)	Point	HR	--	Bft	0.00	0.178	0.45	0.16	0.18	0.08
PI.DS(HN,Bft).MR(Bft)	Point	HN	Bft	Bft	0.34	0.150	0.49	0.16	0.19	0.07
PI.DS(HR,Bft).MR(Bft)	Point	HR	Bft	Bft	0.41	0.145	0.50	0.16	0.21	0.08
PI.DS(HN).MR(Bft)	Point	HN	--	Bft	0.85	0.116	0.45	0.16	0.19	0.07
PI.DS(HR).MR(fac(Bft))	Point	HR	--	fac(Bft)	1.28	0.094	0.52	0.17	0.21	0.09
PI.DS(HN,fac(Bft)).MR fac(Bft))	Point	HN	fac(Bft)	fac(Bft)	1.37	0.090	0.52	0.17	0.18	0.08

A comparison of the detection functions $d(y)$ and $p(y)$ revealed different shapes, suggesting there was additional unmodelled heterogeneity in the data (compare shape of solid black lines in Figure 3). This provided further evidence – besides the lower AIC values – that point independence was the appropriate level of independence.

The estimated trackline probability $p(0)$ for the detection of rough-toothed dolphins by the visual team was estimated for Beaufort sea states 1-5 (Table 2). The estimates of $p(0)$ ranged from 0.71 for Beaufort sea state 1 to 0.09 for Beaufort sea state 5. This study did not include detections during Beaufort sea state 0; hence, we note that the estimated trackline probability of 0.84 for Beaufort sea state 0 was predicted outside the observed range of values.

Table 2. Estimated trackline detection probability $p(0)$ under the trial configuration for the visual team of detecting groups of rough-toothed dolphins shown for individual Beaufort sea states. We note that estimates of $p(0)$ for Beaufort sea state 0 are predicted outside the range of observed values.

Beaufort	**0**	**1**	**2**	**3**	**4**	**5**
p(0) visual	0.84	0.71	0.53	0.34	0.18	0.09

3.2. Independent observer configuration

The best model under the independent observer configuration included the hazard-rate key function without covariates for the DS model (assuming point independence) and linear *Beaufort* and *observer* as covariates for the MR model (Table 3). Two additional models scored Δ AIC values < 2 in comparison to the best model; these included fitting *observer* as a covariate for the MR model or linear *Beaufort* as a covariate in the DS model. We found that the hazard-rate key function typically provided a better fit for any given covariate combination in the DS model, and that fitting *Beaufort* as a linear term was preferred over the corresponding factor term. Including *observer* in the MR model was always preferred over not including it, regardless of other covariates in the MR model. According to the best model, average $p(0)$ estimates across all Beaufort sea states for rough-toothed dolphins were 0.37 (SE=0.14) for visual teams, 0.77 (SE=0.15) for acoustic teams, and 0.84 (SE=0.14) for both teams combined. The average detection probability $\widehat{P}_a$ within the covered area was 0.40 (SE=0.10) for both teams combined.

Table 3. MRDS models under the independent observer configuration, including Δ AIC values (calculated as the AIC – minimum AIC) and AIC weights (Wagenmakers and Farrell 2004), estimates and standard errors (SE) of the trackline detection probability ($p(0)$) for visual, acoustic and combined visual/acoustic, as well as average detection probability (P_a) within the covered area. Only models with Δ AIC < 2 are shown; results for all models provided in supplementary table S3. Note that $p(0)$ values are averaged across all Beaufort sea states and that models which did not converge are not listed here. Key to abbreviations: DS = distance sampling, MR = mark recapture, AIC = Akaike information criterion, SE = standard error, key = key function, HR = hazard-rate, HN = half-normal, Bft = linear *Beaufort*, obs = *observer,* fac(Bft) = *Beaufort* fitted as factor.

Model	**Level of independence**	**DS model**		**MR model**	**Δ AIC**	**AIC weight**	***p(0)* visual**	**SE**	***p(0)* acoustic**	**SE**	***p(0)* combined**	**SE**	P_a ***combined***	**SE**
		Key	**Covariates: distance +**	**Covariates: distance +**										
PI.DS(HR).MR(Bft,obs)	Point	HR	--	Bft+obs	0.00	0.303	0.37	0.14	0.77	0.15	0.84	0.14	0.40	0.10
PI.DS(HR).MR(obs)	Point	HR	--	obs	1.26	0.162	0.41	0.13	0.83	0.10	0.90	0.08	0.43	0.08
PI.DS(HR,Bft).MR(Bft,obs)	Point	HR	Bft	Bft+obs	1.43	0.148	0.38	0.14	0.78	0.14	0.85	0.12	0.41	0.09

The best model for the independent observer configuration showed different shapes in the detection functions $d(y)$, which combines the DS and MR models, and $p(y)$, the MR model only (Fig. 4), again supporting the choice of a model with point independence.

As the best model in the independent observer configuration included both *Beaufort* and *observer* as covariates in the MR model, we were able to predict $p(0)$ values for each observer and Beaufort sea scale (Table 4). The best model predicted that $p(0)$ values were generally lower for the visual team compared to the acoustic team and declined with increasing sea state. Neither team had perfect detection on the trackline; however, combining both teams improved detection to 0.97 for Beaufort sea state 1 to 0.56 in Beaufort sea state 5. We note that this study did not include detections during Beaufort sea state 0; hence, the estimated trackline detection probability of 0.99 for both teams combined for Beaufort sea state 0 was predicted outside the observed range of values.

Table 4. Estimated trackline detection probability $p(0)$ under the independent observer configuration for visual, acoustic, and combined detection of groups of rough-toothed dolphins shown for individual Beaufort sea states. We note that estimates of $p(0)$ for Beaufort sea state 0 are predicted outside the range of observed values.

Beaufort	**0**	**1**	**2**	**3**	**4**	**5**
visual	0.73	0.60	0.46	0.32	0.21	0.13
acoustic	0.95	0.91	0.85	0.77	0.64	0.50
combined	0.99	0.97	0.92	0.84	0.72	0.56

4. DISCUSSION

In our analyses, the best fitting model for both the independent observer and trial configuration assumed point independence and used the hazard-rate detection function (without covariates) for the DS model. For the trial configuration, the best model included Beaufort as a linear covariate for the MR Model (Table 1); the best model for the independent observer configuration included linear Beaufort and observer as covariates for the MR model (Table 3). The average trackline detection probability for rough-toothed dolphins was less than one for both the trial configuration (average $p(0)$ = 0.45) and independent observer configuration (average $p(0)$= 0.37 for the visual, $p(0)$= 0.77 for the acoustic and $p(0)$=0.84 for both teams combined). Furthermore, the trackline detection probability was found to decrease with an increase in Beaufort sea state (Table 2, 4). This study is limited in scope and sample size, and our intention was to explore the potential of these methods. While we do not represent the final values as true, these are reasonable (see Barlow 2015) and add validation to the methods considered.

For shipboard line transect cetacean surveys, estimating the fraction of animals missed by the primary observation team $(1 - p(0))$, is critical to the accuracy of the overall abundance estimate. This study suggests that passive acoustic methods may offer a strong alternative for estimating $p(0)$ for cetaceans. We expect this to be especially true for species with small group sizes and cryptic surfacing behavior, such as rough-toothed dolphins, or in regions where inclement weather may affect sighting conditions. In order to apply these methods to combined visual and acoustic shipboard surveys, a number of considerations must be made.

The first consideration is the type of MRDS configuration: independent observer, trial, or removal configuration. The less desirable removal configuration is generally not appropriate as the different detection probabilities of the visual and acoustic methods violate a fundamental assumption (Laake and Borchers 2004). The preferred independent observer configuration requires that the two observation platforms, in this case the visual and acoustic teams, work independently of each other. In our study, the visual team was independent of the acoustic team; however, the acoustic team was informed of visual observations. Given this one-way independence in our protocol, the trial configuration was the most conservative approach, with the fewest assumptions. Nonetheless, it could be argued that visual and acoustic methods have fundamentally different cues that are not expected to be causally dependent on one another and therefore acoustic detection of animals should not be improved when their presence is identified using visual methods. Application of independent observer configuration allows estimating $p(0)$ for visual and acoustic platforms as well as both platforms combined. Future surveys may consider a protocol that ensures full independence of the platforms.

Next, we must identify the level of independence. Even when the platforms seem independent on first inspection, the detection probabilities for the visual and acoustic platforms may be correlated, which may lead to induced dependence. These correlated variables must be considered in the detection probability model, otherwise they may lead to unmodelled heterogeneity in detection probability and biased abundance estimates (Laake et al. 2011). If unmodelled heterogeneity remains, then the assumption of full independence is not appropriate and must be replaced with the weaker point independence assumption. We can identify potential unmodelled heterogeneity by comparing the shapes of the MR and DS models; the shapes of these models should be the same when no unmodelled heterogeneity remains. In our study, the differences in the shapes of the MR and DS models suggest that additional unmodelled heterogeneity remains and therefore the point independence model is appropriate unless we can identify and include measures for all correlated variables into the MR model.

While variables affecting the detection probability of visual observation platforms are well studied, less is known about the variables affecting acoustic detection of cetaceans. One variable that may explain our apparent unmodelled heterogeneity is group size. There is strong evidence that the probability of visual detection increases with large group sizes (Barlow 1995). Likewise, there is some evidence that small dolphin schools may be less vocal, effectively decreasing the probability of detecting them using acoustic methods (Rankin et al. 2008c). If the detection probability for both observer platforms (here, visual and acoustic detection methods) are affected by group size, then this variable should be included in the MR model. Unfortunately, acoustic estimation of group size is a notoriously difficult endeavor and may be impractical for groups detected using only acoustic methods. In fact, it may be unrealistic to assume that all covariates will be identified and adequately addressed in the model. If unmodelled heterogeneity is identified through the different shapes of the MR and DS models, as is the case in our study, then the study should consider apply point independence.

Abundance estimation of dolphin species during cetacean surveys commonly assumes that all animals on the trackline are observed, or that $g(0)$= 1 (see Barlow 2015). While this assumption may hold for species found in large, conspicuous groups, it may be problematic with species found in smaller group sizes, or those with cryptic surfacing behavior. Rough-toothed dolphins exhibit both of these characteristics, and it has been suggested that the abundance of rough-toothed dolphins may be underestimated using visual observation methods (Rankin et al. 2008b; Barlow 2015). Gerrodette et al. (2008) found that the probability of detection of rough-toothed dolphins should consider group size as a covariate. Likewise, sightability is known to decrease with an increase in Beaufort sea state, especially for animals with these same behavioral characteristics. Barlow (2015) developed models for estimating trackline detection probabilities from distance sampling data without the mark-recapture component and estimated that the fraction of rough-toothed dolphins sighted on the trackline ranged from 0.5 in Beaufort 1 (excellent conditions) to 0.04 in Beaufort 5 (marginal conditions). In his study, Barlow assumed perfect detection on the trackline in Beaufort sea state 0 conditions. While our results were similar for Beaufort sea states 1-5; we estimated $p(0)$ for the visual team in Beaufort 0 to be 0.84 and 0.73 for trial and independent configurations, respectively. Our results suggest imperfect trackline detection in Beaufort 0 conditions. If this is indeed the case, then the trackline detection probabilities estimated by Barlow (2015), which are already lower than our estimates, may in fact be overestimated for rough-toothed dolphins. While these methods show great potential, there are complications related to passive acoustics (in general) and towed array methods, in particular, that require consideration. Here we discuss cross-platform availability bias, range estimation prior to vessel response, understanding covariates related to acoustic detection, differences between passing and closing mode, and complications related to acoustic species classification.

Availability bias is of particular concern when using a visual and acoustic observer for MRDS methods. For many deep-diving species, they may be ‘available’ for acoustic detection while submerged and unavailable for visual detection. Conversely, many of these species are silent during their relatively short surface intervals, when they are available for visual detection. For these situations, the trial configuration, where the acoustics team sets up trials for the visual team, may be more appropriate than the independent observer configuration if acoustic detection distance is sufficiently large to detect most animals at distance. An increased range of availability by the acoustics team may help reduce correlation due to availability by separating the time animals are first available for the acoustics and visual observation team (see Burt et al. 2014). While availability bias between visual and acoustic platforms is well documented for beaked whales and sperm whales (Barlow and Taylor 2005, Marques et al. 2013), we do not expect a cross-platform availability bias for most dolphin species (Rankin et al. 2008b).

A fundamental assumption of distance sampling is that animals are detected prior to any responsive movement to the vessel; this typically requires detection of the group ahead of the vessel. Towed hydrophone arrays are at a disadvantage for detecting and localizing cetaceans directly ahead of the ship due to masking by propeller cavitation, physical interference caused by the ships' hull, and problems inherent with towing hydrophones behind a vessel (Rankin et al. 2008a). Limitations in localization within 30 degrees of the bow are related to the localization methods; these could be mitigated through improved hardware design (e.g., volumetric arrays) and more sophisticated localization methods.

Due to the intrinsic differences in localizing detections using visual and acoustic methods, it is important to identify potential responsive movement of the animals. Responsive movement has potential to affect both the DS and MR models, muddying the effects of unmodelled heterogeneity (Burt et al. 2014). The limited sample size of this study is insufficient for identifying possible responsive movement; however, data pooled from additional SWFSC surveys (Rankin et al. 2008b) suggest rough-toothed dolphins do not exhibit responsive movement to these vessels in these geographic regions (Fig. 5). As visual distances were measured on first detection (ahead of the ship) and acoustic distances were measured as the group passed the beam, acoustic distances larger than visual distances would give an indication of ship avoidance while acoustic distances smaller than visual would give an indication of ship attraction. Consideration of acoustic data for MRDS should include a simple linear regression where acoustic perpendicular distances are regressed against visual perpendicular distances to identify potential responsive movement of animals (see Fig. 5 for details).

To prevent introducing biases in MRDS methods, it is important to include variables that affect acoustic detection probability. The influence of covariates will likely vary by species, and to date only piecemeal descriptions of the limitations to acoustic detection of cetaceans has been examined (e. g., Rankin et al. 2008b). Future work should include examination of these covariates for shipboard surveys using towed hydrophone arrays.

This study examined survey data collected in passing mode, which is atypical for surveys conducted by SWFSC. For closing mode surveys (where visually detected cetacean schools are approached for species identification and group size estimation), it may be possible for acoustic detections of non-sighted groups to be used following the less desirable removal configuration of MRDS if the detection probabilities of the visual and acoustic methods are equivalent (a fundamental assumption for removal configuration). In this scenario, the ‘first observer makes detections of which the second observer is fully aware and the second observer detects observations that are missed by the first observer’ (Laake and Borchers 2004). Considering that equal detection probability by these two methods is unlikely, there may be a need for development of novel statistical methods to apply in these cases.

Consideration of acoustic methods in the MRDS framework requires reasonable acoustic species classification. In this exercise, we assume that all acoustic detections were accurate. While this may be a simple task for some species, especially sperm whales, it is imperfect for most species. While ongoing research is continually improving acoustic classification methods (e.g, Rankin et al. 2017), analytical methods should include addressing errors in acoustic species classification of detections.

Despite the limitations of this study, our results suggest that passive acoustic detection may serve as an independent observer in MRDS studies to estimate $p(0)$ for shipboard cetacean line-transect surveys. In fact, for some species it may be a preferred method to estimate $p(0)$. Studies that intend to apply these methods should take measures to insure full independence of acoustic and visual teams, as the combined results may greatly improve the precision of abundance estimates, especially for cryptic species such as rough-toothed dolphins. Realizing this potential requires further examination of the covariates relevant to model acoustic detections, improvement of acoustic localization and acoustic species identification, and possibly implementing double-blind methods for mark-recapture where the acoustics team is unaware of the visual detections and vice versa.

LITERATURE CITED

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. *International Symposium on Information Theory, 2nd edition* (eds B.N. Petran & F. Csàaki), pp. 267–281. Budapest, Hungary.

Archer FI, Henry AE, Ballance LT (2008) *STenella* Abundance Research LIne Transect and Ecosystem (STAR-LITE) 2007 Cruise Report. National Marine Fisheries Service Center Technical Memorandum NOAA-TM-NMFS-SWFSC-433 (unpublished), 19 pp. Available from SWFSC, 8901 La Jolla Shores Drive, La Jolla, CA 92037, USA.

Barlow J (1995) The abundance of cetaceans in California waters. Part I: Ship surveys in summer and fall of 1991. *Fish Bull* 93:1–14.

Barlow J (2015) Inferring trackline detection probabilities, g (0), for cetaceans from apparent densities in different survey conditions. *Mar Mamm Sci* 31(3):923-943.

Barlow J, Gerrodette T, Forcada J (2001) Factors affecting perpendicular sighting detection distances on shipboard line-transect surveys for cetaceans. *J Cetacean Res & Manag* 3:201-212.

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

Barlow J, Tyack PL, Johnson MP, Baird RW, Schorr GS, Andrews RD, Aguilar de Soto N (2013) Trackline and point detection probabilities for acoustic surveys of Cuvier's and Blainville's beaked whales. *J Acoust Soc Am* 134(3):2486-2496.

Borchers D, Laake J, Southwell C, Paxton C (2006) Accommodating unmodelled heterogeneity in double-observer distance sampling surveys. *Biometrics,* 62, 372-378.

Borchers DL, Zucchini W, Fewster RM (1998) Mark-recapture models for line transect surveys. *Biometrics* 1207-1220.

Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) *Introduction to Distance Sampling*. Oxford University Press, Oxford.

Buckland ST, Laake JL, Borchers DL (2009) Double-observer line transect methods: levels of independence. *Biometrics*, 66, 169-177.

Burt ML, Borchers DL, Jenkins KJ, Marques TA (2014) Using mark–recapture distance sampling methods on line transect surveys. *Methods Ecol Evol* 5(11):1180-1191.

Center for Conservation Bioacoustics. (2011). Raven Pro: Interactive Sound Analysis Software (Version 1.4) [Computer software]. Ithaca, NY: The Cornell Lab of Ornithology. Available from http://ravensoundsoftware.com/.

Gerrodette T, Taylor BL, Swift R, Rankin S, Jaramillo-Legoretta AM, Rojas-Bracho L (2011) A combined visual and acoustic estimate of 2008 abundance, and change in abundance since 1997, for the vaquita, *Phocoena sinus*. *Mar Mamm Sci* 27(2):E79-E100.

Gerrodette T, Watters G, Perryman W, Ballance L (2008) Estimates of 2006 dolphin abundance in the eastern tropical Pacific, with revised estimates from 1986-2003. National Marine Fisheries Service Center Technical Memorandum NOAA-TM-NMFS-SWFSC-422 (unpublished), 39 pp. Available from SWFSC, 8901 La Jolla Shores Drive, La Jolla, CA 92037, USA.

Kinzey D, Olson P, Gerrodette T (2000) Marine mammal data collection procedures on research ship line-transect surveys by the Southwest Fisheries Science Center. National Marine Fisheries Service Center Administrative Report LJ-00-08 (unpublished), 32 pp. Available from SWFSC, 8901 La Jolla Shores Drive, La Jolla, CA 92037, USA.

Laake JL, Borchers DL (2004) Methods for incomplete detection at distance zero. *Advanced Distance Sampling* (eds S.T. Buckland, D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers, & L. Thomas), pp. 108-189. Oxford University Press, Oxford.

Laake JL, Borchers DL, Thomas L, Miller, D, Bishop J (2018) Package 'mrds'. URL https://cran.r-project.org/web/packages/mrds/mrds.pdf.

Laake JL, Collier BA, Morrison ML, Wilkins RN (2011) Point-based mark-recapture distance sampling. *Journal of Agricultural, Biological, and Environmental Statistics,* 16(3), 389-408.

Marques TA, Thomas L, Fancy SG, Buckland ST (2007) Improving estimates of bird density using multiple-covariate distance sampling. *The Auk*, 124(4),1229-1243.

Marques TA, Thomas L, Martin SW, Mellinger DK, Ward JA, Moretti DJ, Harris D, Tyack PL (2013) Estimating animal population density using passive acoustics. *Biological Reviews*, 88(2), 287-309.

Mellinger DK (2001) ISHMAEL 1.0 User's Guide: Integrated System for Holistic Multi-channel Acoustic Exploration and Localization. National Marine Fisheries Service Center Technical Memorandum OAR-PMEL-120, available from Pacific Marine Environmental Laboratory, 2115 SE OSU Dr., Newport, OR 97365, USA.

Pérez JM, Jensen FH, Rojano-Doñate L, Aguilar de Soto N (2017) Different modes of acoustic communication in deep-diving short-finned pilot whales (*Globicephala macrorhynchus*). *Mar Mamm Sci* 33:59-79.

R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

Rankin S, Archer F, Keating JL, Oswald JN, Oswald M, Curtis A, Barlow J (2017) Acoustic classification of dolphins in the California Current using whistles, echolocation clicks, and burst pulses. *Mar Mamm Sci* 33:520-540.

Rankin S, Barlow J (2005) Source of the North Pacific "boing" sound attributed to minke whales. *J Acoust Soc Am* 118:3346-3351.

Rankin S, Barlow J, Oswald J (2008a) An assessment of the accuracy and precision of localization of a stationary sound source using a two-element towed hydrophone array. National Marine Fisheries Service Center Technical Memorandum NOAA-TM-NMFS-SWFSC-416 (unpublished), 29 pp. Available from SWFSC, 8901 La Jolla Shores Drive, La Jolla, CA 92037, USA.

Rankin S, Barlow J, Oswald JN, Ballance L (2008b) Acoustic studies of marine mammals during seven years of combined visual and acoustic line-transect surveys for cetaceans in the eastern and central Pacific Ocean. National Marine Fisheries Service Center Technical Memorandum NOAA-TM-NMFS-SWFSC-429 (unpublished), 58 pp. Available from SWFSC, 8901 La Jolla Shores Drive, La Jolla, CA 92037, USA.

Rankin S, Oswald JN, Barlow J (2008c) Acoustic behavior of dolphins in the Pacific Ocean: Implications for using passive acoustic methods for population surveys. *Can Acoust* 36(1):88-92.

Rankin S, Oswald JN, Simonis AE, Barlow J (2015) Vocalizations of the rough-toothed dolphin, *Steno bredanensis*, in the Pacific Ocean. *Mar Mamm Sci* 31:1538-1548.

Van Parijs SM, Clark CW, Sousa-Lima RS, Parks SE, Rankin S, Risch D, Van Opzeeland IC (2009) Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales. *Mar Ecol Prog Ser* 395:21-36

Wagenmakers EJ, Farrell S (2004) AIC model selection using Akaike weights. Psychonomic Bulletin & Review, 11(1):192-196

FIGURE HEADINGS

Figure 1 Map of STARLITE study area and tracklines, with detections of rough-toothed dolphins by visual (red), acoustic (black), or both methods (green)

Figure 2 Detection frequencies made by a) either platform (visual and/or acoustics), b) visual (observer 1) conditional on detection by acoustics (observer 2), and c) acoustics conditional on detection by visual

Figure 3 Detection functions for best fitting model for the trial configuration (point independence with the hazard-rate key function and no covariates besides distance for the DS model and linear Beaufort as a covariate for the MR model). Left: solid line shows the average detection function $d(y)$ across all Beaufort sea states and circles represent the detection probabilities $d(y)$ for the individual detections. Right: solid line shows the average conditional detection function $p(y)$ across all Beaufort sea states and circles represent the $p(y)$ for the individual detections

Figure 4 Detection functions for best fitting independent observer model (point independence with a hazard-rate key function/no covariates for the DS model and linear Beaufort and observer as covariates for the MR model). Top two rows: black lines are the detection function $d_j(y) = p_j(0)g(y)$ averaged over all sea states scaled by the different $p_j(0)$ for the different observers, i.e. (a) for visual only, (b) acoustic only, (c) both teams pooled and (d) duplicates between teams; circles indicate the detection probabilities $d(y)$ for each observation, histogram bars represent the relative frequencies of detections. Bottom row: black lines are the conditional detection function $p(y)$ averaged over all sea states and circles indicate the detection probabilities $p(y)$ for each observation, histogram bars represent the proportion of all groups detected by observer 2 that were also detected by observer 1 (e) and vice versa (f)

Figure 5 Scatterplot of perpendicular trackline distance estimated using acoustic methods (x-axis) and visual methods (y-axis) for combined visual/acoustic sightings of rough-toothed dolphins during passing mode on SWFSC (truncated within 3 km of the vessel). The green line is represents the best fit of a linear model to the data where acoustic perpendicular distances were regressed against visual perpendicular distances; the black line identifies a potential 1:1 relationship. Points that fall below the black line represent potential vessel avoidance; points that fall above the line represent potential vessel attraction

FIGURES

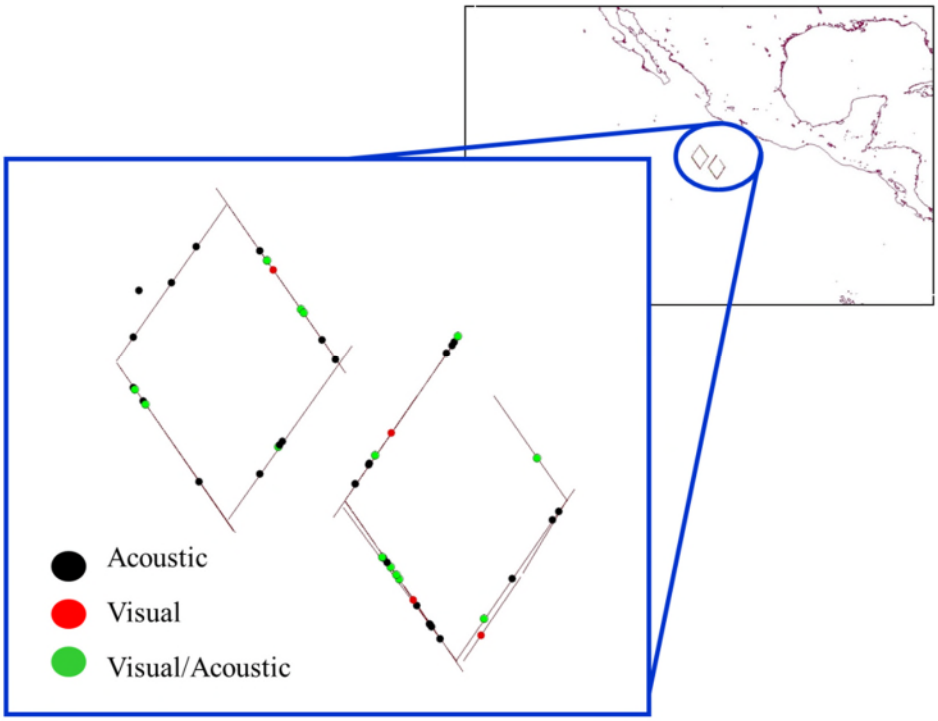


Figure 1

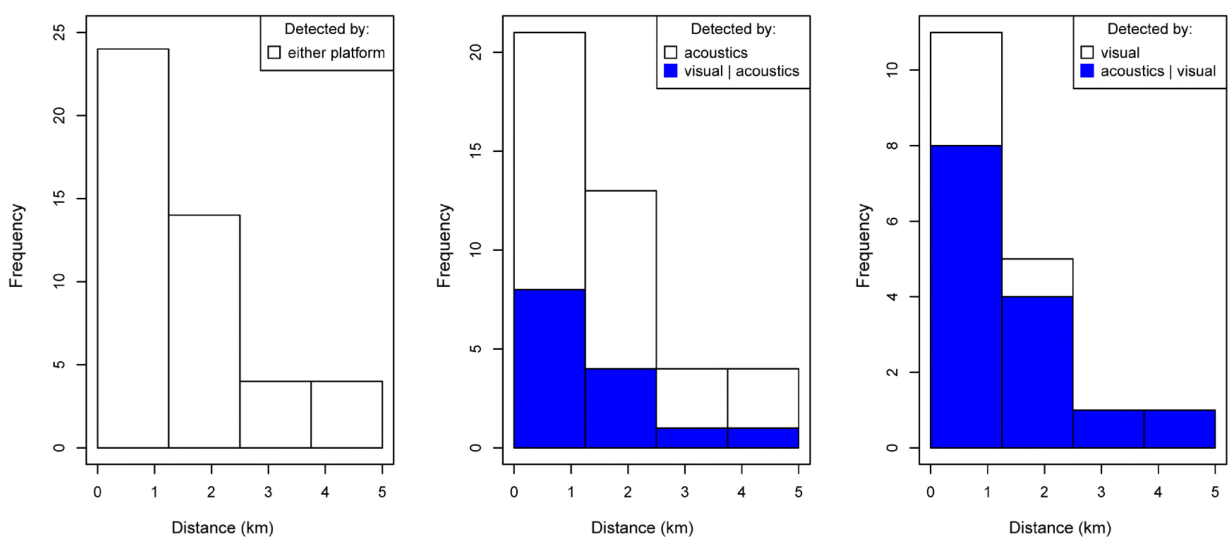


Figure 2

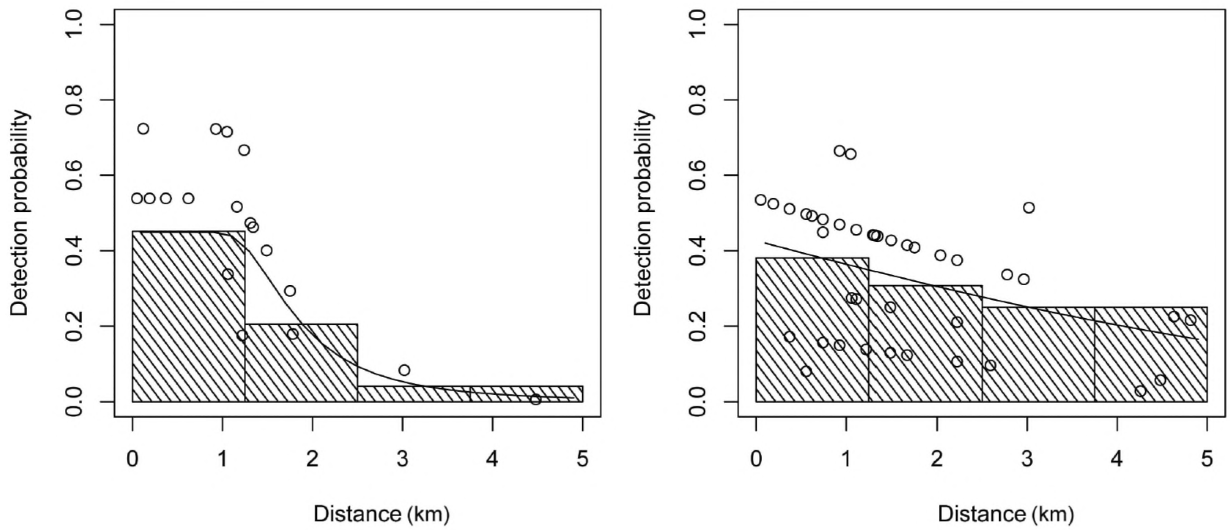
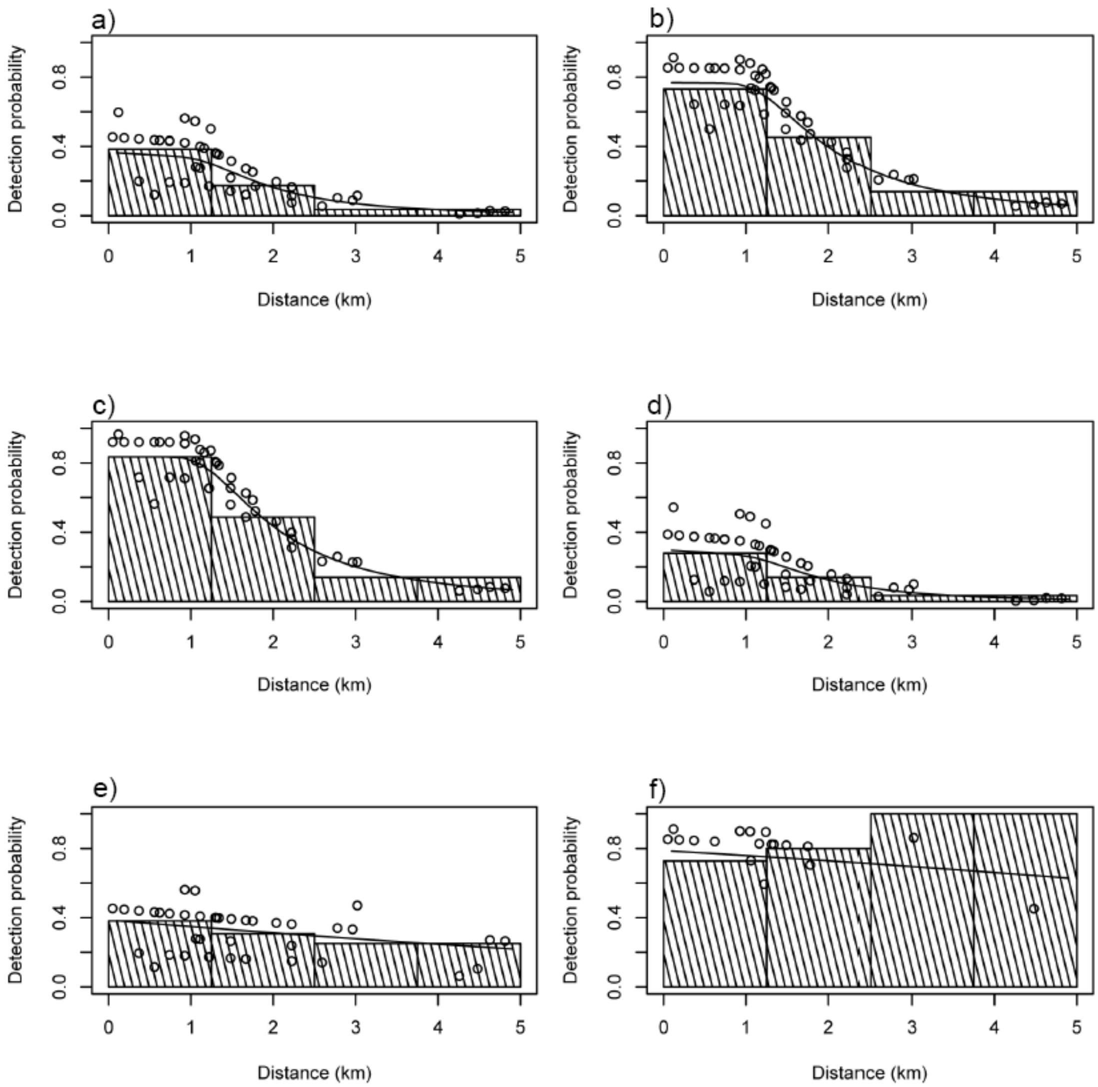

Detection probability
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
Distance (km)
Detection probability
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
Distance (km)

Figure 3



639 **Figure 4**

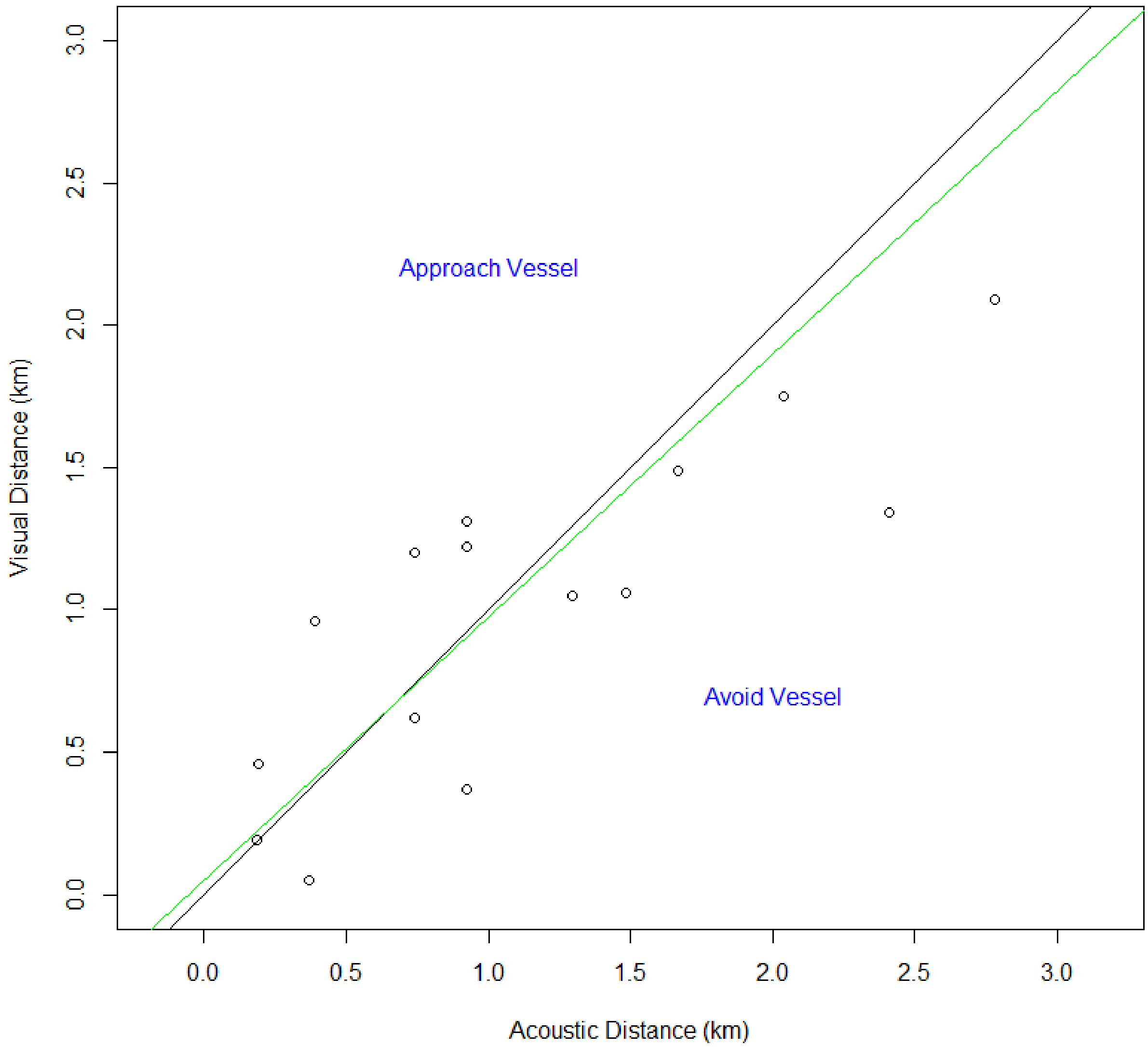

Approach Vessel
Avoid Vessel
Visual Distance (km)
Acoustic Distance (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Figure 5