

## Scaling the Laws of Thermal Imaging–Based Whale Detection

DANIEL P. ZITTERBART,<sup>a,b,c</sup> HEATHER R. SMITH,<sup>d</sup> MICHAEL FLAU,<sup>a</sup> SEBASTIAN RICHTER,<sup>b,c</sup>  
ELKE BURKHARDT,<sup>a</sup> JOE BELAND,<sup>d</sup> LOUISE BENNETT,<sup>e</sup> ALEJANDRO CAMMARERI,<sup>b,f</sup>  
ANDREW DAVIS,<sup>d</sup> MEIKE HOLST,<sup>d</sup> CATERINA LANFREDI,<sup>g</sup> HANNA MICHEL,<sup>a</sup> MICHAEL NOAD,<sup>e</sup>  
KYLIE OWEN,<sup>e,h</sup> AUDE PACINI,<sup>i</sup> AND OLAF BOEBEL<sup>a</sup>

<sup>a</sup> Ocean Acoustics Group, Alfred-Wegener-Institut Helmholtz Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany

<sup>b</sup> Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

<sup>c</sup> Biophysics Lab, University of Erlangen-Nuremberg, Erlangen, Germany

<sup>d</sup> LGL Limited, Toronto, Ontario, Canada

<sup>e</sup> Cetacean Ecology and Acoustics Laboratory, School of Veterinary Science, University of Queensland, Gatton, Queensland, Australia

<sup>f</sup> Fundacion MaryBio, Buenos Aires, Argentina

<sup>g</sup> Consorzio Nazionale Interuniversitario per le Scienze del Mare, and Department of Civil and Environmental Engineering, Politecnico di Milano, Milan, Italy

<sup>h</sup> Australia Aquatic Megafauna Research Unit, School of Veterinary and Life Sciences, Murdoch University, Murdoch, Western Australia, Australia

<sup>i</sup> Hawaii Institute of Marine Biology, University of Hawai'i at Mānoa, Honolulu, Hawaii

(Manuscript received 10 April 2019, in final form 6 February 2020)

### ABSTRACT

Marine mammals are under growing pressure as anthropogenic use of the ocean increases. Ship strikes of large whales and loud underwater sound sources including air guns for marine geophysical prospecting and naval midfrequency sonar are criticized for their possible negative effects on marine mammals. Competent authorities regularly require the implementation of mitigation measures, including vessel speed reductions or shutdown of acoustic sources if marine mammals are sighted in sensitive areas or in predefined exclusion zones around a vessel. To ensure successful mitigation, reliable at-sea detection of animals is crucial. To date, ship-based marine mammal observers are the most commonly implemented detection method; however, thermal (IR) imaging–based automatic detection systems have been used in recent years. This study evaluates thermal imaging–based automatic whale detection technology for its use across different oceans. The performance of this technology is characterized with respect to environmental conditions, and an automatic detection algorithm for whale blows is presented. The technology can detect whales in polar, temperate, and subtropical ocean regimes over distances of up to several kilometers and outperforms marine mammal observers in the number of whales detected. These results show that thermal imaging technology can be used to assist in providing protection for marine mammals against ship strike and acoustic impact across the world's oceans.

### 1. Introduction

Ship strikes of large whales are becoming an increasing problem for whale populations as ship traffic is increasing globally [Frisk 2012; Fais et al. 2016; Dawson et al. 2018; United Nations Conference on Trade and Development (UNCTAD); UNCTAD 2018], and are particularly problematic for highly threatened populations where each individual is crucial for survival of the species (Cates et al. 2017). High-level underwater acoustic sources for marine geophysical prospecting have the potential to elicit injuries or negative physiological or

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JTECH-D-19-0054.s1>.

Corresponding author: Daniel P. Zitterbart, dpz@whoi.edu

DOI: 10.1175/JTECH-D-19-0054.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

behavioral responses in marine mammals (Richardson et al. 1995; Erbe et al. 2018; Southall et al. 2019). Naval midfrequency sonar is criticized for its potentially negative effect on marine mammals and has been implicated in several whale stranding events. To minimize possible adverse impacts on individuals and their populations (e.g., D'Amico et al. 2009; Miller et al. 2012), competent authorities commonly require the implementation of mitigation measures, including vessel speed reductions and shutdown of acoustic sources, if marine mammals are sighted in high-risk areas or in a predefined exclusion zone around the vessel (Weir and Dolman 2007; Laist et al. 2014; Constantine et al. 2015).

For successful mitigation, reliable detection of the animals at sea is crucial. Currently there are two established methods to detect marine mammals from vessels: 1) visual detections that are made by human marine mammal observers (MMOs) scanning the ocean's surface with the naked eye or binoculars and 2) acoustic detections of underwater vocalizations made using passive acoustic monitoring (PAM). Both methods have recognized weaknesses. MMOs are unable to make detections in darkness, and are likely to miss animals when they are fatigued or looking in the wrong direction. PAM is only effective when marine mammals vocalize frequently, and when vocalizations are not masked by vessel or other background noise. With the goal of improving marine mammal detections beyond using these traditional methods, studies in recent years have evaluated the use of thermal imaging cameras to detect marine mammals at the ocean's surface (Verfuss et al. 2018) and to make such detections automatically (Santhaseelan et al. 2012; Zitterbart et al. 2013). Thermal imaging systems have been used to detect marine mammals during nighttime hours for a few decades (Perryman et al. 1999; Schoonmaker et al. 2008), and an automatic detection system has been described and used in recent years (Zitterbart et al. 2013; Smith et al. 2020), but methods have not yet been standardized.

Thermal [infrared (IR)] imaging-based detection relies on detecting the whale's blow during surfacing. A whale's blow is visible as a transient, apparently warm feature at the water's surface (Horton et al. 2017). We have previously described the development of a fully automated thermal (IR) imaging-based marine mammal detection system for use in polar and subpolar waters (Zitterbart et al. 2013). But its detection performance across the spectrum of environmental conditions encountered remained unknown. The environmental conditions expected to impact thermal (IR)-based marine mammal detection performance include sea surface temperature, relative humidity, visibility and wind force. In this study, we examine the effects of

environmental conditions on the capability of a human to perceive a whale cue (e.g., blow, back, splash, breach) in a thermal (IR) imaging data stream at different distances. In addition, we evaluate the influence of environmental conditions on the performance of automatic thermal (IR) imaging-based detection, derive an algorithm to facilitate automatic detection under varied environmental conditions, and compare the detection function of the automatic detection system with an experienced MMO (eMMO). To assess the detection algorithm, we investigate how many detections are made at different distances. Finally, to assess the whole system in a real-time setting, we test the performance of the automatic thermal (IR) imaging whale detection system against eMMOs in a dual platform approach.

## 2. Materials and methods

### a. Field sites

This study includes experiments at four field sites during three consecutive field programs in 2014, 2015, and 2016 (Table 1). The field sites are North Stradbroke Island (NSI), Cape Race (CR), Poipu Shores (PO), and Princeville (PR). In 2014 the experiment was conducted on NSI, Queensland, Australia, where visual observation data were collected from 17 June to 14 July. Point Lookout on NSI provides a suitable location for such observations, as the eastern Australian population of humpback whales (*Megaptera novaeangliae*) passes within a few kilometers of this point during their northward migration (Noad et al. 2019). Observations were made from the decks of a house situated on a cliff (Fig. S1 in the online supplemental material). An 80° sector of the ocean could be observed. The thermal imager was mounted at a height of 51.3 m above mean sea level (MSL). The height was measured using a theodolite and a known GPS location at sea level.

The second field program was conducted at CR, Newfoundland, Canada. In 2015, data were collected from 18 July to 23 August; and in 2016, from 1 to 31 July. The likelihood of spotting marine mammals, primarily humpback whales, is high during the summer months at Cape Race. Data were collected from a vantage point on the edge of a cliff ~26 m MSL (determined using a handheld GPS). An ~198° sector of the ocean could be observed. In 2015, visual data were collected by human observers using a theodolite. In 2016, visual data were collected by MMOs using binoculars from inside purpose-built observation booths. In both years, the thermal (IR) imager was mounted on a level platform near the cliff edge (Fig. S2).

The third field program was conducted on Kauai, Hawaii, from 18 January to 1 March 2016. Two field sites

TABLE 1. Field sites where data were collected to investigate the influence of environmental conditions on thermal (IR) imaging-based whale detection performance.

Location	Latitude	Longitude	Data collection period	Thermal regime air temperature (AT)/sea surface temperature (SST) <sup>a</sup> (°C, mean $\pm$ SD)	Metrics quantified	Elevation (m, MSL)
North Stradbroke Island (NSI)	27°25.912'S	153°32.540'E	9 Jun–14 Jul 2014	Subtropical AT: $17.16 \pm 1.68^b$ SST: $21.84 \pm 0.94^c$	Thermal perceptibility	51.3
Cape Race (CR)	46°39.472'N	53°04.404'W	22 Jul–23 Aug 2015 1–31 Jul 2016	Temperate AT: $13.21 \pm 1.99^d$ SST: $14.32 \pm 1.55$ AT: $12.20 \pm 2.28$ SST: $10.56 \pm 2.11$	Thermal perceptibility Integrated system performance	26.0
Poipu Shores (PO)	21°52.231'N	159°26.873'W	18 Jan–6 Feb 2016	Tropical AT: $21.83 \pm 2.62^e$ SST: $25.72 \pm 0.41$	Automatic detection	16.0
Princeville (PR)	22°13.689'N	159°28.937'W	10–19 Feb 2016	Tropical SST: $24.98 \pm 0.20^f$	Automatic detection Integrated system performance	49.8

<sup>a</sup> Mean AT and SST for the data collection period at each site.

<sup>b</sup> Bureau of Meteorology (BOM), Australian Government 2018. Climate data online. Accessed at <http://www.bom.gov.au/climate/data/>. Accessed 11 December 2018. Point Lookout Station 40209 at 27.44°S, 153.55°E. Minimum and maximum daily air temperatures (mean of daily mean).

<sup>c</sup> Office of Environment and Heritage (OEH), New South Wales Government 2018. Manly Hydraulics Laboratory Waverider buoys—Sea surface temperature data. Accessed at <https://portal.aodn.org.au/>. Accessed 11 December 2018. 28.871°S, 153.694°E. Data logged in 1-h intervals.

<sup>d</sup> <http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/waves-vagues/data-donnees/data-donnees-eng.asp?medsid=C44251>. 46.44°N, 53.39°W. Data logged at 60-min intervals.

<sup>e</sup> [https://www.ndbc.noaa.gov/station\\_history.php?station=nwwh1](https://www.ndbc.noaa.gov/station_history.php?station=nwwh1). 21.954°N, 159.353°W. Data logged at 6-min intervals.

<sup>f</sup> [https://www.ndbc.noaa.gov/station\\_history.php?station=51208](https://www.ndbc.noaa.gov/station_history.php?station=51208). 22.285°N, 159.574°W. Data logged at 30-min intervals.

were chosen to sample the different environmental conditions experienced at the north and south shores of the island. The PO site (Fig. S3) in the south is usually on the leeward side during winter and therefore characterized by low sea states and good visibility (camera 16.0 m MSL). The PR site (Fig. S4) in the north is characterized by high swell and significantly more rain and wind than in the south (camera 49.8 m MSL). All field programs were geared toward observation of humpback whales for comparability.

### b. Metrics

We quantified three metrics:

- 1) Thermal perceptibility—How well a whale cue (e.g., blow, splash, back, breach) in the thermal (IR) video stream is perceived by an informed human observer.
- 2) Automatic detection performance—How well the automated detection system performs as a function of distance.
- 3) Integrated system performance—How the performance of a thermal (IR) imaging-based, automated whale detection system compares to a human observer.

The quantification of each metric required its own experimental protocol. To quantify *thermal perceptibility*,

we measured how many whale cues were detected and localized by human observers (marked) and could retrospectively be perceived as an unambiguous thermal anomaly in the thermal (IR) data stream (recaptured). A team of human observers scanned the ocean's surface for whales using binoculars [Fujinon 7  $\times$  50 FMTRC-SX; field of view (FOV) = 7°30'] or the naked eye. When a whale was spotted, a theodolite was used to measure the bearing and elevation angles to subsequent cues (marked). Timing of sightings of subsequent whale cues was also recorded. After each observation shift, the human observer team reviewed the thermal (IR) data stream. The team fast forwarded to the time stamp of each recorded cue and searched for a thermal anomaly that they could unambiguously identify (recapture). Thermal (IR) data review was performed using Fedallah software, specially designed to allow for ad lib navigation (forward, pause, backward, zoom, rotate) within the thermal (IR) data stream. Thermal anomalies initially classified as being created by a whale were then classified further as being either an aerial display (i.e., breach, half breach, pectoral slap, tail slap, back) or whale blow. Thermal perceptibility review typically required less time than the visual observation shift and was

conducted immediately (i.e., no later than 24 h) after the observation shift.

Performance was quantified as conditional probability of detection (perception) for each cue (i.e., the probability that a cue was perceived in the IR stream given that it was observed (marked) by an observer using the theodolite). We also calculated the conditional probability of detecting groups of whales, which is a more relevant metric in the mitigation context because it is only necessary to unambiguously detect one cue produced by a group of whales within the mitigation zone. Multiple cues by individuals within a group highly decrease the availability bias of the group relative to a single cue. To calculate conditional probabilities for groups of whales, each cue was assigned to a specific group of whales tracked by the visual observer team.

The human observers were instructed to not record all available whale cues, but to instead focus on making fewer but higher accuracy geolocations. Cues missed by the human observers were not of interest and not accounted for in this experiment.

To assess *automatic detection performance*, the detection algorithm was run in real time on the thermal (IR) data stream. An audible alert was played if the algorithm classified a thermal anomaly as a whale cue. The location of the detection was projected onto an image of the thermal imagers FOV, and a “video snippet” of the thermal anomaly was displayed. The 6 s video snippet had a 5° FOV centered on the thermal anomaly. The video snippet and detection location were displayed until a human observer classified the detection as a true or false positive. These data were used to generate a detection function (Thomas et al. 2002), as would be done with data collected following a point-transect distance sampling protocol. Assuming whales are equally distributed across the observational area, one would expect the number of detections to grow linearly with distance (Buckland et al. 2001), as the area of a ring grows linearly with increasing distance. Performance is characterized by the detection function, and the location of its peak.

The detection system consists of several components that can influence its performance. Assessing the *integrated system performance* is important because it integrates the overall benefits and shortcomings. To assess the integrated system performance, we conducted double-blind experiments using three independent platforms. All platforms recorded time, bearing, and distance of sightings. Platform one was an eMMO who scanned the ocean surface with the naked eye or binoculars. Platform two was the thermal (IR) system with a human operator. The human operator validated automatic detections as true or false positive in real time. The third

platform was an assisted MMO (aMMO) equipped with a tablet computer that received unvalidated automatic thermal (IR) imaging-based whale detections along with metadata information (time stamp, computed distance and bearing to the detection). The aMMO could decide to use the thermal (IR) detections or not, just as it would be implemented for mitigation purposes during shipboard operation. Performance was measured as conditional probability  $P(A|B)$ , denoting the probability of method *A* detecting a cue (recapturing) under the condition that method *B* detected (marked) the same cue. Sighting matches between methods (*A* and *B*) were performed on a basis of geolocation of the whale or group. Thresholds for defining matches at PR were 500 m distance between localizations and 3 min between the time of sightings. The 3 min time window between sightings allows for the MMO to verify the observation, which is usually done by the MMO when observing the next cue. MMOs often did not note the time of first cue sighted as it would have distracted them from detecting the next cue. Thresholds for defining matches at CR are 20% of the distance at which the cue was detected, to allow for a buffer zone larger than the error of the distance estimation using the binoculars (~8% at 3 km distance; see Zitterbart et al. 2013). The protocol was designed to mimic current marine mammal mitigation approaches. All platforms were visually and acoustically separated so information about detections were not passed between observing platforms in order to avoid bias.

At the PR site, we evaluated conditional probability as a function of the distance up to which detections were considered; only detections that were in the FOV common to all three methods were analyzed. Pairwise conditional probabilities were calculated for the three methods. At the CR site, comparisons were made between detections made by the thermal (IR) system and the eMMOs. We evaluated conditional probability as a function of distance for baleen whales with and without minke included, and for all marine mammal species observed. We also evaluated conditional probability on a daily basis.

### c. Thermal (IR) imager

Thermal imaging data were acquired using a rotating line scanner (FIRST-Navy, Rheinmetall Defense Electronics, Bremen, Germany) mounted on a tripod on a stable platform. All experiments were conducted from land so no active stabilization of the thermal imager was required. Data acquisition and processing were performed with a custom software (Tashtego). The cryogenic sensor is cooled to 84 K using a Sterling cooler. It scans 360° horizontal × 18° vertical at



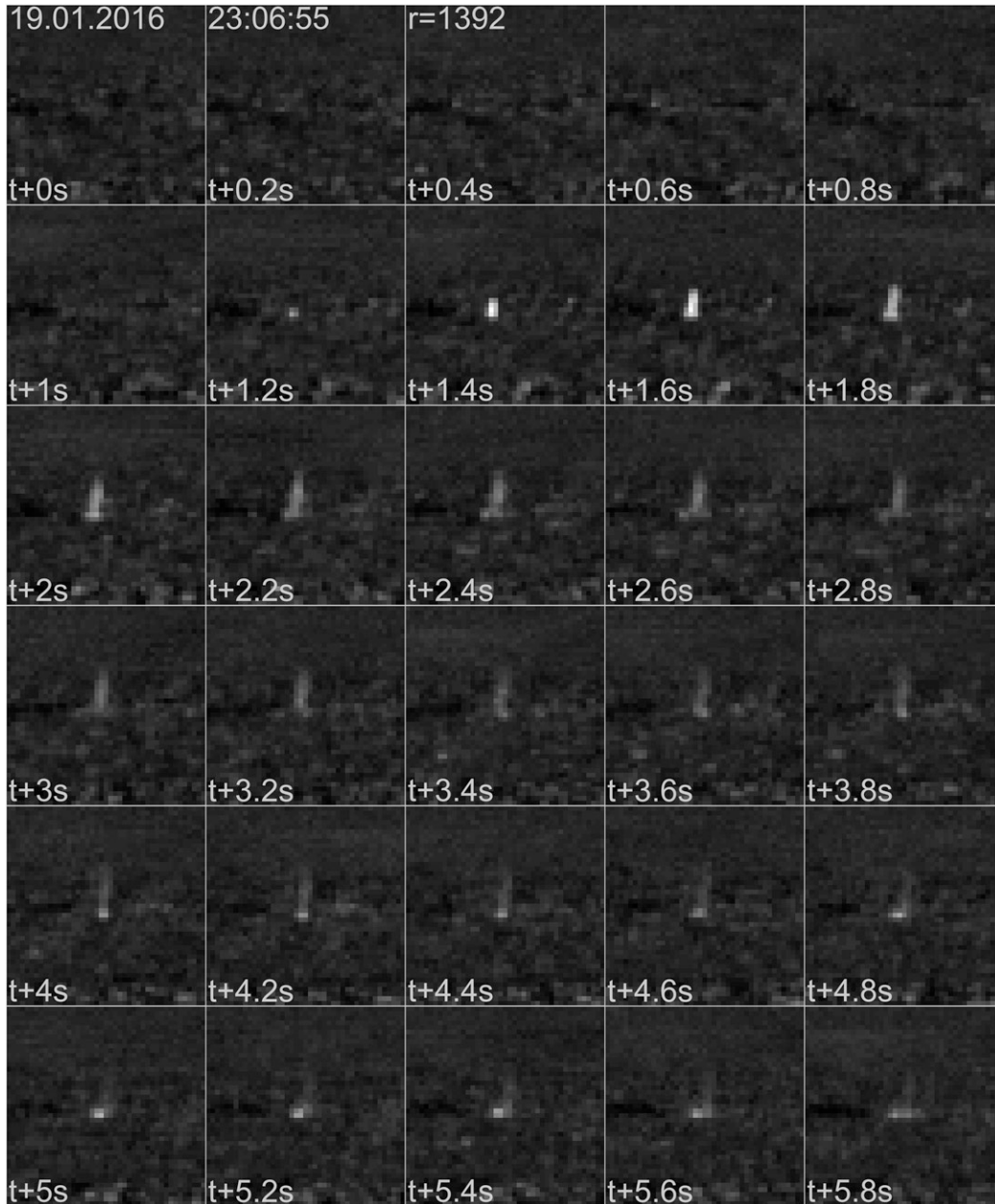


FIG. 1. Thermal image of a humpback whale blow in 1392 m distance at night. Subwindows are 0.2 s apart and centered around the thermal anomaly (i.e., whale blow). With time (top left to bottom right) the whale blow rises in size and intensity (perceptible as a white feature) for about 0.8 s before diminishing and fading away.

5 revolutions per second, providing a 5 Hz video stream of the thermal field of the sensors environs at horizontal and vertical resolutions of  $0.05^\circ$  and  $0.03^\circ$  per pixel, respectively.

#### d. Detection algorithm

The detection software exploits the fact that a whale blow, under the condition that the observer is moving

slowly, is transient in time, but stable in space. The thermal video stream is divided in “subwindows” of different sizes (Rowley et al. 1995). In each of these subwindows, the contrast radiance  $I_{\langle \max, \max-N \rangle} - I_{\text{median}}$  is calculated and tracked over 6 s (Fig. 1). The number of pixel  $N$  used to calculate  $I_{\langle \max, \max-N \rangle} - I_{\text{median}}$  is selected dependent on tile size. A subwindow is marked as a candidate  $s(i, t) = 1$  if the contrast  $C_i$  in that

subwindow is significantly ( $2\sigma$ ) greater than that in the horizontally adjacent subwindows [Eq. (1)]. A detection is made  $d(i, t) = 1$  if more than  $th$  subwindows within the 6-s tracking time are marked as a candidate [Eq. (2)]. The threshold  $th$  can be user-defined or auto-determined to yield to a constant false alert rate:

$$s(i, t) \begin{cases} 1 & c_i > \langle c_{i-1}, c_{i+1} \rangle + 2\sigma \\ 0 & c_i < \langle c_{i-1}, c_{i+1} \rangle + 2\sigma \end{cases} \quad (1)$$

$$d(i, t) \begin{cases} 1 & s_{i,t(0,\dots,30)} \geq th \\ 0 & s_{i,t(0,\dots,30)} < th \end{cases} \quad (2)$$

This algorithm performed well with regard to the consistent detection of whale blows (see section 3), but it proved to be rather susceptible to false positives caused by semistatic objects (e.g., moving palm tree leaves) or slow-moving objects in large swell (e.g., stand-up paddlers, small fishing vessels). To reduce the number of false positives, we created false alert suppression maps by applying a set of heuristic rules. Alerts from recurrent locations (e.g., objects on shore, breaking waves on rocks) were suppressed by simply recording the number of alerts in each subwindow and removing subwindows with regularly recurrent alerts. Alerts from nontarget moving objects (i.e., ships and small watercraft and birds) were suppressed by tracking these objects using a combination of Kalman filters and the Hungarian algorithm. While this worked well for most individual objects, the simple tracking algorithm is not capable of following multiple overlapping tracks or tracing diving birds with flight paths perpendicular to the water surface.

#### e. Time synchronization

Time synchronization was crucial for all experimental protocols. Most equipment (IR scanner, weather station, MMO watch) used GPS-based time (all set to local time) during all field programs. Other equipment (e.g., the theodolite) was synchronized by comparing time stamps at the beginning of each observation shift.

#### f. Environmental parameters

Wind speed data for the NSI field site were downloaded from the Cape Moreton Lighthouse Station 40043 at 27.0314°S, 153.4661°W (Australian Bureau of Meteorology); data were logged at 30 min intervals. Wind speed and relative humidity data for the CR field site were downloaded from the Cape Race Weather Station 8401000 at 46.660°N, 53.076°W (Environment Canada); data were logged at 30 min intervals. Wind speed data for PO were downloaded from Lihue Airport Weather Station at 21.98389°N,

159.34056°W (NOAA); data were logged each 15 min. Wind speed data for PR were downloaded from NOAA offshore buoy 51WH0 (WHOTS) at 22.759°N, 157.917°W (NOAA); data were logged hourly. MMOs at CR recorded precipitation type (rain, drizzle, fog, none) and sightability (subjective judgement of overall viewing conditions: excellent, good, moderately impaired, severely impaired, impossible) every half hour while on effort. A Vaisala FS11 sensor recorded visibility at the CR site.

### 3. Results

#### a. Thermal perceptibility

At the NSI site, we found that humpback whale cues at up to 10 km distance can be perceived as a thermal anomaly in the IR data stream. The conditional probability of perceiving a thermal anomaly matched to a visual observation, that is,  $P(\text{IR}|\text{VIS})$  decreases with increasing distance between the observer and the whale. The probability of perceiving a group of humpbacks reduced from approximately 0.95 for groups closer than 1 km, to 0.22 to for groups at 10 km (Fig. 2).

The cue-based analysis at the NSI site reveals that  $P(\text{IR}|\text{VIS})$  of cues caused by the displacement of relatively large amounts of water (like whale breaches and slaps) is less affected by distance. The perceptibility of whale blows (Fig. 3a) shows a linear decay, while increased wind force affects perceptibility negatively with increased decay in perceptibility per distance unit. Breaches are generally very visible in the thermal imaging data and  $P(\text{IR}|\text{VIS})$  can be as high as 1.00 at up to 7 km distance (Fig. 3d); there was no clear correlation between  $P(\text{IR}|\text{VIS})$  and distance. Slaps (either with the fluke or pectoral flippers, denoted as whale surface display) are the second most perceptible cue; their  $P(\text{IR}|\text{VIS})$  is larger than 0.95 at a 2–3 km range, but drops significantly at greater distances (Fig. 3c). The conditional probability  $P(\text{IR}|\text{VIS})$  for breaches and slaps is not affected by increased wind force. The cue least likely to be perceived in thermal (IR) data is a humpback whale back (Fig. 3b). Note the cue-based analysis only included data up to 8 km as sample sizes farther out were small.

Probability of perception at the CR site (Fig. 4) follows the same pattern. The cue-based analysis reveals that the maximum distance at which cues were perceived in significant numbers was 5 km. At 3 km, strong blows were 1.3 times as likely to be detected as weak blows. During the CR field season, we also made detections of minke whales (*Balaenoptera acutorostrata*;  $N = 102$ ). Minke whales could be perceived at up to

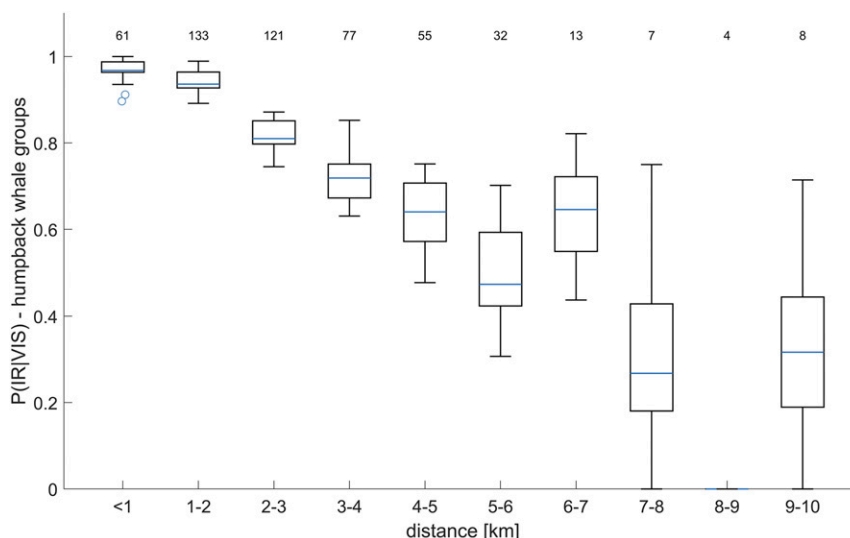


FIG. 2. Probability of perception for whale groups depending on their distance to the observer at the NSI site. Error bars represent standard deviations obtained by bootstrapping. Numbers in top line give the number of encounters in each bin.

800m distance, but with much reduced probability of perception ranging between 0.2 and 0.5.

#### b. Detectability

The detection function describing the performance of the detection algorithm follows a shape that is expected from a point-transect distance sampling survey design. With increasing distance, the number of true positive detections increases due to the increase in area monitored at that distance. We find that the peak of the detection function varies significantly across sites. At the CR field site, peaks are found at 0.5 km during the day and 2 km during the night (Fig. 5a). At the PO and PI field sites, the detection functions show an increase up to 2 and 3 km, respectively (Fig. 5a). Farther out the number of detected cues decreases, with the furthest cue detected at 6 km.

We analyzed the influence of Beaufort wind force (BFT) on detectability at the PO and PR sites. At the PO site, cues were detected at similar distances (~2 km) at wind forces BFT1 and BFT2. Cues were sighted at closer distances as wind force increased (Fig. 5b). We did not encounter wind forces greater than BFT4 at this location. At the PR site, we did not find any influence of wind force on the shape of the detection function (Fig. 5c). During periods with wind force  $\geq$  BFT6, only 8 automatic detections were classified as true positive and were therefore omitted from the analysis (data not shown).

At the CR site we observe a significant difference in the shape of the detection function for the IR system between day and night (Fig. 5a). The detection function during the day is similar to the detection function for the

eMMOs (Fig. 5d), which shows a peak at 0.5–1 km, and another peak at 6 km. The thermal (IR) detection function shows a similar peak at 500 m, but does not show the peak at 6 km. This is consistent with the perceptibility results at CR site, which showed that the maximum distance any cue was perceived by the IR system was 5 km. The second peak observed in the thermal (IR) detection function at 2 km (Fig. 5d), corresponds to nighttime detections (Fig. 5a), during which no MMO data are available. MMO observations show different detection functions for humpback and minke whales (Fig. 5e), with very few sightings beyond 2 km for minke whales.

Comparison of the average detection function with visibility measured with the FS11 visibility sensor shows that in visibility conditions  $<5$  km, the average detection distance for MMOs and the automatic thermal (IR) detections is  $<500$  m. In visibility conditions  $>7$  km, the mean thermal (IR) detection distance increases to 2 km, and MMO detection distances increase to 1 km. In conditions with visibilities  $>10$  km, the average MMO detection distance increases to 2 km (Fig. 6a). When sightability (see section 2) was classified as impossible or severely impaired, both thermal (IR) imaging and MMOs had average detection distances  $<500$  m. In moderate, good and excellent sighting conditions, detection ranges were comparable with the mean around 2 km (Fig. 6b).

#### c. False positive analysis

Automatic detection algorithms unavoidably produce false alerts. The false positive (fp) rate (or false alarm rate)

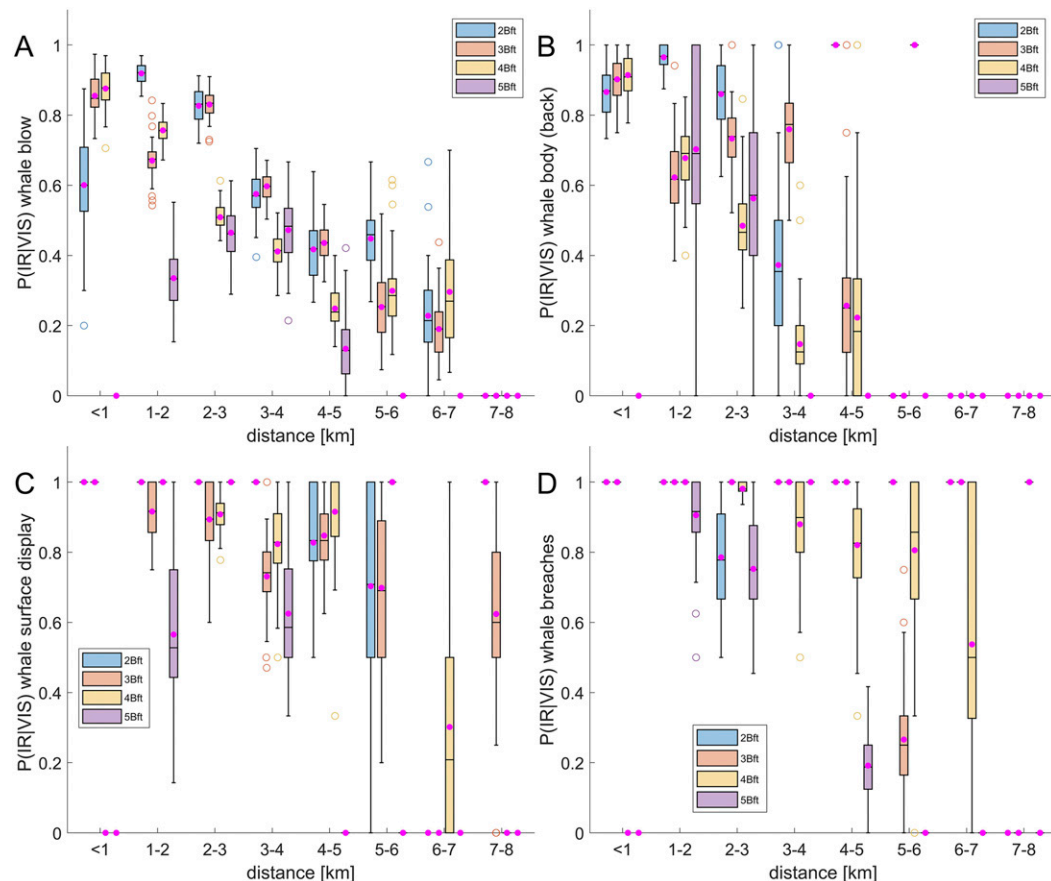


FIG. 3. Probability of perception for different cues and wind speeds. Pink dots denote the mean value of the distribution in each bin. Cues: (a) whale blows No. 1560, (b) whale back No. 137, (c) slaps No. 162, and (d) breaches No. 241. Error bars denote standard deviations obtained by bootstrapping.

is an essential parameter in describing the performance of a detection algorithm. For the whale detection algorithm described and utilized in this study, the temporal distribution of false alerts exhibits a nonnormal distribution, with very few false alerts throughout most hours, but some hours with a high false alert rate of more than 1 false alert per minute. At the PO location the median false alert rate was 0 false positives per hour ( $\text{fph}^{-1}$ ). The absolute mean was  $8 \text{ fph}^{-1}$  with a maximum of  $>50 \text{ fph}^{-1}$  (Fig. 7a). At the PR location the median and mean were 13 and  $30 \text{ fph}^{-1}$ , respectively (Fig. 7a). The distribution of false positives per hour did not correlate with wind force (Fig. 7b).

#### d. Integrated system performance

##### 1) PR SITE

The relationship in detection performance between the automatic thermal (IR) system and the MMO is similar for both experienced and assisted MMOs (eMMOs and aMMOs, respectively; Figs. 8a,b). At distances  $<3 \text{ km}$ , conditional probability increases with increasing distance

up to which sightings are considered, regardless of which methods are compared. As well, the IR system outperforms the eMMOs (Fig. 8a) and aMMOs (Fig. 8b) (i.e., the probability that a cue detected by an MMO was also detected by the IR system is greater than the probability that a cue detected by the IR system was also detected by an MMO). There is a smaller difference in performance between the IR system and the MMO's at distances  $>3 \text{ km}$ ; the thermal IR system and the eMMO perform similar, while the thermal IR system outperforms the aMMO, but to a lesser degree than at distances  $<3 \text{ km}$ . The overlap in detections is greater for the thermal IR system and the eMMO than for the thermal IR system and the aMMO. In the comparison of eMMO and aMMO, the eMMO outperforms the aMMO at all distances except  $1.5 \text{ km}$  (Fig. 8c). At distances  $<3 \text{ km}$ , conditional probability generally decreases as a function of distance to sighting for both MMOs.

##### 2) CR SITE

The eMMO marginally outperforms the thermal (IR) system in detecting large baleen whales over the range



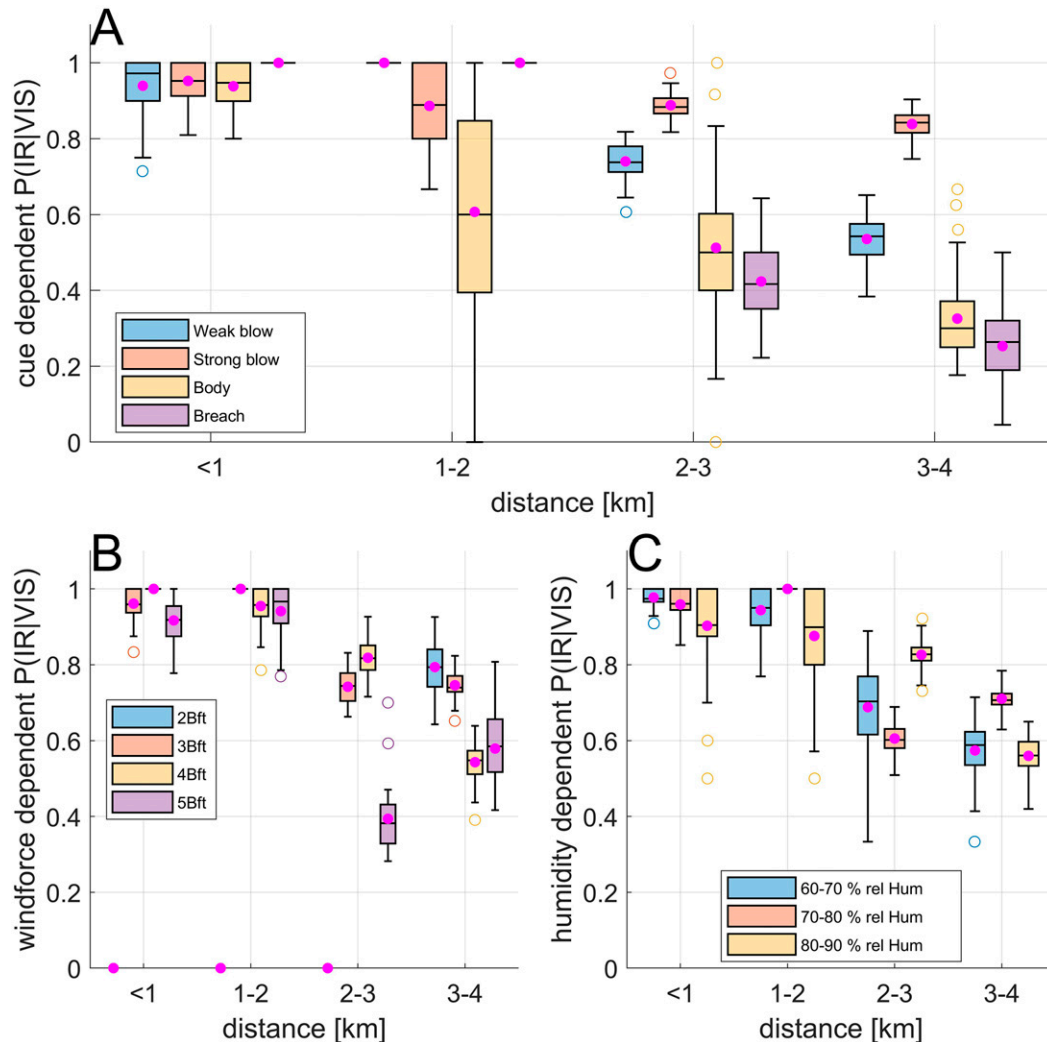


FIG. 4. (a) Probability of perception of different humpback whale cues at the CR site. Perceptibility drops significantly at 2–3 km. Note that strong whale blows have a significantly higher probability of perception than weak whale blows in the 3–4 km distance bin (0.85 vs 0.55). (b) Influence of wind force on perceptibility. Perceptibility is only weakly affected by wind force up to 3 km, but drops off quickly at 2–4 km and BFT5. (c) Relative humidity shows no obvious correlation with perceptibility. Error bars denote standard deviations obtained by bootstrapping.

of distances investigated at CR site (Fig. 8d); the performance of the two methods is nearly identical for cues sighted at 0.5 km, though the difference gradually increases with increasing distance up to which cues sighted are considered such that the eMMO performs approximately twice as well as the thermal (IR) system at 6 km. The difference in performance between the two methods increases when cues produced by other species are included in the analysis such that the eMMO performs approximately 3 times as well as the thermal (IR) system (Figs. 8e,f). The detection performance of the thermal (IR) system also increased, compared to the comparison for baleen whales without minkes, though the increase was slight. On a day-to-day basis, detection

performance was found to be quite variable (Fig. 9), though overall, the eMMO was found to outperform the thermal (IR) system.

#### 4. Discussion

Detecting whales in the open ocean is notoriously difficult, and the performance of any detection system is highly variable. For human observers, detection performance is affected by environmental conditions such as daylight, sea state, wind force, and glare, as well as observer experience level, fatigue, and whale behavior. Here we evaluated how environmental conditions affect the performance of a thermal (IR) imaging-based whale

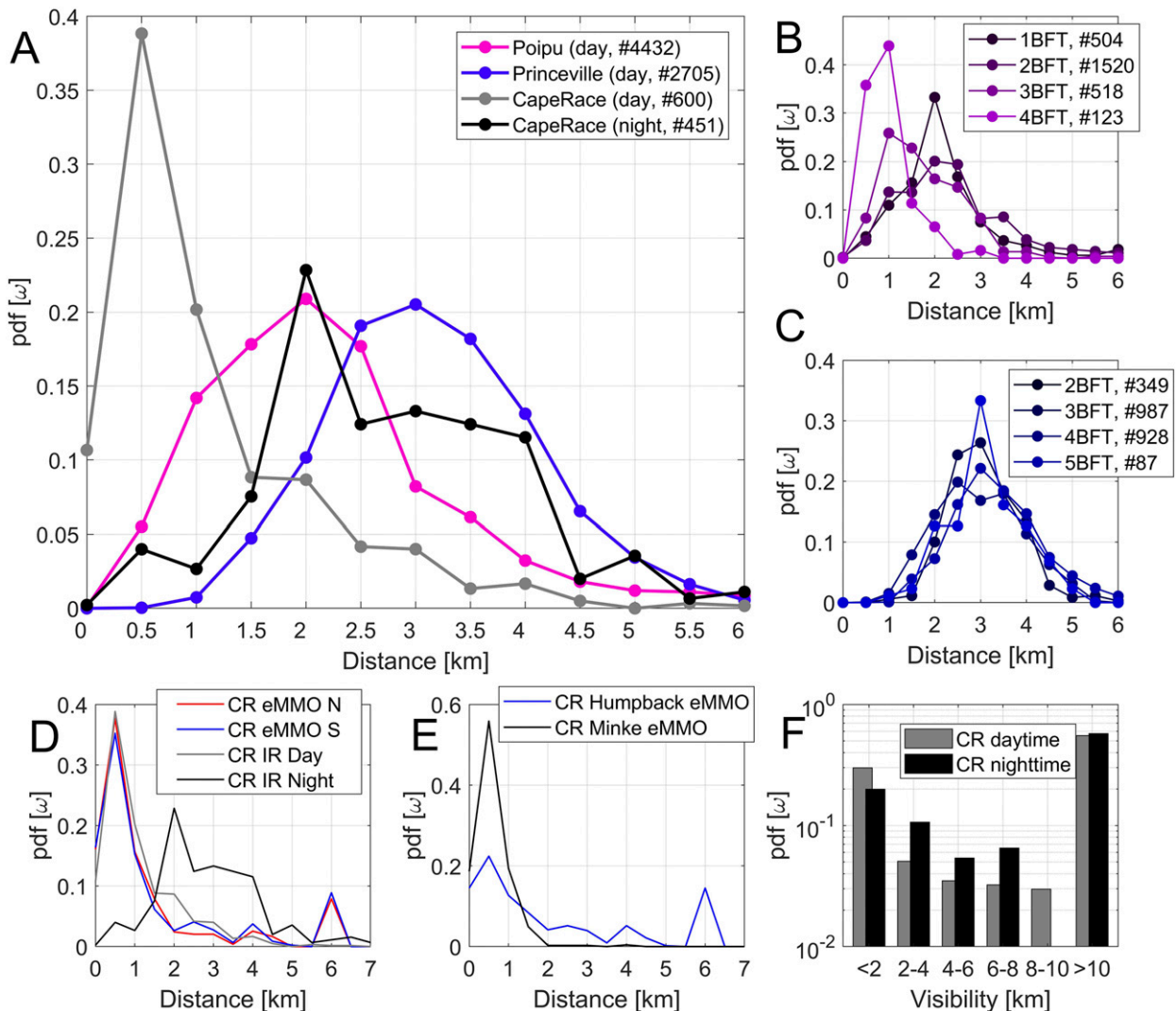


FIG. 5. Detection function for the automatic detection algorithm. (a) Detection functions at three different locations featuring different sensor heights and environmental regimes. CR data were analyzed for day and night independently. (b) Detection function for different wind forces at the PO site. (c) Detection function for different wind forces at the PR site. (d) Distance-dependent detection functions for marine mammals made by MMOs and the thermal IR system at the CR site. (e) Detection functions made by the MMOs at CR separated by species. Minke whales account for the majority of whale sightings up to 2 km, while humpback whales were sighted up to 8 km. (f) Differences in visibility between day- and nighttime at the CR field site. Visibility was generally higher during night- than daytime.

detection system and under which conditions such a system can be a suitable component of a marine mammal monitoring program.

#### a. Thermal perceptibility

The two locations where we evaluated thermal perceptibility were very different. The NSI field site in Australia is characterized by a high elevation (56 m), and relatively dry and windy conditions. In contrast, the CR field site in Canada is at a lower elevation (26 m) and is in an area characterized by high humidity and thick

fog in the summer months (Table S1). Whales were successfully perceived by the thermal (IR) system at both sites despite these differences in environment.

At NSI we find that for the most common cue, a whale's blow, there is a linear decay of perceptibility with increased distance. The perceptibility is higher than 0.8 for individual cues at distances <3 km. Only in wind forces BFT5 and higher is the perceptibility at NSI significantly reduced (Fig. 3a). At NSI, groups of whales could be detected with a chance of more than 90% within 2 km (Fig. 2). Surface behaviors involving

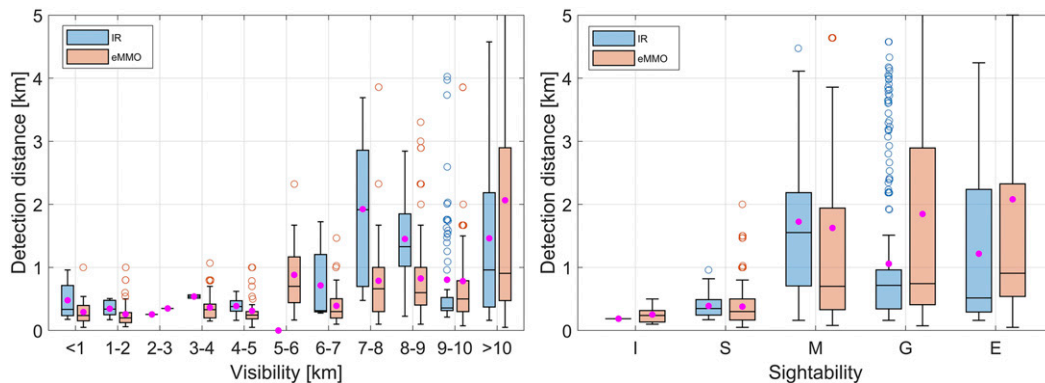


FIG. 6. Influence of (a) measured visibility conditions and (b) human estimated sightability conditions (I = impossible, S = severely impaired, M = moderately impaired, G = good, E = excellent), on the average detection range of MMOs and thermal (IR) automatic detections.

large water displacements, like breaches and slaps, were perceptible even at 8–10 km, far beyond the distances usually relevant for mitigation.

At the CR site, where the camera was mounted at a much lower height relative to NSI, the probability of perception was  $>80\%$  for strong and weak whale blows at distances  $\leq 3$  km. The lower platform height at Cape Race is more comparable to platform heights on large seismic vessels, and detections made from this height indicate the utility of the thermal (IR) system in detecting blows at distances that would allow an MMO to “track” whales in the vicinity of the vessel beyond the safety zone. Blows were subjectively classified as strong or weak by the observer in the field. Though we did not attempt to follow individual whales for the purpose of making detailed behavioral observations at Cape Race, it is likely that the strong blows were produced by whales surfacing immediately following longer dives, while weak blows were produced during shallower dives and surface activity. Because whales

would produce both weak and strong blows over time, the decrease in perceptibility of weak blows (though still  $>50\%$ , Fig. 4a) at distances greater than 3 km, is not overly concerning in a mitigation context especially because mitigation and monitoring zones generally have radii  $<3$  km (Verfuss et al. 2016). Perceptibility was shown not to be influenced by relative humidity (Fig. 4c), and was only reduced in the presence of wind force  $>\text{BFT4}$  (relative to perceptibility at wind force  $\leq \text{BFT3}$ ) at distances  $>3$  km (Fig. 4b), suggests that this technology is suitable for mitigation applications.

## b. Detectability

### 1) DETECTION FUNCTION

With increasing distance from the observer, the detection function increases continuously up to a peak. After this peak, the detection function value follows a nonlinear decay in the number of detected animals. We found detection functions peaking at 0.5 km (CR, daytime),

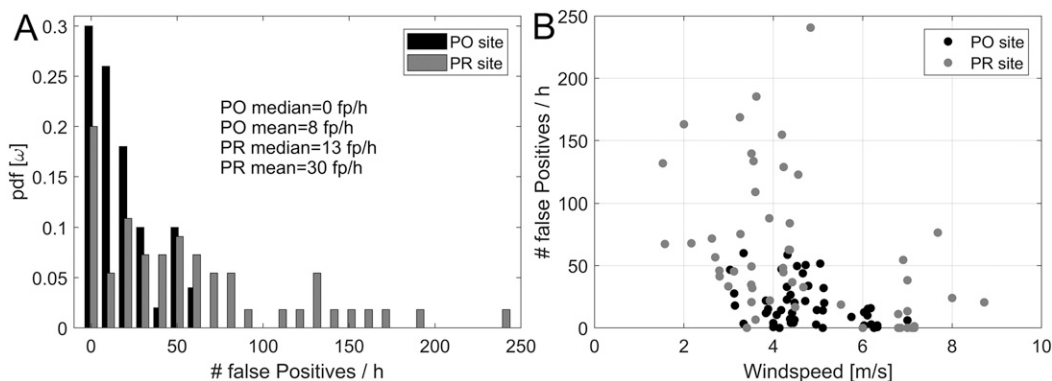


FIG. 7. (a) Distribution of false positive detections per hour for the PO and PR locations. (b) The respective correlation of false positives per hour with the wind speed.

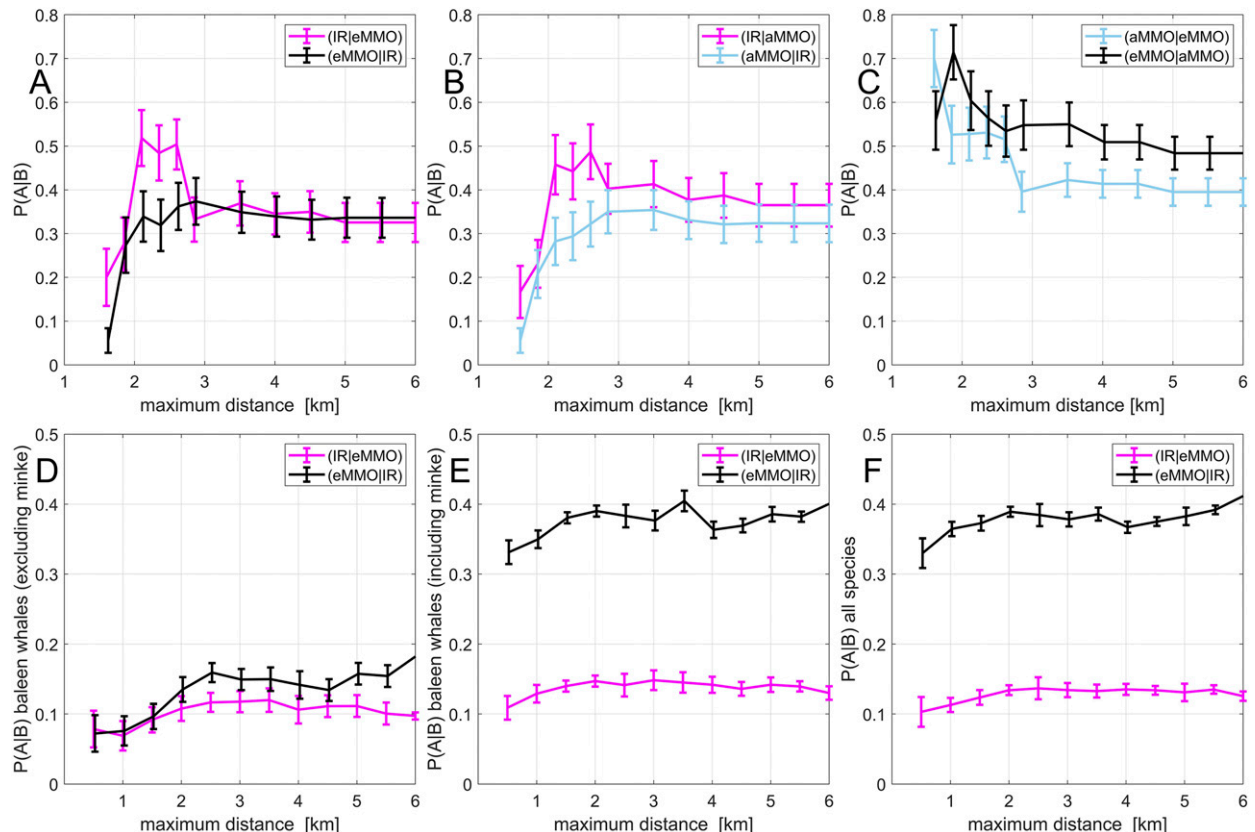


FIG. 8. Integrated system performance evaluation of the automatic whale detection system. Conditional probability of detection by (a) thermal (IR) imaging system and experienced MMO (eMMO), (b) thermal (IR) imaging system and assisted marine mammal observer (aMMO), and (c) eMMO and aMMO at PR site. Conditional probability of detection by thermal (IR) imaging system and experienced MMO (eMMO) at CR site for (d) baleen whales excluding minke whales, (e) baleen whales including minke whales, and (f) all observed species. Error bars denote standard errors across observational shifts in (a)–(c) or standard deviation obtained by bootstrapping in (d)–(f).

2 km (CR nighttime and PO) and 3 km (PR). The increase in the number of detections up to that peak distance can be explained by the linear increase in ocean surface surveyed per increase in differential unit distance (between 1–2 and 2–3 km the observed area is doubled). Therefore, more animals and thus more detections are expected farther out. The peak of the detection function marks the point where (assuming an equal distribution of animals throughout the observation area) the detection method starts to miss cues and can serve as an indicator up to which distance a method can be reliably utilized for mitigation purposes. The position of this peak depends on several different factors, including platform height and whale behavior. The large difference in the range where the thermal (IR) detection functions peaks at the CR site in 2016 (2 km at nighttime, and 0.5 km at daytime) may be attributed to humpbacks tracking capelin (*Mallotus villosus*; Whitehead and Carscadden 1985) in nearshore waters. These different peaks cannot be attributed to

different visibility conditions during the day versus night, as these were shown not to differ significantly (Fig. 5f). While there was less dense fog during the night (the probability for visibility below 1 km is lower during the night), the probability only increased for visibilities of up to 5 km, which still does not allow detections at 3 km distance (Fig. 6a). Both the MMO and the thermal (IR) detection function peak at 0.5 km during the daytime (Fig. 5d), and whales were also often observed at this distance during the perceptibility experiment in 2015 at the CR site. As whales were sighted at 2–3 km during nighttime in 2016 at CR site, it is likely that both the MMOs and the thermal (IR) detection system would have spotted whales at 2–3 km during the day, had they been present at that distance. Such a distance-dependent availability bias is not to be expected from a moving platform like a ship.

Detection functions obtained at the PO site peak at 2 km compared to 3 km at the PR site. This difference can be explained by the lower platform height of 16 m

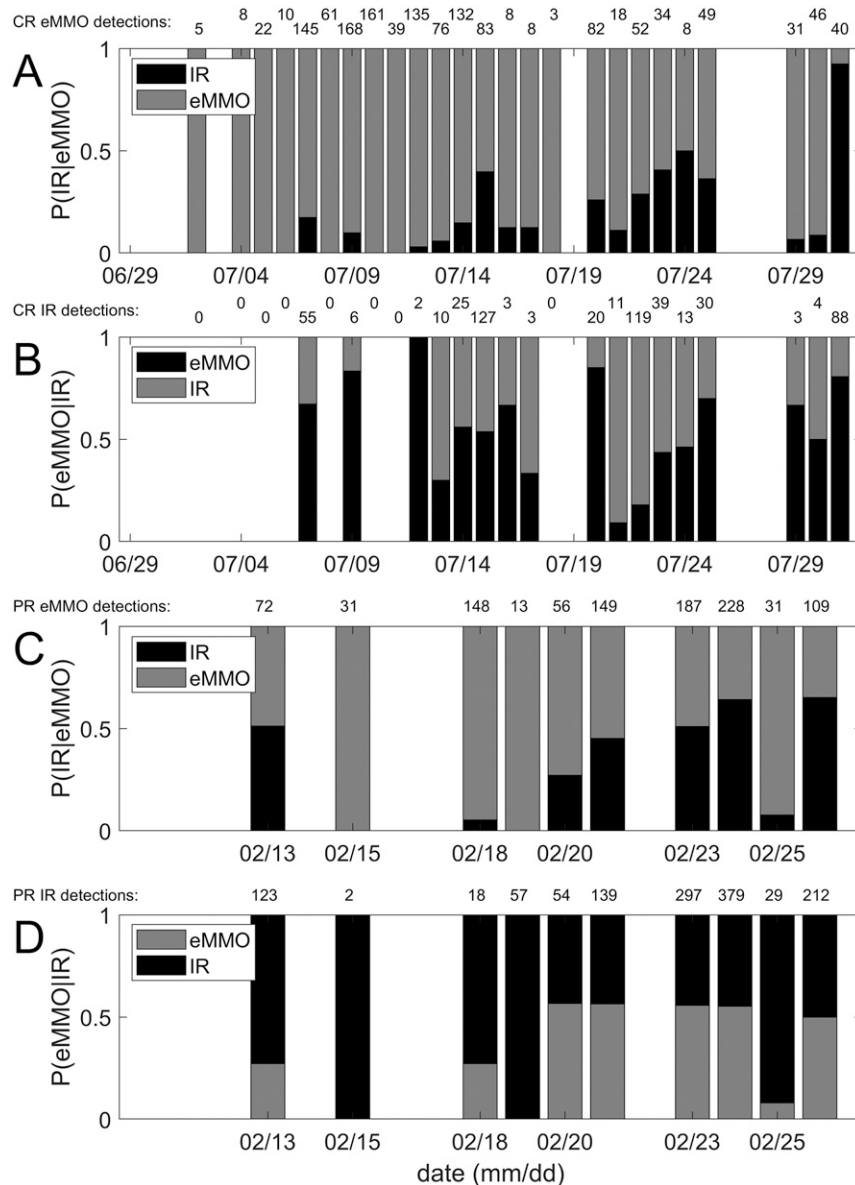


FIG. 9. Daily conditional probabilities that (a),(c) marine mammals were detected by the thermal (IR) system given that they were sighted by MMOs, and (b),(d) marine mammals were sighted by MMOs given that they were detected by the thermal (IR) system, at (a),(b) CR and (c),(d) PR site, respectively, in 2016.

compared to 49.8 m. This result indicates that even at low platform heights whales can be detected out to distances relevant for mitigation with the use of the thermal (IR) detection system.

## 2) WIND

Wind force had an effect on the detection function at the PO site, where increased wind force leads to a reduction of the range of the peak of the detection function. This can be explained with the adaptive nature of

the detection algorithm. Higher wind force levels such as BFT3 and BFT4 are associated with the formation of whitecaps. Whitecaps lead to an increased contrast across the image, thus increased noise, and reduced signal to noise ratio for the individual whale blow. The algorithm is designed this way to reduce false positives that would result from increasing wind force. We do not observe such a reduction in the location of the detection function peak at PR site. We attribute this to the different height of the sensor platform, resulting in



observation of the sea surface at a steeper angle. At steeper angles whitecaps do not resemble whale blows as much as they do from a lower height platform. Our observation is in line with the results obtained from the perceptibility experiment. Wind starts to affect perceptibility at ranges above 3 km, which are already beyond the peak of the detection function, and therefore, little influence on the peak of the detection function is expected.

### 3) FOG

It has previously been shown that fog has a significant impact on the transmission of longwave infrared (LWIR) radiation due to an increase in Mie scattering with large fog droplets (Verfuss et al. 2018). Unfortunately, the impact fog has on the ability of the system to detect whales is impossible to measure within the thermal perceptibility protocol as human observers are also strongly impacted by fog. Therefore, we conducted an analysis on how visibility (including fog) affects detection within the detectability protocol. We find that in dense fog the thermal (IR) imaging system was as impeded as a human observer (Fig. 6b), which is, on basis of the average detection distance, comparable to visibility conditions <5 km as measured by the FS11. It is to be noted that the FS11 sensor measures visibility in units standardized for use in aviation. Therefore, FS11 measurements are generally greater than human-estimated visibility values. We use the FS11 measured visibility values here for comparability with other studies. We also find that in visibility <10 km, the average detection distance is higher for the thermal (IR) system than for an MMO. We interpret these results as hazy or misty conditions, which IR radiation can penetrate better than visual spectrum light, leading to larger detection distances (Fig. 6a). This can be interpreted that in hazy or conditions the thermal (IR) detection systems would allow for greater detection ranges than an MMO, but in dense fog conditions, it is equally affected.

### 4) FALSE ALERTS

Managing the number of false alerts is crucial for the usefulness of any automatic detection system. In a mitigation setup, false negatives (missed whale cues) are not wanted, as this increases the risk to the animals. Ensuring that all cues would be detected with certainty, would imply to classify each detection as a true positive, therefore rendering the detection algorithm pointless. One needs to balance the number of false positives with the number of false negatives (prioritizing recall over precision).

Our experiments showed that the occurrences of false alerts were not normally distributed. At the PO

and PR sites, long phases of hours without a single false alert were followed by some hours where more than one false alert per minute was detected. Sometimes, when the algorithm encountered objects that were not familiar and fast moving (like a tour company vessel traveling along the Hawaiian coast) tens of alerts within a short period of time resulted until the tracking algorithm started track and remove tracked objects.

False alerts were mainly produced by three different object categories: small fishing vessels, birds, and breaking waves. Small fishing vessels often disappeared behind the swell and reappeared for a few seconds, perfectly mimicking a whale blow signature. The IR operators often had difficulties assessing if a detection was a whale blow or small fishing vessel until the fishing vessel came closer (<3–4 km). Individual birds were usually tracked by the object tracking algorithm and filtered before creating a false alert. In cases where several birds flew through the field of view at the same time, the tracking algorithm most often could not separate the tracks, and therefore, did not remove the birds from the detection process, leading to very high numbers of false alerts. At the CR field site in 2016, the majority of false alerts were caused by diving Northern gannets (*Morus bassanus*). Breaking waves also created high false alert rates, especially in the PR location where large swell is common in the winter on the Hawaiian north shore. Many of these false alert sources were due to our shore-based field site locations and we did not attempt to create an algorithm to remove these further, as they are less likely to occur on an open ocean ship-based scenario. For a ship-based setup, waves and birds (and small ice chunks in polar waters) are the most likely source of false alerts. Breaking waves caused a significant number of false alerts (Zitterbart et al. 2013; Smith et al. 2020), but only occur during higher sea states. Seabirds in the air and on the ocean's surface have been reported to be a major source of false alerts on ship-based deployments and need more attention in the automatic detection algorithm development (Smith et al. 2020). A low but constant false alert rate is desirable, as it can help to keep the observers alert.

### c. Integrated system performance

The integrated system performance analysis reveals that no detection method was capable of detecting all cues detected by other methods [i.e., no method has a conditional probability of detection  $P(A|B) = 1$ ]. It is to be noted that at the PR location, a large reef at the shore kept the water quite shallow for about 1 km and thus there are very few detections within 1 km (Fig. 5a, blue

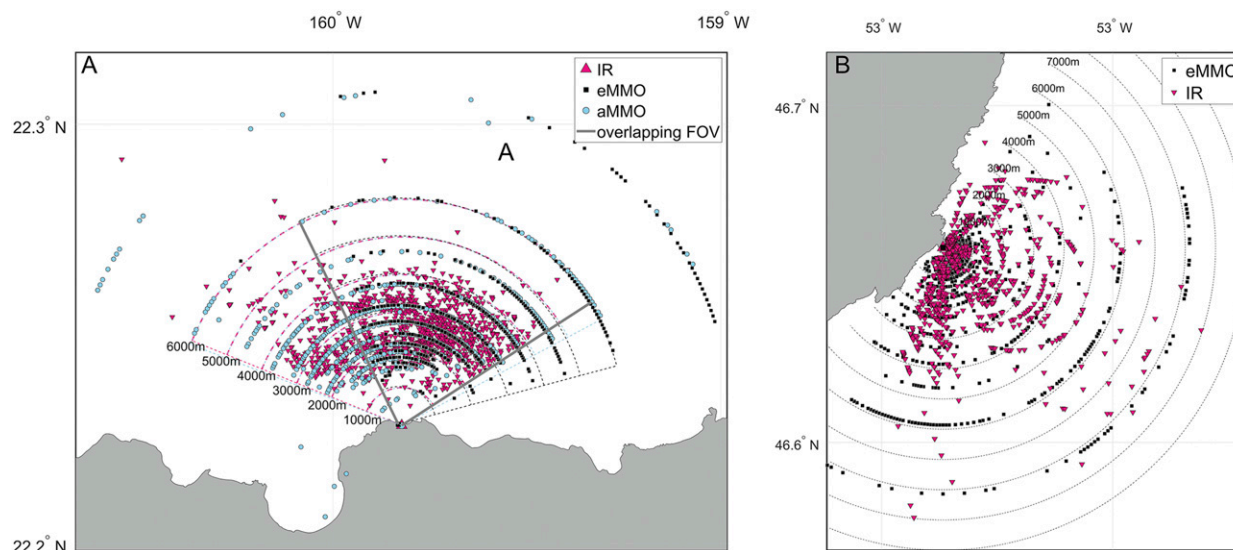


FIG. 10. (a) Absolute detection performance comparison across three methods at PR site. The gray lines denote the common field of view across all three methods. (b) Locations of all automatically detected whale cues (magenta triangles) and all visually detected whale cues by the eMMO (black squares) during Cape Race (CR) 2016 field campaign.

line). Therefore, we compare  $P(A|B)$  only from 1.5 km onward. In the analysis we decided to gradually increase the range up to which detections are considered for the estimation of  $P(A|B)$ . In a mitigation context, it is not crucial to detect whales very far off the vessel; therefore, depending on the application, different ranges are to be considered. When comparing the thermal (IR) system with the eMMO, in short range (2–3 km), the IR system outperforms the eMMO. This result is expected as the detection functions peak is just at this range, indicating the IR system should detect most of the cues. By having a larger concurrent ocean coverage than an eMMO (as it is a 360° system) one would expect it to detect more than the eMMO. The maximum  $P(\text{IR}|\text{eMMO})$  of 0.5 is therefore a rather low value and we think it is underestimated due to the quantization of the distance estimation with the binoculars (see Fig. 10). The concentric rings in the detections made by eMMO and aMMO are a result of the distance estimation using reticules in the binoculars. At 3 km distance, the distance estimation error with a binocular is approx. 8% (Zitterbart et al. 2013) and approx. 5% with the IR system, 250 and 150 m, respectively. This can lead to the localization being outside of the 500 m radius of each other, leading to an underestimation of  $P(A|B)$  of both detection methods. At larger distances than 3 km, the eMMO outperforms the IR system, which is expected as the perceptibility and detection function starts to drop at 3 km, thus the IR system inevitably will miss sighting cues.

Comparing the IR with the aMMO, it can be seen that the IR outperforms the aMMO at all distances considered. During the experiment it became clear that the aMMO is overwhelmed by the amount of information provided by both modalities, especially by false positives provided by the thermal (IR) system. While the thermal (IR) operator can deal with significant amounts of false positives, an aMMO has a reduced software interface where the aMMO could not handle significant numbers of false positives at the same time. Furthermore, the aMMO usually tried to verify a thermal (IR) automatic detection visually, thus focusing on one spot and missing whales at other spots. Furthermore, the aMMO often could not estimate or verify the distance of an IR sighting, leading to a total of 362 aMMO sightings being removed from the analysis because they could not be localized. We realized that the workflow and software interface for the aMMO has to be improved to increase the aMMO performance above the performance of the IR observer. The aMMO summarized the utility of the IR system according to periods of low and high wind force. During periods of low wind force, it allowed the observer to focus on scanning and looking far out, while the close by area was surveilled by the IR system. During periods of high wind force, the aMMOs claimed they would only watch the system and then verify, and found this helpful. A direct comparison between the aMMO and the eMMO reveals that at short distances (<2 km) the aMMO outperforms the eMMO. Between 2 and 3 km, both

systems perform equally well, and with increasing distance, the eMMO outperforms the aMMO. This can be again explained with the detection function dropping off at 3 km.

Overall, these results are very encouraging. Even at distances between 2 and 3 km around the observer, thus beyond usual mitigation zones in most countries, an eMMO will benefit from assistance by an IR system with a potential increased detection rate.

Recapture rates for thermal (IR) and eMMOs are rather low at the CR site compared to the PR site. We attribute this relatively low recapture rate to the complex detection scenario present at the CR site, where multiple and elusive species like minke whales are frequently present, causing the focus of the MMO to shift between close-by elusive animals and animals farther out. For the thermal (IR) system, the often-large numbers of diving gannets in the area, apparently feeding on the same prey as the humpbacks, often lead to the detection algorithm running into its computational limit and subsequently dropping frames, thus missing detections. We anticipate that such problems are owed to the land-based setting of our study, and are unlikely to occur on a ship-based scenario.

Comparison of performance,  $P(A|B)$ , on a daily basis at CR (Figs. 9a,b) and PR (Figs. 9d,e) shows high variability for both methods. The recapture performance of the thermal (IR) system,  $P(\text{IR}|\text{eMMO})$ , varies from 0 to 1 (Fig. 9a) over the study period, while  $P(\text{eMMO}|\text{IR})$  varies from 0.1 to 1. We interpret this variability as indicating that overall performance can be increased by employing both detection methods (thermal (IR) and eMMO) rather than a single method.

## 5. Conclusions

We have shown that a thermal (IR) imaging-based system is a useful tool for detecting whales across a wide range of environmental conditions. Whale blows, produced primarily by humpbacks, were perceptible in >70% of thermal (IR) images up to distances of 3 km, across a range of platform heights (16 to 49.8 m MSL), air and sea surface temperature combinations (12°C vs 10°C and 21°C vs 25°C; AT vs SST, respectively), and at wind force levels <BFT4. Detection functions for the thermal (IR) system generally followed the pattern expected for point transect sampling. Distance to detection peaks ranged from ca. 0.5 to 3 km. Differences in peak location were attributed to platform height (lower platforms resulted in closer detection peaks) and animal behavior. Wind force was also found to influence peak location, though this was also shown to be a function of platform height (greater wind force resulted in a closer

peak at the lowest platform height only). Dense fog was found to impede the detection performance of the thermal (IR) system and MMOs equally, while during hazy and misty conditions, the thermal (IR) can “see” farther than an MMO. False alerts were caused primarily by small vessels, birds and breaking waves; false alert rates were variable across sites, and site-specific knowledge of expected triggers needs to be used to create filters or modify the algorithm to decrease the false positive rate. We found that the thermal (IR) system and MMOs complement one another in terms of making detections: with few exceptions, a single method detected <40% of the detections made by another method. The IR system was found to outperform MMOs in Hawaii, whereas the MMOs outperformed the IR system in Newfoundland. This difference in comparative performance may be attributed to the wider range of species detected in Newfoundland and the vast number of diving gannets that caused significant false positives, limiting the number of detections that could be processed. These results indicate that thermal imaging systems can be a valuable addition to marine mammal monitoring programs including during seismic surveys and for the mitigation of ship strike of large whales. While our results were obtained on land at locations where whale sightings are very common, thermal imaging-based whale detection systems can be used from vessels as previously shown (Zitterbart et al. 2013) and the results published here should be transferable to ship-based installations. Prerequisites are 1) that the thermal imaging camera is sufficiently stabilized (Smith et al. 2020) against pitch and roll, and 2) that the detection algorithm operates on time scales that are small relative to the changes in the background image (as caused by the ship’s steaming). Both conditions are met for the thermal imager employed in this study.

**Acknowledgments.** This work was funded by the Office of Naval Research (ONR) under Award N000141310856, by the Environmental Studies Research Fund (ESRF; [esrfunds.org](http://esrfunds.org)) under Award 2014-03S and by the Alfred-Wegener-Institute Helmholtz Zentrum für Polar- und Meeresforschung. DPZ and OB declare competing financial interests: 1) Patent US8941728B2, DE102011114084B4: A method for automatic real-time marine mammal detection. The patent describes the ideas basic to the automatic whale detection software as used to acquire and process the data presented in this paper. 2) Licensing of the Tashtego automatic whale detection software to the manufacturer of IR sensor. The authors confirm that these competing financial interests did not alter

their adherence good scientific practice. We thank P. Abgrall, J. Coffey, K. Keats, B. Mactavish, V. Moulton, and S. Penney-Belbin for data collection or IR image review. We thank S. Besaw, J. Christian, A. Coombs, P. Coombs, W. Costello, T. Elliott, E. Evans, I. Goudie, C. Jones, K. Knowles, R. Martin, A. Murphy, D. and J. Shepherd; and the staffs at the Irish Loop Express, the Myrick Wireless Interpretive Centre, the Mistaken Point Ecological Reserve, and the lighthouse keepers for logistical assistance at our remote field site. We thank D. Boutilier and B. McDonald (DFO) for assisting us in obtaining license to occupy permits for Cape Race. We thank D. Taylor (ESRF Research Manager) for his support.

**Authors' contribution:** DPZ, HS, OB conceived the study. DPZ, HS, EB, MN, AP, OB organized the field trips. DPZ, MF, SR wrote the thermal imaging detection software. DPZ analyzed the data. DPZ, HS, MF, EB, JB, LB, AC, AD, MH, CL, HM, KO, AP collected the data and conducted retrospective IR data review. DPZ and HS wrote the manuscript, all coauthors edited and proofed the manuscript.

## REFERENCES

- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas, 2001: *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press, 448 pp.
- Cates, K., D. P. DeMaster, R. L. Brownell Jr., G. Silber, S. Gende, R. Leaper, F. Ritter, and S. Panigada, 2017: Strategic plan to mitigate the impacts of ship strikes on cetacean populations: 2017-2020. International Whaling Commission, <https://iwc.int/ship-strikes>.
- Constantine, R., M. Johnson, L. Riekkola, S. Jervis, L. Kozmian-Ledward, T. Dennis, L. G. Torres, and N. Aguilar de Soto, 2015: Mitigation of vessel-strike mortality of endangered Bryde's whales in the Hauraki Gulf, New Zealand. *Biol. Conserv.*, **186**, 149–157, <https://doi.org/10.1016/j.biocon.2015.03.008>.
- D'Amico, A., R. C. Gisiner, D. R. Ketten, J. A. Hammock, C. Johnson, P. L. Tyack, and J. Mead, 2009: Beaked whale strandings and naval exercises. *Aquat. Mamm.*, **35**, 452–472, <https://doi.org/10.1578/AM.35.4.2009.452>.
- Dawson, J., L. L. Pizzolato, S. E. L. Howell, L. Copland, and M. E. Johnston, 2018: Temporal and spatial patterns of ship traffic in the Canadian Arctic from 1990 to 2015. *Arctic*, **71**, 15–26, <https://doi.org/10.14430/arctic4698>.
- Erbe, C., R. Dunlop, and S. Dolman, 2018: Effects of noise on marine mammals. *Effects of Anthropogenic Noise on Animals*, H. Slabbekoorn et al., Eds., Springer Handbook of Auditory Research, Vol. 66, Springer Science and Business Media, 277–309.
- Fais, A., T. P. Lewis, D. P. Zitterbart, O. Álvarez, A. Tejedor, and N. Aguilar Soto, 2016: Abundance and distribution of sperm whales in the Canary Islands: Can sperm whales in the archipelago sustain the current level of ship-strike mortalities? *PLOS ONE*, **11**, e0150660, <https://doi.org/10.1371/journal.pone.0150660>.
- Frisk, G. V., 2012: Noiseconomics: The relationship between ambient noise levels in the sea and global economic trends. *Sci. Rep.*, **2**, 437, <https://doi.org/10.1038/srep00437>.
- Horton, T. W., A. Oline, N. Hauser, T. M. Khan, A. Laute, A. Stoller, K. Tison, and P. Zawar-Reza, 2017: Thermal imaging and biometrical thermography of humpback whales. *Front. Mar. Sci.*, **4**, 424, <https://doi.org/10.3389/fmars.2017.00424>.
- Laist, D. W., A. R. Knowlton, and D. Pendleton, 2014: Effectiveness of mandatory vessel speed limits for protecting North Atlantic right whales. *Endanger. Species Res.*, **23**, 133–147, <https://doi.org/10.3354/esr00586>.
- Miller, P. J. O., and Coauthors, 2012: The severity of behavioral changes observed during experimental exposures of killer (Orcinus orca), long-finned pilot (Globicephala melas), and sperm (Physeter macrocephalus) whales to naval sonar. *Aquat. Mamm.*, **38**, 362–401, <https://doi.org/10.1578/AM.38.4.2012.362>.
- Noad, M. J., E. Kniest, and R. A. Dunlop, 2019: Boom to bust? Implication for the continued rapid growth of the eastern Australian humpback whale population despite recovery. *Popul. Ecol.*, **61**, 198–209, <https://doi.org/10.1002/1438-390X.1014>.
- Perryman, W. L., M. A. Donahue, J. L. Laake, and T. E. Martin, 1999: Diel variation in migration rates of eastern Pacific gray whales measured with thermal imaging sensors. *Mar. Mamm. Sci.*, **15**, 426–445, <https://doi.org/10.1111/j.1748-7692.1999.tb00811.x>.
- Richardson, W. J., C. R. Greene, C. I. Malme, and D. H. Thomson, 1995: *Marine Mammals and Noise*. Academic Press, 576 pp.
- Rowley, H. A., S. Baluja, and T. Kanade, 1995: Human face detection in visual scenes. Carnegie Mellon University Tech. Rep. CMU-CS-95-158R, 26 pp.
- Santhaseelan, V., S. Arigela, and V. K. Asari, 2012: Neural network based methodology for automatic detection of whale blows in infrared video. *Advances in Visual Computing*, Springer, 230–240.
- Schoonmaker, J., J. Dirbas, Y. Podobna, T. Wells, C. Boucher, and D. Oakley, 2008: Multispectral observations of marine mammals. *Proc. SPIE*, **7113**, 711311, <https://doi.org/10.1117/12.800024>.
- Smith, H. R., D. P. Zitterbart, M. Flau, T. F. Norris, E. L. Ferguson, C. Jones, O. Boebel, and V. D. Moulton, 2020: A field comparison of marine mammal detections via visual, acoustic, and infrared (IR) imaging methods offshore Atlantic Canada. *Mar. Pollut. Bull.*, **154**, 111026, <https://doi.org/10.1016/j.marpolbul.2020.111026>.
- Southall, B. L., and Coauthors, 2019: Marine mammal noise exposure criteria: Updated scientific recommendations for residual hearing effects. *Aquat. Mamm.*, **45**, 125–232, <https://doi.org/10.1578/AM.45.2.2019.125>.
- Thomas, L., S. T. Buckland, K. P. Burnham, D. R. Anderson, J. L. Laake, D. L. Borchers, and S. Strindberg, 2002: Distance sampling. *Encyclopedia of Environmetrics*, A. H. El-Shaarawi and W. W. Piegorsch, Eds., Vol. 1, John Wiley and Sons, 544–552.
- UNCTAD, 2018: Review of maritime transport. United Nations Tech. Rep., 116 pp., [https://unctad.org/en/PublicationsLibrary/rmt2018\\_en.pdf](https://unctad.org/en/PublicationsLibrary/rmt2018_en.pdf).
- Verfuss, U. K., and Coauthors, 2016: Low visibility real-time marine mammal monitoring techniques review. OGP Rep. SMRUOGP2015-002, 224 pp.

- , and Coauthors, 2018: Comparing methods suitable for monitoring marine mammals in low visibility conditions during seismic surveys. *Mar. Pollut. Bull.*, **126**, 1–18, <https://doi.org/10.1016/j.marpolbul.2017.10.034>.
- Weir, C. R., and S. J. Dolman, 2007: Comparative review of the regional marine mammal mitigation guidelines implemented during industrial seismic surveys, and guidance towards a worldwide standard. *J. Int. Wildl. Law Policy*, **10**, 1–27, <https://doi.org/10.1080/13880290701229838>.
- Whitehead, H., and J. E. Carscadden, 1985: Predicting inshore whale abundance—Whales and capelin off the Newfoundland coast. *Can. J. Fish. Aquat. Sci.*, **42**, 976–981, <https://doi.org/10.1139/f85-122>.
- Zitterbart, D. P., L. Kindermann, E. Burkhardt, and O. Boebel, 2013: Automatic round-the-clock detection of whales for mitigation from underwater noise impacts. *PLOS ONE*, **8**, e71217, <https://doi.org/10.1371/journal.pone.0071217>.