Shark detection probability from aerial drone surveys within a temperate estuary

Martin T. Benavides, F. Joel Fodrie, and David W. Johnston

Abstract: Drones are easy to operate over metres-to-kilometre scales, making them potentially useful to monitor species distributions and habitat use in shallow estuaries with widely varying environmental conditions. To investigate the utility of drones for surveying bonnethead sharks (*Sphyrna tiburo*) across estuarine environmental gradients, we deployed decoys, fashioned to mimic sharks, in the field. Decoys were placed in two flight areas (0.8 km^2 each) in shallow (<2 m) water near Beaufort, N.C., on five days during 2015–2016. Survey flights were conducted using a fixed-wing drone (senseFly eBee) equipped with a digital camera. Images were indexed for combinations of six environmental factors across flights. Images representative of all ($N = 36$) observed environmental combinations were sent to a group of 15 scientists who were asked to identify sharks in each image. Non-parametric rank-sum comparisons and regression tree analysis on resultant detection probabilities highlighted depth as having the largest, statistically reliable influence on detection probabilities, with decreasing detection probabilities at increased depth. Detection probabilities were higher during midday flights, with notable effects of wind speed and cloud presence also apparent. Our study highlights depth as a first-order factor constraining the temperate estuarine habitats over which drones may reliably quantify sharks (i.e., <0.75 m).

Key words: unoccupied aircraft systems, UAS, visibility bias, bonnethead shark, *Sphyrna tiburo*.

Résumé : Les drones sont faciles à utiliser sur des échelles de mètres à kilomètres, ce qui fait qu'ils peuvent potentiellement servir pour surveiller la répartition des espèces et l'utilisation de l'habitat dans les estuaires peu profonds où les conditions environnementales varient considérablement. Dans le but d'étudier l'utilité des drones pour la surveillance de requins marteau tiburo (*Sphyrna tiburo*) dans divers gradients environnementaux estuariens, nous avons déployé des leurres, conçus pour imiter les requins, dans l'eau. Des leurres ont été placés dans deux zones de vol (0,8 km^2 chacune) en eau peu profonde (2 m) près de Beaufort, en Caroline du Nord, pendant cinq jours en 2015–2016. Les vols pour les levés ont été effectués à l'aide d'un drone à voilure fixe (senseFly eBee) équipé d'un appareil photo numérique. Les images ont été indexées pour des combinaisons de six facteurs environnementaux d'un vol à l'autre. Des images représentatives de toutes les combinaisons environnementales observées ($N = 36$) ont été envoyées à un groupe de 15 scientifiques à qui on a demandé d'identifier les requins dans chaque image. Les comparaisons de la somme des rangs non paramétriques et l'analyse de l'arbre de régression sur les probabilités de détection résultantes ont mis en évidence la profondeur comme ayant la plus grande incidence statistiquement fiable

Received 8 January 2019. Accepted 29 November 2019.

M.T. Benavides and F.J. Fodrie. Institute of Marine Sciences, University of North Carolina at Chapel Hill, Morehead City, NC 28557, USA.
D.W. Johnston.* Division of Marine Science & Conservation, Nicholas School of the Environment, Duke University Marine Lab, Beaufort, NC 28516, USA.
Corresponding author: Martin T. Benavides (e-mail: benavides@unc.edu).
*David Johnston currently serves as an Associate Editor; peer review and editorial decisions regarding this manuscript were handled by David Bird.

sur les probabilités de détection, avec des probabilités de détection décroissantes à une profondeur accrue. Les probabilités de détection étaient plus élevées pendant les vols de mi-journée, avec des effets notablesde la vitesse du vent et de la présence de nuages également apparents. Notre étude met en évidence la profondeur comme un facteur de premier ordre limitant les habitats estuariens tempérés sur lesquels les drones peuvent quantifier les requins de façon fiable (c.-à-d. < 0,75 m). [Traduit par la Rédaction]

Mots-clés : systèmes d'aéronefs sans pilote (UAS), biais de visibilité, requin marteau tiburo, *Sphyrna tiburo*.

Introduction

Distribution and abundance estimates of sharks have typically been obtained from capture methods, such as netting or hook-and-line, often in combination with tagging studies, which together have guided our understanding of shark population dynamics and movement patterns (Kohler and Turner 2001). While valuable, there are challenges to interpretation of data gathered by these methods related to the relatively low density and high patchiness of sharks compared to other taxa, and the need to sample over relatively large areas to reduce uncertainty with respect to shark numbers (Peterson et al. 2017). Additionally, capture methods may be inappropriately invasive in some situations for sampling sharks (e.g., mortality of endangered species), which has inspired the use of less-invasive methods, such as photo identification (Bansemer and Bennett 2008). Aerial visual surveys have also been employed over large spatial scales for estimating shark distribution and abundance (Rowat et al. 2009).

Visual surveys via manned aircraft have also been utilized extensively to study other large marine animals. In the case of marine mammals and seabirds, aerial visual surveys, along with shipboard surveys, are perhaps the most widely used means of obtaining information on distribution and abundance globally (Buckland et al. 2001; Kaschner et al. 2012). With recent technological advances, the use of digital imagery has become competitive with visual methods in manned aerial surveys for these animals, resulting in similar to substantially larger estimates of abundance (Buckland et al. 2012; Koski et al. 2013). Importantly, the use of manned aircraft has a number of logistical and scientific drawbacks, such as prohibitive cost, disturbances to wildlife, and the difficulty of covering smaller survey areas (Christie et al. 2016).

Recently, there have been considerable advances in the use of unoccupied aircraft systems (UASs), creating an attractive platform for both terrestrial and marine ecological surveys (Anderson and Gaston 2013). These UASs are advantageous with respect to aerial manned visual surveys due to the remotely controlled, smaller, and quieter aircraft, as well as the digital imagery component, which could potentially lead to more reliable, reviewable estimates. Marine mammal surveys, which have traditionally been carried out via manned aircraft for many species, have been conducted with UASs for several species, such as dugongs, seals, and sea lions (Jones et al. 2006; Hodgson et al. 2013; Sweeney et al. 2015). UASs are also used in a broad range of ecological studies on marine mammals from estimating size or body condition of individuals to collecting exhaled breath condensate for DNA and hormonal analyses (reviewed in Johnston 2019). Surveys for seabirds and sea turtles appear to benefit from the use of drones, particularly with respect to time and (or) costs when compared to ground- or water-based counts (McClellan et al. 2014; Rees et al. 2018). Finally, Kiszka et al. (2016) examined shark and ray densities by drone surveys in shallow-water reef systems off Moorea, French Polynesia, demonstrating the potential value of this

approach to survey for sharks and showing how the technology was not limited to only those species that are required to surface for respiration.

The bonnethead shark, *Sphyrna tiburo*, is a small coastal shark species often found in estuaries, shallow bays, and channels, where pupping females are most common (Compagno et al. 2005). Bonnetheads are also commonly found in high densities, with multiple individuals encountered within an area of 50 m^2 (Myrberg and Gruber 1974). The resulting patchiness in their distribution makes UASs an attractive survey platform. Drones are easy to operate over mesoscale ranges (<3 km) and at low altitude (<100 m), making them potentially useful to monitor bonnethead distributions and habitat use in shallow-water estuarine habitats. Due to the widely varying environmental conditions found in temperate estuaries, determining the effects of particular environmental variables on detection rates of sharks from drone surveys is important for understanding the efficacy of this approach in estimating patterns of distribution and abundance. There is also mounting interest in utilizing drones in nearshore waters for public safety to help minimize interactions between larger sharks and humans, and thus understanding potential limitations of this approach in different environmental contexts also has very practical applications (Colefax et al. 2019).

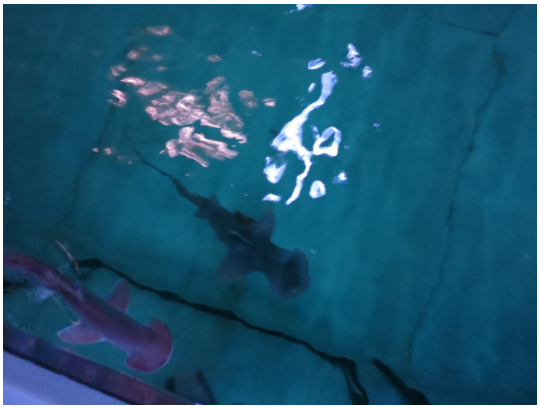
Visibility bias, which results from observers missing animals, has been a fundamental problem in the use of observer-based surveys, particularly in aerial surveys (Caughley 1974). The missing animals are either potentially visible to observers, but not seen (perception bias) or are concealed, often by turbid water (availability bias), although these two biases are difficult to separate in practice (Marsh and Sinclair 1989; Pollock et al. 2006; Barlow 2015). In mid-Atlantic estuaries, turbidity extremes due to frequent resuspension of sediment and plankton by wind and tides have obvious, large effects on light penetration throughout the water column (Kirby-Smith and Costlow 1989). We designed a series of field experiments using shark decoys photographed from overhead by drones to test effects of environmental parameters on visibility bias. Given the aforementioned effect of turbidity on potential visibility bias, even in shallow water columns, we hypothesized that the interaction between turbidity and decoy depth would have the greatest effect on detection probability.

Materials and methods

Shark decoys

To investigate the utility of UASs in surveying bonnethead sharks, we deployed decoys that were fashioned to have the appearance of bonnetheads from overhead. The decoys ($n = 9$) were cut from plywood using the outline from a ~1 m bonnethead shark that did not survive the transition to captive display at the North Carolina Aquarium at Pine Knoll Shores (NC Aquarium). This particular specimen was a gravid female and thus representative of the size range of bonnetheads typically found within the Newport River Estuary, N.C. The plywood decoys were epoxied (Nos. 105 & 207, West System, Mich.) to resist water damage. Decoys were then sanded and spray-painted to mimic the shark's countershading pattern from above using a combination of colors (Nos. 86014, 68181, 84230, and 63000, Valspar, Minn.) To confirm that the decoys had the appearance of bonnetheads, one was placed in a holding tank at the NC Aquarium with a live bonnethead while photos were taken from overhead with Cannon Powershot S110 digital cameras used during drone surveys (Fig. 1). Finally, one decoy was made to have the shape of a more generic shark species, an Atlantic sharpnose, *Rhizoprionodon terraenovae*. This was accomplished by simply trimming the "rostrum" of the bonnethead-shaped decoy to produce a conical snout (i.e., without cephalophoil), thereby allowing us to assess the potential of identifying decoys as bonnetheads or non-bonnetheads. Decoys were positively buoyant, and had to be

Fig. 1. Photograph of live bonnethead (bottom left corner) and bonnethead decoy (slightly off-center) taken at the North Carolina Aquarium at Pine Knoll Shores.



anchored during deployment by 20 cm lines at the head and caudal region, which were connected to standard bricks that rested on the seabed.

Drone flights

Bonnethead sharks are commonly found within the Newport River Estuary, a shallow water body (<3 m average depth), so decoys were placed in two flight areas in shallow waters surrounding Pivers Island in Beaufort, N.C., on five separate days during the fall of 2015 as well as in the spring and fall of 2016 (Fig. 2). Selected quadrats (0.0001° latitude × 0.0001° longitude, approximately 10 m × 10 m) within our flight areas targeted a depth range of 0–2 m. Within the flight areas, we haphazardly positioned decoys across the available range of depths. Depth measurements (to the nearest 0.1 m) were taken using transect tape for each decoy that was deployed (eight or nine decoys per flight day) at the time of deployment. GPS coordinates (decimal degrees) were also recorded for each decoy.

Each day, environmental variables, cloud cover, and secchi disk depth were recorded. Cloud cover was recorded as a categorical variable, with either not cloudy (no clouds visible overhead on the days we conducted the surveys) or cloudy (cloud cover >37% overhead on the days we conducted surveys, based on National Oceanic and Atmospheric Administration (NOAA) definition of partly cloudy), using the Weather Underground Forecast Android phone application, which used data from the Dakota station (KNCBEAUF23), approximately 2 km from our flight areas (Weather Underground 2011). Secchi depth was measured once within the flight area immediately before or after deploying decoys, using a 20 cm secchi disk, which was lowered by a string with marks every 0.1 m into the water until the disk was no longer visible, at which point the depth measurement was recorded. Mean wind speed was also recorded for each flight using Weather Underground data (Weather Underground 2011). Each day, to the extent possible, we scheduled three flights (one per spectral filter, see following paragraph) at low-, mid-, and high-tides to make full use of the local tidal amplitude (~1 m) and expand our depth interval coverage (Table 1). Depth measurements for decoys in subsequent flights were recorded as the sum of the original depth measurement and change in tidal height, estimated from NOAA water level data for the Beaufort, Duke Marine Lab, N.C., station (8656483), <0.5 km from our flight areas (NOAA 2018).

Fig. 2. Map of study area in Eastern North Carolina with flight areas highlighted in yellow. Map created using QGIS (QGIS Development Team 2018). Map data: © 2018 Google; © OpenStreetMap contributors.

Table 1. Summary of environmental conditions and flight times for each flight date.

Flight date	Cloud cover	Wind speed (m s^{-1})	Secchi depth (m)	Approximate local flight times
22 October 2015	N	2–5	1.24	0930, 1230, 1600 EDT
11 March 2016	Y	4–9	1.6	1000, 1200 EST
16 May 2016	N	2–5	1.15	1040 EDT
29 September 2016	Y	3–5	0.7	0830, 1130 EDT
27 October 2016	Y	4	0.97	1300 EDT

Note: Times are either reported in Eastern Daylight Time (EDT) or Eastern Standard Time (EST).

A total of 30 UAS flights were conducted using a fixed-wing drone (eBee, senseFly, Switzerland), equipped with either a Cannon IXUS 127 HS or Cannon Powershot S110 digital camera with one of three spectral filters: regular (RGB), red edge (RE), and near-infrared (NIR). Flight missions were designed and automated using the flight management software included with the eBee (eMotion, senseFly, Switzerland). Each flight area (approximately 0.8 km^2) was divided into eBee overpass transects that were 400 m long and 25 m apart. Flight altitude was 60 m, flight speed was 13 m s^{-1}, and flights lasted about 15 min. eBee cameras captured a downward-facing image roughly every 4 s along each transect with an

on-the-ground resolution of <2.7 cm/pixel. Individual footprint area for digital images ranged from 4200 to 8478 m^2.

Image assessments

Images were indexed for factor levels using four continuous variables: time of day (<1030, 1030–1330), mean wind speed (<4 m s^{-1}, 4–8 m s^{-1}, >8 m s^{-1}), secchi depth (<1 m, 1–1.5 m, >1.5 m), and decoy depth (<0.6 m, 0.6–1 m, >1 m). Images were also indexed by two categorical variables: filter (RGB, RE, NIR) and cloud presence (cloudy, not cloudy). For continuous variables, values were discretized into two- or three-level classifications, based on natural breaks in the data, which was done because the limited range of values observed during the 30 UAS flights did not allow for full exploration of these variables. This index was used to construct a matrix of 43 photographs, containing 0–9 shark decoys. Across the 43 photos, there were 144 bonnethead decoys and 15 non-bonnethead decoys, representative of the full spectrum of combinations of factor levels (36 unique combinations, largely driven by multiple depths within any single photo) present within the days of sampling, with at least two replicates for each level of each factor. The matrix was then utilized to construct a PDF file containing the 43 images for distribution to be scored. Images were sent out to a group of fisheries and estuarine scientists ($n = 15$) who volunteered to score each photo for presence of sharks. Without being provided any prior information regarding the number or identity of decoys that were deployed in the field of view of each image, each scorer was asked to place symbols directly on top of where they thought sharks were in each image, with separate symbols denoting bonnethead or Atlantic sharpnose sharks. Scorers were also given the option to place a mark in a box denoting no sharks were present in the image. To standardize scoring efforts, a quadrant grid denoting maximum zoom frame as well as a 5 min time limit per photograph were specified.

We used a hierarchical coding system to evaluate the series of possible outcomes for each decoy and (or) image after scoring. For images that contained decoys, each decoy was assigned a code of 0 if not detected, or a code of 1 if detected (symbol correctly placed). For a symbol to be considered correctly placed it could not be more than one body length away from the decoy (per instructions to scorers). Any symbol placed at a greater distance than 1 m from any decoy was considered a false detection. For decoys that were detected, a second layer of coding was applied to indicate if the species identification was correct (0 — incorrect, 1 — correct). Finally, images that contained no decoys were only evaluated for the number of false detections in each image.

Data analysis

Detection probability (number of times detected/number of scorers) was calculated for each decoy, a metric for the probability of detection by the "average" observer, which was the value that all subsequent tests were applied to, except in the case of false detections. To examine the range of detection probabilities, mean detection probability and standard error was computed across all decoys (across factor-levels) using the R package *psych* (Revelle 2017). The effects of five parameters on detection probability were further explored via the Mann–Whitney U test (two-level) or Kruskal–Wallis H test (three-level) among factor-level groupings: time of day, filter type, cloud presence, wind, and decoy depth, using the R package *coin* (Hothorn et al. 2008). Because wind and tide conditions changed across survey flights on each flight day, which would affect turbidity, and we failed to sample this variable frequently enough, we decided to exclude secchi depth from our analyses. These non-parametric rank-sum tests were utilized because detection probabilities could not be assumed to be normally distributed within groupings. We considered p values, patterns of

detection probability, and variances to evaluate strength of evidence for environmental conditions on detection probability (sensu Murtaugh 2014).

We used regression tree analysis (in R package *rpart*, Therneau et al. 2015) to rank the relative importance of environmental factors in explaining the variance in detection probabilities. In addition to their flexibility (i.e., non-parametric), these models have strengths in their robustness as well as their relative ease of use and interpretation, complementing traditional statistical techniques (De'Ath and Fabricius 2000). We considered five factors and chose continuous input for numerical variables (time of day, depth, and wind speed) as this provided more informative (i.e., variance reducing) splits of detection probabilities, along tree branches. We pruned the tree using the 1–SE rule (as in Breiman et al. 1984).

To determine if detected sharks could be reliably identified as bonnethead or non-bonnethead, misidentification rates (number of times incorrectly identified/number of scorers who detected decoy) were calculated for all decoys detected by at least one scorer. Misidentification rates were segregated by species (bonnethead or non-bonnethead) to determine if misidentified decoys would lead to "class 1" or "class 2" misidentification. In this context, "class 1" would be the misidentification of a bonnethead as a non-bonnethead (Atlantic sharpnose), which would lead to a bias of underestimation of bonnethead abundance; whereas "class 2" would be the misidentification of a non-bonnethead as a bonnethead and lead to a bias of overestimation of bonnethead abundance. These rates were then aggregated by factor-level to look at effects of environmental parameters on misidentification. These groupings were also compared using non-parametric rank sum tests.

False detections were summed across scorers for each image, aggregated by factor-level and compared using non-parametric rank sum tests to examine possible environmental effects on perceiving sharks when they were not actually present, excluding decoy depth as we had no way to determine at what depth a falsely identified decoy was perceived. All statistical analyses and plotting of data were conducted in R (R Core Team 2016), using the following packages: *diplyr* (Wickham and Francois 2016), *tidyr* (Wickham 2017), *ggplot2* (Wickham 2009), and *rpart.plot* (Milborrow 2017).

Results

Detection probability for all 159 individual decoys ranged from 0 (never detected) to 1 (always detected), with an overall mean value of 0.27 ± 0.03 (mean and standard error). Mean detection probabilities for environmental factor combinations ranged from 0 to 0.96 (Table 2). For the 73 decoys that were detected by at least one observer, individual misidentification rates also ranged from 0 (correctly identified by all scorers who detected) to 1 (misidentified by all scorers who detected), with an overall mean value of 0.24 ± 0.03 SE. Mean false detections for individual images ranged from 0 to 0.4, with an overall mean value of 0.04 ± 0.01 SE across 15 inspections of each photo.

Mean detection probability was negatively related to decoy depth ($\chi^2 = 49.61$, df = 2, $p < 0.001$), from 0.55 ± 0.05 SE at depths <0.6 m to 0.03 ± 0.02 SE at depths >1 m (Fig. 3). Mean detection probability increased from 0.14 ± 0.04 SE in the early morning period (before 1030) to 0.38 ± 0.04 SE in the midday period (1030–1330; $Z = -4.34$, $p < 0.001$) (Fig. 3). Overall mean detection probability was higher on not cloudy days, 0.40 ± 0.08 SE compared to 0.26 ± 0.04 SE on cloudy days, although not well supported statistically ($Z = 1.5$, $p = 0.134$) (Fig. 3). Conversely, mean detection probability trended lower with increasing mean wind speed, from 0.4 ± 0.08 SE at winds below 4 m s^{-1} to 0.14 ± 0.06 SE at winds above 8 m s^{-1}, although due to the high overall variability in the data, we failed to detect a statistically

Table 2. Summary of treatment factor-level combinations with mean and standard error computed across all decoys within each treatment.

Treatment	Time of day	Filter	Clouds	Mean wind speed (m s^{-1})	Secchi depth (m)	Decoy depth (m)	Mean detection probability	Standard error
1	<1030	NIR	N	<4	1–1.5	<0.6	0.17	0.17
2	1030–1330	NIR	N	<4	1–1.5	<0.6	0.87	0.11
3	<1030	RE	N	<4	1–1.5	<0.6	0.56	0.28
4	1030–1330	RE	N	<4	1–1.5	<0.6	0.87	0.13
5	<1030	NIR	N	<4	1–1.5	0.6–1	0.00	0.00
6	1030–1330	NIR	N	<4	1–1.5	0.6–1	0.24	0.21
7	<1030	RE	N	<4	1–1.5	0.6–1	0.00	0.00
8	1030–1330	RE	N	<4	1–1.5	0.6–1	0.00	0.00
9	1030–1330	NIR	N	<4	1–1.5	>1	0.02	0.02
10	1030–1330	RE	N	<4	1–1.5	>1	0.00	0.00
11	1030–1330	RGB	Y	>8	>1.5	<0.6	0.18	0.09
12	<1030	RGB	Y	4–8	>1.5	0.6–1	0.01	0.01
13	1030–1330	RGB	Y	>8	>1.5	0.6–1	0.05	0.03
14	<1030	NIR	Y	4–8	>1.5	0.6–1	0.03	0.03
15	<1030	RE	Y	4–8	>1.5	0.6–1	0.12	0.12
16	<1030	RGB	Y	4–8	>1.5	>1	0.00	0.00
17	1030–1330	RGB	Y	>8	>1.5	>1	0.23	0.23
18	<1030	NIR	Y	4–8	>1.5	>1	0.05	0.05
19	<1030	RE	Y	4–8	>1.5	>1	0.00	0.00
20	1030–1330	RGB	N	<4	1–1.5	<0.6	0.96	0.02
21	1030–1330	RGB	N	<4	1–1.5	0.6–1	0.07	0.07
22	<1030	RGB	Y	4–8	1–1.5	0.6–1	0.67	0.33
23	<1030	RGB	Y	4–8	1–1.5	>1	0.01	0.01
24	<1030	RE	Y	4–8	1–1.5	0.6–1	0.67	0.33
25	<1030	RE	Y	4–8	1–1.5	>1	0.00	0.00
26	<1030	NIR	Y	4–8	1–1.5	0.6–1	0.64	0.32
27	<1030	NIR	Y	4–8	1–1.5	>1	0.05	0.04
28	1030–1330	RGB	Y	4–8	<1	<0.6	0.80	0.20
29	1030–1330	RGB	Y	4–8	<1	0.6–1	0.07	0.05
30	1030–1330	RE	Y	4–8	<1	<0.6	0.92	0.06
31	1030–1330	RE	Y	4–8	<1	0.6–1	0.10	0.08
32	1030–1330	NIR	Y	4–8	<1	<0.6	0.68	0.14
33	1030–1330	NIR	Y	4–8	<1	0.6–1	0.11	0.08
34	1030–1330	RGB	Y	4–8	1–1.5	<0.6	0.51	0.18
35	1030–1330	NIR	Y	4–8	1–1.5	<0.6	0.26	0.12
36	1030–1330	RE	Y	4–8	1–1.5	<0.6	0.51	0.20

consistent difference ($\chi^2 = 3.08$, df = 2, $p = 0.215$) (Fig. 3). The only factor that did not affect mean detection probability was filter ($\chi^2 = 0.67$, df = 2, $p = 0.713$) (Fig. 3).

Higher detection probabilities (0.55 ± 0.05 SE) were associated with shallow depths (<0.72 m). Within the shallow depths, the highest detection probabilities (0.78 ± 0.05 SE) were associated with low wind speed (<4.2 m s^{-1}). At higher wind speeds (≥4.2 m s^{-1}), there were also relatively high detection probabilities (0.62 ± 0.14 SE) associated with the shallowest depths (<0.35 m). Detection probabilities were lower (0.22 ± 0.06 SE) when depths were intermediate (<0.72 m, ≥0.35 m), and with high wind speed (≥4.2 m s^{-1}). The lowest detection probabilities (0.05 ± 0.01 SE) were associated with the deepest depths (≥0.72 m) (Fig. 4).

Misidentification rates yielded no clear patterns among factor-level comparisons or between "species": time of day (class 1, $Z = 0.99$, $p = 0.320$; class 2, $Z = 0.09$, $p = 0.932$), filter (class 1, $\chi^2 = 1.08$, df = 2, $p = 0.582$; class 2, $\chi^2 = 0.22$, df = 2, $p = 0.896$), cloud presence (class 1, $Z = 0.95$, $p = 0.342$; class 2, $Z = 0.26$, $p = 0.798$), wind (class 1, $\chi^2 = 1.6$, df = 2, $p = 0.450$; class 2, $\chi^2 = 0.09$, df = 2, $p = 0.958$), and decoy depth (class 1, $\chi^2 = 0.92$, df = 2, $p = 0.631$; class 2,

Fig. 3. Factor-level comparisons for detection probabilities related to each of the five factors. Data are presented as mean detection probability ± 1 SE.

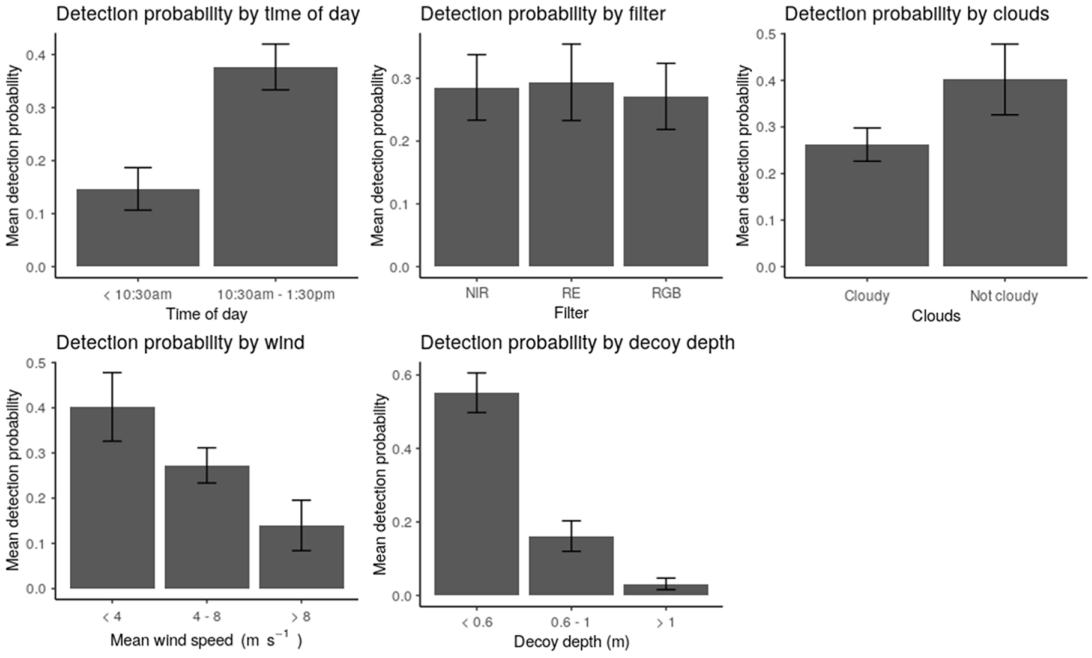
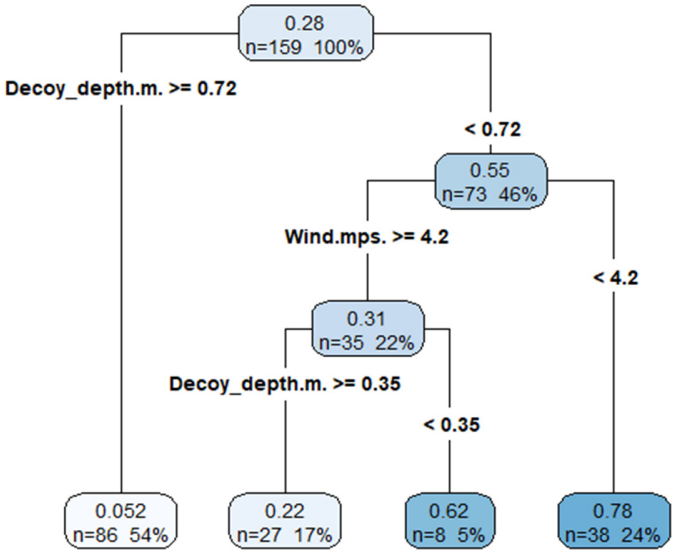


Fig. 4. Regression tree showing split decisions as well as mean detection probability (#.##) at each node and leaf. Also shown are the number of cases in each node as a raw number (*n*) and percentage (%) out of 159 total decoys in images.



$Z = -0.09$, $p = 0.932$). We also failed to detect any clear patterns or meaningful differences in false detections by factor-levels: time of day ($Z = 0.76$, $p = 0.450$), filter ($\chi^2 = 1.63$, df = 2, $p = 0.442$), clouds ($Z = 0.79$, $p = 0.427$), and wind ($\chi^2 = 1.47$, df = 2, $p = 0.478$).

Discussion

By deploying shark decoys across multiple environmental contexts in a temperate estuary we demonstrated that UAS surveys, with the ability to target smaller areas with greater precision and at higher sampling frequencies relative to manned aircraft, may have potential for answering specifically targeted ecological questions about sharks in this and similar environmental systems. The main factor influencing detection probabilities in our study was decoy depth, constraining surveys to shallow water to reliably detect sharks. This is likely due to visibility bias from turbidity, as increases in turbidity increase the rate of light attenuation throughout the water column (Brown 1984), presumably leading to greater concealment of decoys at depth. Robbins et al. (2014) used shark decoys that were slowly raised from depths of at least 5 m until they became visible to estimate the depth at which the decoy could be seen from aerial surveys conducted via manned aircraft. In that study, water turbidity measurements were taken across flight days using a secchi disk and were deeper than the average depth at which the decoys were observed, suggesting turbidity may not be the only factor affecting visibility bias (Robbins et al. 2014). Our study suggests that time of day, wind, and cloud cover may be additional factors affecting visibility bias.

The comparison of time of day (morning versus midday) showed significant differences in detection probability, with mean detection probability during midday over two times as high as during the morning. This result was somewhat surprising as we had hypothesized that the high solar altitude at midday would create more glare when photos were taken from overhead, thereby increasing visibility bias as decoys become concealed beneath the glare. Total solar irradiance reaches a maximum at noon and the reflectance of incident solar radiation increases with increasing zenith angle of incidence (Kirk 1994). This means that while there may be more glare from overhead during midday solar angles, there is also more light available and greater penetration into the water, which could increase visibility. There was a notable effect of wind on detection probabilities during midday, however, with high winds leading to mean detection probabilities less than half of those at lower winds; this could possibly be explained by the increased scattering of light at the surface and thus lower availability and penetration of light in the water column. The availability and penetration of light into the water also likely explains the increased detection probabilities on days with fewer clouds in the sky.

If a decoy was detected, it generally could be identified as a bonnethead or not in ~75% of cases, insensitive to environmental conditions at each decoy. Misidentification rates also do not appear to vary across bonnethead and non-bonnethead decoys, which means that biases towards overestimation and underestimation of bonnethead sharks would be driven mainly by imbalances in abundance of bonnethead versus non-bonnethead species. Likewise, environmental variability does not appear to significantly alter the possibility of a decoy being spotted where it does not actually exist. These results suggest that the main obstacle to reliable estimation of species abundance from aerial drone surveys is visibility bias due to shark depth, and the likely underestimation of true shark abundance in temperate estuaries based solely on aerial surveys.

While our secchi depth measurements provided a description of the range of visibility across our flight days, the frequency at which they were taken (once per flight day) was not sufficient to provide a proxy for turbidity that could be correlated with each of our survey flights, not to mention the potential for spatial differences across our flight areas. Nonetheless, our minimum secchi depth value (0.7 m) roughly coincides with the first split decision in our regression tree (0.72 m decoy depth). There is roughly a 5% chance that a decoy would be spotted at depths >0.7 m; this is not surprising considering that this depth was the visibility minimum for our flight days.

Our study is bounded by some constraints that guide the foci of our broader conclusions regarding the role of UASs in shark surveys. Due to our focus on bonnethead sharks, we only included decoys of small sharks (~1 m), which could have an effect on detection probabilities. In addition, we chose to use still images rather than video, which, especially in the case of surveying living sharks, could potentially influence rates of detections and (or) false detections. We also were unable to test for the effects of different types of substrate beneath our decoys on the detection probability. Presumably, different colors or textures would influence visibility bias depending on how they contrasted with the shark's countershading pattern; however, it should be noted that in tropical high-transparency water, benthic characteristics had no effect on shark decoy detectability from drone surveys (Hensel et al. 2018). While our study was experimental in terms of our control of decoy placement, it was observational in terms of susceptibility to unpredictable environmental changes, which limited our sample sizes for some environmental variables. Finally, mainly due to our study focusing on one UAS platform (fixed-wing), the flight altitude was a variable we kept constant, which could certainly have an effect on detection probabilities due to changes in visibility and image resolution at increased altitudes.

In summary, our decoys demonstrated that drone surveys for sharks in a turbid, temperate estuary, such as the Newport River Estuarine System, probably only work in very shallow water (<0.7 m). Because turbidity increases the rate of attenuation of light at depth, visibility bias of sharks is increased, particularly at depths that exceed the minimum visibility or secchi disk depth. Wind could be a mechanism that exacerbates this visibility bias as it causes further resuspension of solids and alters reflection and refraction of light at the surface. Increasing solar altitude, while potentially causing increased glare in photographs taken from overhead, also leads to increased light availability and penetration in the water column, which could positively affect the detection of sharks from UAS surveys. Our results are in agreement with Kiszka et al. (2016), who suggested that UASs are particularly attractive for investigating population trends and habitat use patterns where visibility enables animal detection from the surface to the bottom of the water column. As interest in this approach to monitor sharks in coastal environments for public safety is increasing, it is important to understand the limitations across different coastal environments, some of which can be quite turbid. We agree with Pollock et al. (2006), who suggest that standardized protocols and strict ceilings on acceptable survey conditions can reduce variation in detection probabilities. We suggest that in temperate estuarine systems, which can have high turbidity, UAS surveys may need to be restricted to areas where the depth is shallower than the visibility minimum.

Acknowledgements

This research project was funded by a North Carolina Aquarium Society Conservation Grant (award No. 2015-03) and the first author was supported by a Ph.D. scholarship from the Peruvian National Council for Science, Technology and Technological Innovation (CONCYTEC). Fieldwork was conducted under Research Permit No. 7-2016 issued by the N.C. Coastal Reserve and National Estuarine Research Reserve. We thank members of the Marine Robotics and Remote Sensing Lab at the Duke University Marine Lab, especially Everette Newton and Julian Dale for UAS flight planning and logistics, as well as Austin Moore for assistance with experimental design and decoy construction. We also thank members of the Coastal Fisheries Oceanography and Ecology lab at the UNC Institute of Marine Sciences, especially Matthew Kenworthy, Shelby Ziegler, Maxwell Tice-Lewis, and Mariah Livernois for assistance with decoy deployment in the field. We thank interns Giada Bargione (Polytechnic University of Marche) and Connor Neagle (North Carolina State University) for assistance with image assessments and data analysis. We also thank

our volunteer scorers (not already named): Amy Yarnall, Cori Lopazanski, Danielle Keller, Lauren Clance, Owen Mulvey-McFerron, James Morley, Omar Fais, Haley Ealey, Ryan Giannelli, Charles Bangley, Claire Pelletier, and Seth Sykora-Bodie. Finally, comments from Stephen Fegley (UNC Institute of Marine Sciences), Nathan Bacheler (NOAA Beaufort Lab), and two anonymous reviewers were most helpful in refining and revising the manuscript.

References

Anderson, K., and Gaston, K.J. 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. Front. Ecol. Environ. **11**(3): 138–146. doi: 10.1890/120150.

Bansemer, C.S., and Bennett, M.B. 2008. Multi-year validation of photographic identification of grey nurse sharks, *Carcharias taurus*, and applications for non-invasive conservation research. Mar. Freshw. Res. **59**(4): 322–331. doi: 10.1071/MF07184.

Barlow, J. 2015. Inferring trackline detection probabilities, g(0), for cetaceans from apparent densities in different survey conditions. Mar. Mammal Sci. **31**(3): 923–943. doi: 10.1111/mms.12205.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.I. 1984. Classification and regression trees. Chapman & Hall, Boca Raton, Fla., USA.

Brown, R. 1984. Relationships between suspended solids, turbidity, light attenuation, and algal productivity. Lake Reserv. Manage. **1**(1): 198–205. doi: 10.1080/07438148409354510.

Buckland, S.T., Anderson, D., Burnham, K., and Laake, J. 2001. Introduction to distance sampling. Oxford University Press, Oxford, UK.

Buckland, S.T., Burt, M.L., Rexstad, E.A., Mellor, M., Williams, A.E., and Woodward, R. 2012. Aerial surveys of seabirds: the advent of digital methods. J. Appl. Ecol. **49**(4): 960–967. doi: 10.1111/j.1365-2664.2012.02150.x.

Caughley, G. 1974. Bias in aerial survey. J. Wildl. Manage. **38**(4): 921–933. doi: 10.2307/3800067.

Christie, K.S., Gilbert, S.L., Brown, C.L., Hatfield, M., and Hanson, L. 2016. Unmanned aircraft systems in wildlife research: current and future applications of a transformative technology. Front. Ecol. Environ. **14**(5): 241–251. doi: 10.1002/fee.1281.

Colefax, A.P., Butcher, P.A., Pagendam, D.E., and Kelaher, B.P. 2019. Reliability of marine faunal detections in drone-based monitoring. Ocean Coast. Manage. **174**: 108–115. doi: 10.1016/j.ocecoaman.2019.03.008.

Compagno, L., Dando, M., and Fowler, S. 2005. Sharks of the world. Princeton University Press, Princeton, N.J., USA.

De'Ath, G., and Fabricius, K.E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology, **81**(11): 3178–3192. doi: 10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2.

Hensel, E., Wenclawski, S., Layman, C., Hensel, E., Wenclawski, S., and Layman, C.A. 2018. Using a small, consumer grade drone to identify and count marine megafauna in shallow habitats. Lat. Am. J. Aquat. Res. **46**(5): 1025–1033. doi: 10.3856/vol46-issue5-fulltext-15.

Hodgson, A., Kelly, N., and Peel, D. 2013. Unmanned aerial vehicles (UAVs) for surveying Marine Fauna: a dugong case study [online]. PLoS ONE, **8**(11): e79556. doi: 10.1371/journal.pone.0079556.

Hothorn, T., Hornik, K., van de Wiel, M.A., and Zeileis, A. 2008. Implementing a class of permutation tests: the coin package. J. Stat. Softw. **28**(8): 1–23. doi: 10.18637/jss.v028.i08.

Johnston, D.W. 2019. Unoccupied aircraft systems in marine science and conservation. Annu. Rev. Mar. Sci. **11**(1): 439–463. doi: 10.1146/annurev-marine-010318-095323.

Jones, G.P.I., Pearlstine, L.G., and Percival, H.F. 2006. An assessment of small unmanned aerial vehicles for wildlife research. Wildl. Soc. Bull. **34**(3): 750–758. doi: 10.2193/0091-7648(2006)34[750:AAOSUA]2.0.CO;2.

Kaschner, K., Quick, N.J., Jewell, R., Williams, R., and Harris, C.M. 2012. Global coverage of cetacean line-transect surveys: status quo, data gaps and future challenges [online]. PLoS ONE, **7**(9): e44075. doi: 10.1371/journal.pone.0044075.

Kirby-Smith, W.W., and Costlow, J.D. 1989. The Newport River Estuarine System. UNC Seagrant College Publication UNC-SG-89-04. Available from the University of North Carolina Sea Grant, Beaufort, N.C., USA.

Kirk, J.T.O. 1994. Light and photosynthesis in aquatic ecosystems. Cambridge University Press, Cambridge, UK.

Kiszka, J.J., Mourier, J., Gastrich, K., and Heithaus, M.R. 2016. Using unmanned aerial vehicles (UAVs) to investigate shark and ray densities in a shallow coral lagoon. Mar. Ecol. Prog. Ser. **560**: 237–242. doi: 10.3354/meps11945.

Kohler, N.E., and Turner, P.A. 2001. Shark tagging: a review of conventional methods and studies. Environ. Biol. Fishes, **60**: 191–224. doi: 10.1023/A:1007679303082.

Koski, W.R., Thomas, T.A., Funk, D.W., and Macrander, A.M. 2013. Marine mammal sightings by analysts of digital imagery versus aerial surveyors: a preliminary comparison. J. Unmanned Veh. Syst. **1**(1): 25–40. doi: 10.1139/juvs-2013-0015.

Marsh, H., and Sinclair, D.F. 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. J. Wildl. Manage. **53**(4): 1017–1024. doi: 10.2307/3809604.

McClellan, C.M., Brereton, T., Dell'Amico, F., Johns, D.G., Cucknell, A.C., Patrick, S.C., et al. 2014. Understanding the distribution of marine megafauna in the English Channel region: identifying key habitats for conservation within the busiest seaway on earth [online]. PLoS ONE, **9**(2): e89720. doi: 10.1371/journal.pone.0089720.

Milborrow, S. 2017. rpart.plot: plot 'rpart' models: an enhanced version of 'plot.rpart'. Available from https://cran.r-project.org/package=rpart.plot [accessed 27 March 2018].

Murtaugh, P. 2014. In defense of *P* values. Ecology, **95**(3): 611–617. doi: 10.1890/13-0590.1.

Myrberg, A.A., and Gruber, S.H. 1974. The behavior of the bonnethead shark, *Sphyrna tiburo*. Copeia, **1974**(2): 358–374. doi: 10.2307/1442530.
NOAA. 2018. Water levels — NOAA tides & currents. Available from https://tidesandcurrents.noaa.gov/waterlevels.html?id=8656483 [accessed 5 June 2019].
Peterson, C.D., Belcher, C.N., Bethea, D.M., Driggers, W.B., Frazier, B.S., and Latour, R.J. 2017. Preliminary recovery of coastal sharks in the south-east United States. Fish Fish. **18**(5): 845–859. doi: 10.1111/faf.12210.
Pollock, K.H., Marsh, H.D., Lawler, I.R., and Alldredge, M.W. 2006. Estimating animal abundance in heterogeneous environments: an application to aerial surveys for dugongs. J. Wildl. Manage. **70**(1): 255–262. doi: 10.2193/0022-541X(2006)70[255:EAAIHE]2.0.CO;2.
QGIS Development Team. 2018. QGIS geographic information system. Open Source Geospatial Foundation Project. Available from http://qgis.osgeo.org [accessed 6 August 2018].
R Core Team. 2016. R: A language and environment for statistical computing. Available from http://www.r-project.org [accessed 27 March 2018].
Rees, A., Avens, L., Ballorain, K., Bevan, E., Broderick, A., Carthy, R., et al. 2018. The potential of unmanned aerial systems for sea turtle research and conservation: a review and future directions. Endanger. Species Res. **35**: 81–100. doi: 10.3354/esr00877.
Revelle, W. 2017. psych: procedures for personality and psychological research. Available from https://cran.r-project.org/package=psych [accessed 27 March 2018].
Robbins, W.D., Peddemors, V.M., Kennelly, S.J., and Ives, M.C. 2014. Experimental evaluation of shark detection rates by aerial observers [online]. PLoS ONE, **9**(2): e83456. doi: 10.1371/journal.pone.0083456.
Rowat, D., Gore, M., Meekan, M.G., Lawler, I.R., and Bradshaw, C.J.A. 2009. Aerial survey as a tool to estimate whale shark abundance trends. J. Exp. Mar. Bio. Ecol. **368**(1): 1–8. doi: 10.1016/j.jembe.2008.09.001.
Sweeney, K.L., Helker, V.T., Perryman, W.L., LeRoi, D.J., Fritz, L.W., Gelatt, T.S., and Angliss, R.P. 2015. Flying beneath the clouds at the edge of the world: using a hexacopter to supplement abundance surveys of Steller sea lions (*Eumetopias jubatus*) in Alaska. J. Unmanned Veh. Syst. **4**(1): 70–81. doi: 10.1139/juvs-2015-0010.
Therneau, T., Atkinson, B., and Ripley, B. 2015. rpart: recursive partitioning and regression trees. Available from https://cran.r-project.org/package=rpart [accessed 27 March 2018].
Weather Underground. 2011. Weather underground: forecasts. Available from https://play.google.com/store/apps/details?id=com.wunderground.android.weather [accessed 30 June 2018].
Wickham, H. 2009. ggplot2: elegant graphics for data analysis. Available from http://ggplot2.tidyverse.org [accessed 27 March 2018].
Wickham, H. 2017. tidyr: tidy messy data. Available from https://cran.r-project.org/package=tidyr [accessed 27 March 2018].
Wickham, H., and Francois, R. 2016. diplyr: a grammar of data manipulation. Available from https://cran.r-project.org/package=dplyr [accessed 27 March 2018].