Remote Sensing in Ecology and Conservation Open Access

RESEARCH ARTICLE

Using Unoccupied Aerial Vehicles to estimate availability and group size error for aerial surveys of coastal dolphins

Alexander M. Brown[1,2], Simon J. Allen[3,4,5], Nat Kelly[6] & Amanda J. Hodgson[1]

[1]Centre for Sustainable Aquatic Ecosystems, Harry Butler Institute, Murdoch University, Perth, Western Australia 6150, Australia
[2]SMRU Consulting, Scottish Oceans Institute, East Sands, University of St Andrews, St Andrews KY16 8LB, United Kingdom
[3]School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, United Kingdom
[4]School of Biological Sciences, University of Western Australia, Crawley, Western Australia 6009, Australia
[5]Department of Anthropology, University of Zurich, 8057 Zurich, Switzerland
[6]Australian Antarctic Division, 203 Channel Highway, Kingston, Tasmania 7050, Australia

Keywords
Abundance estimation, availability bias, cetaceans, drone, UAS, UAV

Correspondence
Alexander M. Brown, SMRU Consulting, Scottish Oceans Institute, East Sands, University of St Andrews, St Andrews KY16 8LB, UK. Tel: +44 1334 466011; E-mail: ab@smruconsulting.com

Funding Information
The Western Australian Department of Biodiversity, Conservation and Attractions and the Sea World Research and Rescue Foundation Inc. funded this research.

Editor: Nathalie Pettorelli
Associate Editor: Nicola Quick

Received: 19 December 2021; Revised: 10 July 2022; Accepted: 27 September 2022

doi: 10.1002/rse2.313

Abstract

Aerial surveys are frequently used to estimate the abundance of marine mammals, but their accuracy is dependent upon obtaining a measure of the availability of animals to visual detection. Existing methods for characterizing availability have limitations and do not necessarily reflect true availability. Here, we present a method of using small, vessel-launched, multi-rotor Unoccupied Aerial Vehicles (UAVs, or drones) to collect video of dolphins to characterize availability and investigate error surrounding group size estimates. We collected over 20 h of aerial video of dive-surfacing behaviour across 32 encounters with Australian humpback dolphins *Sousa sahulensis* off north-western Australia. Mean surfacing and dive periods were 7.85 sec (SE = 0.26) and 39.27 sec (SE = 1.31) respectively. Dolphin encounters were split into 56 focal follows of consistent group composition to which example approaches to estimating availability were applied. Non-instantaneous availability estimates, assuming a 7 sec observation window, ranged between 0.22 and 0.88, with a mean availability of 0.46 (CV = 0.34). Availability tended to increase with increasing group size. We found a downward bias in group size estimation, with true group size typically one individual more than would have been estimated by a human observer during a standard aerial survey. The variability of availability estimates between focal follows highlights the importance of sampling across a variety of group sizes, compositions and environmental conditions. Through data re-sampling exercises, we explored the influence of sample size on availability estimates and their precision, with results providing an indication of target sample sizes to minimize bias in future research. We show that UAVs can provide an effective and relatively inexpensive method of characterizing dolphin availability with several advantages over existing approaches. The example estimates obtained for humpback dolphins are within the range of values obtained for other shallow-water, small cetaceans, and will directly inform a government-run program of aerial surveys in the region.

Introduction

Aerial surveys are used around the world for estimating the abundance of marine mammals, the results of which can inform conservation and management strategies (e.g. Grech et al., 2011; Hammond et al., 2013; Heide-Jørgensen et al., 2008). In order to provide robust abundance estimates, it is also important to estimate and account for the proportion of animals that are not detected. This probability of detection has two components: (i) the probability of an animal being close enough to the surface to be visible (availability); and (ii) the probability of an animal being detected by the observers, conditional upon it being available for detection (perception) (Marsh & Sinclair, 1989; Pollock et al., 2006). Perception bias can be estimated from survey data using independent teams of observers (Pollock et al., 2006). However, estimating availability—defined as the proportion of time that animals are available for detection—requires data additional to that collected during the survey itself.

As a multiplier, an availability correction has a major influence on abundance estimates and, accordingly, inaccurate measures of availability can mean that abundance is significantly over- or under-estimated. For example, population estimates of dugongs *Dugong dugon* in Torres Strait were recently shown to be 6–7 fold greater following the application of more accurate, depth-specific availability correction factors than those used previously (Hagihara et al., 2018). Existing methods used to estimate availability include: (i) recording the surface times of focal animals from a land- or boat-based platform or helicopter (e.g. Forcada et al., 2004; Slooten et al., 2004); (ii) combining tag-derived dive-depth data with estimates of visibility at depth (e.g. Heide-Jørgensen et al., 2012; Pollock et al., 2006) and, (iii) the comparison of aerial survey population estimates to concurrent land-based survey/census results (e.g. Hedley et al., 2011). These methods have various limitations (e.g. cost, personnel safety, logistical constraints, an unrepresentative observation platform, or tags influencing animal behaviour), which may render them impractical or result in inaccurate availability corrections. In particular, observations of surfacing behaviour made from land- or boat-based platforms do not reflect the perspective of an aerial observer, and are likely to under-estimate availability as the angle of observation limits the extent to which submerged animals in the upper few metres of the water column may be visible.

Technological developments over the past decade have made Unoccupied Aerial Vehicles (UAVs, or drones) an increasingly suitable tool for wildlife research (Christie et al., 2016; Nowacek et al., 2016). UAVs are now demonstrably effective platforms for aerial surveys of marine mammals (Hodgson et al., 2013; Seymour et al., 2017), and therefore also tools for gathering data on dive/surfacing behaviour to assess availability from an aerial perspective, while avoiding many of the limitations of alternative approaches (Hodgson et al., 2017).

Here, we present a method for collecting aerial observations of dive/surfacing behaviour of coastal dolphins using small, multi-rotor UAVs deployed from a small vessel. We focused on Australian humpback dolphins (*Sousa sahulensis*, hereafter 'humpback dolphins') off north-western Australia - a species subject to a recent program of aerial surveys in the region, conducted by the Western Australian (WA) state government. Video observations were analysed to estimate the visual availability and group size error of humpback dolphins to aerial observers for application to the estimation of dolphin abundance according to aerial survey data. Using an estimator of the probability that a group is available at least once during an example observation window, we assess inter-group variability in availability. To assess the reliability of our results, we undertook data re-sampling exercises to explore the influence of sample size on availability estimates and their precision.

Materials and Methods

Study area

This study was conducted in the coastal waters of the Dampier Archipelago and the North West Cape, WA (Fig. S1.1). These locations lie within the boundaries of the WA state government's aerial survey program and are representative of the habitats that humpback dolphins occupy in the region (Brown et al., 2012; Hanf et al., 2016). Dolphins in both locations are habituated to small vessel traffic (Allen et al., 2012; Hunt et al., 2017).

UAV system

We used the DJI Phantom 4 Pro, a light (1.4 kg), multi-rotor UAV with an integrated, gimbal-controlled camera. This model was a compromise between affordability (c. US$2000 per unit in 2017), endurance (realistic max flight duration of c. 20–25 min), video quality (3840 × 2160 pixels) and ease of operation. When flying at 80 m above sea level (asl) and with the camera oriented directly perpendicular to the sea surface (nadir), the UAV camera's fixed focal length lens provided a visible area of 116 m (horizontally) × 87 m (vertically). Polarizing filters were fitted to the camera lens throughout data collection. Further details of the UAV system and data collection are provided in Appendix S1 – UAV reporting protocol (following Barnas et al. (2020)).

Data collection

To locate focal dolphin groups, a crew of four people surveyed the coastal waters of each study site from a small (6 m length) research vessel. Survey effort was focused in areas where dolphins had been observed in previous studies (Allen et al., 2012), and where sea conditions for sighting dolphins were most favourable, that is, Beaufort sea state ≤2. Upon sighting dolphins, the date, time, GPS location, water depth, secchi depth, cloud cover (oktas) and group size, composition (adults, juveniles, calves) and behaviour were recorded. We commenced UAV video capture and used an mp3 audio recorder to record verbal notes on missed surfacings, changes in group composition and environmental conditions. Audio was synchronized with the video using a handclap before launching the UAV from a purpose-built platform secured to the bow of the vessel. The UAV was flown to a typical altitude of 80 m (±10 m) and a position offset (towards the sun) from the dolphin group, with the camera angled at 45° (±10°) down from horizontal to: (i) minimize sea surface glare; (ii) provide a larger field of view than a nadir orientation; (iii) keep the vessel in the periphery of the frame and, (iv) provide a perspective comparable to that of an aerial survey observer (Fig. 1). In the case of (iv), the UAV operator aimed to position dolphins in the bottom two-thirds of the frame; animals in the upper third of the frame were distant and often only visible when breaking the surface (particularly in overcast conditions), and therefore unlikely to be representative of the majority of sightings in an aerial survey. Regarding (iii), when a group dived out of view, it was important to maximize the likelihood of all members of the group being in the video frame upon their next surfacing. While the group was submerged, the UAV operator aimed to ensure the video frame included an area of sea along their anticipated path from their last surface position, and the research vessel as a reference point (Fig. 1). Throughout each follow the vessel driver maintained a position c. 100–400 m from the dolphin group, either with the motor in neutral or slow idle speed to minimize any potential disturbance.

Our operational altitude was limited to between 20 m (research permit) and 120 m (aviation regulation); a pilot study had shown that c. 80 m (±10 m) altitude represented a suitable compromise between a field of view that maximized our chance of capturing the moment a group surfaced after a dive, and being able to see the dolphins on the remote controller (RC) screen. When it was

Figure 1. Main image: screenshot (cropped) of representative footage collected by the UAV during focal follows to illustrate the typical positioning of the dolphin group (lower left) relative to that of the vessel (upper right) for orientation purposes. Altitude = 55 m; position in frame = 1; glare = 1; turbidity = 3; cloud cover = 8. Inset: typical positioning of a UAV during a focal follow, captured by a second UAV further from the vessel and at a higher altitude.

proving difficult to capture the dolphins surfacing following a dive (e.g. because they were travelling quickly or undertaking long dives with variable headings), the UAV was flown more consistently at an altitude of c. 100 m.

We aimed to collect 40–60 min of video per dolphin encounter (i.e. two to three flights) to capture multiple dive-surface cycles of a group while maximizing the number of different groups we could observe. Where multiple changes in group composition and/or environmental conditions were apparent within an encounter, we extended the data collection to 120 min (research permit limit). Some encounters ended prematurely when dolphins were lost or sea conditions were too challenging. For each UAV flight, we recorded the apparent response of dolphins at the surface to the presence of the UAV on a scale where responses ranged from zero (no detectable reaction) to four (strong reaction e.g. breaches, tail slaps), as per Weinrich et al. (1991). The dolphins' response was recorded upon the initial positioning of the UAV above the animals, and we noted any subsequent behavioural changes which appeared to coincide with the presence of the UAV.

We used two UAVs operated by two trained personnel in order to capture continuous footage of dolphins beyond the max flight duration of a single UAV. The two UAVs were rotated by having a small period of overlap (dual recording of the dolphin group by both UAVs).

Video review

Video was reviewed on a 27-inch monitor of 3840 × 2160 pixel resolution (LG 27UD69-W). We avoided using the zoom function so that the perspective of the video reviewer was as similar to an aerial survey observer as possible. Each time the number of dolphins visible changed, we recorded the timestamp (to the second), along with the presence of calves, the turbidity of the water column and the predominant position of the dolphins in the frame (split into upper, middle and lower thirds). Turbidity was categorized following Pollock et al. (2006) as either: (1) clear water, shallow depth, seabed clearly visible; (2) variable water quality, variable depth, seabed visible but unclear; (3) clear water, depth unknown, seabed not visible; (4) turbid water, variable depth, seabed not visible. Additionally, at two-min intervals, we recorded: UAV altitude (provided by video subtitle file), glare and Beaufort sea state. Glare was categorized following standard practise in aerial surveys as either: (0) no glare; (1) <25% of frame reflecting light to the extent that objects can only be seen when breaking the surface; (2) 26%–50% of frame; or (3) >50% of frame.

Two assistants each reviewed different videos, but both independently reviewed two of the same encounters to investigate any potential reviewer-bias. The lead author subsequently reviewed all data and split encounters into component focal follows wherever group composition changed, and noted visible changes in water depth (i.e. seabed visible or not). A 'group' was defined using a 'chain rule' whereby any individuals within ~100 m of another individual and engaged in the same or similar behaviour were considered part of the same group (modified from Smolker et al. 1992). Group composition was deemed to have changed when one or more individuals clearly separated from or joined the group according to our group definition. During data review, the lead author also assigned segments of data into either dive or surface phases to identify complete dive-surface cycles. Surface phases typically included several inter-breath intervals, during which animals could disappear from view for a few seconds. Therefore, a dive phase was defined as all focal animals being submerged from view for ≥15 s; following an exploration of the data, this value was seen as appropriate for differentiating shorter inter-breath intervals (more frequent) from longer/deeper dives (less frequent). Incomplete dive-surface cycles were excluded from the data. It is acknowledged that removing incomplete dive cycles potentially introduces bias as longer dives will be more likely to be removed than shorter dives (which may result in positive bias in the proportion of time available); however, there was more likely to be a bias (either positive or negative) as a result of using follows where we only captured part of the dive or surface phase of a dive-surface cycle, particularly for short focal follows. When characterizing the availability process and estimating availability (see below), animals were considered unavailable whenever submerged from view, regardless of duration or assigned phase.

Characterizing the availability process and estimating availability

Our approach to characterizing the availability process and estimating availability largely follows that described by Hodgson et al. (2017) and Laake et al. (1997). All statistical analyses were performed in the R statistical environment (R Core Team, 2017). It was not possible to consistently track individual dolphins across multiple surfacing bouts, so availability was characterized and estimated at the group level. A group was considered visible/available for detection whenever at least one individual was visible. The availability process for the group was characterized by the sequence of a 'surface' period (i.e. when at least one individual of the group was visible, for any length of time) and a 'dive' period (i.e. when no individuals were visible at the surface, for any length of time), where times of surface/dive changes were recorded to the nearest second, and where each surface/dive event was defined by a start and end time. The sequence of surfacing and dive durations were used to estimate mean surfacing and mean dive times, and an associated variance for durations of both phases, for a given component focal follow.

Several approaches have been developed to estimate availability and develop correction factors from line-transect surveys, with approaches differing in terms of data requirements, sources and extent of bias, and suitability to different survey platforms (Borchers et al., 2013). It is important to consider different approaches and select that which is most appropriate to the specific application. Here, having characterized the availability process, we present two example approaches to estimating availability. The first is a simple instantaneous availability correction factor, which is appropriate when only animals at a specified forward distance are included (e.g. distance zero, as animals come abeam). This instantaneous availability was estimated as the total time a group was visible during a follow divided by the total duration of that follow. To minimize the bias that can be introduced via ratios, a jackknife approach (recommended by Choquet et al. 1999) was used to estimate the mean instantaneous availability across all follows and the variance between follows.

The second is an estimate of non-instantaneous availability, which accounts for the period of time an observer can scan an area during an aerial survey. For each follow, group availability ($\widehat{a}$) was estimated as the average probability that a group is available at least once while in view for a specified period of time (Laake et al., 1997). This non-instantaneous availability estimator is given by the following equation:

$$\widehat{a} = \frac{\widehat{E}(s)}{\widehat{E}(s) + \widehat{E}(d)} + \frac{\widehat{E}(d)\left[1 - \widehat{e^{\frac{-t}{E(d)}}}\right]}{\widehat{E}(s) + \widehat{E}(d)},$$

where $\hat{E}(s)$ is the average duration of uninterrupted group visibility within a single follow, $\hat{E}(d)$ is the average time the group was not visible before becoming visible again, and t is the window of time an observer has to see the group if it is available (the 'observation window'). All parameter values were in seconds. The variance of the availability estimate for each follow was estimated via the delta method (Ver Hoef, 2012), assuming that s and d are independent exponential random variables (a two-state Markov process) (Borchers et al., 2002). The assumption that s and d are exponentially distributed was tested using a Kolmogorov–Smirnov goodness-of-fit test; results are reported in Appendix S2. A weighted mean availability estimate across all (or a given subset of) focal follows was produced using focal follow duration as the weight.

The calculation of t has been approached in different ways in the literature; one approach is to calculate a theoretical time based on the width of the aircraft's window, altitude of the plane, and distance from the aircraft (e.g. McLaren, 1961). Here, we present availability and group size correction estimates for an example t of 7 sec, as this is an approximate estimate of t for our particular application. Estimates for different values of t can be produced using this study's code and data (see Data availability statement).

It is acknowledged that this approach of estimating $\widehat{a}$ is sensitive to the value of t, which can be challenging to estimate. Applying incorrect values of t will result in biased estimates. The resulting estimates from this approach may also be biased due to not properly accounting for detection probability changing while animals are within detectable range, although such bias is lower for aerial surveys than vessel-based surveys (Borchers et al., 2013). It is important to consider different approaches and the reader is encouraged to assess the approach which is most suited to their specific application. Nonetheless, in surveys where animals are in view for a very short time relative to their dive cycle duration (e.g. aerial surveys for small cetaceans), the approach to estimating non-instantaneous availability presented here has been considered appropriate (ASCOBANS, 2015), and can easily be adjusted to different values of t. Borchers et al. (2013) present several alternative approaches to estimating availability, including a Hidden Markov Model approach which can exhibit little or no bias, but requires 'forward' (along the transect line) distance data on the moment at which animals are first sighted by observers in addition to perpendicular distance data. Due to logistical challenges associated with the survey platform speed and short window of observation, these data are not routinely collected in aerial surveys for smaller cetaceans, and were not collected in the program of aerial surveys to which this study relates.

Group size error

Animals within a group rarely dive precisely synchronously; therefore, a group size estimate made during a relatively short observation window is likely to be, on average, less than the number of animals actually present (Hodgson et al., 2017). Group size error was estimated using a non-parametric bootstrap approach (Hodgson et al., 2017). For each follow, 1000 continuous segments of an example duration of 7 sec were randomly sampled (with replacement) from the entire follow duration and we recorded the max number of dolphins that were visible during that 7 sec segment (observed group size, including if zero). We then compared mean observed group sizes with their corresponding actual group size ('known' from the entire follow), averaged across all follows, to provide a measure of error which could be used to reduce bias in abundance estimates.

Time of day, environmental and video attributes

Plots of follow-specific availability estimates, using the example $\hat{a}$ estimator, versus time of day, environmental (e.g. turbidity, sea state) and video (e.g. altitude, position in frame) characteristics were produced to illustrate the variability in each attribute and graphically explore their potential for introducing biases in availability estimates.

Behavioural state

The behavioural state of dolphins is generally not recorded during aerial surveys because of the short window of observation, and so cannot be considered in the availability process. However, to ensure our data (and resulting availability estimates) were not biased towards dolphin groups exhibiting particular behaviours (that were, for example, easier to see and follow), we further subdivided focal follows of consistent group composition into sub-follows of consistent behavioural state. This allowed an assessment of the spread of the data across different behavioural states, and behavioural-state-specific availability to be estimated. A full description of the approach and results is provided in Appendix S3.

The influence of sample size: follow lengths and number of follows

Availability estimates can vary widely between focal follows within the same study (Bilgmann et al., 2018; Hodgson et al., 2017), raising concerns over potential biases and whether sufficient data have been collected to produce reliable estimates. Using the example non-instantaneous approach described above (Laake et al., 1997), we investigated how availability estimates and resultant precision varied with focal follow length and number of groups followed. To address the former, we truncated each follow to the first three dive-surface cycles and estimated availability and its CV. This was then repeated for increasing increments of three dive-surface cycles until the full follow length was reached. To address the number of groups followed, we randomly sub-sampled complete follows from the total available, starting with three follows, then increasing in increments of six. For each sub-sample, the mean availability and CV across the selected focal follows were estimated by using the follow-specific estimates of $\hat{E}(s)$ and $\hat{E}(d)$, with a weighting based on duration of the focal follow, to estimate mean surface and dive time, which were then also used with the non-instantaneous availability estimator; the variance of the availability estimate was derived using the delta method.

Previous studies on availability have pooled all data across focal follows to produce one availability estimate (rather than producing an estimate from each follow and obtaining a mean). To further explore the influence of sample size, we pooled data from all our follows in chronological order, as if they were one single 'super' follow, and then estimated availability in increasing increments of 10 min of data (rounded to the nearest complete dive-surface cycle).

Cost and efficiency comparison

We summarized the financial costs associated with this study, and calculated both the cost and time invested in data collection per min of data analysed; for comparison, these statistics were also estimated for two other studies of availability which used helicopters as aerial observation platforms (Bilgmann et al., 2018; Sucunza et al., 2018; Appendix S4).

Results

Data were collected over 19 days between 06 April and 05 May 2017, between the hours of 06:00 and 17:00. From a total of 45 encounters with humpback dolphins, 32 resulted in usable data (Dampier Archipelago $n = 30$; North West Cape $n = 3$). The remaining 12 encounters did not result in usable data because: animals were lost before a UAV could be launched ($n = 3$) or footage of a single complete dive-surface cycle was not obtained ($n = 7$); Indo-Pacific bottlenose dolphins *Tursiops aduncus* were also present within the same group ($n = 2$); or, poor sea conditions prohibited safe UAV operation ($n = 1$). After splitting each encounter according to group composition changes, 56 follows were available for analysis.

Our focal follows were recorded across a total of 107 UAV flights (1–8 flights per encounter; mean = 3.3; mode = 2). Most successful flights had a duration of 20–22 min (max = 23 min 34 sec), providing c. 20 min of video while allowing the UAV to land with c. 20% battery remaining as a safety margin. We successfully operated UAVs in succession with overlap and obtained continuous footage (i.e. without missed surfacings) of focal groups up to a max 90 min.

A behavioural response score of zero was recorded for all flights upon first approach to dolphins, and we detected no evidence of subsequent responses to the UAV throughout data collection.

Once data had been filtered to remove overlapping and unusable sections and ensure complete dive-surface cycles, a total of 20 h 11 min 58 sec of video footage was available for analysis. Follows ranged from 1 min 9 sec to 91 min 51 sec in duration, with a mean of 21 min 39 sec (SE = 3 min 06 sec). Dolphin group sizes ranged from 1 to 9 (median = 3; mode = 2). A total of 35 follows (corresponding to 67.8% of footage) contained one or more calves, of which nine were a single mother-calf pair. Follows were conducted across a range of environmental conditions (Appendix S3).

Potential reviewer bias

The two initial reviewers each independently reviewed the same two randomly chosen focal observations, totalling 1 h 42 min 27 sec of footage. For each follow, the two resulting availability estimates were identical for each observer to within two decimal places. As such, we assumed there was no reviewer bias.

Availability process

Surfacing periods (periods of continuous visibility) varied from 1 to 179 sec (median = 4, mean = 7.85, SE = 0.26) and diving periods (periods of continuous non-visibility) varied from 1 to 245 sec (median = 18, mean = 39.27, SE = 1.31). The average duration of uninterrupted group visibility within a single follow, $\hat{E}(s)$, varied from 3.88 to 123.33 sec and the average time the group was not visible before becoming visible again, $\hat{E}(d)$, varied from 12.12 to 148.43 sec (values for all follows and their variance are provided in Table S2.1).

Availability estimates

The mean instantaneous availability across all follows was 0.39 (CV = 0.07). Values for the non-instantaneous ($\hat{a}$) availability estimator based on an example 7 sec observation window ranged between 0.22 (CV = 1.40) and 0.88 (CV = 1.48) per follow (Table S2.1, Fig. 2), with a mean estimate of 0.46 (CV = 0.34).

Availability was higher for groups with calves (0.52, CV = 0.05, n = 35) than without (0.37, CV = 0.06, n = 21), and there was some evidence of a trend of increasing availability with larger groups (Fig. 2). When data were pooled into two group size categories, the availability estimate for small groups (≤4 individuals) was lower (0.43, CV = 0.05, n = 45 follows) than for large (≥5 individuals) groups (0.65, CV = 0.06, n = 11 follows).

Focal follows covered a variety of behavioural states, with wide variation in availability estimates within and between behavioural states (Appendix S3). Plots of availability

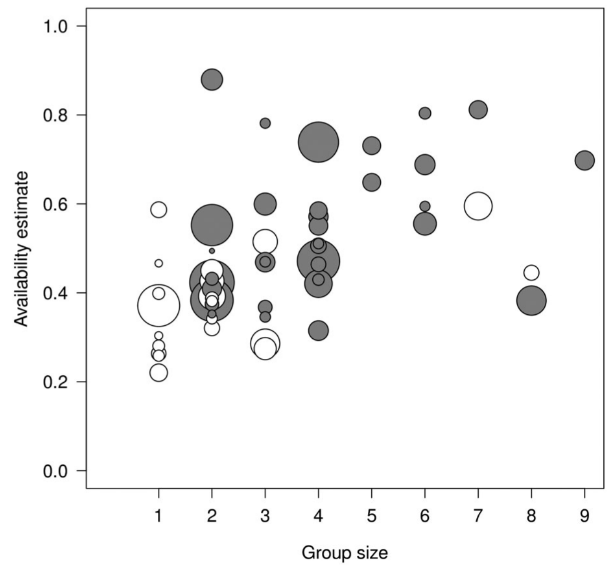

Figure 2. Non-instantaneous availability estimates according to group size, for an example 7 sec observation window. Each circle corresponds to an individual follow. Grey circles indicate groups that included one or more calves. The circle radius is proportional to the duration of each follow.

estimates *versus* time of day, environmental and video attributes did not suggest any clear biases (Appendix S4).

Group size error

The distribution of bootstrapped observed versus actual max group sizes suggests that the mean actual group size was approximately one individual larger than the observed group sizes of 1–7 (Fig. 3; Table 1). There were too few data to adequately estimate the mean actual group size for observed group sizes of 8–9 (n = 3 follows).

The influence of sample size: follow lengths and number of follows

Truncating our follows in increments of three dive-surface cycles revealed pronounced fluctuations in availability estimates between increments at c. < 9 dive-surface cycles (Fig. 4A). At > 9 dive-surface cycles, the availability estimates of individual follows generally varied less between increments, suggesting a stationarity in the result—an indication that the information content of subsequent focal follow data diminished. A similar but less pronounced pattern was observed for the CVs of individual follows (Fig. 4B), with one notable exception (follow #22, labelled in Fig. 4B) where a prolonged surface resting period caused a spike in the CV.

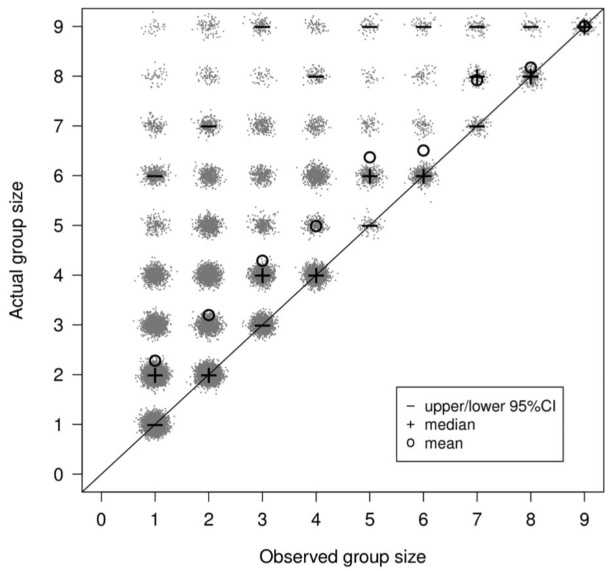

Figure 3. Distribution of observed versus actual group sizes for an example 7 sec observation window across all follows. Each dot corresponds to a realization of a nonparametric bootstrap; dots have been deliberately jittered to illustrate densities. Plus symbols indicate the median actual group size per observed group size; 'o' symbols indicate the mean actual group size per observed group size; minus symbols indicate the 95% CIs, as generated via a non-parametric bootstrap. The diagonal line indicates a 1:1 relationship.

Table 1. Observed versus actual group sizes for an example 7 sec observation window across all follows.

Observed group size	Mean actual group size (±SD)	Mean actual group size underestimate
1	2.26 (1.44)	1.26
2	3.21 (1.57)	1.21
3	4.31 (1.64)	1.31
4	4.93 (1.26)	0.93
5	6.37 (1.08)	1.37
6	6.55 (0.96)	0.55
7	7.72 (0.77)	0.72
8	8.19 (0.39)	0.19
9	9.00 (0.00)	0.00

The influence of the number of follows conducted on availability estimates is shown in Figure 5A. For a given number of follows, the scatter of mean availability estimates was largely unbiased with respect to the overall availability estimate (Fig. 5A), and, as expected, variability in the mean values decreased with increasing number of follows (Fig. 5B).

Truncating our data to include only follows of ≥ 9 dive-surface cycles reduced the total number of follows to 21, but only represented a c. 30% reduction in the total data by duration to 14 h 23 min. The mean non-instantaneous availability estimate across these 21 follows, assuming a 7 sec time window, was 0.52 (CV = 0.06).

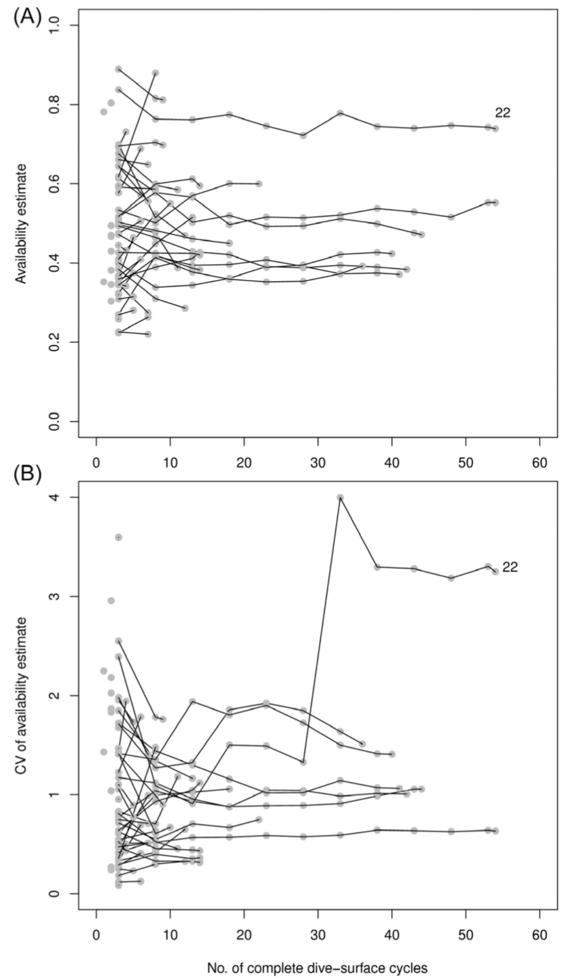

Figure 4. Changes in non-instantaneous availability estimates (A) and their CV (B), assuming an example 7 sec observation window, for each follow, in increasing increments of three dive-surface cycles. Each line represents an individual follow, with lone points representing follows with three or less dive-surface cycles in total. Follow No. 22 is labelled (see results text). The end point of each line corresponds to the availability estimates given in Appendix S2.

Examining the overall availability estimate from all data pooled in increasing increments of 10 mins further illustrated that variability decreases with increasing sample size, but did appear to approach a plateau (Fig. 6). We would have obtained an overall availability estimate up to c. 10% higher than that obtained from the full dataset were our data truncated at 200–400 min.

Cost and efficiency comparison

Considering only the initial outlay for equipment and training (US$12 200), vehicle and vessel hire and fuel (US

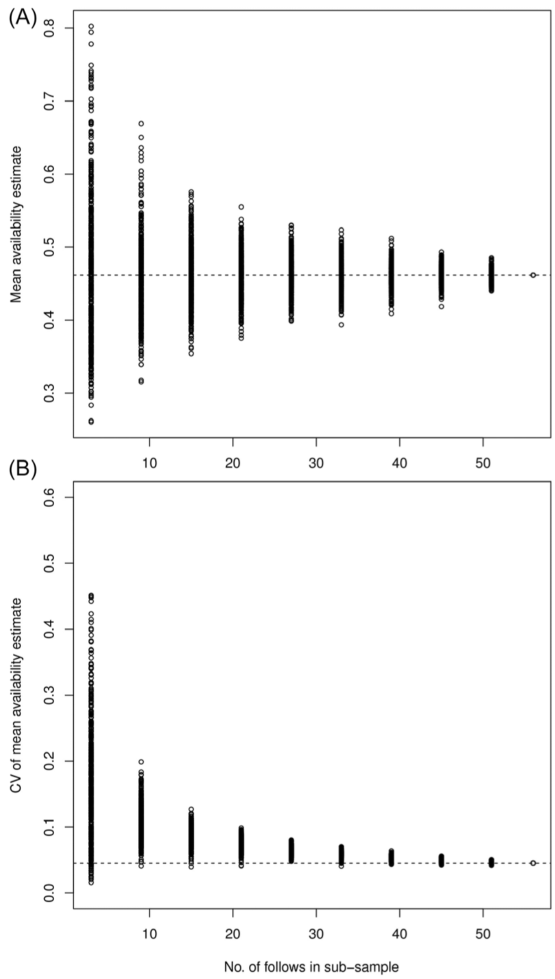

Figure 5. Mean non-instantaneous availability estimates (A) and CV on the standard error of the mean estimates (B), assuming an example 7 sec observation window, with increasing number of follows subsampled from the total 56 available. The dashed line indicates the mean availability estimate (A) and CV of the standard error of the mean (B) across all 56 follows.

$200/day), our study cost c. US$15/min of data analysed (Table S4.2). By comparison, two studies using helicopters resulted in costs of c. US$50–70/min of data analysed, assuming an indicative charter rate of US$1100/hour. When generic salary, food and accommodation rates were added to all studies, the cost of our study rose to US$34/min; however, this was still ≥ 38% cheaper than the two helicopter-based studies.

In our study, approximately 120 h were spent on the water in pursuit of data collection, equating to 5.9 min for every min of data collected and analysed; this compared to estimated ratios of 2.6:1 and 3.9:1 in Bilgmann et al. (2018) and Sucunza et al. (2018) respectively. Details are reported and discussed in Appendix S4.

Discussion

Abundance estimates from aerial surveys are routinely used to inform conservation and management strategies for cetaceans (e.g. Barlow et al., 1988; Hammond et al., 2013), and not correcting for availability provides low-biased abundance estimates (e.g. Allen et al., 2017; Bilgmann et al., 2019). Applying proxy availability correction factors derived for other regions and/or species may provide an abundance estimate closer to true abundance, but can introduce additional uncertainty and large biases, as can previous methods used to empirically derive estimates of availability. Additionally, potential errors in group size estimates during aerial surveys conducted in passing mode (i.e. where no time is taken to circle sightings and verify group size) have been largely ignored (Hodgson et al., 2017). We have presented a method of using small, vessel-launched, multi-rotor UAVs to collect aerial video of small cetaceans to characterize availability and estimate group size error. Our approach was characterized by an aerial perspective similar to that of an aerial survey observer, the collection of high-resolution video to provide a permanent record of observations, and modest daily running costs associated with small vessel operations (Appendix S5). Here, we: (i) appraise our approach relative to existing methods; (ii) compare estimates of availability for humpback dolphins to those presented for other dolphin species; and, (iii) discuss what our data reveal about sample size and subsequent recommendations for future data collection.

Methodology appraisal

An advantage of our method was the aerial perspective from which observations were made. Alternatives, such as estimating availability based on observations from a vessel, land or dive-tag data, rely on assumptions about the visibility of animals at different stages of the dive-surface cycle (e.g. inter-breath intervals) or water depths. The effects of these assumptions were exemplified by observations of franciscana dolphins *Pontoporia blainvillei*, Sucunza et al.'s (2018) helicopter observations providing a mean surfacing interval 13 times longer than previously recorded from a surface platform. Application of the updated availability estimate resulted in a 38% lower abundance estimate for one management area.

Aerial video also provides a permanent record of dive-surface behaviour which can be re-evaluated and manipulated as often as required and for various purposes. Recording availability in situ is typically limited to recording the times at which any individual from a focal group is visible. By contrast, the review of video data allowed us to accurately record the times at which

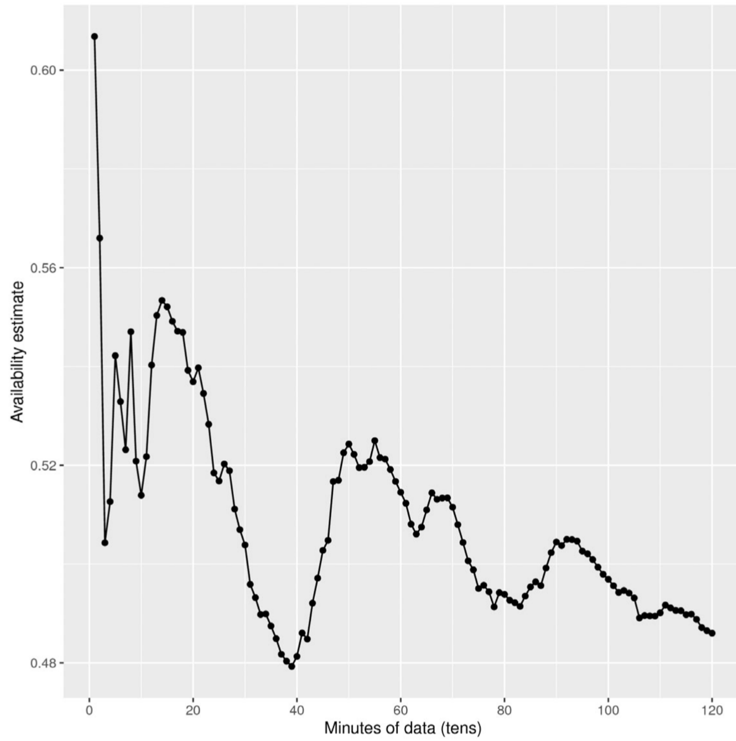

Figure 6. Mean non-instantaneous availability estimates, assuming an example 7 sec observation window, with increasing quantities of data in approx. 10-min increments (rounded to nearest complete dive-surface cycle). Data are pooled across all 56 follows and in chronological order as collected, such that each availability estimate reflects what the final estimate would be had we ceased data collection at that point.

different numbers of individuals were visible and address group size estimation error. We show that aerial surveys of humpback dolphins in passing mode would likely underestimate group size, biasing population estimates downwards. Similarly, Hodgson et al. (2017) suggested that this largely overlooked source of bias could be quite common in abundance estimation from aerial survey data.

Our study was markedly less time-efficient than two examples in which helicopters were used (Bilgmann et al., 2018; Sucunza et al., 2018), likely because helicopters minimize search times. However, the cost and endurance (max 3–4 h flight time) of helicopters limits the field time and, consequently, the volume of data collected. We collected and analysed 1230 min of data, c. 3–6 times that collected in studies using helicopters (Bilgmann et al., 2008; Slooten et al., 2004, 2006; Sucunza et al., 2018). While there was a financial cost to spending more time in the field, our vessel-launched UAV approach was more cost-effective than example helicopter studies under almost all scenarios examined (Appendix S4). Furthermore, with equipment and training in-hand, the cost per minute of UAV data collection in ongoing studies would be lower still. As part of a broader study on detection probability, Hodgson et al. (2017) used a large, fixed-wing UAV to estimate the availability of humpback whales (*Megaptera novaeangliae*); this is unlikely to be cost-effective for a dedicated availability study, but may prove an effective option where the UAV is also being used for the survey itself.

We found it challenging to track dolphins without a visual reference and this was addressed by positioning the vessel in the corner of the video frame, which meant the vessel remained within hundreds of metres of the focal group. Short-term behavioural responses by cetaceans to small vessel approaches are widely documented (review in Senigaglia et al., 2016), and there is potential that the presence of the vessel had some influence on the behaviour of the dolphins. While we did not specifically attempt to assess behavioural response of animals to the vessel, as was done for the UAVs, we do not consider that the vessel's presence had an influence on the behaviour of these humpback dolphins to the extent of biasing availability estimates (i.e. through strong or persistent changes in behaviour). In addition to the study population being habituated to small vessel traffic, the experienced crew ensured that the vessel was operated to avoid actions likely to elicit behavioural responses, such as close approaches, leapfrogging, direct pursuit or rapid changes in speed or heading (e.g. Nowacek et al., 2001; Williams et al., 2002). Any individuals which may have exhibited strong avoidance or increased changes in directions were unlikely to have contributed to the data due to the difficulty in following these animals (noting that data collection was unsuccessful on c. 20% of groups encountered, although a proportion of these were lost due to poor sighting conditions). We were able to complete focal follows of > 1 h on multiple groups and sample a variety of different behavioural states (Fig. S3.1), providing further support for the data being representative of undisturbed behaviour. Nonetheless, this potential source of bias is an important consideration for future applications of our method. For more cryptic/evasive species and populations, maintaining such proximity without influencing dolphin behaviour may not be possible.

The initial data review took almost 12 h for every h of raw video, plus time to check and prepare for statistical analysis. The detailed video analysis may not be necessary if, for example, estimating group size error was not an objective. Simply recording when any member of the group is visible, rather than every time the number visible changes, would provide the same availability estimates. While using overlapping UAVs often allowed us to conduct long, uninterrupted focal follows, it (a) placed additional demands on the field team; (b) occasionally caused us to lose animals; (c) duplicated effort in the initial video review and, (d) was time-consuming to merge the data. Instead, we recommend using one UAV at a time and aiming to capture as many complete dive-surface cycles as possible during each flight.

Availability estimates for humpback dolphins

The mean instantaneous availability estimate we obtained for humpback dolphins (0.39 (CV = 0.07)) is within the range of those obtained for some other shallow-water, small cetaceans. For example, observations from helicopters produced group-based availability estimates of 0.46 (CV = 0.04) and 0.56 (CV = 0.06) for Hector's (*Cephalorhynchus hectori*) and Maui's (*C.h. maui*) dolphins, respectively (Slooten et al., 2004, 2006), and 0.36 (±0.22 SE) for franciscanas (0.39 assuming a 6 s observation window) (Sucunza et al., 2018). By contrast, group-based availability was estimated at 0.94 (±0.02 SE) for common dolphins *Delphinus delphis* off South Australia (helicopter follows; Bilgmann et al., 2018) and 0.77 (±0.10 SE) for bottlenose dolphins *Tursiops truncatus* in the north-west Mediterranean (vessel follows; Forcada et al., 2004). The relatively high estimates for the latter two species may be a result of differences in behaviour and habitats (e.g. larger group sizes and pelagic, low turbidity environments), or may reflect methodological differences. For example, during vessel-based observations, Forcada et al. (2004) assumed groups were available for detection throughout surfacing periods, defined as any period with one or more dolphins surfacing less than 30 sec apart, whereas other studies (including ours) considered animals not visible for any duration to be unavailable. Methods that rely on assumptions about availability can result in extremely inaccurate population estimates, emphasizing the need to estimate availability from the same perspective as the survey itself.

Within our data, there was some evidence of a pattern of increasing availability with group size, as has been reported for other cetaceans (Hodgson et al., 2017; Mobley et al., 2001; Sucunza et al., 2018). When estimating group-based availability (where the group is available if any individual is visible) for animals which surface asynchronously, it is unsurprising that the group is visible for a greater proportion of time when more individuals are present. Given this effect, it is important that availability data are derived from focal groups that are representative of the population to which it will be applied. While we aimed to collect data on a range of group sizes, we were potentially biased towards groups that were easier to locate and follow (e.g. larger and/or resting or socializing groups) than others (e.g. individual foragers). It is perhaps telling that, of the 10 sightings where animals were lost before usable footage was obtained, six were singletons and four were pairs. This bias could be addressed by producing availability estimates for different categories of group composition; for example, ≤ 4 individuals and > 4 individuals (*cf.* Sucunza et al., 2018), or groups with and without calves (*cf.* Hodgson et al., 2017). An alternative is to produce a mean availability estimate weighted according to the proportions of each group composition seen during the aerial survey, as applied by Mobley et al. (2001) to humpback whales, which could offer the least biased way to implement our data for humpback dolphins.

Investigating sample size

Availability estimates for humpback dolphin focal groups were highly variable, as has been reported for cetaceans in previous studies (e.g. Hodgson et al., 2017; Mobley et al., 2001; Sucunza et al., 2018). To minimize potential bias in availability estimates, it is therefore essential that a sufficient quantity of data is collected (both length and number of follows) and that data are representative of the study population. Sub-sampling simulations *within* a follow suggested that the minimum length of a focal follow should be nine dive-surface cycles; above this value, availability stabilizes, and further data give diminishing returns. While it is challenging to monitor in the field, total number of dive-surface cycles is considered a more widely applicable metric than total follow time because of the variability of dive-surface cycle durations. Sub-sampling simulations *across* the total 56 follows showed, as expected, increased precision in the overall availability estimate as more follows were included. However, from 30 to 35 follows, the rate of increase in precision with an increasing number of follows plateaued, suggesting this is an appropriate minimum number of follows in order to achieve a reliable availability estimate.

These recommended minimum sample sizes are based on data for north-western Australian humpback dolphins, and their applicability to other species or regions needs further investigation. Thirty-five focal follows of dolphins in shallow water, each covering nine dive-surface cycles, could represent c. 800 min of data. We acknowledge that collecting such a volume of data may be beyond the resources of some studies; in such cases, prior information on the group size distribution from the aerial survey would be valuable to focus efforts on collecting availability data relevant to the study population.

Concluding Remarks

We present a vessel-launched UAV method for collecting availability data and investigating error in group size estimates which was effective for coastal dolphins in north-western Australia. Our method provides an alternative to existing approaches that: offers a more appropriate aerial perspective and fewer assumptions than land-, boat-based or telemetry-derived observations of dive-surface behaviour, and, in most circumstances, is considerably less costly than using a helicopter. The method could be more broadly applied to cetaceans in a variety of habitats.

As a multiplier, availability corrections have a marked influence on abundance estimates and, therefore, subsequent decisions regarding conservation and management. However, estimates of availability vary widely between focal groups within studies, and little attention has been given to examining the relationship between sample size and potential bias. Our results suggest that accurately characterizing the availability process likely requires more data than is routinely collected. We encourage those planning future aerial surveys to ensure sufficient resources are allocated to collecting the necessary data to produce reliable availability estimates.

Acknowledgements

The Western Australian Department of Biodiversity Conservation and Attractions (DBCA) and the Sea World Research and Rescue Foundation Inc. funded this research. We are grateful for the generous in-kind support of Dan Pederson and Pilbara Ports Authority (Dampier), along with Hampton Harbour Boat & Sailing Club. AB received support from Hartley Anderson Ltd. during analyses and write-up. We were afforded favourable training rates by Global Unmanned Systems Pty Ltd, and vessel and vehicle use rates from Shark Bay Dolphin Research. Drs Delphine Chabanne and Julian Tyne (Murdoch University) were instrumental in data collection, and we are grateful to Ryan Douglas and Corrine Douglas (DBCA) for performing the initial video review. We also thank Holly Raudino and Kelly Waples for helping to facilitate this study and their constructive comments on earlier drafts of this manuscript, along with those of two anonymous reviewers. This research was conducted under permits from DBCA, with approval from Murdoch University Animal Ethics Committee, notifications to the Civil Aviation Safety Authority and the Department of Agriculture and Food, and following consultation with Murujuga Aboriginal Corporation. We acknowledge the traditional custodians of the country on which this research was carried out.

Authors' contributions

AMB, SJA and *AJH* conceived the ideas and designed methodology; *AMB* and *SJA* collected the data; *AMB* and *NK* analysed the data; *AMB, SJA* and *AJH* led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Data availability statement

This article's data and R code are available from the Dryad Digital Repository: doi:10.5061/dryad.qbzkh18mq.

References

Allen, S.J., Cagnazzi, D.D., Hodgson, A.J., Loneragan, N.R. & Bejder, L. (2012) Tropical inshore dolphins of north-Western Australia: unknown populations in a rapidly changing region. *Pacific Conservation Biology*, **18**(1), 56–63.

Allen, S.J., Pollock, K.H., Bouchet, P.J., Kobryn, H.T., McElligott, D.B., Nicholson, K.E. et al. (2017) Preliminary estimates of the abundance and fidelity of dolphins associating with a demersal trawl fishery. *Scientific Reports*, **7** (1), 1–11.

ASCOBANS. (2015) Project Report: preparations for SCANS-III. 22nd ASCOBANS Advisory Committee Meeting, The Hague, Netherlands, 29 September–1 October 2015. AC22/ Inf.8.1.b.

Barlow, J., Oliver, C.W., Jackson, T.D. & Taylor, B.L. (1988) Harbor porpoise, *Phocoena phocoena*, abundance estimation for California, Oregon, and Washington: II. Aerial surveys. *Fishery Bulletin*, **86**(3), 433–444.

Barnas, A.F., Chabot, D., Hodgson, A.J., Johnston, D.W., Bird, D.M. & Ellis-Felege, S.N. (2020) A standardized protocol for reporting methods when using drones for wildlife research. *Journal of Unmanned Vehicle Systems*, **8**(2), 89–98.

Bilgmann, K., Möller, L.M., Harcourt, R.G., Gales, R. & Beheregaray, L.B. (2008) Common dolphins subject to fisheries impacts in Southern Australia are genetically differentiated: implications for conservation. *Animal Conservation*, **11**(6), 518–528.

Bilgmann, K., Parra, G.J., Holmes, L., Peters, K.J., Jonsen, I.D. & Möller, L.M. (2019) Abundance estimates and habitat preferences of bottlenose dolphins reveal the importance of two gulfs in South Australia. *Scientific Reports*, **9**(1), 1–14.

Bilgmann, K., Parra, G.J. & Möller, L.M. (2018) Occurrence, distribution and abundance of cetaceans off the western Eyre Peninsula in the Great Australian Bight. *Deep-Sea Research Part II: Topical Studies in Oceanography*, **157–158**, 134–145.

Borchers, D., Zucchini, W., Heide-Jørgensen, M.P., Cañadas, A. & Langrock, R. (2013) Using Hidden Markov Models to deal with availability bias on line transect surveys. *Biometrics*, **69**(3), 703–713.

Borchers, D.L., Buckland, S.T. & Zucchini, W. (2002) *Estimating animal abundance: closed populations*. Gateshead, UK: Springer-Verlag.

Brown, A., Bejder, L., Cagnazzi, D., Parra, G.J. & Allen, S.J. (2012) The North West Cape, Western Australia: a potential hotspot for Indo-Pacific humpback dolphins *Sousa chinensis*? *Pacific Conservation Biology*, **18**(4), 240–246.

Choquet, D., L'ecuyer, P. & Léger, C. (1999) Bootstrap confidence intervals for ratios of expectations. *ACM Transactions on Modeling and Computer Simulation*, **9**(4), 326–348.

Christie, K.S., Gilbert, S.L., Brown, C.L., Hatfield, M. & Hanson, L. (2016) Unmanned aircraft systems in wildlife research: current and future applications of a transformative technology. *Frontiers in Ecology and the Environment*, **14**(5), 241–251.

Forcada, J., Gazo, M., Aguilar, A., Gonzalvo, J. & Fernández-Contreras, M. (2004) Bottlenose dolphin abundance in the NW Mediterranean: addressing heterogeneity in distribution. *Marine Ecology Progress Series*, **275**, 275–287.

Grech, A., Sheppard, J. & Marsh, H. (2011) Informing species conservation at multiple scales using data collected for marine mammal stock assessments. *PLoS One*, **6**(3), e17993.

Hagihara, R., Jones, R.E., Sobtzick, S., Cleguer, C., Garrigue, C. & Marsh, H. (2018) Compensating for geographic variation in detection probability with water depth improves abundance estimates of coastal marine megafauna. *PLoS One*, **13**(1), 1–15.

Hammond, P.S., Macleod, K., Berggren, P., Borchers, D.L., Burt, L., Cañadas, A. et al. (2013) Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. *Biological Conservation*, **164**, 107–122.

Hanf, D.M., Hunt, T. & Parra, G.J. (2016) Chapter Eight - Humpback dolphins of Western Australia: a review of current knowledge and recommendations for future management. *Advances in Marine Biology*, **73**, 193–218.

Hedley, S.L., Bannister, J.L. & Dunlop, R.A. (2011) Abundance estimates of Southern Hemisphere Breeding Stock "D" humpback whales from aerial and land-based surveys off Shark Bay, Western Australia, 2008. *Journal of Cetacean Research and Management (Humpback Whale Special Issue)*, **3**, 209–221.

Heide-Jørgensen, M.P., Borchers, D.L., Witting, L., Laidre, K.L., Simon, M.J., Rosing-Asvid, A. et al. (2008) Estimates of large whale abundance in West Greenland waters from an aerial survey in 2005. *Journal of Cetacean Research and Management*, **10**(2), 119–129.

Heide-Jørgensen, M.P., Laidre, K.L., Hansen, R.G., Burt, M.L., Simon, M., Borchers, D.L. et al. (2012) Rate of increase and current abundance of humpback whales in West Greenland. *Journal of Cetacean Research and Management*, **12**, 1–14.

Hodgson, A., Kelly, N. & Peel, D. (2013) Unmanned aerial vehicles (UAVs) for surveying Marine Fauna: a dugong case study. *PLoS One*, **8**(11), e79556.

Hodgson, A., Peel, D. & Kelly, N. (2017) Unmanned aerial vehicles for surveying marine fauna: assessing detection probability. *Ecological Applications*, **27**(4), 1253–1267.

Hunt, T., Bejder, L., Allen, S., Rankin, R., Hanf, D. & Parra, G. (2017) Demographic characteristics of Australian humpback dolphins reveal important habitat toward the southwestern limit of their range. *Endangered Species Research*, **32**, 71–88.

Laake, J.L., Calambokidis, J., Osmek, S.D. & Rugh, D.J. (1997) Probability of detecting harbor porpoise from aerial surveys: estimating g(0). *Journal of Wildlife Management*, **61**(1), 63–75.

Marsh, H. & Sinclair, D.F. (1989) Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. *Journal of Wildlife Management*, **53**(4), 1017–1024.

McLaren, I. (1961) Methods of determining the numbers and availability of ringed seals in the eastern Canadian Arctic. *Arctic*, **14**, 162–175.

Mobley, J., Spitz, S. & Grotefendt, R. (2001) *Abundance of humpback whales in Hawaiian waters: results of 1993–2000 aerial surveys. Report for the National Oceanic and Atmospheric Administration*. Hawaii, USA: U.S. Department of Commerce and the Department of Land and Natural Resources, p. 16.

Nowacek, D.P., Christiansen, F., Bejder, L., Goldbogen, J.A. & Friedlaender, A.S. (2016) Studying cetacean behaviour: new technological approaches and conservation applications. *Animal Behaviour*, **120**, 235–244.

Nowacek, S.M., Wells, R.S. & Solow, A.R. (2001). Short-term effects of boat traffic on bottlenose dolphins, *Tursiops truncatus*, in Sarasota Bay, Florida. *Marine Mammal Science*, **17**(4), 673–688.

Pollock, K.H., Marsh, H.D., Lawler, I.R. & Alldredge, M.W. (2006) Estimating animal abundance in heterogeneous environments: an application to aerial surveys for dugongs. *Journal of Wildlife Management*, **70**(1), 255–262.

R Core Team. (2017) *R: a language and environment for statistical computing. Version 3.4.0*. Vienna, Austria: R Foundation for Statistical Computing.

Senigaglia, V., Christiansen, F., Bejder, L., Gendron, D., Lundquist, D., Noren, D.P. et al. (2016) Meta-analyses of whale-watching impact studies: comparisons of cetacean responses to disturbance. *Marine Ecology Progress Series*, **542**, 251–263.

Seymour, A.C., Dale, J., Hammill, M., Halpin, P.N. & Johnston, D.W. (2017) Automated detection and enumeration of marine wildlife using unmanned aircraft systems (UAS) and thermal imagery. *Scientific Reports*, **7**, 45127.

Slooten, E., Dawson, S., Rayment, W. & Childerhouse, S. (2006) A new abundance estimate for Maui's dolphin: what does it mean for managing this critically endangered species? *Biological Conservation*, **128**(4), 576–581.

Slooten, E., Dawson, S.M. & Rayment, W.J. (2004) Aerial surveys for coastal dolphins: abundance of Hector's dolphins off the South Island west coast, New Zealand. *Marine Mammal Science*, **20**(3), 477–490.

Smolker, R.A., Richards, A.F., Connor, R.C. & Pepper, J.W. (1992) Sex differences in patterns of association among Indian Ocean bottlenose dolphins. *Behaviour*, **123**, 38–69.

Sucunza, F., Danilewicz, D., Cremer, M., Andriolo, A. & Zerbini, A.N. (2018) Refining estimates of availability bias to improve assessments of the conservation status of an endangered dolphin. *PLoS One*, **13**(3), 1–16.

Ver Hoef, J.M. (2012) Who invented the delta method? *The American Statistician*, **66**(2), 124–127.

Weinrich, M.T., Lambertsen, R.H., Baker, C.S., Schilling, M.R. & Belt, C.R. (1991) Behavioural responses of humpback whales (*Megaptera novaeangliae*) in the southern gulf of Maine to biopsy sampling. *Reports of the International Whaling Commission Special Issue*, **13**, 91–97.

Williams, R., Bain, D.E., Ford, J.K.B. & Trites, A.W. (2002) Behavioural responses of male killer whales to a 'leapfrogging' vessel. *Journal of Cetacean Research and Management*, **4**(3), 305–310.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. UAV reporting protocol.
Appendix S2. Availability parameters and estimates for individual follows.
Appendix S3. Assessment of behavioural state and availability estimates for follows of consistent behavioural state.
Appendix S4. Plots illustrating non-instantaneous availability estimates (for an example 7 sec observation window) and their corresponding environmental characteristics, video collection attributes and time of day.
Appendix S5. Operational costs and efficiency of this study, with a comparison to other studies of availability using aerial platforms.

