Boyd Charlotte (Orcid ID: 0000-0003-4095-9390)
Shelden Kim (Orcid ID: 0000-0002-9831-0933)

Bayesian estimation of group sizes for a coastal cetacean using aerial survey data

CHARLOTTE BOYD,[1] School of Aquatic and Fishery Sciences, 1122 NE Boat Street, University of Washington, Seattle, Washington 98195, U.S.A. and Marine Mammal Laboratory, NOAA Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, Washington 98115, U.S.A.; **RODERICK C. HOBBS**, Retired from Marine Mammal Laboratory, NOAA Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, Washington 98115, U.S.A.; **ANDRÉ E. PUNT**, School of Aquatic and Fishery Sciences, 1122 NE Boat Street, University of Washington, Seattle, Washington 98195, U.S.A.; **KIM E. W. SHELDEN**, **CHRISTY L. SIMS**, **PAUL R. WADE**, Marine Mammal Laboratory, NOAA Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, Washington 98115, U.S.A.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mms.12592

ABSTRACT

Many small cetacean, sirenian, and pinniped species aggregate in groups of large or variable size. Accurate estimation of group sizes is essential for estimating the abundance and distribution of these species, but is challenging as individuals are highly mobile and only partially visible. We developed a Bayesian approach for estimating group sizes using wide-angle aerial photographic or video imagery. Our approach accounts for both availability and perception bias, including a new method (analogous to distance sampling) for estimating perception bias due to small image size in wide-angle images. We demonstrate our approach through an application to aerial survey data for an endangered population of beluga whales (*Delphinapterus leucas*) in Cook Inlet, Alaska. Our results strengthen understanding of variation in group size estimates and allow for probabilistic statements about the size of detected groups. Aerial surveys are a standard tool for estimating the abundance and distribution of various marine mammal species. The role of aerial photographic and video data in wildlife assessment is expected to increase substantially

with the widespread uptake of unmanned aerial vehicle technology. Key aspects of our approach are relevant to group size estimation for a broad range of marine mammal, seabird, other waterfowl, and terrestrial ungulate species.

Key words: abundance, aerial photograph, aerial video, availability bias, beluga whale, *Delphinapterus leucas,* distribution, perception bias, UAV, visibility bias.

Groups are a fundamental unit for many social species (Royle 2008). For species that aggregate in groups of large or highly variable size, including many species of small cetaceans, sirenians, pinnipeds, seabirds, and other waterfowl, accurate estimation of group sizes is essential for accurate estimation of abundance and density distribution patterns based on survey data. Estimates of group size and group size dynamics are also important for understanding various aspects of foraging, reproductive, and behavioral ecology, and responses to anthropogenic disturbance (*e.g.*, Elgar 1989, Baird and Dill 1996, Giraldeau and Caraco 2000, Heithaus and Dill 2002, Gowans *et al*. 2008, Orbach *et al.* 2014, Koper *et al.* 2016). Yet, estimating group size can be extremely challenging, especially for species that are highly mobile and often submerged or only partially visible (*e.g.*, Gilpatrick 1993, Clement *et al.* 2017).

Two major sources of visibility bias can lead to underestimation of the size of detected groups (Marsh and Sinclair 1989, Laake and Borchers 2004): availability bias occurs when animals are missed because they are underwater, underground, or concealed by vegetation or other animals;

perception bias occurs when available animals are missed because they are not seen by observers due to factors such as distance, size, or coloration.

Availability bias is typically estimated by comparing the length of time that a location is observed with the proportion of time that animals are expected to be unavailable, estimated from ancillary behavior data (*e.g.*, Williams *et al.* 2017). Diving animals, for example, may spend a large proportion of time underwater—information on diving patterns may be collected from a second survey platform (*e.g.*, Laake *et al.* 1997, Hiby and Lovell 1998, Sucunza *et al.* 2018) or using time-depth recorders (*e.g.*, Pollock *et al.* 2006, Thomson *et al.* 2012).

Substantial research effort has been dedicated to estimating and correcting perception bias in the group detection process, using distance sampling and mark-recapture methods (*e.g.*, Alpizar-Jara and Pollock 1996, Buckland *et al.* 2001, Marques and Buckland 2003, Laake *et al.* 2008, Royle 2008, Barlow 2015). However, much less attention has been focused on correcting for perception bias when estimating group sizes (Clement *et al*. 2017), despite the importance of accurate group

size estimates for estimating abundance and distribution patterns in species that occur in large or highly variable groups (see Hobbs *et al.* 2000, Clement *et al*. 2017, Gerrodette *et al.* 2018 for exceptions).

Our research objective was to develop a modeling approach for estimating group sizes from aerial survey data, including aerial photographic or video data. Aerial surveys are a standard tool for estimating the abundance and distribution of various marine and terrestrial vertebrate species, providing essential information for conservation, management, and ecological understanding (Caughley 1979, Garner *et al.* 1999, Gowans *et al.* 2008). The role of aerial photographic and video data in wildlife assessment, monitoring, and research is expected to increase substantially with the widespread uptake of unmanned aerial vehicle (UAV) technology (Hodgson *et al.* 2016).

Our approach builds on previous methods for estimating availability bias developed by McLaren (1961), Laake *et al*. (1997), Hobbs *et al*. (2000), and others, but provides a new way to estimate perception bias due to small image size in aerial photographs or video that is analogous to distance sampling.

Variation in image size may reflect true variation in the physical size of individuals and/or apparent variation attributable to the exaggerated perspective in the wide-angle images used to capture large or dispersed groups within a single frame (*i.e.*, distant objects appear much smaller in wide-angle images than in standard 50 mm images). Group sizes will be underestimated if the analysis does not account for individuals that are not detected because of their small image size. Our focus was on small coastal cetaceans, specifically beluga whales (*Delphinapterus leucas*; Pallas, 1776), but our approach is broadly applicable to aerial survey data and UAV imagery for marine and terrestrial wildlife.

Cook Inlet beluga whales are a small population, resident year-round in Cook Inlet, Alaska (Fig. 1). The population is geographically and genetically isolated from the nearest neighboring populations in Bristol Bay and Yakutat Bay, Alaska (O'Corry-Crowe *et al.* 1997, 2015; Laidre *et al.* 2000). The best available estimates of historical abundance, based on aerial surveys in 1979 and 1991, indicate a population of more than a thousand (Shelden *et al.* 2015). This population suffered severe

declines due to unregulated and unsustainable hunting in the 1990s. The hunt was reduced to 0–2 individuals per year from 1999, following a comanagement agreement between the U.S. National Marine Fisheries Service (NMFS) and Alaska Native organizations (Mahoney and Shelden 2000), and there has been no documented hunt since 2005. The population was expected to start increasing within a few years of the reduction in hunting pressure, but there has been limited evidence of recovery to date (Hobbs *et al.* 2015). Consequently, the Cook Inlet population of beluga whales was listed as Critically Endangered on the IUCN Red List of Threatened Species in 2006 (Lowry *et al.* 2012) and endangered under the U.S. Endangered Species Act (ESA) in 2008 (U.S. Federal Register 2008).

Group size estimates based on aerial survey data provide the foundation for estimates of abundance and distribution patterns for Cook Inlet beluga whales (*e.g.*, Goetz *et al.* 2012, Hobbs *et al*. 2015). These estimates, in turn, play a key role in research and management decision-making. The status of the population under the ESA, identification of critical habitat, and management of anthropogenic activities that might jeopardize

the population (including the future of the subsistence hunt) all depend on estimates of abundance and distribution patterns.

METHODS

Aerial Survey

Survey design—Since 1994, NMFS has conducted dedicated annual or biennial aerial surveys of the Cook Inlet beluga whale population (Hobbs *et al*. 2015). We focused our analysis on data collected from 2004 through 2016, as several changes were made to the survey design and data processing in 2004.

The survey was conducted in early June, when the population is usually highly aggregated in coastal areas near river mouths, foraging on anadromous salmon and eulachon (Moore *et al.* 2000). Group sizes are highly variable at this time, ranging from a few individuals to several hundred, with the majority of the population sometimes aggregated in a single group.

In each survey year, the aerial survey encompassed the entire known range of Cook Inlet beluga whales. The survey was partitioned into two sectors: the Upper Inlet (Fig. 1) and the Lower Inlet, but no belugas have been sighted in the Lower Inlet during these surveys since 2001. The objective for the Upper

Inlet survey was to count the entire population on each of five or more survey days, by locating and estimating the size of all groups. (Note that this differs from a conventional line-transect or strip-transect survey, in which the objective is to detect and count groups within a representative sampling area.) The survey design was strategic, rather than systematic, combining a comprehensive survey of coastal areas on each day, with relatively sparse north-south or sawtooth transects of offshore areas. The survey was also adaptive, with a gradual reduction in survey effort in areas where belugas were not found on previous days. Consequently, the pattern of survey effort varied among survey days and years.

The timing of surveys in relation to the tide is important. Cook Inlet is characterized by an extreme tidal range with extensive mudflats exposed at low tide in areas of the Upper Inlet such as the Susitna Delta (Moore *et al*. 2000). The ability to detect and count all beluga groups increases substantially with the reduction in effective survey area at low tide (Rugh *et al.* 2000). Each survey day was therefore scheduled, to the extent possible, so that the Susitna Delta was surveyed at low

tide, and other areas on the falling tide.

Aircraft and observer configuration—Aerial surveys were flown in twin-engine aircraft with high wings and bubble windows at 244 m (800 ft) and 185 km/h (100 kn). The aircraft model was changed periodically.[2] On each Upper Inlet survey day, coastal areas were surveyed in a clockwise direction, using an inclinometer to maintain the shoreline 10° below the horizon (*i.e.*, at a distance of 1.4 km from the shoreline).

Two observers were positioned on the shoreward-side of the aircraft behind the pilot, with one observer and a data recorder on the other side of the aircraft behind the copilot. In all survey aircraft, the two forward observers had a bubble window. In 2004–2006 and 2008–2010, the shoreward-rear observer also had a bubble window and similar field-of-view to the forward observer. In other years, the shoreward-rear observer had a flat window and more limited field-of-view. The data recorder used custom software on a laptop computer to record encounter data, viewing conditions, changes in effort, and comments. All data entries included date, time, and location information from a portable GPS system connected to the laptop.

Count protocol—Belugas are usually found in dense clusters constituting distinct groups (Rugh *et al*. 2000). When whales were found in looser aggregations, groups were delineated for convenience of counting (Table S1). On some occasions, groups split or merged during the counting process. All such groups were included in the analysis, but data collected either before or after a group split or merger were excluded from the analysis to avoid double-counting.

When a beluga group was detected, the survey aircraft closed on the group and a series of counting passes was initiated based on an oval race-track pattern around the group (Fig. S1) and standardized counting protocols. On each pass, counts were made along the long axis of the oval, with two observers and a videographer on the same side of the aircraft. For medium-large groups (*e.g*., >20 individuals), around eight counting passes were typically implemented, with two observers estimating group size on each pass (*i.e*., up to 16 counts total). After several counting passes, observers switched roles with the recorder and videographer, so that a large group could be counted by four independent observers. Observers

independently rated each of their counts as excellent, good, fair, poor, or unusable depending on viewing conditions. Only counts rated good or excellent were used in this analysis. Counts were not shared among observers until the end of the survey season to maintain the independence of each observer's counting process.

Video data were also collected for medium-large groups, using two parallel-mounted identical video cameras held in fixed position on a hand-held board. Video cameras were upgraded periodically, with a switch to high-definition video in 2011.[3] The "zoom" video camera was set to maximum zoom to ensure that all belugas within the narrow frame of view were detected. In contrast, the "wide-angle" video camera was set to a wide angle designed to ensure that the entire group was captured within the video frame. The specific focal length varied depending on the group size and spread and the aircraft altitude and distance. During video analysis, each video clip was rated as excellent, good, fair, poor, or unusable. Only video clips rated good or excellent were used in this study.

Digital video data were analyzed using a customized

computer program, “Beluga Dots,” that enabled analysts to mark and track individual whales in video clips and measure their maximum size (in terms of pixels) and velocity (in terms of lines-of-resolution moved per unit of time). Size measurements were generally made by two analysts independently, then averaged. The program also allowed analysts to identify individuals detected in matched wide-angle and zoom video clips (*i.e.*, clips recorded simultaneously at different magnifications by the paired zoom and wide-angle cameras) and take image size measurements of the same individual in wide-angle and zoom video clips at exactly the same point in time.

Modeling Approach

Group sizes for Cook Inlet belugas have previously been estimated using point estimation methods to obtain a series of correction factors to correct video and observer count data for various types of visibility bias. Correction factors were then applied sequentially to calculate an average estimated group size for each group (Hobbs *et al*. 2000, 2015).

We developed a Bayesian approach to group size estimation. Bayesian methods are better suited for multistep analyses, in

which estimates from one step are used as inputs in a subsequent step, because uncertainty is automatically propagated from one step to another. Our approach was designed to address the same four types of bias as previous methods (Hobbs *et al*. 2000, 2015): (1) availability bias due to diving behavior (individuals unavailable because submerged, hereafter "availability bias"); (2) availability bias due to proximity (individuals unavailable because concealed by another animal, hereafter "proximity bias"); (3) perception bias (individuals not detected because of small image size) in video data; and (4) individual observer bias (*i.e*., the tendency for individual observers to under- or over-count whales) in visual observer data.

Model Overview

Ecological processes—The model is configured to estimate independent group sizes, sampled as follows:

$$\hat{n}_{g,d} \sim \text{Poisson}(\hat{\lambda}_{g,d}) \quad (1a)$$

where $\hat{n}$ is the true unobserved (or latent) group size, g is an index for group, and d is an index for day, and $\hat{\lambda}_{g,d}$ conveys a

vague prior on the expected size of each group. The sum of group sizes (or total group size) for all detected groups on each day, $\hat{N}_d$, is calculated as follows:

$$\hat{N}_d = \sum_{g=1}^{G_d} n_{g,d} \quad (1b)$$

where G_d is the total number of groups detected on each survey day.

Observation processes—The observation processes are summarized by two likelihood equations (Eq. 2, 3). The likelihood for the count data in each wide-angle video clip was estimated using a standard *N*-mixture model structure (Royle 2004), in which group size is a latent variable that can be estimated from a series of imperfect counts. A basic assumption underpinning the binomial distribution in this model structure is that individuals cannot be over-counted. (An advantage of video over still photography is that object movement patterns can reduce the potential for false positives. In this case study, for example, belugas can be readily distinguished from whitecaps or birds by their distinctive movement patterns.)

Here, the likelihood of the number of individuals counted in a wide-angle video clip, *nv*, depends on the group size, $\hat{n}$, and the probability that individuals are available at the surface, *pa* (availability bias), the probability that they are not concealed, 1 – *pc* (proximity bias), and the average detection probability for available individuals given the image size distribution in the video clip, *pd* (perception bias):

$$nv_{p,g,d} \sim \text{binomial}\left[pa_{p,g,d} * (1 - pc_{p,g,d}) * pd_{p,g,d}\hat{n}_{g,d}\right] \quad (2)$$

where *p* is an index for video or counting pass (see further model details below).

Video data were not available for all groups and passes, so group size estimation for some groups depends on counts by visual observers. Even experienced observers may have inherent biases in how they count—most observers tend to underestimate group sizes, but some tend to overestimate (*cf*., Gerrodette *et al*. 2018). The potential for over-estimation implies that an *N*-mixture model with binomial likelihood cannot be used here. The likelihood of the observer count data in each pass was therefore assumed to follow a negative binomial distribution:

$$no_{i,p,g,d} \sim \text{negative.binomial}(\delta_i * \hat{n}_{g,d}, \vartheta_i) \quad (3)$$

where *no* is the number of whales counted by an individual observer during a counting pass; *i* is an index for individual observer; δ_i is a correction factor for the tendency of each individual observer to over- or undercount group sizes; and ϑ_i captures the dispersion of each individual observer's counts around the expected value. The model gains information on individual observer bias (*i.e.*, δ_i and ϑ_i) from groups and passes for which both video and observer counts are available (*cf.*, Gerrodette and Forcada 2005). (See Appendix S1 Section A for an alternative version of Eq. 3 to account for a flat observation window.)

Figure 2 provides a graphical summary of the model.

Model Details

Availability bias—Cook Inlet waters are extremely turbid, such that individual belugas in a group are only visible from the air when a portion of the body is above the water surface (Hobbs *et al*. 2000). The model component for estimating availability bias builds on previous work by McLaren (1961), Laake *et al*. (1997), and Hobbs *et al*. (2000). The probability of

availability, *pa*, is estimated for each pass, *p*, as follows (see Appendix S1 Section B for derivation):

$$pa_p = \frac{\bar{s}_p}{\bar{s}_p + \bar{b}_p} + \frac{w_p}{\bar{s}_p + \bar{b}_p} + \varepsilon_p \quad (4)$$

where $\varepsilon_p \sim \text{normal}\left(0, \sigma_\varepsilon^2\right)$ and $\bar{s}$ is the measured mean time spent visible at the surface; $\bar{b}$ is the unknown mean time spent invisible underwater; and *w* is the measured video clip duration. Each clip is indexed by group and day, but group and day indices have been suppressed in Equation 4 and subsequent equations for clarity. The probability of availability, *pa*, is constrained to be ≤1 in the model.

Most of the variables in Equation 4 could be measured directly from the video data for each video clip. The exception was time spent invisible underwater, $\bar{b}$, as it is not generally possible to track consecutive surfacings of individual belugas using aerial video data. Data on surface-dive intervals (*i.e.*, *s* + *b*) from a single Cook Inlet beluga radiotracked in early June by Lerczak *et al.* (2000) were therefore used to construct an

informative prior for the mean surface-dive interval (*i.e.*, $\bar{s}_p + \bar{b}_p$).

Proximity bias—In the case of belugas, an individual at the surface may also be unavailable because it surfaced close to another individual and was concealed from view in the wide-angle video. In the Cook Inlet beluga survey, proximity bias was estimated by comparing matched zoom and wide-angle video clips (Fig. 3a, b). Individuals that were detected in the zoom video but missed in the wide-angle video because it was not possible to distinguish two distinct individuals were identified. The probability that an available individual was concealed in the wide-angle video, *pc*, was then calculated as follows:

$$pc = \frac{nc}{nz} \tag{5}$$

where *nz* is the number of individuals detected in zoom video clips; and *nc* is the number of individuals that were detected in zoom video clips, but missed in matched wide-angle video clips because it was not possible to distinguish two individuals. A single probability was calculated for each survey year (i.e.,

$pc_{p,g,d}$ = pc), as this occurred rarely in the Cook Inlet beluga data set.

Perception bias—In aerial photographic or video surveys based on wide-angle imagery, accounting for perception bias due to small image size can be a major challenge. Perception bias occurs because the probability of detection declines as image size approaches zero, such that individuals with smaller image sizes (*e.g.* <4 pixels long) are less likely to be detected by analysts (Fig. 3c, d). Perception bias is thus a function of (1) the distribution of image sizes in an aerial photograph or video clip, and (2) the detectability of objects given their image size. Specifically, for each aerial photograph or video clip, the average detection probability for available individuals in a video clip, pd_p in Equation 2, depends on the probability that individuals fall into each image size class, $\pi_{k,p}$, and the probability that individuals in each size class are detected, ψ_k:

$$pd_p = \frac{\sum_{k=1}^{K} \psi_k * \pi_{k,p}}{\sum_{k=1}^{K} \pi_{k,p}} \quad (6)$$

where k is in index for size class. The denominator in Eqation 6

serves to scale the estimate of detection probability so that it would equal 1 in the case of perfect detection of individuals in all size classes.

We developed an approach for estimating perception bias (*i.e.*, detection probability as a function of image size; Fig. 4a) that is analogous to standard distance sampling methods for estimating detection probability as a function of distance (Buckland *et al*. 2001; see also Kéry and Royle 2016 Chapter 8 for a Bayesian approach).

The main difference between distance sampling and the method developed here is that distance sampling makes use of the simplifying assumption that the underlying distances of all objects (detected and not detected) are homogenously distributed, simplifying estimation of π_k. In contrast, we assume a nonhomogenous distribution of underlying image sizes, which must be estimated. This is facilitated by ancillary data on ψ_k.

Detection Function

The first step in estimating the average detection probability (Eq. 6) is to estimate the probability that individuals in each size class are detected, ψ_k. In the case of

the Cook Inlet beluga survey, information on this detection function can be gained from matched zoom/wide-angle video clips (Fig. 3c, d). It is assumed that all animals in the zoom video frame were detected, but the same individuals were either detected or not in the matched wide-angle video ($q_{j,p}$ = 1 or $q_{j,p}$ = 0 for the jth individual), depending on image size. (Individuals concealed because of their close proximity to another were excluded from this analysis.) The detection function can then be estimated from the image sizes in matched wide-angle video clips, $\hat{ss}$:

$$\text{logit}(pv_{j,p}) = \alpha_0 + \alpha_1 * \hat{ss}_{j,p} \quad (7a)$$

$$q_{j,p} \sim \text{Bernoulli}(pv_{j,p}) \quad (7b)$$

where pv is the probability of detecting the jth individual in a matched wide-angle video clip given its image size in the wide-angle video clip, and and are parameters. The image sizes of individuals that were not detected in the wide-angle video are unknown and must be estimated (see Appendix S1 Section C). The detection probability for each image size class, ψ_k (see Eq. 6), can be estimated using the parameters estimated in Eqation

7a:

$$\text{logit}(\psi_k) = \alpha_0 + \alpha_1 * sk_k \quad (7c)$$

where sk_k is the midpoint of the kth size class (see below).

Image Size Distribution

The second step in estimating the average detection probability (Eq. 6) is to estimate the distribution of image sizes for all individuals (detected and not detected) in each wide-angle video clip (i.e., $\pi_{k,p}$). For each survey year, we fitted models based on the assumption that the distribution of images sizes followed either a lognormal distribution or a zero-truncated normal distribution. (The Weibull distribution was also considered but rejected following preliminary analysis.) We used a hierarchical framework to combine information on the size distribution parameters from video clips with more data (*cf*., Schaub and Kery 2012; Appendix S1 Section D), as the number of individuals detected and measured in wide-angle video clips ranged from <10 to >100.

Based on the parameters μ_p and σ_p^2, the probability that an individual in a video clip falls into the kth size class is:

$$\pi_{k,p} = \phi\left(u_k, l_k \mid \mu_p, \sigma_p^2\right) \quad (8a)$$

where $\phi(u_k, l_k \mid \mu, \sigma^2)$ specifies the cumulative density of the lognormal or zero-truncated normal distribution with mean, μ, and variance, σ^2, between the upper limit, u_k, and lower limit, l_k, of the kth size class. (Measured sizes for detected individuals in standard wide-angle images are the average of two measurements by independent analysts for all years except 2004, so naturally fall into half-pixel size classes. Whole-pixel size classes were used for 2004.)

The mean and variance of the image size distribution (*i.e.*, μ_p and σ_p^2) cannot be estimated directly because the distribution of image sizes is only partially observed—the sizes of undetected individuals are unknown (Fig. 4a). The parameters μ_p and σ_p^2 must therefore be inferred from the image sizes of detected individuals. This is achieved using a construction analogous to binned distance sampling, with size classes replacing distance classes. The likelihood of the video counts of detected individuals in each size class, $y_{k,p}$:, is estimated

based on a multinomial distribution:

$$\boldsymbol{y}_p \sim \text{multinomial}(\boldsymbol{\phi}_p, nv_p) \quad (8b)$$

where nv_p is the total number of individuals counted in each video clip. The vector $\boldsymbol{\phi}$ gives the multinomial cell probabilities based on the expected distribution of detected individuals among the various size classes, *i.e.*, $\phi_{k,p}$ is the joint probability that an individual occurs in the kth size class and is detected:

$$\varphi_{k,p} = \psi_k * \pi_{k,p} \quad (8c)$$

Model Implementation

The posterior distributions of model parameters were approximated using Markov chain Monte Carlo (MCMC) sampling, implemented in JAGS (Plummer 2003). (See Appendix S2 for model code.) Priors were generally vague (Table S2), with the exception of the prior for the surface-dive interval, as noted above. For each survey year, two separate chains were run for 1,000,000 iterations with an initial burn-in of 500,000 and a thinning rate of 1,000 to generate 1,000 saved parameter sets. Convergence was assessed by visual inspection of trace plots for

detected group sizes ($\hat{n}_{g,d}$), total group size ($\hat{N}_d$), and key parameters (*i.e.*, the detection function parameters, α_0 and α_1, the grand mean, M, and grand variance for the size distributions, Σ^2, and the individual observer parameters, δ_i and ϑ_i), and the Gelman-Rubin diagnostic test (Gelman and Rubin 1992). Model fit was assessed using posterior predictive checks (Gelman *et al.* 2004).

Results

For the 11 survey years from 2004 to 2016, there were 5–8 Upper Inlet survey days per year. The total number of groups observed per survey year ranged from 13 (2008; 7 survey days) to 44 (2006; 7 survey days). Observers made an average of 12.1 good or excellent counts per group (range: 1–27). In addition, video data were processed for 3–7 survey days per year, and for an average of 17.8 groups per year (range: 9–31), with an average of 4.0 (range: 1–10) good or excellent wide-angle video clips per group. The total number of individuals counted in video passes per year averaged 2,396 (range: 1,495–3,476).

All models converged, as indicated by visual inspection of

trace plots, posterior predictive checks, and the Gelman-Rubin statistic (Table S3), with the exception of the model for the 2009 survey data with a zero-truncated normal distribution for image sizes. Multivariate proportional scale reduction factors (PRSF, Brooks and Gelman 1998) were <1.06 for all sets of observed population sizes ($\hat{N}_d$), and less than or equal to 1.12 for all sets of detected group sizes ($\hat{n}_{g,d}$) after removing groups for which all observer counts were 0. Higher PRSFs were generally associated with very small groups, reflecting the generic difficulties of estimating discrete variables close to the limit.

The deviance information criterion (DIC; Spiegelhalter *et al.* 2002, Gelman *et al.* 2003) indicated substantially greater support for the lognormal model for image size distributions in some years (2005–2006, 2008–2009, 2011) and the zero-truncated normal model in others (2004, 2007, 2010, 2012–2016) (Table S4). For the five years in which the shoreward-rear observer had a flat window, DIC indicated greater support for models that included an estimated correction factor for a flat window in

2007, 2011, 2012, 2016, but not 2014.

Bias Correction Factors

First, we present the results for a single group to demonstrate how the model combines data on video counts, availability bias, and perception bias to estimate a single group size. Next, we review the estimated correction factors over all groups using standardized plots to check for temporal pattern.

Variation in correction factors for a single group—For many groups, there was substantial variation in the number of individuals counted in video passes. The model is based on the assumption that group size does not change between passes, so variation in video counts must be attributable to variation in availability or perception bias. Close inspection of results for one group provides an example of how the model translates divergent video counts into a single group size estimate. In Figure 5a, the lowest layer shows the number of individuals counted in each video clip (*i.e.*, 66, 66, 103). The median estimated availability in clip 1 was substantially lower than for clip 2 or 3 (*i.e.*, 0.32, 0.50, 0.53, respectively), due, in

part, to a shorter video clip duration (*w* in Eq. 4). This variation in estimated availability accounts for a large part of the difference in video counts between clip 1 and clip 3; whereas the difference in video counts between clip 2 and clip 3 is mostly attributed to variation in perception bias.

Figure 4b shows the image size distribution of individuals detected in each of the three passes represented in Figure 5. The model estimates that the mean image size was smaller in clip 2 than in clips 1 and 3 (*i.e.*, 5.9, 4.9, 5.8, respectively), which may be attributable to shorter surface intervals in clip 2. Consequently, the model estimates a lower detection probability for available individuals in clip 2 than in clips 1 and 3 (*i.e.*, 0.85, 0.54, 0.78, respectively), as indicated by the gap between the inferred distributions of image sizes for all individuals (dotted line) and detected individuals (dashed line).

The limited remaining variation in the video counts is attributed to sampling variation (Fig. 5a). Posterior predictive checks indicate that the divergent counts in the three video clips are well-estimated by the model (Fig. 5b).

Availability bias—Figure 6a shows the median estimated correction factors for availability, proximity, and perception bias for each survey year from 2004 to 2016. Results have been standardized to facilitate comparison, based on 100 individuals counted in a hypothetical wide-angle video clip in each year. The results reveal considerable interannual variation in availability bias. The median estimate of the percentage of individuals available at the surface during video clips was 54.8% (range: 48.6%-72.3%). This median estimate is consistent with median availability calculated from prior information for each year.

Proximity bias—Estimates of proximity bias (*i.e.*, the percentage of whales at the surface that were concealed by other animals) were calculated from matched video clips (Eq. 5) and were not estimated in the model. Proximity bias was fairly minor in all years except 2014, with a median of 1.9% of available whales concealed by other animals (range: 0.0%–10.9%; or 0.0%–3.9% excluding 2014; Fig. 6a).

Perception bias—Figure 6a also indicates considerable interannual variation in estimated perception bias. In

particular, there was a reduction in median perception bias from 2011 onwards, corresponding to the introduction of high-definition video. The median estimate of perception bias indicates that, on average, 74.8% of available individuals were detected in video clips in years prior to 2011 (range: 70.3%–80.9%) and 83.3% (range: 79.6%–87.4%) thereafter. The introduction of high-definition video in 2011 increased the image size of belugas in wide-angle video. This effectively shifted the observed image size distribution to the right. Consequently, a greater proportion of belugas were detectable than in earlier years, and the distribution of observed image sizes provides more information about the shape of the underlying image size distribution, leading to a reduction in uncertainty about the parameters μ and σ (Eq. 8a). For each year, the median estimate of perception bias falls within the range of perception bias calculated from matched clips for all years (*i.e.*, the scale of estimated perception bias is consistent with the scale of perception bias calculated from observed data).

Observer bias—Figure 6b shows observer counts compared to a

standardized 100 individuals counted in a hypothetical wide-angle video clip in each year. The figure shows considerable interannual variation in the ratio of observer counts to video counts. (Note that both observer counts and video counts are data, not model estimates.) This may reflect turnover in the observer team or interannual variation in the dispersion and behavior of beluga groups that makes it easier or harder for observers to count in some years. For example, only one member of the observer team changed between 2008 and 2009; for each of the four observers who participated in both years, the observer:video count ratio was 16% to 42% lower in 2008 than in 2009, but the reasons for this remain unclear.

There was an apparent increase in the observer:video count ratio from 2011 onwards: the median observer:video count ratio was 1.41:1 (range: 0.91:1–1.83:1) prior to 2011 and 1.83:1 (range: 1.67:1–2.03:1) for 2011 onwards. The aircraft and video cameras were upgraded in 2011. The combination of larger cameras and a smaller window for the videographer required the aircraft to fly at a greater distance from groups during counting passes. The most plausible explanation for the increased observer:video

count ratio is the associated expansion in the forward observer's field-of-view relative to the wide-angle video frame.

Figure 6b also shows how the median estimated correction factor for observer bias across all observers varies among years. The median estimated correction factor ($1/\delta$) was 2.56 for 2008, when the observer:video count ratio was exceptionally low, and 1.74 in the following year. Consequently, the standardized corrected observer estimates are similar for 2008 and 2009 despite the much lower observer:video count ratio in 2008. Similarly, the observer correction factor is generally lower for 2011 onwards, following the change in video cameras and aircraft.

Group Size Distribution

Overall, group sizes for Cook Inlet belugas were highly variable, ranging from single digits to the low hundreds (Fig. S2). The number and size of groups varied substantially among years (Fig. 7), with just a few large groups in some years (*e.g.*, 2008) and a larger number of small groups in others (*e.g.*, 2007). There was also considerable variation in estimated group sizes among survey days in each survey year (*e.g.*, Fig.

8), which may reflect group splits and mergers or individuals moving between groups.

Total Group Size Estimates

Total group size estimates (*i.e.*, the sum of detected group sizes) for each survey day, N_d, is shown in Figure 9. (See Fig. S3 for a comparison of total group size estimates based on the lognormal and zero-truncated normal model for image size distributions.)

Total group size estimates vary considerably among survey days in each survey year (Fig. 9). This could be attributable to (1) uncertainty in group size estimates or (2) imperfect group detection. Bayesian methods provide useful information on uncertainty in group size estimates. In 2016, for example, the posterior distributions for the daily sums of detected group sizes overlap (Fig. 10), but none of the 1,000 saved posterior samples gives the same or similar estimates for all survey days, indicating that this variation is more likely attributable to imperfect group detection. The median difference between the highest and lowest daily estimate over the 1,000 saved posterior samples was 118 whales (range: 67–196) or 36.6% of the highest

daily estimate (range: 21.7%–48.3%).

For 2016, the median of the posterior distribution of total group size on the day with the highest median was 314 (mean: 317; 95% credible interval: 284–370). This is consistent with the previous point estimate of 328 whales (95% confidence interval: 279–386), which includes a correction factor of 1.022 for missed groups (Shelden *et al.* 2017).

Discussion

Visibility bias can vary substantially due to variation in animal behavior, environmental conditions, and other factors despite standardized survey protocols (Anderson 2001). Accurate estimation of abundance and distribution patterns depends on recognition of spatiotemporal variation in visibility bias. Failure to recognize spatiotemporal variation in detectability and the application of constant correction factors can result in misleading estimates of abundance and distribution patterns (see Anderson 2001, Royle 2008 for further discussion). Our results highlighted substantial interannual variation in correction factors for availability, perception, and individual observer bias in the Cook Inlet beluga aerial survey. A weak point in

group size estimation for Cook Inlet belugas is the limited ancillary data on dive behavior to support estimation of availability bias. In particular, if dive patterns are characterized by systematic spatial variation, then interannual variation in the spatial distribution of belugas could contribute to interannual variation in total group size estimates (see Ashford *et al.* 2013). Further research on Cook Inlet beluga diving behavior is therefore needed to strengthen understanding of possible variation in availability bias across habitats (see Thomson *et al*. 2012).

Belugas are highly social throughout their global range but are not known to associate in stable or semistable groups (O'Corry-Crowe 2017). The interdiel variation in beluga group sizes shown here is thus consistent with our current understanding of beluga behavior, although further research on association patterns based on individual photographic identification data may provide new information on group structure.

The predictable aggregation of Cook Inlet belugas in river mouths in early June is associated with foraging on anadromous

salmon and eulachon (Moore *et al*. 2000). Interannual variation in the group size distribution likely reflects interannual variation in the distribution of prey resources, with belugas aggregated in a few large groups when prey is concentrated and dispersed among a larger number of smaller groups when prey is more dispersed. In some cases, aggregation in large groups may also be a response to predation risk (*e.g*., by mammal-eating killer whales, *Orcinus orca*; Shelden *et al.* 2003).

Cook Inlet beluga whales are geographically isolated (Laidre *et al*. 2000), so variation in population size cannot be attributed to immigration or emigration. Belugas are long-lived (O'Corry-Crowe 2017), and the population size is assumed to be effectively constant during each survey period. Interdiel variation in total group size estimates must therefore be attributable to either uncertainty in group size estimates or variation in the group detection process. Bayesian methods provide a clear representation of uncertainty in group size estimates and allow for probabilistic statements about variation in group sizes. Posterior distributions indicate a very low probability that interdiel variation in total group size

estimates is attributable solely to uncertainty in the group size estimation process, and hence, a high probability that it is attributable to variation in the group detection process.

Previous analysis of the group detection process in the Cook Inlet beluga aerial survey assumed that all groups were available to the observers and that group detection is a function of group size only, resulting in very small correction factors for missed groups (*e.g.*, 1.022 for 2016; Shelden *et al.* 2017). However, such small correction factors are inconsistent with the interdiel variation in total group size estimates shown here, which indicate that a much higher percentage of individuals is missed on some days.

Interannual variation in the total group sizes may be attributable to variation in population size. However, comparison of total group size estimates on the survey day with the highest estimates in each survey year suggests that imperfect group detection contributes, at least in part, to interannual variation in the total size of detected groups. In particular, it is not plausible that the population nearly doubled between 2006 and 2010 or between 2011 and 2012 through

demographic growth alone (Fig. 9). The low total group size estimates for all survey days in 2011 are most likely attributable to tidal cycles that prevented the team from surveying the Susitna Delta at low tide, leading to an increase in the effective survey area that made it harder to detect groups. (See Reilly and Barlow [1986] for estimates of maximum plausible demographic growth rates for small cetaceans.)

For species that occur in large or highly variable group sizes, accurate estimation of group sizes is essential for estimating both abundance and distribution patterns. Total group size estimates could be used as a direct measure of abundance if it could be assumed that all groups are detected, or as an index of abundance if consistent survey design ensures consistent group detection probabilities (Anderson 2001). Neither of those conditions applies to Cook Inlet belugas. We therefore urge caution in interpreting total group size estimates as a measure or index of abundance, and recommend a more systematic survey design to maximize survey consistency in the future. For other species or populations, abundance could be estimated by combining group size estimates, based on the model presented

here, with estimates of group detection probabilities, based on surveys designed to support estimation of the group detection process, such as replicated counts (Royle 2004) or distance sampling surveys (Buckland *et al.* 2001). Aerial photography and video are expected to play an increasing role in wildlife research, assessment, and monitoring with the widespread uptake of UAV technology. Key aspects of the methods developed here, especially estimation of perception bias, are relevant for other species for which group size estimates are based on aerial video or photography, including small cetaceans, sirenians, pinnipeds, seabirds and other waterfowl.

ACKNOWLEDGMENTS

Research surveys in Cook Inlet were conducted under NMFS Scientific Research Permits No. 782–1438 (2004), 782–1719 (2005-2010), and 14245 (2011-2016). Aircraft and pilots were provided by Commander NW Ltd., NOAA's Aircraft Operations Center, Northern Commanders LLC, and Clearwater Air Inc. Aircraft. Aerial survey observers for this project were K. Goetz, B. Mahoney, J. Mocklin, D. Rugh, T. Ruszkowski, K. Shelden, O. Shpak, C. Sims, B. Smith, and L. VateBrattström. Laboratory

assessment of video clips was provided by K. Goetz, M. Heider, K. MacIntyre, S. Norman, C. Rosevelt, C. Sims, L. VateBrattström, and J. Waite. We thank R. Angliss, B. Brost, P. Clapham, E. Jacobson, M. Ferguson and three anonymous reviewers for their comments, which helped improve the clarity of the manuscript. The views implied or expressed here are those of the authors and do not necessarily represent the views of the National Marine Fisheries Service, NOAA. Reference in this document to trade names does not imply endorsement by the National Marine Fisheries Service, NOAA.

LITERATURE CITED

Alpizar-Jara, R., and K. H. Pollock. 1996. A combination line transect and capture recapture sampling model for multiple observers in aerial surveys. Environmental and Ecological Statistics 3:311–327.

Anderson, D. R. 2001. The need to get the basics right in wildlife field studies. Wildlife Society Bulletin (1973–2006) 29:1294–1297.

Ashford, J. R., T. Ezer and C. M. Jones. 2013. River discharge predicts spatial distributions of beluga whales in the

Upper Cook Inlet, Alaska, during early summer. Polar biology 36:1077–1087.

Baird, R. W., and L. M. Dill. 1996. Ecological and social determinants of group size in transient killer whales. Behavioral Ecology 7:408–416.

Barlow, J. 2015. Inferring trackline detection probabilities, g (0), for cetaceans from apparent densities in different survey conditions. Marine Mammal Science 31:923–943.

Brooks, S. P., and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7:434–455.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers and L. Thomas. 2001. Introduction to distance sampling: Estimating abundance of biological populations. Oxford University Press, Oxford, U.K.

Caughley, G. 1979. Design for aerial census. Aerial surveys of fauna populations. Australian National Parks and Wildlife Service Special Publication 1:15–20.

Clement, M. J., S. J. Converse and J. A. Royle. 2017. Accounting for imperfect detection of groups and individuals when

estimating abundance. Ecology and Evolution 7:7304–7310.

Elgar, M. A. 1989. Predator vigilance and group size in mammals and birds: A critical review of the empirical evidence. Biological Reviews of the Cambridge Philosophical Society 64:13–33.

Garner, G. W., S. C. Amstrup, J. L. Laake, B. F. J. Manly, L. L. Mcdonald and D. G. Robertson. 1999. Marine mammal survey and assessment methods. A. A. Balkema, Rotterdam, The Netherlands.

Gelman, A. and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7:457–511.

Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin. 2003. Bayesian data analysis. Chapman and Hall, London, U.K.

Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin. 2004. Bayesian data analysis. 2nd edition. Chapman and Hall, London, U.K.

Gerrodette, T., and J. Forcada. 2005. Non-recovery of two spotted and spinner dolphin populations in the eastern tropical Pacific Ocean. Marine Ecology Progress Series

291:1–21.

Gerrodette, T., W. L. Perryman and C. S. Oedekoven. 2018. Accuracy and precision of dolphin group size estimates. Marine Mammal Science 35:22–39.

Gilpatrick, J. W. 1993. Method and precision in estimation of dolphin school size with vertical aerial photography. Fishery Bulletin 91:641–648.

Giraldeau, L. A., and T. Caraco. 2000. Social foraging theory. Princeton University Press, Princeton, NJ.

Goetz, K. T., R. A. Montgomery, J. M. Ver Hoef, R. C. Hobbs and D. S. Johnson. 2012. Identifying essential summer habitat of the endangered beluga whale, *Delphinapterus leucas*, in Cook Inlet, Alaska. Endangered Species Research 16:135–147.

Gowans, S., B. Wursig and L. Karczmarski. 2008. The social structure and strategies of delphinids: Predictions based on an ecological framework. Advances in Marine Biology 53:195–294.

Heithaus, M. R., and L. M. Dill. 2002. Food availability and tiger shark predation risk influence bottlenose dolphin habitat use. Ecology 83:480–491.

Hiby, L. and P. Lovell. 1998. Using aircraft in tandem formation to estimate abundance of harbour porpoise. Biometrics 54:1280–1289.

Hobbs, R. C., J. M. Waite and D. J. Rugh. 2000. Beluga, *Delphinapterus leucas*, group sizes in Cook Inlet, Alaska, based on observer counts and aerial video. Marine Fisheries Review 62:46–59.

Hobbs, R. C., K. Shelden, D. J. Rugh, C. L. Sims and J. M. Waite. 2015. Estimated abundance and trend in aerial counts of beluga whales, *Delphinapterus leucas*, in Cook Inlet, Alaska, 1994–2012. Marine Fisheries Review 77:11–32.

Hodgson, J. C., S. M. Baylis, R. Mott, A. Herrod and R. H. Clarke. 2016. Precision wildlife monitoring using unmanned aerial vehicles. Scientific Reports 6:22574.

Kéry, M., and J. A. Royle. 2016. Applied hierarchical modeling in ecology. Academic Press, Waltham, MA.

Koper, R. P., L. Karczmarski, D. Du Preez and S. Plon. 2016. Sixteen years later: Occurrence, group size, and habitat useof humpback dolphins (*Sousa plumbea*) in Algoa Bay, South Africa. Marine Mammal Science 32:490–507.

Laake, J. L., and D. L. Borchers. 2004. Methods for incomplete detection at distance zero. Pages 108–189 *in* S. T.Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers and L. Thomas eds. Advanced distance sampling. Oxford University Press, Oxford, U.K.

Laake, J., M. J. Dawson and J. Hone. 2008. Visibility bias in aerial survey: mark-recapture, line-transect or both? Wildlife Research 35:299–309.

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

Laidre, K. L., K. E. Shelden, D. J. Rugh and B. A. Mahoney. 2000. Beluga, *Delphinapterus leucas*, distribution and survey effort in the Gulf of Alaska. Marine Fisheries Review 62:27–36.

Lerczak, J. A., K. E. Shelden and R. C. Hobbs. 2000. Application of suction-cup-attached VHF transmitters to the study of beluga, *Delphinapterus leucas*, surfacing behavior in Cook Inlet, Alaska. Marine Fisheries Review 62:99–111.

Lowry, L., G. O'Corry-Crowe and D. Goodman. 2012. *Delphinapterus leucas* (Cook Inlet subpopulation). The IUCN Red List of Threatened Species 2012. e.T61442A17691385. Available at http://dx.doi.org/10.2305/IUCN.UK.2012.RLTS.T61442A17691385.en.

Mahoney, B. A., and K. E. Shelden. 2000. Harvest history of belugas, *Delphinapterus leucas,* in Cook Inlet, Alaska. Marine Fisheries Review 62:124–133.

Marques, F. F. C., and S. T. Buckland. 2003. Incorporating covariates into standard line transect analyses. Biometrics 59:924–935.

Marsh, H., and D. F. Sinclair. 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. Journal of Wildlife Management 53:1017–1024.

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

Moore, S. E., K. E. Shelden, L. K. Litzky, B. A. Mahoney and D. J. Rugh. 2000. Beluga, *Delphinapterus leucas*, habitat associations in Cook Inlet, Alaska. Marine Fisheries Review

62:60–80.

O’Corry-Crowe, G. M. 2017. Beluga whale. Pages 93–96 *in* B. Würsig, J. G. M. Thewissen and K. M. Kovacs, eds. Encyclopedia of marine mammals. Academic Press, London, U.K.

O’Corry Crowe, G. M., R. S. Suydam, A. Rosenberg, K. J. Frost and A. E. Dizon. 1997. Phylogeography, population structure and dispersal patterns of the beluga whale *Delphinapterus leucas* in the western Nearctic revealed by mitochondrial DNA. Molecular Ecology 6:955–970.

O’Corry-Crowe, G. M., Lucey, W., Archer, F. I. and B. Mahoney. 2015. The genetic ecology and population origins of the beluga whales, *Delphinapterus leucas*, of Yakutat Bay. Marine Fisheries Review 77:47–58.

Orbach, D. N., J. M. Packard and B. Würsig. 2014. Mating group size in dusky dolphins (*Lagenorhynchus obscurus*): Costs and benefits of scramble competition. Ethology 120:804–815.

Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical

Computing, Vienna, Austria.

Pollock, K. H., H. D. Marsh, I. R. Lawler and M. W. Alldredge. 2006. Estimating animal abundance in heterogeneous environments: An application to aerial surveys for Dugongs. Journal of Wildlife Management 70:255–262.

Reilly, S. B., and J. Barlow. 1986. Rates of increase in dolphin population size. Fishery Bulletin 84:527–533.

Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60:108–115.

Royle, J. A. 2008. Hierarchical modeling of cluster size in wildlife surveys. Journal of Agricultural Biological and Environmental Statistics 13:23–36.

Rugh, D. J., K. E. Shelden and B. A. Mahoney. 2000. Distribution of belugas, *Delphinapterus leucas*, in Cook Inlet, Alaska, during June/July 1993–2000. Marine Fisheries Review 62:6–21.

Schaub, M., and M. Kery. 2012. Combining information in hierarchical models improves inferences in population ecology and demographic population analyses. Animal

Conservation 15:125–126.

Shelden, K. E. W., D. J. Rugh, B. A. Mahoney and M. E. Dahlheim. 2003. Killer whale predation on belugas in Cook Inlet, Alaska: Implications for a depleted population. Marine Mammal Science 19:529–544.

Shelden, K. E., K. T. Goetz, D. J. Rugh, D. G. Calkins, B. A. Mahoney and R. C. Hobbs. 2015. Spatio-temporal changes in beluga whale, *Delphinapterus leucas*, distribution: Results from aerial surveys (1977–2014), opportunistic sightings (1975–2014), and satellite tagging (1999–2003) in Cook Inlet, Alaska. Marine Fisheries Review 77:2–31.

Shelden, K. E., R. C. Hobbs, C. L. Sims, L. Vate Brattström, J. A. Mocklin, C. Boyd and B. A. Mahoney. 2017. Aerial surveys, abundance, and distribution of beluga whales (*Delphinapterus leucas*) in Cook Inlet, Alaska, June 2016. AFSC Processed Report 2017-09, 62 pp. Alaska Fisheries Science Center, NOAA, National Marine Fisheries Service, 7600 SandPoint Way NE, Seattle, WA 98115. Available at http://www.afsc.noaa.gov/Publications/ProcRpt/PR2017-09.pdf.

Spiegelhalter, D. J., N. G. Best, B. R. Carlin and A. Van Der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B: Statistical Methodology 64:583–616.

Sucunza, F., D. Danilewicz, M. Cremer, A. Andriolo and A. N. Zerbini. 2018. Refining estimates of availability bias to improve assessments of the conservation status of an endangered dolphin. PloS ONE 13(3):e0194213.

Thomson, J. A., A. B. Cooper, D. A. Burkholder, M. R. Heithaus and L. M. Dill. 2012. Heterogeneous patterns of availability for detection during visual surveys: Spatiotemporal variation in sea turtle dive-surfacing behaviour on a feeding ground. Methods in Ecology and Evolution 3:378–387.

U.S. Federal Register. 2008. Endangered and Threatened Species; Endangered Status for the Cook Inlet Beluga Whale, Final Rule. FR 73(205):62919–62930 (22 October 2008). National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Department of Commerce, Washington, DC.

Williams, P. J., M. B. Hooten, J. N. Womble and M. R. Bower.

2017. Estimating occupancy and abundance using aerial images with imperfect detection. Methods in Ecology and Evolution 8:1679–1689.

Received: 13 August 2018

Accepted: 8 February 2019

Supporting Information

The following supporting information is available for this article online at http://

Appendix S1. Additional model details.

Appendix S2. Model code.

Table S1. Beluga groups detected in NOAA Cook Inlet beluga aerial surveys, June 2004–2016.

Table S2. Prior distributions.

Table S3. Multivariate proportional scale reduction factors (PSRF) for group size, ***n*** and total group size ***N***.

Table S4. Model comparison based on differences in the deviance information criterion (ΔDIC).

Figure S1. Schematic of the racetrack pattern used to count beluga whale groups in Cook Inlet.

Figure S2. Histogram of the median of the posterior

distribution of group sizes for detected groups on the survey day with the highest total group size estimate in each survey year.

Figure S3. Comparison of total group size estimates on the survey day with the highest estimates in each survey year, based on the lognormal model (pale gray) and zero-truncated normal model (dark gray) for image size distributions. Points indicate the median of the posterior distribution; filled points indicate the model with the greater support from the data. Error bars indicate the 20th and 80th percentiles. Dashed lines indicate video camera upgrades.

Figure 1. Upper Cook Inlet, Alaska. Since 2001, all groups encountered during NMFS June aerial survey were located in the Upper Inlet.

Figure 2. Graphical summary of the model structure. Elements in shaded boxes represent data; elements in circles represent parameters to be estimated; a cloud represents a latent variable. Arrows indicate the dependence of parameters and variables on data.

Figure 3. Example of proximity bias: two whales can be clearly distinguished inside the circle in the zoom video (a) but not in the matched wide-angle video (b). Example of perception bias due to image size: two whales can be clearly distinguished inside the circle in the zoom video (c) but not in the matched wide-angle video (d). (Numbers shown in images are used to track surfaced whales in video analysis.)

Figure 4. (a) Combining information on the underlying distribution of image sizes ($\pi_{k,p}$, dotted line) and the detection function (ψ_k, *dashed* line) to estimate the average detection probability for available individuals in each video pass (gray shaded bars). (b) Histograms of the image sizes of detected

individuals in three wide-angle video clips for one group (the same group as shown in Fig. 5). The dotted line shows the inferred underlying distribution of image sizes (detected and nondetected) ($\pi_{k,p}$); the dashed line shows the expected distribution of images sizes of detected individuals given the estimated detection function (ψ_k).

Figure 5. (a) Representation of median correction factors for video counts for one group in 2010, based on video counts (*nv*) in three wide-angle video clips. 1/*pa* is the correction factor for availability due to diving behavior; and 1/*pd* is the correction factor for perception bias; the correction factor for proximity bias was 0 in 2010 and is not shown. (b) Posterior predictive checks for video counts for the same group. Filled points indicate the observed video count; boxplots summarize the posterior predictive distribution for video counts.

Figure 6. (a) Standardized representation of median correction factors for video counts by survey year, based on 100 hypothetical individuals detected in a wide-angle video clip (*nv*) in each survey year. (b) Standardized representation of the median correction factor for observer bias by survey year, based

on 100 hypothetical individuals detected in a wide-angle video clip (*nv*) in each survey year. *no*/*nv* is the ratio of observer counts to video counts; and $1/\delta$ is the median correction factor for observer bias.

Figure 7. Interannual variation in the number and size of detected groups. Results are shown for the survey day with the highest total group size estimate in each survey year. Point sizes are proportional to the log of the median posterior prediction of group size. Variation in gray tone serves only to differentiate between survey years.

Figure 8. Interdiel variation in estimated sizes of detected groups (*i.e.*, the posterior distribution of group sizes for detected groups) on five complete survey days in 2016. For each day, groups have been sorted in descending of size.

Figure 9. Interdiel variation in total group size estimates. For each year, days have been sorted by total group size estimate; the model represented (*i.e.*, lognormal or truncated normal model for image size distributions) is the one with greater support from the data; variation in gray tone simply serves to differentiate between survey years. Points

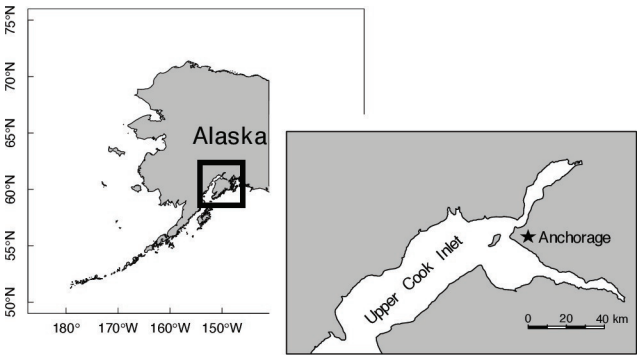
indicate the median of the posterior distribution; open points indicate survey days that were considered incomplete due to weather or other factors. Error bars indicate the 20th and 80th percentiles. Dashed lines indicate video camera upgrades.

Figure 10. Interdiel variation in total group size estimates for detected groups on five complete survey days in 2016. Each histogram summarizes the posterior distribution.

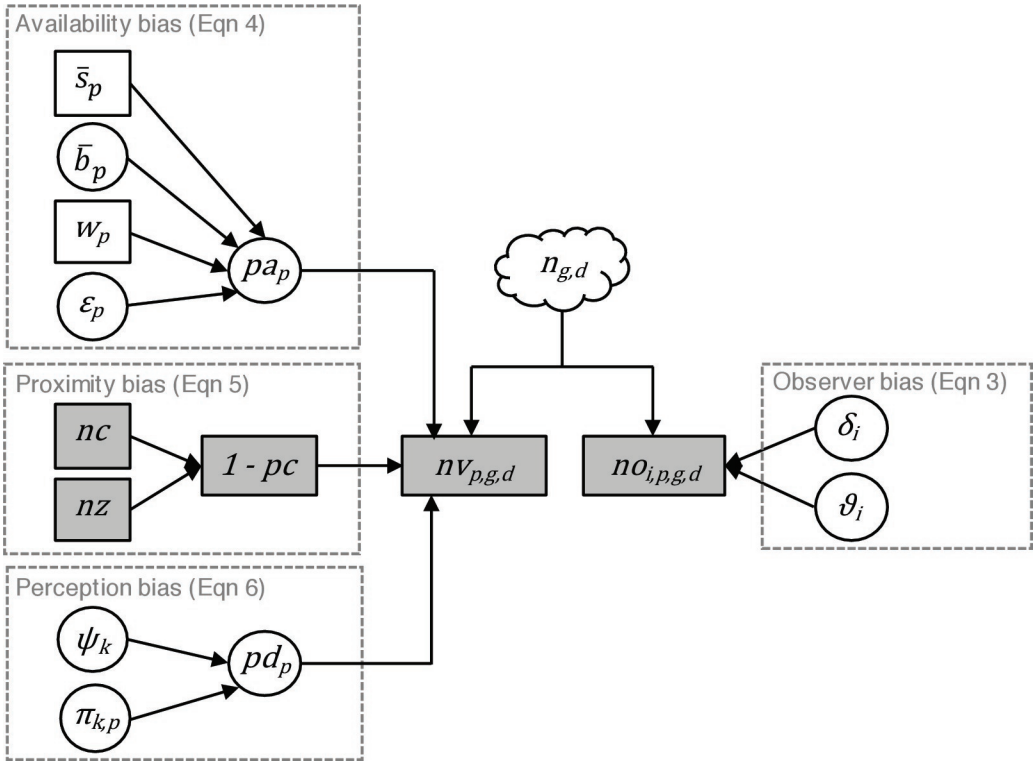
[1] Corresponding author (e-mail: boydchar@u.washington.edu).

[2] Aero Commander 680 in 2004–2006 and 2008–2010; Twin Otter in 2007; Aero Commander 690 in 2011–2016.

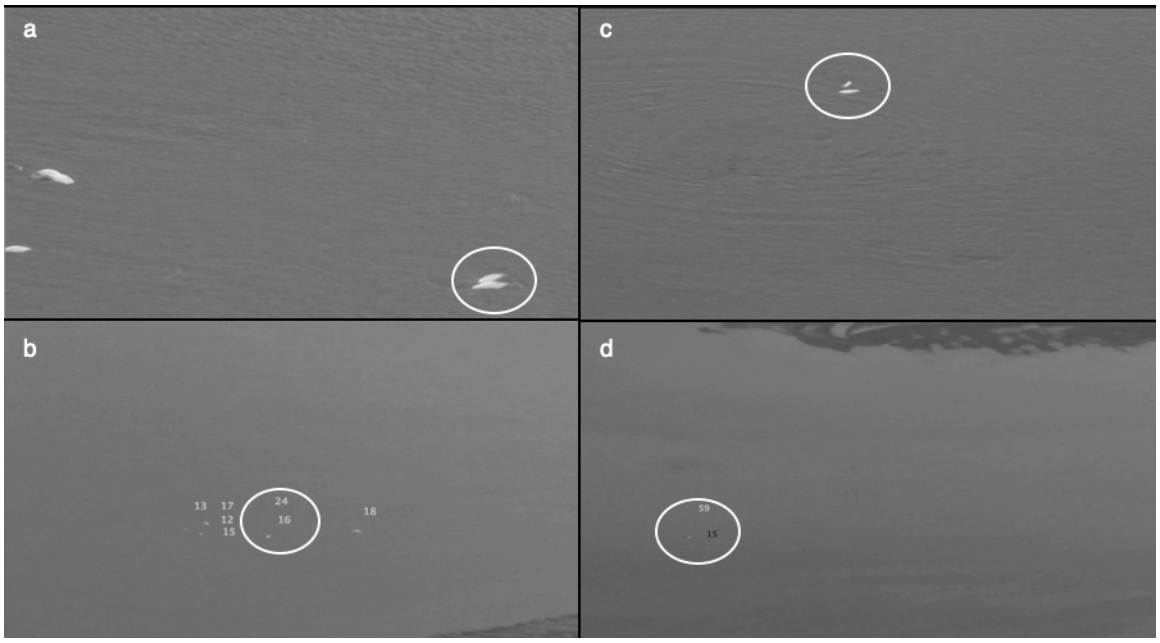
[3] Sony DVCAM (720 x 480 pixels) in 2004–2005; JVC HD 720p (1,280 × 720 pixels) in 2006–2010; and Sony HXR-Nx5U HD (1,920 × 1,080 pixels) in 2011–2016.



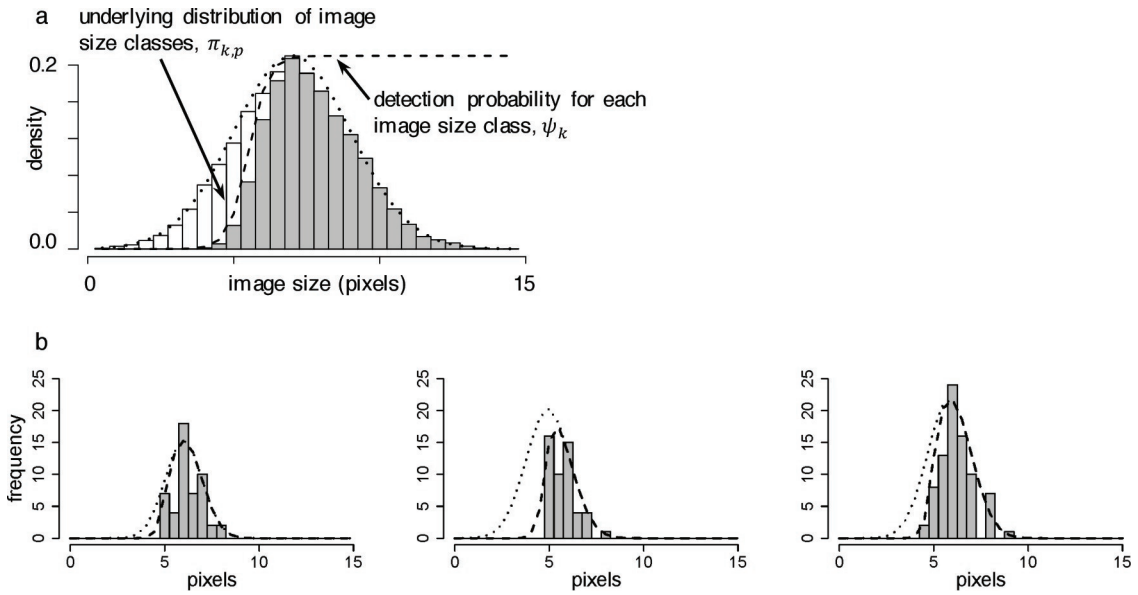
mms_12592_4676_fig 1.eps



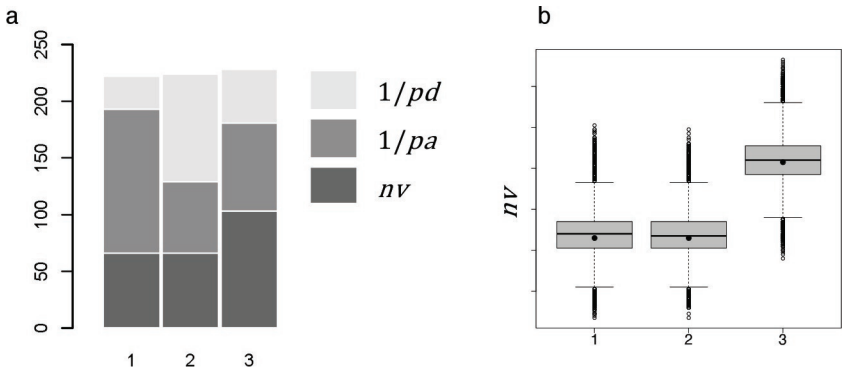
mms_12592_4676_fig 2.eps



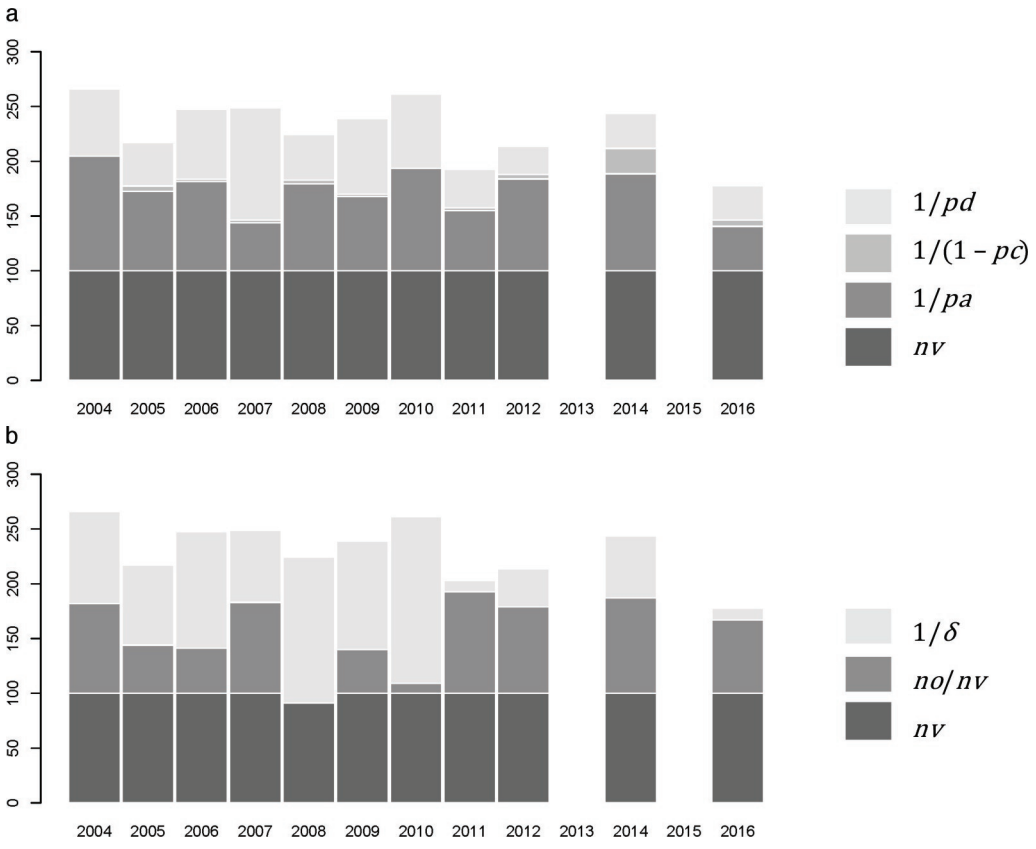
MMS_12592_4676_Fig 3.tiff



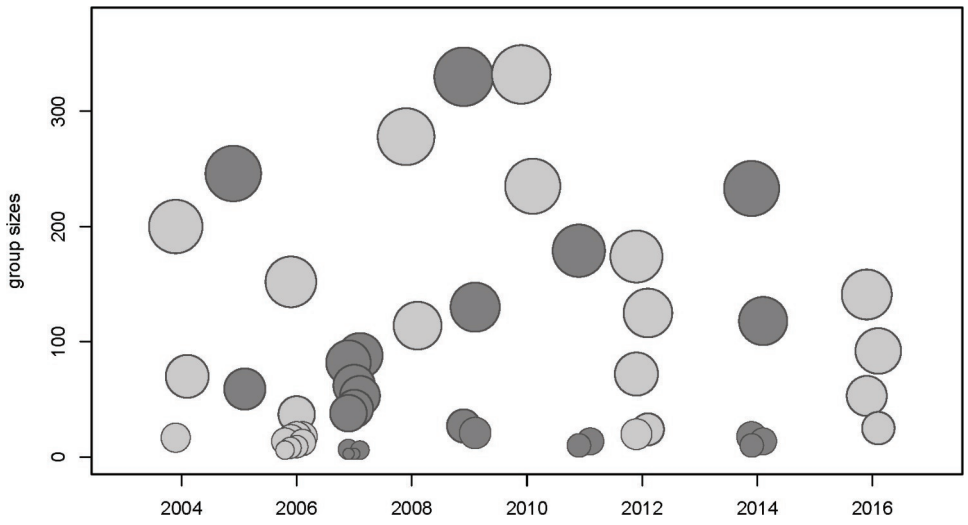
mms_12592_4676_fig 4.eps



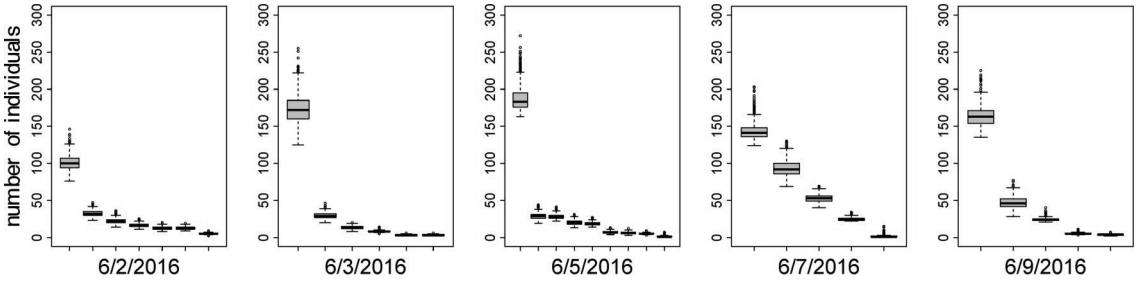
mms_12592_4676_fig 5.eps



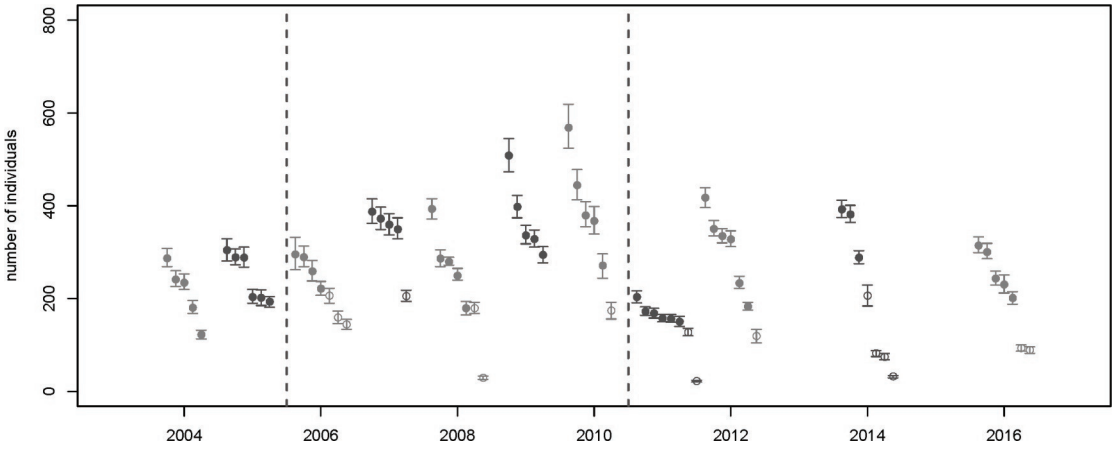
mms_12592_4676_fig 6.eps



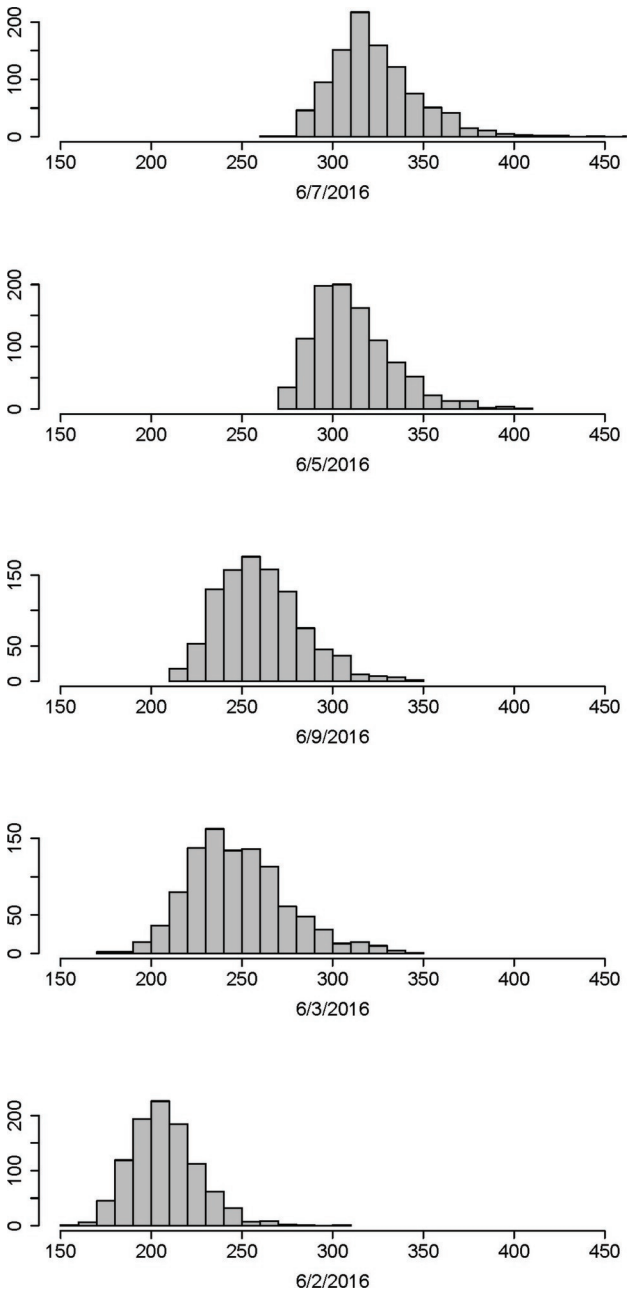
mms_12592_4676_fig 7.eps



mms_12592_4676_fig 8.eps



mms_12592_4676_fig 9.eps



mms_12592_4676_fig 10.eps