Estimating morphometric attributes of baleen whales with photogrammetry from small UASs: A case study with blue and gray whales

JONATHAN D. BURNETT, Aerial Information Systems Laboratory, Forest Engineering, Resources and Management, Oregon State University, 280 Peavy Hall, Corvallis, Oregon, 97331, U.S.A.; LEILA LEMOS and DAWN BARLOW, Marine Mammal Institute, Department of Fisheries and Wildlife, Oregon State University, 2030 SE Marine Science Drive, Newport, Oregon, 97365, U.S.A.; MICHAEL G. WING, Aerial Information Systems Laboratory, Forest Engineering, Resources and Management, Oregon State University, 280 Peavy Hall, Corvallis, Oregon, 97331, U.S.A.; TODD CHANDLER and LEIGH G. TORRES, Marine Mammal Institute, Department of Fisheries and Wildlife, Oregon State University, 2030 SE Marine Science Drive, Newport, Oregon, 97365, U.S.A.

ABSTRACT

Small unmanned aircraft systems (sUASs) are fostering novel approaches to marine mammal research, including baleen whale photogrammetry, by providing new observational perspectives. We collected vertical images of 89 gray and 6 blue whales using low cost sUASs to examine the accuracy of image based morphometry. Moreover, measurements from 192 images of a 1 m calibration object were used to examine four different scaling correction models. Results indicate that a linear mixed model including an error term for flight and date contained 0.17 m less error and 0.25 m less bias than no correction. We used the propagation uncertainty law to examine error contributions from scaling and image measurement (digitization) to determine that digitization accounted for 97% of total variance. Additionally, we present a new whale body size metric termed Body Area Index (BAI). BAI is scale invariant and is independent of body length ($R^2 = 0.11$), enabling comparisons of body size within and among populations, and over time. With this study we present a three program analysis suite that measures baleen whales and compensates for lens distortion and corrects scaling error to produce 11 morphometric attributes from sUAS imagery. The program is freely available and is expected to improve processing efficiency and analytical continuity.

Key words: drone, sUAS, UAV, morphometric, photogrammetry, gray whale, blue whale.

1Corresponding author (e-mail: jonathan.burnett@oregonstate.edu).
Across the animal kingdom, ecologists develop and analyze various metrics to gauge response to environmental and anthropogenic change. A key measure of physiological health and response is animal body condition assessed through morphometrics. Morphometrics are the numerical expression of animal morphological characteristics that facilitate evidenced based quantitative analysis of individual and population trends (Stower et al. 1960). Baleen whale morphometrics can be examined at an individual level to describe energy stores and reproductive capacity, and also at a population level to describe pervasive influences on population health and viability. Typically, larger individuals are considered to be in a better health state due to increased capacity for energy storage (Clutton-Brock and Sheldon 2010, Christiansen et al. 2016). Baleen whales (mysticetes) are long lived, capital breeders that rely on energy stores to support reproductive and migratory life history stages. Therefore, morphometric comparison of baleen whale body condition, across individuals and over time can reveal reproductive state, offspring growth rates, energetic capacity, body size demographic structure, and incidents of compromised health due to injury (Lockyer 1986, Perryman and Lynn 2002, Lockyer 2007, Miller et al. 2012, Christiansen et al. 2016). With synoptic data on prey availability, ecosystem state, and acute impacts, such as entanglements or vessel strike, body condition data can reveal important information on response levels and recovery rates.

Measurements that facilitate morphometric analysis can be generally categorized as direct or indirect. Direct measurement of noncaptive subjects typically involves destructive sampling such as whaling, or samples of opportunity such as stranding events (Norris 1961, Forrester et al. 1980, Finley and Darling 1990). Direct measurements are the gold standard of measurement because morphological parameters can be directly recorded using a measuring device. However, the opportunistic nature of direct measurement leads to inherently small sample sizes and in the case of destructive or stranding samples, remeasurement to discern changes in body condition is impossible. Conversely, indirect methods of photogrammetric morphometric acquisition use geometric principles to estimate parameters based on a scaling reference. This method provides a means of noninvasive data collection, the potential for collecting larger sample sizes, and the ability to remeasure individuals. However, indirect methods yield estimated rather than direct measures and thus are subject to more sources of error, which are detailed later.

Photogrammetry dates to 1829 and has been defined as “…the science or art of obtaining reliable measurements by means of photographs” (Konecny 1985, McGlone 2013). Reliable or precise photogrammetric measurements are possible when certain physical parameters relating to the camera(s) are known. While a limited number of photogrammetric techniques can be applied to a single image (e.g., single photo resection), stereo photogrammetry and 3D reconstruction relies on multiple, overlapping images of a scene acquired using either a moving camera (e.g., on an aircraft) or multiple cameras offset by a known baseline distance. Camera viewing geometries can range from highly oblique to nadir (i.e., direct overhead viewing). These viewing geometries can be achieved via camera placement on boat deck, a crow’s nest, or on an
aircraft. The multicamera method has been used to estimate lengths of hammerhead sharks (Klimley and Brown 1983), bowhead whales (Cubbage and Calambokidis 1987), and sperm whales (Dawson et al. 1995). Single image methods pioneered by Whitehead and Payne (1981) are more common and have been used to estimate morphometry of southern right whales (Best and Rüther 1992, Miller et al. 2012), dolphins (Perryman and Lynn 1993), gray whales (Perryman and Lynn 2002), sperm whales (Jaquet 2006) killer whales, (Fearnbach et al. 2011), and whale sharks (Rohner et al. 2011).

Several sources of uncertainty can influence the precision and accuracy of morphometric measurements from photogrammetry. These have been previously documented (Perryman and Lynn 2002, Jaquet 2006, Fearnbach et al. 2011, Christiansen et al. 2016), and include: body flex, nonhorizontal body position, light refraction on submerged body, deviations in camera roll, pitch (i.e., pointing angle) and yaw, as well as errors in reported range (e.g., distance from camera to whale). Jaquet (2006) used a wooden plank of known length to create a regression model for scaling altitude dependent pixel length measurements to real world units while simultaneously calibrating out ranging error and inherent error sources in the camera and lens system. These errors were determined to be negligible based on the low coefficient of variation (CV) of repeated measurements. Jaquet (2006) also examined effects of angular error and determined that a camera position that deviated by 10° off perpendicular resulted in <2.5% underestimation of fluke width. Perryman and Lynn (2002) and Fearnbach et al. (2011) minimized the influence of uncertainty from body flex, body attitude, body submersion, and camera tilt by filtering out images where these sources were evident to a substantial degree.

Cetacean photogrammetry was traditionally conducted via expensive, time and resource consuming manned aerial surveys, which can be cost prohibitive, thus limiting repeated flights. However, recent technological advances have resulted in the affordable miniaturization of aircraft and camera systems culminating in the advent of small unmanned aircraft system (sUAS) technology (Wing et al. 2013) in the early 2000s. With the advent and accessibility of low cost sUAS technology, photogrammetric measurement of cetaceans has become more accessible, safe, cost effective, and repeatable.

Durban et al. (2015) demonstrated a single camera vertical photogrammetry method for measuring killer whales using a hexacopter equipped with a 25 mm focal length consumer grade camera, and subsequently applied the method to blue whales (Balaenoptera musculus; Durban et al. 2016). Durban et al.’s method used an independent scaling object and achieved length estimates with coefficients of variation from 3.0% to 4.3% for blue whales. Christiansen et al. (2016) applied a similar method to assess the body composition of humpback whales with a low cost Splash Drone (SwellPro Technology Co., Ltd., ShenZhen, China). They concluded that resulting measurements appeared to be robust to error within and between images; however, accurate image scaling required the scaling object (a ship) to be proximal to the whale such that both objects appeared in the same image frame. Another leap in the
evolution the sUAS whale photogrammetry field was presented by Dawson et al. (2017) who used a DJI Inspire Pro sUAS to conduct photogrammetric surveys of southern right whales (*Eubalaena australis*) with the added innovation of a lightweight precision LIDAR instrument to estimate altitude. The improved precision and accuracy of altitude measurements resulted in whale length estimates having 0.5%–1.8% coefficient of variation. These three studies represent the beginning of a new era for baleen whale morphometric research driven by the increased data collection capacity offered by sUASs.

Given the variety of methods used to assess whale morphology and the coming tide of sUAS data resulting from ubiquitous application of low cost (<US$2,000), reliable, consumer grade sUASs, such as the DJI Phantom (DJI Co., Ltd., ShenZhen, China), DJI Inspire, or Splash Drone to cetacean studies, there is a need to develop a standardized and repeatable method of conducting photogrammetric surveys and subsequent morphometric analyses. These low cost systems are relatively simple to operate; however, they are typically characterized as using short focal length lenses with severe lens distortion (Carbonneau and Dietrich 2017) and imprecise altimetry sensors. As such, detailed field and analytical methodology must be developed and disseminated to minimize the propagation of error associated with these characteristics, and ensure that aerial survey operations are being conducted safely and in a manner that is commensurate with drawing inference from the data.

The primary goal of the presented study is to establish methods for conducting accurate and repeatable photogrammetric surveys with low cost sUASs that do not require scaling objects to be coimaged with the survey subject. Incidental to this goal are several secondary objectives: (1) thoroughly evaluate sources for measurement uncertainty, (2) examine strategies to reduce measurement uncertainty, (3) develop standardized methods for extracting whale morphometrics from vertical sUAS imagery, and (4) disseminate these methods in the form of freely available MATLAB and R scripts. We develop these methods with vertical sUAS imagery of eastern North Pacific gray whales (*Eschrichtius robustus*) foraging off the coast of Oregon, and pygmy blue whales (*B. m. brevicauda*) foraging in the South Taranaki Bight of New Zealand. We collect the commonly assessed morphometrics for evaluating whale body condition (e.g., length and width: Perryman and Lynn 2002) and add additional width measurements at percentages of total length similar to the method presented by Christiansen et al. (2016). In addition, we introduce a length normalized surface area index that we refer to as body area index (BAI) that allows comparison of body size among whales similar to body mass index (BMI) in humans.

**Methods**

**Study Area and Collection Methods**

Small UAS overflights of blue whales occurred in the South Taranaki Bight region of New Zealand during January and February 2016 as part of a larger project to describe the ecology of this population (Torres 2013;
Barlow et al. 2018). Field methods are thoroughly documented in Torres et al. (2017). Six blue whales were imaged over four separate flights during this period. Small UAS overflights of gray whales occurred off the Oregon coast during August–October 2016; 89 gray whales were imaged over 43 flights. The primary survey equipment for this study was a DJI Phantom 3 Pro sUAS (P3 Pro) and DJI Phantom 4 sUAS (P4). The cameras on both aircraft have a 3.61 mm focal length and 0.0015 mm pixel size (i.e. pixel pitch). Manual flight control of the aircraft was through the included remote control. Small UAS configuration and real time camera output were available through an Apple iPad Mini tablet ground station that was operating the DJI Go application.

The DJI sUASs were chosen because the systems are robust to cross winds even when traveling at 40 km/h, pilot training is intuitive, the aircraft can safely initialize on a moving platform (e.g., boat), and as is demonstrated later, the cameras have lower distortion than commonly used action cameras like the GoPro Hero. The camera is stabilized by a three axis brushless gimbal, is capable of 4K video output, and can transmit a high definition real time video sample to the pilot/observer. The video contains altitude and geolocation metadata recorded at 1 Hz and camera directional pointing is controlled via remote control.

The sUAS was navigated such that the whales were centered in the camera field of view at altitudes between 25 m and 40 m above sea level (ASL) with a flight duration of <10 min. A calibration object of known length was centered in the frame and imaged from 10 m to 40 m during takeoff and landing for all flights, for correcting barometric altimeter error in postprocessing (Jaquet 2006, Durban et al. 2015). Object lengths were 4.40 m and 1.00 m for the blue and gray whale flights, respectively. The product of the aerial survey is a 4K video of individual whales and calibration objects. Video format was chosen instead of still images because individual frames are high resolution (e.g., 8 MP) and video increased the likelihood of capturing a whale in an ideal presentation (e.g., at water surface, fully elongated, not rotated on any axis in reference to the water surface, and body outline is clear). Additionally, the use of video format facilitated postflight behavioral analysis for use in a follow on study.

Body Area Index

Christiansen et al. (2016) demonstrated that intraseasonal body condition changes in humpback whales can be assessed with a body condition index (BCI). BCI captures width variation along the length of the whale by segmenting trapezoids along percentiles of body length and using sums of trapezoids to estimate flat dorsal surface area. Although BCI offers an approximation of surface area (SA), here we develop a more complete estimate of the flat dorsal surface area of a whale than trapezoid sums, by assuming a parabolic shape for approximately 40% of a whale’s total length and combining the area under parabolas representing each side of a whale (Fig. S1). An initial analysis indicated that manual analyst detection of whale body edge resulted in very high uncertainty. As such, we chose to constrain these edge detections with a parabolic model because
(1) the model fit the test set of whales well ($R^2 > 0.9$), and (2) the parabolic model constrains the impact of interpretive error by providing a least squares optimized fit that adds objectivity to the subjectivity of clicking points to delineate the edge of the whale body. Furthermore, parabolic models have been successfully used to evaluate body condition in other large bodied mammals, such as cows (Halachmi et al. 2008, 2013). The suitability of using parabolic models to derive surface area for the BAI calculation was examined by estimating $R^2$, as well as visual analysis of the agreement between each parabola and the side of the whale. We also examined the independence of BAI and WL by estimating Pearson correlation coefficient between WL and BAI and comparing them to those between SA and BAI.

We evaluated the parabolic shape of each whale by orienting each whale image along a horizontal axis from rostrum to tail where length (pixels) is along the x-axis and width (pixels) is along the y-axis. An analyst manually placed 11 points on the outline of each whale side between 15% and 65% of length in approximately 5% intervals. Parabolas were independently fit to each side, due to a lack of symmetry caused by mild variation in body presentation (e.g., curvature). Parabolas were fit to the points using Equation 1:

$$BW_i = \beta_0 + \beta_0(WL_i^2) + \epsilon_i$$  \hspace{1cm} (1)

where $BW_i$ is width at $WL_i$, where $WL_i$ is length in units of pixels (p) at the $i$th percentage of WL. The goodness of parabolic fit was also evaluated between 10% and 70% to determine if more area can be included without compromising the quality of the model. Models were evaluated using $R^2$ and $P$ values. Surface area was estimated from the fitted parabolas using Equation 2:

$$SA_p = \left( \int_{WL_{20\%}}^{WL_{60\%}} (\text{Eq.1}_{s1}) dx \right) + \left( \int_{WL_{20\%}}^{WL_{60\%}} (\text{Eq.1}_{s2}) dx \right)$$  \hspace{1cm} (2)

where $SA_p$ is the surface area in pixels (p) between 20% and 60% of WL and Eq. $1_{s1}$ and Eq. $1_{s2}$ are parabolic models for side 1 and side 2, respectively.

We assert that a length normalized index of body composition is appropriate for our study, because it facilitates body condition comparison among individuals and between observations of the same individual in the way that body mass index (BMI) is used to compare body condition among humans (Flegal et al. 2012). BMI = mass (kg)/height (m)$^2$ (Gallagher et al. 1996); to emulate this we used surface area as a surrogate for body mass and estimated body area index (BAI) using Equation 3:

$$BAI = \frac{SA_p}{(0.4 \times WL_p)^2} \times 100$$  \hspace{1cm} (3)
where $S_{A_p}$ is the estimated surface area of the whale in units of p, $WL_p$ is the estimated length of the whale and Equation $1_{s1}$ and Equation $1_{s2}$ represent the parabolic models fit for side 1 and side 2 of the whale, respectively. The length is multiplied by 0.4 because surface area is only captured across 40% of WL (between 20% and 60% body length). The multiplication by 100 allows the BAI estimate to be a more intuitive value >1.0. There are two distinct advantages to the proposed BAI metric. BAI is a unitless index because $SA_i$ is divided by $(0.4 \times WL_p)^2$, canceling the units. A unitless index has the benefit of being scale invariant and is not influenced by scaling errors that may arise during photogrammetry efforts, as described later. As a unitless normalized index, BAI can be used to compare size of an individual between two time periods, size between two individuals, size relative to the larger population, and size relative to an established standard. Mean BAI for a population can be derived from Equation 4:

$$SA_{2i} = \beta_1 (0.4 \times WL_p)^2 + \varepsilon_i$$  \hspace{1cm} (4)

where $SA_{2i}$ is the $SA \times 100$ of whale $i$, $\varepsilon_i$ is random error, and the intercept is zero to ensure $SA2$ is zero when $WL = 0$. Mean BAI (mBAI) = $\beta_1$ and is the mBAI for the population of whales used to train the model.

We developed five simulated change scenarios to examine the sensitivity of BAI to change at the population and individual level. We established a desired change sensitivity threshold at 10% SA because 10% is the low end range of gray whale seasonal change observed by Rice and Wolman (1971). The 10% WL change amount was chosen to reduce confusion during analysis and interpretation. We simulated changes by applying the respective changes in WL and SA specified in Table 1 to the raw (i.e., uncorrected) morphometric data. The simulation scenarios were intended to represent biologically meaningful change events. BAI1 is the unchanged scenario representing the current BAI of the whale. BAI estimates for each scenario and whale were analyzed for significant difference using a pairwise Bonferroni adjusted $t$-test of BAIs with the pairwise.t.test function in R. Population sensitivity significance was assessed by comparing differences in mBAI using the SE of $\beta_1$ in the fitted models for each of the simulations.

Table 1. Simulated change scenarios devised to discern how changing whale length (WL) and surface area (SA) influence body area index (BAI) estimates.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>WL</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAI1</td>
<td>Unchanged</td>
<td>Unchanged</td>
</tr>
<tr>
<td>BAI2</td>
<td>0.1</td>
<td>Unchanged</td>
</tr>
<tr>
<td>BAI3</td>
<td>Unchanged</td>
<td>0.1</td>
</tr>
<tr>
<td>BAI4</td>
<td>Unchanged</td>
<td>-10%</td>
</tr>
<tr>
<td>BAI5</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Photogrammetric Method

We employed a vertical image photogrammetric method that used camera to subject distance (i.e., range) to scale images from pixels to
meters, similar to the method described in Jaquet (2006) and Fearnbach et al. (2011). Aircraft altitude above sea level (ASL) was used as a surrogate for range with the assumption that images were captured at nadir, the subject was imaged at the sea surface, and ASL was zero at sea level. Morphometric attributes were measured in pixels in each of the whale images using the *Whale Measurements* program developed in MATLAB (MATLAB and Image Processing Toolbox Release 2016b, The MathWorks, Inc., Natick, MA). Pixel lengths were converted to metric lengths via ground sampling distance (GSD) using Equation 5:

\[
GSD_i = \frac{d_c (H'_i + \varepsilon_{hi})}{f_c}
\]

where \(GSD_i\) was the ground projected horizontal surface distance represented by one side of a square image pixel (Comer et al. 1998), and \(H'_i\) was the ASL of the uAS at the \(i\)th observation. \(d_c\) was the physical dimension of one side of a square pixel on the sensor chip in units of mm (i.e., pixel pitch) and \(f_c\) was focal length in units of mm; both parameters were fixed and specific to camera \(c\). The term \(\varepsilon_{hi}\) is the bias in ASL at the \(i\)th observation.

![Diagram](image.png)

*Figure 1.* A graphical depiction of the pertinent parameters used to estimate scaled object lengths from measurements. GSD = ground sampling distance (i.e., ground distance of one pixel), \(d_c\) = pixel pitch, \(H'\) = altitude, and \(f_c\) = focal length.
and accounts for the likelihood that aircraft barometer was zeroed above sea level. Figure 1 is a graphical depiction of these parameters.

Once GSD was calculated, objects were scaled from a pixel length measurement to a metric length estimate with Equation 6:

$$L'_{ki} = GSD_i \times OL_{pki}$$

where $L'_{ki}$ was the scaled length of object $k$ at observation $i$, GSD$_i$ was calculated for observation $i$, and $OL_{pki}$ was the pixel length of object $k$ at the $i$th observation.

Sources of Uncertainty in Photogrammetric Measurements

The primary sources of uncertainty in measurement estimates derived from vertical imagery (i.e., nadir pointing) are those related to (1) lens distortion, (2) the assumption of zero camera tilt, (3) the uncertainty in the above ground level (AGL) flying height (and, hence, the image scale), and (4) the uncertainty in the analyst digitization of the whale body edges on the imagery (Dolan et al. 1978).

Lens distortion is the result of passing light through an aspherical onto a two dimensional sensor (Zhang 2000). The type and severity of lens distortion is primarily correlated to focal length, but also materials and manufacturing methods (Zhang 2000). Tilt uncertainty is the degree of uncertainty of the true pointing angle of the camera when the camera was assumed to be pointing nadir. Camera tilt was excluded from this analysis after five aircraft initializations on a level surface resulted in $<3^\circ$ of tilt error, which is under the conventional threshold warranting explicit correction (Philpot and Philipson 2012, U.S. Army Corps of Engineers 2015). Tilt error was estimated with a protractor on a level surface and a straight edge aligned with the pointing angle of the camera. We expect the tilt error to be similar at sea because the gimbal uses accelerometers and electrical gyroscopes to identify the gravity vector rather than establishing a reference from the calibration surface.

Ranging uncertainty in the context of this study refers to the uncertainty associated with estimating true sUAS AGL ($H'$). $H'$ is the only variable required for estimating GSD (Eq. 5) and ultimately scaling images. Several factors contribute to ranging uncertainty and include: wind driven changes in local barometric pressure, ocean swells, zeroing aircraft altitude above sea level (e.g., ship deck), imaging a subsurface whale, and imprecision inherent in low cost barometric sensors. We define analyst digitization uncertainty as the uncertainty associated with estimating true object length in image space. Based on this definition, there are two primary drivers of digitization uncertainty: (1) the deviation from the assumption that the object being measured is flat level and perfectly orthogonal to the camera, and (2) the uncertainty associated with an analyst manually measuring an object on an image.
Mitigating the Influence of Uncertainty Sources with Linear Modeling

A commonly employed ranging (e.g., altitude) correction model is described in (Jaquet 2006) and requires an object of known length (e.g., calibration object) to calculate what we refer to as empirical GSD (eGSD). eGSD is calculated by dividing known calibration object length (OL) in meters by estimated calibration object length in pixels OLp and calculates the ASL ($H_0$) from which the object must have been imaged based on the geometric relationship between fixed camera parameters (e.g., $f_c$) and OLp in the image. The correction model accounts for systematic error in ranging that results from zeroing the barometric altimeter above sea level and from an altimeter that exhibits bias. We calculated eGSD and regressed it against observed ASL ($H_0^i$) using Equation 7 to estimate a corrected GSD (cGSD) using the lm function in R:

$$cGSD_i = \beta_0 + \beta_1 (H_0^i) + \epsilon_i.$$  (7)

The cGSDs estimated from this model can be considered unbiased and observations from multiple flights and days can be aggregated to increase sample size and thus power, when certain underlying assumptions are met. In addition to the conventional assumptions associated with linear modeling (e.g., independent, and normal distribution of observations), Equation 7 assumes (1) that the altimeter is always zeroed to the same height above sea level, (2) local environmental barometric pressure does not change over the duration of the flight, and (3) the pixel length measurements used to calculate eGSD are unbiased and precise. Violations of assumptions 1 and 2 lead to bias in the resulting GSD estimates.

Error in Equation 7 may not be completely random due to unique conditions associated with range estimation for each individual flight per day (abbreviated as Date-Flight), such as different take off locations, wind conditions, and ocean swell. We examined the possibility of a Date-Flight effect on the relationship between GSD and altitude to determine if calibration should be performed on a per day and flight basis by assessing significance and explained variance with a repeated measures analysis of variance (ANOVA) (Girden 1992, Weinfurt 2000). To account for the possibility that Date-Flight has a significant effect on GSD, we modified Equation 7 to be a linear mixed model (LMM) that includes a Date-Flight error term (Eq. 8), using the lmer function in the lme4 package in R (De Boeck et al. 2011, Bates et al. 2014):

$$cGSD_{ji} = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j})(H_0^i) + \epsilon_{ji}.$$  (8)

where cGSD is estimated at $H_0^i$ for the $i$th observation in Date-Flight group $j$, $u_{0j}$ is a random effect to account for changing intercept (i.e., bias) by Date-Flight, and $u_{1j}$ is a random effect accounting for changing slope (e.g., barometric instability) for each Date-Flight group.
Smoothed-corrected GSD (scGSD) was evaluated as a potential mitigation of highly variable OL\(_p\) estimates on cGSD precision by regressing on a smoothed eGSD. Equations 7 and 8 are both sensitive to violations of the assumption that OL\(_p\) estimates are accurate. Since digitization error is one of the primary error sources in photogrammetric measurements, it is highly likely that this assumption is violated in most instances. The geometric relationship between \(H_0\) and OL\(_p\) is such that log\(_{10}\) transformed pixel length decreases as log\(_{10}\) transformed \(H_0\) increases. Any deviation from 1:1 linearity suggests the presence of digitization error. We smoothed OL\(_p\) by regressing the pixel length of the calibration object to the observed \(H_0\):

\[
\log_{10}(OL_{pi}) = \beta_0 + \beta_1 \times \log_{10}(H_{0i}) + \epsilon_i \tag{9}
\]

This model is appropriate for estimating OL\(_p\) if no substantial deviations from assumptions 1 and 2 described previously are evident. If these assumptions are violated, then a modified model form that includes a Date-Flight error term is necessary. To account for this possibility, we created a linear mixed model specified in Equation 10:

\[
\log_{10}(OL_{pji}) = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j}) \times \log_{10}(H_{0ji}) + \epsilon_{ji} \tag{10}
\]

scGSDs were estimated by dividing OL by modeled OL\(_p\) at \(H_0\) from Equations 9 and 10, respectively, and regressing the result against \(H_0\) using Equations 11 and 12, respectively. The scGSDs estimates have the potential to be biased when \(H_0\) or OL\(_p\) measurements are consistently over or under estimated. However, the estimates will be inherently less variable and potentially increase the likelihood of detecting trends in the data.

\[
scGSD_i = \beta_0 + \beta_1 (H'_0) + \epsilon_i \tag{11}
\]

\[
scGSD_{ji} = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j}) (H'_{0ji}) + \epsilon_{ji} \tag{12}
\]

**Evaluation of Linear Model Corrections with Calibration Data**

The linear models were applied to the calibration data in different combinations to facilitate comparison of estimated mean OL (Eq. 3) among the five correction methods specified in Table 2. The precision of mean estimated length was assessed using mean root mean squared error (Eq. 13) and bias. Coefficient of variation (CV) was calculated to assess relative measurement variation by correction method. 95% confidence intervals (CIs) were estimated for all estimates using 2.5% and 97.5% quantiles from a nonparametric bootstrap of \(n = 1,000\) replicates. Non-parametric bootstrapping has been shown to be an effective estimator of standard error when sample sizes are small and true variance is unknown (Efron and Tibshirani 1986). CIs for M2 through M5 included GSD model prediction uncertainty. Equations 7 and 9 used a SE derived from the
Table 2. Different correction methods examined for estimating GSD and eGSD using the regression models developed previously. eGSD describes how eGSD was calculated for the purpose of training the model listed under the GSD column and applies strictly to the calibration object estimates. GSD describes the GSD estimation method being used in Equation 6 to estimate scaled length and is applicable in the context of both the calibration data and the whale morphometric measurement data.

<table>
<thead>
<tr>
<th>Method</th>
<th>eGSD</th>
<th>GSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>—</td>
<td>Eq. 5</td>
</tr>
<tr>
<td>M2</td>
<td>OL ÷ OL_p</td>
<td>Eq. 7</td>
</tr>
<tr>
<td>M3</td>
<td>OL ÷ OL_p</td>
<td>Eq. 8</td>
</tr>
<tr>
<td>M4</td>
<td>OL ÷ Eq. 7</td>
<td>Eq. 11</td>
</tr>
<tr>
<td>M5</td>
<td>OL ÷ Eq. 8</td>
<td>Eq. 12</td>
</tr>
</tbody>
</table>

The bootstrapping method described in Davison and Hinkley (1997) in conjunction with the boot function in R (Carlson and Ripley 1997) and the LMMs (Eq. 8, 10, 11, 12) used predictinterval function within merTools (Knowles and Frederick 2016, Knowles et al. 2016). The CIs for the LMMs do not account for the variance associated with the random error term (e.g., Date-Flight). CIs were used in conjunction with an ANOVA to discern difference of mean OL among correction methods:

\[ m\text{RMSE}_{s} = \frac{\sum_{j=1}^{m_s} \sqrt{\sum_{x_i=1}^{n_{sj}} \frac{(x_{sij} - x_t)^2}{n_{sj}}}}{m_s} \] (13)

where \( m\text{RMSE}_{s} \) is the mean RMSE for the \( s \)th correction method, \( m_s \) is the number of flights for a given method \( s \), \( n_{sj} \) is the number of observations within a given correction method and Date-Flight (\( j \)), \( x_{sij} \) is the \( i \)th observation for a given method and \( j \)th Date-Flight, and \( x_t \) is the true length of object \( t \). The linear models were evaluated for substantial departures of the assumptions of constant variance and symmetric error distribution using residual plots.

Comparison of Linear Model Corrections on Whale Morphometrics

The purpose of the linear corrections was ultimately to improve the precision and accuracy of whale morphometric measurements. The performance of a given correction method with the calibration data was expected to be a good indication of how the correction method would estimate any given morphometric whale measurements. The GSD correction models (Eq. 7, 8, 11, 12) trained in the correction methods in Table 2 were applied to the whale observation data to estimate cGSD. The means of scaled morphometric attributes were calculated for each whale. True morphometric attribute length was unknown, so the five GSD correction methods were evaluated with CV, CIs, and graphically with boxplots. CIs were calculated using the nonparametric
bootstrapping method previously described. The linear models were evaluated for substantial departures of the assumptions of constant variance and symmetric error distribution using residual plots.

**Mitigating Error from Lens Distortion**

Lens distortion is conventionally quantified with a polynomial equation (Brown 1971, Zhang 2000). Although Adobe Lightroom (Adobe Systems Inc., San Jose, CA) provides calibration profiles for the DJI Phantom cameras used in this study, the profiles were estimated with full resolution image stills and we used video frames. Therefore, we estimated and corrected distortion using the Mathworks Single Camera Calibrator application and the method described in the associated tutorial (Mathworks 2017). The Mathworks method uses a variation on Brown’s distortion equation based on the multiview distortion coefficient estimation method developed by Zhang (2000).

We included the similarly priced DJI Phantom 4 Pro in this distortion analysis as a convenience to the community because the Phantom 3 Pro and Phantom 4 have already been phased out of production; future DJI-based photogrammetry will be conducted with the Phantom 4 Pro or newer models. For each camera we captured 4K video while rotating the x- and y-axis of a checkerboard at various angles. Fifty frames were extracted from the video and ingested into the Single Camera Calibrator application to estimate the three parameter polynomial model that best estimates lens distortion for a given camera model.

The lens distortions of the DJI Phantom cameras were compared to cameras similar to those used in Christiansen et al. (2016) and Durban et al. (2016). We extracted calibration information from existing Adobe Lightroom lens calibration profiles for the GoPro Hero 4 Black (Christiansen et al. 2016). A profile for the Canon Powershot D30 used in Christiansen et al. was not available so we used the profile for the Canon Powershot S100 at 5.2 mm focal length. This seemed appropriate since both cameras have similar sensor sizes and lenses. Similarly, we used the available profile for the 25 mm Voigtlander lens calibrated on a micro 4/3s Panasonic DMC camera because it was expected to be analogous to Durban’s (2016) Olympus Zuiko lens given the identical focal length and sensor size, and similar cost.

**Contributions of Nondistortion Error Sources to Total Uncertainty**

As ranging error (e.g., AGL altitude) and analyst digitization error (e.g., pixel length measurements) are two of the largest sources of uncertainty in vertical photogrammetric measurement estimates, we examined the relative impact of each of these components on object length estimates in the calibration data for both the blue and gray whale data sets. For this analysis, we used the special law of propagation of variance to estimate total propagated uncertainty when estimating object length using Equations 5 and 6, assuming $H^*$ and $OL_p$ are independent (Ghilani 2011). Total propagated uncertainty was evaluated using Equation 14. We estimated total propagated uncertainty at 15 m and 40 m because approximately 95% of the imaging of the calibration objects occurred
within this altitude range and these two extremes were expected to illu-
minate potential altitude dependent trends in individual parameter influ-
ence on total uncertainty. Direct estimation of $\sigma$ was not possible
because Equation 14 must be evaluated at a fixed altitude and the cali-
bration objects were imaged across a range of altitudes. The variance of
$H'$ was determined using the SE of the predictor in Equation 11 and the
variance of OL was determined using the SE of the predictor in
Equation 9. LMMS (Eq. 10, 12) were not used due a dependence on
Date-Flight. Equation 11 was chosen for estimating $H'$ variance because
it is expected to be less influenced by variability within OL measurements:

$$\sigma_{OL'} = \pm \sqrt{\left( \frac{1}{f_c} \times d_c \right)^2 \times \sigma^2_{H'} + \left( \frac{H'}{f_c} \times d_c \right)^2 \times \sigma^2_{OL'}} \quad (14)$$

where $\sigma_{OL'}$ is 1 standard deviation of the estimated object length, $\sigma^2_{H'}$ is
the square of the standard deviation of ASL, and $\sigma^2_{OL}$ is the square of
the standard deviation of the number of pixels.

**Image Extraction**

The basis for measurement of individual whales in this study is the
nadir pointing 4K video taken during sUAS flight. For each whale sight-
ing, five full resolution frames were extracted from the video using the
“snapshot” functionality in VLC Media Player (version 2.2.4) following
the pertinent recommendations of the error mitigation strategies
described above. Five calibration images were also extracted during
both take off and landing of each flight. Effort was made to ensure the
survey objective was centered, in focus, and that the full range of the
altitude gradient was represented within the set of 10 images per flight.

**Image Analysis**

We developed a three-program analytical framework (Appendix S2) for
photogrammetric whale morphometric analysis to minimize sources of ana-
lytical error and standardize morphology measurements across multiple ana-
lysts and images. The first program is titled Whale Calibration Object
Measurement and was developed in MATLAB to standardize the measure-
ment of GSD calibration objects as well as the pertinent outputs that facilitate
mitigating the effects of uncertainty through linear modeling. The program
prompts the user for the following inputs related to the sUAS camera, cali-
bration object, and specific flight: $f_c$, $d_c$, sighting number, flight number,
date, object name, known OL in mm, observe $H'$, and height difference
between sUAS initialization location and the calibration object (parameter-
ized as $\epsilon_{hi}$ in Eq. 5). The user is guided through an interactive measuring pro-
cess and a summary table is produced that includes the prompted inputs,
OL, and GSD (as calculated from Eq. 5). This program also incorporates the
lens distortion corrections for the three DJI cameras, but also allows users to
provide a custom lens distortion correction for any camera that was calibrated with the MATLAB Camera Calibrator application.

The second program, titled *Whale Measurements*, was also developed in MATLAB, for the purpose of standardizing the measurement of morphometric attributes in vertical whale images collected from sUASs, and standardizing the output measurements for subsequent scaling and error assessment in a third program described later. The program interactively requests an input image, and guides the user through several processes in the following sequence: (1) prompts user for details pertinent to the specific camera, date, flight, sighting and individual whale, (2) crops the image to the subject of interest, (3) aligns the whale lengthwise across a horizontal axis with an origin of 0,0 to simplify calculations, (4) guides the user through a series of measurements to collect the morphometric attributes specified in Table 3, (5) prompts the user for flight ASL ($H^0$) and the vertical difference between sUAS initialization location and sea level parameterized as in Eq. 5), and (6) outputs a table of results and an image (e.g., Fig. S1) containing the subject with an overlay of pertinent metrics. This program also incorporates the lens distortion correction models for the three DJI cameras.

The third program, titled *Whale Quantitative Analysis*, was developed in R 3.3.1 (R Core Team 2017) to process the summary tables produced in the first two programs to provide measurement data and error estimates. This program queries the user for directories that contain the summary tables produced in the previous two programs. Data are automatically grouped by date, flight, and sighting under the assumption that all data shares a common calibration object and camera. The program builds the GSD correction models from the calibration data and applies the corrections to both the calibration data and whale morphometric data to create the five GSD correction methods in Table 2 for each whale. Summary tables of the calibration data include RMSE, bias, CIs, CVs, and the estimates of total propagated uncertainty (Eq. 14) grouped by correction method. Summary tables of each whale and associated morphometric measurements include the estimated scaled value, CIs, and CVs. The results presented below are based primarily on data collected during sUAS survey of 89 gray whales in

Table 3. The morphometric attributes produced by the analytical programs and their descriptions.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>WL</td>
<td>Whale length: rostrum to notch in tail</td>
</tr>
<tr>
<td>MW</td>
<td>Manual width: manual measurement of width at widest point</td>
</tr>
<tr>
<td>OW</td>
<td>Optimized width: width at point on parabola nearest MW</td>
</tr>
<tr>
<td>FW</td>
<td>Tail width: tip to tip fluke width</td>
</tr>
<tr>
<td>W20</td>
<td>Width at 20% of WL from rostrum</td>
</tr>
<tr>
<td>W30</td>
<td>Width at 30% of WL from rostrum</td>
</tr>
<tr>
<td>W40</td>
<td>Width at 40% of WL from rostrum</td>
</tr>
<tr>
<td>W50</td>
<td>Width at 50% of WL from rostrum</td>
</tr>
<tr>
<td>W60</td>
<td>Width at 60% of WL from rostrum</td>
</tr>
<tr>
<td>SA</td>
<td>Surface area between 20% and 60% of WL</td>
</tr>
<tr>
<td>BAI</td>
<td>Body Area Index</td>
</tr>
</tbody>
</table>
Oregon, collected in 2016 and include results from the data collected during sUAS survey of six pygmy blue whales in New Zealand collected in 2016.

RESULTS

Gray Whale Calibration Object Correction Method Comparisons

Image measurements of the 1 m calibration object were examined in the context of the GSD correction methods. The blue whale calibration data was similarly analyzed but the results were excluded from this manuscript to economize space since the trends were largely similar. Data were filtered due to high variability (CV > 15%) of estimated scaled object length. Twenty-two of the 193 observations, were removed due to uncorrected estimated lengths >1.97 SDs ($t = 0.025$, df = 192) from the mean of all measurements.

The linear models used for the GSD correction methods (Table 2) were visually evaluated for violations of non constant variance and non random error using plots of predicted vs. residuals. Equation 7 and 11 residuals displayed structural trends that indicate a violation of the assumption of error non heteroscedasticity. Nonheteroscedasticity indicators were not visible in the LMMs (Eq. 8, 12) residuals, which suggests Date-Flight grouping accounted for the non random error.

The effect of a Date-Flight grouping of calibration data was examined to determine if a per flight imaging of the calibration object was necessary. A Date-Flight grouping conceptually accounts for systematic variances in altitude that ultimately increase uncertainty in GSD and subsequent scaled length estimates. Results of the repeated measures ANOVA indicated a significant effect of Date-Flight on GSD at the $P < 0.05$ level ($F_{1,39} = 2.93 \times 1,031$, $P < 0.0001$). Mean squared error of Date-Flight accounted for approximately 27% of the overall variance in GSD. The remaining variance in GSD is attributed to Altitude which was 73% of the variance in GSD at the $P < 0.05$ level ($F_{1,132} = 2.64 \times 1,032$, $P < 0.0001$). These results suggest that it is appropriate to include Date-Flight as a random error term in the linear model correction of GSD as is reflected in Equation 8. The same per flight variance can be expected to influence the pixel length smoothing model (Eq. 9), justifying the inclusion of the Date-Flight term in Equation 10. The statistical model comparison results are corroborated by the lower mRMSE (Table S1), and the near zero bias of calibration object lengths (Fig. 2) in addition to the relatively narrow CIs relative to correction methods 2 and 3. These results suggest that a GSD correction including Date-Flight as an explicit error term is appropriate (e.g., M3 and M5).

Measurement error (RMSE) and bias (over/under estimation) are the primary metrics for examining model performance among the five alternatives in Table 2. M1 (uncorrected) resulted in the largest RMSE and bias of the five alternatives. The inclusion of the Date-Flight term reduced RMSE in M5 and M3 compared to M4 and M3, respectively. CIs of M2–M5 were larger than M1 due to model induced uncertainty. CIs for M4 and M5 were narrower than those for M2 and M3 because of the improved
predicted performance when using smoothed OLₚ estimates for calculating eGSD. Figure 3B further corroborates the improved estimation performance of including Date-Flight error term by the absence of the skew evident in the M2 and M4 observations. M5 has the lowest RMSE and bias of the alternatives as a result of accounting for pixel length variation when estimating the eGSD used to train the model as well as accounting for the Date-Flight effect. Operationally, these results suggest that per flight systematic sources of uncertainty (e.g., initialization height, swell, etc.) are influencing ASL (H') estimation at a level of significance that warrants continuing per flight imaging of a calibration object.

Comparison of Correction Methods on Gray Whale Measurements

Whale morphometric attribute estimates (Table 3) based on the five correction methods in Table 2 displayed similar trends as calibration object results above. Figure 3A shows estimated WL by correction method for 9 of the 89 whales imaged. The associated CV and mean of WL appear in Table S2. The no correction (i.e., uncorrected) method (M1) estimated a longer WL for six of the nine gray whales depicted in Figure 3A, which is consistent with the relationship between M1 and the
M2–M5 in the calibration data. The CV of the WLs for these six whales is <5%. However, for Whale 1, Whale 4, and Whale 5, WL in M3 is greater than WL in M4. This break in the trend is attributed to high CV (>5%) that is indicative of a significant and uncorrected digitization and/or ranging error. The mean CV of WL for the population of gray whales was 2.10% for M1 and 2.24% for M5. For these three whales, M1 WLs are not significantly different from M5. In contrast, the M1 WLs for the other six whales is always greater than M5. This trend is consistent with the relationship between M1 and M5 in the calibration object data (Fig. 2).

**Figure 3.** Estimated WL for each of the five GSD correction methods (Table 2) for each of nine arbitrarily selected gray whales (A) and each of the six blue whales imaged (B). Measurements for eight whales appear in Figure 3A because Whale 1 and Whale 2 are the same individual imaged over two flights; the same is true for Whale 3 and Whale 4. Bars are the 95% bootstrapped standard errors.
and provides evidence that M5 is appropriately correcting the morpho-
metric estimates.

Blue Whale Morphometric Measurement Correction Method
Comparisons

We conducted an identical analysis on the six blue whales surveyed in
New Zealand to demonstrate the applicability of the software tools to
another baleen whale species. WL estimates of each whale for the five
correction methods appear in Figure 3B. Methods 2–5 produced effectively
identical WL estimates for each whale which indicates a consistent zeroing
of the aircraft altitude and very little discernible bias in the digitization pro-
cess among whales and flights. These results further support the evidence
that Method 5 produced accurate results even when Date-Flight influences
are negligible. Eight individuals are depicted in Figure 3B because Whale
2 and Whale 4 are duplicates of Whale 1 and Whale 3, respectively. Whale
1 Method 1 WL was 18.77 m and Whale 3 Method WL was 18.20 m, simi-
larly, Whale 2 Method 1 estimated length was 18.33 m and Whale 4 Method
1 was 19.33 m. Differences were significant (at the $P < 0.05$ level) and just
outside of the 95% confidence intervals. Whales 3 is a reimage of Whale
1 during a different flight and Whale 4 is a resurvey of Whale 2. These sur-
veys were kept separate to illustrate the how a small number of image
observations ($n < 4$) of a single whale can produce significantly different
means with a false sense of certainty when measurements have low varia-
tion and undetectable bias, reinforcing the need to analyze a minimum of
five good frames (or images) per whale.

Effects of Lens Distortion

Camera distortion was estimated and quantified in terms of pixels and
compared to camera and lens models similar to those in Christiansen
et al. (2016) and Durban (2016) (Fig. 4A). All cameras evaluated exhib-
ited measurable distortion. When isolated to the cameras in this study,
Phantom 3 Pro (P3 Pro) exhibited the most severe distortion, followed
by the Phantom 4 (P4) with the Phantom 4 Pro (P4 Pro) having low dis-
tortion across most of the video frame. When expanded to incorporate
cameras and lenses like those used in previous studies, the GoPro Hero
4 Black (Christiansen et al. 2016) exhibited the most severe lens distor-
tion, with the PowerShot S100 (similar to Christiansen et al. 2016) and
Voigtlander (similar to Durban 2016) exhibiting more distortion than
either of the DJI P4 cameras. The distortion map for each of the DJI
cameras (Fig. 4B) demonstrate the importance of keeping the measure-
ment subject centered in the image frame, especially when using the P3
Pro and not correcting the image for distortion. However, applying dis-
tortion correction largely alleviates this concern because residual distor-
tion in corrected images was less than one pixel in 99% of pixels.

Total Propagated Uncertainty

The results of the total propagated uncertainty analysis appear in
Table 4. As has been observed in the previous results, the gray whale
data are more variable than the blue whale data. We attribute this difference in variability to two key differences: (1) blue whales tended to be centered and fully elongated in the images more frequently than gray whales, likely due to behavioral differences between species (foraging gray whales are more bendy than blue whales at the surface) and (2) the calibration reference used during the blue whale study was an object on the vessel at water level that was less susceptible to pitching and yawing from ocean swells than the 1 m board used for the gray whale study. The relative contribution of $\sigma^2_H$ was lower than $\sigma_{OL}$ for both gray whale and blue whale data, which indicates that the barometric altimeter used to estimate $H'$ is linear and relatively stable. The large OL values are a

Figure 4. Effect of camera specific lens distortion for the Phantom 3 Pro (P3 Pro), Phantom 4 (P4), and the Phantom 4 Pro (P4 Pro). (A) Estimated distortion coefficients for each camera, as well as cameras similar to those used in Christiansen et al. (2016) and Durban (2016). (B) is a lens distortion map in context of each DJI Phantom camera’s video frame. Although distortion increases gradually from frame center, the shaded areas are generalized to encompass the pixel area in an uncorrected video frame that contains five or fewer pixels of distortion related displacement.
result of analyst digitization error and poor quality images (e.g., glare, off center imaging). The variability in OLp is clearly discernible in the pixel length vs. altitude plot (Fig. S2). If bias in $H_0$ had not been corrected from Equation 7, $\sigma^2_{H_0}$ would have been more influential at 15 m than 40 m since the influence of $\sigma^2_{H_0}$ increases as bias/$H'_0$ increases.

**BAI Assessment**

Mean $R^2$ values of parabolas for gray whale ranged from 0.80 to 0.98, with a group mean $R^2$ of 0.92, suggesting that the parabola fit the analyst defined points well. One whale out of the 89 gray whales had a mean $R^2$ of 0.51. Low $R^2$, arbitrarily defined as below 25th percentile ($R^2 < 0.91$), was generally associated with poor whale edge visibility. BAI is scale invariant and thus unaffected by uncertainty associated with ranging error, so a comparison among correction methods was unnecessary. Pearson correlation coefficients for WL and BAI ($R^2 = 0.11$) is evidence that BAI is substantially more independent of WL than SA ($R^2 = 0.90$). A comparison of BAIIs between two whales having different lengths but similar SAs (Fig. 5) demonstrates the potential utility of BAI for making inference. A long but thin whale (Fig. 5A) has a much lower BAI and presumably less fat reserves than a short, thick whale (Fig. 5B).

We examined the sensitivity of BAI to detecting a 10% change in individual whale size using the change simulation scenarios in Table 1 and pairwise $t$-tests. Individual BAI change from BAI1 to BAI2, BAI3, BAI4, and BAI5 is discernible in 62.2%, 28.1%, 28.1%, and 27.0% of the 89 whales, respectively (Table 5). Despite the inconsistent performance of BAI for detecting individual change, using SA directly to discern change from SA1 to SA3, SA4, and SA5 was only successful for 17% of the whales (Table 6). SA2 was not evaluated because SA does not change in this scenario. The generally poorer performance in SA alone

**Table 4.** Results of the total propagated uncertainty analysis and supporting input parameters for Equation 14 in context to estimate the length of the calibration object (OL) from vertical imagery. Heading names refer to the respective data set (e.g., Gray = gray whale; Blue = blue whale) and numbers in the headings represent the altitude (e.g., 15 and 40 m) at which total propagated uncertainty was evaluated. Actual object length is abbreviated as OL'. % uncertainty is Total uncertainty/OL'.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Gray 15</th>
<th>Gray 40</th>
<th>Blue 15</th>
<th>Blue 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>H' (m)</td>
<td>15</td>
<td>40</td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>OLp</td>
<td>164.2</td>
<td>69.7</td>
<td>701.2</td>
<td>238.8</td>
</tr>
<tr>
<td>$\sigma_{OLp}$</td>
<td>51.5</td>
<td>19.4</td>
<td>37.5</td>
<td>12.9</td>
</tr>
<tr>
<td>$\sigma_{H_0}$</td>
<td>0.15</td>
<td>0.15</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>$\sigma^2_{OLp}$</td>
<td>99%</td>
<td>99%</td>
<td>97%</td>
<td>97%</td>
</tr>
<tr>
<td>$\sigma^2_{H_0}$</td>
<td>1%</td>
<td>1%</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>OL' (m)</td>
<td>1</td>
<td>1</td>
<td>4.41</td>
<td>4.41</td>
</tr>
<tr>
<td>Total uncertainty (m)</td>
<td>0.29</td>
<td>0.32</td>
<td>0.24</td>
<td>0.21</td>
</tr>
<tr>
<td>% uncertainty</td>
<td>29</td>
<td>32</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
is potentially the result of the added uncertainty from scaling (e.g., GSD error). However, the one case where SA is superior to BAI is distinguishing SA4 from SA5, which was more frequently discernible than BAI4 from BAI5. In terms of BAI, these two scenarios produce a nearly identical BAI but very different SAs. These results suggest that using BAI to detect 10% change in individual whale size may not be reliable but will likely perform better than using SA directly. However, the fact that SA could adequately capture the simulated difference between BAI4 and BAI5 suggests that the SA metric should not be negated from evaluation when examining individual trends.

BAI exhibited limitations when attempting to differentiate change across individuals (Fig. 6) as evidenced by the large SEs obscuring individual whale changes across the scenarios (Fig. 7). However, BAI did appear to perform well when examining change at the population level (Fig. S3). It is evident that the mBAI (gray dashed line Fig. 7) of scenarios BAI2–BAI5 are distinguishable from the mBAI of BAI1. The increased sensitivity at the population level is primarily a function of

<table>
<thead>
<tr>
<th>Whale</th>
<th>Whale Length (m)</th>
<th>60% Width (m)</th>
<th>Surface Area (m²)</th>
<th>Body Area Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td>11.66 +/- 1.31</td>
<td>1.15 +/- 0.21</td>
<td>7.61 +/- 1.59</td>
<td>34.98 +/- 0.80</td>
</tr>
<tr>
<td>b)</td>
<td>9.85 +/- 0.68</td>
<td>1.62 +/- 0.23</td>
<td>7.95 +/- 1.44</td>
<td>51.21 +/- 1.72</td>
</tr>
</tbody>
</table>

Figure 5. Visual comparison of body area index of a long gray whale with a low BAI (A), and a shorter gray whale with a high BAI (B).
Table 5. Comparison of body area index (BAI) for all five change scenarios in Table 1. Percentage is the ratio of the 89 whales exhibiting a significant change ($P < 0.05$ at 95% significance) in BAI between the specified body condition determined by the scenario annotated in the column and the scenario annotated in the row (i.e., column one, row one indicates that 65.2% of whales exhibited a statistically significant BAI change when body condition changed from the as measured condition to the simulated change determined by scenario BAI2.

<table>
<thead>
<tr>
<th></th>
<th>BAI2</th>
<th>BAI3</th>
<th>BAI4</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAI2</td>
<td>65.2%</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>BAI3</td>
<td>28.1%</td>
<td>77.5%</td>
<td>—</td>
</tr>
<tr>
<td>BAI4</td>
<td>28.1%</td>
<td>13.5%</td>
<td>70.8%</td>
</tr>
<tr>
<td>BAI5</td>
<td>27.0%</td>
<td>19.1%</td>
<td>70.8%</td>
</tr>
</tbody>
</table>

discriminatory power offered by the larger sample size (e.g., $n = 89$). However, there were limitations to the type of body composition changes that could be discerned. Body composition changes from BAI1 to BAI4 and change from BAI1 to BAI5 could not be differentiated from each other. This confusion is only possible when WL changes, which is limited to immature whales, but highlights the importance of collecting SA and WL.

DISCUSSION

This study develops and presents repeatable photogrammetric methods for measuring whale morphometrics from sUAS imagery, with robust examination of methods to reduce uncertainty, which is especially important when using low cost sUASs that have cameras with non rectilinear lenses and imprecise altimeters. The analytical backbone of this study is the two MATLAB programs and the summary tables produced from the Whale Quantitative Analysis program in R. The real power of this tool set is the ability to quickly turn image measurements into scaled metrics of

Table 6. Comparison of two dimensional body surface area (SA) for all five change scenarios in Table 1. Percentage is the ratio of the 89 whales exhibiting a significant change ($P < 0.05$ at 95% significance) in SA between the specified body condition determined by the scenario annotated in the column and the scenario annotated in the row (i.e., column one, row one indicates that 65.2% of whales exhibited a statistically significant SA change when body condition changed from the as measured condition to the simulated change determined by scenario BAI2.

<table>
<thead>
<tr>
<th></th>
<th>SA1</th>
<th>SA2</th>
<th>SA3</th>
<th>SA4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>SA3</td>
<td>16.9%</td>
<td>16.9%</td>
<td>59.6%</td>
<td>—</td>
</tr>
<tr>
<td>SA4</td>
<td>16.9%</td>
<td>16.9%</td>
<td>59.6%</td>
<td>—</td>
</tr>
<tr>
<td>SA5</td>
<td>16.9%</td>
<td>16.9%</td>
<td>70.8%</td>
<td>59.6%</td>
</tr>
</tbody>
</table>
Figure 6. Estimated BAI for the five simulated change scenarios described in Table 1. The nine whales depicted are the nine gray whales examined in Figure 3A.

Figure 7. The relationship between body area index (BAI) and estimated whale length (WL) for the gray whales (n = 89), ordered by the five change scenarios listed in Table 1. Headings denote the scenario number. The gray dashed line is estimated mean BAI (mBAI) for the scenario and is derived from the slope term in Equation 4. Error bars are the 95% CIs around the mBAI for each whale. Points missing bars have insufficient observations to derive a CI.
the 11 morphometric attributes listed in Table 3 (Whale Quantitative Analysis processing time for the 89 whales was < 180 s) and calculation of BAI for length independent comparisons of body condition. We are confident the combination of the three programs developed for this study will reduce time spent processing image data and increase time spent making ecological inferences on the resulting morphometric attributes. More importantly, this tool set offers the community an analytical method to standardize morphometric measurements and error estimation when using sUAS photogrammetric methods, increasing the continuity of data among studies and researchers.

In comparison to the aircraft used in Christiansen et al. (2015), Durban et al. (2015, 2016), and Dawson et al. (2017), the DJI Phantom series employed in this study are much smaller and cheaper aircraft. The smaller size of the DJI Phantom is inherently more compatible with space constraints associated with small vessel operations, and the low-cost makes the financial risk of a total loss more palatable to researchers with limited budgets. The P4 lens distortion is generally comparable to the optics of the other systems (Fig 4.); however, the 3.61 mm focal length of the P4 system requires much lower altitude flights to achieve a similar GSD as those reported in Durban et al. (2015, 2016) and Dawson et al. (2017). In terms of ranging accuracy, the DJI Phantom sUAS barometric altimeter is less precise than the LIDAR employed by Dawson et al. (2017). In terms of ranging accuracy, the DJI Phantom sUAS barometric altimeter is less precise than the LIDAR employed by Dawson et al. (2017), yet the mean coefficients of variation for the WL of gray whales and blue whales were 2.24% and 3.64%, respectively. These are in the range of those reported by both Durban et al. and Christiansen et al., but larger than the mean of 1% reported by Dawson et al. For sUAS photogrammetry systems without LIDAR, measurement error can be reduced through adherence to the flight protocol provided below.

Based on our results, it is imperative that studies account for ranging error with some type of calibration object. Unrected object lengths contained substantial bias (Fig. 2) that exceeds the bias of the calibration object reported in Durban et al. (2015), although their method used a much longer calibration object that is more robust to confounding scaling caused by movement on sea surface, and the precision of the barometric altimeter was likely superior to that of the DJI Phantom. Even when using a LIDAR for range estimation we recommend using an independent calibration object to make estimates on measurement bias and RMSE. The CV metric that is frequently used in whale photogrammetry (e.g., Jaquet 2006, Christiansen et al. 2016, Durban et al. 2016, Dawson et al. 2017) is useful for understanding consistency of measurements, but CV does not quantify error since it does not require an absolute reference.

The results from the M5 estimates compared to the other four correction methods further suggest that smoothing pixel lengths of the calibration object prior to creating the GSD correction model results in less erroneous estimates of scaled length. While there were instances where M5 estimates of WL were not significantly different from WL estimates in M1, these exceptions tended to be associated with high levels of
variability in the observations. High variability in the observations is likely a function of nonstrict adherence to the optimal imaging recommendations presented by Christiansen et al. (2016) (e.g., whale not centered in the camera during flight), but the error associated with off-center imaging is partially mitigated by employing our developed lens distortion corrections. Although developed for DJI Phantom series systems, the software is designed to ingest any camera calibration profile created with MATLAB's Camera Calibrator application. It is likely that higher precision barometric (Durban et al. 2015) or laser altimeters (Dawson et al. 2017) would reduce the dependency on calibration objects, but integrating nonstandard instruments onto commercially built sUASs reduces flight time, and increases complexity.

Although lens distortion can induce significant error into measurements, it is systematic and thus the effects are correctable inside of the far edges of the video frames. The distortion calibrations for DJI P3, DJI P4, and DJI P4 Pro are incorporated into the freely distributed MATLAB software and heavily minimizes the impact of off-center imaging. However, corrections do typically crop the image to an extent, so effort should still be taken to center the subject in the frame whenever possible. Additionally, the results show that although low in cost, the P4 Pro has a very low distortion lens and thus we would recommend that spendthrift researchers consider this platform when deciding between discounted P3 and P4 or paying full price for the P4 Pro.

The analysis of total propagated uncertainty shows that digitization error remains the largest source of uncontrolled error, reinforcing the need for multiple observations per whale and ensuring images are of high quality. Our uncertainty analysis also shows the importance of using a calibration object to correct altitude-related bias in GSD estimates. These findings lead us to recommend that surveys should be conducted at the highest reasonable altitude to achieve the measurement objective because the influence of altitude error on GSD reduces as the altitude error/absolute altitude ratio decreases.

We condensed our recommendations in the form of an uncertainty mitigation protocol (Table 7) as a convenience for what we believe will be a rapidly growing community of whale photogrammetrists. This protocol is not specific to our study or even whales, but rather, is broadly applicable to any study where the subject is a surfacing animal and the survey aircraft is a sUAS with a nadir pointing camera.

The final objective of this study was to develop and present a length-independent body condition metric, which we term Body Area Index (BAI), which facilitates comparison of whale body condition over time, among and between populations. Results show that BAI is more independent of WL than SA and is sensitive to population level changes in WL and SA at the 10% threshold, with exception of situations where the population is one that contains individuals growing in length between measurements.

The length normalized characteristic of BAI makes it a useful metric for making inference about individual fat reserves in relation to each other (Fig. 5–6) as well as individual fat reserves relative to the population mean (gray dashed line in Fig. 7). The status of calves can be similarly
Many of the whales <10 m in length, which are presumed to be calves or sub adults, have mean BAIs equal to or higher than the population mean (gray dashed line in Fig. 7) suggesting that these whales have slightly elevated fat reserves compared to the population.

The scale invariant property of BAI is especially valuable in surveys where scaling error cannot be controlled with a calibration object or through the implementation of a high precision LIDAR, as was evidenced by the increased sensitivity of BAI to detecting change in body size compared to using body surface area directly. The results of the individual BAI change sensitivity analysis were inconclusive because change could be detected in some whales and not others. We attributed this inconsistency to high within whale BAI variability relative to the low sample size (e.g., fewer than five images per whale). When individual change detection is necessary, we recommend conducting multiple flights over the same whale and performing a power analysis to determine how many observations (e.g., images/frames) will be necessary to discern change at the desired level of sensitivity. The other advantage to BAI is simply dependent on minimally distorted vertical images of the subject and thus utility is independent of camera, UAS, or ranging precision.

A potential limitation to BAI is the underlying assumption of the parabolic shape that is used to estimate surface area. We demonstrate that the parabolic model used to estimate the surface area, which is fundamental to BAI estimation, accounted for >80% of width variations across evaluated. Many of the whales <10 m in length, which are presumed to be calves or sub adults, have mean BAIs equal to or higher than the population mean (gray dashed line in Fig. 7) suggesting that these whales have slightly elevated fat reserves compared to the population.

Table 7. Uncertainty Mitigation Protocol based on recommendations in literature and best practices learned during field operations.

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Power up (i.e., initialize) the sUAS from the same location on the watercraft every time to minimize influence of random error in ranging uncertainty.</td>
</tr>
<tr>
<td>2.</td>
<td>Measure the vertical distance between power–up location and water level and add that distance to reported altitudes to account for bias in ranging caused by initializing above sea level (Durban et al. 2015).</td>
</tr>
<tr>
<td>3.</td>
<td>As much as feasible, image over flat water in nonwindy to minimize ranging uncertainty.</td>
</tr>
<tr>
<td>4.</td>
<td>Only measure images/frames where the whale is at water surface, fully elongated with no curvature (Perryman and Lynn 1993, Fearnbach et al. 2011, Christiansen et al. 2016) to minimize digitization error.</td>
</tr>
<tr>
<td>5.</td>
<td>Measure 5+ images of the same subject from each flight to evaluate variation.</td>
</tr>
<tr>
<td>6.</td>
<td>Keep subject centered in the camera to minimize error associated with lens distortion, scale nonuniformity induced by camera tilt error, and geometric distortion from non orthogonal viewing.</td>
</tr>
<tr>
<td>7.</td>
<td>Image a calibration object every flight. Object should be rigid and located as close to sea level and as long as possible. Longer objects are more robust to scaling error caused by vertical motion in ocean swell.</td>
</tr>
<tr>
<td>8.</td>
<td>Image from the highest safe and legal altitude that ensures adequate level of detail, to minimize the influence of altitude error on GSD estimation.</td>
</tr>
<tr>
<td>9.</td>
<td>Apply a GSD correction model like that in Equation 12 that accounts for altitude variances on a per flight basis.</td>
</tr>
</tbody>
</table>
the length of all whales but one. The exceptional whale did not exhibit signs of emaciation or pregnancy. The poor fit was attributed to only having a single measurement on this whale, poor whale outline visibility, and the whale being located in the lower right corner of the video frame. Pregnant or severely emaciated whales may present forms that deviate from that of a parabola and the $R^2$ of associated parabolas would likely be lower. However, low $R^2$ was also attributed to poor edge visibility, so low $R^2$ should only be used as a guide for discerning when SA and BAI metrics are reasonably valid. Further investigation is necessary to discern the appropriateness of the method on those individuals with atypical body shapes. The method presented here would potentially be improved upon by incorporating convex hull algorithms (Barber et al. 1996) or automatic segmentation algorithms (Misimi et al. 2008).

Future studies will determine the broader applicability of our provided framework, but we are confident that similar results can be achieved on any species exhibiting similar body shape characteristics as gray and blue whales. Although our method was developed on the DJI Phantom series, the methods presented are applicable to any sUAS-based whale photogrammetric campaign where images are taken at a known altitude. We expect that future studies will focus on automatic whale edge delineation in images and further investigate the applicability of BAI for ecological inference.

**Conclusion**

This study presents a length normalized body size index (BAI) that facilitates comparison among individuals and populations to describe trends in body condition. Additionally, we examined the effectiveness of models used to correct error in scale image measurement and determined the most precise and accurate model was a LMM containing a Date-Flight error term and regressed on eGSD values that were derived from smoothed pixel length estimates. We subsequently determined that analytical digitization error was the largest source of uncertainty in scaled measurement estimates and developed an Uncertainty Mitigation Protocol to help future studies avoid controllable sources of uncertainty. We also developed a three program analytical suite for obtaining 11 morphometric attributes of free swimming baleen whales from vertical sUAS imagery. Our findings suggest that sUAS photogrammetry from a DJI Phantom 3 Pro and Phantom 4 is a precise method to assess baleen whale body size when there are sufficient observations of an individual whale and uncertainty from ranging error is controlled by imaging an object of known length every flight.

**ACKNOWLEDGMENTS**

Gray whale overflights were authorized under NOAA/NMFS permit #16011 issued to John Calambokidis. We are thankful for support by Brazil's Science Without Borders program, Brazil's CNPq, and the Harvard Laspau Institute for LL. Special thanks to Dr. Chris Parrish for editing and advice on analyzing
photogrammetric accuracy. We also acknowledge the contributions of Alessandra Hossley, Ashley Leupold, and Taylor Mock for their efforts in data processing.

**LITERATURE CITED**


condition and growth in right whales *Eubalaena glacialis* and *E. australis*. Marine Ecology Progress Series 459:135–156.


Received: 9 May 2017
Accepted: 3 May 2018
The following supporting information is available for this article online at http://onlinelibrary.wiley.com/doi/10.1111/mms.12527/suppinfo.

Figure S1. Output image from Whale Measurements program displaying the morphometric attributes that are measured to include fluke width (FW) and whale length (WL). Parabolas fit by Equation 1 are also depicted. Optimized Width (OW) is described in Table 3 and the non-orthogonality highlights the susceptibility of individual measurements to interpretive error.

Figure S2. The relationship between calibration object length in pixels and altitude. (a) Depicts a wide variation in object length around a single altitude when altitude is not corrected for bias. The relationship between altitude and empirical ground sampling distance (eGSD) (b) shows the wide range in eGSDs observed from a single altitude when altitude bias is uncorrected.

Figure S3. Mean estimated body area index (BAI) for the entire population of gray whales \( (n = 89) \) in the context of each of the simulated whale length and surface area change scenarios as specified in Table 1.

Table S1. Mean estimated calibration object length and supporting metrics resulting from each of the scaling error correction methods described in Table 2. mRMSE is the mean of the root mean squared errors Equation 4 across all the flights contributing to the estimate of the mean, CV\% is the coefficient of variation in units of percentage. Mean.lwr and Mean.upr designate the lower and upper bounds of the 95% confidence intervals on the mean.

Table S2. Results of the five correction methods listed in Table 2 for analysis of the nine whales identified in Figure 3. Whale Length (WL) mean is the mean estimated WL and WL coefficient of variation (CV) is the CV of the individual observations that contributed to the mean.

APPENDIX S2

Whale Measurements and Whale Calibration Object Measurement MATLAB programs: Whale_Measurements.m and Whale_Calibration_Object_Measurement.m are packaged within a ZIP file containing supporting scripts.

Whale Quantitative Analysis R program: Whale_Quantitative_Analysis.r is a single R script file.

These programs are available from Leigh Torres (e-mail: leigh.torres@oregonstate.edu).