



Bayesian approach for predicting photogrammetric uncertainty in morphometric measurements derived from drones

K. C. Bierlich^{1,2,*}, R. S. Schick¹, J. Hewitt³, J. Dale¹, J. A. Goldbogen⁴,
A. S. Friedlaender⁵, D. W. Johnston¹

¹Division of Marine Science and Conservation, Nicholas School of the Environment, Duke University Marine Laboratory, Beaufort, North Carolina 28516, USA

²Marine Mammal Institute, Department of Fisheries, Wildlife, & Conservation, Oregon State University, Hatfield Marine Science Center, Newport, Oregon 97365, USA

³Department of Statistical Science, Duke University, Durham, North Carolina 27708, USA

⁴Department of Biology, Hopkins Marine Station of Stanford University, Monterey, California 93950, USA

⁵Institute of Marine Sciences, Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, California 95604, USA

ABSTRACT: Increasingly, drone-based photogrammetry has been used to measure size and body condition changes in marine megafauna. A broad range of platforms, sensors, and altimeters are being applied for these purposes, but there is no unified way to predict photogrammetric uncertainty across this methodological spectrum. As such, it is difficult to make robust comparisons across studies, disrupting collaborations amongst researchers using platforms with varying levels of measurement accuracy. Here we built off previous studies quantifying uncertainty and used an experimental approach to train a Bayesian statistical model using a known-sized object floating at the water's surface to quantify how measurement error scales with altitude for several different drones equipped with different cameras, focal length lenses, and altimeters. We then applied the fitted model to predict the length distributions and estimate age classes of unknown-sized humpback whales *Megaptera novaeangliae*, as well as to predict the population-level morphological relationship between rostrum to blowhole distance and total body length of Antarctic minke whales *Balaenoptera bonaerensis*. This statistical framework jointly estimates errors from altitude and length measurements from multiple observations and accounts for altitudes measured with both barometers and laser altimeters while incorporating errors specific to each. This Bayesian model outputs a posterior predictive distribution of measurement uncertainty around length measurements and allows for the construction of highest posterior density intervals to define measurement uncertainty, which allows one to make probabilistic statements and stronger inferences pertaining to morphometric features critical for understanding life history patterns and potential impacts from anthropogenically altered habitats.

KEY WORDS: Bayesian modeling · Photogrammetry · Uncertainty analysis · Drones · Unoccupied aircraft system · UAS · Morphometrics · Marine mammals · Cetaceans

1. INTRODUCTION

The morphology of an animal is one of the most fundamental factors affecting its habitat use and foraging performance, and can reflect details of the current

health, likelihood of survival, and potential reproductive success of an individual (Arnold 1983, Irschick 2003). Collecting accurate morphological measurements of individuals is often essential for monitoring populations, and recent studies have demonstrated

*Corresponding author: kevin.bierlich@oregonstate.edu

how animal morphology has profound implications for conservation and management decisions, especially for populations inhabiting anthropogenically altered environments (De Meyer et al. 2020, Miles 2020). However, obtaining accurate morphometric measurements of marine megafauna is challenging, particularly for cetaceans, as they are often found in remote locations, spend little time at the surface of the water, and their large size can preclude safe capture and live handling (Johnston 2019).

Recently, unoccupied aircraft systems (UAS), or drones, have proven to be a valuable, non-invasive tool for collecting high-resolution photogrammetric data on cetaceans across a variety of ecosystems. Durban et al. (2015) first demonstrated the utility of using UAS for acquiring morphometric measurements of killer whales *Orcinus orca* in a remote location with limited occupied aircraft support. Since then, UAS have been used from tropical to polar environments and applied to several cetacean species of vastly different sizes and body shapes, including blue whales *Balaenoptera musculus*, humpback whales *Megaptera novaeangliae*, southern right whales *Eubalaena australis*, North Atlantic right whales *E. glacialis*, gray whales *Eschrichtius robustus*, fin whales *B. physalus*, Antarctic minke whales *B. bonaerensis*, and Bryde's whales *B. brydei* (Christiansen et al. 2016, 2018, 2020a,b, Durban et al. 2016, Gough et al. 2019, Lemos et al. 2020). However, a broad range of UAS platforms, sensors, and altimeters were applied in these studies, and there is no unified way to predict photogrammetric uncertainty across this methodological spectrum. As such, it is difficult to make robust comparisons among studies, disrupting collaborations amongst researchers using platforms with varying levels of measurement accuracy. As the capacity to collect morphometric data on various species via UAS continues to grow, there is a need for standardization of measurements across studies and minimization and quantification of errors (Castrillon & Bengtson Nash 2020). This will ultimately build a greater capacity to better monitor populations exposed to a variety of environmental and anthropogenic stressors.

Traditional methods for acquiring morphometric measurements of cetaceans have previously been limited to carcasses collected from scientific whaling operations (Ichii & Kato 1991) or opportunistically from commercial whaling (Lockyer 1981, Christiansen et al. 2013), subsistence hunting (Lambertsen et al. 2005), stranding events (Palacios et al. 2004), and bycatch (Read 1990, Koopman et al. 2002). Aerial photogrammetry from occupied aircraft was adopted as a non-invasive technique for estimating the length of free-

living cetaceans by Whitehead & Payne (1981). Photographs of southern right whales next to a boat with a known-sized disc were collected to set the scale of the photograph (Whitehead & Payne 1981). This method was later enhanced by using altimeters to record the altitude and set the scale of the photograph by calculating the ground sampling distance (GSD), the distance on the ground that each pixel represents (Cubbage & Calambokidis 1987, Best & Rüther 1992). These methods have been commonly used to obtain length measurements of odontocetes and mysticetes (Perryman & Lynn 1993, Ratnaswamy & Winn 1993, Perryman & Westlake 1998, Fearnbach et al. 2011, 2018), as well as width measurements to assess nutritive condition related to reproduction in gray whales (Perryman & Lynn 2002) and southern and North Atlantic right whales (Miller et al. 2012).

Compared to occupied aircraft, UAS have greatly enhanced opportunities to more efficiently collect high-resolution aerial photogrammetric data on cetaceans, as they offer a more affordable (Arona et al. 2018) and immediate option that is less limited by weather and infrastructure (e.g. Cosens & Blouw 2003, Fearnbach et al. 2011), and, importantly, presents less risk to wildlife biologists (Sasse 2003). However, many studies use different UAS platforms equipped with various cameras and focal length lenses, which have inherent differences in lens distortion (i.e. Burnett et al. 2019) and GSD (see Fig. 1). These different platforms and sensors vary in performance, quality, and photogrammetric accuracy; to date, a rigorous analysis of the impact of these factors has not been undertaken.

In the USA, the National Institute of Standards and Technology (NIST) provides guidelines put forth by the International Committee for Weights and Measures (CIPM) for how to report measurements and uncertainty, as a measurement result is complete only when accompanied by a quantitative statement of its uncertainty (Taylor & Kuyatt 1994). In the context of collecting aerial imagery of cetaceans for morphometric analysis, measurement errors can be introduced by environmental conditions (e.g. glare, wave refraction, water clarity) and animal behavior (e.g. curved vs. straight body at depth/at surface). These errors are largely uncontrollable, but they can be mitigated, such as by avoiding flying during times with high glare and filtering for high-quality images where the animal is straight at the surface with minimal refraction from waves (Christiansen et al. 2018, Burnett et al. 2019, Raoult et al. 2020). It then becomes important to understand components in the UAS photogrammetric workflow that can be controlled to

minimize systematic errors associated with the UAS platform and analysis.

In general, assessment of measurement error has focused on measuring the length of a known-sized object either on land (Best & Rüther 1992, Perryman & Lynn 1993, Christiansen et al. 2018), floating at the surface (Perryman & Lynn 2002, Burnett et al. 2019, Gough et al. 2019, Kahane-Rapport et al. 2020), or on a research vessel (Durban et al. 2015, 2016) to quantify errors associated with altimeters and analyst measurements. These different forms of data collection can lead to significant biases in reported measurement errors associated with altimeters on UAS, which can have different accuracies depending on altitude and ground elevation, and whether the UAS are flown over land or sea (Dawson et al. 2017). Many studies have accounted for altimeter error by using correction coefficients from a linear equation (Best & Rüther 1992, Perryman & Lynn 1993) or by applying a percent error to all measurements (Ratnaswamy & Winn 1993), e.g. error value of <5% for all measurements. Since the true length of the whale is unknown, the coefficient of variation (CV) is often used to determine the within-frame and between-frame measurement precision. Within-frame precision compares measurements of the same whale in a single image multiple times, often by multiple analysts, while between-frame precision compares measurements of the same whale between different images (Cosens & Blouw 2003).

Recently, Christiansen et al. (2018) used a frequentist statistical approach to first measure a known-sized object at altitudes between 5 and 120 m and then used resampling methods to build an error distribution around measurements of southern right whales described by laser altimeter error, image quality, and the CV of within- and between-image measurements. Similarly, Burnett et al. (2019) used known-sized objects to first independently estimate the variance around altitude recorded from barometers on small low-cost UAS and measured length, and then estimated total measurement error via variance propagation. While these methods greatly improve error estimation, both assume that errors are independent and adequately described by their standard deviations and CVs (Christiansen et al. 2018, Burnett et al. 2019). The apparent length of an object is dependent on altitude, and errors may have more complicated features (e.g. skew, and heavy tails or outliers). Methods that account for these issues may further improve estimates of measurement error across UAS platforms, and facilitate error propagation to additional, derived quantities.

Here, we built off the studies of Burnett et al. (2019) and Christiansen et al. (2018) and developed a Bayesian statistical model that propagates the combined impact of measurement errors to UAS photogrammetric measurements and derived quantities, such as length-based age classifications (e.g. juvenile/adult). Similar to Racine-Poon (1988), we used a designed experiment to generate training data for the Bayesian statistical model. The experiment used a known-sized object floating at the surface to study how measurement error scales with altitude for several different UAS platforms equipped with different cameras, focal length lenses, and altimeters. We then applied the fitted model to 2 ecological scenarios in which we (1) predicted the length and measurement uncertainty around unknown-sized humpback whales to assign maturity classification, and (2) predicted the population-level morphological relationship between rostrum to blowhole distance and total length of Antarctic minke whales. The Bayesian model offers several components that improve uncertainty predictions. First, the model provides a framework that jointly estimates errors from altitude and length measurements. Second, the model combines altitude measured with a barometer and laser altimeter while incorporating the different errors specific to each. Third, rather than a single point-estimate, the model outputs a posterior predictive distribution of measurements around an object of unknown length (e.g. a whale). Fourth, this approach allows for the construction of highest posterior density (HPD) interval to define measurement uncertainty, which allows one to make probabilistic statements and reach stronger conclusions, e.g. classification of the maturity of an animal based on its estimated length. Finally, we provide a framework that can easily incorporate additional modeling layers. For example, at the data collection level, the effect of additional covariates on measurement error can be studied; or, at the scientific level, morphological relationships can be jointly estimated alongside measurement error.

2. MATERIALS AND METHODS

2.1. Calculating Ground Sampling Distance

The GSD sets the scale of the photo in order to convert measurements made in pixels into standard units (i.e. meters) by using the following equations (see Torres & Bierlich 2020 for review):

$$\text{GSD} = \frac{a}{f_c} \times \frac{S_w}{I_w} \quad (1)$$

$$\text{Length} = \text{GSD} \times L_p \quad (2)$$

where a is altitude, which is the distance (m) from the camera to the object of interest (i.e. the whale is assumed to be flat at the surface of the water), f_c is the focal length of the camera (mm), S_w is the sensor width of the camera (mm), I_w is the image width (pixels, px), and L_p is the length (px) of the object of interest. As altitude increases, GSD also increases, ultimately decreasing the resolution of the image, which can influence L_p measurement accuracy as specific features and edges of the whale become more difficult to identify (see Figs. 1 & 2). Furthermore, different methods of data collection (i.e. collecting still images vs. video) will also change the GSD for the same camera, as video uses fewer pixels in the image width (I_w) compared to still images, ultimately decreasing resolution and increasing the GSD (Eq. 1, Fig. 1).

In Eq. (1), the camera parameters (S_w , I_w , f_c) are fixed and can be accounted for. The other parameter, altitude (a), has the greatest influence on determining GSD. This is potentially a major source of measurement error due to discrepancies in the correct scaling of the pixels in the image caused by altimeter errors (Burnett et al. 2019). All UAS are equipped with a barometer, which is a pressure sensor for recording altitude, but each will have some level of inaccuracy when measuring and recording true altitude. Recording the offset from launch height to the water is critical, as the barometer sets the zero pressure at the takeoff point. Flying from land also adds complications with allowance for altitude offsets due to tidal fluctuations. Generally speaking, lower accuracy usually arises from the use of low-cost sensors commonly found on small UAS (Wei et al. 2016). Sev-

eral studies have improved the accuracy of the altitude recorded by using a laser altimeter, e.g. LightWare SF11/C LIDAR (Dawson et al. 2017), or by measuring a known-sized object during each flight to correct for barometric altitude inaccuracies using linear equations (Burnett et al. 2019, Lemos et al. 2020).

2.1.1. Training and calibration data (x)

Seven UAS flights were conducted on 26 June 2019 at the Duke University Marine Lab in Beaufort, NC, USA (34° 43' 0.156" N, 76° 40' 24.42" W) and are detailed in Table 1. Each aircraft was launched from a dock ~100 m from a known-sized floating calibration object made of PVC pipe (1.48 m × 1.15 m). The calibration frame was foam-filled to maintain flotation at the surface of the water and anchored via rope with a buoy to prevent drifting. The site was chosen because it is sheltered from ocean swell and thus ensured the calibration frame remained relatively flat at the surface during data collection. Prior to takeoff, the launch height was measured from the water surface to the camera lens and then later added to the recorded barometer altitude to account for the bias introduced from the barometer zeroed at launch height (see Durban et al. 2015) and the local rising tide. Each aircraft collected imagery of the calibration frame at altitudes between 10 and 120 m in 10 m increments, similar to Christiansen et al. (2018). For the P4Pro flight with video, a still frame was captured using the snapshot function in VLC Media Player Software (Version 3.08, VideoLAN), as in Lemos et al. (2020). To test for possible effects of lens distortion at various altitudes, the calibration frame

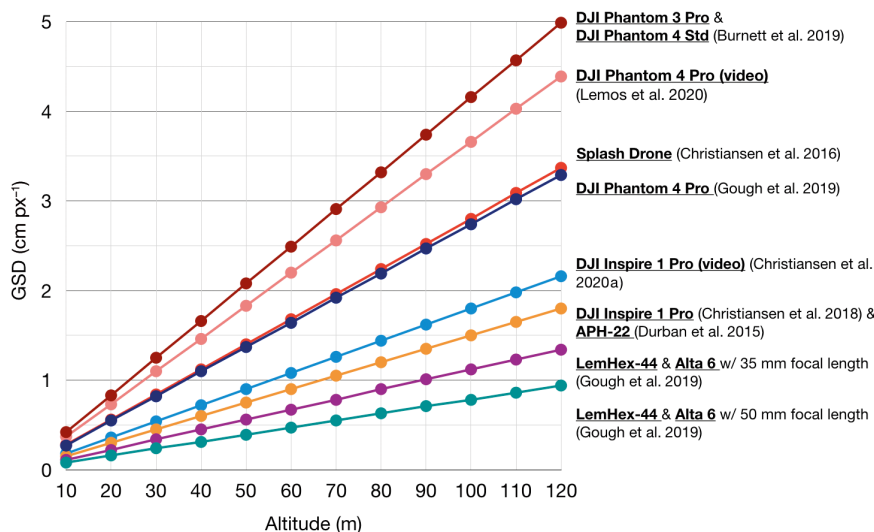


Fig. 1. Ground sampling distance (GSD) for unoccupied aircraft system (UAS) platforms commonly used for cetacean photogrammetry. The GSD displayed is exact, meaning it does not account for distortion or altitude errors. GSD increases with increasing altitude, lowering image resolution. See Table 1 for list of UAS platforms used in this current study; px: pixel

Table 1. Camera specifications for each flight and aircraft. All aircraft contained a barometer and laser altimeter, except for Phantom 3 Std. The Olympus EPM2 camera has also been used on other UAS platforms, e.g. APH-22 (Durban et al. 2015, 2016). px: pixel

Flight	Aircraft	Data collected	Altimeter	Camera	Focal length (mm)	Camera sensor (mm)	Image dimensions (px)	Pixel dimensions (mm px ⁻¹)	f-number
1	Alta 6	Images	Barometer & Laser	Sony a5100	50	23.5 × 15.6	6000 × 4000	0.00392 × 0.00390	f1.8
2	Alta 6	Images	Barometer & Laser	Sony a5100	35	23.5 × 15.6	6000 × 4000	0.00392 × 0.00390	f1.8
3	Lemhex-44	Images	Barometer & Laser	Sony a5100	35	23.5 × 15.6	6000 × 4000	0.00392 × 0.00390	f1.8
4	CineStar	Images	Barometer & Laser	Olympus EPM2	25	17.3 × 13	4608 × 3456	0.00375 × 0.00376	f1.8
5	Phantom 4 Pro	Images	Barometer & Laser	Phantom 4 Pro	8.8	13.20 × 8.80	5472 × 3078	0.00241 × 0.00286	f/2.8–f/11
6	Phantom 4 Pro	Video	Barometer & Laser	Phantom 4 Pro	8.8	13.20 × 8.80	4096 × 2160	0.00322 × 0.00407	f/2.8–f/11
7	Phantom 3 Std	Images	Barometer	Phantom 3 Std	3.61	6.16 × 4.62	4000 × 3000	0.00154 × 0.00154	f/2.8

was positioned in the center of the image during the ascent and in the corner during the descent. We also tested the distortion of each camera using the MathWorks Single Camera Calibrator App in MATLAB, following the provided tutorial (MathWorks 2017) and Burnett et al. (2019).

In addition to an onboard barometer, each aircraft except for the P3Std was equipped with a LightWare SF11/C laser altimeter that simultaneously recorded altitude along with the barometer (Table 1). A custom designed housing was created to support and power the laser altimeter on the P4Pro (installation instructions at https://github.com/marrs-lab/DJI_PH4_LaserAltimeter). The LightWare SF11/C laser altimeter is rated for altitude measurements up to 120 m above land and 40 m above moving water with ± 0.1 m of error (LightWare Optoelectronics 2018). Each platform contained the laser altimeter and camera co-located on a 2-axis gimbal with pitch angle controlled via remote control to ensure image collection at nadir, except for P4Pro, which had the laser fixed on the aircraft frame, and we accounted for the pitch and roll to calculate the vertical altitude during image collection (Dawson et al. 2017). Timestamp drift can lead to improper scaling of images due to incorrect altitudes used in calculating GSD (Voges et al. 2018, Raoult et al. 2020). To correct for this, we took images of an iPad screen connected to a Bad Elf GPS unit displaying the current GPS time prior to take off. This ensures we can accurately link the timestamp of an image to the altitude recorded by the barometer and laser altimeter.

2.1.2. Photogrammetry

The length of the calibration frame was measured in pixels (L_p') using the straight-line tool in ImageJ 1.5i (Schneider et al. 2012) by 3 separate analysts (2 considered 'expert', one considered 'novice'). To assess measurement bias between analysts, the coefficient of variation (CV%) was calculated between the 3 L_p' measurements for each image, following the approach of Christiansen et al. (2016).

To compare the measurement error amongst different UAS platforms, i.e. to derive the uncorrected measurement error, we used Eqs. (1) and (2) to convert each L_p' measurement into 2 length measurements (m), one using the altitude recorded from the barometer and one from the laser altimeter. The uncorrected percent error was then calculated as:

$$\text{uncorrected \% error} = \frac{(L_{p,ijk}' - L_{co})}{L_{co}} \times 100 \quad (3)$$

where L_{co} is the true length of the calibration object (1.48 m) and $L_{p,ijk'}$ is the length (px) of the calibration object in each image i , measured by analyst j , using altimeter k (barometer or laser) (Christiansen et al. 2018, Burnett et al. 2019).

2.2. Model development

2.2.1. Overview of Bayesian approach

Unlike frequentist statistical theory, the Bayesian approach views both data and the underlying parameters (i.e. variances) that generated the data as random (see Austin et al. 2002 and Ellison 2004 for reviews). Using Bayes' theorem, a model of the observed data, called the likelihood function, is combined with prior knowledge pertaining to the underlying parameters, called the prior probability distribution, to form the posterior probability distribution. The posterior probability distribution serves as updated knowledge about the underlying parameter and can be used as prior information for subsequent studies.

Following this framework, we first estimated the posterior probability distribution for a vector of photogrammetric error parameters (θ) from different UAS platforms using calibration data (x) of a known sized object via:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} \quad (4)$$

where $f(x|\theta)$ is the likelihood function, $f(\theta)$ is the prior probability distribution that defines the potential range for θ , $f(x)$ is the marginal distribution of the measurement data, and $f(\theta|x)$ is the posterior distribution that defines the likely range of θ given data x . We designed the likelihood function to account for sources of UAS measurement error (see Section 2.2.2). Taking a similar approach as Racine-Poon (1988), we then used the posterior probability distributions for θ , specific to UAS platform, as prior information to form posterior predictive distributions for measurements of an unknown-sized object (L_{new}) from observations (x_{new}) collected in the field (e.g. of whales) using that UAS platform via:

$$f(L_{new}|x, x_{new}) = \int f(L_{new}|\theta, x_{new}) f(\theta|x) d\theta \quad (5)$$

where $f(L_{new}|\theta, x_{new})$ is a conditional likelihood, and $f(\theta|x)$ is the posterior probability distribution from Eq. (4), which effectively serves as an updated prior probability distribution. The posterior predictive distribution $f(L_{new}|x, x_{new})$ quantifies uncertainty in measurements of an unknown-sized object based on the

experimental study of measurement errors for a given UAS platform.

2.2.2. Error estimation

We used Eqs. (1) and (2) to design the likelihood function. We assumed that L , a , and L_p represent the length of a target object, the exact altitude at which an image is taken, and the length of the object in the image, in pixels, respectively. The data $x = (a_L', a_B', L_p')$ are the altitude as measured by a laser altimeter (a_L') and barometer (a_B'), and the measured length of the object in the image, in pixels (L_p'), all of which represent noisy versions of the exact values a and L_p , respectively. We assigned a uniform prior distribution to a (min = 5 m and max = 130 m), which restricts the model to the altitude range of UAS image collection but is otherwise uninformative. We expect most altimeter measurement errors are smaller than a few meters at worst, and that the errors are normally distributed around the true altitude a such that:

$$a_L' \sim N(a, \sigma_L^2) \quad (6)$$

$$a_B' \sim N(a, \sigma_B^2) \quad (7)$$

The altimeters have separate variance parameters σ_L^2 and σ_B^2 , to each of which we assign an inverse gamma prior distribution (shape = 2, rate = 1). Measurement error is assumed to have constant variance over unknown altitudes, which simplifies modeling and estimation. Impacts of the constant variance assumption are evaluated in model validation (Section 2.3.1).

We also assumed the measured length of the object, in pixels, L_p' is normally distributed around the exact value L_p by specifying the prior distribution:

$$L_p' \sim N(L_p, \sigma_{L_p}^2) \quad (8)$$

to which we assigned an inverse gamma prior distribution (shape = 5, rate = 4) for the variance parameter $\sigma_{L_p}^2$. Eqs. (1) and (2) imply that the exact length of the object in pixels, L_p , depends on the length L of the target object and the altitude at which the image is taken (a) via:

$$L_p = \frac{L \times f_c \times I_w}{a \times S_w} \quad (9)$$

The length L of the target object is known in the calibration experiment, yielding the likelihood:

$$f(\theta|x) = \int f(a) f(a_L'|a, \theta) f(a_B'|a, \theta) f(L_p'|a, \theta) da \quad (10)$$

in which $f(a)$ is the uniform prior density for the true altitude, $f(a_L'|a, \theta)$ and $f(a_B'|a, \theta)$ are the densities for

the altimeter measurement errors (Eqs. 6 and 7), $f(L_p|a, \theta)$ is the measurement error distribution of the pixel length (Eq. 8), the parameter vector is $\theta = (\sigma_L^2, \sigma_B^2, \sigma_{LP}^2)$, and the calibration object length L_{co} is substituted for L in Eq. (9). We then can use measurements of L_{co} as training data to estimate error parameters θ specific to UAS platforms.

2.2.3. Measurement predictions

After estimating θ , we can use the posterior predictive distribution (Eq. 5) to make inferences about the size of an unknown object (L_{new}), e.g. length of a whale, which is conditional on a new set of measurements $x_{new} = (a_{L,new}', a_{B,new}', L_{p,new}')$ and our error parameter estimates from the UAS platforms used in data collection. The conditional likelihood $f(L_{new}|\theta, x_{new})$ in Eq. (5) is formed from Bayes' rule. The rule implies the known-length likelihood (Eq. 10) is weighted by the prior distribution for the unknown object length $f(L_{new})$, and is proportional to:

$$f(L_{new}|\theta, x_{new}) \propto f(x_{new}|\theta, L_{new}) f(L_{new}) \quad (11)$$

in which $f(x_{new}|\theta, L_{new})$ is the likelihood Eq. (10), but for which the unknown object length L_{new} is substituted for L in Eq. (9). We assigned a uniform prior distribution (min = 0 m, max = 30 m) to the unknown lengths of humpback whales, which restricts the model to reasonable size ranges but is otherwise uninformative. We discuss an alternate choice for $f(L_{new})$ in Section 2.3 that allows for estimation of morphological relationships between multiple length measurements.

Model development and analyses were conducted in R (version 4.0.2, R Core Team 2020). Estimation and prediction were performed using Markov chain Monte Carlo sampling in NIMBLE (de Valpine et al. 2017) with 1000 burn-in followed by 1 000 000 iterations. Three independent chains were run to confirm consistency between runs and inspected visually for convergence. Model code is freely available at the Duke University Research Data Repository (<https://doi.org/10.7924/r4sj1jj6s>).

2.3. Testing data (L_{new})

2.3.1. Model validation

We validated the ability of the model to estimate unknown lengths by studying the out-of-sample prediction error of the model in 2 scenarios. In the first

scenario, we randomly split the LemHex-44 and Alta 6 data (collected in Section 2.1.1) into equal-sized training and testing subsets. In the second scenario, we ensured the training and testing subsets were also balanced with respect to images taken at low, medium, and high altitudes (respectively 0–30, 31–60, and 61–120 m). To be precise, the LemHex-44 and Alta 6 data included 42 images nominally taken at altitudes between 0 and 30 m; the second scenario ensured that the training and testing subsets each had 21 of these images, while the first scenario did not. The first scenario is more general, while the second scenario allows out-of-sample prediction error to be studied with respect to altitude.

In both scenarios, we used the training subset to estimate the error parameters 3 separate times: when altitude information is taken from (1) both the barometer and laser altimeters, (2) only the barometer, and (3) only the laser. We used each set of parameter estimates to predict the length measurements of the testing subset. Next, we compared the predicted lengths to the known length of the calibration object (L_{co}), using root mean square error (RMSE) and mean absolute error (MAE) to summarize the prediction errors. Comparing the 3 sets of predictions lets us test how altimeter choice influences uncertainty.

2.3.2. Ecological scenarios

We then tested the model in 2 ecological scenarios to (1) predict total body length measurements of humpback whales ($n = 48$) in order to assign maturity class, and (2) estimate the morphological relationship between total length and rostrum to blowhole distance of Antarctic minke whales ($n = 27$). Both scenarios used images collected with the LemHex-44 (35 mm f_c) (referred to as 'LemHex') and Alta 6 (referred to as 'Alta') with either a 35 or 50 mm f_c . Both scenarios compare predicted lengths under 2 models: Model 1 assumes that each platform contains only a barometer to record altitude for length measurements, whereas Model 2 assumes that each platform has both a barometer and a laser altimeter. Images were selected for each individual and ranked for quality in measurability following Christiansen et al. (2018), where a score of 1 (good quality), 2 (medium quality), or 3 (poor quality) was applied to 7 attributes: camera focus, straightness of body, body roll, body arch, body pitch, total length measurability, and body width measurability. Images with a score of 3 in any attribute were removed from analysis, as well as any images that received a score of 2 in both

roll and arch, roll and pitch, or arch and pitch (Christiansen et al. 2018). Data were collected along the Western Antarctic Peninsula between 2017 and 2019 as part of long-term ecological research projects focused on understanding population dynamics and the impacts of interannual environmental variability on whale body condition. We used MorphoMetriX open-source photogrammetry software (Torres & Bierlich 2020) to measure (in pixels) (1) the total length (TL), as tip of rostrum to fluke notch, of humpback and Antarctic minke whales, as well as (2) rostrum to blowhole distance (RB) of Antarctic minke whales. Both scenarios used the error parameters for the LemHex and Alta to generate posterior predictive distributions for each measurement. Measurement uncertainty was defined by constructing 95 % HPD intervals, which is an interval that represents the region with a 95 % probability of encompassing the parameter.

2.3.2.1. Scenario 1: Predicting length-based maturity class. A single image was used for each individual to obtain length measurements. Length is often used to classify individuals as mature or immature (e.g. Christiansen et al. 2016, 2020a, Lemos et al. 2020). Individuals were assigned as 'mature' if >50 % of their predicted TL posterior distribution was greater than the average length at maturity of 11.2 m, as used in previous studies (Christiansen et al. 2016, 2020a) that were based on data collected at whaling stations (Chittleborough 1955a,b). We then compared measurement uncertainty (as 95 % HPD interval widths) and maturity classification between Model 1 and 2, as well as how uncertainty changes with altitude.

2.3.2.2. Scenario 2: Predicting morphological relationships. In Scenario 2, we demonstrate how to use the Bayesian uncertainty model to estimate standard morphological relationships between TL and RB of a whale. Morphological relationships are commonly used to identify differences between species (e.g. Leslie et al. 2020), track ontological growth (Christiansen et al. 2016), and estimate TL when a TL measurement is unobtainable (e.g. Ratnaswamy & Winn 1993, Fearnbach et al. 2018, Groskreutz et al. 2019). We used measurements from up to 4 images per individual, for a total of 40 images.

Within our hierarchical Bayesian framework, the morphological model replaces the independent prior distribution $f(L_{\text{new}})$ for the unknown lengths in Eq. (11), allowing the morphological model to be estimated with respect to photogrammetric measurement errors. Using the morphological model within our framework allows an explicit, population-level relationship between TL and RB to be estimated from

photogrammetric data. The prior distribution is set to restrict the support of measurements and parameters, but is also large enough to be practically uninformative. We assumed RB measurements are linearly related to TL measurements. The i th RB measurement, RB_i , is modeled as a truncated normal random variable (min = 0 m, max = 30 m) with location parameter μ_i and scale parameter σ_{RB}^2 via:

$$RB_i | \mu_i, \sigma_{RB}^2 \sim \text{trunc} - N(\mu_i, \sigma_{RB}^2; 0, 30) \quad (12)$$

$$\mu_i = \beta_1 + \beta_2 TL_i \quad (13)$$

The location μ_i models the morphological relationship between RB_i and TL_i , and the scale σ_{RB}^2 quantifies the amount of population-level variability around the morphological relationship. The truncated range for RB_i is the same as the original prior distribution $f(L_{\text{new}})$.

The prior distribution for the morphological relationship is designed to be largely uninformative via:

$$\beta_1 \sim U(-3, 30) \quad (14)$$

$$\beta_2 \sim \text{Beta}(1, 1) \quad (15)$$

$$\sigma_{RB}^2 \sim \text{Inv-Gamma}(0.01, 100) \quad (16)$$

The prior distribution for the intercept β_1 is mostly non-negative, which strongly assumes that the intercept can be interpreted as a baseline size parameter. The slope β_2 quantifies the population-level ratio between RB and TL, which is *a priori* known to be <1 since TL is the largest length measurement of a whale. The prior distribution for σ_{RB}^2 is designed to be conjugate, heavy-tailed, and uninformative.

The TL measurements are modeled as truncated normal random variables with location parameter μ_{TL} and scale parameter σ_{TL}^2 . The truncation is wide enough so that the location μ_{TL} describes the population-level mean TL value well, and the scale σ_{TL}^2 quantifies the amount of population-level variability around the mean:

$$TL_i | \mu_{TL}, \sigma_{TL}^2 \sim \text{trunc} - N(\mu_{TL}, \sigma_{TL}^2; 0, 30) \quad (17)$$

Within our context, both parameters μ_{TL} and σ_{TL}^2 are modeled with uninformative priors:

$$\mu_{TL} \sim U(0, 30) \quad (18)$$

$$\sigma_{TL}^2 \sim \text{Inv-Gamma}(0.01, 100) \quad (19)$$

The posterior distribution for the morphological relationship is a variation of the posterior predictive distribution, specified via:

$$\begin{aligned} & f(\beta_1, \beta_2, \sigma_{RB}^2, \mu_{TL}, \sigma_{TL}^2 | x, x_{RB}, x_{TL}) \propto \\ & \int \left\{ \prod_i [f(x_{RB,i} | \theta, RB_i) f(x_{TL,i} | \theta, TL_i)] \times \right. \\ & \left. f(RB_i | TL_i, \beta_1, \beta_2, \sigma_{RB}^2) f(TL_i | \mu_{TL}, \sigma_{TL}^2) \right. \\ & \left. f(\beta_1, \beta_2, \sigma_{RB}^2, \mu_{TL}, \sigma_{TL}^2) f(\theta | x) \right\} dRB_i dTL_i d\theta \end{aligned} \quad (20)$$

in which x and θ represent the collection of training data and measurement error parameters as before, respectively; $x_{RB,i}$ and $x_{TL,i}$ represent the altitude and pixel-length measurements for RB and TL measurements, respectively; RB_i and TL_i represent the unknown RB and TL measurements, respectively, taking the place of the generic object length L in previous equations; and the distributions comprise the likelihood and prior for a Bayesian analysis. We then compared the estimated morphological relationship and associated measurement uncertainty between Model 1 (barometer only) and Model 2 (barometer and laser altimeter).

3. RESULTS

3.1. Training data

The max altitude recorded by the laser altimeter varied between platforms in the training dataset, but each platform recorded to an altitude of at least 62 m (Table S1 in the Supplement at www.int-res.com/articles/suppl/m673p193_supp.pdf). Of the 1068 total measurements (across 3 analysts), 189 measurements had nulled laser values (i.e. altitudes greater than the max altitude recorded by the laser altimeter for each UAS) and were removed from the training data in the analysis. This resulted in a total of 534 length measurements in the training dataset using the barometer and 345 using the laser altimeter.

There were clear resolution differences amongst platforms due to GSD, especially with the DJI P4Pro with video (Fig. 2). Differences in L_p' measurements of the same image amongst the 3 analysts were low, with $CV < 5\%$ for all measurements (min = 0.01 %, max = 4.02 %, mean = 0.87 %, SD = 0.83 %). Thus, the analyst variable (within-frame precision) was excluded from the model.

Overall, each aircraft had greater uncorrected measurement error (before uncertainty model applied) when using the barometer for altitude compared to using the laser altimeter (Fig. 3; Table S2). The Alta with 50 mm f_c had the lowest average uncorrected % error compared to other platforms with the barometer, but also had the third highest SD (Fig. 3; Table S2), likely due to an outlier at a low altitude (Fig. 3). The mean uncorrected % error was reduced to $< 1\%$ when the laser was used for all aircraft, except CineStar (Fig. 3; Table S2). Interestingly, the uncorrected % error was also slightly negatively biased for each platform when the laser altimeter was used. The P4Pro with video displayed wide variation in uncorrected % error for both barometer and laser (Fig. 3, Table S2), and was removed from the analysis. These results show that the choice of altimeter can have significant impacts on measurement error, with altitude measured by the barometer resulting with greater error (Fig. 3).

Some differences were observed between uncorrected measurements of the calibration frame in the center vs. corner of the image, but further inspection revealed that these differences in error were likely driven by inaccuracies of the barometer, rather than

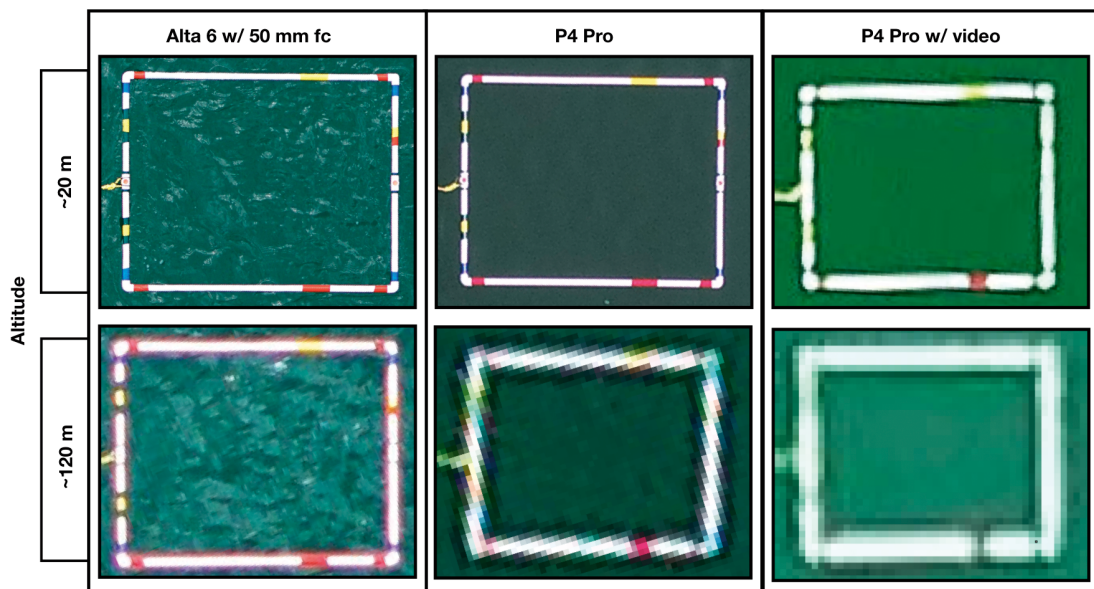


Fig. 2. Image of the calibration frame (1.48×1.15 m) to demonstrate differences in resolution due to ground sampling distance (GSD) amongst the Alta 6, P4Pro, and P4Pro with video at altitudes of ~20 and ~120 m; fc: focal length of the camera

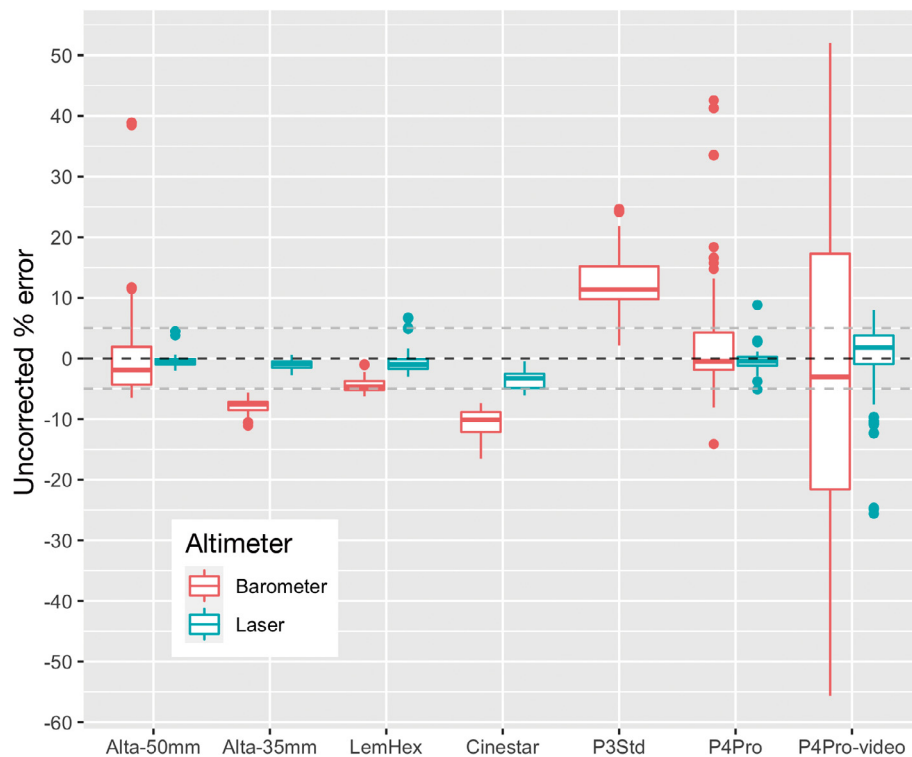


Fig. 3. Uncorrected % error for measurements from each unoccupied aircraft system (UAS) platform. Black dashed line represents 0% uncorrected error (true length = 1.48 m). The gray dashed lines represent the under- and over-estimation of the true length by 5% (1.41 and 1.55 m, respectively). The middle line in each box represents the median (50th percentile), the lower and upper hinges of the box represent the first and third quartiles (25th and 75th percentiles), respectively, and the lower and upper whiskers represent the smallest and largest values that extend at most 1.5× the interquartile range. Any data beyond these whiskers are considered outlying points and are plotted individually. Three analysts measured *n* images for each aircraft using a barometer and laser: Alta 6 with 50 mm focal length lens ('Alta-50mm', $n_{\text{barometer}} = 25$, $n_{\text{laser}} = 16$), Alta 6 with 35 mm focal length lens ('Alta-35mm', $n_{\text{barometer}} = 24$, $n_{\text{laser}} = 9$), LemHex-44 ('LemHex', $n_{\text{barometer}} = 25$, $n_{\text{laser}} = 23$), Cinestar ($n_{\text{barometer}} = 23$, $n_{\text{laser}} = 20$), P3Std ($n_{\text{barometer}} = 24$, $n_{\text{laser}} = 0$), P4Pro ($n_{\text{barometer}} = 21$, $n_{\text{laser}} = 11$), P4Pro w/ video ('P4Pro-video', $n_{\text{barometer}} = 36$, $n_{\text{laser}} = 36$)

object positioning (see Fig. S1). For this reason, object positioning was removed from the model. Each camera displayed low distortion (mean error ≤ 1.03 pixels) and was thus assumed negligible for incorporating in the model.

3.2. Measurement predictions

3.2.1. Model validation

The model validation study compared the model when altitude is collected using different altimeters (barometer only, laser only, and both barometer and laser) under 2 scenarios. Results from the first scenario, which looks at overall prediction error, shows that errors are typically small across altimeters, with RMSE and MAE < 0.08 m, although using only the barometer had larger RMSE and MAE values (Table 2). Results from the second scenario, which allows us to study

prediction error with respect to changing altitude, indicate that errors tend to decrease with altitude, but that the differences are scientifically negligible for large animals (Table S3). In both scenarios, errors were nearly identical when using only a laser altimeter and using both a barometer and laser altimeter, indicating that the laser altimeter results in lower measurement uncertainty (Table 2; Table S3).

Table 2. Results from validation test comparing how altimeter choice influences uncertainty predictions by comparing the root mean square error (RMSE) and mean absolute error (MAE) of the predicted measurements using altitude from only the barometer, only the laser altimeter, and both the barometer and laser altimeter

	RMSE (m)	MAE (m)
Barometer only	0.072	0.062
Laser altimeter only	0.021	0.014
Barometer and laser altimeter	0.021	0.016

3.2.2. Ecological Scenario 1

We predicted length distributions of humpback whales ($n = 48$) using altitude from only a barometer (Model 1) and from both a barometer and laser altimeter (Model 2) to assign sexual maturity. An example of an individual's posterior predictive length distribution and 95 % HPD intervals from Models 1 and 2 is shown in Fig. 4. This individual would be classified as 'mature' if using Model 1, but as 'immature' if using Model 2. This result highlights the importance

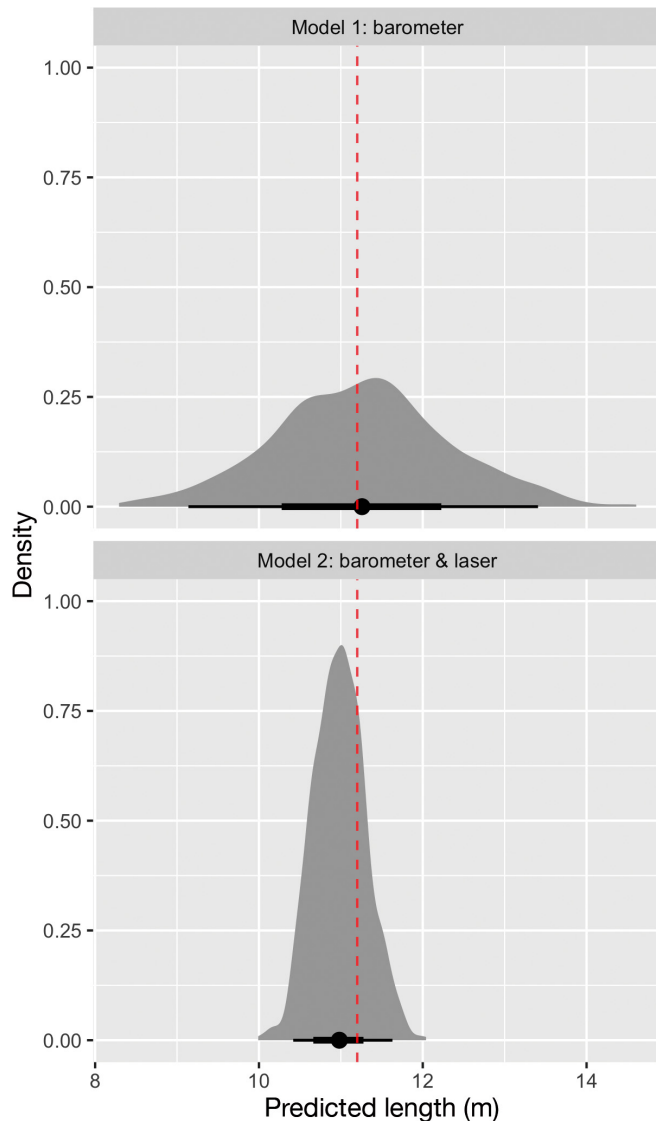


Fig. 4. Example of predicted posterior length distributions from Models 1 and 2 for the same individual humpback whale. The longer black bars represent the 95 % highest posterior density (HPD) intervals, the thicker shorter black bars represent the 65 % HPD interval, and the black dot represents the mean value. The red dashed line represents the average length at maturity for humpback whales (11.2 m)

of altimeter choice and incorporating uncertainty into maturity classification. Overall, Model 1 (barometer only) predicted longer measurements with much greater uncertainty compared to Model 2 (barometer and laser altimeter), across all individuals (Fig. 5). Each 95 % HPD interval quantifies the most likely range in which an exact measurement lies. As such, we can represent the total uncertainty in each measurement using the width of the estimated 95 % HPD interval. Model 1 predicted wider 95 % HPD intervals (mean $\pm$ SD = 3.81 ± 1.47 m, min = 1.52 m, max = 6.84 m) compared to Model 2 (1.10 ± 0.42 , min = 0.42 m, max = 2.00 m) (Figs. 5 & 6). These wider predictive distributions from Model 1 resulted with more overlap across the 11.2 m maturity cutoff length, with several points in particular that would be considered 'mature' if using Model 1, but 'immature' if using Model 2 (Fig. 5). These results suggest that when measurements are made solely with the barometer, extra caution needs to be taken to classify measured individuals into different life-history classes. The widths of the 95 % HPD intervals decreased as altitude increased for both models (Figs 5 & 6). All uncorrected total length measurements (before applying the uncertainty model) fit within the respective 95 % HPD interval of each individual (Fig. S2).

3.2.3. Ecological Scenario 2

We used the Bayesian uncertainty model to estimate the population-level morphological relationship between RB and TL for Antarctic minke whales ($n = 27$) when using altitude from only a barometer (Model 1) compared to from both a barometer and laser altimeter (Model 2). Linear regression confirmed a linear relationship between empirical RB and TL estimates ($r^2 = 0.73$, $p < 0.001$). The mean posterior distribution for the slope parameter (β_2) indicates the population-level ratio between RB and TL. Model 1 predicted a slightly smaller β_2 with much wider uncertainty (mean $\pm$ SD = 0.17 ± 0.045 [95 % HPD interval: 0.08, 0.25]) compared to Model 2 (0.18 ± 0.021 [95 % HPD: 0.13, 0.22]) (Table 3, Fig. 7). Overall, Model 1 predicted larger TL measurements with greater uncertainty compared to Model 2 (Fig. S3), resulting in a larger population-level mean TL value (μ_{TL}) for Model 1 (mean = 8.53 ± 0.196 m) compared to Model 2 (7.67 ± 0.218 m) (Table 3, Fig. 7). Model 2 more reliably captured the population variability in TL, σ_{TL}^2 (mean = 1.171 ± 0.383) compared to Model 1 (0.635 ± 0.308) (Table 3, Fig. 7). There was also a decrease in the estimated value and posterior uncertainty in the popula-

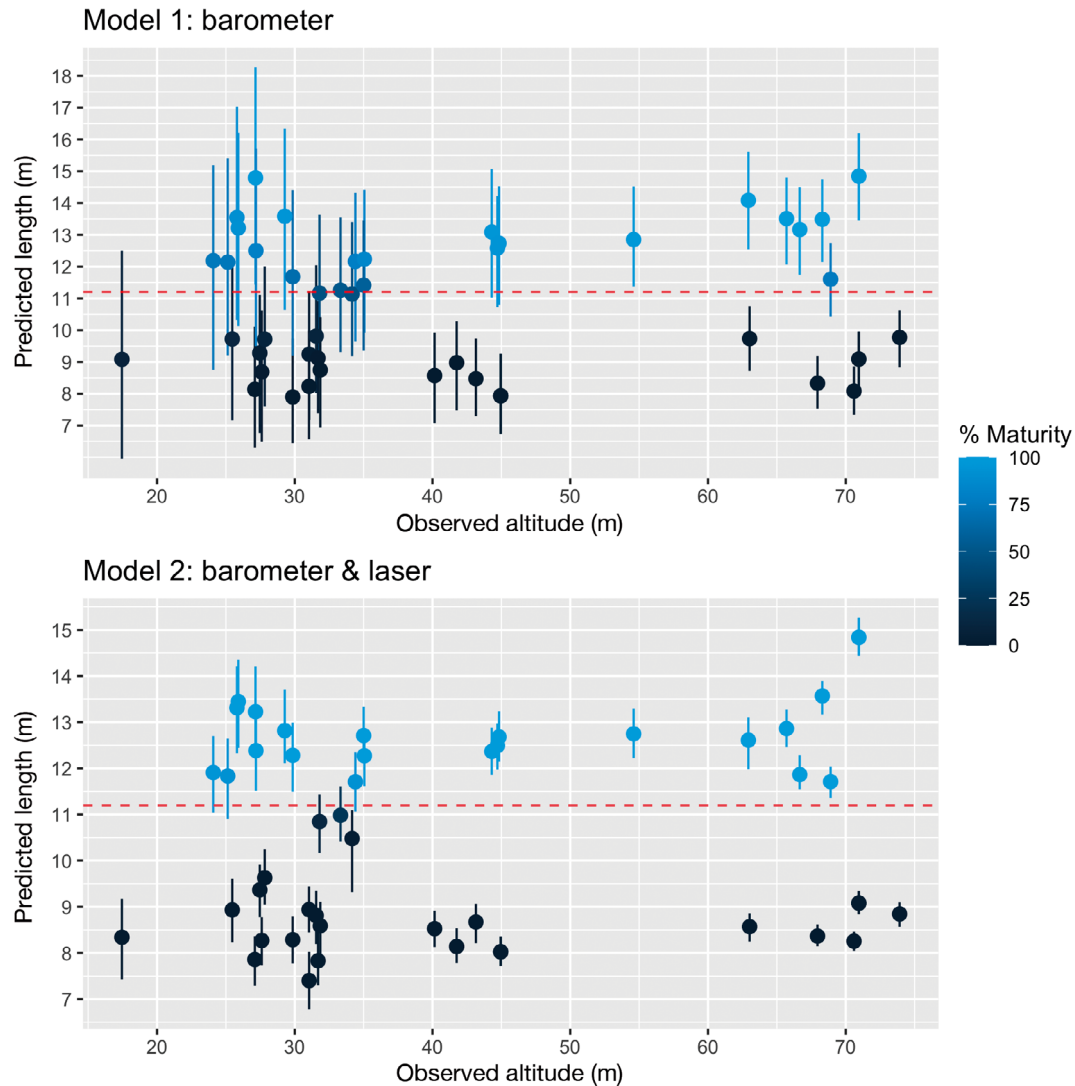


Fig. 5. Comparison of the predictive length measurements of humpback whales ($n = 48$) using Model 1 and Model 2. Each point represents the mean of the predictive posterior length distribution and the lines around each point represent the 95% highest posterior density (HPD) interval. Each individual is colored by the % probability of being mature, defined as the proportion of their predictive posterior length distribution that is greater than 11.2 m, the average length of maturity for humpback whales (represented by the red dashed line) used by Christiansen et al. (2016, 2020a)

tion variability of RB, σ_{RB}^2 , in Model 2 (mean = 0.010 ± 0.003) compared to Model 1 (0.013 ± 0.005). Altogether, the regression parameters (β_1, β_2) had larger uncertainty in Model 1 compared to Model 2 (Table 3, Fig. 7).

4. DISCUSSION

4.1. Novelty

Here we built off of previous methods for quantifying photogrammetric error (e.g. Christiansen et al. 2018, Burnett et al. 2019) and present a Bayesian statistical framework for predicting photogrammetric

uncertainty around UAS-derived measurements. We developed this Bayesian statistical model for researchers using a range of UAS platforms containing different cameras, focal length lenses, and altimeters (barometer vs. both a barometer and laser altimeter). We applied this framework to unknown-sized objects (e.g. whales) that can be used for probabilistic assessments of ecological importance, e.g. determining sexual maturity, or assessing morphological relationships. Our results represent a standardized practice for reporting measurement uncertainty and facilitating collaboration amongst researchers using different UAS platforms and sensors. Importantly, these results also allow us to better measure and predict

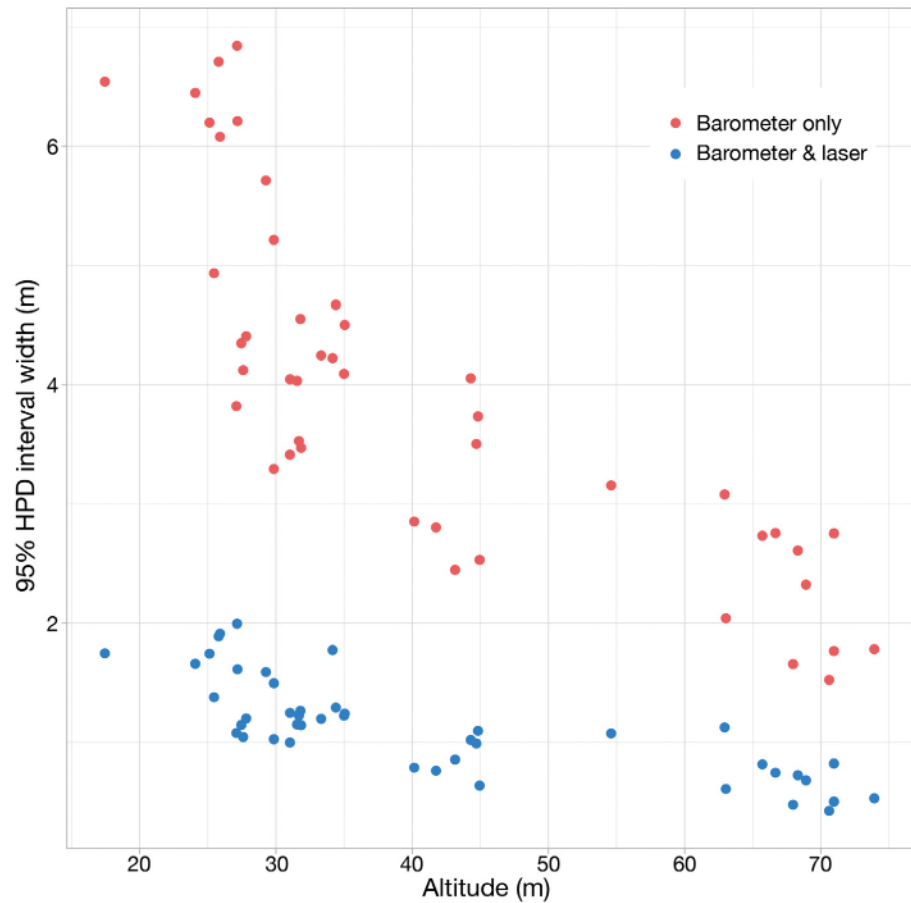


Fig. 6. Uncertainty decreases with altitude. The 95 % highest posterior density (HPD) interval widths (the width of the black bars in Figs. 4 & 5) of the predictive distribution decreases with increasing altitude for both Model 1 and Model 2. Each point represents the difference between the upper and lower bound of the 95 % credible interval for each measurement

Table 3. Results from Model 1 (barometer only) and Model 2 (barometer and laser altimeter) in ecological Scenario 2 for estimating the population-level morphological relationship between total body length (TL) and rostrum to blowhole distance (RB) for Antarctic minke whales ($n = 27$). The mean, standard deviation (SD), and 95 % highest posterior density (HPD) intervals of the posterior distribution are shown for each parameter: the intercept (β_1), slope parameter (β_2) which indicates the population-level ratio between RB and TL, the population-level mean TL value (μ_{TL}), and the scale parameter which quantifies the amount of population-level variability around the mean TL (σ_{TL}^2) and RB (σ_{RB}^2)

	Mean	SD	95 % HPD interval		
			Lower	Upper	Width
Model 1					
β_1	-0.175	0.380	-0.93	0.58	1.51
β_2	0.166	0.045	0.08	0.25	0.17
μ_{TL}	8.531	0.196	8.14	8.91	0.77
σ_{RB}^2	0.013	0.005	0.01	0.02	0.01
σ_{TL}^2	0.635	0.308	0.15	1.25	1.1
Model 2					
β_1	-0.223	0.161	-0.54	0.10	0.64
β_2	0.175	0.021	0.13	0.22	0.09
μ_{TL}	7.670	0.218	7.24	8.10	0.86
σ_{RB}^2	0.010	0.003	0.00	0.02	0.02
σ_{TL}^2	1.171	0.383	0.56	1.93	1.37

morphological features critical to understanding life history patterns and potential impacts from anthropogenically altered habitats.

4.2. Laser altimeter reduces measurement uncertainty

The results of the present study demonstrate that using a laser altimeter for altitude (Scenarios 1 and 2: Model 2) reduces measurement uncertainty compared to only using a barometer (Scenarios 1 and 2: Model 1). This was also confirmed in the validation study (Table 2; Table S3). While 95 % HPD interval widths decreased with altitude for both models in Scenario 1, all mean widths for Model 2 were <2.0 m compared to <6.8 m for Model 1 (Figs. 4–6). The validation study also revealed that while uncertainty is highest when using only a barometer, using only a laser altimeter yields similar results to using a barometer and laser altimeter, indicating that the laser altimeter drives the reduction in uncertainty. Thus, we recommend that whenever feasible, a laser altimeter should be used for recording altitude. We

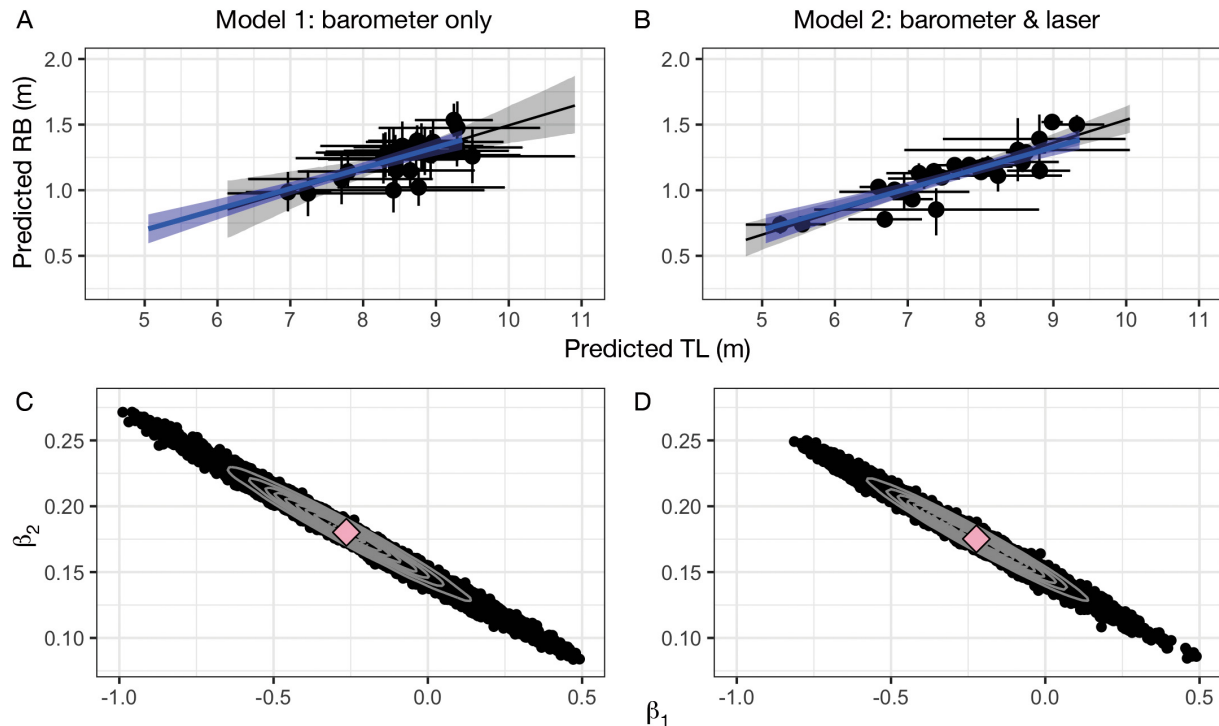


Fig. 7. Comparison of the morphological relationship between total body length (TL) and rostrum to blowhole distance (RB) for Antarctic minke whales (n = 27) between (A,C) Model 1: barometer only and (B,D) Model 2: barometer and laser altimeter. The predictive posterior estimates of RB and TL and their 95 % highest posterior density (HPD) intervals (as bars) for each individual are shown along with the confidence band for the population-level morphological relationship for Model 1 (A) and Model 2 (B). The standard linear regression between the empirical, or 'uncorrected', measurements of RB and TL is represented by the blue line with confidence bands in both (A) and (B). Posterior samples for regression parameters highlight the strong posterior correlation between regression parameters, intercept (β_1) and slope (β_2), for Model 1 (C) and Model 2 (D). The pink diamond represents the mean population-level morphological relationship for each model: $\beta_2 = 0.166$ in Model 1 (C) and $\beta_2 = 0.175$ in Model 2 (D)

demonstrate how this can be achieved for 'off-the-shelf' products, such as the DJI P4Pro, with a custom-made housing (https://github.com/marrs-lab/DJI_PH4_LaserAltimeter). However, because DJI P4Pro with video displayed lower resolution and wider variation in uncorrected % error compared to DJI P4Pro with still images (Figs. 2 & 3; Table S2), we recommend using still images instead of video to reduce measurement uncertainty.

While the maximum altitude recorded with the laser in the training data varied across platforms (Table S2), all platforms obtained a minimum reading of 62 m (non-nulled). Thus, flying at altitudes below 62 m may help ensure consistent laser altimeter readings to yield tighter uncertainty predictions when using this particular laser altimeter. If only using a barometer for recording altitude, then flying at altitudes greater than 30 m may help reduce the risk of obtaining measurements with larger uncertainty (Fig. 6; Table S3). Researchers should thus consider collecting data at altitudes that will yield an appropriate GSD (i.e. Eq. 1, Fig. 1) and 95 % HPD

interval width (Fig. 6; Table S3) that will best accomplish research objectives based on the UAS aircraft, camera, focal length lens, altimeter, and expected length of the target species. For example, to measure changes in body condition of a small cetacean, such as a harbor porpoise *Phocoena phocoena*, smaller 95 % HPD interval widths will be required compared to measuring a larger species, such as a blue whale, and the altitude required to achieve this will depend on the UAS aircraft, camera, focal length lens, and whether a laser altimeter is used in addition to a barometer.

Because barometers are susceptible to rapid changes in pressure unrelated to a shift in aircraft altitude, e.g. from gusts of wind, temperature, and changes in barometric pressure during flight, in general they have been noted to have poor accuracy (Sabatini & Genovese 2014) and problems with drift and delay. These problems are more severe with low-cost sensors commonly found on small UAS (Wei et al. 2016). Barometric altimeters convert changes in aerostatic pressure, the difference between the

atmospheric pressure at a given altitude and the pressure set as the zero point, to altitude measurement (Jan et al. 2008). Variations in temperature and humidity can impact the aerostatic pressure, and thus the recorded altitude (m) (Bao et al. 2017). These factors may be particularly influential when flying over water and may have contributed to the greater predicted uncertainty, especially at lower altitudes (Fig. 6). Jech et al. (2020) noted differences in barometer accuracies compared to Durban et al. (2015, 2016), despite using the same UAS platform (APH-22), that were likely influenced by differences in temperature. Jan et al. (2008) highlighted how a normal procedure for airplanes before arrival at a destination is to adjust the onboard barometer to local barometric pressure provided by air traffic control. These studies suggest that UAS flown in different environments may be susceptible to different levels of barometric error.

Model 2 in Scenario 1 also displayed greater 95 % HPD interval widths at lower altitudes, though not nearly as wide as Model 1 (Fig. 6). The LightWare user manual states an accuracy of ± 0.1 m on a 70 % reflective target at 20°C (LightWare Optoelectronics 2018), so perhaps a combination of temperature and a less reflective surface at lower altitudes caused this increased uncertainty. We recommend that future studies record environmental data on each UAS flight to further explore how different oceanic environments influence barometric and laser accuracy. This may be particularly important for studies comparing migrating populations with foraging and breeding grounds in polar vs. tropical regions.

It is important to note that the training data used in both models were collected in a much warmer climate compared to the testing data (North Carolina vs. Western Antarctic Peninsula). Future studies should also record location and date of the training data to see how changes in environment influences measurement predictions. An advantage of the Bayesian framework is that it allows for the integration of new training data. Thus, researchers working in different environments or in similar places at different times of the year can collect, share, and update training data to further improve predictive distributions for future work.

4.3. Bias

No measurement bias was observed across the 3 different analysts, but these images were all pre-selected by a single analyst and were all of high

quality in terms of measurability (a flat object floating at the surface). Future studies should further examine the potential for bias introduced when analysts have to choose which images to measure. This would likely have little influence in the present study, but could have an effect when selecting images of whales with larger variation in image quality, such as glare, refraction, water visibility, and different body orientations (e.g. such as straight, fluke-down, etc.) and depths, as shown by Christiansen et al. (2018). The Bayesian framework allows for straightforward integration of additional covariates to the model for studies with length predictions that are more heavily influenced by different analysts, image quality, and repeated measurements of individuals in different body orientations.

4.4. Interpreting predicted measurements

4.4.1. Ecological Scenario 1

Overall, measurements using only the barometer (Model 1) had larger mean predicted length posterior distributions compared to using the barometer and laser altimeter (Model 2). Lengths are a common metric used to classify organisms into demographic units, such as sexual maturity (i.e. Christiansen et al. 2016, 2020a, Lemos et al. 2020) and to determine population-level impacts of stressors on growth (Stewart et al. 2021). Figs. 4 & 5 demonstrate the importance of including uncertainty around a measurement in these types of analyses, as the conclusion for maturity classification changes depending on which altimeter is used. Several measurements could be classified differently, and Fig. 4 provides an example of an individual that would be classified as immature if using the laser and barometer, but as mature animals if using only the barometer (Fig. 4).

By generating posterior distributions around measurements in these types of analyses, the probability that an animal exceeds a specific length can be computed to assist in classification and allow ages to be discussed and studied under uncertainty. For example, classification decisions can incorporate uncertainty through rules, as this present study classified individuals as sexually mature if 50 % of their predictive distribution was >11.2 m. In studies with greater uncertainty around each measurement (i.e. only using a barometer), stricter classification rules can be applied, such as 'individuals were classified as sexually mature if 80 % of their predictive distribution was >11.2 m'. This further demonstrates the advan-

tages of using a Bayesian approach to predict uncertainty, where a frequentist approach is limited by standard deviations and confidence intervals, which may be less accurate and precise.

4.4.2. Ecological Scenario 2

As in Scenario 1, Model 1 had larger mean predicted TL posterior distributions with greater uncertainty compared to Model 2 (Fig. S3). This led to a larger population-level mean TL (μ_{TL}) for Model 1 compared to Model 2 (Table 3, Fig. 7). Because the laser altimeter provides more precise and informative measurements than the barometer, Model 2 better captures the variability in the population-level TL measurements (σ_{TL}^2) compared to Model 1 (Table 3, Fig. 7). This is evident in Fig. 7A,B, where Model 1 measurements are larger, have greater uncertainty, are more closely clustered together, and are shifted to the right and up from the empirical relationship between RB and TL (blue line) compared to Model 2. Since RB is modeled as a dependent variable relative to TL, σ_{RB}^2 models the deviation from the linear relationship between RB and TL. The estimated σ_{RB}^2 value is smaller for Model 2 compared to Model 1, and with less uncertainty (Table 3), which yields the narrower model-based confidence bands (grey shading) in Fig. 7A,B.

Overall, the regression parameters (β_1 , β_2) had greater uncertainty in Model 1 compared to Model 2 (Table 3, Fig. 7). The mean population-level morphological relationship between RB and TL of Antarctic minke whales was smaller with much greater uncertainty for Model 1 compared to Model 2 (Table 3, Fig. 7). This further highlights the importance of using a laser altimeter when estimating morphological relationships from UAS-based imagery. Morphological relationships pertaining to the skull are commonly used for distinguishing species (Leslie et al. 2020), tracking ontological growth (Christiansen et al. 2016), and estimating TL when direct measurements of TL are not obtainable (Ratnaswamy & Winn 1993, Fearnbach et al. 2018, Groskreutz et al. 2019). The Bayesian framework we present here demonstrates how morphological relationships are simultaneously estimated with photographically derived measurements, rather than in 2 stages, in which physical measurements are first extracted from photographs and then used to estimate morphological relationship. This hierarchical Bayesian framework naturally allows complex correlations and dependencies between the measurement error parameters to be accounted for

during estimation. This framework also allows repeated observations of the same lengths to be directly incorporated, rather than being manually averaged or otherwise summarized before analysis, and incorporates altitude measurements from multiple altimeters (barometer and laser).

The present study provides a robust method for predicting photogrammetric measurement uncertainty specific across UAS platforms. This approach will help researchers set protocols to minimize measurement errors during data collection to yield scientifically robust conclusions through an analytical workflow. Standard frameworks like the one presented here can facilitate collaboration amongst researchers using different UAS platforms to pool resources for comparative studies using past and future photogrammetric data to better monitor species and populations.

Data availability. The data and uncertainty model code can be found at the Duke University Research Data Repository (<https://doi.org/10.7924/r4sj1jj6s>), and instructions for DJI Phantom 4 Pro laser altimeter installation can be found at https://github.com/marrs-lab/DJI_PH4_LaserAltimeter.

Acknowledgements. We thank Clara Bird and Allison Duprey for assistance in data collection/management; the Duke University Marine Lab for project support; and Dr. Andrew Read for project design and manuscript feedback. Antarctic humpback whale imagery was collected as part of National Science Foundation Office of Polar Programs Grants 1643877 and 1440435 to A.S.F. under NMFS permits 14809 and 23095, ACA Permits 2015-011 and 2016-024, UCSC IACUC Fria1706.

LITERATURE CITED

- ✦ Arnold SJ (1983) Morphology, performance and fitness. *Am Zool* 23:347–361
- ✦ Arona L, Dale J, Heaslip SG, Hammill MO, Johnston DW (2018) Assessing the disturbance potential of small unoccupied aircraft systems (UAS) on gray seals (*Halichoerus grypus*) at breeding colonies in Nova Scotia, Canada. *PeerJ* 6:e4467
- ✦ Austin PC, Brunner LJ, Hux JE (2002) Bayeswatch: an overview of Bayesian statistics. *J Eval Clin Pract* 8:277–286
- ✦ Bao X, Xiong Z, Sheng S, Dai Y, Bao S, Liu J (2017) Barometer measurement error modeling and correction for UAH altitude tracking. In: 29th Chinese Control And Decision Conference (CCDC), Chongqing, China. IEEE, p 3166–3171. <https://doi.org/10.1109/CCDC.2017.7979052>
- ✦ Best PB, R  ther H (1992) Aerial photogrammetry of southern right whales, *Eubalaena australis*. *J Zool (Lond)* 228: 595–614
- ✦ Burnett JD, Lemos L, Barlow D, Wing MG, Chandler T, Torres LG (2019) Estimating morphometric attributes of baleen whales with photogrammetry from small UASs: a case study with blue and gray whales. *Mar Mamm Sci* 35:108–139

- ✦ Castrillon J, Bengtson Nash S (2020) Evaluating cetacean body condition; a review of traditional approaches and new developments. *Ecol Evol* 10:6144–6162
- ✦ Chittleborough RG (1955a) Aspects of reproduction in the male humpback whale, *Megaptera nodosa* (Bonnaterre). *Aust J Mar Freshw Res* 6:1–29
- ✦ Chittleborough RG (1955b) Puberty, physical maturity, and relative growth of the female humpback whale, *Megaptera nodosa* (Bonnaterre), on the Western Australian Coast. *Aust J Mar Freshw Res* 6:315–327
- ✦ Christiansen F, Vikingsson GA, Rasmussen MH, Lusseau D (2013) Minke whales maximise energy storage on their feeding grounds. *J Exp Biol* 216:427–436
- ✦ Christiansen F, Dujon AM, Sprogis KR, Arnould JPY, Bejder L (2016) Noninvasive unmanned aerial vehicle provides estimates of the energetic cost of reproduction in humpback whales. *Ecosphere* 7:e01468
- ✦ Christiansen F, Vivier F, Charlton C, Ward R, Amerson A, Burnell S, Bejder L (2018) Maternal body size and condition determine calf growth rates in southern right whales. *Mar Ecol Prog Ser* 592:267–281
- ✦ Christiansen F, Sprogis KR, Gross J, Castrillon J, Warick HA, Leunissen E, Bengtson Nash S (2020a) Variation in outer blubber lipid concentrations does not reflect morphological body condition in humpback whales. *J Exp Biol* 223:jeb.213769
- ✦ Christiansen F, Dawson SM, Durban JW, Fearnbach H and others (2020b) Population comparison of right whale body condition reveals poor state of the North Atlantic right whale. *Mar Ecol Prog Ser* 640:1–16
- ✦ Cosens SE, Blouw A (2003) Size- and age-class segregation of bowhead whales summering in northern Foxe Basin: a photogrammetric analysis. *Mar Mamm Sci* 19:284–296
- ✦ Cabbage JC, Calambokidis J (1987) Size-class segregation of bowhead whales discerned through aerial stereophotogrammetry. *Mar Mamm Sci* 3:179–185
- ✦ Dawson SM, Bowman MH, Leunissen E, Sirguy P (2017) Inexpensive aerial photogrammetry for studies of whales and large marine animals. *Front Mar Sci* 4:366
- ✦ De Meyer J, Verhelst P, Adriaens D (2020) Saving the European eel: how morphological research can help in effective conservation management. *Integr Comp Biol* 60:467–475
- ✦ de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Lang DT, Bodik R (2017) Programming with models: writing statistical algorithms for general model structures with NIMBLE. *J Comput Graph Stat* 26:403–413
- ✦ Durban JW, Fearnbach H, Barrett-Lennard LG, Perryman WL, LeROI DJ (2015) Photogrammetry of killer whales using a small hexacopter launched at sea. *J Unmanned Vehicle Syst* 3:131–135
- ✦ Durban JW, Moore MJ, Chiang G, Hickmott LS and others (2016) Photogrammetry of blue whales with an unmanned hexacopter. *Mar Mamm Sci* 32:1510–1515
- ✦ Ellison AM (2004) Bayesian inference in ecology. *Ecol Lett* 7: 509–520
- ✦ Fearnbach H, Durban JW, Ellifrit DK, Balcomb KC III (2011) Size and long-term growth trends of Endangered fish-eating killer whales. *Endang Species Res* 13:173–180
- ✦ Fearnbach H, Durban JW, Ellifrit DK, Balcomb KC (2018) Using aerial photogrammetry to detect changes in body condition of endangered southern resident killer whales. *Endang Species Res* 35:175–180
- ✦ Gough WT, Segre PS, Bierlich KC, Cade DE and others (2019) Scaling of swimming performance in baleen whales. *J Exp Biol* 222:jeb204172
- ✦ Groskreutz MJ, Durban JW, Fearnbach H, Barrett-Lennard LG, Towers JR, Ford JKB (2019) Decadal changes in adult size of salmon-eating killer whales in the eastern North Pacific. *Endang Species Res* 40:183–188
- ✦ Ichii T, Kato H (1991) Food and daily food consumption of southern minke whales in the Antarctic. *Polar Biol* 11:479–487
- ✦ Irschick DJ (2003) Measuring performance in nature: implications for studies of fitness within populations. *Integr Comp Biol* 43:396–407
- ✦ Jan S, Gebre-Egziabher D, Walter T, Enge P (2008) Improving GPS-based landing system performance using an empirical barometric altimeter confidence bound. *IEEE Trans Aerosp Electron Syst* 44:127–146
- ✦ Jech JM, Johnson JJ, Lutcavage M, Vanderlaan ASM, Rzhantov Y, LeROI D (2020) Measurements of juvenile Atlantic bluefin tuna (*Thunnus thynnus*) size using an unmanned aerial system. *J Unmanned Vehicle Syst* 8: 140–160
- ✦ Johnston DW (2019) Unoccupied aircraft systems in marine science and conservation. *Annu Rev Mar Sci* 11:439–463
- ✦ Kahane-Rapport SR, Savoca MS, Cade DE, Segre PS and others (2020) Lunge filter feeding biomechanics constrain rorqual foraging ecology across scale. *J Exp Biol* 223:jeb.224196
- ✦ Koopman HN, Pabst DA, McLellan WA, Dillaman RM, Read AJ (2002) Changes in blubber distribution and morphology associated with starvation in the harbor porpoise (*Phocoena phocoena*): evidence for regional differences in blubber structure and function. *Physiol Biochem Zool* 75:498–512
- ✦ Lambertsen RH, Rasmussen KJ, Lancaster WC, Hintz RJ (2005) Functional morphology of the mouth of the bowhead whale and its implications for conservation. *J Mammal* 86:342–352
- ✦ Lemos LS, Burnett JD, Chandler TE, Sumich JL, Torres LG (2020) Intra- and inter-annual variation in gray whale body condition on a foraging ground. *Ecosphere* 11: e03094
- ✦ Leslie MS, Perkins-Taylor CM, Durban JW, Moore MJ and others (2020) Body size data collected non-invasively from drone images indicate a morphologically distinct Chilean blue whale *Balaenoptera musculus* taxon. *Endang Species Res* 43:291–304
- ✦ LightWare Optoelectronics (2018) LightWare SF11 Laser altimeter product manual. www.mouser.com/datasheet/2/321/28054-SF11-Laser-Altitude-Manual-Rev8-1371857.pdf
- ✦ Lockyer C (1981) Growth and energy budgets of large baleen whales from the Southern Hemisphere. In: Gordon Clark J (ed) *Mammals in the seas*. FAO Fish Ser 5, Vol 3: General papers and large cetaceans. FAO, Rome, p 379–487
- ✦ MathWorks (2017) Single camera calibration app. <https://www.mathworks.com/help/vision/ug/single-camera-calibrator-app.html>
- ✦ Miles DB (2020) Can morphology predict the conservation status of iguanian lizards? *Integr Comp Biol* 60:535–548
- ✦ Miller CA, Best PB, Perryman WL, Baumgartner MF, Moore MJ (2012) Body shape changes associated with reproductive status, nutritive condition and growth in right whales *Eubalaena glacialis* and *E. australis*. *Mar Ecol Prog Ser* 459:135–156
- ✦ Palacios DM, Salazar SK, Day D (2004) Cetacean remains and strandings in the Galápagos Islands, 1923–2003. *Lat*

- Am J Aquat Mamm 3:127–150
- ✦ Perryman WL, Lynn MS (1993) Identification of geographic forms of common dolphin (*Delphinus delphis*) from aerial photogrammetry. Mar Mamm Sci 9:119–137
- ✦ Perryman WL, Lynn MS (2002) Evaluation of nutritive condition and reproductive status of migrating gray whales (*Eschrichtius robustus*) based on analysis of photogrammetric data. J Cetacean Res Manag 4:155–164
- ✦ Perryman WL, Westlake RL (1998) A new geographic form of the spinner dolphin, *Stenella longirostris*, detected with aerial photogrammetry. Mar Mamm Sci 14:38–50
- R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>
- ✦ Racine-Poon A (1988) A Bayesian approach to nonlinear calibration problems. J Am Stat Assoc 83:650–656
- ✦ Raoult V, Colefax AP, Allan BM, Cagnazzi D and others (2020) Operational protocols for the use of drones in marine animal research. Drones 4:64
- ✦ Ratnaswamy MJ, Winn HE (1993) Photogrammetric estimates of allometry and calf production in fin whales, *Balaenoptera physalus*. J Mammal 74:323–330
- ✦ Read AJ (1990) Estimation of body condition in harbour porpoises (*Phocoena phocoena*). Can J Zool 68:1962–1966
- ✦ Sabatini AM, Genovese V (2014) A sensor fusion method for tracking vertical velocity and height based on inertial and barometric altimeter measurements. Sensors (Basel) 14:13324–13347
- Sasse B (2003) Job-related mortality of wildlife workers in the United States, 1937–2000. Wildl Soc Bull 31:1015–1020
- ✦ Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675
- ✦ Stewart JD, Durban JW, Knowlton AR, Lynn MS and others (2021) Decreasing body lengths in North Atlantic right whales. Curr Biol 31:3174–3179.e3
- Taylor BN, Kuyatt CE (1994) Guidelines for evaluating and expressing the uncertainty of NIST measurement results. NIST Tech Note 1297, 1994 edn. US Government Printing Office, Washington, DC
- ✦ Torres W, Bierlich K (2020) MorphoMetriX: a photogrammetric measurement GUI for morphometric analysis of megafauna. J Open Source Softw 5:1825–1826
- ✦ Voges R, Wiegardt CS, Wagner B (2018) Finding timestamp offsets for a multi-sensor system using sensor observations. Photogramm Eng Remote Sens 84:357–366
- ✦ Wei S, Dan G, Chen H (2016) Altitude data fusion utilising differential measurement and complementary filter. IET Sci Meas Technol (Singap) 10:874–879
- Whitehead H, Payne R (1981) New techniques for assessing populations of right whales without killing them. In: Gordon Clark J (ed) Mammals in the seas. FAO Fish Ser 5, Vol 3: General papers and large cetaceans. FAO, Rome, p 189–211

Editorial responsibility: Peter Corkeron,
Woods Hole, Massachusetts, USA
Reviewed by: P. Conn and 1 anonymous referee

Submitted: October 18, 2020
Accepted: July 1, 2021
Proofs received from author(s): August 24, 2021