# Inferring sperm whale (*Physeter macrocephalus*) sex and age classes using aerial photogrammetry

Ana Eguiguren1\*, David Gaspard1, Christine Konrad1, Hal Whitehead1

1Biology Department, Dalhousie University, Halifax, Nova Scotia, Canada

\*Corresponding author: [anaeguibur@gmail.com](mailto:anaeguibur@gmail.com)

Table of Contents

[Inferring sperm whale (Physeter macrocephalus) sex and age classes using aerial photogrammetry 1](#_Toc201083895)

[1 | INTRODUCTION 2](#_Toc201083896)

[2 | METHODS 2](#_Toc201083897)

[2.1 | Data Collection 2](#_Toc201083898)

[2.2 | Morphometric measurements 2](#_Toc201083899)

[2.3 | Inferring Age/sex class 5](#_Toc201083900)

[3. Results 6](#_Toc201083901)

[Discussion 7](#_Toc201083902)

## 1 | INTRODUCTION

## 2 | METHODS

### 2.1 | Data Collection

We carried out dedicated surveys in the deep waters (> 1000 m) off the Galápagos Islands aboard a 12.03 m sailboat (*Balaena*) between January and May 2023 (research permit No. PC-86-22). We searched for sperm whales acoustically using a 100 m towed hydrophone and visually during daylight hours. When we encountered groups of females and juveniles, we followed them for as long as possible at a cautious distance to collect behavioural, acoustic, and photo-identification data. When single males were detected, we approached for a brief data collection but abandoned them after they dove.

If conditions were adequate (windspeed < 10 kts and no rain), we conducted 1 – 2 hour flight sessions using a DJI Mini 2 drone (249 g) equipped with propeller guards and landing gear. We conducted sessions in the morning and afternoon when glare in the water interfered the least with visibility. Once we approached a group of whales with the drone, we flew between 15 - 120 m above the water and pointed the camera perpendicularly (i.e., nadir) over the whales. During flights, we recorded continuously at 29.79 fps at 1080p or 4K resolution. We alternated a group-follow protocol–during which we kept visual contact with a group of whales by flying high enough to fit all whales in the frame (Altmann 1974)–with brief moments of close approach (15 - 20 m)–to capture individual whales’ distinctive marks and allow for more accurate size estimates. At the end of each flight, we hovered over the research vessel to collect a calibration image (see 2.2.1).

### 2.2 | Morphometric measurements

#### 2.2.1 | Estimating and correcting measurement error

Errors in aerial photogrammetry arise from several sources, of which the most impactful are imprecise altitude estimates (Burnett et al. 2019, Bierlich et al. 2021, Glarou et al. 2022, Napoli et al. 2024). Drones that derive altitude measurements from inbuilt barometers, as was our case, can be inaccurate due to changes in meteorological conditions and internal biases (Burnett et al. 2019, Bierlich et al. 2021). We used measurements of our research vessel collected throughout the field season at various altitudes (27 – 120 m) to quantify the uncertainty in morphometric measurements and correct altitude estimates. We quantified percent measurement error using a modified version of Bierlich et al. 2021 as:

Where is the known length of the calibration object (12.03 m), and is the estimated length in meters of the calibration object in each image . To compute a corrected altitude estimate, we first calculated the true altitude given the for each still image of the research vessel and its known length .

We used MorphoMetriX V2 (Torres & Bierlich 2020) to measure the length in pixels () of the research vessel and whales. Next, we applied a mixed-effects linear regression with date as a random effect to estimate a corrected altitude ( given the barometer altitude across measurements for each day:

We used corrected altitude estimates to estimate whale morphometry.

#### 2.2.2 | Measuring whales

Drone footage was quality-rated on a scale of 0 – 8, with 0 being high quality and 8 being low quality, based on the level of glare, sea-surface disruption, focus, and exposure. Only recordings with a quality rating ≤ 4 were included in the analysis. Within high-quality videos, extracted still images using the behavioural analysis software BORIS (Friard & Gamba 2016). We selected frames where whales were lying mostly flat at the water surface, located near the center of the frame, and where the drone camera was positioned at nadir relative to the water surface.

For each whale, we measured total length (TL), head-to-flipper length (HF), and head-to-dorsal fin length (HD) (Figure 1).

A whale swimming in the water

Description automatically generated

s

d

t

n

f

Figure . Landmarks used to measure sperm whale morphometry. s = snout; f = flipper; d = dorsal fin; t = tail-stalk; n = fluke notch. Total length (TL) measures the piece-wise distance from s to d, to t, to n. Head to flipper (HF) measures the length from s to f. Head to dorsal fin (HD) measures the length from s to d.

We converted measures to true lengths (*L*) in meters by applying the following equation, modified from Burnett et al. (2019):

where *H* is the drone altitude above sea level, and α is a scaling corresponding to the DJI Mini 2 drone camera. While this value can be computed based on known camera parameters (i.e., focal length and pixel dimensions), these values were unavailable for our drone model. We therefore empirically estimated α by obtaining measurements of a known object of known length *L* and known distance *H* using equation 2.

We also identified measured whales based on observable markings—including visible fluke marks, indentations, rake marks, white patches, and sloughed skin patterns (O’Callaghan et al. 2024). We rated still images used for photo-identification on a scale of 1 – 5 (1 = poor, 5 = good) based on focus, contrast, and saturation (Modified from Arnbom 1987). Initial identifications were made using images rated ≥ 3. In cases where multiple still images of the same individual were taken from a video recording, we also assigned identifications to lower-quality images if contextual evidence supported the match to a higher-quality image.

Whales were photo-identified

#### 2.2.1 | Assessing the reliability of morphometric measurements

We evaluated the degree to which TL, HD, and HF could be measured reliably between frames and across days by estimating the coefficient of variation (CV) in measurements taken from the same individual across different frames, flights, and days (Christiansen et al. 2018). We examined the relationship between observed CV and altitude and image Q rating to determine an optimal cutoff at which morphometric features could be measured reliably.

### 2.3 | Inferring Age/sex class

Based on Nishiwaki et al. (1963), the relationship between sperm whale length (*L*) and nose-to-body ratio (*R*) can be modelled by separate logistic curves for males and females. For females, the nose-to-body ratio () can be approximated as:

Where is the maximum (asymptote) nose-body-ratio of female whales and is the initial rate of change in R with increasing length. For males, the relationship between body length and nose-to-body ratio for young (i.e., small individuals) would be the same as for females, but would diverge after a given length threshold () such that:

Where is the maximum difference of nose-to-body ratio of a male compared to a female of the same size, and is the initial rate of change in nose-to-body ratio with length following the point of divergence.

We inferred the probability that individual whales were female by first finding the parameter values for *maxF, fr, maxM,* and *mr* that minimized the total sum-of-squares given our data, using the *optim* function with the default Nelder-Mead algorithm in base R (R Core Team 2019). Next, the posterior probability that each whale was female was estimated based on how close each point fell to the ‘female curve’ using the following equation:

And converted to a probability of an individual being a female as:

While we were primarily interested in a quantitative representation of the likelihood that sperm whales belong to one sex or the other, we explored the classification abilities of our algorithm by applying it to simulated datasets (1000 simulations for each scenario) with known sexes associated with different levels of measurement error. We also used simulations to determine a size cutoff above which individuals could be more confidently assigned a sex.

## 3. Results

#### 3.1 | Error estimation and correction

We used 343 measurements of *Balaena* taken at nadir over 18 days at varying altitudes. Length estimates based on barometric altitudes underestimated the boat length by 0.55 m (*SD* = 0.37). This error was associated with an average 2.35 m underestimation of the barometric altitude (*SD* = 1.94). The modelled corrected altitude () resulted in an average 0.12 % error (*SD = X,* Figure 2).

A graph with a bar graph and text

AI-generated content may be incorrect.

Figure . Distribution of percent error in the length estimate of a calibration object based on uncorrected (i.e. barometric) and corrected altitude measurements using a DJI Mini2 drone. Dashed gray lines indicate ± 5% errors.

#### 3.2 | Whale measurements

We took measurements from 504 still images between 14.7 – 132 m altitude (corrected). Of these, 310 could be assigned to 89 individuals. Only still images of video taken under 70 m had high enough quality (Q3 – 5) to be identified reliably (Figure 3).

A graph of a number of different levels

AI-generated content may be incorrect.

Figure . Corrected altitude (m) distribution across photo quality ratings (Q) of across still images. The 70 m threshold is shown for reference.

We had >= 2 measurements of HD for 88 individuals (2 – 13 times), and 74 of HF, with each individual being measured between 2 – 8 times.

#### 3.3 | Age/sex inference

## Discussion

* Different nose/body ratios may influence length estimation based on IPI’s (Christine)
* Measurement of uncertainty can be incorporated into demographic models based on aerial photogrammetry
* How can one transfer our findings when using other drone models