# Inferring sperm whale (*Physeter macrocephalus*) sex 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 6](#_Toc201083900)

[3. Results 8](#_Toc201083901)

[Discussion 12](#_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.

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 most flights, 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 (12.03 m) 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 in meters, and is the estimated length in meters of the calibration object in each image . We used MorphoMetriX V2 (Torres & Bierlich 2020) to measure the length in pixels () of the research vessel in still images taken from video recordings, and converted length measurements in pixels ( to length () in meters by applying equation (2), modified from Burnett et al. (2019):

where *H* the drone altitude above sea level, and α is a scaling corresponding to the DJI Mini 2 drone camera. While , can be computed based on known camera parameters (i.e., focal length and pixel dimensions), these values were unavailable for our drone model from the manufacturer. We therefore empirically estimated on land α by obtaining measurements of a known object of known length *L* and known distance (*H*)in the lab using equation (3).

To estimate the bias in the drone’s barometric altitude, we first computed the true altitude given the for each still image of the research vessel and its known length .

We then applied a linear regression to estimate a corrected altitude ( given the barometer altitude:

#### 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 the total length (*TL*), snout-to-flipper length (*SnF*), and snout-to-dorsal fin length (*SnD*) in pixels (Figure 1). *TL* was measured from the tip of the rostrum to the fluke notch, *SnF* was measured from the tip of the snout to the transversal intersection of the base of the flippers with the spine, and *SnD* was measured from the tip of the snout to the caudal base of the dorsal fin. To estimate nose proportions, we calculated the nose-to-body ratio (*R*) by dividing *SnF* or *SnD* by *TL*, resulting in two metrics: *R – Flipper* and *R – Dorsal*, respectively. *TL* was converted from pixels to meters using Equation 2 and incorporating the corrected drone altitude () calculated using Equation 5.

To capture inter-image variability, we attempted to measure each whale at least three times per recording. However, obtaining *SnD* and *SnF* measurements was sometimes hindered by whale positions and visibility. As sperm whales often tuck their flippers against their body, the insertion point of the flipper could not always be observed from the drone’s perspective, which impeded measuring SnF. SnD measurements were limited by light and water conditions, or when the dorsal fin gradually tapered into the body without a clear boundary.

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.

### 2.3 | Photo-identifying whales

We 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 (for example, if the same whale could be tracked throughout a recording).

### 2.4 | Inferring age classes and sex

#### 2.4.1 Age classes

We associated individual length (*TL*) to age classes defined in **Table 1**, which we delineated based on whaling-based research relating length measurements to analyses of gonadal development, stomach contents, and dentin layer based age estimates (Best 1979, Best et al. 1984, Rice 1989).

**Table 1.** Age classes for male and female sperm whales based analyses of whaling data (Best 1979, Best et al. 1984, Rice 1989).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age class | Sex | Length (m) | Age range | Life stage traits |
| Neonates | Both | < 4.10 | Few days - months | Unhealed umbilical regions, likely recently born (Best et al. 1984). |
| Calves | Both | 4.10 – 5.5 | < 1 year | Young individuals almost exclusively dependent on milk (Best et al. 1984). |
| Juveniles | Both | 5.5 – 7.6 | 1 – 2 years | Primarily depend on milk for sustenance, although solid foods have been found in their gut contents (Best et al. 1984). |
| Sub-Adult | Male | 7.6 – 10.0 | 2 – 7 years | While milk may still present in the stomach, solid foods are more frequently found. Still, the majority if individuals have not attained sexual maturity (Best et al. 1984). |
| Sub-Adult | Female | 7.6 – 8.5 | 2 – 7 years | While milk may still present in the stomach, solid foods are more frequently found. Still, the majority if individuals have not attained sexual maturity (Best et al. 1984). |
| Adult | Male | 10.0 – 13.7 | 7 – 20 years | Sexual maturity in males (sperm production) starts at 10 m long, between 7 – 11 years of age, which matches with them leaving their natal unit (Best 1979, Best et al. 1984, Mendes et al. 2007). During this period, individuals continue to grow. |
| Adult | Female | 8.5 – 10.m | 7 – 20 years | Females achieve gonad development between 8.2 – 9.2 m and are able to conceive shortly after (Rice 1989). |
| Mature | Male | > 13.7 | > 20 | Almost all males of this size range are physiologically fertile, as defined by the concentration of sperm in their seminal fluid and will be either solitary or form bachelor schools. Although physiologically fertile, males will likely only start mating when they’ve reached < 15.7 m (35 years) (Best et al. 1984, Rice 1989). |
| Mature | Female | 10 – 12 m | > 20 years | Females attain full size at this age. |

#### 2.3.2 Sex

The relationship between sperm whale length (*L*) and nose-to-body ratio (*R*) shown by Nishiwaki et al. (1963) 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) *R* of female whales, and is the initial rate of change in *R* with increasing length. For males, the relationship between body length and *R* for young (i.e., small individuals) is expected to follow the same trend as that of females, diverging after a length threshold () such that:

Where is the maximum difference of *R* ratio of a male compared to a female of the same size, and is the initial rate of change in *R* ratio with length following the point of divergence (*chm*), which we set at 6 m based on Nishiwaki et al. 1963.

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). We initialized the optimizing algorithm using parameter estimates based on Nishiwaki’s data (based on tip of snout to eyeball). 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 for the likelihood of being female ():

And converted to a posterior probability of an individual being a female by:

To propagate the uncertainty associated with inter-image variation of our individual estimates, we applied a stratified bootstrap by individual ID (Dixon 2001, Napoli et al. 2024). In each of 1000 iterations, we randomly sampled measurements collected from one still image for each individual whale. Sampled data was used to obtain optimized parameter values and individual . Resulting parameter values and individual P(Fi)s were then used to calculate mean values and 95th percentile confidence intervals (95% *CI*). For this analysis, we included only individuals with at least three measurements of *TL, R – Flipper*, and *R - Dorsal*.

#### 2.3.3 Behavioural context

We then inspected whether individual whales performed or received *peduncle dives*. *Peduncle dives* are short (few seconds) and shallow dives performed by a calf or juvenile onto the base of the peduncle (fluke stalk) and under a usually larger whale, during the calf/juvenile often presses its snout onto the larger whale’s genital region (Gero & Whitehead 2007). They can be detected on drone base recordings when calves arch their backs and dive under a larger whale’s body repeatedly. Previously, peduncle dives were assumed to indicate suckling (Gero & Whitehead 2007), but recent underwater footage suggests that peduncle dives may not be associated with suckling and that this rather represents a form of affiliative behaviour between calves and mothers/allomothers (Sarano et al. 2023). Although peduncle dives may not necessarily involve suckling, all published reports of *peduncle dives* in which the sex of the receiving whales is known involve female whales (Gero & Whitehead 2007, Sarano et al. 2023). For each measured whale, we recorded whether it had been observed performing or receiving a *peduncle dive* in any of the video footage from which still images for measurements were taken.

## 3. Results

### 3.1 | Error estimation and correction

We obtained 343 measurements of *Balaena* across 18 days in the field at varying altitudes. Length estimates based on barometric altitudes underestimated the boat length by 0.55 m on average (*SD* = 0.37 m), corresponding to a -4.55% measurement error (*SD* = 3.15%). This bias was associated with an average 2.35 m underestimation in barometric altitude (*SD* = 1.94 m). Using the model corrected altitude () reduced average length error to 0.12 % length error (*SD = 3.15%*).

### 3.2 | Whale measurements and photo-identification

We were able to extract *R - Dorsal* metrics more frequently than *R - Flipper* (491 and 297, respectively). Only images captured at altitudes up to 70 m altitude had high enough quality (Q3 – 5) for initial identification (Figure 2), resulting in a subset of 168 still images assigned to 51 individuals for which *TL, R- Dorsal,* and *R – Flipper* could be measured at least three times.

A graph with a number of boxes

AI-generated content may be incorrect.

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

### 3.3 | Age and sex inference

#### 3.3.1 Parameter optimization

Optimal values for *fr, maxf, mr and maxm* varied more across bootstrap iterations in models fit with *R – Dorsal* than *R – Flipper* (**Table 1**), resulting in generally higher levels of uncertainty associated with models based on *R – Dorsal* than *R – Flipper* (**Figure 3**). *Fr* values were highly variable in both models, resulting in a high degree of uncertainty in modeling *R* of smaller (< 6 m) whales (**Figure 3**). Still, the divergence in *R* between males and females after *chm* was consistently more pronounced in models based on *R – Flipper* metrics (**Figure 3**). Estimates of asymptote parameters (*maxf* and *maxm*) were generally more stable than growth parameters (*fr* and *mr*), although some iterations of the *R – Dorsal* model resulted in distant outliers of the male-specific parameters (*mr* and *maxm*; **Figure 4**)

**Table 2**. Bootstrapped means and 95th percentile confidence intervals (95% CI) based on 1000 iterations for parameters relating sperm whale length (m) and nose-to-body ratio (R) metrics based on snout to the caudal base of the dorsal fin (R – Dorsal) and on snout to the flipper insertion point (R – Flipper). Parameters reflect the growth rate of females and small males (≤ 6 m) (fr), the female asymptote of R (maxf), the growth rate of larger males (> 6 m) (mr), and the male asymptote of R (maxm).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *R* Metric | *fr* [95% CI] | *maxf*[95% CI] | *mr* [95% CI] | *maxm*[95% CI] |
| *R* - *Dorsal* | 2.8 (0.63 - 14.9) | 0.65 (0.64 - 0.65) | 0.2 (0.01 - 0.62) | 0.89 (0.22 - 4.79) |
| *R - Flipper* | 2.26 (0.5 - 33.64) | 0.3 (0.3 - 0.3) | 0.05 (0.01 - 0.16) | 2.05 (0.45 - 6.89) |

A comparison of different colored lines

AI-generated content may be incorrect.

**Figure 3.** Bootstrapped logistic curves of the total length (m) and the nose-to-body ratio of sperm whales based on measures of the snout to the caudal base of the dorsal fin (A) and snout to the base of the flipper (B). Theoretical male curves are shown in violet and theoretical female curves are shown in green. The average R values across iterations are shown by light violet and green lines for males and females respectively. Dashed vertical lines indicate the minimum body lengths associated with sperm whale sex and age classes as follows: length at birth (4 m; NB), juvenile (J; 5.5 m), sub-adult (SA; 7.6 m ), adult female (AF – 8.5 m), adult male and mature female (AM/MF – 10 m), maximum female length (Fmax – 12 m), and mature male (MM – 13.7). (Best 1979, Best et al. 1984, Rice 1989, Mendes et al. 2007).

A screenshot of a graph

AI-generated content may be incorrect.

**Figure 4.** Distribution of bootstrapped parameter estimates for R – Flipper and R – Dorsal models.

#### 3.3.2 Posterior probabilities

Models based on *R – Dorsal* metrics resulted in higher uncertainty for individual *P(f)* estimates than *R – Flipper* models (**Figure 5**). Models fit with *R – Flipper* consistently assigned a high likelihood of an individual being female (*P(f) > 0.95*) to 21 individuals ranging from 9.8 – 12.5 m and *R – flipper* measures between 0.28 – 0.31. This length range coincides with the 10 – 12 m range of mature females based on whaling data (Best et al. 1984). *R -Flipper* models also resulted in a near-zero *P(f)* values for 5 individuals 5 individuals between 12.9 – 16.1 m and *R flipper* 0.38 – 0.41, which can be classified as males based on their length and behavioural context. Conversely, in models fit with *R – Dorsal*, only two individuals that could be assumed to be mature males based on their sizes (ID = 01 & 81) were consistently assigned low *P(f)* values. No individuals were consistently assigned a high *P(f)* value based on *R – Dorsal* models.

Models based on different *R* metrics did not always assign similar *P(f)* values to individual whales. For example, while individual 74, which was observed being suckled by a smaller whale, was estimated to have a high *P(f)* with high certainty (0.99) based on *R – Flipper* metrics, had a low mean *P(f)* (0.12, CI = 0 – 0.44). Likewise, individual 04 is a large male observed alone and known be a male based on behavioural context. While *R – flipper* based models consistently assigned a near-zero *P(f)* estimate, *R – Dorsal* models assigned non-classifyiable values (~0.5; CI = ).

A diagram of a diagram of a variety of colored circles

AI-generated content may be incorrect.

**Figure 5.** Bootstrapped mean Length (m) and nose-to-body ratio (R) for individual sperm whales based on (A) rostrum – dorsal fin and (B) rostrum – flipper. Point colours reflect the mean posterior probability of individuals being female, P(fem), and point sizes indicate the 95th percentile confidence interval width for P(Fem). Point shape denotes whether individuals were observed being suckled on by other individuals. Individuals that were observed suckling and known males (> 13.7 m) are labeled for reference. Dashed vertical lines indicate the minimum body lengths associated with sperm whale sex and age classes as follows: calf (4 m; NB), juvenile (J; 5.5 m), sub-adult (SA; 7.6 m ), adult female (AF – 8.5 m), adult male and mature female (AM/MF – 10 m), maximum female length (Fmax – 12 m), and mature male (MM – 13.7). (Best 1979, Best et al. 1984, Rice 1989, Mendes et al. 2007).

- Peduncle dive patterns.

We found 6 of the measured individuals onto which other whales performed suckling dives whales. The p(fem) based on HD ratio of these individuals ranged from bootstrapped averages xxx – xxx. The p(fem based on HF ratio of the same individuals was higher, ranging from 0.60 – 0.99. Indivi

## Discussion

* Parameter estimates for both models could be variable resulting from inter-image variability in measurements.
* But the model based on R – Flipper showed that whales > 9.8 length and R <0.31 could be reliably identified as females, helping refine existing sex/age classes for sperm whales. Models based on R – Dorsal had too much uncertainty, even in cases of known sex (i.e. very large males), likely as a result of how reliably this measurement can be taken. Even if R – Flipper can be measured under more specific conditions, it better captures sexual dimorphism.
* The threshold at which we can identify females coincides with xxx. maturity level based on whaling data.
* Analyzing behaviour in at this sex classification (although not perfect) can produce valuable insights: e.g., suckling was most often detected on individuals with a high probability of being mature females, but also observed in individuals that had uncertain classification status. Not observed in small individuals, and not observed in adults. This is preliminary data, but provides an example of what can be inferred.

Future work:

* 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
* Ground-trughing would be really cool