# Inferring sperm whale (*Physeter macrocephalus*) sex and age class 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 class using aerial photogrammetry 1](#_Toc202983197)

[1 | INTRODUCTION 3](#_Toc202983198)

[2 | METHODS 5](#_Toc202983199)

[2.1 | Data Collection 5](#_Toc202983200)

[2.2 | Morphometric measurements 5](#_Toc202983201)

[2.3 | Photo-identifying whales 8](#_Toc202983202)

[2.4 | Inferring sex and age class 9](#_Toc202983203)

[3. RESULTS 13](#_Toc202983204)

[3.1 | Error estimation and correction 13](#_Toc202983205)

[3.2 | Whale measurements and photo-identification 13](#_Toc202983206)

[3.3 | Sex and age class inference 14](#_Toc202983207)

[4. DISCUSSION 19](#_Toc202983208)

## 1 | INTRODUCTION

Demographic data (i.e., sex and age) provide a key dimension to our understanding of the biology and conservation status of species. At an individual level, animals of different sex/age classes adopt distinct social and ecological behaviours that reflect differences in their reproductive strategies and metabolic needs (Bleich et al. 1997, Griffiths et al. 2014). From a population conservation standpoint, knowledge of demographic structure is essential for assessments of the reproductive potential and overall conservation status (Volis & Deng 2020). But, while some sex and age classes can be easily discerned in the field based on their size, appearance, and behaviour—like newborns and mature males in sexually dimorphic species—others—like physically immature individuals—may be hard to tell apart. Sexing live individuals can involve genital inspection, which may not be feasible for some species, or relatively costly molecular analyses of samples collected using variably invasive techniques (Shaw et al. 2003). Likewise, delineating informative age-classes of live animals requires tracking individuals over time or implementing molecular analyses (Le Clercq et al. 2023), both of which can be financially and logistically challenging.

Identifying sex and age classes is particularly hard to do in wild cetaceans, as their bodies are often submerged making visually based assessments of age/sex impractical. This has resulted in field-based research on cetaceans classifying individuals into coarse age classes without sex distinctions (calves, juveniles, and adults), or into age/sex classes that lump together immature males with mature females (e.g., Gowans et al. 2001, Denkinger et al. 2020). The behaviour of sperm whales (*Physeter macrocephalus*)off the Galápagos Islands has been the focus of a multi-decade research project since the 1980s (Whitehead 2003, Cantor et al. 2017). To date, individuals have been classified into four broad age/sex classes: calves are considerably small (ca. < 5.5m) individuals found near other larger whales; mature males are considerably large (ca. > 12 m) individuals; bachelor males are other individuals found in small (< 4) groups; and mature females/immature individuals are all other whales found in larger groups (Christal & Whitehead 1997, Whitehead 2003, Cantor et al. 2017). Although mature males can be reliably identified in the field, as they can be 40% longer and weigh three times as much as mature females (Rice 1989), the distinction between bachelor males, immature individuals and mature females is less clear. As the behaviours of mature females and immature males/females are shaped by different social and ecological processes (Eguiguren et al. 2023), this grouping likely masks information about their behaviour and population structure.

The emergence of uncrewed aerial vehicles (UAVs) has allowed researchers to extract precise morphometric measurements of free-ranging cetaceans with minimal impact on their behaviour (Burnett et al. 2019, Bierlich et al. 2021, Glarou et al. 2022). More recently, UAV-derived morphometric measurements have been used to delineate age-classes (Vivier et al. 2024) and even reproductive status of wild cetaceans (Fernandez Ajó et al. 2023, Robinson & Visona-Kelly 2025). However, these methods have been based and tested on known details of individuals’ age and reproductive status, which are not available for the Galápagos sperm whales.

Here, we developed a method to infer the age/sex of sperm whales based on morphometric measurements, together with prior knowledge of sperm growth and development, for sperm whales of unknown age/sex observed off the Galápagos Islands. We synthesized existing data on the development and length-age relationship of sperm whales to refine age-class groupings. To distinguish sexes within age-classes, we relied on the male’s disproportionately larger nose (Cranford 1999). This feature accounts for most of the difference in size between males and females and, although most pronounced in mature males, can be detected before adulthood via direct measurements (Nishiwaki et al. 1963, Cranford 1999). Using AUV-derived morphometric measurements of Galápagos sperm whales of unknown sex/age, we investigated the uncertainty associated with measurements of total length and nose-to-body ratio and developed a model-optimizing algorithm to estimate the probability that individuals are females based on their total body length and nose-to-body ratio. To demonstrate the application of our methods, we explored individuals’ involvement in *peduncle dives*—a stereotyped interaction which has thus far been reported only between calves/juveniles and females—in light of our sex/age class inferences.

## 2 | METHODS

### 2.1 | Data Collection

We carried out dedicated surveys in the deep waters (> 1000 m) off the Galápagos Islands aboard a 12 m sailboat (*Balaena*) between January and May 2023 (Galápagos National Park 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 1080 pixel 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 the equation of Bierlich et al. (2021):

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 α by obtaining measurements of a known object of known length *L* at a known distance (*H*)in the lab and then 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, we 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*), and two alternative nose length measures—snout-to-flipper length (*SnF*) and snout-to-dorsal-fin length (*SnD*)—in pixels (Figure 1). *TL* was measured piecewise (see Figure 1) from the snout to the fluke notch, *SnF* was measured from the snout to the transversal intersection of the base of the flippers with the spine, and *SnD* was measured from the snout to the caudal base of the dorsal fin. To estimate nose proportions, we calculated the nose-to-body ratio (*NR*) by dividing *SnF* or *SnD* by *TL* (all in pixels), resulting in two metrics: *NRflipper* and *NRdorsal,* respectively. *TL* was converted from pixels to meters using Equation 2 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.



**Figure 1.** Aerial images of a sperm whale showing landmarks used to measure sperm whale morphometry (s = snout; f = flipper insertion point; 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 (a). Snout-to-dorsal fin length (SnD) measures the distance from s to d (b). Snout-to-flipper length (SnF) measures the distance from s to the midpoint of the spine that intersects perpendicularly with f (c).

### 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 sex and age class

#### 2.4.1 Sex

Much of this size difference arises from the disproportionately larger ‘noses’ of males, which can account for c.a. 40% of their total length, compared to up to 30% of the female’s (Cranford 1999). Sperm whale’s noses encase organs involved in the production of *clicks*, which they use for echolocation and communication (Norris & Harvey 1972). Sperm whale clicks have a particular multi-pulse structure that result from sound waves emitted near the blow hole and bouncing against the base of the skull (Norris & Harvey 1972). Because the time elapsed between pulses is proportional to the size of the sperm whale’s nose, it has been theorized that the hypertrophy of male sperm whales’ nose serves as a means to acoustically exaggerate and advertise their size to females and other males (Cranford 1999, Madsen et al. 2002). Although the hypertrophy of male sperm whales’ noses is most notable when they achieve maturity (> 20 years), it becomes detectable in older juveniles (ca. 2 years – 6 m) and intensifies with age (Nishiwaki et al. 1963).

The relationship between sperm whale length (*L*) and nose-to-body ratio (*NR*) 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 modeled as follows:

Where is the maximum (asymptote) *NR* of female whales, and is the initial rate of change in *NR* with increasing length. For males, the relationship between body length and *NR* 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 *NR* of a male compared to a female of the same size, and is the initial rate of change in *NR* with length when length is greater than the threshold (*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 et al.’s (1963) figure showing the relationship between total body length and *NR* estimates in which nose length was measured from the tip of the snout to the 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 ():

Where *n* is the total number of individuals, *resfi*is the residual between an individual’s observed *NR* and that predicted by the female curve, and *SS* is the sum of the residuals (both for the female and male curves).We then computed the posterior probability of an individual being a female (, assuming males and females are equally likely a priori, by:

To propagate the uncertainty associated with inter-image variation of in estimates of individual probability of being female, we applied a stratified bootstrap simulation 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 probabilities of being female. Resulting estimates 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, NRflipper*, and *NRdorsal*.

#### 2.4.2 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 be present in the stomach, solid foods are more frequently found. Still, the majority of 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 of 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. |

#### 2.4.3 Behavioural context

We then inspected whether individual whales performed or received *peduncle dives* across age classes and differing probabilities of being female. *Peduncle dives* are short (a few seconds) and shallow dives performed usually by a calf or juvenile onto the base of the peduncle (fluke stalk) and under a 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-based recordings when calves arch their backs and dive under a larger whale’s body repeatedly. For each measured whale, we recorded whether it had been observed performing or receiving a *peduncle dive* in any of the video recordings from which still images for measurements were extracted.

Peduncle dives were previously assumed to indicate suckling (Gero & Whitehead 2007), but recent underwater footage suggests that they may not be associated with suckling or milk extraction, and instead represent a form of affiliative behaviour between young whales and mothers/allomothers (Sarano et al. 2023). Although peduncle dives may not necessarily involve suckling, all reports of *peduncle dives* in which the sex of the receiving whales is known involve females (Gero & Whitehead 2007, Konrad et al. 2019, Sarano et al. 2023).

## 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%*). These length estimates resulted in an observed 95% CI width of 0.2 m, corresponding to 1.6% of the vessel’s true length. Equivalent to CV = 3.15%.

### 3.2 | Whale measurements and photo-identification

We were able to extract *NRdorsal* metrics more frequently than *NRflipper* (491 and 297, respectively). Only images captured at altitudes up to 70 m had high enough quality (Q3 – 5) for initial identification (Figure 2), resulting in 504 still images assigned to 90 individuals for which *TL* could be measured at least once, and a subset of 168 still images assigned to 51 individuals for which *TL, NRdorsal,* and *NRflipper* could be measured at least three times.

A graph with a number of boxes

AI-generated content may be incorrect.

**Figure 2.** Corrected drone altitude (m) distribution across photo quality ratings (Q1 -5) of still images. The 70 m threshold is shown for reference.

### 3.3 | Age class and sex inference

#### 3.3.1 Length estimation and age class assignment

The 95% CI width in bootstrapped estimates of sperm whale *TL* had a median of 0.35 m (mean = 0.42, *SD =* 0.32). This represented a median of 3.29% of the mean *TL* for each individual (mean = 4.18%, *SD =* 3.34%). Resulting *TL* estimates ranged from 4.1 -16.1 m, with most individuals measuring between 8 – 13 m (**Figure 3**).

A graph of lines and dots

AI-generated content may be incorrect.

**Figure 3.** Total body length (m) and nose-to-body ratio (NRflipper) estimates of individual sperm whales. Point locations show the bootstrapped mean for each individual (N simulations = 1000), horizontal error bars show the corresponding 95% CI length range, and vertical error bars show the 95% CI NRflipper range. The dashed vertical lines indicate the minimum body lengths associated with sperm whale sex and age classes based on Best 1979, Best et al. 1984, and Mendes et al. 2007 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).

Bootstrapped *NRdorsal*measurements were associated with wider and more variable 95% CI’s than *NRflipper* measurements (*NRdorsal* mean = 0.023, SD = 0.02; *NRflipper* mean = 0.018, 0.012).

#### 3.3.2 Parameter optimization

Optimal values for *fr, maxf, mr and maxm* varied more across bootstrap iterations of the models fit with *NRdorsal* than *NRflipper* (**Table 2**), resulting in generally higher levels of uncertainty associated with models based on *NRdorsal* than *NRflipper* (**Figure 3**). *fr* values were highly variable in both models, resulting in a high degree of uncertainty in modelling the *NR* of smaller (< 6 m) whales (**Figure 3**). Still, the divergence in *NR* between males and females after *chm* was consistently more pronounced in models based on *NRflipper*, partly because large males (> 13. 7 m) had disproportionately higher *NRflipper* than the rest of individuals, while their *NRdorsal* are closer to those the general population (**Figure 5**). Estimates of asymptote parameters (*maxf* and *maxm*) were generally less variable than growth parameters (*fr* and *mr*), although some iterations of *NRdorsal* models resulted in distant outliers of the male-specific parameters (*mr* and *maxm*; **Figure 4**). For adult males, *NRflipper* seems to increase linearly with length (Figures 3 and 4), and thus the logistic model is an unnecessary elaboration (also indicated by *mr*=~0; Table 2), and *maxm* is meaningless (indicated by *maxm* estimated to be impossibly greater than 1.0).

**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 (NR) metrics based on snout to the caudal base of the dorsal fin (NRdorsal) and on snout to the flipper insertion point (NRflipper). 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] |
| *NRdorsal* | 2.8 (0.63 - 14.9) | 0.65 (0.64 - 0.65) | 0.2 (0.01 - 0.62) | 0.89 (0.22 - 4.79) |
| *NRflipper* | 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 a line graph

AI-generated content may be incorrect.

**Figure 4.** 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 NR values across iterations are shown by light violet dashed and green solid lines for males and females, respectively. c

A diagram of a graph

AI-generated content may be incorrect.

**Figure 5.** Distribution of bootstrapped parameter estimates (x axis) for NRdorsal and NRflipper models.

#### 3.3.2 Posterior probabilities of being female

Models based on *NRdorsal* metrics resulted in higher uncertainty for individual estimates of the probability of being female than *NRflipper* models (**Figure 5**). Models fit with *NRflipper* consistently (which we here define as cases when the individual bootstrapped 95% *CI* width for *P(f)* estimates< 0.05) assigned a high probability (*P(f) > 0.95*) of an individual being female to 21 individuals ranging from 9.8 – 12.5 m and *NRflipper* 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). The *NRflipper* model also resulted in a consistently low probability (*P(f)* < 0.05) of individuals being female for 5 individuals between 12.9 – 16.1 m and *NRflipper* 0.38 – 0.41, which can be classified as males based on their length considerably exceeding the maximum recorded female length. Conversely, in models fit with *NRdorsal*, only two individuals that could be assumed to be mature males based on their sizes (ID01 & ID81) were consistently assigned low probabilities of being females. No individuals were consistently assigned a high probability of being female consistently based on *NRdorsal* models.

Models based on *NRflipper* metrics resulted in more reasonable individual estimates of the probability of being female than *NRdorsal* based models. For example, individual ID74 (mean *TL =* 10.78 m, 95% *CI =* 10.63 – 11.06 m), which was observed receiving peduncle dives, was classified with high confidence as female by *NRflipper* models (mean *P(f)* = 0.99, 95% *CI =* 0.99 – 1.00), yet received low and uncertain estimatesfrom the *NRdorsal* model(mean = 0.12, 95% *CI* = 0 – 0.44). Similarly, individual ID04, a large male (mean *TL =* 15.2 m, 95% *CI =* 14.9 – 15.5 m), was confidently assigned a near-zero probability of being female by the *NRflipper* model (mean < 0.001, 95% *CI width =* 0), but received an uncertain and intermediate probability estimate based on the *NRdorsal* models (mean = 0.50, 95% *CI* = <0.001 – 0.97).Given that *NRflipper* models resulted in more certainty and consistency with context, behaviour, and whaling data, we explore the peduncle dive patterns in the following section considering the predictions made by *NRflipper* models.

A diagram of a graph

AI-generated content may be incorrect.

**Figure 6.** Bootstrapped mean Length (m) and nose-to-body ratio (NR) for individual sperm whales based on (a) snout – dorsal fin distance (NRdorsal) and (b) snout – flipper distance (NRflipper). The solid green line and dashed violet line show the bootstrapped mean modele NR for females and males, respectively. Point colours show the mean posterior probability of individuals being female (P(f)), and point sizes indicate the bootstrapped 95th percentile confidence interval (95% CI) width for P(f). Point shape denotes whether individuals were observed involved in peduncle dives (triangles = receiving, squares = doing, circles = none). Individuals that were observed receiving peduncle dives and mature males (> 13.7 m) are labelled for reference. Dashed vertical lines indicate the minimum body lengths associated with sperm whale sex and age classes based on Best 1979, Best et al. 1984, and Mendes et al. 2007 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).

#### 3.3.3 Peduncle dive patterns

We inspected xxx mins of the footage from which we extracted whale measurements. Within this footage, we found three individuals doing and 12 individuals receiving peduncle dives out of the 90 individuals for which we had at least one total length measurement (**Figure 6**). We were able to measure more individuals receiving peduncle dives than those performing them because the frequent diving involved in performing peduncle dives often resulted in an arched body position which was not suitable for accurate length measurements.

Length measurements of individuals that performed peduncle dives either fell within the total length ranges corresponding to calves (n = 1) or juveniles (n = 3; **Figure 6**). Individuals that received peduncle dives ranged from 8.9 – 12.5 m length, corresponding to the overlapping age/sex classes that include adult to mature females and subadult – adult males. Four out of seven individuals for which we could measure *NRflipper* –and thus estimate their probability of being female—had a high probability and certainty of being female. The remaining three individuals had slightly lower probabilities of being female associated with a high degree of uncertainty based on *NRflipper* models (ID11 P(f) = 0.64, 95% CI [<0.01, 1.00]; ID75 P(f) = 0.62, 95% CI [0.13, 0.97]; ID76 P(f) = 0.88, 95% CI [0.55, 1.00]; **Figure 5**).

A diagram of a graph

AI-generated content may be incorrect.

**Figure 7.** Mean total length (m) distribution of individual sperm whales observed doing, receiving, or not involved in peduncle dives (PD). Points are colored by the bootstrapped mean P(f) estimated using NRflipper models or are shown as asterisks if NRflipper could not be measured.

## 4. 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. OTHER EXAMPLES: MATING BEHAVIOUR; SAME/DIFFERENT SEXES TOUCHING; ….

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