# Inferring sperm whale (*Physeter macrocephalus*) sex and developmental stage using aerial photogrammetry

Ana Eguiguren1\*, David Gaspard1, Christine M. K. Clarke1, Hal Whitehead1

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

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

## 1 | INTRODUCTION

Demographic data (i.e., sex and age) provide a key dimension for understanding behaviour and evaluating reproductive status of a population. At an individual level, animals of different developmental stages and sexes may adopt distinct social and ecological behaviours that reflect differences in their reproductive strategies and metabolic needs (Bleich et al. 1997, Ruckstuhl 2007, Griffiths et al. 2014). From a population conservation standpoint, knowledge of the demographic structure is essential for assessing life history parameters and their changes over time (Volis & Deng 2020). In some cases, sex and developmental stages can be easily discerned in the field based on body size, appearance, and behaviour—such as categorizing small, young developmental stages like newborns and determining sex in mature individuals in sexually dimorphic species. But, physically immature individuals can be hard to distinguish from mature ones, and the sexes may be hard to tell apart at this stage. Thus, identifying developmental stages of live animals has often required tracking individuals over time or implementing hormonal analyses (Le Clercq et al. 2023), both of which can be invasive and financially and logistically challenging. Likewise, sexing live individuals in the absence of obvious sexual dimorphism 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).

Identifying sex and developmental stage is particularly hard to do in wild cetaceans, as their bodies are often submerged, which makes visually based assessments impractical. As a result, studies of cetaceans in the wild frequently classify individuals into coarse developmental classes without sex distinctions (calves, juveniles, and adults), or, in species with male-biased sexual size dimorphism, such as sperm whales, into developmental/sex classes that lump together immature males with mature females (e.g., Gowans et al. 2001, Denkinger et al. 2020).

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 and wellbeing (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 the reproductive status of wild cetaceans (Cheney et al. 2022, Fernandez Ajó et al. 2023, Robinson & Visona-Kelly 2025).

Sperm whales (*Physeter macrocephalus*)off the Galápagos Islands have been the focus of a multi-decade research project spanning 1985 - 2023 (Whitehead 2003, Cantor et al. 2017). Because sperm whales in the region are highly mobile (Whitehead et al. 2008) individuals are only feasible to track over a few days and are rarely re-sighted over several decades, making observation-based assessments of their age impractical. Thus, individuals have been classified into four broad developmental/sex classes: calves are considerably small (ca. < 5.5m) individuals found near other larger whales; mature males are considerably large (ca. > 12 m) individuals; immature bachelor males are other individuals found in small (< 4 individuals) 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 not 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 masks important details about their behaviour and population structure.

Informed by age and sex specific morphological data, we developed a method to infer the developmental stage and sex of sperm whales based UAV-derived morphometric measurements, in the absence of known individual demographic data. We first defined size ranges corresponding to finer-scale developmental stages using existing data on sperm whale growth and length-age relationships derived from analyses of thousands of individuals killed during industrial whaling. To distinguish individual sex, we relied on the male’s extreme sexual dimorphism and particularly their disproportionately larger nose, which (when measured from the base of the skull to the tip of the snout) can account for c.a. 40% of their total length, compared to up to 30% of the females’ (Cranford 1999; Box 1). Although the hypertrophy of male sperm whales’ noses is most notable when they reach physical maturity (> 20 years), it can be detectable in older juveniles (ca. 2 years – 6 m) and intensifies with age (Nishiwaki et al. 1963). We 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 developmental stage/sex 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 visually during daylight hours and acoustically (using a 100-m towed hydrophone). When we encountered groups of females and juveniles, we followed them for as long as possible at a cautious distance (c.a. 50 m) to collect behavioural, acoustic, and photo-identification data.

If conditions were adequate (windspeed < 10 kt and no rain), we conducted 1 – 2 hour flight sessions, composed by a series of consecutive flights, 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 from direct sunlight on the water interfered the least with visibility. Over the whales, we flew between 15 - 120 m above the water and pointed the camera down perpendicularly (i.e., nadir). During flights, we recorded continuously at a resolution of 1080 x 1902 or 3840 x 2160 px (4K) resolution, both at 29.79 fps. We alternated between 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)–and 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 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 is altitude measurement error, which impacts the scaling factor used to estimate true object lengths (Burnett et al. 2019, Bierlich et al. 2021, Glarou et al. 2022, Napoli et al. 2024a). Our morphometric estimates were based on altitude measurements from the built-in barometric altimeter, which is less precise (i.e., has more variation) and less accurate (i.e., approaches true values less often) than measurements obtained from laser-based altimeters that are typically used in cetacean photogrammetric research (Burnett et al. 2019, Bierlich et al. 2021). We did not incorporate a laser altimeter on our UAV as it would have exceeded the payload capacity of the airframe and significantly reduced flight time, given the UAV’s small size. We opted for this UAV model because our initial attempts to fly and retrieve a larger UAV (Phantom 4 Pro) equipped with a laser altimeter from the research vessel were largely unsuccessful due to a combination of pilot inexperience (AE) and the inherent challenges of landing a drone on a sailboat at sea. As our main interest was to collect video recordings for analyzing sperm whale behaviour, we opted for the UAV system with reduced of accuracy and precision in altitude measurement in exchange for longer flight times, reliable retrieval, and relatively low replacement cost.

To quantify the uncertainty in morphometric measurements and correct barometric altitudes of our UAV system, we used measurements of our research vessel (12.03 m) collected throughout the field season at various altitudes (27 – 120 m). 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* is the drone altitude above sea level, and α is a scaling factor 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:

To account for the possibility that barometric altitude biases would vary on different days as a result of changes in weather conditions, we also fit random effects models with date as a random intercept and slope. Although we found evidence for variation in the intercept and slope across different dates, this had a negligible effect on measurement error. This likely results from our study area being having very little variation in barometric pressure within and between days. The results of random effects models are shown in **Supplementary Material 1**.

#### 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 (described in **Supplementary Material 2**). 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, where the possible effects of lens distortion are negligible (Burnett et al. 2019), and when the drone camera was positioned at nadir relative to the water surface, as confirmed by the flight log data. We attempted to capture a broad size range of individuals, so we note that measured whales are not a random sample of the population.

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 from the snout to the fluke notch, *SnF* was measured from the snout to the transversal intersection of the anterior base of the flippers with the spine (when at least one flipper was visible), 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* (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 aimed 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. To compare the variability across images, we obtained average coefficients of variance (% CV), calculated by dividing the standard deviation (SD) by the mean x 100 for measures taken from the same individual.

A whale in the water

AI-generated content may be incorrect.

**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 | Photoidentification

We identified measured whales from their aerial photographs based on observable markings—including visible fluke marks, scars, indentations, rake marks, white patches, and sloughed skin patterns (O’Callaghan et al. 2024). We rated images used for measurement 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 developmental stage

#### 2.4.1 Sex

The relationship between sperm whale length (*L*) and nose-to-body ratio (*NR*) depicted 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 modelled 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 *NR* for male sperm whales 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 found 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 BFGS optimizing 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. We constrained the optimizing to *mr* values > 0 to ensure that modelled *NR* values based on male growth models would be higher than those of the female growth model.

The posterior probability that each individual was female was estimated based on Bayes’ theorem under the prior assumption that individuals would be equally likely of either sex. Estimates of individuals’ likelihood of being female () were computed based on how close each point fell to the ‘female curve’ following equation (8).

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).The likelihood of being male () was calculated similarly. We then computed the posterior probability of an individual being a female ( using equation (9).

We report our results in terms of *P(f)* but note that the probability of an individual being male under this framework is the direct opposite (*P(m) = 1 – P(f)).*

To propagate the uncertainty associated with inter-image variation in estimates of individual probability of being female, we applied a stratified bootstrap simulation by individual ID (Dixon 2001, Napoli et al. 2024a). In each of 1000 iterations, we randomly sampled a set of measurements 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*.

We carried out robustness checks evaluate the effect of our modelling assumptions on individual’s *P(f)* estimates to our modelling decisions. Specifically, we systematically varied *chm* values within a reasonable range (5 – 7 m) to compare resulting posterior *P(f)* estimates. We also computed posterior *P(f)* estimates using on a prior *P(f)* set to 0.79, which corresponds to the proportion of females in breeding groups off the Galápagos Islands genetically determined in 1991 (Richard et al. 1996).

#### 2.4.1 Developmental stages

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

**Table 1.** Developmental stages 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 | < 4.1 | Few days - months | Unhealed umbilical regions, likely recently born (Best et al. 1984). |
| Calves | Both | 4.1 – 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 corresponds with them leaving their natal unit (Best 1979, Best et al. 1984, Mendes et al. 2007). |
| 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 usually performed by a calf or juvenile next to the base of the peduncle (fluke stalk) and under a larger whale, during which 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 have been assumed to indicate suckling or stimulate the release of milk (Gero & Whitehead 2007). Underwater footage suggests that they may not be associated with direct suckling, although the role of this behaviour in promoting milk letdown remains unclear (Sarano et al. 2023). Alternatively peduncle dives may represent a form of affiliative behaviour between young whales and mothers/allomothers (Sarano et al. 2023). Still, 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. 2019a, Sarano et al. 2023).

#### 2.4.4 Testing predictive performance

To examine the performance our method on whales observed in other populations, we also included morphometric measurements of two individuals, one observed in the North Atlantic (The Gully Canyon off the Scotian Shelf, Canada) and the other in the Arctic Ocean (Baffin Bay, Canada). The North Atlantic individual was recorded using a DJI Phantom 4Pro V2 + and the Arctic individual was recorded with a DJI Inspire 2 equipped with a DJI Zenmuse X5S micro four-thirds camera with Olympus M.Zuiko lens. UAV altitude measurements were obtained through laser altimetry using a LidarBoX system (Bierlich et al. 2024). We computed the posterior probabilities of these individuals being female based on their *TL* an *NRflipper* measurements and the optimized parameter values and individual likelihoods of being male or female from our original Galápagos dataset. Because these individuals were located in high latitudes (Rice 1989, Whitehead et al. 1992) we expected them to have near-zero probabilities of being female.

## 3. RESULTS

### 3.1 | Error estimation and correction

We obtained 343 measurements of *Balaena* across 18 days in the field between February 1st and April 29, 2023, 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 estimate error to 0.12 % (*SD = 3.15%,* 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 of 70 m or less were 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 | Developmental stage and sex inference

#### 3.3.1 Uncertainty in individual measurements and developmental stage assignments

Observed *TL* measurements of the same individual had an average 2% CV (*SD* = 1.39%). 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* (mean = 4.18%, *SD =* 3.34%). Resulting *TL* estimates ranged from 4.1 -16.1 m, with 80% of individuals measuring between 7.4 – 12.6 m (**Figure 3**). These estimates resulted in no individuals categorized as neonates, three as calves, three as juveniles, one as a subadult, and four as mature males. The remainder (n = 40) fell within age classes with overlapping ranges between males and females (i.e., AF, AM, and MF).

*NRflipper* measures ranged from 0.27 – 0.41 (mean = 0.31, SD = 0.03). On average, *NRflipper* measures had an observed 2.95% CV (SD =2.12%).



#### **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 bootstrapped 95% CI NRflipper range. The dashed vertical lines indicate the minimum body lengths associated with sperm whale sex and age defined in **Table 1** as follows: calf (C), juvenile (J), sub-adult (SA), adult female (AF adult male and mature female (AM/MF), maximum female length (Fmax), and mature male (MM – 13.7).

#### 3.3.2 Parameter optimization

We found the divergence between mature males and the rest of the measured whales was much less pronounced for *NRdorsal* measurements than *NRflipper* measurements, contributing to greater uncertainty in *P(f)* based on this metric. We describe results for models fit using *NRflipper* below and those for models fit using *NRdorsal* in **Supplementary Material 3**. Optimal *fr* values were variable across bootstrap iterations, resulting in a high degree of uncertainty in modelling the *NRflipper* of smaller (< 6 m) whales (**Figure 4** & **Figure 5**). Still, the divergence in *NRflipper* between males and females after *chm* was consistently pronounced, partly because large males (> 13. 7 m) had disproportionately higher *NRflipper* than the rest of individuals (**Figure 6**). Estimates of asymptote parameters (*maxf* and *maxm*) were less variable than growth parameters (*fr* and *mr*) (**Figure 4**). Additionally, for adult males *NRflipper* seems to increase linearly with length (**Figures 5**), and thus the logistic model may be an unnecessary elaboration. This is also supported by *maxm* often estimated to be impossibly greater than 1.0; **Figure 4**).

A diagram of a diagram

AI-generated content may be incorrect.

**Figure 4.** Distribution of bootstrapped parameter estimates modelling the growth rate of females and small males (≤ 6 m) (fr), the female asymptote of NR (maxf), the growth rate of larger males (> 6 m) (mr), and the male asymptote of NR (maxm).

A graph of a line

AI-generated content may be incorrect.

**Figure 5.** 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 anterior base of the dorsal fin (NRflipper). Theoretical male curves are shown in pink and theoretical female curves are shown in green. The average NRflipper values across iterations are shown by light violet dashed pink and green solid lines for males and females, respectively. The vertical line indicates the point of divergence between males and females (chm = 6 m) based on Nishiwaki et al. (1963).

#### 3.3.2 Posterior probabilities of being female

Models fit with *NRflipper* consistently—defined here as having bootstrapped 95% *CI* widths for *P(f)* <0.05—assigned a high probability (*P(f) > 0.95*) of being female to 21 individuals ranging from 9.8 – 12.5 m *TL* and between 0.28 – 0.31 *NRflipper* (**Figure 6)**. 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 12 m (Evans & Hindell 2004). Both individuals from the North Atlantic fell within the size range expected for males (> 12.5 m), relatively large *NRflipper* measurements (>0.38), and *P(f)* estimates <0.001 (**Figure 6**).

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 snout –– flipper distance (NRflipper). The solid green line and dashed pink line show the modeled NR for females and males, respectively. Point colours show the posterior probability of individuals being female (P(f)). **Points with black outlines have 95% CI ranges ≤ 0.05 for bootstrapped estimates of P(f)**. Point shape denotes whether individuals were observed involved in peduncle dives (triangles = receiving, squares = doing, circles = none). The rhombi show males observed in the North Atlantic and Arctic Oceans. 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).

*P(f)* estimates were robust to varying *chm* and prior *P(f)* values for most individuals, particularly for those that had consistently either high or low probabilities of being female (**Supplementary Material 4**). Individuals for which varying parameter values had a more considerable effect (i.e., > 0.05 difference in *P(f)* between scenarios) had generally intermediate posterior *P(f)* estimates (0.25-0.80), and wide bootstrapped 95% confidence intervals. Images of a sample of individuals and their corresponding *P(f)* values are shown in **Supplementary Material 5**.

#### 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 7**). 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 high probability and certainty of being female (*P(f) =* 1, 95% CI [0.99, 1.00]. The remaining three individuals had slightly lower probabilities of being female associated with a high degree of uncertainty (ID11 *P(f)* = 0.38, 95% CI [<0.01, 1.00]; ID75 *P(f)* = 0.61, 95% CI [0.12, 0.96]; ID76 *P(f)* = 0.91, 95% CI [0.52, 1.00]; **Figure 7**).

A graph showing a number of points

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

We developed a minimally invasive method of inferring sperm whale developmental stage and sex by leveraging prior knowledge on sperm whale morphometric development and sexual dimorphism. UAV-based body length (*TL*) estimates provide useful proxies for developmental stages and can help refine the traditionally used ‘calf/mature female-immature/mature male’ classification system. Applying Bayesian theory, we estimated the posterior probabilities of individuals belonging to either sex given their *TL* and *NRflipper*. Despite uncertainty arising from different sources of measurement error, we found that nose-to-body ratio measurements based on snout to flipper distances (*NRflipper*) reliably captured the development of sexual dimorphism in sperm whales’ noses, with male and female growth curves diverging with increasing length (Nishiwaki et al. 1963, Cranford 1999). Some individuals could be classified as males or females with high confidence based on their posterior probability estimates, while others lacked the certainty to be assigned as either. Our inspection of peduncle dive patterns (PD) illustrates how our numeric representation of morphological ‘femaleness’ and developmental stage inferences can inform behavioural analyses. Based on simple photogrammetric measurements and a low-cost UAV system, our approach can add key demographic information into sperm whale behavioural analyses and population models.

### 4.1 | Developmental stage inferences

Our estimates of total body length were within previously reported size ranges for sperm whales obtained through direct measurements (Best et al. 1984, Evans & Hindell 2004). The uncertainty in *TL* estimates of our UAV system (CV = 2.0%) was comparable to most boat-based photogrammetric methods relying on laser photogrammetry(CV = 1.3 - 5.1%; Gordon 1990, Dawson et al. 1995, Jaquet 2006). However, our UAV system had higher uncertainty than state-of-the-art approaches for measuring sperm whales with UAV systems equipped with laser altimeters (CV = 1.0%; Dickson et al. 2021). While some research objectives, like detecting individual changes in morphometry over time, may require a higher level of precision, some uncertainty may be acceptable in studies looking at general patterns across a population (e.g., Waters & Whitehead 1990). Additionally, information on measurement error can be incorporated in statistical analyses, allowing for a measured interpretation of resulting patterns and parameter estimates (e.g., Bierlich et al. 2021).

The size-based developmental stage classes we propose refine the existing field-based classification typically used for sperm whales across the globe. The size ranges of our proposed developmental stages (**Table 1**) are based on the size distributions at given developmental milestones (e.g., most individuals that rely exclusively on milk are under 5.5 m long; individuals that incorporate solid foods but still primarily rely on milk (i.e., juveniles) are between 5.5 – 7.6 m long) that are well grounded on anatomical, dietary, and gonadal analyses (Best et al. 1984). Inferences of age or developmental stages based on body size have been used in the past to model population parameters (Waters & Whitehead 1990). These inferences generally rely on growth curves that relate individual length measurements to age estimates based on dentin layer counts of killed or stranded individuals (Ohsumi 1977, Best et al. 1984).However, it is important to note that growth curves are accompanied by uncertainty arising from individual variation in size and development with age (Martin 1980). For example, observed *TL* measures for sperm whales have a standard deviation of up to 0.96 m at a given age (Waters & Whitehead 1990).

Recent work attempting to identify age-classes based on UAV-derived morphometric measures in common bottlenose dolphins, *Tursiops truncatus,* has shown that size-based age-class assignments perform poorly when age-bin definitions are narrow (Cheney et al. 2022, Vivier et al. 2023). This the inherent of converting a continuous measure (*TL*) into a categorical one (age-class). But, when bins are wide enough, they perform well (with 2 – 3 age class bins having >72.5% individuals assigned within 2 years of actual age-class ) (Vivier et al. 2023, 2025). Moreover, they found that size-based age classification was most accurate for individuals < 2 years old, which corresponds to the exponential phase of their growth curve, and less accurate once decreasing and stabilizing growth rates are achieved (Vivier et al. 2023). For sperm whales, the initial exponential growth phase takes place between 0 – 7 years (4.1 – 7.6 m), which would allow for an accurate distinction between calves (<5.5 m) and juveniles (5.5 – 7.6 m). Additionally, males experience a ‘secondary growth spurt’ after attaining sexual maturity (>10 m), which would make adult males (10 – 13.7 m) reliably distinguishable from mature males (> 13.7 m).

### 4.2 | Sex inferences

The overall shape of *NRflipper – TL* growth curves and the resulting posterior probability estimates of individuals being female were generally consistent with previous knowledge on sperm whale sexual dimorphism (Nishiwaki et al. 1963, Cranford 1999). This was true despite optimum parameter estimates being sensitive to measurement uncertainty (**Figure 4).** Namely, all whales > 13 m—corresponding to the adult/mature male size range (Best et al. 1984)—had consistently low probabilities of being female and whales with low *NRflipper* (<0.32) between 8.5 – 12 m—corresponding to the mature female size range—had a consistently high probability of being female (Best et al. 1984). Similarly, smaller individuals (<7.6 m) consistently had *P(f)* ~ 0.5, which is consistent with the expectation that in immature individuals, sexual dimorphism, although present, is harder to detect (Nishiwaki et al. 1963). Hence, our methods allow reliable sex identification (as adult/mature females) of the majority of individuals within the traditional female-immature age-sex class Likewise, our approach correctly assigned a *P(f) ~ 0* to the whales observed in the North Atlantic and Arctic Oceans, where the vast majority of individuals are known to be adult and mature males (Martin 1980). Although this represents a very small sample size (n = 2), it demonstrates the applicability of our methods for data collected in other regions and using different UAV platforms.

Still, our approach resulted in some individuals having high levels of uncertainty and intermediate (i.e. ~ 0.5) *P(f)* values, despite having *TL* ranges (>8.5 m) at which sexual dimorphism should be detectable based on direct measurements (Nishiwaki et al. 1963). The uncertainty in *P(f)* estimates may partly be due to the variability associated with our measurement system, particularly for individuals with wide 95% CI estimates. Intermediate *P(f)* values may also reflect individual variation in levels of sexual dimorphism in secondary sex traits (McLaughlin et al. 2023), which would make distinguishing subadult males from adult and mature females particularly challenging. Unfortunately, *NR-TL* curves in Nishiwaki et al. (1963) are based on mean measurements, so we don’t have a baseline for the naturally occurring variation across individuals. Additionally, there are reports across cetacean species of individuals with partial or full hermaphroditism in their genital organs, which in some cases, is linked to intersex chromosome arrangements (Einfeldt et al. 2019). Whether these variations translate to ‘intermediate’ secondary sex traits has not been explored. Still, these findings highlight that some caution should be taken when assuming a direct link between phenotype and chromosome arrangement. In the future, individuals found to have uncertain *P(f)* values could be targeted for genetic sampling.

Notably, we found that very few individuals 7.5 - 12.5m had consistently low *P(f)* values—i.e., were likely males—and that those that did fell below the modelled male *NRflipper*curve (**Figure 5**). The absence of individuals with higher *NRflipper* ratios within this size range may partly reflect the expected departure of young males from their natal units, with most individuals expected to leave when they attain slightly under 10 m (between 7 – 11 years old; **Table 1** (Best et al. 1984, Waters & Whitehead 1990)). Because our fieldwork was focused on large groups which are generally composed of mature females and immature individuals (Best 1979), it is likely that adult (i.e., sexually mature) males were underrepresented in our sample. Despite this we conservatively assumed equal prior probabilities of observing each sex. This may have underestimated the probability of being female for some intermediate subjects. When we changed the prior expected sex ratio to 0.79 (the proportion of females in the breeding groups of Richard et al.’s (1996) study off the Galápagos Islands), the posterior probability of individuals with originally intermediate *P(f)* values increased considerably (**Supplement 4 – Figure S4-2)**. However, implementing this informed prior resulted in unrealistically high probabilities of individuals being female for calves (original = 0.5, updated = 0.79), given the even sex ratio of sperm whales in birth and early years (Best et al. 1984). Thus, we consider our conservative prior to produce a better representation of the morphometric different between males and females throughout their development. In the future, including known sex individuals, especially immature males, may help develop models that can better distinguish the sex of individuals within this size range.

We found consistent support for a constant (i.e., linear) increase of *NRflipper* with respect to *TL* for males between 6 – 16.1 m (**Figure 5**). This linear trend in *NRflipper* growth emerged despite our initial implementation of a logistic model. The observed pattern aligns with the nose-to-body ratiorelationship with body length reported by Nishiwaki et al. (1963). Although our dataset did not cover the full length span of mature males, which can reach over > 18 m (Kasuya 1991), our findings support a continued growth of the male sperm whales’ nose despite their total body length growth rate tapering off at 15 m (40 years). This pattern may that the growth in larger males, rather than indeterminate skeletal growth, is primarily driven by the growth of soft tissues encasing the sperm whales’ spermaceti organ, which in older males visibly protrudes beyond the lower jaw when seen from the side (Cranford 1999).

Sustained growth of secondary sexual traits well beyond sexual maturity has also been observed in other mammal species with high degrees of sexual dimorphism, including giraffes and elephants (Simmons & Scheepers 1996, Hollister-Smith et al. 2007). There is direct evidence that male giraffes with longer necks and larger-bodied elephants have higher reproductive success (Simmons & Scheepers 1996, Hollister-Smith et al. 2007). While the contribution of larger noses to male sperm whales’ reproductive success remains untested, our findings further indicate that strong sexual selective pressures are acting on this trait as it continues to grow despite the potentially high energetic cost of building lipid-rich tissue (Cranford 1999).

### 4.3 | Peduncle dive patterns

Our inspection of PD patterns in relation to inferences of sperm whale age and sex illustrates the applicability of our methods for investigating behavioural patterns. Most (4 of 7) individuals that received PD had consistently high probabilities of being female and ranged between 9.8 – 12.5 m (**Figure 5**), which suggests these are most likely mature females (Best et al. 1984). One individual receiving PD within this size range had relatively low probability of being female,but this estimate was associated with very low certainty (*P(f)* = 0.36, 95% CI width = 1), which would make inferences about its sex and developmental stage unwarranted.The remaining individuals observed receiving PDs, for which *P(f)* could not be estimated (n = 5), also fell within this size range (**Figure 6**). Two of the individuals observed receiving PDs fell in the a ‘adult female’ size range (8.5 – 10 m – **Table 1**) had high *P(f)* estimates (>0.6), albeit a low associated certainty (95% CI width > 0.80). While young females of this size range are capable of conceiving (Rice 1989), receiving PDs is not exclusive to parous females (Konrad et al. 2019b). Although our approach did not have the resolution to confidently assign all individuals receiving PDs as females, our results generally agree with previous work showing that only females receive peduncle dives (Gero & Whitehead 2007, Konrad et al. 2019b, Sarano et al. 2023).

We also found all individuals performing peduncle dives were under 7.6 m (**Figure 6**), corresponding to the size range of juveniles (n = 2) and calves (n = 1), which is also congruent with previous work (Gero & Whitehead 2007, Konrad et al. 2019b, Sarano et al. 2023). Our methods for detecting participation of PD were not exhaustive, as we only inspected a subset of available footage, and thus cannot rule out the participation of any of the remaining individuals in this behaviour. Still, our findings generally aligned with the expectation that this behaviour is limited to calves/juveniles performing the dives, and females receiving them, even if its direct association with suckling remains unclear (Konrad et al. 2019b, Sarano et al. 2023).

### 4.4 | Applications

Refined definitions of developmental stages can contribute to our understanding of behavioural development. For instance, investigating the interactions and spatial arrangement between immature individuals and their mothers or caregivers can provide insights into the behavioural development of immature individuals and corresponding changes in maternal care (Mann & Smuts 1998, 1999, Fellner et al. 2013). Until now, the systematic study of these behavioural changes has been mostly limited to research on captive individuals or wild populations with extraordinary conditions that allow for longitudinal research approaches (i.e., repeated observations over time of few individuals) to individual behaviour (Mann & Smuts 1998, Fellner et al. 2013, Sakai et al. 2013, Eguiguren et al. 2025b). Using UAV-derived *TL* estimates, either as continuous or categorical proxies for development, could yield similar insights through a cross-sectional (i.e., observations at a given time across several individuals) approach. This method would be particularly valuable in cases where long-term monitoring and age-determination is impractical, as is the case for highly mobile populations found far offshore. Cooperative care of the young is a central feature and driver of sperm whale sociality (Gero et al. 2013, Cantor et al. 2019). Being able to infer the developmental stage and sex of individuals from UAV-derived footage would allow us to better understand the extent to which care behaviours are driven by calves or juveniles seeking care versus adults providing care, and how these change overtime. It would also reveal the degree to which each sex may contribute to different aspects of calve care.

Likewise, *TL* and *NRflipper* measurements can provide valuable information for interpreting the interactions between adult or mature males and groups of females. It is hypothesized that only mature males (> 13.6 m) participate significantly in reproduction, and that larger males with relatively larger noses have a competitive reproductive advantage (Cranford 1999), however this has not been empirically tested. By analyzing the interactions between adult/mature females with known males of different sizes and nose-to-body ratio, we would be able to explore if *TL,* *NRflipper*, and raw nose measurements correlate with the frequency with which females approach or interact with males and vice versa. While this would not directly measure reproductive success, patterns of female-male interactions could clarify the drivers of female choice (Eguiguren et al. 2023).

Length-based inferences of developmental stage and sex obtained through UAV photogrammetry can also provide a relatively inexpensive and quick method for quantifying the age structure of a population, and inferring its reproductive potential (Waters & Whitehead 1990, Vivier et al. 2025). Usually, estimating the age distribution of a population requires large-scale sampling (e.g., hunting or commercial harvesting), mark-recapture methods and long-term monitoring. But, photogrammetric estimates of size distribution, informed by ground-truthing data, can provide useful estimates (Waters & Whitehead 1990, Vivier et al. 2025). This is a particularly useful means of monitoring the reproductive potential of a population over time, for instance by comparing reproductive parameters obtained near the end of whaling by traditional methods with present day parameters, which can inform our assessments of populations’ vulnerability with changing conditions. Updating reproductive parameters for sperm whales would contribute to existing knowledge gaps in the different populations’ vulnerability in the face of compounding anthropogenic threats (Eguiguren et al. 2025a). Still, some care should be taken to make sure that individuals measured are a representative and unbiased sample of the population.

Our methods produce a quantitative representation of the likelihood that an individual is either male or female, which contributes essential information for interpreting behavioural observations. Because differences in the needs between males and females shape their behaviours and dictate their social relationships, the social interactions of males and females can be quite different, especially in sexually dimorphic species. Thus, behavioural studies of social interactions (e.g., affiliative/aversive behaviours, decision-making, cooperation) have been most informative when individual sex can be distinguished (e.g., Connor et al. 2006, Harvey et al. 2017, Zwamborn et al. 2023). This added layer of knowledge can help us make more useful inferences when investigating social interactions. For example, are there social behaviours that are exclusive or predominantly engaged in by mature females? Are some behaviours more frequent among immature males? These questions help elucidate the nature of relationships in sperm whales and the proximate mechanisms by which their societies are maintained and established ( Eguiguren et al. 2025b).

### 4.5 | Limitations and methodological considerations

Our work is chiefly limited by the absence of known data on the developmental stage and sex of measured individuals. This means that we can’t provide evaluations of classification performance equivalent to those presented by Cheney et al. (2022), Vivier et al. (2024), and Robinson & Visona-Kelly (2025). Future applications of our methods could overcome this limitation by collecting measurements from individuals of known sex and developmental stage in cases where this is feasible. For the present study, we evaluated the ability of our methods to infer individual developmental stages and sexes by comparing our findings to those based on direct measurements of thousands of killed individuals (Nishiwaki et al. 1963, Ohsumi 1977, Best et al. 1984) or mass strandings (Evans & Hindell 2004). While these sources provide a useful baseline, there are some caveats to extrapolating these findings to our sample. Beyond individual variation in growth rates, population-level growth rates can change in response to resource availability and human impacts (Adamczak et al. 2023). For instance, Clarke et al. (2012) found that female sperm whales killed before the whaling moratorium (1959 – 1962) in the South Eastern Pacific sexually matured earlier (6.5 years) and at smaller sizes (8.2 m) than in other regions, presumably as a result of prolonged whaling in the region. Similarly, Waters & Whitehead (1990) found that growth curves and overall lengths of Galápagos sperm whales in 1985 and 1987 were slightly smaller than those generated in previous decades using whaling data. While some of the differences in the latter case may reflect a bias in whaling data towards larger and more lucrative individuals in whaling data, the differences between growth curves remained within the expected variation of age – length relationships (Waters & Whitehead 1990). There is also evidence that size distributions among female sperm whales vary geographically, with whales in lower latitudes being generally smaller than those in higher latitudes (Best et al. 2016). Thus, while our general appraisal of developmental stage and sex is informative, the precise parameters describing the *TL* and *NRflipper* curves may not be directly applicable to whales from other regions. Applying this method to other datasets requires will require estimating optimal parameters for a given population.

We chose an UAV system that is relatively inexpensive (<500 USD vs > 2,000 USD for other frequently used systems) and user-friendly, which may be ideal for projects that are budget and/or experience-limited, allowing them to collect valuable demographic data that would otherwise not be attainable. If higher accuracy and precision are needed, simple improvements can be made by implementing laser-based altimeters. There are several open-sourced resources for installing lidar systems on commercially available UAVs frequently used in cetacean monitoring (Bierlich et al. 2024). This would improve accuracy and precision in *TL* estimates (Bierlich et al. 2021, Napoli et al. 2024b), but would not resolve the uncertainty associated with measuring *NRflipper* (or other body ratios) as they are independent of altitude estimates. Because whales are not rigid, there is some unavoidable uncertainty in this. We suggest, as we have done here, taking several measures for the same individuals to quantify the uncertainty associated with these metrics, which can then be propagated through further analytical steps using frequentist or Bayesian approaches (e.g., Bierlich et al. 2021, Napoli et al. 2024b)