Table of Contents

[3. DISCUSSION 1](#_Toc204255586)

[3.1 | Measurement error and uncertainty 1](#_Toc204255587)

[3.1 How reliable are our raw measurements? 3](#_Toc204255588)

[3.2 Despite uncertainty, measurements are congruent with sperm whale body size: 4](#_Toc204255589)

[3.3 And our approach allowed us to make some well-supported inferences about individual sex/age 5](#_Toc204255590)

[3.2 Despite the uncertainty, how helpful are our inferences? 5](#_Toc204255591)

[3.1 Inferring developmental stages 7](#_Toc204255592)

[3.2 Our methods allowed us to pick apart mature females from juvenile males, but there are some uncertainties 9](#_Toc204255593)

[3.4 Linear NR growth in males? 10](#_Toc204255594)

[3.5 Case study on peduncle diving – matches our expectations and demonstrates use 10](#_Toc204255595)

[3.6 Demographic structure applications 10](#_Toc204255596)

[4. CONCLUSIONS 11](#_Toc204255597)

## 

## 3. DISCUSSION

We developed a minimally invasive method of inferring individual sperm whale developmental stage and sex leveraging prior knowledge on sperm whale morphometric development and sexual dimorphism. AUV-based body length (*TL*) estimates provide useful proxies for developmental stages and can help refine the traditionally used ‘calve/mature female-immature/mature male’ classification system. 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 (Nishiwaki et al. 1963, Cranford 1999). Applying Bayesian theory, we estimated the posterior probabilities of individuals belonging to either sex given their *TL* and *NRflipper*. While some individuals could be classified as females/females with high confidence based on their posterior probability estimates, others lacked the certainty to be assigned as either. Still, our inspection of peduncle dive patterns (PD) illustrates how the numeric representation of morphological ‘femaleness’ and developmental stage 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.

An important limitation of our work is the absence of data on the known developmental stages and sexes of measured individuals. This means that we can’t provide evaluations of performance equivalent to those used for similar research (Cheney et al. 2022, Vivier et al. 2024, Robinson & Visona-Kelly 2025). We provide interpretations on the ability of our method to infer individual developmental stages and sexes by comparing our findings to those primarily based on direct measurements from industrial whaling operations (Nishiwaki et al. 1963, Ohsumi 1977, Best et al. 1984) or mass strandings (Evans & Hindell 2004). While these sources provide thorough analyses based on hundreds – thousands of individuals (Nishiwaki et al. 1963, Ohsumi 1977, Best et al. 1984), there are some caveats to extrapolating their findings to our data. First, growth curves that relate length to age may shift as a response to changes in sources of mortality and survival, including the removal of whaling pressure and local anthropogenic/environmental pressures. Additionally, there may be geographic differences in general size and developmental trajectories. Clarke et al. found that sperm whales mature younger and smaller than others in the SEP, but this may be due to different methods of assessing maturity.

We can overcome this because….

Acoustic measurements have also shown that whales in different regions have different sizes (Caribbean vs. Galapagos), meaning that the precise parameters describing the *TL* and *NRflipper* curves we found may not be directly applicable to whales from other regions. Applying this method to other datasets would involve finding the corresponding optimal parameters for a given location.

### 3.1 | Inferring developmental stages from *TL*

Inferences on the age from *TL* measures have been carried out for sperm whales since the industrial whaling period, during which growth curves were developed to estimate age based on length by relating direct length measurements and age estimates based on dentin layer counts of killed individuals. When measured individuals are a representative sample of the population, *TL* estimates have been used to estimate life history parameters, like survival and reproductive rates, which provide valuable information for evaluating a population’s vulnerability over time (Waters & Whitehead 1990, Evans & Hindell 2004). Although the individuals we measured do not represent a random sample of the population, the size distribution we found resembles that typical of groups of females in tropical/subtropical waters (Waters & Whitehead 1990, Evans & Hindell 2004). Given existing length-age keys, AUV-based length estimates could therefore easily be applied to calculate demographic parameters from wild live populations, without the need for costly and time-consuming longitudinal methods (like mark-recapture). We note, however, that previous attempts to categorize individual age classes based on photogrammetric estimates have resulted in poor accuracy in small cetaceans (Cheney et al. 2022, Vivier et al. 2023). However, broad age class groupings in dolphins have been accurate (Vivier et al. 2023).

Additionally, the *TL-based* developmental stage assignments we present provide a useful level of information for behavioural and demographic analyses. For example, differentiating calves (<xx m), juveniles (xx – xx m), and sub-adults can provide a means to explore the development of calf-mother/allomother behaviours. Studies of individual Indopacific dolphins have shown that maternally initiated care behaviours (contact and following) are highest for younger calves but decrease considerably as individuals age. Observing general behavioural interaction rates at a group level can reveal similar processes. Likewise, the ability to differentiate adult males from mature males can inform variation in behaviour and reactions of female groups. Questions that can be explored include: Are females more likely to approach/interact with mature males than adult males? While there is evidence for individual variation in developmental trajectories, these general classes may still be useful when investigating overall patterns of behaviour and population parameters.

### 3.2 | Sex inferences – what can we define and what is uncertain?

### 3.1 | Measurement error and uncertainty

#### 3.1.1 TL estimates

AUV-derived *TL* estimates were more precise (i.e. had less variability) than those obtained using earlier boat-based photogrammetric methods (CV of present study = 2% vs previous work = 4.4 – 5.1% (Gordon 1990, Dawson et al. 1995). However, our *TL* had greater uncertainty compared to more recent techniques used for measuring sperm whales (1 – 1.3% CV (Jaquet 2006, Dickson et al. 2021) and other cetaceans. The higher uncertainty in our length estimates, particularly when compared to AUV-based photogrammetry (Dickson et al. 2021), results from our reliance on a barometric altimeter to estimate GDS, which is considerably less precise than laser-based altimeters used frequently for cetacean photogrammetry (Dickson et al. 2021, Bierlich et al. 2021). Surprisingly, even though whale body position is constantly changing (Dickson et al. 2021), we found that body length estimates had slightly less uncertainty (CV = 2%) than measurements of the calibration object (CV = 3.15%). This most likely is due to still images of the calibration object being taken over four months, while images from individual whales were almost always collected on the same day (maximum 2 days apart). Because barometric altitude estimates are affected by meteorological conditions, the CV for the calibration object likely captures variability induced by a broader range of conditions than those we encountered while recording each whale.

Despite the uncertainty associated with morphometric measurements, UAV-based morphometric estimates of sperm whale total body length were consistent with previously reported ranges (Waters & Whitehead 1990, Evans & Hindell 2004, Jaquet 2006). While a measurement error of 4.18% may not be suitable for analyses that require high precision, such as those interested in detecting changes in individual morphometry over time, it can be acceptable towards delineating general developmental stages (Vivier et al. 2023).

#### 3.1.2 NR measurements

Direct measurements of stranded individuals indicate that the soft tissue that spans from the base of the skull to the tip of the snout, which encapsules the spermaceti organ, is the region of the sperm whale’s body that is most sexually dimorphic (Cranford 1999). Nishiwaki et al.’s (1963) estimates of nose-to-body ratio also support this, with ratios based on distances from the tip of the nose to the center of the eye or the mouth angle showing more sexual dimorphism than those extending to the tip of the flippers. However, we were unable to consistently measure the distance from the tip of the snout to the base of the skull from AUV images as the base of the skull was only distinguishable in whales with lower body condition (i.e., skinnier). Conversely, *NRflipper* measurements could be obtained whenever a flipper was visible and reliably captured the expected sexual dimorphism of mature males better than the alternative *NRdorsal* measures, despite being associated with some uncertainty (average CV = 2.9%). The uncertainty in *NRflipper*measurements likely reflects the inherent difficulty in measuring the snout-to-flipper distance, and the fact that it relies on two measurements (nose length and *TL*), which makes it more variable (Vivier et al. 2024). While we attempted to standardize this measurement by drawing a cross-sectional reference line from the flipper’s insertion point across the body (**Figure 1**), this still was subject to annotator error, particularly when only one flipper was visible.

#### 3.1.2 Growth curve parameters and posterior probability estimates

Parameter estimates for male and female growth curves were sensitive to measurement error in *TL* and *NRflipper*, which contributed to some uncertainty in *P(f)* estimates, particularly for some whales of intermediate body size (*TL = 8 – 12*). Still, general growth curve shapes were consistent across bootstrapped iterations (**Figure 5),** providing further evidence that the relationship between *NRflipper* and *TL* provides a useful means of capturing the sexual dimorphism in sperm whales.

### 3.2 Despite uncertainty, measurements are congruent with sperm whale body size:

Despite the uncertainty associated with morphometric measurements, UAV-based morphometric estimates of sperm whale total body length were consistent with previously reported ranges (Waters & Whitehead 1990, Evans & Hindell 2004, Jaquet 2006). Although measured individuals do not represent a random sample of the population, the size distribution we found resembles that typical of groups of female groups in tropical waters, composed mostly of adult/mature females, some calves/juveniles and occasional mature males (Waters & Whitehead 1990, Evans & Hindell 2004). Likewise, Although Nishiwaki et al. (1963) did not measure nose-to-body ratio using the same morphological landmarks we did, our *NRflipper* range (0.27 – 0.41) was congruent with their measurements, as it was higher than ratios estimated from the tip of the snout to opening of the jaw (0.11 – 0.15) and eyeball (0.16 – 0.26), and lower than ratios estimated from the tip of the snout the tip of the flipper (0.35 – 0.43).

### 3.3 And our approach allowed us to make some well-supported inferences about individual sex/age

### 3.2 Despite the uncertainty, how helpful are our inferences?

-Still NRflipper models identified individuals that with consistently high likelihood of being female, despite measurement error. These individuals would have traditionally been categorized as fem/juvs.

- Inferrences don’t have real ground truthing, but are based on known biological differences – mechanism and pattern are well documented.

-Applying this to other populations would benefit from known data.

-parameters should be adjusted for different populations (e.g., Atlantic known too have different sizes)

- uncertainty in length – developmental stage assumptions: definitions usually based on general patterns, not good for capturing individual variation (for which sperm whales are famous).

- still informative for general patterns: behavioural differences/ demographic changes over time/reproductive rates/male departure ages

**Sandbox**

### 3.1 Inferring developmental stages

The smallest individual we measured (mean *TL* = 4.1 m, 95% CI = 3.7 – 4.3 m) fell within the length range of sperm whales at birth (3.92 - 4.05 m), suggesting it was a few days to a few weeks old (Best et al. 1984). The *TL* of largest individual that could be assigned a high probability of being female also fell within the maximum range for females ((ca. 12 m; Waters & Whitehead 1990). The largest male we found 16.1 m (95% CI = 15.8 -16.2 m), making him at least 25 years old (Ohsumi 1977).

#### Length estimates are reasonable

Developmental stages are defined in principle using arbitrary size cutoff that indicate general developmental patterns (e.g., how big are most whales when they start consuming food? At what size are most whales achieving sexual maturity). Sperm whales notoriously though can have quite some variability in their development – e.g., weaning can be very prolongued! This means that, in principle the developmental stages are imperfect. Still, in the absence of true data on individual status, they represent valid generalizations that can inform our understanding on their behaviour. Study by dolphins shows that length estimates are good representations of age only at 2 – 3 age class bins. But whether age is a good predictor of developmental in the species is uncertain. We consider that development (more closely associated with size) is likely a more important variable than just age in some contexts (e.g., behaviour, reproductive potential, etc.). Our length estimates can help assign individuals to developmental stages, particularly differentiating within young individuals and within Mature males.

#### Uncertainty is reasonably represented – higher than better drones, sources have been captured

Our bootstrapped estimates of uncertainty associated total length estimates seem to realistically reflect true error. The corrected measurement error for the DJI Mini2 drone based on calibration image analysis were similar to that of higher end DJI drones using the inbuild barometric altimeter, but had more variability overall (Napoli et al. 2024). Likewise, sd of bootstrapped total whale length estimates (average =. 0.17 m, SD = 0.12) are wider than those for the calibration object and for higher end drones (Napoli et al. 2024). As the 95% CI estimates for total whale lengths (mean = 4.18%, SD = 3.34 % of average length) are slightly higher than that of the measurement of the calibration object (1.6% of true length). Uncertainty in estimating our calibration object likely results from remaining barometric error estimates, which likely change with conditions, as well as shifts in the boat’s altitude and angle over the water. Error estimates for our calibration object are similar to those reported for other drone models using a laser altimeter, despite us using the inbuild barometric altitude (Bierlich et al. 2021). The additional uncertainty in total length estimates and NR ratio measurements likely comes changes in the body position of sperm whales and annotator error in locating landmarks. Still, we found that this level of error did not exceed that of researchers using other more sophisticated drone models.

1. Error estimates are also consistent with other works – despite us using a very cheap drone.
   1. Altimeter-error (captured by *Balaena* measurements) is similar to that of other models.
   2. Bootstrapping likely captured realistic measure of uncertainty, as it is slightly broader than *Balaena* (reflecting error from differences in whale position/visibility of morphological landmarks).

### 3.2 Our methods allowed us to pick apart mature females from juvenile males, but there are some uncertainties

1. NRflipper for known mature males is highly divergent from that of the rest of the population – despite using a different landmark than previous work.
2. Several individuals had high uncertainty and intermediate p(fem) values
   1. May have some measurement error (TL and ratio)
   2. May be ‘intermediate-nosed whales’ - natural variation + intersex whales
   3. No individuals within the SA – Fmax range had high confidence of being male
      1. Males at these ages leave – we primarily followed large groups which have primarily females, calves and juveniles
3. Limitations
   1. Parameter estimates for male and female *NRflipper*growth curves were sensitive to measurement error (between images/within individuals), resulting in some uncertainty in p(fem) estimates, especially young ones.
      1. Not many little individuals used! – hard to observe.
      2. Still able to tease apart some age/sex classes.
      3. Are there other alternatives to our optimizing algorithm?
   2. Used a cheap drone with built in barometer – not great
   3. Error estimates seem to resemble those taken with laser (?) a bit odd.
   4. Sperm whale populations have different growth trajectories – Caribbean/Pacific – may not be directly translatable.
   5. Laser altimeter may improve length estimates – narrow ci for lengths, increase certainty of pf
   6. No ground truthing available – could try on population with known sex/age (based on biopsy/genital inspection – not merely on behaviour).

### 3.4 Linear NR growth in males?

1. Linear growth of NR for males was surprising
   * 1. Nose continues to grow disproportionately for as long as body continues to grow? – does pressure for larger noses remain even when whales are huge?
     2. Larger > 16 m males would help clarify this

### 3.5 Case study on peduncle diving – matches our expectations and demonstrates use

1. Peduncle dive patterns not surprising
   1. Mature females receive, calves & juveniles do, big males not involved.
   2. Some uncertain individuals (high mean p being fem, but high uncertainty) – can’t tell what they are at this point. Likely female given what we know about peduncle diving, but could be a mistake, or reflect non-nursing function (some males do baby sit)
   3. Note it doesn’t represent all the peduncle dive patterns – would be cool to explore, very accessible with drone.

### 3.6 Demographic structure applications

1. Opens door to investigating other social behaviours, including tactile interactions, mating, etc. (like the shark bay dolphins)/ demographic/population structure estimates if sampled individuals are representative of the population (which ours are not!)
   1. If used in longitudinal studies, can track changes in demographic composition (important to inform conservation).

## 4. CONCLUSIONS

We provided a simple approach to infer sex/age classes of sperm whales, allowing finer-grained sex-age classifications.

* 1. Using priors on sexual dimorphism can be incorporated for remote measurement to infer demography of species for which this is hard.
  2. Doing this can provide key information for field behaviour and population research while being minimally invasive.

**Sandbox**

Here, we developed a minimally invasive method of inferring individual sperm whale age-class and sex leveraging prior knowledge on sperm whale morphometric development. Using a low-cost, commercially available UAV, we obtained total body length estimates that allowed for more narrow age-class assignments than traditional field work observations.

Nose-to-body ratio measures based on snout to flipper distances (*NRflipper*) reliably captured the sexual dimorphism in sperm whales’ noses, providing a useful means of inferring individual sex. While parameter estimates for male and female *NRflipper*growth curves were sensitive to measurement error (between images/within individuals), optimal models were consistently able to differentiate likely mature females (MF) from males. Still, some individuals between 8.5 – 12 m long were assigned ambiguous probabilities of being female. Additionally, we found that the relationship between *NRflipper*and *TL* for males <17 m is linear, rather than logistic. Our observations of individuals engaging in peduncle diving generally fit our expectations; only calves and juveniles were observed doing peduncle dives, and most individuals receiving peduncle dives fell within the female size range and had a high probability of being female.

The smallest individual we measured (mean *TL* = 4.1 m, 95% CI = 3.7 – 4.3 m) fell within the length range of sperm whales at birth (3.92 - 4.05 m), suggesting it was between a few days and a few weeks old (Best et al. 1984). The *TL* of largest individual that could be assigned a high probability of being female also fell within the maximum range for females ((ca. 12 m; Waters & Whitehead 1990). The largest male we found 16.1 m (95% CI = 15.8 -16.2 m), making him at least 25 years old (Ohsumi 1977).