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. 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 males/females with high confidence based on their posterior probability estimates, 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 UAS system, our approach can add key demographic information into sperm whale behavioural analyses and population models.

Inferring developmental stages based on TL measurements

The uncertainty in *TL* estimates of our UAV system (CV = 2.0%) fell in the lower end of boat-based photogrammetric methods relying on laser photogrammetry used in the past (CV = 1.3 - 5.1%; Gordon 1990, Dawson et al. 1995, Jaquet 2006). However, our UAV system had higher uncertainty compared to state-of-the art approaches for measuring sperm whales with UAV systems equipped with laser altimeters (CV = 1.0%; Dickson et al. 2021). Still, we found resulting morphometric estimates of total body length remained within previously reported size ranges for sperm whales obtained through direct measurements (Best et al. 1984, Evans & Hindell 2004). 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). This is particularly valuable for a population of highly mobile individuals that is impractical (if not impossible) to track over time. Additionally, information of measurement error can also 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, thereby providing details that contribute towards estimating demographic parameters and deepening our understanding of behavioural development. The size range 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 (i.e., calves) 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; etc.) 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 widely to model population parameters. 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 recognize that growth curves are accompanied by a degree of uncertainty that arises from individual variation in size and development with age. For example, observed *TL* measures for sperm whales a standard deviation of 0.9 m at a given age (Waters & Whitehead 1990), meaning that size-based age inferences are associated to inherent uncertainty. Additionally, behavioural patterns, such as the age at which males are expected to leave their natal unit, may be better predicted by developmental stage and size rather than age alone (Best et al. 1984).

Recent work attempting to identify age-classes based on AUV-derived morphometric measures in common bottlenose dolphins *Tursiops truncatus* showed that size-based age-class assignments perform poorly, particularly when age-bin definitions are narrow (Cheney et al. 2022, Vivier et al. 2023). This highlights the presence of individual variability in size at a given age, as well as the inherent difficulty 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 bins) (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 for age classes corresponding to decreasing and stabilizing growth rates (Vivier et al. 2023).

Some of the uncertainty in relating size to developmental stage may be overcome by parameterizing size as a continuous proxy for development. In fact, it may be that size is a better predictor of development than age given the individual variation in maturation rates. Alternatively, if developmental stages are required, we propose that informative distinctions can be made for distinguishing between general developmental stages. For sperm whales, the initial growth period (i.e., exponential growth phase) takes place between 0 – 7 years (4.1 – 7.6 m), which can be divided into two developmental phases; calves and juveniles (Best et al. 1984). At this stage of development, the growth trajectories of males and females are expected to be nearly equivalent (Nishiwaki et al. 1963, Best et al. 1984). 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).

Inferring sex based on NRflipper – TL relationship

We found that, despite optimum parameter estimates being sensitive to measurement uncertainty (**Figure 4)**, the overall shape of *NRflipper – TL* growth curves and the resulting posterior probability estimates of were generally consistent with previous knowledge on sperm whale sexual dimorphism (Nishiwaki et al. 1963, Cranford 1999). Namely, all whales > 13 m—corresponding to the adult/mature male size ranges (Best et al. 1984)—had consistently low probabilities of being female; whereas 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 sexual dimorphism in immature individuals, although present, is harder to detect (Nishiwaki et al. 1963). Moreover, all the individuals that were observed receiving peduncle dives had a higher probability of being female (**Figure 6**), which is consistent with all previous reports in which all individuals receiving peduncle dives being known to be female (Gero & Whitehead 2007, Konrad et al. 2019, Sarano et al. 2023). Thus, our methods allow for some individuals within the traditional female-immature age-sex class to be reliable identified as adult/mature females.

Still, our methods resulted in many 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 to our measurement system, particularly for individuals with high 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 males particularly challenging. Unfortunately, we could not access the raw data used to build the *NR-TL* curves in Nishiwaki et al, so we don’t have a grounded comparison of the naturally expected variation across individuals. Additionally, there are reports across cetacean species cetaceans 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 also 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.

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, it is likely that adult (i.e., sexually mature) males were underrepresented in our sample.

We found consistent support for a linear relationship between *TL* and *NRflipper* for males between 6 – 16.1 m (**Figure 5**). This indicates that the relative size of male sperm whales’ noses continue to increase throughout their lives. The growth of this portion of the body is primarily driven by the growth of soft tissue encasing the sperm whales’ spermaceti organ, which in older males visibly protrudes beyond the lower jaw when seen from the side (Cranford 1999)

Sex inferences

* 1. Still, many individuals had low uncertainty and intermediate p(F) at a size range where dimorphism is supposed to be noticeable.
     1. Partly reflect inter-individual variation in body ratios at given sizes: different levels of femaleness (e.g., lions with manes?)
     2. No individuals with very p(F) at size range 8 – 12: results from males at that stage leaving maternal units, which we were focused on following (big groups).
  2. Found evidence that NRf continues to grow linearly, at least until 16 m.
     1. Males just keep growing all their lives, most of the growth may happen in the nose!
     2. Makes sense because the growth is the soft tissue, that often visibly juts above the lower jaw.
     3. Some pressure to keep expanding this soft tissue (if it is full of fat, it must be very expensive energetically)

Applications & future directions

1. Future directions
   1. PD study demonstrates how inferences can be used while accounting for uncertainty associated from measurement error and lack of ground-truthing data
      1. Behavioural patterns can be explored by looking at relationships between continuous p(F) values without necessarily binarizing these outcomes.
      2. Ideally, ground truth data would allow for optimal thresholds to be determined for classification.
   2. Developmental studies of behaviour based on cross-sectional approach (rather than longitudinal studies) – changes in calve vs. juvenile vs. subadult.
   3. Estimates of population parameters: assess change (post whaling, etc.)
   4. Sex inference: differences in surface behaviour between males and females
      1. Male departure age?
      2. Participation in social behaviours

Regardless of whether size-based inferences of developmental stage are made based on a continuous or categorical variable, refined developmental stages can contribute to our knowledge of behavioural development and a population’s reproductive capacities. For example, investigating the interactions and spatial arrangement between immature individuals and their (allo)mothers can provide insights into the behavioural development of immature individuals and the corresponding changes in maternal care (Mann & Smuts 1998, 1999, Fellner et al. 2013). Until now, these insights have 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). Using AUV-derived *TL* estimates as proxies for developmental stages can yield similar insights based on a cross-sectional (i.e., observations at a given time across several individuals) approaches, which may be applicable in cases where long-term monitoring and age-determination is impractical (e.g., deep divers, nomads, little money). 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). This can contribute towards answering questions to contribute our understanding on sperm whale cooperative care. Likewise, valuable information can be gained by investigating the behavioural patterns and interactions of adult and mature males with increasing length. While it is suspected that only mature males (> 13.6 m) participate significantly in reproduction, it is uncertain what role female choice has in the matter. By analyzing the interactions between adult/mature females with known males of different sizes, we may be able to answer some of these questions. E.g., do female/immature males approach/interact more frequently with larger males? Are males of different sizes more or less likely to seek proximity/contact with groups of females?

1. Limitations
   1. We don’t have ground truth data, and we are generalizing based on whaling data that was often collected in other parts of the world.
      1. There are documented size differences in different regions
      2. Growth rates can change in response to changing environment – e.g., no more whaling/new anthropogenic sources (how much do they really vary)
      3. Should be taken as general appraisals – but represent the best we can do for this population
   2. Method can be fine-tuned in the future: laser-based altimeter and incorporating known data (biopsies, known life histories).
   3. Still, our methods are a valuable tool for doing a transversal assessment of behavioural patterns and demographic parameters.
   4. Using a cheap drone is a very low entry point that can be accessible to more researchers around the world