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 UAS system, our approach can add key demographic information into sperm whale behavioural analyses and population models.

Inferences on developmental stage 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 techniques used in the past employing laser photogrammetry (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 using UAVs 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 ranges from direct measurement methods (Best et al. 1984, Evans & Hindell 2004). The validity of our *TL* estimates is also supported by the estimated size range of individuals involved in peduncle dives (PD), as only individuals whose sizes fell within the calves/juveniles class performed them, while only those within the size range expected for adult and mature females received them (**Figure x).** , While some research objectives may require a higher level of precision (e.g. detecting changes in individual morphometry over time), 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.

Recent work shows that TL-based age estimates can be inaccurate, particularly when age bins are narrow (Cheney et al. 2022, Vivier et al. 2023). But, when bins are wide enough, they perform well and can provide insights into the demographic structure of a population (Vivier et al. 2025).

by Vivier et al. (2024) showed that size-based age-classifications were accurate (80 – 90% true accuracy across classes) when few (2 – 3) class bins were delineated, and used this method to estimate age-class structure and make inferences about the population status.

* 1. Finer than used traditionally, which can be useful to make group-level insights.
  2. Easier to capture compared to other photogrammetric estimates (e.g. nest-based photos), and gathered together with behavioural recordings.
  3. While don’t have ground truth data, some size-based age classifications can be accurate enough.
     1. Especially in smaller individuals
  4. Developmental stages were delimitated based on size distributions at developmental milestones
     1. There is known individual variation in age-size relationships and developmental milestones
     2. May not have perfect accuracy, still represent overall group-level patterns and demographic parameters.

1. Sex inferences
   1. Estimates of P(f) were consistent with previous findings
      1. Big whales (> 13 m) were given a near-zero p of being female consistently – size range falls well outside the female range
      2. Little whales (calves and juveniles) consistently had p ~ 0.5, consistent with finding that even if sexual dimorphism is present even in birth, it is not as pronounced in earlier life stages
      3. The individuals that had high p(f) values had smaller NRs and ranged the typical mature female range. They also represented the majority of individuals seen receiving PDs.
      4. But we are limited in our ability to evaluate the accuracy of our methods in the absence of true data (discussed below).
   2. 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).
   3. 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)
2. 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
3. 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