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 ‘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 (Nishiwaki et al. 1963, Cranford 1999). 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.

4.1 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 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 note that growth curves are accompanied by uncertainty arising 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).

Recent work attempting to identify age-classes based on AUV-derived morphometric measures in common bottlenose dolphins *Tursiops truncatus* have shown that size-based age-class assignments perform poorly 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).

4.2 Inferring sex based on the nose-to-body – 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 individuals being female 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 range (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). Our methods allow for some individuals within the traditional female-immature age-sex class to be reliable identified as adult/mature females.

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 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 constant (i.e., linear) growth rate of *NRflipper* with respect to *TL* for males between 6 – 16.1 m (**Figure 5**), indicating that the relative size of male sperm whales’ noses continues to increase throughout their lives. This linear trend in *NRflipper* growth emerged despite our initial implementation of a logistic model, as evidenced by unrealistically high asymptotic values (maxm> 1; supplementary table). The observed pattern aligns with the nose-to-body ratiorelationship with body length reported by Nishiwaki et al. (1963). At the same time, it contrasts with male *TL-*age growth curves that show decelerating growth in total body length as individuals approach 15 m (~40 years) (Nishiwaki et al. 1963, Ohsumi 1977, Waters & Whitehead 1990). Together, these results suggest that the growth in larger males are 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, instead of indeterminate skeletal growth (Cranford 1999). Sustained growth of secondary sexual traits well beyond sexual maturity has also been observed in other mammal species, including giraffes and elephants (Simmons & Scheepers 1996, Hollister-Smith et al. 2007). In these species, 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 reproduction 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 Case study on Peduncle diving

Our inspection of peduncle dive (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). The remaining individuals observed receiving PDs, for which *P(f)* could not be estimated (n = 5), also fell within this size range (**Figure 6**). These findings align with previous work in which individual sex was known based on molecular methods or genital inspections, showing that only females receive peduncle dives (Gero & Whitehead 2007, Konrad et al. 2019, Sarano et al. 2023). Two of the individuals observed receiving peduncle dives fell in the a ‘adult female’ size range (8.5 – 10 m – **Table 1**) and had a high degree of uncertainty associated to their *P(f)* estimates (**Figure 6**). Research off the Caribbean found that nulliparous females receive peduncle dives, but whether these individuals were younger females or simply had not given birth within the study period is unknown (Konrad et al. 2019). Females at this size range have most likely attained sexual maturity and are capable of conceiving (Rice 1989). However, our methods did have the resolution to confidently discern these individuals as females. 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 also is congruent with previous work (Gero & Whitehead 2007, Konrad et al. 2019, Sarano et al. 2023). We caution that our methods for detecting participation of PD where not exhaustive, as we only inspected a subset of available footage, and thus don’t 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. 2019, Sarano et al. 2023).

4.4 Future directions

Refined definitions of developmental stages can contribute to our understanding of behavioural development. 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. 2025). Using AUV-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 of individuals from AUV-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.

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* and *NRflipper* correlate with the frequency with which females approach or interact with males and vice versa. While this would not directly measure the effect of reproductive success, patterns of female-male interactions could clarify the drivers of female choice (Eguiguren et al. 2023).

Length-based inferences of developmental stage obtained through UAV photogrammetry can also provide a relatively inexpensive and quick method quantifying the age structure of a population, and inferring its reproductive potential (Waters & Whitehead 1990, Vivier et al. 2025). Usually, estimating the age distribution oof a population requires 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, shortly after the end of whaling vs 50 years later, which can inform our assessments of populations’ vulnerability with changing conditions. Updating these numbers for sperm whales would contribute to existing knowledge gaps in the different populations’ vulnerability to a changing ocean. Still, some care should be taken to make sure that individuals measured are a representative and unbiased sample of the population.

Our methods also 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. Thus, behavioural studies of social interactions (e.g., affiliative/aversive behaviours, decision-making, cooperation) are most informative when individual sexes can be discriminated (e.g., Connor et al. 2006, Harvey et al. 2017, Zwamborn et al. 2023). While we have been able to discern mature males using boat-based observations, we have traditionally not been able to differentiate mature females from other immature individuals. 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 individuals? These questions help understand the nature of relationships in sperm whales, which until now we have only been able to glean from proximity-based assessments of association (Christal et al. 1998, Gero et al. 2014).

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 work applications of our methods could overcome this limitation if they collect measurements from individuals of known sex and developmental stage, which can be obtained through genetic sampling, visual inspections, and longitudinal monitoring. 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 their findings to the whales we sampled off the Galápagos Islands in 2023. First, growth rates can change in response to resource availability and can be density dependent (Adamczak et al. 2023). Changes in growth rates have been documented in sperm whales during/post-whaling (Kasuya 1991). 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. 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.

We chose an AUV system that is relatively inexpensive and user-friendly, which provides a useful means for projects that are budget and/or experience-limited to make valuable inferences that would otherwise not be available. Still, if higher accuracy and precision are required, simple improvements can be made by implementing laser-based altimeters. There are several open-sourced resources for installing lidar systems to 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. 2024), but would not resolve the uncertainty associated with measuring *NRflipper* (or other body ratios) as they are independent of altitude estimates. Because whales move, this may just be an unavoidably uncertain metric. We suggest that taking several measures for the same individuals would help capture 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. 2024)

Conclusions