## 3. DISCUSSION

Here, we developed a minimally invasive method of inferring individual sperm whale age-class and sex leveraging prior knowledge on sperm whale morphometric development and sexual dimorphism. Despite the uncertainty associated with our photogrammetric measurements, we were able to refine developmental stage and sex definitions that have been traditionally used for sperm whales. Namely, nose-to-body ratios across body length 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.

### 3.1 How reliable are our raw measurements?

Although our total body length estimates were more precise than those from earlier photogrammetric methods—2% CV vs previous work = 4.4 – 5.1% CV (Gordon 1990, Dawson et al. 1995)—they showed greater average uncertainty compared to more recent techniques used for measuring sperm whales (1 – 1.3% CV (Jaquet 2006, Dickson et al. 2021). Additional uncertainty in our estimates when compared to AUV-based photogrammetry likely results from our use of the inbuilt barometric altimeter, which is considerably less precise and accurate than laser-based altimeters used in recent work (Dickson et al. 2021, Bierlich et al. 2021). Surprisingly, despite the fact that whale body position is constantly – even if slightly – changing (Dickson et al. 2021), we found a higher degree of variability in measurements of the calibration object (CV = 3.15%) than body length estimates (CV = 2%). Likely this results because still images of the calibration object were taken over a period of four months, while images from individual whales were at most separated by one day. Because barometric altitude estimates are affected by meteorological conditions, it is likely that the CV for the calibration object is capturing a variability induced by a broader range of conditions than those existing for individual whales.

The uncertainty in measuring HF ratio was higher than that for total length (CV = 2.93%), even if this value was not affected by altitude estimates. This possibly reflects a difficulty in measuring the snout-to-flipper distance. While we tried to measure this consistently by drawing a reference line, the angle of the reference line may have not fallen in the right place, particularly when only one flipper was visible (makes it hard to draw the real cross-section).

But, despite some uncertainty, 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 were not sampled randomly, the size distribution we found, with the majority of whales measuring between 7.4 – 12.6 m and the rest ranging from 4.1 – 16.1 m (**Figure 3**), resembled that typical of groups of female groups with calves/juveniles and occasional visits from mature males (Evans & Hindell 2004). 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).

### 3.1 How does uncertainty propagate in our model?

We found that

- uncertainty emerging from barometric error as well as whale position (inter-image) – resulted in variable parameter estimates in the NRflipper-TL models. Difference particularly intense in small individuals (could also be because we had so few).

### 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)

- general length measurements are consistent with previous methods. So UAV-based length estimates can also help assign individuals to developmental stage. But:

- uncertainty in measurement (greatly improved using laser altimeter)

- 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

#### 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 assing individuals to developmental stages, particularly differentiating within young individuals and within Mature males. However,

#### 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 wide 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. While a ~0.5 m error may not be suitable for analyses that require high precision, like those interested in detecting changes in individual morphometry over time or estimating age, it is an acceptable step towards an age-class determination (see dolphin paper) and inferring sex of some whales.

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).