Supplementary Material 1

-Results summary for sex inferences based NRdorsal measurements in comparison with NRflipper measurements-

# S1.1 Variability in nose-to-body ratio measurements

Measurements of *NRdorsal* were generally more variable than measurements of *NRflipper* ***Figure S 1****,* with nearly twice as many individuals having wider 95% CI’s for *NRdorsal* measurements (n = 33) than *NRflipper* measurements (n = 18). This likely reflects the lack of a distinct boundary for the base of the dorsal fin in several whales, which can also be hard to distinguish depending on light and water conditions.

A graph with black and white lines

AI-generated content may be incorrect.

**Figure S1 1**. Distribution of 95% Confidence interval widths for bootstrapped NRdorsal and NRflipper measurements.

# S1.2 Parameter optimization

Optimal values for *fr*, *maxf*, *mr* and *maxm* varied more across bootstrap iterations of the models fit with *NRdorsal*than *NRflipper* (**Table S 1**), resulting in generally higher levels of uncertainty associated with models based on NRdorsal than *NRflipper* (***Figure S 1***), although some iterations of *NRdorsal* models resulted in distant outliers of the male-specific parameters (*mr* and *maxm*; **Figure S 2**).

**Table S1 1.** Bootstrapped means and 95th percentile confidence intervals (95% CI) based on 1000 iterations for parameters relating sperm whale length (m) and nose-to-body ratio (NR) metrics based on snout to the caudal base of the dorsal fin (NRdorsal) and on snout to the flipper insertion point (NRflipper). Parameters reflect the growth rate of females and small males (≤ 6 m) (fr), the female asymptote of R (maxf), the growth rate of larger males (> 6 m) (mr), and the male asymptote of R (maxm).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *R* Metric | *fr* [95% CI] | *maxf*[95% CI] | *mr* [95% CI] | *maxm*[95% CI] |
| NRdorsal | 2.8 (0.63 - 14.9) | 0.65 (0.64 - 0.65) | 0.2 (0.01 - 0.62) | 0.89 (0.22 - 4.79) |
| NRflipper | 2.26 (0.5 - 33.64) | 0.3 (0.3 - 0.3) | 0.05 (0.01 - 0.16) | 2.05 (0.45 - 6.89) |

A comparison of a graph

AI-generated content may be incorrect.

**Figure S1 2.** Bootstrapped logistic curves of the total length (m) and the nose-to-body ratio of sperm whales based on measures of the snout to the caudal base of the dorsal fin (a) and snout to the base of the flipper (b). Theoretical male curves are shown in violet and theoretical female curves are shown in green. The average NR values across iterations are shown by light violet dashed and green solid lines for males and females, respectively. The vertical line indicates the point of divergence between males and females (chm = 6 m) based on Nishiwaki et al. (1963).

A diagram of a graph

AI-generated content may be incorrect.

**Figure S1 3.** Distribution of bootstrapped parameter estimates (x axis) for NRdorsal and NRflipper models.

# S1.3 Posterior probabilities of being female

Models based on *NRflipper* metrics resulted in more reasonable individual estimates of the probability of being female than *NRdorsal* based models. For example, individual ID74 (mean *TL =* 10.78 m, 95% *CI =* 10.63 – 11.06 m), which was observed receiving peduncle dives, was classified with high confidence as female by *NRflipper* models (mean *P(f)* = 0.99, 95% *CI =* 0.99 – 1.00), yet received low and uncertain estimatesfrom the *NRdorsal* model(mean = 0.12, 95% *CI* = 0 – 0.44; ***Figure S 3***). Similarly, individual ID04, a large male (mean *TL =* 15.2 m, 95% *CI =* 14.9 – 15.5 m), was confidently assigned a near-zero probability of being female by the *NRflipper* model (mean < 0.001, 95% *CI width =* 0), but received an uncertain and intermediate probability estimate based on the *NRdorsal* models (mean = 0.50, 95% *CI* = <0.001 – 0.97).Additionally, in models fit with *NRdorsal*, only two individuals that could be assumed to be mature males based on their sizes (ID01 & ID81) were consistently assigned low probabilities of being females. No individuals were consistently assigned a high probability of being female based on *NRdorsal* models.

A graph of a diagram

AI-generated content may be incorrect.

**Figure S1 4.** Bootstrapped mean Total Length (m) and nose-to-body ratio (NR) for individual sperm whales based on (a) snout – dorsal fin distance (NRdorsal) and (b) snout – flipper distance (NRflipper). The solid green line and dashed pink lines show the bootstrapped mean modeled NR for females and males, respectively. Point colours show the mean posterior probability of individuals being female (P(f)). **Points with black outlines have 95% CI ranges ≤ 0.05 for bootstrapped estimates of P(f).** Point shape denotes whether individuals were observed involved in peduncle dives (triangles = receiving, squares = doing, circles = none). Individuals that were observed receiving peduncle dives and mature males (> 13.7 m) are labelled for reference. Dashed vertical lines indicate the minimum body lengths associated with sperm whale sex and age classes based on Best 1979, Best et al. 1984, and Mendes et al. 2007 as follows: calf (4 m; NB), juvenile (J; 5.5 m), sub-adult (SA; 7.6 m ), adult female (AF – 8.5 m), adult male and mature female (AM/MF – 10 m), maximum female length (Fmax – 12 m), and mature male (MM – 13.7).

Supplementary Material 2

-*Sample photographs of measured sperm whales-*

A group of whales in the water

AI-generated content may be incorrect.

**Figure S2 1.** Sample still image close-ups of individual sperm whales analyzed. The circle at the top left of each photograph shows the bootstrapped mean posterior probability of that individual being female (P(f)). ID number, bootstrapped mean total length in meters (TL), mean NRflipper measurements, and the bootstrapped 95% CI range of P(f) are shown below each individual photograph. Individual ID054 is observed doing a peduncle dive onto another whale, and individual ID067 is shown receiving a peduncle dive.

Supplementary Material 3

-Error estimation for DJI Mini Drone 2-

# S3.1 Altitude error estimation and correction

We ran three candidate hierarchical generalized linear models to correct drone altitude using the *nlme* R package (Pinheiro et al. 2000, Pinheiro et al. 2023):

1. True altitude ~ barometric altitude (no random effects)
2. True altitude ~ barometric altitude (random intercept = Date)
3. True altitude ~ barometric altitude (random intercept & slope = Date)

Adding date as a random effect was intended to capture the effect that different weather conditions could have on barometric altitude measurements. We found that models including date as a random variable improved overall model performance, as demonstrated by the significantly lower AIC for models 2 and 3 (Table S3 1). Model 3 also resulted in corrected altitude measurements with narrower confidence. Still, hierarchical models decreased measurement error and uncertainty marginally, with model 3 resulting in 95% CI error width being only 20 cm smaller than model 1. While this difference may be important in some contexts (e.g., when attempting to detect growth or changes in body condition for individuals), we considered it negligible to our goal of inferring general developmental stages and sex.

Table S3 Summary statistics (mean, standard deviation (SD), 2.5th percentile, 97.5th percentile, and 95% CI width) for the error (m) estimates of drone altitude and calibration object length based on the corrected drone altitude. dAIC shows the difference in AIC from the model with lowest AIC. Raw, uncorrected altitude values are shown in the first row. The final model (1) used in this study is highlighted in gray.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Altitude error (m)** | | | | | **Measurement error (m)** | | | | |
| **Correction model** | dAIC | Mean | SD | 2.5th p | 97.5th p | 95% CI width | Mean | SD | 2.5th p | 97.5th p | 95% CI width |
| uncorrected | NA | 2.4 | 1.9 | -1.2 | 0.2 | 1.4 | -0.55 | 0.38 | -1.23 | 0.16 | 1.39 |
| 1. No random effects | 65.6 | 0.00 | 1.9 | -3.9 | -2.1 | 1.8 | 0.02 | 0.38 | -0.65 | 0.74 | 1.39 |
| 2. Date as random **intercept** | 50.2 | 0.00 | 1.8 | -3.8 | -1.8 | 1.9 | 0.01 | 0.36 | -0.69 | 0.70 | 1.38 |
| 3. Date as random **intercept** and **slope** | 0 | 0.00 | 1.5 | -2.7 | -1.7 | 1.0 | 0.01 | 0.33 | -0.61 | 0.59 | 1.20 |