**Supporting information (S1): Details for building HHMM**

**Summary**

The following details the characteristics of individual dives of killer whales recorded on drone video, as well as the procedures used to build the hidden hierarchical Markov model (HHMM) used to predict the behaviours of killer whales from time-depth recorders without any information from video (unknown dives).

**Individual dive characteristics from drone video used to build HHMM**

Our use of an HHMM was motivated by first examining individual dive characteristics and drone video labels of each killer whale’s behavioural state. Note that only a subset of dives had drone video matched to the CATS tag data.

We first examined how representative the dives on the drone footage were relative to common dive characteristics of killer whales. Table S1 represents the dive characteristics of the individual dives that had drone video matched to CATS dataloggers.

As expected, logging behaviour had shallower dive depths than foraging or travelling for both male and female killer whales. Foraging had an extremely high standard deviation in dive duration and maximum dive depth relative to the other behavioural states of females. The standard deviation in dive duration and maximum dive depth attained by females was more than twice the mean values for these dive characteristics (Table S1). This indicates that foraging was less stereotypical overall relative to the other behavioural states examined in females — and was more so than for male killer whales. Males had deeper maximum dive depths and longer dive durations while travelling than while foraging. Foraging dives of males had lower standard deviations relative to the same behaviors in females. However, our sample sizes of dives with drone video was smaller for males, and we recorded limited logging or travelling behavioural states relative to resting and foraging.

**Details for building Hidden Markov Models to predict behaviour of tracks**

The HHMM was fit using individual killer whale dives as fine-scale observations. We assumed that there were a total of four (fine-scale) dive types: shallow, medium, deep, and logging. The first three dive types were selected from visual inspection of dive profiles as well as histograms of maximum depth. Logging behaviour was unique in that it was characterized by much longer post-dive intervals than the other three dive types. We summarized each killer whale dive with three dive characteristics (maximum depth in meters, dive duration in seconds, and post-dive surface interval in seconds) from the CATS tags. Each characteristic was assumed to follow a gamma distribution that was a function of the underlying dive type. All three characteristics were assumed to be independent of one another after conditioning on its corresponding underlying dive type and behavioural state.

For the coarse scale, dives were separated into 10-minute “dive tracks” to define the HHMM’s coarse-scale hidden Markov chain (Leos-Barajas et al., 2017). We defined each “dive track” as the shortest sequence of dives (after the previous “dive track”) that was at least 10 minutes in length. This ensured that coarse-scale behavioural state lasted for at least 10 minutes. We selected a total of three behavioural states (resting, travelling, and foraging), which is in line with the behaviours observed via drone footage.

We assumed that the distribution of each dive type (shallow, medium, deep, and logging) was identical across dive track behavioural states (resting, travelling, and foraging). This choice is in line with Leos-Barajas et al. (2017), and we found that the HHMM produced more biologically meaningful results when the dive type-dependent distributions were shared across behavioural states.

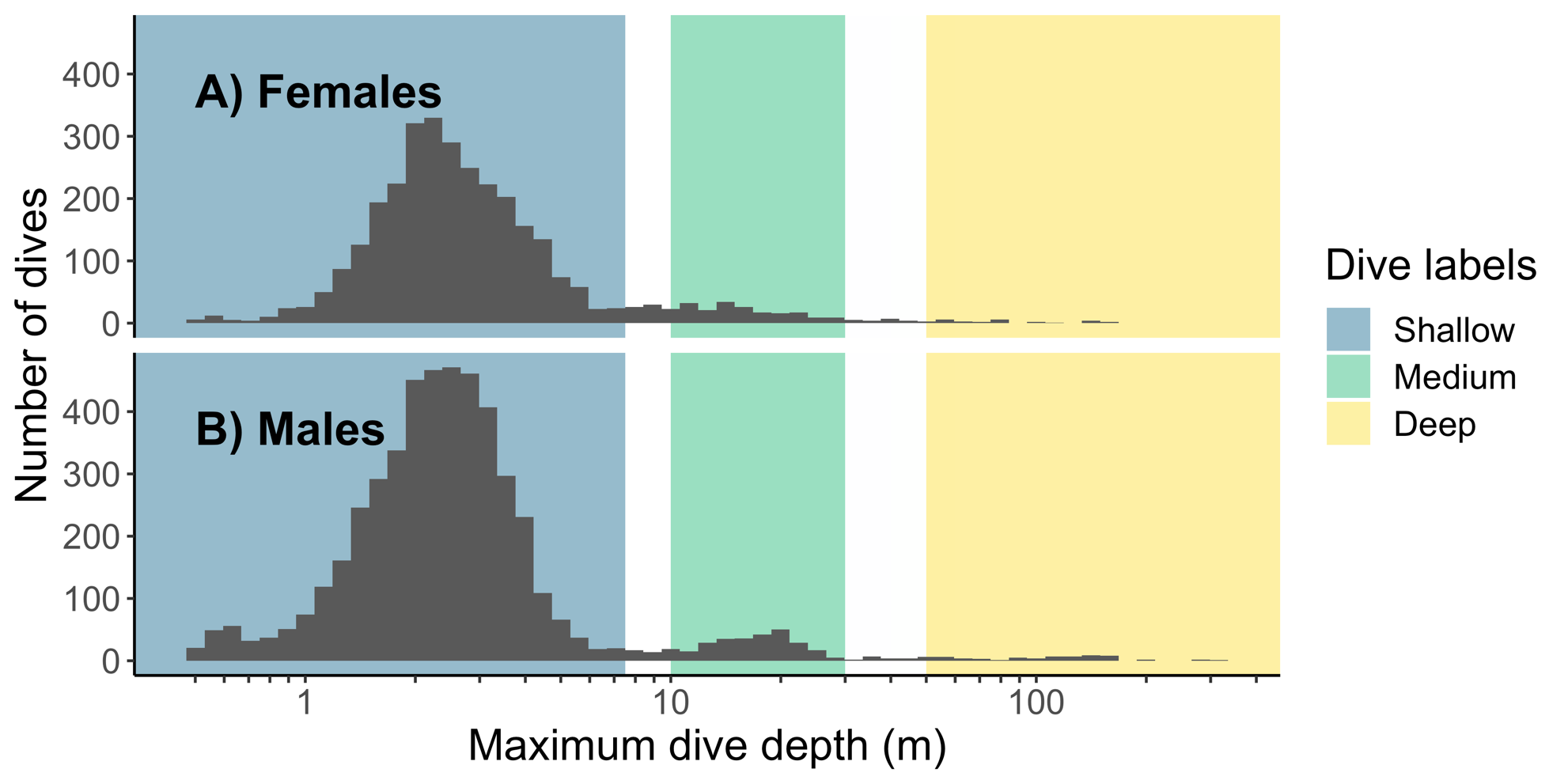
We incorporated labels for the dive types and dive track behavioural states using a method similar to that of Li et al. (2021). For the dive track behavioural states, we used drone videos to directly label killer whale behaviour associated with individual dives, and therefore the dive track corresponding to that dive (i.e., these were observed labels, not unobserved or predicted). , but identifying foraging behaviour took priority over identifying other behaviours. For example, if a track contained both a foraging label and either a travelling or resting label, then the travelling and resting labels were ignored and that track was labelled as foraging. Alternatively, if a track containedboth a travelling label and a resting label, butno foraging labels, then the travelling and resting labels were also ignored, and the HHMM predicted the behaviour without a label. This prevented us from assigning contradictory labels associated with a particular track, but prioritized identification of foraging behaviour. All deep dives were given a “foraging” track label because research has shown that dives > 30m are more often linked with foraging (Wright et al., 2017). Namely, we fixed the probability of starting a “foraging” dive track in a “deep” dive as zero. We also fixed the probability of jumping from any other dive to a “deep” dive as zero during “foraging” dive tracks.

There was a subset of dives that we did not have a drone video label for, but we had reason to believe that they were foraging due to field observations. When an animal did a longer duration dive, the drone often lost sight of where the whale would surface and did not capture the full post-dive surface interval of that dive. We assumed that these specific dives were foraging because video indicated that they were longer in duration, often had fluking present, and the CATS tag data confirmed deeper depths compared to synchronized dives. Within the HHMM, any dive that was deeper than 30 meters and also within 2 minutes of a “foraging” dive (as observed on the drone video) was also labelled as “deep” and “foraging”. This assumed that 2 deep dives close together in time were both foraging if one of them was labelled by drone video as foraging.

Dive type labels were determined using the maximum dive depth and post-dive surface interval duration according to the following criteria:

1. Dives with post-dive surface intervals **≥** 10 seconds were labeled as “logging” dives. This threshold was used because all video-identified logging dives had post-dive intervals of **≥** 10 seconds, and all other video-labelled dives had post-dive surface intervals < 10 seconds.
2. Dives < 7.5 meters were labelled as “shallow”. This threshold was determined via visual inspection of a histogram of maximum depths. If a dive was > 7.5 m and < 10 m, the HHMM model assigned the depth category without explicitly being defined as either shallow or medium.
3. Dives between 10 meters and 30 meters were labelled as “medium”. If a dive was > 30 m and < 50 m, the HHMM model assigned the depth category without explicitly definition as either medium or deep. The 10 m threshold was determined via visual inspection, and the 30m threshold was determined because previous studies have concluded that foraging does not occur at depths shallower than 30 meters (Wright et al., 2017).
4. Dives > 50 meters were labelled as “deep”. This threshold was determined via visual inspection of dive depths. In addition, it is also rare for any behaviour besides foraging to occur deeper than 50 meters (Wright et al., 2017), and deep dives were associated with foraging in this model.

Separate HHMMs were fit for the male and female killer whales. All HHMM analyses were done using the *momentuHMM* package in R (McClintock and Michelot, 2018). We found that the model produced biologically meaningful results, and we validated the accuracy of the model using cross validation. Future studies may incorporate the dive type thresholds above by using truncated gamma distributions to model maximum depth and post-dive surface intervals. However, truncated gamma distributions were not accessible in the *momentuHMM* package (McClintock and Michelot, 2018).

****

**Fig S1. Histogram of maximum dive depth categories.** Behavioural state labels were determined using the maximum dive depth and post-dive surface interval duration according to the following criteria. The x-axis is on a log scale with X m binwidths .

**Table S1. Dive characteristics of 8 northern resident killer whales recorded by CATs tags and categorized by behavioural state observed on drone videos.** Behavioural states observed on drone videos, sample size of individual dives per behavioural sate on drone (N), and summary statistics for dive durations, surface interval durations, and maximum depths from dives that had overlapping drone video and CATS dataloggers. Behavioural states are defined in Table 2. Sample size of whales for all behaviours was n=6 females (389 dives) and n=2 males (87 dives).

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Behavioural state** | **Dives** | **Dive duration (min)** | | | **Surface interval (min)** | | | **Max dive depth (m)** | | |
| **N** | **Mean** | **S.D.** | **Range** | **Mean** | **S.D.** | **Range** | **Mean** | **S.D.** | **Range** |
| **Female** | Logging | 7 | 0.33 | 0.35 | 0.08-1.10 | 0.52 | 0.46 | 0.05-1.41 | 1.78 | 1.03 | 0.89-3.43 |
| Resting | 50 | 0.35 | 0.42 | 0.08-2.38 | 0.06 | 0.01 | 0.03-0.08 | 2.34 | 2.02 | 1.10-12.20 |
| Foraging | 28 | 0.53 | 0.95 | 0.09-4.97 | 0.04 | 0.01 | 0.03 - 0.05 | 8.15 | 14.07 | 1.50-57.93 |
| Travelling | 304 | 0.46 | 0.44 | 0.01 – 4.67 | 0.04 | 0.01 | 0.02 - 0.08 | 2.51 | 2.00 | 0.55-26.13 |
| **Male** | Logging | 1 | 0.18 | NA | NA | 1.73 | NA | NA | 1.13 | NA | NA |
| Resting | 33 | 0.40 | 0.54 | 0.05-2.42 | 0.06 | 0.01 | 0.04-0.08 | 2.73 | 2.58 | 0.79-12.30 |
| Foraging | 51 | 0.32 | 0.14 | 0.02-0.61 | 0.06 | 0.02 | 0.03-0.13 | 2.12 | 0.61 | 0.59-4.17 |
| Travelling | 2 | 0.50 | 0.15 | 0.39-0.60 | 0.05a | NA | NA | 4.19 | 1.68 | 3.00-5.37 |

a For males travelling behavioural state, both individual surface intervals were 0.05 minutes.

References

LEOS-BARAJAS, V., GANGLOFF, E. J., ADAM, T., LANGROCK, R., VAN BEEST, F. M., NABE-NIELSEN, J. & MORALES, J. M. 2017. Multi-scale modeling of animal movement and general behavior data using hidden Markov models with hierarchical structures. *Journal of Agricultural, Biological and Environmental Statistics,* 22**,** 232-248.

LI, J., LEE, J.-Y. & LIAO, L. 2021. A new algorithm to train hidden Markov models for biological sequences with partial labels. *BMC Bioinformatics,* 22**,** 1-21.

MCCLINTOCK, B. T. & MICHELOT, T. 2018. momentuHMM: R package for generalized hidden Markov models of animal movement. *Methods in Ecology and Evolution,* 9**,** 1518-1530.

WRIGHT, B. M., FORD, J. K., ELLIS, G. M., DEECKE, V. B., SHAPIRO, A. D., BATTAILE, B. C. & TRITES, A. W. 2017. Fine-scale foraging movements by fish-eating killer whales (*Orcinus orca*) relate to the vertical distributions and escape responses of salmonid prey (*Oncorhynchus spp.*). *Movement Ecology,* 5**,** 1-18.