Killer whale respiration rates

Tess M. McRae1\*, Beth L. Volpov1, Evan Sidrow2, Sarah M. E. Fortune1,3, Marie Auger-Méthé2,4, Nancy Heckman 2, Andrew W. Trites1

1 Marine Mammal Research Unit, Institute for the Oceans and Fisheries,University of British Columbia, Vancouver, BC, Canada

2 Department of Statistics,University of British Columbia, Vancouver, BC, Canada

3 Current address: Department of Oceanography, Dalhousie University, Halifax, NS, Canada

4 Marine Mammal Research Unit, Institute for the Oceans and Fisheries,University of British Columbia, Vancouver, BC, Canada

\* Corresponding author

E-mail: t.mcrae@oceans.ubc.ca (TMM)

**Abstract**

Knowing the rate at which killer whales breathe while resting, travelling and foraging is a means to determine their energy expenditures and ultimately their food requirements. However, it is uncertain whether killer whales take only single breaths between dives, and how their breathing rates vary among behaviours. We tested the single breath assumption, and calculated activity-specific respiration rates using video from UAV drones and concurrent biologging data from southern and northern resident killer whales equipped with CATS tags. We identified behavioural states from drone footage and matched them to time-depth data from the CATS tags to label different dive types. We then used the labelled dives and a minimum dive threshold of 0.5 m to develop hierarchical hidden Markov models (HHMM) to define tracks and estimate respiration rates for predicted behavioural states of unknown tracks (series of dives without video). We found that killer whales breathe once per surface interval for all behavioural states except when logging (2-3 breaths per surface interval), which did not occur for durations **≥** 10 min. Breaths per minute for females averaged 1.6 while travelling, 1.5 while foraging, and 1.2 while resting. In contrast, respiration rates of males were 1.8 breaths min-1 while travelling, 1.7 while foraging, and 1.3 while resting. Assuming tidal volume of exchanged air is mass-specific, the females in our study consumed 2.5–18.3 L O2 min-1 compared with 14.3–59.8 L O2 min-1 for males across all behaviours. Our findings confirm that killer whales take single breaths between dives and indicate that energy expenditure derived from respirations requires using sex- and behavioural-specific respiration rates. These findings can be applied to bioenergetics models on an activity-specific basis, and contribute towards obtaining better predictions of dive behaviours, energy expenditure and the food requirements of apex predators.

**Introduction**

Respiration rates have been used to calculate field metabolic rates (FMR), oxygen consumption (VO2) and energy expenditure of cetaceans (Williams and Noren, 2009, Roos, 2015). Breathes can be counted each time an animal surfaces (Williams and Noren, 2009, Noren, 2011, Noren and Hauser, 2016) — and can be averaged over a series of dives to calculate respiration rates (Kriete, 1995). Respiration rate can then be used to calculate oxygen consumption (VO2) using measures of tidal volume (VT) and percentage of O2 extracted from inhaled air to calculate field metabolic rates that are ultimately needed to estimate prey requirements.

Killer whales have been assumed to only take a single breath between dives. Previous studies have estimated VO2 of free-ranging killer whales based on qualitative field observations (Miller et al., 2010) or acoustic recordings that captured respirations (Wright et al., 2017). All available information suggest that killer whales breathe only once per surface interval. However, this has not been explicitly quantified or tested. Nor is it known whether the assumed single breath applies to all behavioural states (e.g., travelling, foraging, logging, and resting).

Verifying the number of breaths taken each time a whale surfaces is needed to obtain accurate estimates of FMR and daily energy expenditure of killer whales. There is also a need to determine the relationship between respiration rates and associated behaviours. Previous observations from boat-based focal follows (Noren and Hauser, 2016) or land-based tracking periods (Williams and Noren, 2009) have calculated respiration rates while individual whales were engaged in different behaviours. However, these studies have been limited in their scope and ability to objectively categorize behaviour and quantify respiration rates.

Recently, hidden Markov models (HMMs) have been used as statistical tools to categorize complex movement data of cetaceans recorded by kinematic devices into behavioural states (DeRuiter et al., 2017, Isojunno et al., 2017, , Tennessen et al., 2019, Holt et al., 2021, Miller et al., 2022). Hierarchical HMMs (HHMMs) are specific types of hidden Markov models that can jointly categorize behaviours at multiple time scales (e.g., jointly define individual fine-scale dive types as well as coarse-scale behaviours of tracks comprised of several dives, Leos-Barajas et al., 2017, Sidrow et al., 2022). Although other studies have used HMMs and HHMMs to categorize movement patterns of cetaceans, most studies do not use drone data to verify that learned behaviours correspond to energetically meaningful behaviours such as foraging, travelling, or resting (Tennessen et al., 2019, Isojunno et al., 2017, Leos-Barajas et al., 2017, Holt et al., 2021). For hidden Markov model predictions to be useful for field application and predicting bioenergetics, the behavioural states predicted need to be verified, have a clearly defined function, and be biologically meaningful.

To accurately predict activity-specific energy expenditure from respiration rates, respirations should be counted over durations long enough for an animal to balance their O2 stores (Kooyman and Ponganis, 1998, Butler and Jones, 1997, Ponganis, 2011). In the case of killer whales, respiration rates should be calculated over durations **≥** 10 minutes to avoid biased and inaccurate estimates of respiration rates (Kriete, 1995, Noren and Hauser, 2016, Williams and Noren, 2009).

Many analyses to date that use hidden Markov models to predict behavioural states of cetaceans do so based on individual dives rather than a longer series of dives and surfacings (DeRuiter et al., 2017, Tennessen et al., 2019, Holt et al., 2021, Isojunno et al., 2017). As a result, they cannot be used to calculate estimates of respiration rates for different behavioural states. Activity-specific respiration rates for killer whales need to be calculated for longer periods of diving (≥ 10 min) and for verified behavioural states (with meaningful functions defined) to obtain meaningful energetics data.

We investigated whether killer whales breathe more than once per surface interval by counting total breaths per surface observed on drone videos for different behavioural states. We then matched these drone video labels to time-depth data from CATS tags to inform hierarchal hidden Markov models (HHMM) to predict resting, travelling, and foraging of dive tracks of unknown behaviour. Output from the HHMM was then used to calculate activity-specific respiration rates and oxygen consumption for both male and female resident killer whales. Our findings can be applied to bioenergetics models on a sex- and behavioural-specific basis, and contribute to making better predictions of dive behaviours, energy expenditure and the effect of these large predators on their ecosystem.

**Materials and methods**

## **Data collection**

We collected data from 11 northern and southern resident killer whales (*Orcinus orca,* NRKW) in August 2020 off the coast of British Columbia in the Queen Charlotte Sound (Table 1). Whale ID, sex, birth year, and age as of 2020 were determined from catalogues of known individuals (Towers et al., 2020, Center for Whale Research, 2023). All 11 animals carried video cameras and time-depth dataloggers (CATS tags, Customizable Animal Tracking Solutions, www.cats.is) — and 8 of them were simultaneously followed using an unmanned aerial vehicle (UAV) drone. All killer whale data were collected under University of British Columbia Animal Care Permit no. A19-0053 and Fisheries and Oceans Canada Marine Mammal Scientific License for Whale Research no. XMMS 6 2019.

**Table 1. Summary of data collected on 11 resident killer whales.**

[Insert table 1 and full caption]

### **Time-depth data collection and processing (CATS tags)**

We tagged animals with suction cup CATS biologgers that were equipped with time-depth recorders, forward-facing animal-borne video, passive acoustic recorders (96 kHz), tri-axial accelerometers, magnetometers, gyroscopes, and satellite telemetry for asset retrieval (Cade et al., 2021). To deploy the CATS tags, a small vessel approached the focal group of killer whales and waited for a killer whale to surface near the vessel. With the CATS datalogger attached to the end of an 8m long pole, the CATS tag was suctioned onto the back of an individual whale, near the base of the animal’s dorsal fin. The CATS datalogger continuously recorded dive movements for the duration of time the tag was attached. The CATS tags remained attached from 1.4 to 19.9 hours (Table 1).

Time-depth data from the CATS dataloggers was zero-offset corrected by calibrating depth data in MATLAB (Cade et al., 2021). CATS dataloggers originally sampled at 50 Hz for synchronization with acceleration data on a concurrent project, but we only included time-depth data that was down-sampled to 2 Hz. All subsequent data processing and analysis were carried out with the statistical software program R 4.1.2 (R Core Team, 2022).

Dives and surface intervals were defined with a 0.5 m minimum dive threshold on time-depth data at 2 Hz (Luque, 2007). We selected a shallow minimum dive depth threshold because subsequent analysis focused on capturing all of the respirations at the surface. We tested different minimum dive thresholds ranging from 0.5 to 2.0 m on a subset of animals and concluded that 0.5 m lined up the best with the surface intervals on drone by visual analysis. Deeper dive thresholds excluded some surface intervals that were respirations on drone videos.

### **Video data collection and processing on drones**

We collected video data from a UAV (i.e., drone) from 8 of the 11 whales carrying the CATS tags. Once an individual whale was successfully tagged with a CATS tag, we deployed a drone to follow the tagged whale and take video footage of the whale at the surface. The drone turned off its video when the tagged whale was no longer visible at the surface during longer duration dives (to conserve battery), or when the drone was taking pictures instead of recording video for concurrent studies. We had occasional instances where the drone battery died midway through observing a focal whale. Consequently, the drone video data collected was not a random subsample of all dives recorded by the CATS tags and was biased towards shallower dives. We collected drone video data on some deeper and longer duration dives (up to 58 m), but these dive types did not have the same probability of being captured on the drone video as more shallow dives.

**Behavioural states observed on video**

The drone video footage was analyzed using focal animal sampling of the tagged whales. Each tagged whale had a different coloured CATS tag that was visible from the drone video to confirm whale identification matching the drone to the correct CATS tag. We used the drone video footage to visually count the total breaths per surface interval and behavioural category of each whale (Fig 1, Table 2). We defined each behavioural state based on video observations using behavioural definitions for resident killer whales (Noren and Hauser, 2016, Miller et al., 2010, Ford, 1989). We excluded socializing from the current analysis because it was not mutually exclusive to the other behavioural states, and it rarely occurred among the 8 tagged animals with drone footage.

**Fig 1.** **Still images from UAV drone video showing datalogger placement and respiration.**

[Insert full caption + uploading figure 1 separately]

**Table 2. Definitions of behavioural states of individual dives observed on drone video for northern and southern resident killer whales.**

[Insert table 2 and caption]

If the behaviour observed at the surface on the drone video did not meet the criteria in Table 2, we classified those surface intervals as “unidentified” and removed them from analysis (n=28 total surface intervals removed as unidentified). These “unidentified” behaviours comprised a comparatively small proportion of all behaviours observed on the drone video (5.6% unidentified out of 504 total respirations were unidentified). The unidentified behaviours appeared to be randomly distributed among 7 of the 8 animals with drone video available.

We measured surface interval duration (min) using elapsed time from the known video start time on the drone video. The start of a surface interval (which is also the end of the preceding dive) was defined as the first point at which any part of the whale’s head broke the surface. The end of the surface interval (i.e., the start of the next dive) was the first point at which the whale’s entire body including dorsal fin was fully submerged.

We measured surface interval duration on both the drone video and the CATS tag. Dives obtained from the CATS tags were defined with a 0.5 m minimum dive threshold, but dives recorded by the drone were defined based on submergence of the whale’s body which varied by body size and tag placement. We used the dive characteristics defined on the CATS tags because CATS tag values were more standardized and had a higher sample rate. More notably, we used the surface interval durations calculated from the time-depth-records of the CATS tag because it was most readily available from all biologging devices to predict respiration rate from surface interval durations.

**Matching dives on CATS dataloggers to individual dives on drone video**

Additional data processing was required to align the drone and CATS tag clocks because they were deployed from different field computers. The time-stamps of the CATS tag and drone video were aligned visually by plotting the raw depth vs time (at 2 Hz) with the dive and surface interval start/end times outputted from *diveMove* (Luque, 2007). Next, we overlaid the start and end times of the dives and surface intervals on the drone. We used unique surface intervals as landmarks (e.g., comparatively longer dives and longer surface interval durations) and drone video field notes to manually calculate an “offset” and correct the drone time-stamps to be accurately synced with the CATS tag time-stamps (offsets ranged from 109 to 281 seconds). This process was repeated for all whales and all dives with drone video. Animals that had drone video on the same day had the same offset calculated independently from each other indicating that the offset calculation and re-alignment were accurate.

The dives and surface intervals identified on the drone videos were matched to corresponding dives on the CATS tag in R based on a common time vector of surface interval start/end times synced to both datasets at the level of the individual dive. We used both the start and the end of the surface interval duration to match dives between CATS tags and drone videos. Matching the dives using 2 checkpoints ensured that the matching was accurate. Accuracy of matching was verified by visual plots for each animal for all dives.

## **Statistical Analysis**

**Building Hidden Markov Models to predict behaviour for tracks**

Drone video determined the behavioural state associated with many dives, but we only had drone video observations on ~ 6% of the total dives that killer whales performed (n=476 dives with video; n=8118 dives total on 11 whales, Table 1). We therefore developed hierarchical hidden Markov models (HHMM) to predict the types of individual dives as well as the behavioural categories associated with tracks, or series of dives (Leos-Barajas et al., 2017). The HHMM was also used to define track structure using CATS time-depth data and a synchronized drone videos. The individual dives and tracks informed each other jointly within the HHMM as a single hierarchical process. The HHMM modeled the sequence of track behaviours as an unobserved Markov chain, and it modeled each sequence of individual dives within each track as another unobserved Markov chain whose dynamics depended upon the behavioural state of that specific track. As such, the sequence of dive types within a track was dictated by the behaviour of that track.

A track was defined as the shortest continuous sequence of dives and surface intervals (Baird et al., 2005) that was **≥** 10 minutes in cumulative duration (sum of all the individual dives and surface intervals). We set 10 minutes as the track threshold to allow sufficient time for balancing of O2 stores (Kooyman and Ponganis, 1998, Butler and Jones, 1997, Ponganis, 2011) and to reduce bias in overestimating respiration rates (Kriete, 1995, Noren and Hauser, 2016, Williams and Noren, 2009). This time limit also incapsulate the potential for whales to carry an oxygen debt over multiple dives that is not fully repaid during a single surface interval.

Examining respiration rates at the level of the track rather than individual dives allowed the HHMM to account for potential oxygen debt incurred during diving. If a track reached the 10-minute threshold duration while at depth, the track terminated at the end of the next complete surface interval (i.e., a track could not end in the middle of a dive at depth or in the middle of a surface interval and always ended when a new dive started). If a track reached the 10 minute threshold during a surface interval, the track ended at the end of that surface interval (i.e., the start of the next dive). Although the HHMM predicted behavioural state at both the level of individual dives and tracks, subsequent analysis only focused on the level of the track to focus on meaningful respiration rate durations calculated over **≥** 10-minute tracks.

Detailed description of building the HHMM model and dive characteristics of individual dives from drone video are described in Supporting Information (S1). First, we summarized each dive using its maximum dive depth (m), dive duration (s), and post-dive surface interval (s) as calculated from the CATS tags, and treated these features as an observation of a dive from the HHMM (Table S1).

Second, we generated labels for dive types and track behaviours. Using maximum depth data, we labeled the type of each dive as shallow (< 7.5 m), medium (10-30 m), or deep (>50 m, see S1 for details). If the maximum depth of a dive was between explicitly defined depth categories, that dive was left unlabelled prior to fitting the HHMM. We also labeled the behaviour of several tracks using drone videos directly (e.g., these are observed labels, not unobserved or predicted). Any track with at least one “deep” dive was labelled as a “foraging” track because previous studies have shown that dives > 30m are more often linked with foraging (Wright et al., 2017). Any dive that was deeper than 30 meters and within 2 minutes of a “deep” dive from a “foraging” track (as observed on drone video) was also labelled as “deep” and its track was labelled as “foraging”. We incorporated labels corresponding to track behaviour and dive type into the HHMM in a similar manner to Li et al. (2021).

We fit separate HHMMs for the male and female killer whales on tracks ≥ 10 min. All HHMM analyses were done using the *momentuHHMM* package in R (McClintock and Michelot, 2018). Maximum dive depth, dive duration, and post-dive surface intervals were used as inputs to fit the HHMM. Therefore, testing for a difference in the means of any dive characteristics among behavioural states predicted by the HHMM would result in highly inflated type I error rate (Gao et al., 2022). Consequently, we did not analyze dive characteristics per behavioural state for tracks predicted by the HHMM, and subsequent plots of individual dive durations are exploratory only.

**Assessment of HHMM model error with confusion matrix**

We assessed the error of the HHMM in predicting track behaviour using k-fold cross validation, where one “fold” in the cross-validation scheme corresponded to a single whale’s dive profile with 11 total whales (Hastie et al., 2001). For each whale, a new HHMM was trained using all of the data except for that whale, and the left-out whale’s behaviour was predicted using the Viterbi algorithm (Viterbi, 1967) with the newly-trained HHMM. The true track labels from drone video were then compared with the predicted track labels generated from the Viterbi algorithm. This procedure predicted the model’s ability to accurately label the behaviour of a new whale if it had no video-generated labels. This procedure yielded one confusion matrix (contingency matrix) per behavioural category per sex at the level of the track for foraging, resting, and travelling (there were 2 males and 6 females with video data, and each contingency matrix included the total number of video-labelled bouts). Error was only assessed on tracks that had video labels (females = 66 tracks video-labelled out of 176 total dive bouts predicted; males = 62 tracks video-labelled out of 243 total dive bouts predicted).

Total counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) were used to calculate detection (TP rate), false positive rate (FP rate), precision, specificity, and accuracy across behavioural states for both sexes (Table 3). Detection (TP rate, sensitivity, recall rate) was calculated as the proportion of tracks observed on drone video that were correctly classified for each specific behaviour as true positives by the HHMM model as TP/(TP+FN). This was repeated separately for each behaviour. For example, the detection rate for foraging females is the proportion of foraging tracks on drone video that were classified as foraging tracks by the HHMM model (and not classified as any other behavioural state). The false positive rate was calculated as the proportion of false positives that were correctly identified as FP/(FP+TP). Precision was the proportion of actual TP classified as TP within all potential true or false positives TP/(TP+FP). Specificity (TN rate) was calculated as TN/(TN+FP). Accuracy was defined as the proportion of the total tracks that the HHMM correctly predicted the same behaviour that matched drone video, and was calculated as (TP+TN)/(TP+TN+FP+FN).

**Table 3. Calculation of contingency matrix used to assess HHMM measures of error predicting behavioural state relative to drone video.**

[Insert table 3 with remaining caption]

**Calculating respiration rates from HHMM predicted tracks**

We used track as the unit of analysis to calculate respiration rate over a physiologically meaningful time span (≥ 10 min, Kriete, 1995, Williams and Noren, 2009, Noren and Hauser, 2016) while also incorporating variation in diving behaviour observed in the field. This 10-minute tracking period has been previously used to reduce bias in overestimating respiration rates (Kriete, 1995). Given that logging behaviour was rare (1.7% of the drone video dives) and did not occur in in durations **≥** 10 min, we excluded logging from respiration rate analysis at the level of the track.

Given that whales breathed once per surface interval for all behaviours at the level of the dive bout (excluding logging), summing total post-dive surface intervals per bout was equivalent to counting respirations for resting, foraging and travelling (Williams and Noren, 2009, Noren and Hauser, 2016, Miller et al., 2010). Respiration rate (breaths min -1) was calculated as the total number of respirations (e.g., total number of post-dive surface intervals) per track divided by the cumulative track duration on the CATS tag. Each track had a single predicted behavioural state which included all of the dives and inter-bout surface intervals until the end of the last post-dive surface interval as defined in the HHMM model. The respiration rates were then compared between behavioural categories within each sex.

**Dive durations**

We divided all individual dive durations on CATS tags into short (< 1 min) or long dives (**≥** 1 min) to explore diving patterns based on behaviours predicted by HHMM. This distinction was for exploratory purposes to give context to calculated respiration rates, and was not included in the HHMM or respiration rate analysis. The dive durations were measured directly by CATS tags, and were grouped according to the behaviours predicted by the HHMM (See S1 for details). Note also that respiration rate analysis was done on a different level of the dataset (only dive bouts/tracks > 10 min, and not individual dives).

**Model analysis on respiration rates and error assessment**

All respiration rate analysis was done at the level of the track on predicted track behaviours generated by the HHMM for all of the CATS dives using linear mixed-effects models (nlme package, LME, Pinheiro et al., 2023). LME models correct for repeated measurements among and within animals, and use individual animal variation relative to the mean of the population (Zar, 2010, Pinheiro and Bates, 2000). Whale ID was treated as a random effect.

We used LME models to determine if respiration rate (breath min-1) varied among behavioural categories at the level of the track. The primary dependent variable was behavioural category predicted from the HHMM model (resting, foraging, or travelling). When the dependent variable is categorical, this performs a repeated measures ANOVA with the important addition of accounting for random effects. The significance of the fixed factor was determined using a conditional ANOVA F-test. Model comparisons were performed using a likelihood ratio test (LRT) on two hierarchically nested models. When models were significant, Tukey post hoc tests with Bonferroni adjusted p-values were used to compare the means between multiple levels and identify the behavioural state(s) that differed (*mvtnorm* and *multcomp* R libraries, Genz et al., 2023, Hothorn et al., 2008). Statistical significance was set at α = 0.05. Sexes were modelled separately for all HHMM and LME models due to differences in predicted energetics related to body mass differences in this sexually dimorphic species.

The LME model and calculation of respiration rates per behaviour predicted by the HHMM assumes that the HHMM is infallible in predicting the behaviour per dive bout. Our cross-validation results showed that the HHMM had an accuracy of 85-98%, deeming this a reasonable assumption.

**Example of predicted oxygen consumption rate**

We calculated predicted oxygen consumption rate (VO2, L O2 min -1 and L O2 sec -1) for comparison with published values, and for use in bioenergetic models (Noren, 2011, Roos et al., 2016). We calculated predicted VO2 using 2 methods based on either fixed tidal volume(VO2 fixed, Roos et al., 2016, Noren, 2011) or mass-specific tidal volume (VO2 mass-specific).

We first calculated VO2 fixed using a maximum fixed tidal volume (VT) per sex as assumed in other studies (females = 149.0 L, males = 258.5 L, Williams and Noren, 2009, Noren, 2011, Roos et al., 2016, Kriete, 1995). Rather than apply the same mean oxygen extraction from inhaled air (EO2) to all behaviours, we used the EO2 values that best matched the behavioural state definitions in our study. As such, resting EO2 came from trained killer whales while they rested (activity level 1, females = 35.1%, males = 38.4% in Table 9, Kriete, 1995). In contrast, foraging and travelling EO2 were estimated from trained animals undertaking light to moderate swimming and shallow diving activities (activity level 2, females = 40.0%, males = 43.3% in Kriete, 1995). For males only, EO2 for activity level 2 was averaged from activity levels 1 and 3 values for both whales because it was not directly measured to yield a single mean EO2 for all males in activity level 2 (43.7%, Hyak = 41.15, Finna = 46.3 in Table 6, Kriete, 1995). This resulted in fixed TO2 per respiration per sex per behavior for VO2 fixed.

To calculate VO2 from VT based on mass-specific tidal volume (VO2 mass-specific), we used predicted body mass per animal per sex per behaviour (see Supporting Information S2). In brief, we predicted body length from ages of animals alive in 2020 using a Gompertz growth model per sex (Fearnbach et al., 2011). We then estimated body mass (kg) as a function of body length (cm) for each individual whale (Bigg and Wolman, 1975)—and matched the 11 tagged whales in our study to whales in Kriete (Table 16, 1995) that had similar body masses per sex to yield predicted mass-specific VT per whale (Table 1, 1995).

The calculations for VO2 mass-specific were the same as for VO2 fixed except that we calculated a mass-specific tidal volume for VO2 mass-specific for each animal per activity level 1 or activity level 2 (Table 1, Table 16, Kriete, 1995). A mass-specific VT also yielded mass-specific EO2 and TO2 estimates as well. This table also only includes tracks that were ≥ 10 min cumulative duration. Animals A113, I129, I145, I107, and D21 were excluded from VO2 mass-specific calculations because their predicted body masses were not within 15% of the sex-specific predicted body masses and VT of the killer whales in Kriete (1995). See Table S2 for details on calculations as well as predicted body length (cm), predicted body mass (kg), and mass-specific tidal volume (VT) per whale used to calculate VO2 mass-specific.

**Results**

**Summary of data collected and individual activity budgets**

Drone video data subsampled 6% of the 8,118 total dives with available time-depth data across all 11 killer whales (Table 1). For females, the majority of behavioural states observed on the drone video at the level of individual dives were travelling (78.1%) followed by resting (12.9%), foraging (7.2%), and logging (1.8%, n=389 total dives). In contrast, foraging was observed more often on drone video for males (58.6%) compared to the other behavioural states (rest=37.9%, travel = 2.3%, logging=1.2%, n=87 dives). It’s important to note that activity budgets from drone video are a discontinuous subsample of the total behaviours on CATS tags. Overall logging was rare when both sexes were combined (1.7%, n=8 out of 476 dives) or analyzed separately for females and males (1.8, 1.2% respectively). For females, 389 behavioural labels from drone video were used to inform the HHMM model that yielded 191 total tracks prior to filtering for only tracks that were ≥ 10 min. For the males, 87 behavioural labels from drone video were used to inform the HHMM model that yielded 279 total tracks prior to 10-minute track duration filtering.

**Total respirations**

The killer whales observed on the drone video breathed once per surface interval for all behavioural states except logging (mean logging = 2.1, resting = 1.0, foraging = 1.0, travelling = 1.0 breaths per surface interval). Total respirations per surface interval varied significantly among behaviours (LRT = 1197.5, p < 0.001) due to the higher rate of respirations during logging (mean = 2.1, SD = 0.4, range 2-3 breaths) compared to all other behaviours (1 breath per 1 surface interval, Tukey, p < 0.001). Half of the 8 whales were observed logging (A113 = 3, D21 = 1, D26 = 3, I145 = 1 total surface intervals with logging observed on drone video). Logging was unique in having more than 1 breath per surface interval compared to the other behavioural states, but logging was comparatively much less common than other behaviours on drone video individual dives (1.7% of 476 total respirations on drone video vs 17.4% resting, 16.6% foraging and 64.3% travelling). However, logging did not occur in durations **≥** 10 minutes and was excluded from respiration rate analysis at the level of a track.

**Predicted track behaviours and activity budgets**

For females, the HHMM predicted 76 resting, 75 travelling, and 40 foraging tracks from 3,163 individual recorded dives (n = 7 whales, Table 4). The male HHMM predicted 100 resting, 76 foraging, and 103 travelling tracks from 4,955 individual dives (n = 4 whales, Table 4, see Fig 2 for an example of a dive profile of a male resident killer whale). After excluding tracks that were < 10 min in cumulative duration, females had 70 resting, 38 foraging, and 68 travelling tracks (n = 176 total tracks, 92% of total tracks). Males had 90 resting, 62 foraging, and 91 travelling tracks (n = 243 total tracks, 87% of total tracks) predicted by the HHMM. Logging was excluded from the predicted track analysis due to logging occurring in durations less than 10 minute durations.

**Table 4**: **Total number of tracks per behavioural state predicted by the hierarchical hidden Markov models (HHMM).**

[Insert table 4 and caption]

**Fig 2. Example of a dive-depth profile from whale D21 illustrating depth categories and track behavioural states predicted by the HHMM.**

[Fig 2 gets uploaded separately+ need to insert caption]

Activity budgets at the level of the track showed similar proportions of time spent resting (females = 39.8%, males = 37.0%) and travelling (females = 38.6%, males = 37.5%) with less time spent foraging for both sexes (females = 21.6%, males = 25.5%, Table 4). These are proportions of predicted behaviours and only include tracks that were **≥** 10 minutes cumulative duration. Foraging was relatively less common than travelling or resting in both sexes at the level of the track.

**Error in predicted behaviours**

A k-fold cross-validation was done to test the HHMM error in predicting behavioural states for tracks. The cross-validation compared the “true” behaviour observed on drone video to the “estimated” behaviour predicted by the HHMM to assess error using a confusion matrix (Table 5). For the females, detection (true positive rate) was highest for foraging (95%), followed by travelling (71%), and then resting (56%). For males, detection rate was slightly higher for resting (100%) compared to foraging (96%), but this was influenced by low or absent numbers of false positives and false negatives across all behaviours. There were not any true positive predictions for males travelling, and consequently no detection rate could be calculated for travelling males. These error metrics are at the level of the track based on HHMM predicted tracks. It is important to note that although we had more female killer whales with video than males (6 vs 2 whales, Table 1), we had more individual dives (3,163 vs. 4,955) and more total tracks for males than females (243 vs 176, Table 3). Overall, the HHMMs reliably predicted when a killer whale was forging, resting, or travelling at the level of the track with an accuracy of 85% for females and 96-98% for males.

**Table 5**: **Summary of HHMM error metrics predicting behavioural state relative to drone video.**

[Insert table 5]

**Respiration rates**

We converted the total number of surface intervals per track predicted by the HHMM model into respiration rate (breaths min-1) by dividing the total number of surface intervals per track by the cumulative track duration (sum of all surface intervals and dive durations). Respiration rate analysis with LME models was done at the level of the track only on tracks **≥** 10 minutes. Respiration rates varied significantly among behaviours for females (Fig 3, LRT = 11.28, p = 0.0035) and males (Fig 3, LRT = 64.03, p < 0.001). Mean respiration rate for females was highest during travelling tracks (1.6) compared to foraging (1.5) and resting (1.2 breaths min-1). Males showed similar patterns with the highest mean respiration rate during travelling (1.8) followed by foraging (1.7) and resting (1.3 breaths min-1).

**Fig 3.** **Respiration rates (breaths min-1) of 7 female and 4 male resident killer whales for different behavioural states (resting, foraging, and travelling).**

[Fig 3 gets uploaded separately+ need to insert caption]

Post hoc Tukey tests indicated that travelling tracks differed from resting for females (Tukey, p = 0.003) and that travelling differed from foraging (Tukey, p < 0.001) and resting (Tukey, p < 0.001) for males. All other comparisons in the Tukey tests for both females and males were not significant. Males and females were modelled separately due to body size differences in this sexually dimorphic species. Comparisons between male and female results were not made due to two separate HHMM models being used to predict track behavioural states, one for males and one for females.

**Dive durations**

Individual dive durations for both sexes of whales ranged from a few seconds to as long as 8.5 minutes for the males and 7.7 minutes for the females (all dives on CATS tags) and averaged 0.5 min (32 sec, SD = 0.78, n=4,955 individual dives) for males and 0.6 min (36 sec, SD = 0.84, n=3,163 individual dives) for females on all CATS tags. The majority of individual dives on CATS tags were < 1 min for both sexes (89% of female dives, and 91% of all individual male dives on CATS tags).

**Fig 4. Individual long dive durations (≥ 1 minute) of 7 female and 4 male resident killer whales for different behavioural states (foraging, resting, and travelling).**

[Fig 4 gets uploaded separately+ need to insert caption]

**Fig 5. Distributions of individual dive durations for short dives (< 1 minute) from CATS tags with behavioural state predicted by HHMM (bins = 4 seconds).**

[Fig 5 gets uploaded separately+ need to insert caption]

In terms of longer dives **≥** 1 minute (Fig 4), the males we tagged made longer foraging dives on average (mean = 3.9 vs 2.9 min), and had shorter travelling dives than did the females (mean = 1.5 vs 2.1 min), but had similar resting individual dives (mean = 2.6 vs 2.7 min). Long dives accounted for 11% of the total dives made by females (n=354 long dives) and 9% of the total dives made by males (n=440 long dives). Individual dive durations for males were also more variable compared to those of females for all behaviours (i.e., while foraging, resting or travelling; Fig 4). Within the long dives only, female dive durations appeared to conform more to a central tendency than did those of males for all behavioural states. However, males and females were consistent in terms of making longer foraging dives on average than resting dives, which were in turn longer on average than when travelling.

Shorter duration dives (< 1 min) tended to be normally distributed while travelling, but were skewed following short foraging dives (Fig 5). The distribution of short resting dives made by males and females fell between those incurred while foraging and travelling. Overall, however, the distribution of short duration dives (< 1 min) for all three behavioural states were consistent between males and females (Fig 5), unlike their long duration dives (>1 min) that showed greater variability among behavioural states and sexes (Fig 5).

**Calculated oxygen consumption**

Overall, calculated VO2 (L O2 min-1) based on fixed tidal volume (VO2 fixed) or mass-specific tidal volume (VO2 mass-specific) followed the same overall trends as respiration rate among behaviours for each sex. This reflects the fact that VO2 calculations were based on respiration rate along with values from the literature. However, differences in calculated energy expenditure among behaviours were more pronounced when tidal volume (and consequently TO2) were mass-specific.

**Oxygen consumption calculated with fixed tidal volume**

VO2 fixed refers to VT and subsequent TO2 being fixed per sex (regardless of age or body mass) for all females at 149.0 L and fixed for all males at 258.5 L as in previous studies (Noren, 2011, Roos et al., 2016). Calculated VO2 fixed rates ranged from 5.8-39.7 L O2 min-1 for females across all behavioural states (Table 6) — and were also calculated in L O2 sec-1 for comparison with similar studies on killer whales (Roos et al., 2016, Kriete, 1995). Within females, mean VO2 fixed was highest for travelling (20.5), followed by foraging (18.1), and were lowest for resting (13.6 L O2 min-1, Table 6). VO2 fixed significantly varied among behavioural states (using L O2 min-1, LRT = 36.64, p < 0.001) due to VO2 fixed being lower during the resting behavioural state. Specifically, resting VO2 fixed significantly differed from travelling (Tukey, p < 0.001) or foraging (Tukey, p < 0.001), but travelling did not differ from foraging (Tukey, p = 0.86).

Males showed the same trends as females with travelling whales expending more energy on average relative to the other behaviours (travelling = 42.1, foraging = 40.4, and resting = 27.9 L O2 min-1, Table 6). VO2 fixed for males significantly varied among behavioural states (using L O2 min-1, LRT = 105.04, p < 0.001), and post hoc Tukey tests showed that all 3 behavioural states significantly varied among each other (Tukey, p < 0.001 for all comparisons). The predicted VO2 fixed values for males were higher than females at the same respiration rate and behaviour because the EO2 and VT values were fixed at a higher VT for males (258.5 L) compared to females (149.0 L) in this sexually dimorphic species.

**Table 6. Example of calculated oxygen consumption rates (VO2 fixed) of 11 northern and southern resident killer whales derived from respiration data with fixed tidal volumes per sex.**

[Insert table 6]

**Oxygen consumption calculated with mass-specific tidal volume**

We calculated VO2 mass-specific only for animals that had predicted body masses within 15% of the body masses available in Kriete (1995, see S2 for details). Similar to VO2 fixed, mean calculated VO2 mass-specific for females was highest for travelling (7.4), followed by foraging (7.3), then lowest for resting (6.7 L O2 min-1, Table 7, LRT = 13.26, p = 0.0013). Post hoc Tukey tests showed that only foraging differed from resting (p < 0.001). All other comparisons in the Tukey tests within females were not significant. For males, VO2 mass-specific significantly varied among behavioural states (using L O2 min-1, LRT = 110.49, p < 0.001) and was significantly higher for travelling (43.6) followed by foraging (35.2), then resting (25.8 L O2 min-1, Table 7, Tukey, p < 0.001 for all comparisons). VO2 mass-specific estimates were less than VO2 fixed for the same sex and behavioural state and also showed greater variability within each behaviour because VO2 mass-specific used lower and more variable values of VT (ranging from 42-149 L, S1) instead of a higher fixed VT for all females (149 L) or males (258.5 L).

**Table 7. Example of calculated oxygen consumption rates (VO2 mass-specific) of 6 northern and southern resident killer whales derived from respiration data at the level of the track with mass-specific tidal volumes per sex.**

[Insert table 7]

**Discussion**

Using drones and simultaneously deployed biologging devices to quantify respiration rates and associated behaviours, confirmed the assumptions of others that killer whales take a single breath per surfacing while travelling, foraging and resting. Overall, we validated the assumption that one breath equals on surface interval for all behavioural states except when logging (mean = 2.1 breaths per surface interval), and found that the mean respiration rates per track for females and males while resting (1.2 and 1.3 breaths min-1), foraging (1.5 and 1.7 breaths min-1), and travelling (1.6 and 1.8 breath min-1) were consistent to those reported for killer whales. We also verified that HHMMs are an accurate tool for predicting behavioural states of unknown dive types, but required video verification to yield biologically meaningful categories. We also found that VO2 predicted from respiration rate significantly varied among behaviours for both sexes with significantly less energy expended during resting and females exhibiting lower calculated energy expenditure than larger males.

**Predicted track behaviours**

Hidden Markov models have previously been used to predict and categorize movement patterns of several cetacean species.Hidden Markov models have been used to study the movement patterns of blue whales (DeRuiter et al., 2017), long-finned pilot whales (Isojunno et al., 2017, Miller et al., 2022), harbour porpoises (Leos-Barajas et al., 2017), sperm whales (Miller et al., 2022), humpback whales (Miller et al., 2022), northern bottlenose whales (Miller et al., 2022), and killer whales (Tennessen et al., 2019, Holt et al., 2021, Sidrow et al., 2022, Mul et al., 2020). These studies categorized movement data and investigated a variety of biological applications such as deriving activity budgets (Isojunno et al., 2017), determining differences in diving patterns between sexes (Tennessen et al., 2019), and assessing impacts of anthropogenic disturbances such as sound (Holt et al., 2021, Miller et al., 2022, DeRuiter et al., 2017, Isojunno et al., 2017, Mul et al., 2020). Additional studies have also investigated the theory and statistics behind hidden Markov models using cetaceans (Sidrow et al., 2022, Leos-Barajas et al., 2017). These studies demonstrate that hidden Markov models are a robust method to categorize behaviours of unknown movement data in cetaceans. However, these studies often lack meaningful behavioural categories and verification of behavioural categories in their HMMs.

Despite the emergence of hidden Markov models as a tool to analyze cetacean diving behaviour, there has not been a study to date that validates the behavioural states predicted by the HMM. Our study is the first to predict biologically meaningful behavioural states of resident killer whales using HHMMs verified with drone video observations. The behavioural states predicted by the HHMM in our study were verified with drone video and assessed with a full cross-validation. This is in contrast to previous studies that lacked behaviour verification and were only able to predict ambiguous or suspected behavioural states. For example, Tennessen et al. (2019) categorized dives of resident killer whales into 5 distinct behavioural states based on AIC metrics, but these states were only numbered states not explicitly defined behavioural states.

Tennessen et al. (2019) incorporated acoustic verification of hunting echolocation to help identify potential foraging, but lacked verification of all behaviours. Their study only speculated based on the characteristics of each behavioural state what the biological function of the state might be, but did not explicitly identify a distinct function for the 5 behavioural states. Furthermore, some behavioural states did not have any single suspected function (e.g., state 3 was likely a variety of multiple behaviours). Other studies also allow the HMMs to define movement types based on best fit, but often lead to vague hypothesized dive states that lack biologically relevant behavioural labels (Isojunno et al., 2017, Holt et al., 2021, Leos-Barajas et al., 2017, Sidrow et al., 2022).

Using HHMMs allowed us to increase our statistical power by expanding the number of total dives analyzed while also defining longer duration dive tracks to calculate respiration rates. Using drone videos, we only had 476 discontinuous individual dives with observed behavioural states. However, by using drone video to inform HHMMs, we predicted behavioural states and activity-specific energetics on over 8,000 individual dives with high accuracy (Table 5). Our HHMMs also allowed us to group the 8,000 dives into tracks of at least 10 minutes to calculate respiration rates over physiologically relevant durations. Thus, using HHMMs as a statistical tool coupled with a subset of drone videos allowed us to predict continuous behavioural states of thousands of previously unknown dives on a free-ranging cetacean.

The HHMM had high accuracy, low false positive rates for most behaviours, and reasonably high measures of detection indicating that this method is a reliable tool to predict behavioural states and respirations from dive-depth data(Table 3, 5). Our cross-validation comparing the HHMM predictions to the “truth” on drone video yielded high rates of accuracy (85% for females and 96-98% for males; Table 5).

Our linear mixed effects (LME) analysis of respiration rates per behavioural state assumed that the behavioural states predicted by the HHMM were 100% accurate (i.e., that all of the predicted behavioural states per track were 100% accurate prior to calculating respiration rate per track). This was a reasonable assumption based on measures of accuracy, detection, and false-positive (FP) rate.

Detection (true positive rate) was relatively high (71-100%) for all behaviours except resting females (56%, Table 5). The lower detection rate for resting females reflects the low total numbers of true positive and false negatives for resting females. Based on detection measures of error, the HHMM was more reliable at detecting foraging compared to travelling in female whales. This makes sense because we prioritized building the HHMM model to detect foraging behaviour because it is the only behaviour where energy expenditure can be offset by prey consumption.

Overall the false positive rate was low except for females that were resting (55%) and males that were travelling (100%, Table 5). A false positive occurred when the HHMM identified a track as a specific behavioural state, but the drone video observed any behaviour except that target behaviour. For males travelling, the high false positive rate was skewed by the absence of true positives and false negatives which prevented calculating detection. Notably, all of the measures of error were influenced by the total number of tracks per behaviour and how many of those individual dives were labelled on drone video per behaviour. In the future, the HHMM could be optimized to maximize or minimize whichever measure or error is of most interest.

**Respiration rates**

Average respiration rates are needed to estimate the average costs of foraging, resting, and travelling. The mean respiration rates for females (1.2-1.6 breaths min-1) and males (1.3-1.8 breaths min-1) were similar to those previously reported for killer whales calculated over similar track durations (Williams and Noren, 2009, Noren and Hauser, 2016). Our study predicted the highest respiration rates while travelling (1.6 females; 1.8 breaths min-1 males), followed by foraging (1.5 females; 1.7 breaths min-1 males), and then resting (1.2 females; 1.3 breath min-1 males). Williams and Noren (2009) identified similar respiration rates of 1.6-1.7 breaths min-1 while travelling and foraging from land-based observations. Similarly, Noren and Hauser (2016) found mean respiration rates of 1.6 breaths min-1 for foraging, 1.4 for travelling, and 1.0 for resting when calculated over a boat-focal follow of 10-15 minutes. The resting activity state they defined combined resting (as we defined it) with “completely stationary” behaviours (i.e., logging) which may explain why the mean resting respiration rate in Noren and Hauser (2016) was slightly lower than in our study.

Comparing foraging and travelling behavioural states shows some inconsistencies between studies.For example, killer whales in our study had slightly higher mean respiration rates while travelling (1.6 females; 1.8 breaths min-1 males) compared to foraging (1.5 females; 1.7 breaths min-1 males, Fig 3). In contrast, southern resident killer whales in Noren and Hauser (2016) found the reverse trend with higher respiration rates for foraging compared to travelling (1.6 breaths min-1 while foraging vs. 1.4 while travelling for males and females combined). The apparent differences between our findings might be attributable to differences in the number of males tested in each study and the proportional number of males that contributed data to each of the behavioral states of respiration rates. It is important to distinguish between respiration rates of males and females during foraging or travelling because whales expend energy during all behavioural states, but it is only during successful foraging that whales can consume prey to offset expended energy. Notably, prey sharing, common among killer whales, would represent an exception to this rule whereas whales are able to consume prey without expending energy to catch prey (Ford and Ellis, 2006, Wright et al., 2016, Beerman et al., 2016).

We incorporated aerial observations of behaviour and breathing with detailed TDR data to yield a more complete picture of the whales underwater behaviours. Previous studies calculating respiration rates have primarily used surface-based observations for identifying diving behaviours based on assumptions from surface behaviours (Noren and Hauser, 2016, Williams and Noren, 2009). Assigning diving behaviours based only on surface observations may be limited due to the inability to directly observe underwater behaviour. Both Noren and Hauser (2016) and Williams and Noren (2009) identified behaviour and counted breaths from either boat observations or land-based observations of killer whales at the surface. As such, these studies only observed surface behaviours that were subject to human judgement in assigning behaviours and counting breaths in real time. In our case, we combined less subjective techniques to identify behaviours and respirations—one from above the whale through drone observations (which allowed viewing the same behavioural clip multiple times and at slower speeds) and another underwater view from the back of the whale through CATS tag data. The combination of these two different approaches to quantify respiration rate and behaviours allowed us to build and verify HHMMs to predict behaviours and therefore calculate respiration rates on a large sample of dives without direct human observation. The similarities in the range of respiration rates obtained in this study and in previous studies highlights the viability of using HHMMs to calculate respiration rates and determine behavioural states.

**Dive durations patterns**

The patterns of individual dive durations differed depending on whether the whale was travelling, resting or foraging (Figs 4 & 5), but further research is needed to determine if these trends are statistically significant. However, the differences in individual dive duration patterns among behavioural states make physiological sense, and give confidence that the HHMM did a good job in correctly predicting the three basic behavioural states. For example, the short travelling dives were normally distributed (Fig 4) consistent with animals swimming in a predictable sustainable manner. In contrast, the distribution of short foraging dives was highly skewed (Fig 4) consistent with whales that extend themselves while searching for prey and seek to replenish their oxygen stores to return to depth and forage (Fig. 3).

Interestingly, all of the dives we recorded from the 11 animals were less than the calculated aerobic dive limit for killer whales. The cADL for killer whales based on scaled measurements from bottlenose dolphins (Noren et al., 2002) was 10.2 minutes for females and 11.8 minutes for males (Miller et al., 2010) — while the resting-surface cADL for a trained adult male killer whale based on serial blood lactate measurements was 13.3 minutes (Noren et al., 2012). Thus, it is unlikely that the whales in our study were diving with a pronounced oxygen debt because all of the dives were < 8.5 minutes.

The differences in individual dive durations between males and females while foraging and travelling potentially reflect differences in age class, social relationships, sample sizes per sex, and oxygen storage capacities related to body size. The majority of the females in our study (86%) were skewed towards immature individuals < 13 years old (oldest female was 14), but all of the males were adults > 13 years (15-28 years, Table 1)—with predicted body masses ranging from 1,252-2,555 kg for females and 3,382-4,172 for males (Table S2). The greater variability in durations of the longer dives of females compared to males may also reflect behavioural differences associated with social responsibilities and gender-based relationships of male and female resident killer whales. Some of the differences between males and females, as well as among individuals (particularly during travelling) could also reflect pod differences given that the 11 tagged killer whales were from different pods and were not travelling together (Table 1).

**Calculated oxygen consumption**

The trends in calculated VO2 values mirror those we observed in the respiration rates. Our calculated VO2 fixed values are similar to those previously reported for killer whales. For females, the mean calculated VO2 fixed values across all behaviours (0.1-0.7 L O2 sec-1, Table 6) were comparable to the ranges on trained killer whales across activity states 1, 2 and 3 (approximately 0.05-0.8 L O2 sec-1, extrapolated from Figs 8 and 9, Kriete, 1995). For males, our calculated VO2 fixed values (0.3-1.4 L O2 sec-1, Table 6) were higher than those observed on trained killer whales across 3 activity states (approximately 0.1-1.2 L O2 sec-1, extrapolated from Figs 6 and 7, Kriete, 1995). Other studies that have calculated VO2 from speed and respiration rates have found VO2 values ranging from 0.3-0.5 for females and 0.5-1.0 L O2 sec-1 for male killer whales (values extrapolated from Figs. 3A and 3D for model 1 in Roos et al. (2016) with fixed TO2 values from Kriete (1995)). Of these calculated VO2 fixed values, we suspect those for females are more robust than for males because we had more female and immature whales compared to males (7 vs 4 whales). However, we did have a bigger sample size of total tracks for males (243 vs. 176). There are a lot of assumptions made in calculating both VO2 types. Consequently, statistical significance for VO2 should be interpreted conservatively.

Mass-correcting VO2 did not change any of the major VO2 trends, but did lower VO2 values and made differences among behaviours more prominent for females. We examined the assumption of a fixed VT and TO2 by calculating both VO2 fixed and VO2 mass-specific. VO2 fixed and VO2 mass-specific showed the same general trends with travelling having higher VO2 than foraging then resting for both VO2 types. The differences among behavioural states were more pronounced for VO2 mass-specific relative to VO2 fixed, but both showed that calculated VO2 varied significantly among behavioural states. Calculating VO2 with mass-specific tidal volume lowered the mean VO2 per behavioural state for females only, but not substantially for males (Table 6, 7). This was because our study included several immature animals assumed to be female that likely had lower maximum tidal lung volumes than the fixed tidal volume of 149.0 assumed for all females with VO2 fixed. It is important to note that VO2 fixed had a larger sample size of animals and total tracks than VO2 mass-specific due to the strict criteria in matching predicted body masses within 15% for VO2 mass-specific.

The applicability of using mass-specific tidal volumes to calculate VO2 mass-specific hinges on the availability of activity specific VT and EO2 from similarly sized animals of the same sex. Currently the only published source of this information in killer whales is Kriete (1995) which lacks a wide scope in predicted body masses, especially for male whales. The body masses of females in our study generally matched up more closely to predicted body masses of animals with available mass-specific VT (S2 Table). The males in our study were unfortunately further away from the predicted body masses of the 2 male killer whales with available mass-specific VT. Calculating VO2 from respiration rate involves many assumptions (see below). Using mass-specific VT and EO2 would ideally improve the accuracy of these calculations, but data are currently lacking to do this on a wide range of body sizes per sex (Kriete, 1995). This highlights the need for more current mass-specific estimates of respiratory and physiological variables on animals with a wider range of body sizes and ages.

**Study limitations**

**Sampling limitations**

In terms of population demographics, our sampled animals were skewed towards females and immature animals.The majority of the animals in our study were immature and of unknown sex (5 out of 11 whales were unknown and < 13 years old, Table 1). For our study, we assumed that animals of unknown sex were females, similar to the assumptions of Roos et al. (2016). We believe this assumption is reasonable because immature male body proportions are more similar to females in length and body mass than to those of larger adult males. Both males and female killer whales follow a similar growth curve until they reach 11-15 years of age, at which time males dramatically increase in size (Olesiuk et al., 2005). Any potential differences in sex (which can partially be attributed to differences in mass in a sexually dimorphic species) would likely not become apparent until after the ages of 11-15 years when males grow much larger than females and develop their characteristic large dorsal fin. Consequently, we built two separate HHMMs, one for males and one for females, to account for differences in body size between sexes—and undertook separate analysis of males and females. We had more female animals than males, but more total dives and total tracks for the males than females, credited largely to an overnight deployment on D21 with 1,812 dives and a longer deployment on I107. The LME models on respiration rate accounted for unbalanced samples between sexes and among behavioural states with whale as a random effect.

The primary constraint we faced in sampling deeper dives with drone video was the challenge of the drone pilot to track and view the surfacing of focal animals following longer and deeper dives (because animals making deep dives usually surfaced out of drone view). Consequently, the drone video data collected was potentially biased towards sampling shallower dives.The maximum dive depths for foraging dives with drone video in our study (range = 0.6 – 58 m, Table S1) were substantially less than the maximum dive depths on our CATS tags (range = 0.5-317 m) and those observed in other free-ranging resident killer whales in the Pacific Northwest (170.3 ± 76.5 m, Noren and Hauser, 2016).

**HHMM limitations**

Our HHMMs make several assumptions that may have been violated in this dataset. First, the HHMMs treated a behavioural track as a the shortest sequence of dives that lasted for at least 10 minutes. However, the true underlying behaviour of a killer whale may have changed within a pre-defined “behavioural track”, which may have resulted in errors in our parameter and behavioural estimates. Future studies may allow the coarse-scale behavioural state to change at more flexible intervals than the strict 10-minute divisions we set here.

The HHMMs also assumed that the behaviours of all whales of the same gender were identical to one another. However, it is reasonable to assume that behaviour varies between individuals, even if they are the same gender. Future studies may incorporate random effects into the HHMMs to account for differences in behaviour between individuals.

Our HHMMs also assumed that a given dive type (shallow, medium, or deep) had the same distribution between all behavioural states. However, we show in Figure (5) that the dive duration of short dives appears to vary between behavioural states. This is concerning since most short dives are also shallow dives, indicating that the fine-scale distribution of dive duration changes depending upon the coarse-scale behavioural type. Future studies can explicitly model the difference in the distribution of dive duration between behavioural types.

Our linear mixed effects (LME) analysis of respiration rates per behavioural state assumed that the behavioural states predicted by the HHMMs were 100% accurate. While this was a reasonable assumption based on measures of accuracy, detection, and false-positive (FP) rate, it is not perfect since there is inherent uncertainty when estimating the hidden states of an HHMM. Future research can develop a model that jointly estimates behavioural states of a killer whale as well as its respiration rate during various behavioural states.

**Oxygen consumption assumptions and limitations**

We made several reasonable assumptions to calculate VO2 fixed from respiration rate as has been done in other studies of large cetaceans (Williams and Noren, 2009, Dolphin, 1987, Blix and Folkow, 1995, Christiansen et al., 2014) (e.g., assuming that every breath had a constant VT, EO2 and consequently fixed TO2). However, using fixed TO2 values (with a constant VT) is known to overestimate VO2 compared to when TO2 is modelled to allow variation on a respiration-by-respiration basis (Roos et al., 2016). The lower VO2 mass-specific results for females relative to VO2 fixed for the same behavioural states partially supports this theory. It is also likely that the physiological assumptions required to calculate VO2 from respiration rate vary by the degree to which they are met for each behavioural state (Fahlman et al., 2016). Assuming that VT and EO2 are constant may be realistic for behaviours that lack substantial diving (i.e., logging or resting), but could well be less realistic for foraging, which included deeper longer dives. We attempted to correct for this by using specific EO2 estimates that corresponded to the behavioural state definitions instead of applying the same EO2 and TO2 to all behaviours.

Additionally, the calculations used to predict oxygen consumption from respiration rate involved making several assumptions that reflect the challenges of measuring physiological variables on free-ranging cetaceans. First, we assumed that the animals of unknown sex in our study were females (which are smaller than males in this sexually dimorphic species) to assign sex-specific values of TO2, VT, or EO2. In our case, the unknown whales were all immature (6-10 years) and showed no signs of the rapid growth phase that males experience as they become sexually mature (13-19 years, Noren, 2011). It was therefore reasonable to assume they were females for the purpose of estimating oxygen consumption given that immature males have similar body shapes as females based on age-growth curves for killer whales (Noren, 2011, Fearnbach et al., 2011).

Estimates of VT and EO2 are sex-specific in killer whales, which also necessitated assigning each unknown animal to a sex (Kriete, 1995). Using adult parameters for EO2 and VT on immature animals of either sex may have also inflated the calculated VO2 fixed values for those animals relative to VO2 mass-specific.

**Application of findings**

For the purposes of estimating respirations from only surface intervals recorded by TDR data, the assumption that one breath is equal to one surface interval is valid for resting, foraging, and travelling behaviours. The only behavioural category that did not have one breath per surface interval was logging, which was rare (1.7%). If focusing on logging, surface intervals can be assigned a mean total respiration of 2.1 breaths per single surface interval. Otherwise all other surface intervals for foraging, travelling, or resting can be assigned 1.0 breath per surface interval to predict respiration rates from TDR data, assuming that the tracks meet the minimum duration of 10 minutes.

The criteria used to define dive phases has a significant influence on estimated surface interval times as well as on the total number of dives, dive duration, and dive depth (Murphy, 2021). It therefore needs to be given careful consideration when interpreting respiration rate and behavioural data. The minimum depth threshold used to define the start and end of a dive is the second step in defining dive phases (after zero-offset correction), but selected values vary widely among studies even for the same species. Studies that focus on surface behaviours and respirations often use a shallower minimum dive threshold than those that focus on dive behaviour at depth.

Our analysis of the drone footage and TDR data indicate that the minimum dive depth threshold for killer whales should be 0.5 m. This is shallower than the deeper minimum dive depth thresholds used in others studies of killer whales of 1.0 m (Wright et al., 2017), 1.5 m (Murphy, 2021), and ~1-2 m (dive threshold visually inferred from Fig 4 “short dives” (Miller et al., 2010, Roos et al., 2016). We chose 0.5 m as the minimum dive threshold because it allowed us to accurately match all of the surfaces on drone video to the CATS tag (tested depths: 0.5-2.0 m). Deeper thresholds excluded some of the surface intervals which would have underestimated respiration rates in our study. We suggest using a shallower minimum dive threshold when estimating respiration rates from TDR data because it captures all respirations for resting, travelling, logging, and foraging.

The HHMM trained in our study can be used to accurately predict behavioural states and total respirations per behaviour of unknown dives (i.e., dives without drone video) from historical or future TDR data. We tested the HHMMs ability to predict behavioural states on unknown dives (including animals 3 that had no drone video available), and the HHMM still had reasonable error metrics indicating the robustness of this statistical tool. This provides a means to re-analyze decades of historical TDR data and future data to identify previously unrecognized behavioural states needed to calculate activity budgets. The HHMM was trained and validated with video, and can now be applied to time-depth datasets to identify resting, travel, foraging, or logging behaviours. Concurrent use of drones, biologging devices, and Hidden Markov models are a useful means to accurately quantify respiration rates, behavioural states, and energetics for killer whales—thus opening the door to determine the historical activity budgets of killer whales from decades of archived time-depth data.

**Conclusions**

The primary goal of our study was to validate whether killer whales take single breaths between dives, as well as to calculate behavioural-specific respiration rates needed to estimate energy requirements of killer whales as a function of time spent resting, travelling and foraging. Combining video of verified behaviours from the air and under water with the predicted behavioural states derived from hierarchical hidden Markov confirmed that killer whales breathe once per surface interval, except when logging; and that killer whales breathe at faster rates while travelling than when foraging and resting.

While our findings provide a means to derive respiration-based estimates of energy expenditure from biologging data, the HHMM methods we developed and validated have broader implications. Most notably, they revealed that the minimum dive threshold when processing dive data (used to define the start and end of a dive, and define dive phases from time-depth-data) should be 0.5 m, which is much shallower than the 1-2 m threshold depths used by others. Deeper dive depth thresholds underestimate total surface intervals, total breaths, VO2, and ultimately underestimates the predicted energy requirements of killer whales. A second important methodological outcome from our study was the development of a refined and validated hierarchical hidden Markov model that has a high degree of accuracy in determining what killer whales do based solely on dive-depth dive data (i.e., resting, travelling, and foraging).

Having concurrent drone video and underwater video to assess error and validate model predictions was key to being able to quantify biological meaningful behaviours from killer whale dive data. It has yielded a powerful statistical tool that can accurately determine activity budgets from dive-depth data — and provides a means to assess how changes in conditions have affected energetic costs and the ways in which killer whales spend their time.

**Acknowledgements**

We thank Mike deRoos and Chris Hall for assistance in the field with tag deployments, Taryn Scarff for assistance in drone deployments, and Keith Holmes for piloting the drone and providing videos. Drone footage was collected in partnership with Hakai Institute. We are also grateful to the M/V Gikumi captain and crew for all their logistical support and to Carling Gerlinsky, Dave Rosen, and Kate Colson for providing feedback on our manuscript. We would also like to thank Dawn Noren and an anonymous reviewer for their feedback on an earlier draft.

**References**

BAIRD, R. W., HANSON, M. B. & DILL, L. M. 2005. Factors influencing the diving behaviour of fish-eating killer whales: sex differences and diel and interannual variation in diving rates. *Canadian Journal of Zoology,* 83**,** 257-267.

BEERMAN, A., ASHE, E., PREEDY, K. & WILLIAMS, R. 2016. Sexual segregation when foraging in an extremely social killer whale population. *Behavioral Ecology and Sociobiology,* 70**,** 189-198.

BIGG, M. A. & WOLMAN, A. A. 1975. Live-capture killer whale (*Orcinus orca*) fishery, British Columbia and Washington, 1962–73. *Journal of the Fisheries Board of Canada,* 32**,** 1213-1221.

BLIX, A. S. & FOLKOW, L. 1995. Daily energy expenditure in free living minke whales. *Acta Physiologica Scandinavica,* 153**,** 61-66.

BUTLER, P. J. & JONES, D. R. 1997. Physiology of diving of birds and mammals. *Physiological Reviews,* 77**,** 837-899.

CADE, D. E., GOUGH, W. T., CZAPANSKIY, M. F., FAHLBUSCH, J. A., KAHANE-RAPPORT, S. R., LINSKY, J. M., NICHOLS, R. C., OESTREICH, W. K., WISNIEWSKA, D. M. & FRIEDLAENDER, A. S. 2021. Tools for integrating inertial sensor data with video bio-loggers, including estimation of animal orientation, motion, and position. *Animal Biotelemetry,* 9**,** 1-21.

CENTER FOR WHALE RESEARCH. 2023. *Orca Identification,* [Online]. USA. Available: <https://www.whaleresearch.com/orca-id> [Accessed 2023].

CHRISTIANSEN, F., RASMUSSEN, M. H. & LUSSEAU, D. 2014. Inferring energy expenditure from respiration rates in minke whales to measure the effects of whale watching boat interactions. *Journal of Experimental Marine Biology and Ecology,* 459**,** 96-104.

DERUITER, S. L., LANGROCK, R., SKIRBUTAS, T., GOLDBOGEN, J. A., CALAMBOKIDIS, J., FRIEDLAENDER, A. S. & SOUTHALL, B. L. 2017. A multivariate mixed hidden Markov model for blue whale behaviour and responses to sound exposure. *The Annals of Applied Statistics,* 11**,** 362-392.

DOLPHIN, W. F. 1987. Dive behavior and estimated energy expenditure of foraging humpback whales in southeast Alaska. *Canadian Journal of Zoology,* 65**,** 354-362.

FAHLMAN, A., VAN DER HOOP, J., MOORE, M., LEVINE, G., ROCHO-LEVINE, J. & BRODSKY, M. 2016. Estimating energetics in cetaceans from respiratory frequency: why we need to understand physiology. *Biology Open,* 5**,** 436-442.

FEARNBACH, H., DURBAN, J. W., ELLIFRIT, D. K. & BALCOMB III, K. C. 2011. Size and long-term growth trends of endangered fish-eating killer whales. *Endangered Species Research,* 13**,** 173-180.

FORD, J. K. 1989. Acoustic behaviour of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia. *Canadian Journal of Zoology,* 67**,** 727-745.

FORD, J. K. & ELLIS, G. M. 2006. Selective foraging by fish-eating killer whales *Orcinus orca* in British Columbia. *Marine Ecology Progress Series,* 316**,** 185-199.

GAO, L. L., BIEN, J. & WITTEN, D. 2022. Selective inference for hierarchical clustering. *Journal of the American Statistical Association***,** 1-11.

GENZ, A., BRETZ, F., MIWA, T., MI, X., LEISCH, F., SCHEIPL, F., BORNKAMP, B., MAECHLER, M. & HOTHORN, T. 2023. Multivariate normal and t distributions. *R package.* 1.2-1 ed.

HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J., HASTIE, T., TIBSHIRANI, R. & FRIEDMAN, J. 2001. Model assessment and selection. *The elements of statistical learning.* New York, NY: Springer.

HOLT, M. M., TENNESSEN, J. B., WARD, E. J., HANSON, M. B., EMMONS, C. K., GILES, D. A. & HOGAN, J. T. 2021. Effects of vessel distance and sex on the behavior of endangered killer whales. *Frontiers in Marine Science,* 7**,** 1211.

HOTHORN, T., BRETZ, F. & WESTFALL, P. 2008. Simultaneous inference in general parametric models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences,* 50**,** 346-363.

ISOJUNNO, S., SADYKOVA, D., DERUITER, S., CURE, C., VISSER, F., THOMAS, L., MILLER, P. J. O. M. & HARRIS, C. M. 2017. Individual, ecological, and anthropogenic influences on activity budgets of long‐finned pilot whales. *Ecosphere,* 8**,** e02044.

KOOYMAN, G. L. & PONGANIS, P. 1998. The physiological basis of diving to depth: birds and mammals. *Annual Review of Physiology,* 60**,** 19-32.

KRIETE, B. 1995. *Bioenergetics in the killer whale, Orcinus orca.* Doctoral Dissertation, University of British Columbia.

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.

LUQUE, S. P. 2007. Diving behaviour analysis in R. *R News,* 7**,** 8-14.

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.

MILLER, P. J. O., ISOJUNNO, S., SIEGAL, E., LAM, F.-P. A., KVADSHEIM, P. H. & CURÉ, C. 2022. Behavioral responses to predatory sounds predict sensitivity of cetaceans to anthropogenic noise within a soundscape of fear. *Proceedings of the National Academy of Sciences,* 119**,** e2114932119.

MILLER, P. J. O., SHAPIRO, A. D. & DEECKE, V. B. 2010. The diving behaviour of mammal-eating killer whales (*Orcinus orca*): variations with ecological not physiological factors. *Canadian Journal of Zoology,* 88**,** 1103-1112.

MUL, E., BLANCHET, M.-A., MCCLINTOCK, B. T., GRECIAN, W. J., BIUW, M. & RIKARDSEN, A. 2020. Killer whales are attracted to herring fishing vessels. *Marine Ecology Progress Series,* 652**,** 1-13.

MURPHY, I. 2021. *Improving dive phase definitions in northern resident killer whales.* MSc Dissertation, University of British Columbia.

NOREN, D. P. 2011. Estimated field metabolic rates and prey requirements of resident killer whales. *Marine Mammal Science,* 27**,** 60-77.

NOREN, D. P. & HAUSER, D. D. 2016. Surface-based observations can be used to assess behavior and fine-scale habitat use by an endangered killer whale (*Orcinus orca*) population. *Aquatic Mammals,* 42.

NOREN, S., WILLIAMS, T., RAMIREZ, K., BOEHM, J., GLENN, M. & CORNELL, L. 2012. Changes in partial pressures of respiratory gases during submerged voluntary breath hold across odontocetes: is body mass important? *Journal of Comparative Physiology B,* 182**,** 299-309.

NOREN, S. R., LACAVE, G., WELLS, R. S. & WILLIAMS, T. M. 2002. The development of blood oxygen stores in bottlenose dolphins (*Tursiops truncatus*): implications for diving capacity. *Journal of Zoology,* 258**,** 105-113.

OLESIUK, P. F., ELLIS, G. M. & FORD, J. K. 2005. Life history and population dynamics of northern resident killer whales (*Orcinus orca*) in British Columbia. *In:* FISHERIES AND OCEANS CANADA (ed.). Nanaimo, British Columiba.

PINHEIRO, J., BATES, D., DEBROY, S., SARKAR, D. & TEAM, R. C. 2023. Linear and nonlinear mixed effects models. *R package.* 3.1-96 ed.

PINHEIRO, J. C. & BATES, D. M. 2000. Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus.* New York, NY: Springer.

PONGANIS, P. J. 2011. Diving mammals. *Comprehensive Physiology,* 1**,** 447-465.

R CORE TEAM 2022. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

ROOS, M. M. 2015. *Respiration timing and underwater activity in killer whales (Orcinus orca).* Doctoral Dissertation, University of St Andrews.

ROOS, M. M., WU, G.-M. & MILLER, P. J. 2016. The significance of respiration timing in the energetics estimates of free-ranging killer whales (*Orcinus orca*). *Journal of Experimental Biology,* 219**,** 2066-2077.

SIDROW, E., HECKMAN, N., FORTUNE, S. M., TRITES, A. W., MURPHY, I. & AUGER‐MÉTHÉ, M. 2022. Modelling multi‐scale, state‐switching functional data with hidden Markov models. *Canadian Journal of Statistics,* 50**,** 327-356.

TENNESSEN, J. B., HOLT, M. M., WARD, E. J., HANSON, M. B., EMMONS, C. K., GILES, D. A. & HOGAN, J. T. 2019. Hidden Markov models reveal temporal patterns and sex differences in killer whale behavior. *Scientific Reports,* 9**,** 14951.

TOWERS, J. R., PIKINGTON, J. F., GISBORNE, B., WRIGHT, B. M., ELLIS, G., FORD, J. & DONIOL-VALCROZE, T. 2020. Photo-identification catalogue and status of the northern resident killer whale population in 2019. *In:* FISHERIES AND OCEANS CANADA (ed.) *Canadian Technical Report of Fisheries and Aquatic Sciences.* Nanaimo, BC.

VITERBI, A. 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory,* 13**,** 260-269.

WILLIAMS, R. & NOREN, D. P. 2009. Swimming speed, respiration rate, and estimated cost of transport in adult killer whales. *Marine Mammal Science,* 25**,** 327-350.

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.

WRIGHT, B. M., STREDULINSKY, E. H., ELLIS, G. M. & FORD, J. K. 2016. Kin-directed food sharing promotes lifetime natal philopatry of both sexes in a population of fish-eating killer whales, *Orcinus orca*. *Animal Behaviour,* 115**,** 81-95.

ZAR, J. H. 2010. *Biostatistical analysis,* Upper Saddle River, New Jersey, Pearson Prentice Hall.

**Supporting Information**

See attached documents for details of HHMM building (S1), mass-specific oxygen consumption (VO2 mass-specific) calculations (S2), and data used for analysis (S3).