Cuvier's beaked whale foraging dives identified via machine learning using depth and triaxial acceleration

David A. Sweeney[1], Gregory S. Schorr[1,*], Erin A. Falcone[1], Brenda K. Rone[1], Russel D. Andrews[1], Shannon N. Coates[1], Stephanie L. Watwood[2], Stacy L. DeRuiter[3], Mark P. Johnson[4], David J. Moretti[2]

[1]Marine Ecology and Telemetry Research, Seabeck, WA 98380, USA
[2]Naval Undersea Warfare Center, Newport, RI 02841, USA
[3]Department of Mathematics and Statistics, Calvin University, Grand Rapids, MI 49546, USA
[4]Aarhus Institute of Advanced Studies, Aarhus University, 8000 Aarhus C, Denmark

ABSTRACT: Knowledge of Cuvier's beaked whale *Ziphius cavirostris* behavior has expanded through the utilization of animal-borne tags. However, many tag types do not record sound — thus preventing echolocation click detections to identify foraging — or have short deployments that sample a limited range of behaviors. As the quantity of such non-acoustic tag data grows, so too does the need for robust methods of detecting foraging from non-acoustic data. We used 692 dives from 5 sound-recording tags on Cuvier's beaked whales in southern California, USA, to develop extreme gradient boosting tree models to detect foraging based on 1 Hz depth and 16 Hz triaxial acceleration data. We performed repeated 10-fold cross validation using classification accuracy to tune 500 models with randomly partitioned training and testing datasets. An average of 99.9 and 99.2 % of training and testing dataset dives, respectively, were correctly classified across the 500 models. Dives without associated sound recordings (n = 2069 from 7 whales including 4 non-acoustic tags) were classified via a model that maximized training information using dive depth and duration, ascent and descent rates, bottom-phase average vertical speed, and roll circular variance during dive descents and bottom phases. Of all long, deep dives (conventionally assumed to include foraging), 2.4 % were classified as non-foraging dives, while 0.3 % of short, shallow dives were classified as foraging dives. Results confirm that conventional depth and/or duration classifiers provide reasonable estimates of longer-term foraging patterns. However, additional variables previously listed enhance foraging detections for unusual dives (notably non-foraging deep dives) for southern California Cuvier's beaked whales.

KEY WORDS: Cuvier's beaked whale · *Ziphius cavirostris* · Foraging · Diving · Machine learning · Tagging

1. INTRODUCTION

Most of what is known about the behavior of Cuvier's beaked whales *Ziphius cavirostris* has come from the use of animal-borne tags, technology that is continually advancing (Evans et al. 2013). The first tags deployed on Cuvier's beaked whales were attached via suction cups, and while these deployments were typically short, the data suggested that Cuvier's beaked whales exhibit a strongly bimodal diving pattern consisting of long, deep dives typically followed by a series of shorter, shallower dives (Baird et al. 2006, Tyack et al. 2006). Some of these short-term tags (notably DTAGs) included acoustic

sensors and accelerometers (Johnson & Tyack 2003). DTAG recordings showed that echolocation clicks and accelerations associated with prey captures only occurred during long, deep dives, leading to the interpretation that the primary purpose of such dives is foraging (Johnson et al. 2004, Tyack et al. 2006).

In addition to their extreme diving behavior (Tyack et al. 2006, Schorr et al. 2014, Quick et al. 2020), Cuvier's beaked whales are best known for their presence in cetacean mass stranding events associated with exposure to military sonar (D'Amico et al. 2009, Filadelfo et al. 2009). While the mechanisms connecting sound exposure to mortality remain unclear, the apparent reliance of Cuvier's beaked whales on a physiologically demanding foraging behavior has been highlighted as a potential risk factor (Hooker et al. 2009, Fahlman et al. 2014). However, the limited recording duration of suction-cup-attached tags, which seldom remain attached for more than a day (Tyack et al. 2006, DeRuiter et al. 2013), may result in an incomplete picture of behaviors under both natural conditions and when exposed to anthropogenic sound sources. Therefore, it is important to acquire data over longer intervals to sample the full variability of natural behaviors, capture responses to anthropogenic activities, and ultimately aid development of response and risk models.

In response to the need for longer periods of data collection, compact tags capable of extended attachments were developed. Termed 'low-impact, minimally percutaneous, external-electronics transmitter (LIMPET)' tags (Wildlife Computers), these small tags are attached to the dorsal fin via barbed darts and provide a record of geographic movements and diving behaviors over periods of weeks and months (Andrews et al. 2008). LIMPET tag data are transmitted via satellite, thus obviating the need to recover the tag but greatly constraining the type and resolution of data that can be collected. Nonetheless, these much longer but lower-resolution datasets have confirmed the prevalence of bimodal diving behavior in this species (Schorr et al. 2014, Joyce et al. 2017, Shearer et al. 2019, Barlow et al. 2020). While LIMPET tags have been broadly used to study Cuvier's beaked whales, including their geographic distributions (Schorr et al. 2014, Foley et al. 2021), diving capacities (Schorr et al. 2014, Quick et al. 2020), social behaviors (Cioffi et al. 2021), and reactions to anthropogenic activities (Falcone et al. 2017), interpretation of the datasets they provide have been constrained by their limited range of sensors and temporal resolution. LIMPET tags typically transmit only depth and temperature data with low temporal

resolution; they lack acoustic sensors and do not transmit summarized or raw accelerometer data. Researchers have traditionally used *K*-means clustering analyses (Schorr et al. 2014, Falcone et al. 2017), depth thresholds (Joyce et al. 2017, Shearer et al. 2019), duration thresholds (Quick et al. 2019, 2020, Cioffi et al. 2021), or a combination of methods (Barlow et al. 2020) to assign dives recorded by LIMPET tags into deep and shallow classes, with deep dives presumed to include foraging. However, the assignment of deep dives as foraging dives and shallow dives as non-foraging dives has relied solely upon observations from short-term acoustic tags that may not include the full range of Cuvier's beaked whale behaviors (Johnson et al. 2004, Tyack et al. 2006).

Despite the quantity of DTAG data from Cuvier's beaked whales that support the classification of foraging behavior based on dive depth and duration, deep dives with little or no foraging are occasionally performed by this species and may constitute an avoidance response to disturbances. DeRuiter et al. (2013) reported a non-foraging deep dive recorded by a sound-recording tag on a Cuvier's beaked whale following exposure to simulated mid-frequency active sonar (MFAS). Additionally, a whale in the Ligurian Sea showed reduced foraging effort during a deep dive coincident with a close vessel pass (Aguilar Soto et al. 2006). These observations support the notion that beaked whales may respond to perceived threats by silencing and remaining at depths beyond the reach of shallower-diving predators (Aguilar de Soto et al. 2020), with implications on both foraging efficiency and gas management when they return to the surface. Therefore, understanding the prevalence of such responses and their impact on normal foraging behavior is a valuable step towards assessing the costs of anthropogenic disturbances.

As beaked whale foraging behavior may vary with time of day (Arranz et al. 2011, Barlow et al. 2020) and location, high-resolution data spanning multiple diel cycles and a wide geographic area are ideally needed to develop and validate foraging classification algorithms. Here we used data from new medium-duration (i.e. up to 2 wk) cetacean tags with acoustic sensors that were deployed on Cuvier's beaked whales in southern California, USA, a region where these animals are regularly exposed to MFAS and other anthropogenic activities (Falcone et al. 2017). Compared to suction-cup tags, these dart-attached archival tags provide greatly extended recordings of baseline behavior while also offering an increased chance of sampling responses to anthropogenic disturbances within a single deploy-

ment. Using data from these tags, we developed a machine-learning algorithm for accurately detecting foraging using regularly sampled depth (1 Hz) and acceleration (16 Hz) data. We applied this method to tag data that includes depth and acceleration data but lacks sound recordings, and we discuss the implications of our findings for the design of studies using long-duration tags with lower-resolution data and no sound recordings.

2. MATERIALS AND METHODS

2.1. Data collection

Medium-duration, dart-attached archival tags (Lander II and SMRT; Wildlife Computers) were deployed on Cuvier's beaked whales in the southern California Anti-Submarine Warfare Range from 2018–2019 (Table 1). These tags are invasive Type A tags (Andrews et al. 2019) that anchor to the whale's dorsal surface with 4 LIMPET-style darts that can penetrate up to 7 cm into tissue. The Lander II is a revision to the Whale Lander tag (Owen et al. 2016); its electronics package has a dorso-ventrally flattened ovoid shape that remains outside the body during tag deployments, with maximum dimensions of 18.2 × 10.3 × 2.5 cm (length × width × height). It includes syntactic foam so that once released from the tagged whale, it floats with its Argos transmitter and GPS receiver antennas exposed to facilitate recovery. The Lander II tags collected depth data at 4 Hz with an effective resolution of 1 m, triaxial

accelerometry at 16 Hz, temperature at 1 Hz, and attempted up to 6 Fastloc GPS snapshots per hour. The electronics package shape of SMRT tags is like that of the Lander II but with maximum dimensions of 19.6 × 7.0 × 3.7 cm. The main difference between the tags is that SMRT tags include a single hydrophone for recording sound at a rate of 192 kHz with 16-bit resolution. Sound data were decimated within the SMRT tag by a factor of 2 followed by loss-less compression (Johnson et al. 2013), resulting in a stored sampling rate of 96 kHz. A one-pole high-pass filter was included with a cut-off frequency of 100 Hz, resulting in an approximate −3 dB recording bandwidth from 100 Hz to 44 kHz. SMRT tag memory capacity allowed continuous sound recording for the first 6 d of each deployment. SMRT tags also contained a pressure sensor (effective resolution of 1 m) sampling at 1 Hz, temperature sensor sampling at 0.5 Hz, triaxial accelerometer sampling at 100 Hz (or 50 Hz for 1 tag), and triaxial magnetometer sampling at 25 Hz.

2.2. Data processing and analysis

All data processing and analyses were performed using RStudio (v1.4.1717; R v3.6.0), MATLAB (v9.8.0.1380330), and PAMGUARD (v2.01.03) using a combination of open-source tag analysis tools (www.animaltags.org) and custom tools. Figures were made using the R packages 'ggpubr' v0.4.0 (Kassambara 2020), 'ggformula' v0.10.1 (Kaplan & Pruim 2020), 'ggplot2' v3.3.3 (Wickham 2016), 'patchwork' v1.1.1 (Pedersen 2020), and 'wesanderson' v0.3.6 (Ram & Wickham 2018) and Microsoft 365 Word (v2201; Build 14827.20158). Depth data from Lander II tags and temperature data from SMRT tags were resampled to 1 Hz to match the pressure sensor sampling rate of the SMRT tags, and the pressure data were corrected for temperature sensitivity. Any submergence >50 m was classified as a dive following prior LIMPET tag analyses from this region (Schorr et al. 2014, Falcone et al. 2017, Barlow et al. 2020), and the maximum depth reached and the total dive duration were calculated for each dive. To facilitate a comparison between dive classifications from the model developed in this study and conventional methods used for Cuvier's beaked whales in southern California

Table 1. Summary of *Ziphius cavirostris* tag deployments and number of dives recorded by each tag. Deep and shallow dives were classified using *K*-means clustering based on maximum dive depth and duration. The numbers of dives with/without acoustic evidence of foraging are provided in parentheses for SMRT tags. Note that not all SMRT tag dives included acoustic data

Deployment date (d/mo/yr)	Tag type	Deployment duration (d)	No. of completed dives (no. of foraging/non-foraging dives)	
			Deep dives	Shallow dives
13/01/2018	Lander II	5.7	50	250
30/03/2018	Lander II	19.3	158	583
31/03/2018	Lander II	5.6	50	199
11/01/2019	Lander II	9.1	71	318
13/01/2019	SMRT	0.5	4 (4/0)	15 (0/15)
12/10/2019	SMRT	5.9	52 (36/10)	198 (1/178)
12/10/2019	SMRT	7.0	58 (42/0)	229 (0/177)
11/11/2019	SMRT	12.1	104 (40/1)	410 (0/176)
17/11/2019	SMRT	0.3	1 (1/0)	11 (1/10)

(Schorr et al. 2014, Falcone et al. 2017, Barlow et al. 2020), *K*-means clustering was performed on a per-individual basis to classify dives as either deep (conventionally assumed to be foraging dives) or shallow (conventionally assumed to be non-foraging dives) using scaled dive depth and duration (mean-centered and scaled by the standard deviation). The bottom phase of each dive was estimated as the time between the first and last inversion in the vertical direction of travel that occurred below 73% of the maximum depth reached during the dive (73% is equivalent to the maximum echolocation clicking start depth relative to the dive depth across all SMRT foraging dives).

Dives during which sound was recorded (a subset of SMRT tag dives) were classified as foraging dives if echolocation clicks and buzzes produced by the tagged whale were recorded during the dive (Johnson et al. 2004, Aguilar Soto et al. 2006, Tyack et al. 2006, DeRuiter et al. 2013, Alcázar-Treviño et al. 2021). Audio files from each SMRT tag were manually reviewed using the PAMGUARD spectrogram module (Gillespie et al. 2009) to determine the start and end time of echolocation clicks and buzzes. Echolocations from the tagged individual were manually distinguished from those of conspecifics or nearby delphinids based on the presence of relatively high spectral energy below 20 kHz and relatively consistent click amplitudes over sequences of clicks (Zimmer et al. 2005, Johnson et al. 2006). Echolocations not assigned to the tagged individual were excluded from the analysis. Due to varying signal-to-noise levels across tag recordings, buzz presence could not be reliably determined throughout all dives from all tags.

Triaxial acceleration data recorded by the SMRT tags were interpolated and then decimated by the necessary factors to resample the data to 16 Hz, thus matching the acceleration sampling rates of the Lander II tags. For 50 Hz data, this required an 8-fold interpolation followed by a 25-fold decimation; for 100 Hz data, it required a 4-fold interpolation followed by a 25-fold decimation. To measure the extent to which Cuvier's beaked whales exhibited postural changes when searching for, pursuing, and capturing prey, we calculated roll circular variances (Berens 2009) over the combined descent and bottom phase of each dive, excluding periods when the whale had a steep pitch angle within 20° of a vertically-oriented posture (i.e. to avoid erroneous roll values due to gimbal lock at high pitch angles). Several additional dive parameters were computed to guide automatic classification of foraging and non-

foraging dives. Descent (start of dive to start of bottom phase) and ascent (end of bottom phase to end of dive) rates (defined as the total change in depth during the ascent or descent divided by the duration of the ascent or descent) were calculated, as these were found to differ between deep and shallow dives in prior studies (Baird et al. 2006, 2008, Tyack et al. 2006). We also calculated the proportion of sign changes in the first difference of the bottom-phase depth time series (Miller et al. 2015) and the bottom-phase average vertical speed (i.e. bottom-phase mean absolute vertical speed ignoring travel direction) to identify changes in the vertical movements of whales when they were likely pursuing prey.

In an attempt to identify individual prey captures, acceleration transients potentially associated with strikes at prey (Ydesen et al. 2014, Sweeney et al. 2019) were detected by first computing the norm-jerk (i.e. the vector magnitude of the triaxial acceleration differential) using data at the common 16 Hz sampling rate. Transients were then detected during the descent and bottom phases when the norm-jerk signal surpassed the maximum norm-jerk value during each respective dive's ascent (disregarding the last 5 s of the ascent when norm-jerk peaks can be abnormally large as whales approach the surface). We used the maximum norm-jerk value during the dive ascent, when the whale was not presumed to be foraging, as the peak detector threshold to curtail false positive detections at norm-jerk levels comparable to non-foraging periods. Adjacent norm-jerk peaks separated by <3.5 s (average buzz duration from SMRT tags) were combined and counted as the same peak (Sweeney et al. 2019).

Due to concerns that 16 Hz was an insufficient sampling rate to consistently identify prey captures (Ydesen et al. 2014), we tested whether jerk transients could be distinguished from the norm-jerk signal before, during, and after each identified buzz by calculating the root-mean-square (RMS) of the norm-jerk signal within 1 s windows (n = 3): 1 window centered at the end time of each buzz, 1 window ended at the start time of each buzz, and 1 window started as many seconds as the buzz was in duration after the end time of each buzz. To test the hypothesis (α = 0.05) that RMS norm-jerk values centered at the end time of buzzes differed from those before and after each buzz, we used the R package 'glmmTMB' v1.0.2.1 (Brooks et al. 2017) to fit a gamma mixed-effects regression with a log link function with RMS norm-jerk as the response variable. We chose gamma rather than standard Gaussian regression since RMS norm-jerk values are always non-nega-

tive and tend to be right-skewed. In addition to the categorical predictor variable differentiating between RMS norm-jerk time windows, we included nested random effects to account for autocorrelation within tagged individuals, foraging dives, and distinct buzzes. Regardless of whether RMS norm-jerk values were greater at the end time of buzzes compared to before and after buzzes, identified norm-jerk peaks could be associated with either prey captures or rapid maneuvering not directly related to prey captures. Therefore, we determined how many identified norm-jerk peaks (using both 16 Hz data from all tags and 100 Hz data from the 4 tags that sampled acceleration at 100 Hz) were within 3.5 s (average buzz duration from SMRT tags) of the end time of identified buzzes.

2.3. Model development and application

We assessed the feasibility of accurately classifying dives as foraging or non-foraging dives via machine learning using variables derived from 1 Hz depth data and 16 Hz triaxial acceleration data by first creating 500 replicates of all dives with concurrent sound recordings (n = 692). Each replicate was then randomly partitioned into training and testing datasets using the R package 'caret' v6.0-86 (Kuhn 2020). Training and testing datasets were partitioned using a roughly two-thirds (n = 462) to one-third (n = 230) split, respectively, where each split dataset had approximately the same percentage of foraging dives as the full dataset of dives with acoustic data (18.1%). Using the 500 partitioned training datasets, we created extreme gradient boosting tree models from the R package 'xgboost' v1.4.1.1 (Chen et al. 2021) to classify dives (foraging or non-foraging) based on dive depth, dive duration, bottom-phase duration, ascent rate, descent rate, proportion of sign changes in the first difference of the bottom-phase depth time series, bottom-phase average vertical speed, the number of norm-jerk peaks during the descent and bottom phase, and roll circular variance during the descent and bottom phase (Code S1 in the Supplement at www.int-res.com/articles/suppl/m692 p195_supp.pdf). For each replicate model, we performed a hyperparameter optimization grid search via 10-fold cross validation repeated 10 times using parallel computing with a socket cluster using the R packages 'caret' v6.0-86, 'snow' v0.4-3, and 'doS-NOW' v1.0.19 (Tierney et al. 2018, Kuhn 2020, Microsoft Corporation & Weston 2020) to find the best-fit model based on overall classification accu-

racy within the training dataset. The following hyperparameters for the optimization grid search were chosen to allow sufficient model complexity without overfitting: eta (step size shrinkage) from 0.5–1 by steps of 0.1, max_depth (maximum tree depth) from 1–3 by steps of 1, min_child_weight (minimum sum of instance weights needed) was 0.25–1 by steps of 0.25, colsample_bytree (subsample ratio of predictors when constructing each tree) was 1, gamma (minimum required loss reduction) was 0.25–1 by steps of 0.25, nrounds (number of trees) from 1–4 by steps of 1, and subsample was 1. Model goodness-of-fit and prediction accuracies were summarized across all 500 model replicates.

To predict the presence of foraging in dives from tags without acoustic data, which consisted of some dives from SMRT tags (due to programmed sensor shutdowns) and all dives from Lander II tags, we refit the model using all SMRT tag dives with acoustic data (n = 692) and the same variables and optimization methods previously listed, thus maximizing the volume of training data (Code S2). We hereafter refer to this model as the comprehensive model. Fractional contributions of each covariate in the comprehensive model were calculated based on the total predictive information gained from the variable's splits (Chen et al. 2021).

3. RESULTS

Tags recorded 2069 dives without acoustic data and 692 dives with acoustic data. *K*-means clustering classified these dives using depth and duration into 2213 shallow dives (558 with acoustic data) and 548 deep dives (134 with acoustic data). Of the 134 deep dives with acoustic data, 11 dives (8.2%) did not contain echolocation clicks from the tagged whale (or any conspecific), whereas 2 shallow dives (0.36% of 558 shallow dives with acoustic data) did. Thus, 98.1% of the 692 dives with acoustic data were correctly classified with respect to presumed foraging by simple *K*-means clustering. The depth and duration ranges of dives from all tags are shown in Fig. 1.

Across the 692 dives with acoustic data from SMRT-tagged whales, 2526 buzzes from the tagged whales were identified. However, only 625 of these buzzes (24.7%) had either 16 Hz or 100 Hz norm-jerk peaks within 3.5 s of the end time of the buzz. Despite the small percentage of buzzes (i.e. probable prey captures) that were associated with above-threshold norm-jerk peaks, moderate associations between jerk peaks and buzz times were still
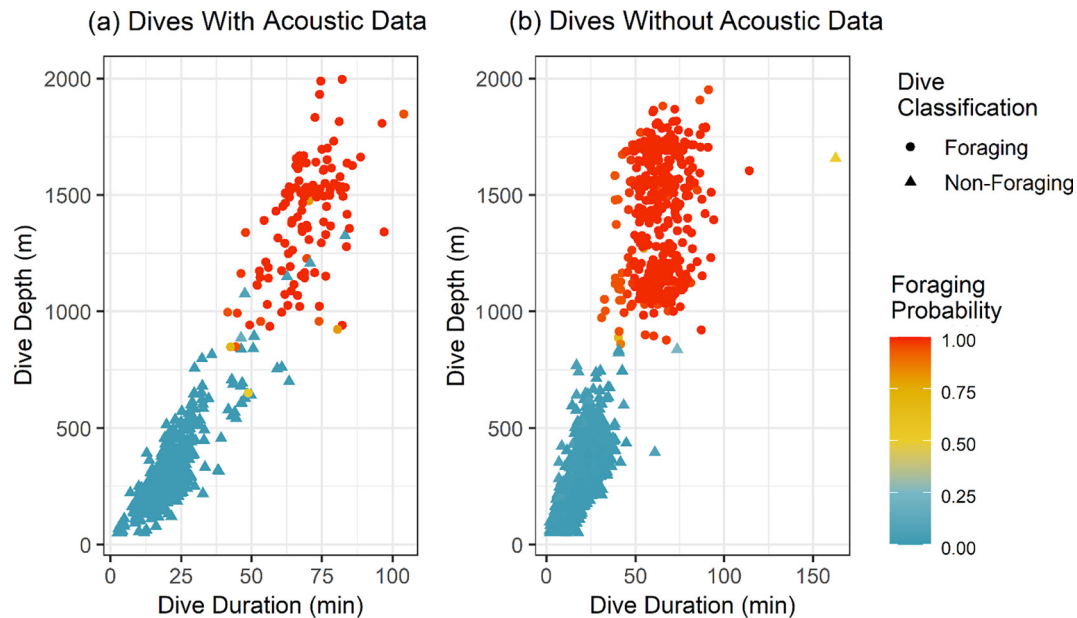
Fig. 1. Comprehensive model classifications of all *Ziphius cavirostris* dives (a) with acoustic data and (b) without acoustic data. Data symbols show the dive classifications by the comprehensive model (which also match the acoustically validated results in panel a), while point colors show the model-fitted probabilities that a given dive was a foraging dive. Note that *K*-means clustering is not presented here and that the horizontal axis scales are different between the 2 panels

observed. RMS norm-jerk values (16 Hz sampling rate) centered around the end time of buzzes (n = 2526) were significantly greater than those both before ($\beta = 0.83$, SE = 0.02, $z = 53.62$, p < 0.001) and after each buzz ($\beta = 0.98$, SE = 0.02, $z = 59.15$, p < 0.001). Of the 125 foraging dives with acoustic data, 105 (84%) had 16 Hz norm-jerk peaks within 3.5 s of the end time of a buzz. In the 125 foraging dives with acoustic data, 620 of 1376 (45.1%) norm-jerk peaks detected during foraging dive descents and bottom phases were within 3.5 s of the end time of a buzz. The median ratio of the number of 16 Hz jerk peaks associated with buzzes to the total number of buzzes during the dive was 0.21 (range = 0–1, inter-quartile range = 0.40). To test if these low percentages of jerk peaks associated with buzzes were a consequence of the low (16 Hz) acceleration sampling rate, the analysis was re-run on the 121 foraging dives with acoustic data from the 4 tags that recorded 100 Hz triaxial acceleration. On these tags, 468 out of 1232 (38.0%) 100 Hz norm-jerk peaks were associated with buzz times.

Although random data partitioning allowed us to test the accuracy of the 500 model replicates, it often reduced the number of deep, non-foraging dives used in model training. Given the biological relevance of accurately determining whether foraging occurred during deep dives or not, we used Kendall's

rank correlation tests to assess how prediction accuracies from the 500 model replicates varied depending on the proportion of deep, non-foraging dives present in the training datasets. These tests revealed that the proportion of deep, non-foraging dives in the

Table 2. Summary of model accuracies from the 500 models using randomly partitioned training and testing datasets comprised of foraging and non-foraging dives by *Ziphius cavirostris* for which definitive classification was possible based on acoustic data. Results for the training datasets reflect the ability of the models to properly fit the training datasets, while testing dataset results reflect how well the fitted models predicted dives from the subset of the data that were not used for training. Combined dataset results showcase the accuracy of the fitted models to classify dives from both the training and testing datasets combined (see Figs. S1 & S2 for further details on these results). IQR: inter-quartile range

	Median proportion of correctly classified dives (IQR; range)	
	Foraging dives	Non-foraging dives
Training datasets	1 (0.012; 0.976–1)	1 (0; 0.995–1)
Testing datasets	0.976 (0.049; 0.829–1)	1 (0.005; 0.968–1)
Combined datasets	0.992 (0.008; 0.944–1)	1 (0.002; 0.988–1)

training datasets was correlated with the proportion of correctly classified deep, non-foraging dives in the testing datasets (Kendall's tau = 0.2, z = 5.8, p < 0.001), but not correlated with the proportion of correctly classified foraging dives in the testing datasets (Kendall's tau = 0.06, z = 1.67, p = 0.094). Due to imposed model complexity limitations, 352 of the 500 (70.4%) models using partitioned data perfectly fit the training data. However, of the 148 models that did not perfectly fit training data, the average number of misclassified dives was only 1.2 (median = 1, range = 1–3). An average of 99.9% of training dataset dives (median = 100%, range = 99.4–100%) and 99.2% of testing dataset dives (median = 99.3%, range = 97.0–100%) were correctly classified across the 500 models (Table 2; Figs. S1 & S2).

The comprehensive model (Fig. 2) fit the training dataset (i.e. all 692 dives with acoustic data) with
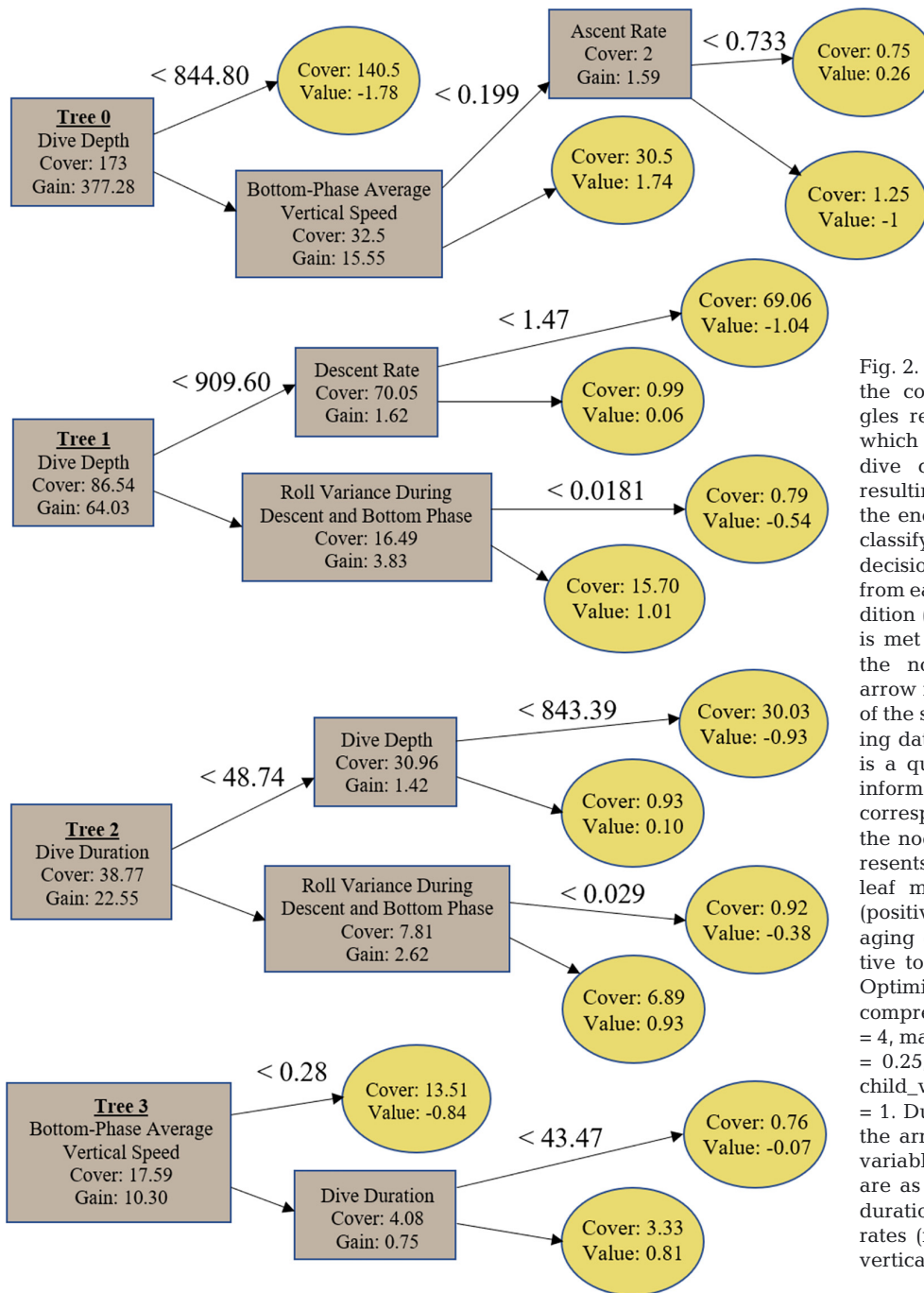
Fig. 2. Optimized decision trees from the comprehensive model. Rectangles represent decision nodes from which a set of arrows point to guide dive classifications along towards resulting leaves (yellow ellipses) at the end of each decision tree. When classifying each dive via these four decision trees, the top arrow coming from each node is followed if the condition (shown above each top arrow) is met for the variable listed within the node. Otherwise, the bottom arrow is followed. 'Cover' is the sum of the second-order gradient of training data classified to the leaf. 'Gain' is a quantification representing the information gained from a split (thus corresponding to the importance of the node in the model). 'Value' represents the marginal value that the leaf may contribute to predictions (positive values contribute to foraging dive classifications and negative to non-foraging classifications). Optimized hyperparameters for the comprehensive model were: nrounds = 4, max_depth = 3, eta = 0.9, gamma = 0.25, colsample_bytree = 1, min_child_weight = 0.75, and subsample = 1. Due to space limitations around the arrows coming from each node, variable units for decision thresholds are as follows: dive depth (m), dive duration (min), descent and ascent rates (m s^{-1}), bottom-phase average vertical speed (m s^{-1}), roll variance (radians)
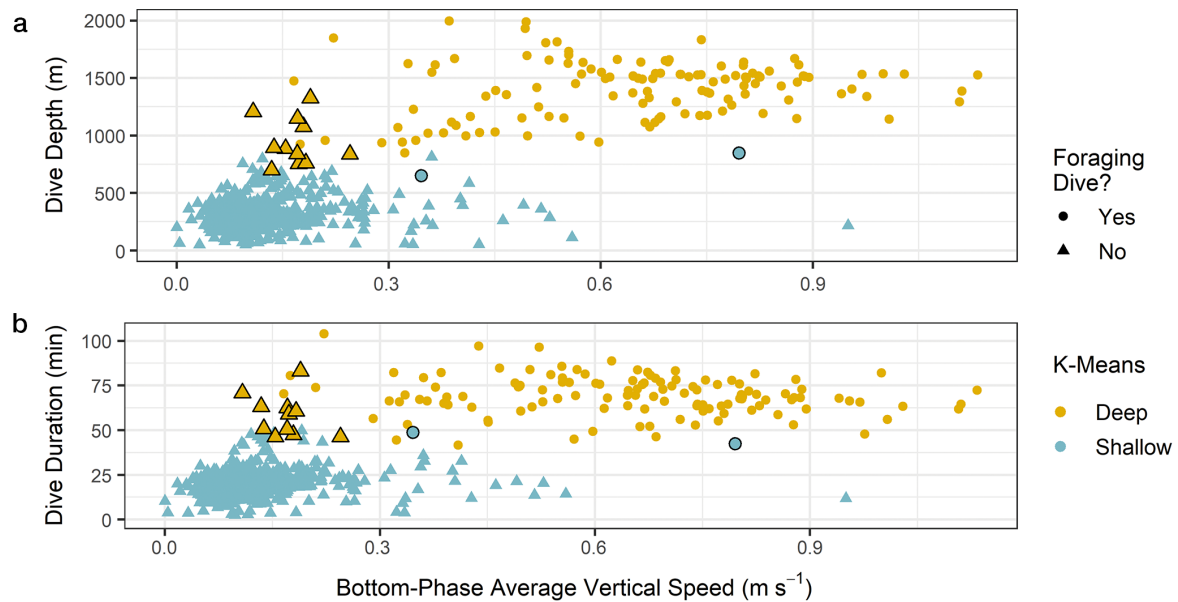
Fig. 3. Plots of *Ziphius cavirostris* dive (a) depth and (b) duration as a function of the bottom-phase average vertical speed (m s^{-1}) in dives used to construct the comprehensive model. Points are shaped based on whether foraging was acoustically detected (i.e. echolocation clicks from the tagged whales were recorded) during the dive, and point colors display the classification of each dive as deep or shallow using *K*-means clustering. Deep, non-foraging dives and shallow, foraging dives (i.e. dives incorrectly classified via *K*-means clustering), are slightly enlarged and outlined in black

perfect accuracy (Fig. 1a). Within the 4 constructed trees (Fig. 2) of the comprehensive model, 11 split nodes were formed that classified dives using dive depth, dive duration, bottom-phase average vertical speed (Fig. 3), roll circular variance during the

descent and bottom phase (Fig. 4), ascent rate (Fig. 5), and descent rate (Fig. 6). The fractional contributions (scale from 0 to 1) of each variable to the comprehensive model were: dive depth = 0.883, bottom-phase average vertical speed =
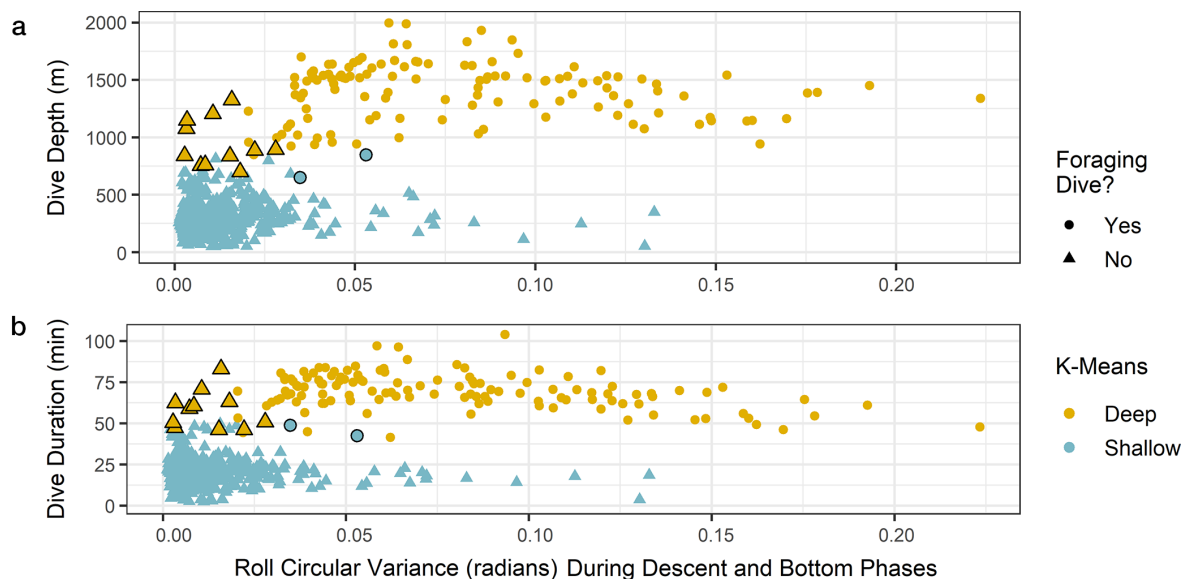

Fig. 4. Plots of *Ziphius cavirostris* dive (a) depth and (b) duration as a function of the roll circular variance (radians) during the descent and bottom phase in dives used to construct the comprehensive model. Other details as in Fig. 3
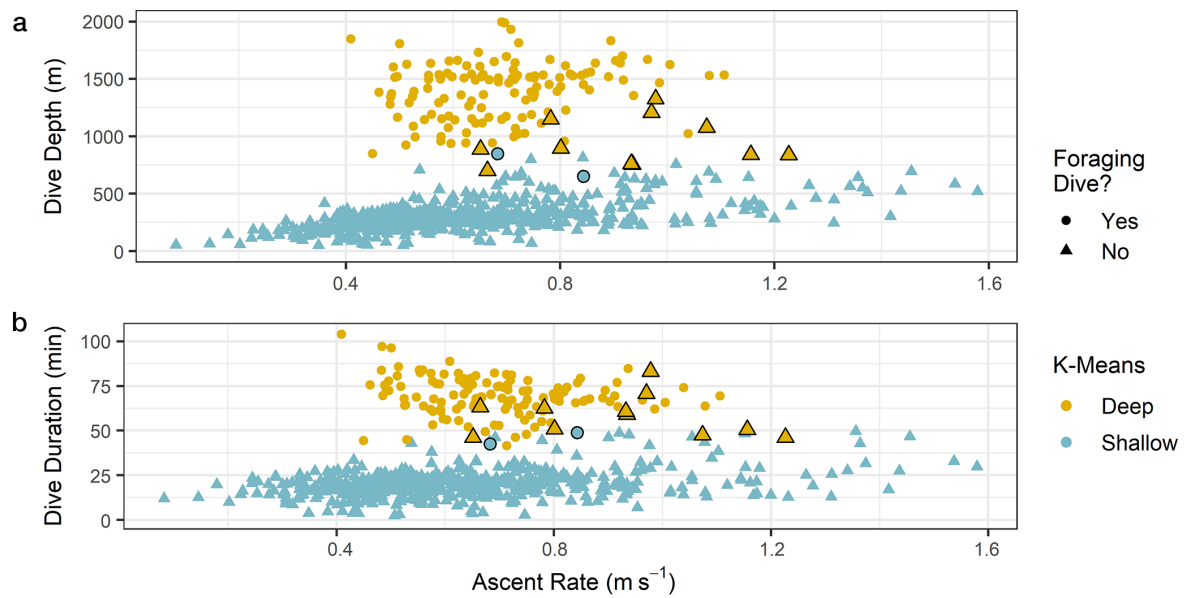
Fig. 5. Plots of *Ziphius cavirostris* dive (a) depth and (b) duration as a function of the ascent rate (m s^{-1}) in dives used to construct the comprehensive model. Other details as in Fig. 3

0.052, dive duration = 0.046, roll circular variance during the descent and bottom phase = 0.013, ascent rate = 0.003, descent rate = 0.003. Upon combining all dives with and without acoustic data across all tags (Fig. 1), 13 of 548 (2.4%) *K*-means clustered deep dives were classified as non-foraging dives, and 7 of 2213 (0.3%) *K*-means clustered shallow dives were classified as foraging dives. These 20 dives (Figs. S4–S23) with differing foraging classifications between *K*-means clustering (where deep are presumed foraging and shallow are presumed non-foraging) and the comprehensive model accounted for 0.72% of all dives from all tags (Table 1).
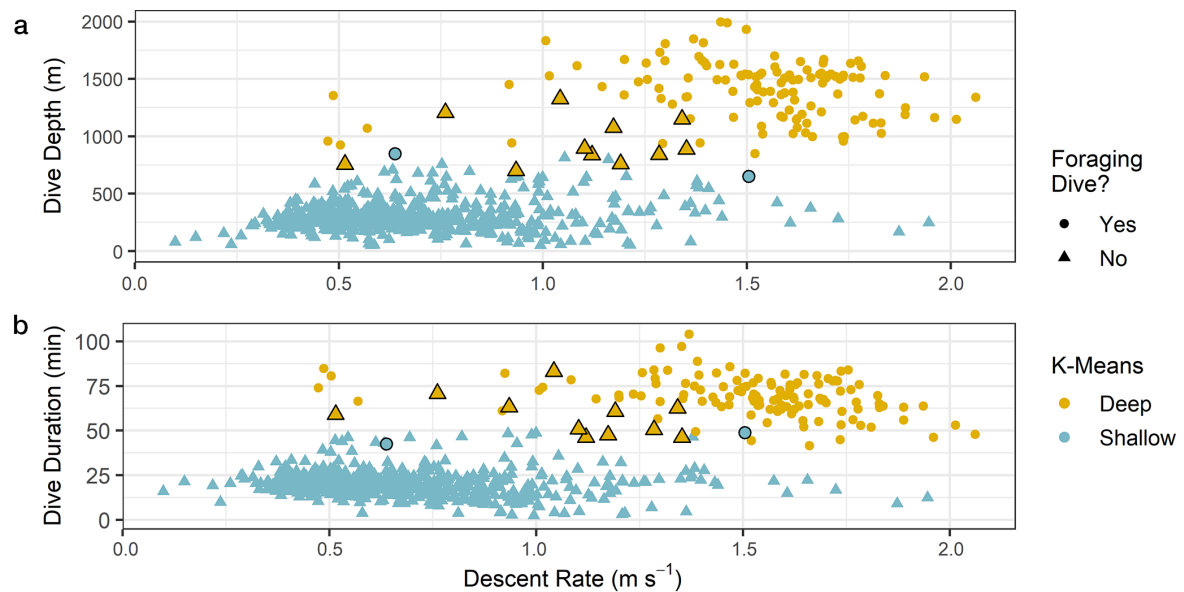

Fig. 6. Plots of *Ziphius cavirostris* dive (a) depth and (b) duration as a function of the descent rate (m s^{-1}) in dives used to construct the comprehensive model. Other details as in Fig. 3

4. DISCUSSION

As the amount of tag data collected to study Cuvier's beaked whale biology and responses to anthropogenic disturbances continues to grow, so too does the need for robust methods of inferring foraging from the relatively low-temporal-resolution data often returned by these tags. Using data from 5 medium-duration archival tags with pressure sensors, accelerometers, and acoustic sensors deployed on Cuvier's beaked whales in a region where these animals are commonly exposed to MFAS and other anthropogenic activities, we developed a model capable of determining if foraging occurred during dives without concurrent sound recordings. The resulting classification algorithm allows us to leverage the full suite of archival tag data available from this project for behavioral studies where foraging disruption is a key response metric. It also provides insight into the accuracy of prior studies (Baird et al. 2006, 2008, Schorr et al. 2014, Falcone et al. 2017, Joyce et al. 2017, Quick et al. 2019, 2020, Shearer et al. 2019, Barlow et al. 2020, Cioffi et al. 2021) that have inferred foraging using only low-resolution depth data.

K-means clustering has been used to infer foraging behavior from low-resolution Cuvier's beaked whale tag data in southern California (Schorr et al. 2014, Falcone et al. 2017, Barlow et al. 2020). Our findings confirm that maximum dive depth and dive duration can be used to accurately infer foraging status for most dives, given that 98.1% of dives with acoustic data were correctly classified with respect to presumed foraging by simple *K*-means clustering. Most of the incorrectly classified dives using *K*-means clustering (11 out of 13) were deep dives that did not contain echolocation clicks. Despite occurring infrequently, such non-foraging deep dives may have major implications for anthropogenic disturbance analyses where foraging disruption is a key metric and where all deep, long dives are assumed to include foraging. The ability to recognize these unusual deep dives will increase our understanding of the circumstances in which they occur.

Although the use of 500 model replicates allowed effective assessment of the capabilities of extreme gradient boosting tree models on these Cuvier's beaked whale tag data, they were inherently limited in their ability to accurately classify dives since they were fit using a subsample of the available dataset, and the unusual dives we are most interested in characterizing represented a very small fraction of the dives. To make the most of the available data

with definitive acoustic foraging indications, we performed model fitting in 2 steps. First, we performed a cross-validation study on random subsets of the data to show that model fits consistently yielded strong classification performance. As expected, our results showed that increasing the proportion of non-foraging deep dives in a training dataset improved the ability of the models to accurately predict such dives in other data. We then fit a final model (the comprehensive model) to the entire dataset with acoustic data under the assumption that, by maximizing the volume of training data in the comprehensive model, model performance should be optimal. Thus, we assume that the accuracy of our predicted foraging classifications for all dives without associated sound recordings is similar to the summarized prediction accuracy from the 500 cross-validation models fit to subsets of the data.

The foundational presence of dive depth in the comprehensive model (the leading split node in the first 2 trees with a fractional contribution to the model of 0.883; Fig. 2) is not surprising given that Cuvier's beaked whales are known to feed primarily on cephalopods and benthic fish that are found at great depths (West et al. 2017). The importance of dive depth in the comprehensive model also explains why only 0.72% of dives were misclassified by simple *K*-means clustering using only dive depth and duration. The strong bimodality exhibited by these whales aided model fitting of dives with acoustic data that occurred towards the ends of certain tag deployments where we suspect the tag had come loose based on amplified signal noise in the acceleration data. There was one such dive without acoustic data that occurred while the tag was seemingly loose and that was classified as a shallow dive according to *K*-means clustering but was predicted to include foraging according to the comprehensive model (Fig. S4). Although the probability of this dive including foraging according to the comprehensive model likely increased due to the exaggerated roll variance from the loose tag, the depth-related covariates in the comprehensive model provided sufficient evidence to support the foraging classification for this dive regardless of the roll variance level.

Despite the inclusion of dive duration in the comprehensive model, the very small fractional importance of dive duration (0.046) relative to that of dive depth suggests that using a dive duration threshold alone would be a less accurate method of classifying dives compared to a depth-only threshold. Descent and ascent rates had the lowest fractional contributions to the comprehensive model (0.003), and the 13

dives with acoustic data for which *K*-means clustering classifications did not accurately detect foraging were all scattered across the observed distributions of these 2 variables (Figs. 5 & 6).

The role of bottom-phase average vertical speed in the optimized model (Fig. 3) suggests that Cuvier's beaked whales in southern California perform frequent vertical excursions when pursuing prey. Only 4 foraging dives with acoustic data (3.2%) had bottom-phase average vertical speeds <0.284 m s^{-1} (Fig. 3), whereas all 11 non-foraging deep dives with acoustic data had bottom-phase average vertical speeds below this same level (threshold in tree 3 of Fig. 2). All 6 non-foraging deep dives that reached depths >844.8 m had average bottom-phase vertical speeds <0.2 m s^{-1} (thresholds in tree 0 of Fig. 2). Shallow dives were strongly clustered with low bottom-phase average vertical speeds, typically <~0.3 m s^{-1}, although shallow dives with high bottom-phase average vertical speeds were occasionally observed with no evidence of foraging in the acoustic record. Several of these active shallow dives were preceded by deep dives with conspecific click detections, suggesting the elevated activity levels may in some cases have been social in nature.

The descents and bottom-phases of foraging dives (during periods with shallow pitch angles) tended to possess far more roll variance than those of non-foraging dives (Fig. 4), likely due to postural changes when searching for and pursuing prey, as seen in other odontocetes (Miller et al. 2004, 2015, Madsen et al. 2005, Stimpert et al. 2014). Within the dataset with acoustic data, 9 of the 11 non-foraging deep dives had roll variance levels lower than any foraging dive (Fig. 4). Abnormally low levels of roll have also been observed in other beaked whale species during non-foraging deep dives coincident with exposure to simulated sonar (Stimpert et al. 2014, Miller et al. 2015).

One compelling application of this dive classification model would be to incorporate it into tag firmware, so that foraging could be accurately detected in dives as they are recorded, potentially obviating the need to transmit or recover the higher-resolution depth and accelerometer data while still capturing the important distinctions they provide. The presence of roll, an orientation-dependent parameter, in the comprehensive model is a potential obstacle, however. Dart-attached tags are typically applied to free-ranging animals either ballistically or with a pole and, in both cases, the location and orientation of the tag on the animal cannot be controlled precisely. The onboard processing algorithm would

therefore need to infer *in situ* the tag orientation on the animal to estimate roll. Although possible, for example using the measured orientation when the animal is breathing at the surface, it would add considerable complexity to the data-processing algorithm of the tag. An exploratory model run using the same methods as the comprehensive model but excluding orientation-dependent variables (Code S3 and Fig. S3) produced a model that fit the training dataset with 99.6% classification accuracy and only predicted 2 dives without associated acoustic data differently than the full comprehensive model. Therefore, with only modest alterations to the existing model framework and input parameters, similar results can be obtained using extreme gradient boosting tree algorithms, although the use of orientation-dependent parameters (as they were calculated in this study) does still improve data fitting accuracy.

Norm-jerk peaks do not appear to be a reliable metric for the identification of foraging in dives by Cuvier's beaked whales in southern California. Although RMS norm-jerk values centered around the end time of buzzes were significantly greater than those both before and after each buzz, most jerk peaks above the predetermined thresholds occurred without an associated buzz. Only 45.1 and 38.0% of 16 and 100 Hz norm-jerk peaks, respectively, were within 3.5 s of a buzz (i.e. high false positive rate), and many buzzes were not associated with jerk peaks (i.e. many missed detections). This could indicate that capturing different species and sizes of prey in southern California (Adams et al. 2015) may require varying degrees of kinematic activity (e.g. rapid cranial motions like those exhibited by other marine predators; Kokubun et al. 2011, Iwata et al. 2012, Ydesen et al. 2014). Alternatively, the limited connection between jerk peaks and buzzes in our dataset could suggest either that the sampling rates were inadequate to detect these motions or that the acceleration transients did not propagate effectively to the tag, possibly due to the location of the tag on the whale body. Ultimately, the moderate performance of the jerk peak detector at identifying buzz times showcases the difficulties in balancing true positive and false positive detections using a predetermined threshold and getting unambiguous signals from a target behavior (Sweeney et al. 2019).

The large dataset available to this study supports the picture of a strongly stereotyped diving behavior in Cuvier's beaked whales from southern California; most dives can be correctly classified with respect to foraging activity by depth and duration alone. However, the longer tag durations available in this study

captured some previously undescribed diving behavior. For example, 1 deep dive was recorded in which the tagged whale clicked for about 3.5 min and performed 2 buzzes during its descent before aborting foraging. Shortly thereafter, the whale performed a 4.2 min bottom phase and ascended towards the surface. The ascent rate (0.450 m s^{-1}) and roll circular variance (0.022 radians) of this dive were relatively low compared to other foraging dives, whereas the bottom-phase average vertical speed (0.323 m s^{-1}) and descent rate (1.52 m s^{-1}) were well within the typical ranges for foraging dives. This was the only foraging dive with acoustic data that possessed less than 13.3 min of clicking (average clicking duration was 30.6 min). Although classified as a foraging dive with a predicted foraging probability of 0.938 (fifth lowest probability of foraging among foraging dives with acoustic data), the entire bottom phase of this dive did not contain clicks. Dive duration was only 44.4 min and maximum depth was 849 m; were this dive only 5 m shallower, it would have been classified as a non-foraging dive. This dive provides a specific example of the ability of the model to highlight abnormal dives that can be further investigated for behavioral state changes. In fact, the cessation of foraging in this unusual dive coincided with an explosive event in the acoustic record of the tag. Had this dive simply been recorded as a deep dive and assumed to include normal foraging behavior as in LIMPET tag studies, true foraging disruption would have been underestimated.

This study provides an assessment of the accuracy that can be achieved in classifying foraging and non-foraging dives based on long-duration datasets from animals within a population of Cuvier's beaked whales regularly exposed to anthropogenic activities. These results confirm that dives classified by only depth and duration provide reasonable estimates of longer-term patterns in foraging effort, even for a population regularly exposed to anthropogenic activities. However, our findings also suggest that the addition of roll circular variance, bottom-phase average vertical speed, and ascent and descent rates enhances detections of unusual dives (e.g. long, deep dives without foraging effort). This is an important outcome because acceleration data could be recorded, summarized, and transmitted by small satellite-linked tags (e.g. LIMPET tags at 16 Hz), whereas such tags cannot currently record and process sound. The decision tree algorithm from this model can help inform future tag designs, potentially suggesting a way of summarizing raw accelerometer and depth data into reliable metrics that can be transmitted via Argos, thus allowing these compact tags to collect long-duration datasets with accurate foraging classifications. Although our model fit is specific to Cuvier's beaked whales in southern California, the same methodology could be applied to develop models for other populations and species using either the same or different variables.

LITERATURE CITED

Adams J, Walker WA, Burton EJ, Harvey JT (2015) Stomach contents of a Cuvier's beaked whale (*Ziphius cavirostris*) stranded in Monterey Bay, California. Northwest Nat 96: 93–98

Aguilar de Soto N, Visser F, Tyack PL, Alcazar J and others (2020) Fear of killer whales drives extreme synchrony in deep diving beaked whales. Sci Rep 10:13

Aguilar Soto N, Johnson M, Madsen PT, Tyack PL, Bocconcelli A, Fabrizio Borsani J (2006) Does intense ship noise disrupt foraging in deep-diving Cuvier's beaked whales (*Ziphius cavirostris*)? Mar Mamm Sci 22:690–699

Alcázar-Treviño J, Johnson M, Arranz P, Warren VE and others (2021) Deep-diving beaked whales dive together but forage apart. Proc R Soc B 288:20201905

Andrews R, Pitman R, Ballance L (2008) Satellite tracking reveals distinct movement patterns for Type B and Type C killer whales in the southern Ross Sea, Antarctica. Polar Biol 31:1461–1468

Andrews RD, Baird RW, Calambokidis J, Goertz CEC and others (2019) Best practice guidelines for cetacean tagging. J Cetacean Res Manag 20:27–66

Arranz P, Aquilar de Soto N, Madsen PT, Brito A, Bordes F, Johnson MP (2011) Following a foraging fish-finder: diel habitat use of Blainville's beaked whales revealed by echolocation. PLOS ONE 6:e28353

Baird RW, Webster DL, McSweeney DJ, Ligon AD, Schorr GS, Barlow J (2006) Diving behaviour of Cuvier's (*Ziphius cavirostris*) and Blainville's (*Mesoplodon densirostris*) beaked whales in Hawai'i. Can J Zool 84: 1120–1128

Baird RW, Webster DL, Schorr GS, McSweeney DJ, Barlow J

(2008) Diel variation in beaked whale diving behavior. Mar Mamm Sci 24:630−642

Barlow J, Schorr GS, Falcone EA, Moretti D (2020) Variation in dive behavior of Cuvier's beaked whales with seafloor depth, time-of-day, and lunar illumination. Mar Ecol Prog Ser 644:199−214

Berens P (2009) CircStat: a MATLAB toolbox for circular statistics. J Stat Softw 31:1−21

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A and others (2017) GlmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J 9:378−400

Chen T, He T, Benesty M, Khotilovich V and others (2021) Xgboost: extreme gradient boosting. https://CRAN.R-project.org/package=xgboost

Cioffi WR, Quick NJ, Foley HJ, Waples DM and others (2021) Adult male Cuvier's beaked whales (*Ziphius cavirostris*) engage in prolonged bouts of synchronous diving. Mar Mamm Sci 37:1085−1100

D'Amico A, Gisner RC, Ketten DR, Hammock JA, Johnson C, Tyack PL, Mead J (2009) Beaked whale strandings and naval exercises. Aquat Mamm 35:452−472

DeRuiter SL, Southall BL, Calambokidis J, Zimmer WMX and others (2013) First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. Biol Lett 9:20130223

Evans K, Lea MA, Patterson TA (2013) Recent advances in bio-logging science: technologies and methods for understanding animal behaviour and physiology and their environments. Deep Sea Res II 88–89:1−6

Fahlman A, Tyack PL, Miller PJO, Kvadsheim PH (2014) How man-made interference might cause gas bubble emboli in deep diving whales. Front Physiol 5:13

Falcone EA, Schorr GS, Watwood SL, DeRuiter SL and others (2017) Diving behaviour of Cuvier's beaked whales exposed to two types of military sonar. R Soc Open Sci 4:170629

Filadelfo R, Mintz J, Michlovich E, D'Amico A, Tyack PL, Ketten DR (2009) Correlating military sonar use with beaked whale mass strandings: What do the historical data show? Aquat Mamm 35:435−444

Foley HJ, Pacifici K, Baird RW, Webster DL, Swaim ZT, Read AJ (2021) Residency and movement patterns of Cuvier's beaked whales *Ziphius cavirostris* off Cape Hatteras, North Carolina, USA. Mar Ecol Prog Ser 660:203−216

Gillespie D, Mellinger DK, Gordon J and others (2009) PAMGUARD: semiautomated, open source software for real-time acoustic detection and localization of cetaceans. J Acoust Soc Am 125:2547

Hooker SK, Baird RW, Fahlman A (2009) Could beaked whales get the bends? Effect of diving behaviour and physiology on modelled gas exchange for three species: *Ziphius cavirostris, Mesoplodon densirostris* and *Hyperoodon ampullatus.* Respir Physiol Neurobiol 167:235−246

Iwata T, Sakamoto KQ, Takahashi A, Edwards EWJ, Staniland IJ, Trathan PN, Naito Y (2012) Using a mandible accelerometer to study fine-scale foraging behavior of free-ranging Antarctic fur seals. Mar Mamm Sci 28:345−357

Johnson MP, Tyack PL (2003) A digital acoustic recording tag for measuring the response of wild marine mammals to sound. IEEE J Oceanic Eng 28:3−12

Johnson M, Madsen PT, Zimmer WMX, Aguilar de Soto N, Tyack PL (2004) Beaked whales echolocate on prey. Proc R Soc B 271:S383−S386

Johnson M, Madsen PT, Zimmer WMX, Aguilar de Soto N, Tyack PL (2006) Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation. J Exp Biol 209:5038−5050

Johnson M, Partan J, Hurst T (2013) Low complexity lossless compression of underwater sound recordings. J Acoust Soc Am 133:1387−1398

Joyce TW, Durban JW, Claridge DE, Dunn CA and others (2017) Physiological, morphological, and ecological tradeoffs influence vertical habitat use of deep-diving toothed-whales in the Bahamas. PLOS ONE 12:e0185113

Kaplan D, Pruim R (2020) Ggformula: formula interface to the grammar of graphics. https://CRAN.R-project.org/package=ggformula

Kassambara A (2020) Ggpubr: 'ggplot2' based publication ready plots. https://CRAN.R-project.org/package=ggpubr

Kokubun N, Kim JH, Shin HC, Naito Y, Takahashi A (2011) Penguin head movement detected using small accelerometers: a proxy of prey encounter rate. J Exp Biol 214:3760−3767

Kuhn M (2020) Caret: classification and regression training. https://CRAN.R-project.org/package=caret

Madsen PT, Johnson M, Aguilar de Soto N, Zimmer WMX, Tyack P (2005) Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*). J Exp Biol 208:181−194

Microsoft Corporation & Weston S (2020) DoSNOW: foreach parallel adaptor for the 'snow' package. https://CRAN.R-project.org/package=doSNOW

Miller PJO, Johnson MP, Tyack PL (2004) Sperm whale behaviour indicates the use of echolocation click buzzes 'creaks' in prey capture. Proc R Soc B 271:2239−2247

Miller PJO, Kvadsheim PH, Lam FPA, Tyack PL and others (2015) First indications that northern bottlenose whales are sensitive to behavioural disturbance from anthropogenic noise. R Soc Open Sci 2:140484

Owen K, Jenner CS, Jenner MNM, Andrews RD (2016) A week in the life of a pygmy blue whale: migratory dive depth overlaps with large vessel drafts. Anim Biotelem 4:17

Pedersen TL (2020) Patchwork: The composer of plots. https://CRAN.R-project.org/package=patchwork

Quick NJ, Cioffi WR, Shearer J, Read AJ (2019) Mind the gap — optimizing satellite tag settings for time series analysis of foraging dives in Cuvier's beaked whales (*Ziphius cavirostris*). Anim Biotelem 7:5

Quick NJ, Cioffi WR, Shearer JM, Fahlman A, Read AJ (2020) Extreme diving in mammals: first estimates of behavioural aerobic dive limits in Cuvier's beaked whales. J Exp Biol 223:jeb222109

Ram K, Wickham H (2018) Wesanderson: a Wes Anderson palette generator. https://github.com/karthik/wesanderson

Schorr GS, Falcone EA, Moretti DJ, Andrews RD (2014) First long-term behavioral records from Cuvier's beaked whales (*Ziphius cavirostris*) reveal record-breaking dives. PLOS ONE 9:e92633

Shearer JM, Quick NJ, Cioffi WR, Baird RW and others (2019) Diving behaviour of Cuvier's beaked whales (*Ziphius cavirostris*) off Cape Hatteras, North Carolina. R Soc Open Sci 6:181728

Stimpert AK, DeRuiter SL, Southall BL, Moretti DJ and others (2014) Acoustic and foraging behavior of a Baird's beaked whale, *Berardius bairdii*, exposed to simulated sonar. Sci Rep 4:7031

Sweeney DA, DeRuiter SL, McNamara-Oh YJ, Marques TA, Arranz P, Calambokidis J (2019) Automated peak detection method for behavioral event identification: detecting *Balaenoptera musculus* and *Grampus griseus* feeding attempts. Anim Biotelem 7:7

Tierney L, Rossini AJ, Li N, Sevcikova H (2018) Snow: simple network of workstations. https://CRAN.R-project.org/package=snow

Tyack PL, Johnson M, Soto NA, Sturlese A, Madsen PT (2006) Extreme diving of beaked whales. J Exp Biol 209: 4238–4253

West KL, Walker WA, Baird RW, Mead JG, Collins PW (2017) Diet of Cuvier's beaked whales *Ziphius cavirostris*

from the North Pacific and a comparison with their diet world-wide. Mar Ecol Prog Ser 574:227–242

Wickham H (2016) Ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, NY

Ydesen KS, Wisniewska DM, Hansen JD, Beedholm K, Johnson M, Madsen PT (2014) What a jerk: prey engulfment revealed by high-rate, super-cranial accelerometry on a harbour seal (*Phoca vitulina*). J Exp Biol 217: 2239–2243

Zimmer WMX, Johnson MP, Madsen PT, Tyack PL (2005) Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*). J Acoust Soc Am 117: 3919–3927