



Continuous-Time Discrete-State Modeling for Deep Whale Dives

Joshua HEWITT[✉], Robert S. SCHICK, and Alan E. GELFAND

Understanding unexposed/baseline behavior of marine mammals is required to assess the effects of increasing levels of anthropogenic noise exposure in the marine environment. However, quantifying variation in the baseline behavior of whales is challenging due to the fact that they spend much of their time at depth, and therefore, their diving behavior is not directly observable. Data collection employs tags as measurement devices to record vertical movement. We focus here on satellite tags, which have the advantage of collection over a time window of weeks. The type of data we analyze here suffers the disadvantage of being in the form of depths attached to an arbitrarily created set of depth bins and being sparse in time. We provide a multi-stage generative model for deep dives using a continuous-time discrete-space Markov chain. Then, we build a likelihood, incorporating dive-specific random effects, in order to fit this model to a set of satellite tag records, each consisting of a temporally misaligned collection of deep dives with sparse binned depths for each dive. Through simulation, we demonstrate the ability to recover true model parameters. With real satellite tag records, we validate the model out of sample and also provide inference regarding stage behavior, inter-tag record behavior, dive duration, and maximum dive depth.

Supplementary materials accompanying this paper appear online.

Key Words: Hierarchical model; Markov chain Monte Carlo; Markov process; Misalignment; Model validation; Satellite tags.

1. INTRODUCTION

Marine mammals are difficult animals to study due to the proportion of time they spend underwater. Beaked whales, in particular, are cryptic since they only spend a small portion of their time at or near the surface. They are deep-diving creatures that can dive to extreme depths for long periods of time (Tyack et al. 2006; Shearer et al. 2019). These deep dives are presumed to have a foraging purpose (Johnson et al. 2006), yet little is known about the underlying behavioral and physiological processes of these dives. To better study these

J. Hewitt (✉) · A. E. Gelfand, Department of Statistical Science, Duke University, Durham, NC 27708-0251, USA (E-mail: joshua.hewitt@duke.edu). R.S. Schick, Nicholas School of the Environment, Duke University, Durham, NC, USA.

© 2021 International Biometric Society

Journal of Agricultural, Biological, and Environmental Statistics, Volume 26, Number 2, Pages 180–199
<https://doi.org/10.1007/s13253-020-00422-2>

dives, scientists have used *on-animal* tags that have revolutionized the collection of data from marine mammals.

Various bio-telemetry devices are used to learn about the physiology of diving behavior. The basic principle is to attach a device onto the back of an animal in order to measure and record features of this behavior. Different devices have been employed, each with their own costs and benefits. Satellite tags are useful for studying movement of the animal in the x - y plane and the diving behavior in the z dimension over long periods of time, e.g., weeks to months. However, these tags, when deployed on the animals, have limited transmission bandwidth, so the data are typically downgraded, summarized, or discretized prior to transmission. This results in data that have coarse temporal resolution (Quick et al. 2019). In particular, we obtain depth data recorded every 5 min and summarized in discrete depth bins (Wildlife Computers Manual - <https://static.wildlifecomputers.com/SPLASH10-TDR10-User-Guide.pdf>, last accessed 8 January, 2020). Following data collection, the challenge from an applied ecology standpoint is to use these data to make inference on the underlying dive processes and to estimate the endogenous and exogenous covariates that drive state persistence and transition.

The need to develop such process-driven behavioral models that can explain these coarse dive data motivates our study. Here we propose a model to infer about the latent diving behavior of beaked whales using such tags. Our model is a continuous-time discrete-space model (CTDS), which is inspired by, but quite different from, modeling work developed and applied to movements of terrestrial (Hanks et al. 2015) and marine mammals (Wilson et al. 2018).

Because an observation in the satellite tag dive data is collected every five minutes and only in the form of a depth bin, inference from such data is inherently limited. So, what is the scientific understanding we can obtain from such data? We are able to learn about baseline behavior with regard to times in three different dive stages; we are able to learn about velocities as well as directional behavior in each of these stages. We are able to learn about the distribution of dive duration and maximum dive depth. Since we consider a collection of satellite tags, we are able to learn about both mean behavior and inter-animal behavioral variability. We also include the sex of the animal as a predictor to see if there are sex effects with regard to the foregoing inference. Finally, all of this inference is with regard to baseline behavior, i.e., in the absence of exposure to anthropogenic noise sources. In follow-on work, inference will be brought to the important environmental issue of assessing the effects on behavior from exposure to anthropogenic sound.

The novelty of our contribution is elaborated in the ensuing paragraphs. Our overall intent is to propose a generative model for deep whale dives where the depth axis has been partitioned into *dive bins*, creating a discrete state space for dive depths. We model time continuously so that there is a discrete state associated with every time during a dive. We characterize a deep dive into three well-identified stages (each of which is associated with a distinct dive phase)—descent, foraging, and ascent—with a separate model specification for each. We acknowledge that whale behavior during deep dives is complex, arguably more complex than a three stage specification. However, with the sparse data that are collected using satellite tags, it is not possible to extract more than three recognizable deep dive stages. Moreover, we are modeling at the level of changes in depth bins, as fine as possible with

our binned data. So our model captures the entire discrete-space behavior for a dive while recognizing that such behavior will depend upon the three evident stages. In different words, our stages are not latent, not hidden. It would not be possible to extend the parametrization for movement within a dive to additional stages (i.e., to model sub-types of the descent, foraging, and ascent stages) without explicitly specifying what stage the dive is in at all observed time points prior to analysis.

We incorporate random times in stages, yielding a semi-Markov process according to stage. That is, we adopt a homogeneous/stationary continuous-time Markov (jump) process specification within each stage (Anderson 2012). Furthermore, we employ a set of satellite tag records that are assumed to be exchangeable, introducing tag-specific random effects for modeling inter-animal variation. The result is a demanding hierarchical model which is elaborated in Sect. 3.

Our generative model provides start and end times in the surface bin (i.e., the depth bin that contains zero depth). A complete realization of a dive is a sequence consisting of a depth bin along with a duration in the bin, then a transition to a new bin followed by a duration in that bin, etc., until the surface bin is returned to. However, there are misalignment issues between the model we propose and the data that are collected, as we detail in Sect. 3. Figure 1 shows an example of the data collected, which reveal an incomplete record of a dive; we only observe the dive depth at sparse time points. Furthermore, we do not observe the exact start time of the dive; we only observe a zero depth bin at a time which provides a neighborhood for the start of the dive. Similarly, we do not observe the exact end time of a dive, and we also do not observe the exact duration of the dive or the exact times when a bin was entered or when it was left. With coarse resolution for the observation times, the effect of all of this misalignment can be consequential.

Misalignment, in some fashion, is at the root of all continuous-time discrete-space (CTDS) model fitting. If we observed the process at the transition times, we could write down the likelihood sequentially in terms of the stochastic specification for the process. However, the likelihood under the model for the data we collect is more challenging since it needs to account for all imaginable trajectories that pass through the entire observed set of time-depth data. Fortunately, our proposed continuous-time Markov process, with its associated infinitesimal generator, using matrix exponential forms, allows us to write down the parametric likelihood, thus enabling model fitting.

Hanks et al. (2015), working in the x - y plane, have provided seminal work in CTDS model fitting. In particular, they assign observations in continuous 2-dimensional space to a set of raster cells to create their discrete space. Then, their model provides parameter estimates for the covariates that drive both residence time in specific raster cells and transitions to neighboring cells. Though the data analyzed by Hanks et al. (2015) are richer in time than the dive data we analyze, they still have only partially observed movement tracks. To create full realizations of the actual movement paths, they offer an imputation-based solution through realizations of an Ornstein–Uhlenbeck (O-U) process model (Johnson et al. 2008). Recently, Scharf et al. (2017) discuss, more generally, approximate imputation distributions (AIDs) as a broad approach to fitting animal movement models in either continuous or discrete spaces.

This reveals an important distinction between imputation-based fitting and ours. Hanks et al. (2015) observe locations in a continuous 2-dimensional space, which they can utilize in their O-U model imputation. They use this to generate a catalogue of imputed trajectories to do model fitting and to develop their inference. By comparison, we only have binned data, so we only imagine a binned generative model. As a result, we work with the infinitesimal generator mentioned above, using matrix exponential forms, to develop a likelihood for binned realizations. We do not generate imputed trajectories for model fitting; we avoid the use of an AID.

We apply our model to discrete depth bin data, in fact to both simulated data and to the satellite tag diving records of seven different Cuvier's beaked whales (*Ziphius cavirostris*) tagged off Cape Hatteras, NC (Southall et al. 2018). Our model has components to capture times in stages, velocity, and directional parameters within each stage along with a potential sex effect in these parameters. Further, we borrow strength through random effects across the satellite tag replicates to better learn about these components and assess inter-individual variability. We can also learn about total dive duration and maximum depth.

We model at the level of the dive, with each dive specific to its tag, and hence, to its individual. We introduce dive-specific random effects to account for the substantial variability across dives within a satellite tag record, as well as to capture the misalignment between the observed data for a specific dive and the actual start and end times of the dive. These random effects are in addition to the tag record random effects. We fit our hierarchical model within a Bayesian framework, obtaining posterior samples of parameters that enable posterior distributions for all parameters and all model components. Combined with our adjustment for misalignment, we obtain posterior predictive distributions for the total duration and maximum depth of potential dives.

We illustrate the performance of our approach using a single simulated tag record to provide some sensitivity analysis of inference to the data sparsity in terms of the length of the interval between observation times. Then, with real data, we obtain the foregoing inference and also investigate out-of-sample model adequacy.

The format of the paper is as follows. In Sect. 2, we offer a descriptive summary of the satellite tag records we analyze. In Sect. 3, we present the generative continuous-time Markov chain model we employ as well as our approach to fitting it to the misaligned observed data replicates. In Sect. 4, we present the results of a brief simulation study. In Sect. 5, we present the results of fitting our model to the real data. We offer a summary and potential future work in Sect. 6.

2. THE DATA

We model data from seven Cuvier's beaked whales (*Ziphius cavirostris*), which were each tagged once, offshore of Cape Hatteras, NC (Supplement, Table 1). Animals were tagged in 2019 using a SPLASH-10-292 tag from Wildlife Computers, Inc. (Redmond, WA). The tag was programmed to record depth data every 5 min for a period of 2 weeks (Fig. 1). Though the depth sensor on the tag is accurate to ± 1 m, the depth data are discretized into coarse bins prior to transmission because bandwidth between the tag and the satellite is limited.

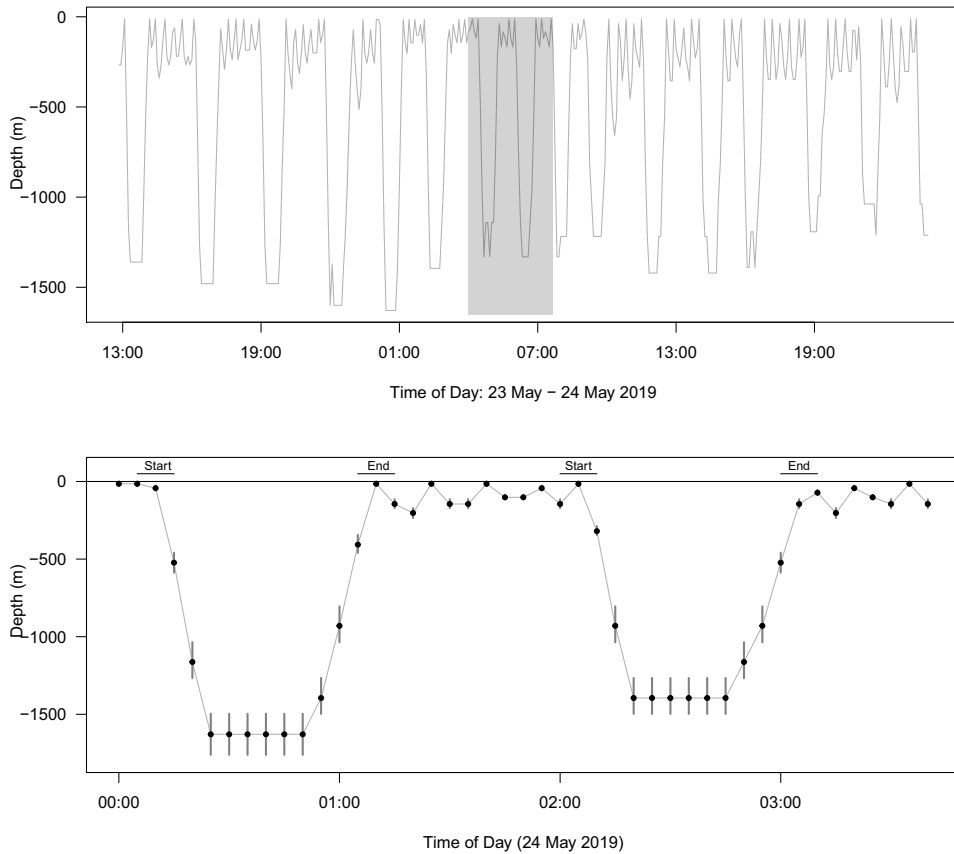


Figure 1. Top panel shows two days of diving behavior from the 14-day dive record for ZcTag084, tagged off Cape Hatteras, NC in May 2019. Gray shading denotes two dives depicted in bottom panel. Black dots show the depth bin midpoints, and the vertical bars show the bin widths. Putative start and end times of the dive are shown in the margin. Light gray line shows linear interpolation of the dive between observations.

For example, for a whale actually recorded at 1399 m in the water column, the observation is reported as the whale having been in a depth bin centered at 1422 m, with a width of 266 m (Fig. 1).

Satellite tag data do not allow dive duration and maximum depth to be studied directly, but the durations and maximum depths observed in the raw data—collected at 5-minute intervals—still broadly characterize deep dives. Mean observed dive duration across all 7 whales was 62 min. Mean maximum observed depth across these animals was 1213 m. While each tag recorded depths for 2 weeks, some animals' dives were recorded following a controlled exposure. We eliminated all post-exposure dives from the analysis, hence the smaller number of dives for certain whales (Supplement, Table 1).

The satellite tags discretize the ocean's water column into 16 depth bins, with the bin widths increasing with depth. SPLASH-10-292 tags have been engineered to use the maximum depth within a sliding time window to adaptively define depth bins. The adaptation minimizes discretization error, and the engineering decision allows the tags to be used on a wide range of animals that span a wide range of diving behaviors. In the data we ana-

lyze, a sliding time window updates depth bins every four hours. The shallowest bins are typically 40 m wide—e.g., 0–40 m, 40–80 m, and the deepest bins are typically over 350 m wide—e.g., 1642–2011 m and 2011–2380 m. The depth bins are fairly consistent across time because Cuvier’s beaked whales exhibit extremely regular diving behavior; accordingly, we map observations to a single, “canonical” set of depth bins as a data pre-processing step. We select the set of 16 depth bins with the greatest maximum depth as our canonical depth bins, and we assign each observation to the canonical depth bin that provides the greatest overlap with the observed depth bin.

3. MODEL SPECIFICATION AND INFERENCE

In this section, first we formally specify a generative model for deep whale dives (Sect. 3.1) and also present some illustrative dives under this model. Then, in Sect. 3.2, we develop the likelihood for the actual recorded observations associated with a given deep dive. Adopting a Bayesian framework for inference, in Sect. 3.3 we turn to model fitting employing a collection of deep dives from a set of satellite tag records. The records are assumed to be exchangeable, and the dives within a tag record are assumed to arise under this model but are only partially observed, i.e., at sparse but regular times. We address the misalignment issues described in the Introduction: unknown start time and end time for the dive, and unknown times in stages for the dive. Finally, in Sect. 3.4, we turn to posterior inference for model parameters along with posterior predictive dive simulation and its use for model assessment.

3.1. MODEL SPECIFICATION

We start with a model for an actual (but unobservable) dive realization. For deep whale dives, we propose a generative, continuous-time discrete-space (CTDS) model with latent stages. CTDS models for terrestrial animal movement are discussed briefly in Introduction. They are continuous-time Markov chains (CTMCs): stochastic processes defined over countable state spaces (Anderson 2012). Here, we define the states for our deep dive model as a fixed set of depth bins that satellite tags use to partition ocean depths. Conceptually, breakpoints $D_0 < D_1 < \dots < D_M$ define M depth bins, where $D_0 = 0$ is the ocean surface and D_M is the maximum depth we model. We use the integers $1, \dots, M$ to label the depth bins, calling $[D_{\ell-1}, D_\ell]$ the ℓ th depth bin and denoting its width by $d_\ell \equiv D_\ell - D_{\ell-1}$. Whales move between adjacent depth bins during a dive. The function $\ell_{ij}(t)$ labels which of the M depth bins whale $i \in \{1, \dots, n\}$ is in at time $t > 0$ from the start of the whale’s j th dive.

Our model uses latent dive stages to describe movement across three distinct dive phases. We assume deep dives begin in stage 1, a descent stage that starts at the surface, then transition to stage 2, a foraging or bottom stage, and finish in stage 3, an ascent stage that ends at the surface. The function $s_{ij}(t)$ indicates which of the stages, $\{1, 2, 3\}$, whale i is in at time $t > 0$ from the start of the whale’s j th dive. The stage function $s_{ij}(t)$ begins in the descent stage, i.e., $s_{ij}(0) = 1$, and switches when the descent stage ends at time denoted by $T_{ij}^{(1)} > 0$. The

stage function switches again when the foraging stage ends at time denoted by $T_{ij}^{(2)} > T_{ij}^{(1)}$. Letting $T_{ij}^{(3)}$ denote the time the ascent stage (and dive) ends, the duration of the dive is, in fact, $T_{ij}^{(3)}$. Since we do not know the exact start and end times of dives, we don't observe the true duration $T_{ij}^{(3)}$. We also do not know $T_{ij}^{(1)}$ or $T_{ij}^{(2)}$, but we anticipate them to be dependent. In the next subsection, we accommodate these unknowns in the development of the likelihood under the observed data for a given dive.

We model the stage progressions through a semi-Markov process, which generates deep dive realizations that spend a realistic amount of time in each stage. Let $\xi_{ij}^{(1)}$ denote the time in stage 1—the descent stage—for the i th satellite tag record's j th deep dive. Similarly, let $\xi_{ij}^{(2)}$ denote the time in stage 2—the foraging stage. The associated stage transition times are $T_{ij}^{(1)} = \xi_{ij}^{(1)}$ and $T_{ij}^{(2)} = T_{ij}^{(1)} + \xi_{ij}^{(2)}$. We use the bivariate distribution G_i , with positive support, to jointly model $(\xi_{ij}^{(1)}, \xi_{ij}^{(2)})$ for each of the i th whale's dives. We offer a partly empirical approach for prior specification of G_i in the next subsection. The time spent ascending $\xi_{ij}^{(3)}$ is determined as the time until arrival in the surface bin. It is completely governed by the ascent stage specification, which (as with the other stages) provides speed and movement between depth bins.

For a given stage within a given tag record, we assume a homogeneous CTMC characterizes movement between depth bins. We specify these within-stage CTMCs via their distribution for holding time in a bin and their distribution for depth bin transition. Suppressing the tag index i and the dive index j for all of the parameters for the remainder of this subsection, the generic holding time $\Delta > 0$ is the duration of time a whale remains in depth bin $\ell \in \{1, \dots, M\}$ before transitioning to an adjacent depth bin $\ell' \equiv \ell \pm 1$ such that $\ell' \in \{1, \dots, M\}$. For a homogeneous CTMC, starting at say time t , Δ is exponentially distributed with a rate that only depends on ℓ and model parameters—the distribution does not depend on time t or on previous transitions. Similarly, the transition distributions only assign probabilities to ℓ' based on ℓ and model parameters. This construction is typically expressed as the distribution of the future given the depth at the present time, $\ell(t)$, and again, it does not depend on t . Because of our three stage specification, the distribution of Δ and $\ell(t)$ across an entire dive will depend upon the stage that time t is in. We use empirically developed characteristics of deep dives to motivate our specifications for the within-stage holding time and depth bin transition distributions.

The distribution for the holding time Δ in bin ℓ is parameterized with respect to positive stage-dependent speed parameters, $\lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)}$, respectively, and the bin width d_ℓ :

$$\Delta | \ell, s \sim \text{Exp} \left(\lambda^{(s)} / d_\ell \right), \quad (1)$$

where s is the dive stage. That is, for any duration that begins in stage s and is in bin ℓ , the expected duration under (1) is $E[\Delta | \ell, s] = d_\ell / \lambda^{(s)}$, proportional to the bin width d_ℓ . Such proportionality implies that, for fixed speed parameters, wider depth bins tend to take more time to transition through than narrower depth bins.

The depth bin transition distribution for ℓ is parameterized with respect to stage-dependent directional preference parameters, $\pi^{(1)}, \pi^{(2)}, \pi^{(3)} \in [0, 1]$. Transitions can only be made to an adjacent bin ℓ' , as defined above, which ensures that dives follow a contigu-

ous path through the discretized water column. If a whale is in one of the boundary depth bins, i.e., $\ell \in \{1, M\}$, then the transition distribution places all of its mass on $\ell' = 2$ or $\ell' = M - 1$, respectively. Additionally, transitions from $\ell = 2$ to $\ell' = 1$ are only allowed in stage 3, and this transition always ends a dive. In all other cases, the probability the whale transitions from ℓ to the next deepest bin $\ell' = \ell + 1$ is modeled via

$$P(\ell' = \ell + 1 | \ell, s) = \pi^{(s)}, \quad (2)$$

where, again, s is the dive stage. So, $1 - \pi^{(s)}$ defines the probability for transitioning instead to the next shallowest bin $\ell' = \ell - 1$. To be clear, $\pi^{(s)}$ applies to a transition that occurs at any time t in stage s .

We treat the directional preference parameters $\pi^{(1)}$ and $\pi^{(3)}$ as unknown, but we assume $\pi^{(2)} = 0.5$. We expect $\pi^{(1)} > 0.5$ so the transition model (2) favors downward movement in the descent stage. Similarly, we expect $\pi^{(3)} < 0.5$ so the transition model (2) favors upward movement in the ascent stage. We set $\pi^{(2)} = 0.5$ in order not to introduce any directional bias during the foraging stage.

As a result, provided with a start time, stage transition times, and model parameters, we can straightforwardly generate dive realizations in a sequential fashion. Formally, a dive realization assigns a depth bin and stage to all times $t > 0$ via $\tilde{\mathcal{Y}}(t) = (s(t), \ell(t))$. The likelihood for the random function $\tilde{\mathcal{Y}}(t)$ depends on the stage transition times of the dive, $T^{(1)}$ and $T^{(2)}$, and the exact sequence of its depth bin transitions and times. Let a dive consist of R total depth bin transitions, which occur at times $\tau_0 < \tau_1 < \dots < \tau_R$. As described above, a dive begins at time $\tau_0 = 0$ in the diving stage at the surface bin, i.e., $\tilde{\mathcal{Y}}(\tau_0) = (1, 1)$. Then, the dive transitions at time $\tau_1 > 0$ to the next bin $\ell(\tau_1)$ after the holding time $\Delta_0 = \tau_1$. The dive ends when the whale returns to the surface bin, and the r th holding time is the duration $\Delta_r \equiv \tau_{r+1} - \tau_r$.

Figure 2 shows realizations of dives under four different sets of parameter values for $\lambda^{(s)}$, for $s = 1, 2, 3$, and $\pi^{(1)}$ and $\pi^{(3)}$. These parameter values resemble ranges that have been seen in real deep dives associated with the given set of bins. Dive (A) is similar to dives seen in data, with parameters $\theta^{(1)} = (0.99, 1.25)$, $\theta^{(2)} = (0.5, 0.6)$, and $\theta^{(3)} = (0.05, 1)$, where $\theta^{(s)} = (\pi^{(s)}, \lambda^{(s)})$. Dive (B) has more variability than Dive (A) because it is simulated with faster transition rates $\lambda^{(1)} = 2.5$, $\lambda^{(2)} = 1.2$, and $\lambda^{(3)} = 2$. Dive (C) is longer than Dive (A) because it is simulated with less extreme ascent and descent directional preference parameters $\pi^{(1)} = 0.8$ and $\pi^{(3)} = 0.2$. Finally, dive (D) has the most variability because it is simulated with both faster transition rates and less extreme directional preference in the parameters $\theta^{(1)} = (0.8, 2.5)$, $\theta^{(2)} = (0.5, 1.2)$, and $\theta^{(3)} = (0.2, 2)$. The simulations (A), (B), and (D) have $T^{(1)} = 10$ min and $T^{(2)} = 25$ min, while simulation (C) has $T^{(1)} = 20$ min and $T^{(2)} = 35$ min.

3.2. DATA AUGMENTATION LIKELIHOOD FOR OBSERVED DIVES

Again suppressing replicate and dive index, formally, for an individual dive, a realization under the generative specification for each stage arises sequentially. Extending the notation of the previous subsection, we denote the dive segment in stage s by $\tilde{\mathcal{Y}}^{(s)} \equiv \{(\ell(\tau_r^{(s)}), \Delta_r^{(s)})\}$,

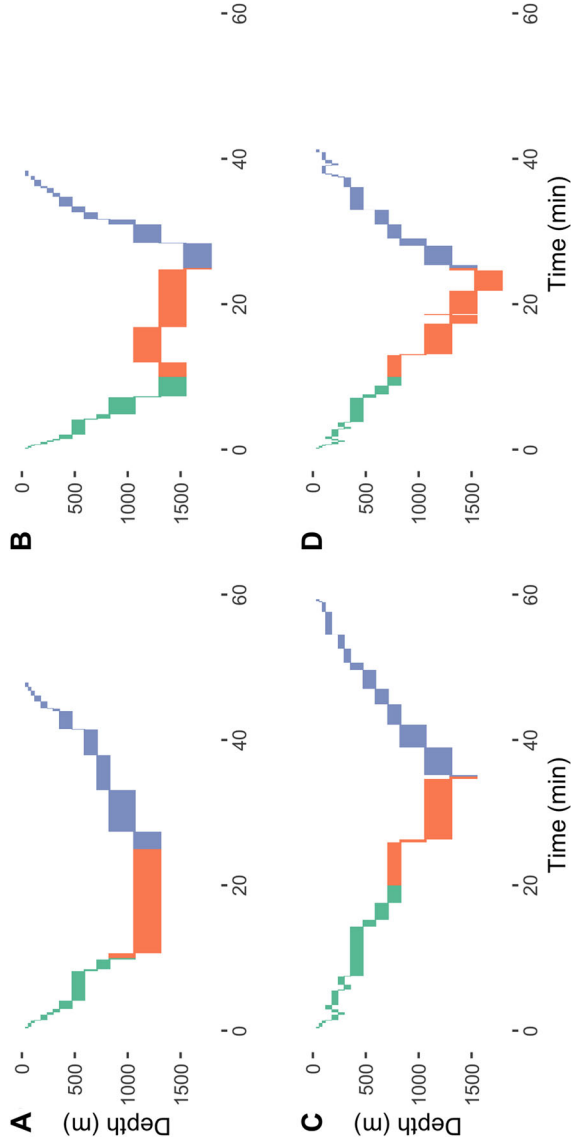


Figure 2. Simulated dives to illustrate the impact of model parameters. Each dive's descent stage is colored in green (■); the foraging stages are colored in orange (■), and the ascent stages are colored in purple (■). Each dive is simulated from the same random seed.

for $r = 0, 1, 2, \dots, R^{(s)}$. That is, stage s begins in bin $\ell(\tau_0^{(s)})$ immediately when the previous stage ends $\tau_0^{(s)} = T^{(s-1)}$, and we draw a $\Delta_0^{(s)} \sim [\Delta_0^{(s)} | \ell(\tau_0^{(s)})]$. Then, we are at $\tau_1^{(s)} = \tau_0^{(s)} + \Delta_0^{(s)}$ and we draw $\ell(\tau_1^{(s)})$ given $\ell(\tau_0^{(s)})$ followed by $\Delta_1^{(s)}$ given by $\ell(\tau_1^{(s)})$, etc. The stage- s segment $\tilde{\mathcal{Y}}^{(s)}$ ends at time $T^{(s)}$, and its final depth bin is used to initialize the next segment $\tilde{\mathcal{Y}}^{(s+1)}$.

A critical point emphasized in the Introduction is that we provide a continuous-time, discrete bin model for a dive realization, $\tilde{\mathcal{Y}} = (\tilde{\mathcal{Y}}^{(1)}, \tilde{\mathcal{Y}}^{(2)}, \tilde{\mathcal{Y}}^{(3)})$, but, we never observe $\tilde{\mathcal{Y}}$. With remote telemetry devices, we observe a sequence of times $t_1 < t_2 < \dots < t_m$ with associated labels $\{\ell(t_j) : j = 1, 2, \dots, m\}$, which we denote by $\mathcal{Y} = \{(t_k, \ell(t_k)) : k = 1, \dots, m\}$.

In general, with telemetry data, the times t_1, \dots, t_m are offset from the start of the dive. So, we cannot directly associate a likelihood with \mathcal{Y} without first *synchronizing* the observation times to corresponding true dive times. We let ε quantify the offset and introduce the dive-aligned observation times $t_k^* = \max(t_k - \varepsilon, 0)$. The offset allows us to develop an augmented data likelihood (Van Dyk and Meng 2001). If the offset ε were known, then telemetry devices would yield an aligned observation $\mathcal{Y}^* = \{(t_k^*, \ell(t_k^*)) : k = 1, \dots, m\}$.

We use a scaled and shifted Beta distribution as the prior for the ε parameter associated with each modeled dive; that is, we have a dive-specific random offset. Suppose the telemetry device records depths every δ seconds. If $\ell(t_1) = 1$, then the true start time for the dive must fall within $[t_1 - \delta, t_1 + \delta]$, i.e., the dive may have started before or after t_1 , so the appropriate support for ε is $[-\delta, \delta]$. More precisely, the definition of t_1^* ensures the dive started after $t_1 - \delta$. If $\ell(t_1) > 1$, then the true start time for the dive must fall within $[t_1 - \delta, t_1]$, so $\varepsilon < 0$, and the prior should be updated accordingly. Through the same observation process, the end of each dive is also only observed with some uncertainty. So, we introduce and define the end dive offset as ψ and, similarly, assign support for ψ to also be $[-\delta, \delta]$ for dives where we observe the final surface state, i.e., $\ell(t_R) = 1$.

We condition on the unknown stage transition times $T^{(1)}$ and $T^{(2)}$ in addition to the offsets ε and ψ to derive an augmented data likelihood for the observed dive record \mathcal{Y} . Let $\Theta = \{\theta^{(1)}, \theta^{(2)}, \theta^{(3)}\}$ be the collection of stage-dependent parameters $\theta^{(s)} = (\pi^{(s)}, \lambda^{(s)})$. CTMC theory enables us to write down the density (distribution) using bracket notation, $[\mathcal{Y}^* | \Theta, T^{(1)}, T^{(2)}, \varepsilon, \psi]$, hence the likelihood, $L(\Theta, T^{(1)}, T^{(2)}, \varepsilon, \psi; \mathcal{Y})$. With the priors, $[(T^{(1)}, T^{(2)}) | G], [\Theta], [\varepsilon]$ and $[\psi]$, we have a full, augmented model specification.

In order to provide $[\mathcal{Y}^* | \Theta, T^{(1)}, T^{(2)}, \varepsilon, \psi]$, we digress briefly to remind the reader of some properties we need under a first-order continuous-time homogeneous Markov process with a finite number of discrete states. Formally, such a process is defined as a stochastic process with right-continuous, piecewise constant paths. In our notation, for any pair of times t and t' and any pair of states ℓ and ℓ' , the process is defined through a stationary transition function $P(t' - t)_{\ell, \ell'}$, which governs the evolution of the process via

$$P(\ell(t') = \ell' | \ell(t) = \ell, \{\ell(u), u < t\}) \equiv P(t' - t)_{\ell, \ell'}. \quad (3)$$

CTMC distribution theory combines the holding time and depth bin transition distributions (1) and (2) to induce the transition function $P(t' - t)_{\ell, \ell'}$ from an infinitesimal generator matrix. The infinitesimal generator matrix $A^{(s)} \in \mathbb{R}^{M \times M}$ for stage s combines the stage

parameters such that, for $\ell \in \{2, 3, \dots, M-1\}$,

$$A_{\ell, \ell'}^{(s)} = \begin{cases} -\lambda^{(s)}/d_\ell & \text{if } \ell' = \ell \\ \pi^{(s)}\lambda^{(s)}/d_\ell & \text{if } \ell' = \ell + 1 \\ (1 - \pi^{(s)})\lambda^{(s)}/d_\ell & \text{if } \ell' = \ell - 1 \\ 0 & \text{otherwise} \end{cases}.$$

As for $\ell = 1$, for stage 1, $A_{11}^{(1)} = -\lambda^{(1)}/d_1$ and $A_{12}^{(1)} = \lambda^{(1)}/d_1$. For stages 2 and 3, there is no probability of transitioning into or out of $\ell = 1$, respectively. As for $\ell = M$, for all stages, $A_{MM}^{(1)} = -\lambda^{(s)}/d_M$ and $A_{M, M-1}^{(s)} = \lambda^{(s)}/d_M$.

For arbitrary depth bins $\ell(t), \ell(t') \in \{1, \dots, M\}$ and times $0 < t < t'$, the within-stage transition function $P^{(s)}(t' - t)_{\ell, \ell'}$ defines the within-stage transition probability via $P(\ell(t') = \ell' | \ell(t) = \ell, s) = P^{(s)}(t' - t)_{\ell, \ell'}$, where $P^{(s)}(t' - t)_{\ell, \ell'}$ is the (ℓ, ℓ') entry of the exponential transition matrix,

$$P^{(s)}(t' - t) = e^{(t' - t)A^{(s)}}. \quad (4)$$

The exponential transition matrix is the convergent series $e^{(t' - t)A^{(s)}} = \sum_{k=0}^{\infty} ((t' - t)A^{(s)})^k / k!$, which can be computed up to desired numerical precision via numerical linear algebra methods (Al-Mohy and Higham 2010).

As a result, we have the full conditional distribution $[\mathcal{Y}^* | \Theta, T^{(1)}, T^{(2)}, \varepsilon, \psi]$ because conditioning on $T^{(1)}$ and $T^{(2)}$ is equivalent to conditioning on the stage function $s(t)$ at all times $t > 0$. The Markov property for CTMCs simplifies the full conditional via

$$\begin{aligned} & [\mathcal{Y}^* | \Theta, T^{(1)}, T^{(2)}, \varepsilon, \psi] \\ &= \prod_{k=1}^m [\ell(t_k^*) | \ell(t_{k-1}^*), \dots, \ell(t_1^*), \Theta, T^{(1)}, T^{(2)}, \varepsilon, \psi] \\ &= \prod_{k=1}^m [\ell(t_k^*) | \ell(t_{k-1}^*), \Theta, T^{(1)}, T^{(2)}, \varepsilon, \psi]. \end{aligned} \quad (5)$$

From above, the transition probability associated with $\ell(t_{k-1}^*)$ to $\ell(t_k^*)$, i.e., $[\ell(t_k^*) | \ell(t_{k-1}^*), \Theta, T^{(1)}, T^{(2)}, \varepsilon, \psi]$ can be written down explicitly using (4), but the transition probability depends on where t_{k-1}^* and t_k^* fall relative to $T^{(1)}$ and $T^{(2)}$. Since the transition matrix $P^{(s)}(t' - t)$ allows $P^{(s)}(t' - t)_{k, k} > 0$, this probability can be a product of terms.

3.3. MODEL FITTING

We fit the CTDS dive model to a set of n satellite tags, with tag i providing n_i dives, i.e., our data are $\mathbf{Y} = (\mathcal{Y}^1, \dots, \mathcal{Y}^n)$ with \mathcal{Y}^i consisting of the partially observed deep dives $\{\mathcal{Y}_j^i : j = 1, 2, \dots, n_i\}$. We treat the dives as conditionally independent given parameters Θ_i for tag i and the dive-specific random effects (start time, end time, and times in stages). That is, we assume that the deep dives for animal i share a common Θ_i because the dives tend to have similar descent/ascent rates, duration, and maximum depth, but we allow Θ_i to vary across i .

The tag-specific parameters $\Theta_i = (\theta_i^{(1)}, \theta_i^{(2)}, \theta_i^{(3)})$, where $\theta_i^{(s)}$ is comprised of parameters for vertical speeds and for directional preferences according to stage are modeled as random effects. Specifically, for vertical speeds, we take $\log \lambda_i^{(s)} = \alpha_{i,0}^{(s)} + \alpha_1^{(s)} \text{sex}_i$ with $\text{sex} = 0$ if adult male, $=1$ otherwise. So, we allow sex to affect speed in each stage. In addition, we introduce tag-specific random effects within stage through the prior distribution $\alpha_{i,0}^{(s)} \sim N(\alpha_0^{(s)}, \sigma_\alpha^{2(s)})$ with $\alpha_0^{(s)} \sim N(\mu_\alpha^{(s)}, 10^4)$. The random effects specification introduces customary hierarchical centering to improve Gibbs sampling behavior (Gelfand et al. 1996). However, we specify common directional preferences for all animals, i.e., $\pi_i^{(s)} = \pi^{(s)}$, because, for the present application, the descent and ascent parameters required to capture deep diving behavior are so extreme (i.e., close to 1 and 0, respectively) that there is no consequential between-animal variation to model (as well as no sex effect). We use Beta distributions as priors for the directional persistence parameters $\pi^{(1)} \sim \text{Beta}(5, 2)$ and $\pi^{(3)} \sim \text{Beta}(2, 5)$. The hierarchical centering for speed uses $\mu_\alpha^{(1)} = 0.4$, $\mu_\alpha^{(2)} = -1.2$, and $\mu_\alpha^{(3)} = 0$. The priors center the parameters around suitable a priori speeds and directional preferences reported in a previous study of Cuvier's beaked whales (Tyack et al. 2006, Table 2). All of the variance parameters $\{\sigma_\alpha^{2(s)}\}$ use Inv-Gamma(2, 1) prior distributions.

We employ a partially empirical approach to specify the joint prior distribution for the stage transition times $[T_{ij}^{(1)}, T_{ij}^{(2)} | \hat{G}_i]$. Specifically, we obtain \hat{G}_i using all of the deep dives within a single satellite tag, i.e., \hat{G}_i is specific to the individual wearing that tag. First, we use linear interpolation to approximate complete dive trajectories $\tilde{Y}_1^i, \dots, \tilde{Y}_{n_i}^i$ from observations $\mathcal{Y}_1^i, \dots, \mathcal{Y}_{n_i}^i$. For each dive, the approximate depth at time t is defined as the linear interpolation between the midpoints of the depth bins at the closest observation times $t_{k-1} < t < t_k$; see, for example, Fig. 1. Then, following suggestions in Hooker and Baird (2001), each interpolated dive is segmented into empirical descent, foraging, and ascent stages according to the first and last times at which the whale crosses 85% of each dive's maximum depth.

We use the empirically segmented dives to estimate \hat{G}_i as a bivariate lognormal, parameterized via a two-step moment-matching scheme. We begin with a smoothing step, in which we compute maximum likelihood fits of the empirical descent and foraging durations to independent Gamma distributions. Then, we parameterize the marginal lognormal distributions so that the marginal lognormal moments match the mean and variance of the smoothed, Gamma fits. The empirical correlation between descent and foraging times provides the correlation in the bivariate log-normal distribution. This moment-matching scheme yields a \hat{G}_i which produces better out-of-sample predictive validation than a direct maximum likelihood fit.

We use the augmented likelihood (5) for each tag record, along with the foregoing priors, to construct a Gibbs sampler to obtain samples from the posterior distribution of the parameters and random effects within and across tag records. The sampler requires initial values for the stage-dependent parameters Θ_i and the random effects across all dives $\mathbf{T}_{i1} = (T_{i1}^{(1)}, \dots, T_{in_i}^{(1)})$, $\mathbf{T}_{i2} = (T_{i1}^{(2)}, \dots, T_{in_i}^{(2)})$, $\mathbf{E}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})$, and $\Psi_i = (\psi_{i1}, \dots, \psi_{in_i})$. Most parameters are updated on unconstrained scales via adaptive random walk Metropolis-Hastings steps (Andrieu and Thoms 2008). This includes the transformed descent parameters $\text{logit } \pi^{(s)}$, animal-specific speeds $\log \lambda_i^{(s)}$, stage transition

times $(\log T_{ij}^{(1)}, \log T_{ij}^{(2)})$ as a blocked update, and start and end offsets ε_{ij} and ψ_{ij} . The regression effects $\alpha_0^{(s)}$ and $\alpha_1^{(s)}$ and random effect scales $\sigma_\alpha^{2(s)}$ are updated with conjugate steps.

3.4. MODEL ASSESSMENT

Our model validation approach uses out-of-sample posterior predictive checks (Gelman et al. 2014, Chapter 6). We randomly split the dataset of dive observations into a single testing and a single training partition, creating the partitions such that they both contain the same number of dives. The model is fitted to the training observations, which yields posterior samples for each tag's model parameters Θ_i , stage transition times \mathbf{T}_{i1} and \mathbf{T}_{i2} , and offsets \mathbf{E}_i and Ψ_i . Then, we compare the empirical distribution of the testing observations to the posterior predictive distribution for simulated observations \mathcal{Y} of a new dive $\tilde{\mathcal{Y}}$ where, as noted in Sect. 2, we confine ourselves to dives that are *observed* to exceed 800 m—a deep dive threshold adopted from analysis of earlier data records from this same geographic area (Shearer et al. 2019).

When we attempt model validation, the features of the dives we employ must be associated with the observations actually used to fit the model, i.e., with the data, \mathbf{Y} . That is, we recreate the observation process for the validation. Inference on dive features such as duration and maximum depth is of primary ecological interest, and we pursue this below. This leads us to compare the observed distribution over the depth bins at nominal times 5 min, 10 min, etc., with the corresponding predictive distribution over these depth bins at these same, nominal times. The predictive distributions are described more precisely below but, conceptually, we take a posterior draw of the parameter vector Θ_i to generate a predictive realization $\tilde{\mathcal{Y}}_{\text{new}}$ of a deep dive. Then, we draw an offset ε , which we use to align the predictive realization with the nominal bin times, yielding observations \mathcal{Y}_{new} of depth bins at each nominal observation time. We do this over a sample of posterior predictive replicates to obtain a depth bin distribution for each nominal observation time. For each tag replicate and for each nominal time, by overlaying, we offer informal visual comparison between the posterior predictive distribution and the empirical distribution of depth bins.

The posterior predictive distribution for the partial observation of a new dive $[\mathcal{Y}_{\text{new}}|\mathbf{Y}]$ is obtained through composition sampling the posterior predictive distribution for complete dives $[\tilde{\mathcal{Y}}_{\text{new}}, \varepsilon_{\text{new}}, \psi_{\text{new}}|\mathbf{Y}]$, i.e.,

$$[\mathcal{Y}_{\text{new}}|\mathbf{Y}] = \int [\mathcal{Y}_{\text{new}}|\tilde{\mathcal{Y}}_{\text{new}}, \varepsilon_{\text{new}}, \psi_{\text{new}}] \times [\tilde{\mathcal{Y}}_{\text{new}}, \varepsilon_{\text{new}}, \psi_{\text{new}}|\mathbf{Y}] d\tilde{\mathcal{Y}}_{\text{new}} d\varepsilon_{\text{new}} d\psi_{\text{new}}. \quad (6)$$

We approximate the posterior (6) by Monte Carlo sampling the finite mixture distribution $\frac{1}{B} \sum_{b=1}^B [\mathcal{Y}_{\text{new}}|\tilde{\mathcal{Y}}_{\text{new}}^{(b)}, \varepsilon_{\text{new}}^{(b)}, \psi_{\text{new}}^{(b)}]$, where B is the number of posterior parameter samples. While the predictive dive distribution $[\tilde{\mathcal{Y}}_{\text{new}}, \varepsilon_{\text{new}}, \psi_{\text{new}}|\mathbf{Y}]$ implicitly averages over all animals, we sample new dives proportionally, so that we match the empirical distribution of the number of validation dives across satellite tags.

The observation distribution $[\mathcal{Y}_{\text{new}} | \tilde{\mathcal{Y}}_{\text{new}}, \varepsilon_{\text{new}}, \psi_{\text{new}}]$ in (6) is degenerate because our model assumes observations are made at times that are exactly offset from the start of a new dive $\tilde{\mathcal{Y}}_{\text{new}}$ by ε_{new} units of time. Similarly, the sign of the end offset ψ_{new} determines whether or not the final surface bin is observed. In our Monte Carlo computation, the offsets $\varepsilon_{\text{new}}^{(b)}$ and $\psi_{\text{new}}^{(b)}$ are uniformly sampled from the b th posterior sample of dive offsets $\mathbf{E}_i^{(b)}$ and $\Psi_i^{(b)}$. Once again, the animal index i is chosen so that our posterior predictive distribution's number of dives across satellite tags matches the empirical distribution seen in the validation dataset.

The rest of the b th posterior sample is used to draw the dive $\tilde{\mathcal{Y}}_{\text{new}}^{(b)}$ from $\tilde{\mathcal{Y}}_{\text{new}}^{(b)} \sim [\tilde{\mathcal{Y}}_{\text{new}}^{(b)} | \Theta_i^{(b)}, \mathbf{T}_{i1}^{(b, I_b)}, \mathbf{T}_{i2}^{(b, I_b)}]$. Here $\Theta_i^{(b)}$ is the b th posterior sample of model parameters, and $\mathbf{T}_{i1}^{(b, I_b)}$ and $\mathbf{T}_{i2}^{(b, I_b)}$ are stage 1 and stage 2 transition times taken from the b th posterior samples $\mathbf{T}_{i1}^{(b)}$ and $\mathbf{T}_{i2}^{(b)}$, respectively. Specifically, the stage transition times $\mathbf{T}_{i1}^{(b, I_b)}$ and $\mathbf{T}_{i2}^{(b, I_b)}$ are the I_b th entry of the sampled random effect vectors $\mathbf{T}_{i1}^{(b)}$ and $\mathbf{T}_{i2}^{(b)}$; the index I_b is sampled uniformly from $\{1, \dots, n_i\}$ for each $b = 1, \dots, B$.

4. SIMULATION INVESTIGATION

We assess the model's ability to recover model parameters through simulation. We simulate 100 dives from the model when the parameters Θ are fixed at their prior mean values for vertical speeds, and at $\pi^{(1)} = 0.97$ and $\pi^{(3)} = 0.03$. The parameter values are obtained from a previous study of deep whale dives (Tyack et al. 2006). We use the empirical priors from our study to specify the distributions G_1 and G_2 from which we simulate stage durations. The effect of time between observations is important for understanding how satellite tags that could record data more frequently may potentially improve parameter recovery. Thus, we estimate model parameters when the simulated dives are observed every 300, 60, and 30s. We observe the simulated dives from $t = 0$, and treat the offsets (ε , ψ) as unknown during estimation.

All parameters are adequately recovered, and results reveal that parameter estimates are moderately improved by increasing sampling frequency (i.e., by reducing time between observations). The Gibbs sampler is run for 10,000 iterations, and the first 5000 samples are discarded for burn-in. Traceplots and effective sample size diagnostics suggest that the chain converges and yields adequate effective sample sizes to draw posterior inference. Table 1 compares posterior estimates for the model parameters Θ to the true parameters used to simulate the data. Decreasing the time between observations also decreases the uncertainty for the directional preference parameters $\pi^{(1)}$ and $\pi^{(3)}$, but the uncertainty in the speed parameters $\alpha_0^{(1)}$, $\alpha_0^{(2)}$, and $\alpha_0^{(3)}$ does not noticeably decrease. However, the posterior means get closer to the true value. The findings suggest that collecting data at 5-minute intervals is adequate for estimating the dive characteristics we consider.

5. ANALYSIS OF REAL DIVE DATA

From the satellite telemetry records described in Sect. 2, we analyze 318 deep dives ($> 800\text{m}$) across seven animals. On average, we observed 45 deep dives per animal. We

Table 1. Recovery of simulation parameters when fitting the dive model to different sets of observation times. Table shows posterior means, standard deviations, and 95% highest posterior density intervals for each parameter. Each parameter's exact value is noted at the start of each group of rows

Parameter	Timestep (s)	Mean	Sd.	HPD Interval
$\pi^{(1)} = 0.97$	300	0.95	0.029	(0.89, 0.99)
	60	0.97	0.009	(0.95, 0.98)
	30	0.97	0.007	(0.95, 0.98)
$\pi^{(3)} = 0.03$	300	0.05	0.015	(0.03, 0.08)
	60	0.04	0.007	(0.02, 0.05)
	30	0.04	0.006	(0.03, 0.05)
$\alpha_0^{(1)} = 0.41$	300	0.61	0.960	(-1.20, 2.50)
	60	0.47	0.990	(-1.50, 2.50)
	30	0.44	1.020	(-1.65, 2.36)
$\alpha_0^{(2)} = -1.20$	300	-1.60	1.100	(-3.70, 0.40)
	60	-1.50	0.940	(-3.40, 0.41)
	30	-1.43	1.070	(-3.45, 0.54)
$\alpha_0^{(3)} = 0.00$	300	0.20	0.960	(-1.90, 2.00)
	60	-0.03	0.990	(-2.00, 1.90)
	30	-0.02	0.974	(-1.99, 1.88)

estimate model parameters for these dives, we use the posterior predictive distribution to conduct out-of-sample model validation, and we draw inference on summary-statistics for dives, as discussed in Sect. 1.

The Gibbs sampler is run for 10,000 iterations, and the first 5000 samples are discarded for burn-in. Traceplots and effective sample size diagnostics suggest that the chain converges and yields adequate effective sample sizes to draw posterior inference. Table 2 presents posterior estimates of the model parameters Θ , and the posteriors show Bayesian learning over the priors derived from Tyack et al. (2006). The posterior distribution for Θ also has generally low correlation between components. The posterior correlations for each intercept-slope pair $(\alpha_0^{(s)}, \alpha_1^{(s)})$ are negative, as expected, and, in fact, for all three stages are -0.65 up

Table 2. Posterior means, standard deviations, and 95% Highest posterior density (HPD) intervals for model parameters Θ when fit to satellite tag data

Parameter	Stage	Mean	Sd.	HPD.Interval
$\pi^{(s)}$	$s = 1$	0.97	0.01	(0.93, 0.99)
	$s = 3$	0.05	0.01	(0.03, 0.06)
$\alpha_0^{(s)}$	$s = 1$	0.55	0.28	(0.04, 1.13)
	$s = 2$	-1.70	0.38	(-2.46, -0.98)
	$s = 3$	0.16	0.28	(-0.36, 0.73)
$\alpha_1^{(s)}$	$s = 1$	-0.09	0.42	(-0.90, 0.78)
	$s = 2$	0.35	0.58	(-0.77, 1.53)
	$s = 3$	-0.06	0.42	(-0.90, 0.77)
$\sigma_\alpha^{2(s)}$	$s = 1$	0.30	0.18	(0.08, 0.62)
	$s = 2$	0.49	0.35	(0.12, 1.06)
	$s = 3$	0.32	0.19	(0.09, 0.67)

to two significant digits. The absolute magnitude of all other posterior correlations between pairs of parameters is less than 0.12.

Out-of-sample model validation shows that our model reproduces characteristics of observed dives. Illustratively, Fig. 3 shows that the posterior predictive distributions of depth bins at the nominal 5-minute observation times match the empirical distributions of the validation dives relatively well for animal tag record Zc084. Results for the other tag records are similar, and are available in the Supplement (Supplement, Figures 1–12). The posterior predictive distributions generally have the same shape and location as the empirical distributions. However, for some time windows the predictive distributions tend to be shifted to the right (i.e., predicted depths are deeper than observed).

The posterior predictive distribution for dives over 800m allows us to draw inference on characteristics of dives that are biologically meaningful. The posterior predictive distributions for descent and foraging durations (Fig. 4) match the validation distributions well, and the predictive distribution for overall dive duration is slightly overdispersed relative to the validation distribution. (Recall that these validation distributions are not “observed” but are extracted empirically.) The posterior predictive distribution for the maximum observed depth of dives over 800m is approximately uniformly distributed across the six depth bins that collectively span 773 m to 2383 m. The distribution is overdispersed with respect to the empirical validation data for Zc084, but is more reasonable for animal tag record Zc085 (Supplement, Figures 13–14).

The random effects model lets us jointly estimate diving rates for each whale. Bivariate posterior distributions for descent $\lambda^{(1)}$ and ascent $\lambda^{(3)}$ speeds are similar across whales, although there is some evidence that two clusters exist—one in which $\lambda^{(1)} \approx \lambda^{(3)}$, and a second in which $\lambda^{(1)} > \lambda^{(3)}$ (Supplement, Figure 14). Similarly, the posterior predictive distribution for overall dive duration shows some variation between animals, especially in the right tails of the distributions (Supplement, Figure 15).

6. SUMMARY AND FUTURE WORK

Our work has been motivated by the practical employment of longer term tags, i.e., satellite tags, for marine mammal diving behavior. Most historical satellite tag data recorded only the maximum depth an animal reached during a dive. In contrast, here we examine data that were recorded at finer temporal resolutions. Using these data, we developed a multi-stage model for the deep dives associated with a satellite tag record. The satellite tag data examined here suffer two strong challenges with regard to learning about diving behavior. First, they are collected very sparsely—in our case, at 5-min intervals—and second, they only records depths in terms of an arbitrarily created set of depth bins.

Our approach to learning about dive behavior formulates a generative model which produces dive realizations incorporating three observable stages, a descent stage, a foraging stage, and an ascent stage. Other models for whale diving behavior use multiple foraging sub-stages (Langrock et al. 2014). However, these models are developed for data with much higher resolution than satellite tags provide and, in addition, view the substages as latent.

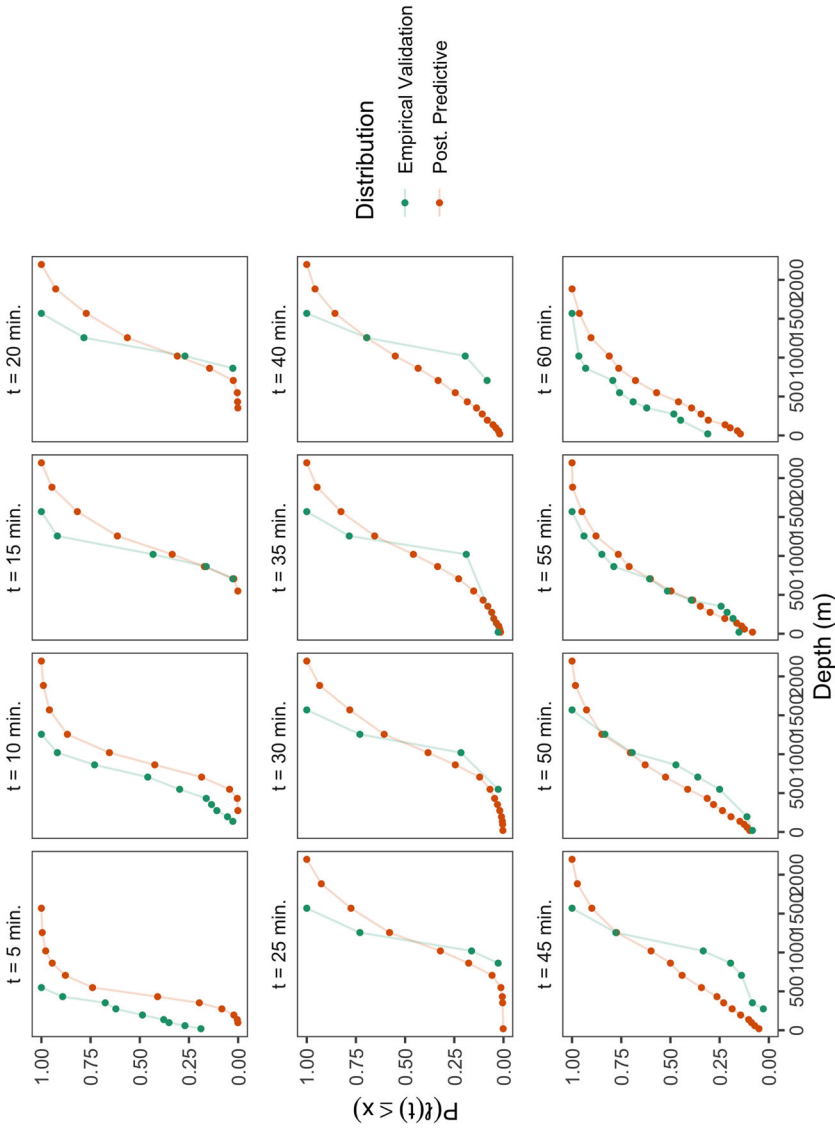


Figure 3. Out-of-sample model validation. Each panel compares the empirical distribution ($n = 37$) of depth bins at the nominal 5-minute observation times to the posterior predictive distribution for animal tag record Zc084 .

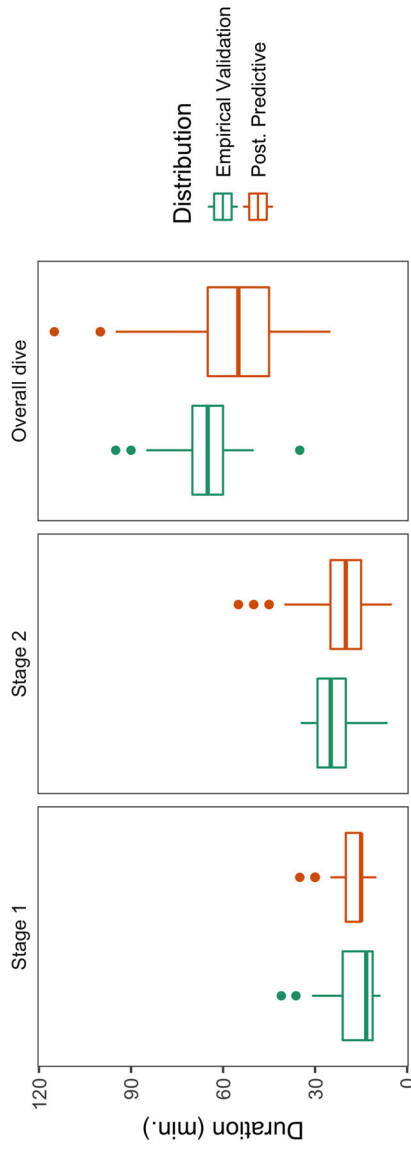


Figure 4. Comparison of predictive and validation distributions for the length of (1) the descent stage, (2) the bottom/foraging stage, and (3) the overall dive length for animal tag record Zc084 .

Our model is a continuous-time Markov chain specification, which supplies a depth bin at every time during the dive. The model proceeds sequentially within a given stage, choosing a duration in bin followed by a transition to a new bin. The deep dives within a satellite tag record share a common stage-wise duration and transition specification. However, dive-specific random effects are introduced. We introduce replicate-specific parameters in order to borrow strength and capture inter-animal variation.

We address misalignment issues to fit the model to observations. The sparse temporal interval at which depth bins are recorded precludes knowing exactly when a dive starts and ends, in addition to times in stages. We handle this issue through the properties of our continuous-time Markov chain specification as well as additional dive-specific random effects that capture the start time and end time of a dive.

The overall model specification is hierarchical and was fitted within a Bayesian framework. The model was implemented in NIMBLE version 0.9.0 with a custom likelihood function (de Valpine et al. 2017). The model runs quickly, fitting the family of dives analyzed in a few hours with modest computing power. Through simulation, we demonstrated the ability of the model to recover the true continuous-time discrete-space model parameters. Then, we applied the model to seven, real satellite tag records. We used a partly empirical approach to specify stage transition time priors. We validated the model, out of sample, by showing the ability to predict dive behavior across the sparse time intervals, and also with regard to dive duration and maximum depth. The validation results suggest that this model is able to capture and reproduce biologically meaningful characteristics of deep dives.

There is a rich agenda of future work. Our current effort captures baseline diving behavior. We seek to learn about departure from such behavior in response to exposure to anthropogenic sound such as navy sonar signals. Using information on exposure to sound can we identify deep dive behavior that differs from baseline?

We will also attempt to model the entire satellite tag record rather than just the deep dive portion. This entails modeling the dives sequentially, including the time between deep dives, which is devoted to shallow dives and surface activity. Such modeling may reveal interesting shallow dive behavior and also inform about subsequent deep dive behavior.

Additionally, we plan to enrich our generative model to allow within-stage, time-dependent durations and bin transition probabilities. Further, we will explore the possibility of bringing our approach to terrestrial animal movement modeling, considering transitions over rectangles in two-dimensional space. The intent is to avoid the use of approximate imputation distributions employed in current CTDS modeling, as described in the Introduction.

ACKNOWLEDGEMENTS

The data analyzed here were collected as part of the Atlantic Behavioral Response Study under NMFS permit #22156, issued to Doug Nowacek. We thank Andy Read of Duke University and Brandon Southall of Southall Environmental Associates for allowing us use of the data. Support for the Atlantic BRS is provided by Naval Facilities Engineering Command Atlantic under Contract No. N62470-15-D-8006, Task Order 18F4036, Issued to HDR, Inc. The research reported here was financially supported by the United States Office of Naval Research grant N000141812807, under the project entitled Phase II Multi-study Ocean acoustics Human effects Analysis (Double MOCHA). This contribution is Double MOCHA White Paper #01. We acknowledge and thank several people for stimulating conversation that spurred our thinking and development of the model, including Richard

Glennie, Catriona Harris, Mark Johnson, Théo Michelot, Len Thomas, and Peter Tyack—all from the University of St Andrews. We also thank Will Cioffi, Nicola Quick from Duke University. Finally, we thank Stacy DeRuiter of Calvin University.

[Accepted October 2020. Published Online January 2021.]

REFERENCES

- Al-Mohy AH, Higham NJ (2010) A new scaling and squaring algorithm for the matrix exponential. *SIAM J Matrix Anal Appl* 31(3):970–989
- Anderson WJ (2012) *Continuous-time Markov chains: An applications-oriented approach*. Springer, Berlin
- Andrieu C, Thoms J (2008) A tutorial on adaptive mcmc. *Stat Comput* 18(4):343–373
- de Valpine P, Turek D, Paciorek C, Anderson-Bergman C, Lang D Temple, R B (2017) Programming with models: writing statistical algorithms for general model structures with NIMBLE. *J Comput Graph Stat* 26:403–417
- Gelfand A, Sahu S, Carlin B (1996) Efficient parametrization for generalized linear mixed models. In: Bernardo J et al (eds) *Bayesian Statistics 5*. Clarendon Press, Oxford, pp 165–180
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2014) *Bayesian data analysis*. CRC Press, USA
- Hanks EM, Hooten MB, Alldredge MW et al (2015) Continuous-time discrete-space models for animal movement. *Ann Appl Stat* 9(1):145–165
- Hooker SK, Baird RW (2001) Diving and ranging behaviour of odontocetes: a methodological review and critique. *Mamm Rev* 31(1):81–105
- Johnson DS, London JM, Lea M-A, Durban JW (2008) Continuous-time correlated random walk model for animal telemetry data. *Ecology* 89(5):1208–1215
- Johnson M, Madsen PT, Zimmer W, De Soto NA, Tyack P (2006) Foraging Blainville’s beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation. *J Exp Biol* 209(24):5038–5050
- Langrock R, Marques TA, Baird RW, Thomas L (2014) Modeling the diving behavior of whales: a latent-variable approach with feedback and semi-markovian components. *J Agric Biol Environ Stat* 19(1):82–100
- 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*). *Animal Biotelemetry* 7(1):5
- Scharf H, Hooten MB, Johnson DS (2017) Imputation approaches for animal movement modeling. *J Agric Biol Environ Stat* 22(3):335–352
- Shearer JM, Quick NJ, Cioffi WR, Baird RW, Webster DL, Foley HJ, Swaim ZT, Waples DM, Bell JT, Read AJ (2019) Diving behaviour of Cuvier’s beaked whales (*Ziphius cavirostris*) off Cape Hatteras, North Carolina. *R Soc Open Sci* 6(2):181728
- Southall B, Baird R, Bowers M, Cioffi W, Harris C, Joseph J, Quick N, Margolina T, Nowacek D, Read A et al (2018) Atlantic Behavioral Response Study (BRS)—2017 Annual progress report. Prepared for US Fleet Forces Command. Submitted to Naval Facilities Engineering Command Atlantic, Norfolk, Virginia, under Contract No N62470-15-D-8006, Task Order 50, issued to HDR Inc., Virginia Beach, Virginia
- Tyack PL, Johnson M, Soto NA, Sturlese A, Madsen PT (2006) Extreme diving of beaked whales. *J Exp Biol* 209(21):4238–4253
- Van Dyk DA, Meng X-L (2001) The art of data augmentation. *J Comp Graphic Stat* 10(1):1–50
- Wilson K, Hanks E, Johnson D (2018) Estimating animal utilization densities using continuous-time markov chain models. *Methods Ecol Evol* 9(5):1232–1240

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.