Inferring Fine-Scale Animal Behaviors using Hidden Markov Models

Evan Sidrow

June 5, 2020

Abstract

In the field of animal movement, recent advances in high-frequency tagging technology have made available a vast amount of data which can exhibit simultaneous behavioral processes occurring at different time scales. One way to model this data is to use a hierarchical hidden Markov model (HHMM), where the system is modeled as a nested structure of hidden Markov models (HMMs). At very short time scales, however, observations can exhibit complicated dependence structures which cannot be easily captured by traditional HMMs. This work investigates how to incorporate fine-scale processes into the larger structure of HHMMs while maintaining computational efficiency. We apply our method to dive data collected from a northern resident killer whale off the coast of British Columbia, Canada.

The field of animal movement is in the midst of a "data renaissance" where advancements in tagging technology have given rise to an explosion of data available for statistical modeling. In particular, tagging technologies are capable of recording observations at rates of tens of hertz, resulting in time series containing millions of observations over the course of several hours. This results in a vast amount of data which often exhibits many different simultaneous behavioral processes occurring at different time scales.

One solution to this issue is to use a hierarchical hidden Markov model (HHMM). HHMMs model the entire system as a nested structure of hidden Markov models (HMM) where each HMM corresponds to one behavioral process. One nice property of HHMMs is that its likelihood is relatively easy to compute, facilitating fast maximum likelihood estimates for its associated parameters.

At the shortest time scales, however, observations often exhibit complicated dependence structures which cannot be easily captured by a traditional HMMs. To address this issue, it is possible to model small-scale animal behavior as the solution to some stochastic differential equation, but these methods tend to be computationally intractable and require approximate inference techniques such as Markov-chain Monte Carlo (MCMC).

This work investigates how to incorporate fine-scale processes into the larger structure of hierarchical hidden Markov models while maintaining computational efficiency. We bridge the gap between discrete hidden Markov models and continuous-time stochastic process models by showing that the two are equivalent under certain conditions. In addition, we extract features from highly structured sub-dive behaviors that otherwise could not be modeled with a simple HMM. Finally, we apply our method to dive data collected from a Northern resident killer whale off the coast of British Columbia, Canada.

1 Background

Hidden Markov models are useful when inferring a single unobserved process, but biological processes often involve multiple simultaneous hidden processes which can occur and at different time scales. For example, a preliminary observation of the killer whale dive data shown in figure (1) shows that the behavior of this killer whale changes between approximately hour-long periods of predominately short, shallow dives and long, deep dives. Leos-Barajas et al. encounter a similar issue when modeling the movement of a harbor porpoise in the North Sea, and use it as a motivating example when they introduce hierarchical hidden Markov models.

1.1 Hidden Markov models

Hidden Markov models (or HMMs) are comprised of an unobserved Markov chain $X = (X_1, \dots, X_T)$ and a sequence of (possibly high-dimensional) observations $Y = (Y_1, \dots, Y_T)$, each of length T. Each random variable in the unobserved chain X_t can take one of N possible values, and X has corresponeding probability transition matrix $\Gamma \in \mathbb{R}^{N \times N}$ and initial distribution $\delta \in \mathbb{R}^N$:

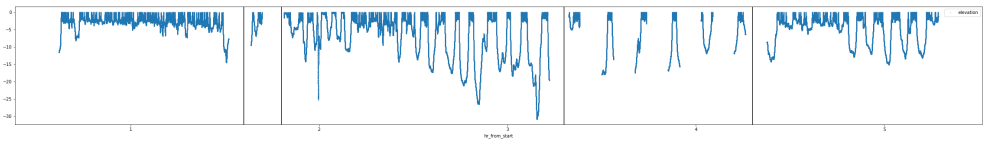

Figure 1: Raw depth data of a Killer Whale off the coast of British Columbia, Canada.

$$\delta_i = Pr(X_1 = i)$$

$$\Gamma_{ij} = Pr(X_{t+1} = j | X_t = i) \qquad \forall t \in \{1, \ldots, T-1\}$$

Further, each random varaible X_t emits an observation Y_t whose distribution depends only on the value of X_t and none of the preceding observations or behavioral states: $p_\theta(y_t|x_t, x_{t-1}, \ldots, x_1, y_{t-1}, \ldots, y_1) = p_\theta(y_t|x_t)$. Note that the emission distribution depends upon parameters θ. A visualization of this dependence structure can be seen in figure (2). In the field of animal movement, the unobserved chain X usually represents the latent behaviour of an animal (e.g. foraging, resting, migrating, etc.), while the observations Y are often a series of step lengths and turning angles for land animals and either depth or accelerometer data (or both) for marine animals.

The probability transition matrix Γ and the parameters of the emission distributions, θ, can be estimated by maximizing the likelihood of the observed data y, $\mathcal{L}_{\text{HMM}}(y)$, with respect to the Γ and θ. In addition, $\mathcal{L}_{\text{HMM}}(y)$ can be calculated using the *forward algorithm*:

$$\mathcal{L}_{\text{HMM}}(y; \theta, \Gamma, \delta) = \delta P(y_1; \theta) \prod_{t=2}^{T} \Gamma P(y_t; \theta) \mathbf{1}$$

where:

$$P(y_t; \theta) = \text{diag}(p_\theta(y_t|X_t = x_1), ..., p_\theta(y_t|X_t = x_N))$$

and $\mathbf{1}$ is an N-dimensional column vector of ones.

In order to enusre identifiability and right-stochasticity after optimizing $\mathcal{L}_{\text{HMM}}(y)$, Γ is parameterized using $\eta \in \mathbb{R}^{N \times N}$ and the following link function:

$$\Gamma_{ij} = \frac{\exp(\eta_{ij})}{\sum_{k=1}^{N} \exp(\eta_{ik})}, \qquad \eta_{ii} = 0 \quad \forall i \in \{1, \ldots, N\}$$

This allows for unconstrained optimization over η and removes the constraint that Γ be right-stochastic. $\mathcal{L}_{\text{HMM}}(y; \theta, \Gamma, \delta)$ can be maximized using any numerical optimizer.

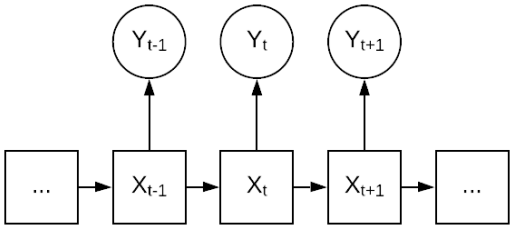

Figure 2: Graphical representation of a traditional HMM.

1.2 Hierarchical HMMs

A hierarchical hidden Markov model (or HHMM) is a variation of a hidden Markov model in which each hidden state of the original HMM X_t emits both an observation Y_t as well as another fine-scale hidden markov model of length T_t^*. This fine-scale HMM is comprised of a Markov chain $X_t^* = (X_{t,1}^*, \ldots, X_{t,T_t^*}^*)$ and observations $Y_t^* = (Y_{t,1}^*, \ldots, Y_{t,T_t^*}^*)$. As before,

each fine-scale observation Y^*_{t,t^*} depends only on the value of its corresponding hidden state, X^*_{t,t^*}. X^*_{t,t^*} can take one of N^* values and is characterized by an initial distribtuion $\delta^{*(X_t)} \in \mathbb{R}^{N^*}$ and probability transition matrix $\Gamma^{*(X_t)} \in \mathbb{R}^{N^* \times N^*}$:

$$\delta_i^{*(x_t)} = Pr(X^*_{t,1} = i | X_t = x_t)$$

$$\Gamma_{ij}^{*(x_t)} = Pr(X^*_{t,t^*+1} = j | X^*_{t,t^*} = i, X_t = x_t) \qquad \forall t^* \in \{1, \ldots, T_t^* - 1\}$$

Finally, the fine-scale emission probabilities $p_{\theta^{*(X_t)}}(y^*_{s,t}|x^*_{s,t})$ are parameterized by $\theta^{*(X_t)}$. Note the parameters of the fine-scale hidden Markov model, $\Gamma^{*(X_t)}$, $\delta^{*(X_t)}$, and $\theta^{*(X_t)}$ all depend upon the hidden state of the *crude-scale* hidden markov model. However, depending upon the discretion of the researcher, it is possible to force any of these parameters to be independent of the crude-scale hidden state X_t. A visualization of the full structure of the HHMM can be seen in figure (3).

Due to the nested structure of a hierarchical hidden Markov model, the likelihood of an HHMM is still easy to calculate using the forward algorithm:

$$\mathcal{L}_{\text{HHMM}}(y, y^*; \theta, \theta^*, \Gamma, \Gamma^*, \delta, \delta^*) = \delta P(y_1, y_1^*; \theta, \theta^*, \Gamma^*, \delta^*) \prod_{t=2}^{T} \Gamma P(y_t, y_t^*; \theta, \theta^*, \Gamma^*, \delta^*)\mathbf{1}$$

where:

$$P(y_t, y_t^*; \theta, \theta^*, \Gamma^*, \delta^*) = \text{diag}\left[p_\theta(y_t|x_t = x_1)\mathcal{L}_{\text{HMM}}\left(y_t^*; \theta^{*(x_1)}, \Gamma^{*(x_1)}, \delta^{*(x_1)}\right), ..., \right.$$
$$\left. p_\theta(y_t|x_t = x_N)\mathcal{L}_{\text{HMM}}\left(y_t^*; \theta^{*(x_N)}, \Gamma^{*(x_N)}, \delta^{*(x_N)}\right)\right]$$

Note that this formualtion assumes that the crude-scale observations at a given time Y_t and the fine-scale observation time series Y_t^* are independent of one another when conditioned on X_t.

For more information on specific considerations for HHMMs such as incorporating covariates into the probability transition matrix, model selection and model checking, see Adam et al [1].

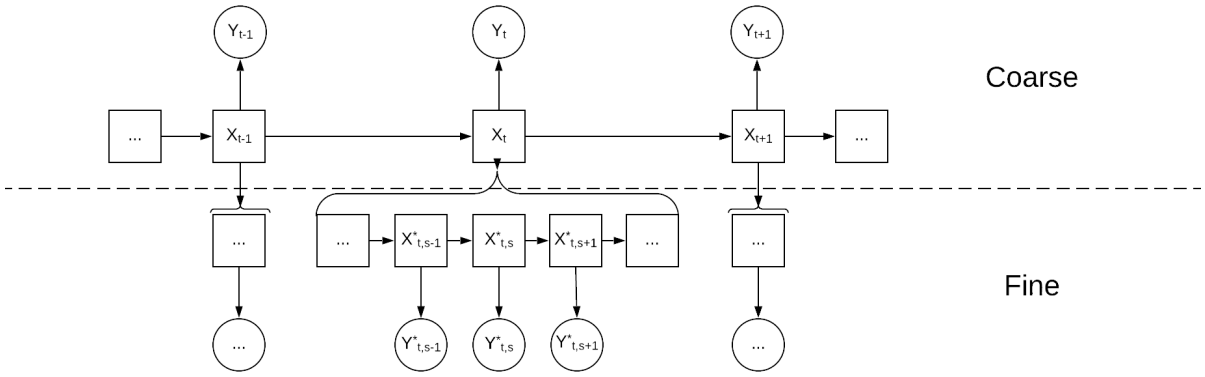

Figure 3: Graphical representation of a traditional HHMM.

1.3 Conditionally autoregressive HMMs

One of the key assumptions of both HMMs and HHMMs is *conditional indpendence* between observations at both the crude and fine scale. Namely, given the state X_t or X^*_{t,t^*}, Y_t or Y^*_{t,t^*} (respectively) is assumed to be independent from all other observations. Therefore, traditional HMMs and HHMMs can fail when the observation sequence Y exhibits significant auto-correlation in time. Examples include fluking in marine mammals in Vancouver, BC (see the results section) and the swiming behavior of horn sharks off the coast of Southern California [1].

One way to deal with autocorrelation in fine-scale behavioral processes is to use a state-switching continuous model such as the one introduced by Michelot et al [5], which models the movement of an animal as an Ornstein-Uhlenbeck process with parameters that depend upon the underlying behaviour state of the animal. Continuous time models are

advantagous because of their flexibiliy: they can be built up from aribitrarily complex stochastic differential equations and they allow for uneven step lengths in the observations sequence Y. However, most continuous time models require MCMC algorithms to perform inference and as a result are not easily incorporated into the HHMM structure.

Another option is to use the CarHMM, or *conditionally auto-regressive hidden Markov model*, introduced by Lawler et al [4], in which autocorrelation is explicity modeled into the emission distributions of the HMM while maintaining the strucutre needed to run the forward algorithm for fast direct likelihood maximization. In particular, if the emission distribution of observation Y_t is parameterized by its mean and variance, i.e. $\theta_{x_t} = \{\mu_{x_t}, \sigma^2_{x_t}\}$, the CarHMM introduces autocorrleation into the HMM by assuming that an observation Y_t has mean $(1 - \phi_{x_t}) \cdot \mu_{x_t} + \phi_{x_t} \cdot y_{t-1}$ rather than μ_{x_t}. Note that the autocorrelation term ϕ_{x_t} depends upon the behavioral state of the animal. This model easily fits into the HHMM structure, but seems to lack the flexibility and natural interpretation of continuous-time models. However, we prove in the following section that under certain conditions these two models are in fact equivalent.

The likelihood of CarHMM is still compatible with the forward algorithm:

$$\mathcal{L}_{\text{CarHMM}}(y) = \delta \prod_{t=2}^{T} \Gamma P(y_t; \theta) \mathbf{1} \tag{1}$$

where:

$$P(y_t; \theta) = \text{diag}(p_\theta(y_t | y_{t-1}, X_t = x_1), ..., p_\theta(y_t | y_{t-1}, X_t = x_N)), \qquad t > 1$$

and the graphical model associated with the strcture of a CarHMM is shown in figure (4). Note that the first observation y_1 is assumed to be fixed as an initial value.

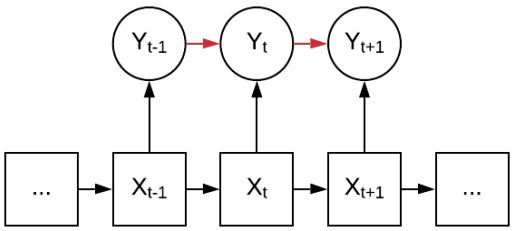

Figure 4: Graphical representation of a traditional CarHMM. The additional arrows representing autocorrelation between observations are shown in red for emphasis.

1.4 State decoding

Once an HMM, HHMM, or CarHMM model is fit using the process described above, it is common to find the most likely sequence of hidden states $\hat{X}$ conditioned on the learned parameters by using a dynamic programing algorithm called the Viterbi algorithm [6]. In the case of HHMMs, this can be followed by running the Viterbi algorithm again on each sub-dive state to find the mostly likely sequence of fine-scale hidden states $\hat{X}^*_t$ conditioned on the learned parameters and the estimated crude-state value $\hat{X}_t$. Note that while $\hat{X}$ is a maximum likelihood estimate of X, $\hat{X}^*_t$ is *not* necessarily a maximum likelihood estimate of $\hat{X}^*_t$ because it is conditioned on the value of $\hat{X}_t$.

While the Viterbi algorithm is the de-facto standard in the current ecology literature, we suggest to instead find the *probability* of each crude-level state (conditioned on the learned parameters) using the *forward-backward algorithm*. The forward-backward algorithm has the same time complexity as the forward algorithm and is also used to find the *psuedoresiduals* of a given model, which is an important tool for model validation. In addition, For HHMMs in particular, the forward-backward algorithm can be used recursively to find the probability of the fine-level states $X^*_{s,t}$ exactly by marginalizing out X_t:

$$P(X^*_{s,t} = x^*_{s,t}) = \sum_{n=1}^{N} P(X_t = x_n) P(X^*_{s,t} = x^*_{s,t} | X_t = x_n)$$

Where $P(X_t = x_n)$ can be found using the forward-backward algorithm on the crude-level markov chain and $P(X^*_{s,t} = x^*_{s,t} | X_t = x_n)$ can be found by running the forward-backward algorithm on the fine-level HMM for every possible value of X_t.

2 Results regarding CarHMMs and the Ornstein-Uhlenbeck process

2.1 Equivalence of CarHMM and OU process

Once the CarHMM is fit by maximizing the likelihood in equation (1), the emission distributions can be interpretted as the solution of a state-switching Ornstein-Uhlenbeck process similar to the one introduced by Michelot et al [5]. In particular, the following two conditions must be met:

1. The underlying behavioral state of the continuous-time model must follow a Markov chain rather than a Markov process.
2. The emission distributions of the CarHMM must be normal.

If both conditions (1) and (2) are met, then the CarHMM is equivalent to a state-switching Ornstein-Uhlenbeck process. This gives new intpretation to the learned parameters of the CarHMM in the context of continuous-time model.

2.1.1 Proof

A one-dimensional state-switching Ornstein-Uhlenbeck process y is the solution to the following stochastic differential equation:

$$dy_t = \beta_{x_t}(\gamma_{x_t} - y_t)dt + \omega_{x_t}dW_t$$

where x_t is the fine-scale behavior of the animal at time t, β_{x_t} relates to rate at which the process returns to its mean value, γ_{x_t} is the long-term mean value of the process, ω_{x_t} is related to short-term variance, and W is a Wiener process. x_t is described by an unobserved Markov process. If the behavioral state x_t does not change between observations (i.e. x_t follows a Markov chain), the solution to this equation is known to be the following [5]:

$$y_{t+\Delta t}|x_{t:t+\Delta t} \sim \mathcal{N}\left((1 - e^{-\beta_{x_t}\Delta t})\gamma_{x_t} + e^{-\beta_{x_t}\Delta t}y_t, \quad \frac{\omega_{x_t}^2}{2\beta_{x_t}}(1 - e^{-2\beta_{x_t}\Delta t})\right)$$

Now, suppose that Δt is constant for all observations, as is the case for hidden Markov models. In addition, introduce the following transformations:

$$\mu_{x_t} = \gamma_{x_t}, \qquad \phi_{x_t} = e^{-\beta_{x_t}\Delta t}, \qquad \sigma_{x_t}^2 = \frac{\omega_{x_t}^2}{2\beta_{x_t}}(1 - e^{-2\beta_{x_t}\Delta t}) \tag{2}$$

Then, we have the following:

$$y_{t+\Delta t}|x_{t:t+\Delta t} \sim \mathcal{N}\left((1 - \phi_{x_t})\mu_{x_t} + \phi_{x_t}y_t, \quad \sigma_{x_t}^2\right)$$

If Δt is fixed and x_t is adjusted to follow a Markov chain rather than a Markov process, then this model is equivalent to the CarHMM with normal emission probabilities. Note that all of the parameter transformations above are one-to-one, so it is easy to go from the CarHMM to the continuous model and back again. This allows for the pricipled construction of the continuous-time model to be combined with the computational convienence of the CarHMM.

2.2 Generalization to unequal time steps

Using this intuition, we generalize the CarHMM to data with unequal time steps: i.e. Δt becomes a function of t. First, it is necessary to assume that the behavioral state x_t only changes once between observations, which is a fair assumption if each time step Δt is sufficiently small. Next, the following definitions are introduced:

$$\Gamma_{\Delta t,\Lambda} = \begin{pmatrix} 1 - e^{-\lambda_1\Delta t} & p_{12}e^{-\lambda_1\Delta t} & \dots & p_{1N}e^{-\lambda_1\Delta t} \\ p_{21}e^{-\lambda_2\Delta t} & 1 - e^{-\lambda_2\Delta t} & \dots & p_{2N}e^{-\lambda_2\Delta t} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1}e^{-\lambda_N\Delta t} & p_{N2}e^{-\lambda_N\Delta t} & \dots & 1 - e^{-\lambda_N\Delta t} \end{pmatrix}$$

where $\Lambda = \{p, \lambda\}$, p_{ij} is the probability that the animal moves from state i to state j given that it left state i, and λ_i is the rate at which the animal leaves behavioral state i. Finally, we just note that the parameters in equation (2) become functions of the time step between observations Δt:

$$\mu_{x_t} = \gamma_{x_t}, \qquad \phi_{x_t}(\Delta t) = e^{-\beta_{x_t}\Delta t}, \qquad \sigma_{x_t}^2(\Delta t) = \frac{\omega_{x_t}^2}{2\beta_{x_t}}(1 - e^{-2\beta_{x_t}\Delta t})$$

Then the likelihood of the CarHMM with unequal time steps is just the following:

$$\mathcal{L}_{\text{CarHMM}}(y;\theta,\Lambda,\delta) = \delta \prod_{t=2}^{T} \Gamma_{\Delta t,\Lambda} P(y_t;\theta)\mathbf{1}$$

where:

$$P(y_t;\theta) = \text{diag}(p_\theta(y_t|y_{t-1},\Delta t, X_t = x_1), ..., p_\theta(y_t|y_{t-1},\Delta t, X_t = x_N)), \qquad t > 1$$

and:

$$p_\theta(y_t|y_{t-1},\Delta t, X_t = x_t) = \mathcal{N}\left((1-\phi_{x_t}(\Delta t))\mu_{x_t} + \phi_{x_t}(\Delta t) y_{t-1}, \quad \sigma^2_{x_t}(\Delta t)\right).$$

now $\mathcal{L}_{\text{CarHMM}}(y;\theta,\Lambda,\delta)$ is simply maximized with respect to $\theta = \{\beta,\omega,\gamma\}$, $\Lambda = \{p,\lambda\}$, and δ to find the maximum likelihood estimate of the CarHMM generative model with uneven time steps.

3 Model Formulation

3.1 Short-time Fourier Transform on Fine-scale process

One issue with HMMs is that they assume markovian dynamics conditioned on the hidden state, i.e. that any observation Y_t^* depends only on the behaviorial state X_t^* (here we focus on the fine-scale process). However, there are many animal movement processes which violate this markov propertey on very fine scales. For example swimming behaviour of marine mammals can be periodic since the animal repeatedly flukes to propel itself forward. Work has been done in the past to model non-markovian dynamics in the *behavioural* process X_t^* [3], but addressing non-markovian dynamics within the observation process Y_t^* is still a relatively unstudied area. With improvements in tagging technology allowing for data collection at very high frequencies, data exhibiting noisy and non-markovian fine scale behavior is likely to persist in future studies.

To address this issue, we recommend borrowing techniques from the signal processing literature to compress the data and summarize its essential elements. In particular, we suggest performing the discrete-time Short-time Fourier Transform (STFT) over each observed fine-scale process Y_t^*:

$$STFT\{Y^*_{t,t^*:t^*+w-1}\}(n) := \hat{Y}^{*(n)}_{t,t^*} = \sum_{n=0}^{w-1} Y^*_{t,t^*+n} e^{-i\frac{2\pi k}{w}n} \qquad \forall n \in \{0,\dots,w-1\}, \quad t^* \in \{1, w+1, 2w+1, \dots, w\left(\lfloor T_t^*/w \rfloor - 1\right)+1\}.$$

The STFT slides a moving window of length w accross the time series Y_t^* with a step size of w and transforms the domain of each window from time to frequency. This allows the spectrum of Y_t^* at time t^* to be summerized by a w-dimensional vector of fourier coeficients. While other step sizes can be used for the sliding window, we select w to avoid serial dependence between windows.

If $Y_t^* \in \mathbb{R}^{T_t^*}$, then $\hat{Y}_t^* \in \mathbb{C}^{\lfloor T_t^*/w \rfloor \times w}$. While this allows Y_t^* to be represneted in a way that eliminates obvious periodic behavior, the dataset itself is still approximately as large as Y_t^* itself. To reduce the size of $\hat{Y}_t^*$, we propose taking summary statistics of each window as follows:

$$Z^{*(1)}_{t,t^*} = \mathcal{R}\left(\hat{Y}^{*(0)}_{t,t^*}\right) \qquad Z^{*(2)}_{t,t^*} = \frac{1}{w}\sum_{n=1}^{\tilde{w}} |\hat{Y}^{*(n)}_{t,t^*}|^2$$

$Z^{*(1)}_{t,t^*}$ is equal to the average value of $Y^*_{t,t^*:t^*+w-1}$ while $Z^{*(2)}_{t,t^*}$ is equal to the squared 2-norm of $Y^*_{t,t^*:t^*+w-1}$ that can be attributed to frequencies in the signal between 1 and $\tilde{w}$ periods per window length. Both the window length w and the max frequency $\tilde{w}$ are tuning parameters that should be tuned in a problem-specific way. w should be long enough to capture the periodic behavior of the underlying process (at least as long as the length of a period), but short enough to avoid over-smoothing of the data and to maintain high resultion in the behavioral process X^*. $\tilde{w}$ should be selected such that the maximum frequency of Y_t^* that makes biological sense is $\tilde{w}$ per window length. Note that these summary statistics are just one possible choice, and future studies can adjust the definitions as needed. A visualization of transforming a one-dimensional sequence Y^* to Z^* can be seen in figure (5).

Finally, note that it is possible to accomodate for unequal time steps within each window by using the **non-uniform discrete Fourier transform (NDFT)**. We do not describe the details of this method in this work, but the generalization is straightforward. Refer to Bagchi et al [2] for details.

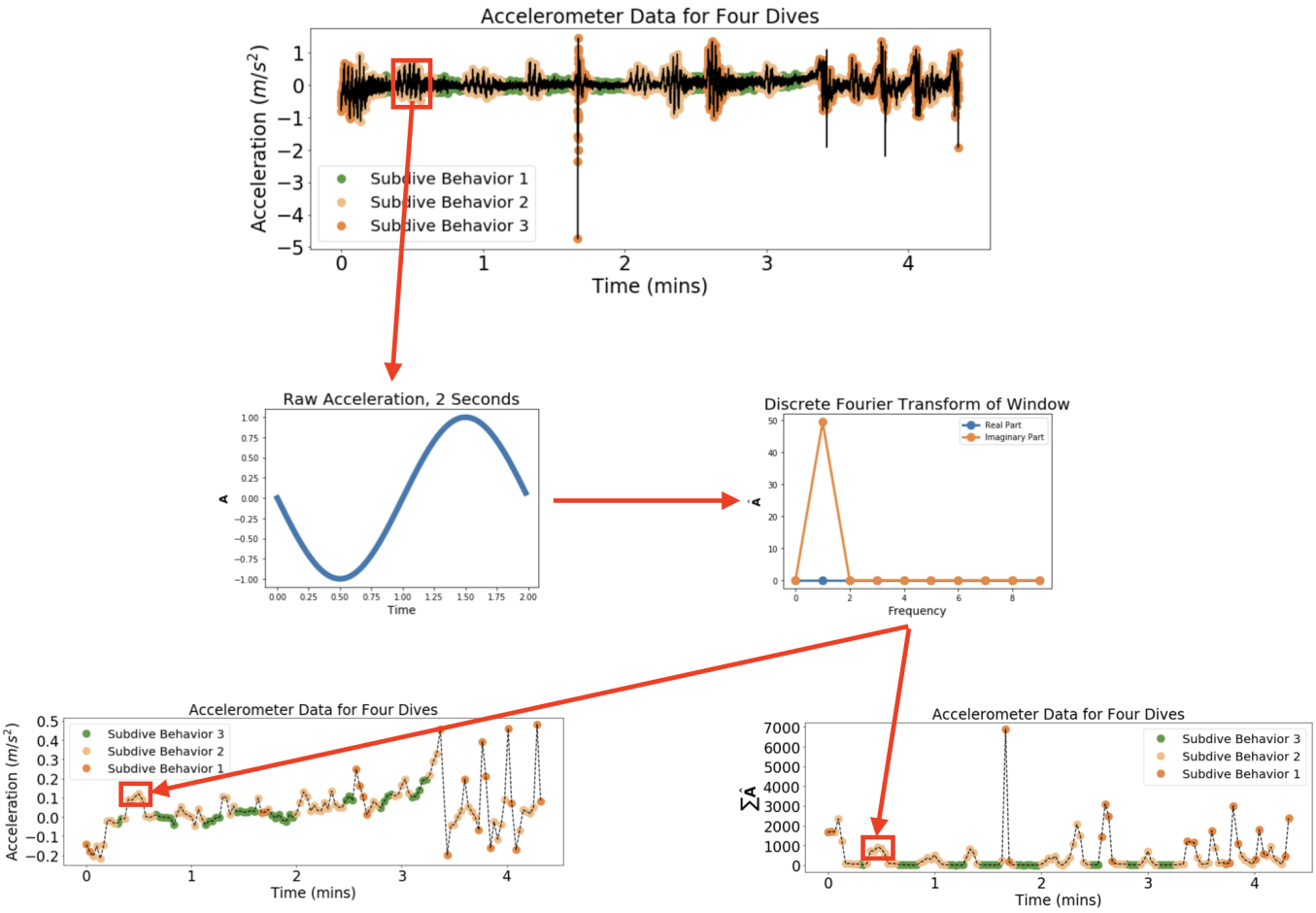

Figure 5: Visualization of transforming Y_t^* into Z_t^* using a sliding window and fourier transform.

3.2 Model Structure: combining the HHMM and CarHMM

Hierarchical hidden Markov models can be used to jointly model simultaneous coarse-scale and fine-scale processes taking place simultaneously. However, as mentioned before, the fine-scale process Y^* can often exhibit autocorrelation and intricate structure. Transforming Y_t^* to Z_t^* removes fine-scale periodic behavior, but Z_t^* can still exhibit autocorrelation, especially in the window averge, $Z_{t,t^*}^{*(1)}$. Therefore, we replace the fine-scale HMM within the Hierarchical HMM with a CarHMM according to figure (6).

The likelihood of this model is still easy to calculate using the forward algorithm:

$$\mathcal{L}_{\text{CarHHMM}}(y, z^*; \theta, \theta^*, \Gamma, \Gamma^*, \delta, \delta^*) = \delta P(y_1, z_1^*; \theta, \theta^*, \Gamma^*, \delta^*) \prod_{t=2}^{T} \Gamma P(y_t, z_t^*; \theta, \theta^*, \Gamma^*, \delta^*)\mathbf{1}$$

where:

$$P(y_t, z_t^*; \theta, \theta^*, \Gamma^*, \delta^*) = \text{diag}\left[p_\theta(y_t|X_t = x_1)\mathcal{L}_{\text{CarHMM}}\left(z_t^*; \theta^{*(x_1)}, \Gamma^{*(x_1)}, \delta^{*(x_1)}\right), ..., \right.$$
$$\left. p_\theta(y_t|x_t = x_N)\mathcal{L}_{\text{CarHMM}}\left(z_t^*; \theta^{*(x_N)}, \Gamma^{*(x_N)}, \delta^{*(x_N)}\right)\right]$$

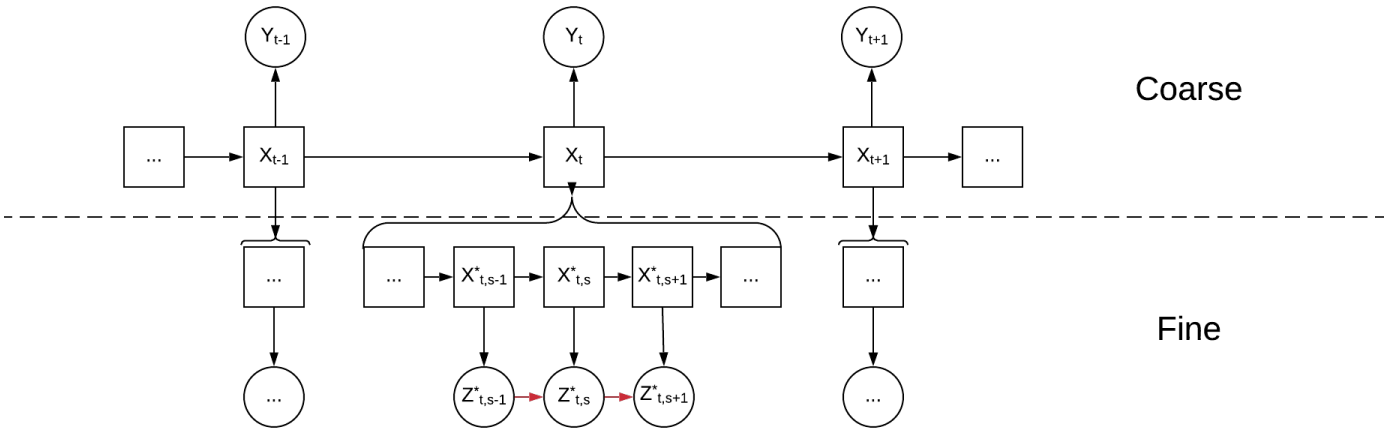

Figure 6: Graphical representation of a CarHHMM. The additional arrows representing autocorrelation between observations are shown in red for emphasis.

4 Simulation Study

4.1 Data Simulation

To test the CarHHMM with STFT, a sequence of 100 marine-animal dives were simulated. The coarse-scale observations Y were set as the duration of each dive, and the fine-scale observations Y^* were set as one dimensional acceleration readings simulated at 50 hertz. Specifically, the following procedure was followed:

1. 100 dive durations were simulated using an HMM generative model with the following parameters:

$$\Gamma = \begin{pmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix}$$
$$Y_t|X_t \sim Gamma$$
$$\mathbb{E}(Y_t|X_t = 1) = 15s, \quad \mathbb{E}(Y_t|X_t = 2) = 60s$$
$$\mathbb{V}(Y_t|X_t = 1) = 25s^2, \quad \mathbb{V}(Y_t|X_t = 2) = 100s^2$$

2. Once the dive durations were calculated, for each $t \in \{1, \ldots, 100\}$, dive t was broken into $\lfloor Y_t/2 \rfloor$ 2-second segments (the end of the dive sequence was discarded). Further, each 2-second segment was assigned a behaviour according to a fine-scale Markov chain X_t^*, where $X_{t,t^*}^* \in \{1, 2\}$ and $t^* \in \{1, 101, 201, \ldots, 100 * \lfloor Y_t/2 \rfloor + 1\}$. The parameters of the fine-scale behaviour Makov chain were set to be as follows:

$$\Gamma^{*(1)} = \begin{pmatrix} 0.25 & 0.75 \\ 0.75 & 025 \end{pmatrix} \qquad \Gamma^{*(2)} = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$$

   where $\Gamma^{*(1)}$ was used for dives where $X_t = 1$ and $\Gamma^{*(2)}$ was used for dives where $X_t = 2$

3. For each 2-second segment, the Fourier modes $\hat{Y}_{t,t^*}^*$ were simulated using the following procedure. Note that the n^{th} Fourier mode of $\hat{Y}_{t,t^*}^*$ is denoted as $\hat{Y}_{t,t^*}^{*(n)}$:

$$
\begin{aligned}
\hat{Y}^{*(0)}_{t,1} &\sim \mathcal{N}\left(\mu = 1, \sigma^2 = 0.01\right) \\
\hat{Y}^{*(0)}_{t,t^*} &\sim \mathcal{N}\left(\mu = 0.9 Y^{*(0)}_{t,t^*-100} + 0.1, \sigma^2 = 0.01\right), && t^* \in \{101, 201, \ldots, 100 * \lfloor Y_t/2 \rfloor + 1\} \\
\hat{Y}^{*(n)}_{t,t^*} &= a^{(n)}_{t,t^*} i \sqrt{b^{(n)}_{t,t^*}}, && n \in \{1, \ldots, 49\}
\end{aligned}
$$

$$
\begin{aligned}
a^{(n)}_{t,t^*} &\sim \begin{cases} -1 & w.p.\ 1/2 \\ 1 & w.p.\ 1/2 \end{cases} \\
(b^{(n)}_{t,t^*} | X^*_{t,t^*} = 1) &\sim Gamma(1/n^2, 1) \\
(b^{(n)}_{t,t^*} | X^*_{t,t^*} = 2) &\sim \begin{cases} Gamma(1/n^2, 1) & \text{for } n \neq 2 \\ Gamma(100, 1) & \text{for } n = 2 \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
\hat{Y}^{*(50)}_{t,t^*} &= 0 \\
\hat{Y}^{*(n)}_{t,t^*} &= -\hat{Y}^{*(100-n)}_{t,t^*}, && n \in \{51, \ldots, 99\}
\end{aligned}
$$

Finally, $Y^*_{t,t^*:t^*+99}$ is set using the inverse discrete fourier transform of $\hat{Y}^*_{t,t^*}$:

$$
Y^*_{t,t^*:t^*+99} = IDFT\left(\hat{Y}^*_{t,t^*}\right)
$$

This gives a strong periodic component to acceleration when the subdive state $X^*_{t,t^*} = 2$. There are several practical reasons behind this construction:

(a) $\hat{Y}^*_{t,t^*}$ is anti-symetric about $\hat{Y}^{*(50)}_{t,t^*}$ so that its inverse fourier transform is real-valued.

(b) $\hat{Y}^{*(n)}_{t,t^*}$ decays like $1/n^2$ to facilitate continutity.

Note that this process idoes not result in a continuous sequence Y^*_t since the average values of consecutive 2-second segments jump. However, note that these average values are highly autocorrelated, so the jumps are not too severe. See figure (7) for details. In addition, Y^*_t has the following desirable properties:

1. $Z^{*(1)}_{t,t^*+1} | Z^{*(1)}_{t,t^*} \sim \mathcal{N}\left(\mu = 0.9 Z^{*(1)}_{t,t^*} + 0.1, \sigma^2 = 0.01\right)$
2. $$
Z^{*(2)}_{t,t^*} \sim \begin{cases} \text{Gamma}(\sum_{n=1}^{\tilde{w}} \frac{1}{n^2}, 1) & \text{for } X^*_{t,t^*} = 1 \\ \text{Gamma}(20 + \sum_{n=2}^{\tilde{w}} \frac{1}{n^2}, 1) & \text{for } X^*_{t,t^*} = 2 \end{cases}
$$

So we can directly compare the results of the CarHHMM with the ground truth.

I still need to put in the actually results from the simulation study- I had to rethink how to do the simulation in the first place since enforcing continuity was REALLY hard to do in a principled way.

5 Results

Results will go here. I have some stuff on CarHMMs and HHMMs from STAT 548 but I need to put it together still.

6 Discussion

THE FOLLOWING IS FROM THE STAT 548 PAPER:

While incorporating autocorrelation within a hidden Markov model to analyze animal movement is not new, Lawler et al. introduce a new formulation of this model in the CarHMM. They also review several useful preprocessing tools such as the lag plot and a method for interpolation that includes dividing the data into separate groups if observations are far apart.

This paper summarizes the CarHMM and provides intuition behind the interpolation scheme laid out by Lawler et al. In addition, it shows that if the emission distributions are normal for the sequence of step-sizes $\mathbf{D}$, then the CarHMM models $\mathbf{D}$ as a one dimensional Ornstein-Uhlenbeck process.

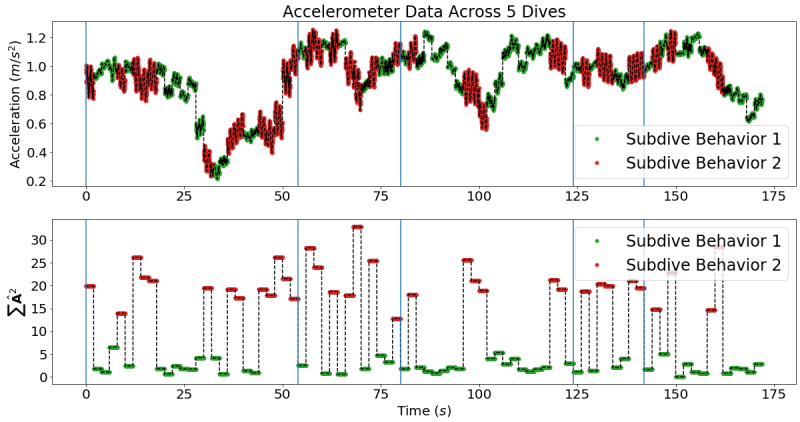

Figure 7: Simulated Acceleration Data.

Finally, the CarHMM is adapted to dive data and used to model the behavior of a killer whale off the coast of British Columbia, Canada. In particular, that whale exhibited auto-correlated velocities in the z-direction as well as sinusoidal behavior in dynamic body acceleration. The adjusted CarHMM is able to capture within-dive behaviors of active swimming and passive gliding, but more analysis is necessary to determine the exact number of within-dive behaviors and generalize the method to a larger number of dive types.

References

[1] Timo Adam, Christopher Griffiths, Vianey Leos Barajas, Emily Meese, Christopher Lowe, Paul Blackwell, David Righton, and Roland Langrock. Joint modelling of multi-scale animal movement data using hierarchical hidden markov models. *Methods in Ecology and Evolution*, 10, 06 2019.

[2] Sonali Bagchi and Sanjit Mitra. The nonuniform discrete fourier transform. 01 2001.

[3] Roland Langrock, Ruth King, Jason Matthiopoulos, Len Thomas, Daniel Fortin, and Juan Morales. Flexible and practical modeling of animal telemetry data: Hidden markov models and extensions. *Ecology*, 93:2336–42, 11 2012.

[4] Ethan Lawler, Kim Whoriskey, William Aeberhard, Chris Field, and Joanna Flemming. The conditionally autoregressive hidden markov model (carhmm): Inferring behavioural states from animal tracking data exhibiting conditional autocorrelation. *Journal of Agricultural, Biological and Environmental Statistics*, 05 2019.

[5] Thèo Michelot and Paul Blackwell. State-switching continuous-time correlated random walks. *Methods in Ecology and Evolution*, 01 2019.

[6] A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, pages 260–269, April 1967.

Appendix

A Dimension Reduction on $\hat{Y}_t^*$

First, we simply use down-sampling and only record every w^{th} time step of $\hat{Y}_t^*$. This both reduces the space of $\hat{Y}_t^*$ to $\mathbb{C}^{\lfloor S_t^*/w \rfloor \times w}$ and ensures that none of the sliding windows overlap when taking the STFT. This is important because HMMs assume temporal independence between observations. Next, the dimension of $\hat{Y}_t^*$ can be cut in half by recognizing that the Y_t^* is real-valued, and therefore $\hat{Y}_{t,k}^*$ is equal to the complex-congugate of $\hat{Y}_{t,w-k-1}^*$. Finally, by Parseval's Thereom we have that:

$$\sum_{n=0}^{w-1} |Y_{t+n}^*|^2 = \sum_{n=0}^{w-1} |\hat{Y}_{t,n}^*|^2$$

B Equivalency of CarHMM and one-dimensional state-switching Ornstein-Uhlenbeck process

If it is the case that (1) the underlying behavioral state of the continuous-time model must follow a Markov chain rather than a Markov process, and (2) the emission distributions of the CarHMM are gaussian, then the CarHMM and the state-switching continuous model are equivalent. This allows the theoretically grounded continuous-time state-switching model to be used in the computational convieneint HMM (and therefore HHMM) framework. In addition, it gives new intpretation to the learned parameters of the CarHMM in the context of an Ornstein-Uhlenbeck process.

A one-dimensional state-switching Ornstein-Uhlenbeck process y^* is the solution to the following stochastic differential equation:

$$dy_t^* = \beta_{x_t^*}(\gamma_{x_t^*} - y_t^*)dt + \omega_{x_t^*} dW_t$$

where x_t^* is the fine-scale behavior of the animal at time t, $\beta_{x_t^*}$ relates to rate at which the process returns to its mean value, $\gamma_{x_t^*}$ is the long-term mean value of the process, $\omega_{x_t^*}$ is related to short-term variance, and W is a Wiener process. As before, x_t^* is described by an unobserved Markov process. The solution to this equation is known to be the following [5]:

$$y_{t+\delta}^* \sim \mathcal{N}\left((1 - e^{-\beta_{x_t^*}\delta})\gamma_{x_t^*} + e^{-\beta_{x_t^*}\delta} y_t^*, \quad \frac{\omega_{x_t^*}^2}{2\beta_{x_t^*}}(1 - e^{-2\beta_{x_t^*}\delta})\right)$$

Now, suppose that δ is constant for all observations, as is the case for hidden Markov models. In addition, introduce the following transformations:

$$\mu_{x_t^*} = \gamma_{x_t^*}, \qquad \phi_{x_t^*} = e^{-\beta_{x_t^*}\delta}, \qquad \sigma_{x_t^*}^2 = \frac{\omega_{x_t^*}^2}{2\beta_{x_t^*}}(1 - e^{-2\beta_{x_t^*}\delta})$$

Then, we have the following:

$$y_{t+\delta}^* \sim \mathcal{N}\left((1 - \phi_{x_t^*})\mu_{x_t^*} + \phi_{x_t^*} y_t^*, \quad \sigma_{x_t^*}^2\right)$$

If δ is fixed and x_t^* is adjusted to follow a Markov chain rather than a Markov process, then this model is equivalent to the CarHMM with normal emission probabilities. Note that all of the parameter transformations above are one-to-one, so it is easy to go from the CarHMM to the continuous model and back again. This allows for the pricipled construction of the continuous-time model to be combined with the computational convienence of the CarHMM.