

**Incorporating multi-scale structures and physiological processes into the modeling of
animal movement**

by

Vianey Caroline Leos Barajas

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Mark S. Kaiser, Major Professor
Petruța Caragea
Philip Dixon
Tracy Heath
Jarad Niemi

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

Copyright © Vianey Caroline Leos Barajas, 2019. All rights reserved.

DEDICATION

To my Latino/a professors:

To say I wouldn't have made it here without you is no understatement. You believed in me every step of the way, you took me under your wing, you provided opportunities so that I could flourish.

Importantly, you understand that these spaces are not made for people like us to succeed and still managed to provide a pathway for me to follow in your footsteps. You reach out to the next generation, you are active in your encouragement, you believe in us. You bring your full selves everyday, even when that means others criticize the way you speak, the way you act, your culture. You cause the good kind of trouble, even when there are consequences. You fight so that the next generation won't have it as hard as you had it. You're fearless.

I'll do my best to live up to your example and make it easier on the next generation. If I manage to cause a fraction of the trouble you all have, I'll know I have succeeded. Para la cultura. Gracias Javier Trigos-Arrieta, Eduardo Montoya y Alicia Carriquiry.

To my friends and to my family, many of which are both to me. To my peeps. And to my wonderfully goofy cats, Strax and Gaara. Thank you and love you.

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
LIST OF FIGURES	vi
ACKNOWLEDGMENTS	ix
ABSTRACT	x
CHAPTER 1. INTRODUCTION	1
1.1 Multi-scale Analyses	1
1.2 Incorporating Physiological Dynamics	2
1.3 References	3
CHAPTER 2. MULTI-SCALE MODELING OF ANIMAL MOVEMENT AND GENERAL BEHAVIOR DATA USING HIDDEN MARKOV MODELS WITH HIERARCHICAL STRUCTURES	5
2.1 Abstract	5
2.2 Introduction	6
2.3 Hidden Markov Models with Hierarchical Structures	8
2.3.1 Basic HMM Framework	8
2.3.2 Extension to allow for hierarchical structures	10
2.4 Applications	12
2.4.1 Harbor Porpoises	12
2.4.2 Garter Snakes	18
2.5 Discussion	22
2.6 References	25
CHAPTER 3. HIDDEN MARKOV MODELS WITH MULTI-SCALE STATE PROCESSES IN MOVEMENT ECOLOGY	29
3.1 Abstract	29
3.2 Introduction	30
3.3 Hidden Markov Models	31
3.4 Multi-Scale State Processes	32
3.4.1 Equal-length internal state duration times	33
3.4.2 Variable-length internal state duration times	36
3.4.3 State-Dependent Distributions	37
3.5 Inference	38
3.6 Diving With a Tiger Shark	39

3.6.1	A day in the life	40
3.6.2	Up and Down	44
3.7	Discussion	47
3.8	References	49
3.9	Appendix	51
3.9.1	Tuning parameter for internal state duration	51
3.9.2	Section 3.6.1 Tables and Plots	53
3.9.3	Section 3.6.2 Tables and Plots	57
CHAPTER 4. INCORPORATING BODY CONDITION INTO THE ANALYSIS OF ANIMAL MOVEMENT		
4.1	Abstract	61
4.2	Introduction	62
4.3	Data Structure	63
4.4	Model	64
4.4.1	Movement	64
4.4.2	Body Condition (% of fat)	66
4.4.3	Joint model	69
4.5	Fitting the Joint Model	70
4.5.1	Likelihood Evaluation	70
4.5.2	Bayesian inference	72
4.6	Simulation	73
4.6.1	Movement Model Results	75
4.6.2	Predicted Condition Results	75
4.7	Discussion	76
4.8	References	78
4.9	Appendix	80
4.9.1	Posterior Predictive Checks	80
4.9.2	HMC Posterior Draws	84
CHAPTER 5. FUTURE WORK SUMMARY AND DISCUSSION		
5.1	Multi-scale animal behavior	88
5.2	Physiology and Movement	89
5.3	References	90

LIST OF TABLES

	Page
Table 3.1	Mapping indices from the original temporal scale to indexing via segments. 33
Table 3.2	Means and 95% credible intervals of production state-dependent distribution parameter estimates. 41
Table 3.3	Marginal means and 95% credible intervals for the production state-dependent distributions for each internal state. 45
Table 3.4	Results of HMC posterior draws for model in Section 3.6.1. Values for the marginal means and chosen quantiles are reported, along with the effective number of samples, n_{eff} , and value of \hat{R} 56
Table 3.5	Results of HMC posterior draws for model in Section 3.6.2. Values for the marginal means and chosen quantiles are reported, along with the effective number of samples, n_{eff} , and value of \hat{R} 60
Table 4.1	Models fitted to the simulated data. 73
Table 4.2	Prior distributions for the parameters associated with the HMM across all four models. 75
Table 4.3	HMC Posterior Draws for Model 1. Values for the marginal means and chosen quantiles are reported, along with the effective number of samples, n_{eff} , and value of \hat{R} 84
Table 4.4	HMC Posterior Draws for Model 2. Values for the marginal means and chosen quantiles are reported, along with the effective number of samples, n_{eff} , and value of \hat{R} 85
Table 4.5	HMC Posterior Draws for Model 3. Values for the marginal means and chosen quantiles are reported, along with the effective number of samples, n_{eff} , and value of \hat{R} 86
Table 4.6	HMC Posterior Draws for Model 4. Values for the marginal means and chosen quantiles are reported, along with the effective number of samples, n_{eff} , and value of \hat{R} 87

LIST OF FIGURES

	Page
Figure 2.1 Dependence structure in hierarchically structured HMMs.	11
Figure 2.2 Fitted state-dependent distributions for the dive duration, the maximum depth and the dive wigginess, the latter together with the estimated point mass on zero.	15
Figure 2.3 Exemplary sequence of the first 25% of the observations of the dive duration, the maximum depth and the dive wigginess. Hourly segments are indicated by vertical grey lines. The states were decoded using the Viterbi algorithm, where the colors refer to the production states.	17
Figure 2.4 Fitted state-dependent distributions for distance traveled.	20
Figure 2.5 Internal state decodings by garter snake. Each garter snake is from one of two ecotypes: Lakeshore or Meadow.	21
Figure 3.1 Comparing activity levels across days reflective of internal state 1 or 2. Top row: the local state probabilities of production state 1 given internal state 1 for day 20 and given internal state 2 for day 4. The black line denotes the median probabilities, the purple shaded area spans the 25 th and 75 th quantile, while the gray spans the 2.5 th and 97.5 th quantile. Bottom row: The value of 2 times the absolute change in depth colored by the median local state probability of production state 1 for day 20 and day 4.	42
Figure 3.2 The internal state probability, $P(H_k = 1)$, across the 21 days. The black line denotes the mean probabilities, the purple shaded area spans the 25 th and 75 th quantile, while the gray spans the 2.5 th and 97.5 th quantile.	42
Figure 3.3 Comparing activity levels across days reflective of internal state 1 or 2. Top row: the local state probabilities of production state 1 given internal state 1 for day 20 and given internal state 2 for day 4. The black line denotes the median probabilities, the purple shaded area spans the 25 th and 75 th quantile, while the gray spans the 2.5 th and 97.5 th quantile. Bottom row: The value of 2 times the absolute change in depth colored by the median local state probability of production state 1 for day 20 and day 4.	43
Figure 3.4 Top row: Point-wise credible intervals for the production state-dependent distributions for each internal state. Bottom row: Internal state-dependent distributions constructed as a weighted sum of the production state-dependent distributions.	46
Figure 3.5 Top row: Mean probability of internal state 1 along with point-wise credible intervals. Bottom row: A subset of the dive profile of the tiger shark colored by the mean probability of internal state 1.	47

Figure 3.6	Histogram of observed values (in red) along with 95% pointwise posterior predictive intervals for observations $y = 0, 1, 2, \dots, 100$	53
Figure 3.7	Distributions of the quantiles (in black) alongside the quantile of the observations (in purple).	54
Figure 3.8	Autocorrelation structure of posterior predictive distributions (in purple) alongside the autocorrelation structure of the observation process (in black).	54
Figure 3.9	Top: QQ-plot of the residuals evaluated at the estimated marginal means of the parameters. Bottom: Autocorrelation structure of the residuals evaluated at the estimated marginal means of the parameters.	55
Figure 3.10	Histogram of observed values (in red) along with 95% pointwise posterior predictive intervals for observations $y = -100, -99, \dots, 99, 100$	57
Figure 3.11	Distributions of the quantiles (in black) alongside the quantile of the observations (in purple).	58
Figure 3.12	Autocorrelation structure of posterior predictive distributions (in purple) alongside the autocorrelation structure of the observation process (in black).	58
Figure 3.13	Top: QQ-plot of the residuals evaluated at the estimated marginal means of the parameters. Bottom: Autocorrelation structure of the residuals evaluated at the estimated marginal means of the parameters.	59
Figure 4.1	Diagram of the joint model for condition and behavior, with the dashed circle representing a predicted value of the condition process.	69
Figure 4.2	Blue line: simulated condition data. Green segments: processed condition data for Model 3.	74
Figure 4.3	Mean t.p.m. under the four possible model constructions, along with 95% credible intervals.	76
Figure 4.4	Top: Predicted body fat percentages compared to the true body fat percentage, taken to be a value between 0 and 100, along with 95% credible intervals. Bottom: 95% credible intervals of distribution of differences between true body fat percentage and predicted body fat percentage. Dashed lines: Days in which body fat percentage is directly observed.	77
Figure 4.5	Top: Posterior predictive step length distributions for model 1 (in light blue) and the marginal distribution of the observed data (in black). Middle: Posterior predictive turning angle distributions for model 1 (in light blue) and the marginal distribution of the observed data (in black). Bottom: Posterior predictive autocorrelation function for model 1 (in light blue) and the observed autocorrelation function of the observed data (in black).	80
Figure 4.6	Top: Posterior predictive step length distributions for model 2 (in light blue) and the marginal distribution of the observed data (in black). Middle: Posterior predictive turning angle distributions for model 2 (in light blue) and the marginal distribution of the observed data (in black). Bottom: Posterior predictive autocorrelation function for model 2 (in light blue) and the observed autocorrelation function of the observed data (in black).	81

Figure 4.7	Top: Posterior predictive step length distributions for model 3 (in light blue) and the marginal distribution of the observed data (in black). Middle: Posterior predictive turning angle distributions for model 3 (in light blue) and the marginal distribution of the observed data (in black). Bottom: Posterior predictive autocorrelation function for model 3 (in light blue) and the observed autocorrelation function of the observed data (in black).	82
Figure 4.8	Top: Posterior predictive step length distributions for model 4 (in light blue) and the marginal distribution of the observed data (in black). Middle: Posterior predictive turning angle distributions for model 4 (in light blue) and the marginal distribution of the observed data (in black). Bottom: Posterior predictive autocorrelation function for model 4 (in light blue) and the observed autocorrelation function of the observed data (in black).	83

ACKNOWLEDGMENTS

I would like to thank my adviser, Mark Kaiser, for allowing me the freedom to explore many avenues of research and his guidance. Thank you to my many collaborators. Thank you Juan Morales for sharing your eternal wisdom and mentoring.

Without many, my dreams of working on projects related to animal movement, specifically shark movement, would not have come true. Thank you.

ABSTRACT

HMMs are commonly used to model animal movement data and infer aspects of animal behavior. Their ability to connect an observation process to an underlying state process, generally serving as a proxy for a finite set of animal behaviors of interest, matches the intuition that the observed movements stem from an underlying (unobserved) behavioral process. We can further extend the HMM framework to consist of multiple state processes to reflect that different behaviors are identified by different compositions of the observed movement processes. We refer to this extension as a multi-scale HMM whereby one state process is connected to the underlying behaviors that generate the movements at the temporal scale at which the data are processed and another is connected to a larger-scale behavioral process, defined as a composition of fine-scale behavioral states. We present two formulations of the multi-scale HMM. We illustrate the application of multi-scale HMMs in four real-data examples, vertical movements of harbor porpoises observed in the field, and garter snake movement data collected as part of an experimental design, in chapter 2 and under two different formulations applied to tiger shark data in chapter 3.

HMMs again play a feature role in chapter 4, where we aim to connect movement and physiology dynamics and their evolution and interaction over time. A long-sought goal in ecology is to connect movement with population dynamics. For many species and especially for ungulates, there is a known link between condition (e.g. fat reserves) and the probability of survival and reproduction. Assuming a particular genetic makeup and physiology, condition reflects the history of behavioral decisions, including movement and habitat use. However, the condition of an animal can also have a direct implication on the types of movements that it performs and the habitats that it visits. Movement data for ungulates are typically collected at a fine temporal scale, e.g. a position recorded by a GPS device every five or ten minutes. However, fat reserves cannot be measured remotely and must be done manually. This in turn creates a mismatch in the temporal scale at

which the two data streams are observed, i.e. every five minutes for movement vs approximately once a month for condition. Further, the temporal mismatch leads to various challenges when jointly modeling the two processes. For the movement model, we use discrete-time, finite-state HMMs with the positional data of the sheep serving as the observation process and the underlying state process serving as a proxy for behaviors of interest. To incorporate condition as a potential covariate affecting the movement, and thus behavioral, process, we make use of the physiological equations that describe the evolution of body fat in Merino sheep in order to predict daily values of the condition process. The physiological equations are expressed as a function of the states inferred by HMM, as well as the distance that the sheep travels.

CHAPTER 1. INTRODUCTION

The work conducted in this thesis lies squarely in the area of statistical ecology, specifically in the area of modeling time series of animal movement data. Due to technological advances in devices that can be affixed to animals, large amounts of data can be collected on a wide range of animals, across marine, terrestrial and aerial environments. As the data reveal patterns and movements with much higher detail than ever before possible, the number of research questions that can be asked and informed by the data is growing rapidly. Common statistical approaches applied in the analysis of animal movement data aim to answer two main questions, “What types of behaviors did the animal(s) exhibit?” and “What are the primary drivers of these behaviors?” (Patterson et al., 2009; Hooten et al., 2017). In this spirit, we delve into the manners in which we can define and identify animal behaviors at multiple temporal scales and how we can incorporate pertinent drivers of these behaviors. We further incorporate physiological processes into the analysis of animal movement, in order to construct a framework in which behavior and physiology interact and evolve simultaneously (Hooten et al., 2019).

1.1 Multi-scale Analyses

State-space models are a common class of time series models applied to animal movement data (Patterson et al., 2008, 2017; Hooten et al., 2017). Their ability to connect an observation process to an underlying state process, generally serving as a proxy for a finite set of animal behaviors of interest, matches the intuition that the observed movements stem from an underlying (unobserved) behavioral process. When applied to positional data, state-space models are commonly used to differentiate between movements encompassing traveling, area-restricted search and resting behaviors (Morales et al., 2004). The signals emitted from these types of behaviors are easily observed in the data as directional, long distances traveled (traveling), short distances and high turning

angles (area-restricted search) and short or zero distances (resting) over some (\sim 5–60 m) window of time. Other common applications include analysis of marine mammal dive data and analysis of accelerometer data. In these applications, we again have clear signals that can be observed in the data. For instance, differentiating shallow dives from deep dives (DeRuiter et al., 2017) to make inferences about the occurrence of non-foraging and foraging events and large spikes in dynamic acceleration signals from small signals, close to zero, to infer activity levels in sharks (Leos-Barajas et al., 2017). While the simplistic approach to identifying different signals in the movement data already provides a wealth of information to the understanding of an animal’s behaviors and its drivers, it has clear limitations in the types of behaviors that can be represented.

Hidden Markov models are a particular class of state-space models with a discrete state-space and are popular time series models commonly applied to time series of animal movement data (Zucchini et al., 2016). Chapters 2 and 3 focus on extending the structure of hidden Markov models (HMMs) in order to capture a larger variety of behavioral patterns and provide a framework in which to make inferences about drivers of behavior at multiple temporal scales.

1.2 Incorporating Physiological Dynamics

While chapters 2 and 3 focus on the principal aim of extending the types of behaviors that are able to be captured via the general class of HMMs, chapter 4 focuses on the manner in which body condition of the animal could be included as a covariate in the analysis of animal movement. Underlying an animal’s movement patterns and behavioral responses lies the physiological condition of the animal and its abilities to perform biologically necessary activities. However, a large hurdle in this approach is that condition data requires manual collection. In free ranging animals, condition data may only be collected when the animal is re-captured, making it difficult to collect condition data at a fine temporal scale.

As a case study, we present an approach to incorporate body condition into the analysis of Merino sheep movement (Wilmshurst et al., 2000; Delgiudice et al., 2001). To obtain values of condition on a daily scale, we use a combination of observed condition values (approximately

obtained every 30 days) along with physiological equations specific to Merino sheep to predict values of body condition when they are not directly observed. We use HMMs for the movement model.

One particular challenge associated with this approach lies in the feedback between condition and movement assumed in the modeling framework. Movement is directly influenced by the underlying condition of the sheep while predictions of condition inherently depend on results of the movement model, such as identifying periods of foraging behavior. We present a joint model that models animal movement, includes body condition as a potential driver of behavior and simultaneously predicts the values of condition when not observed for inclusion in the movement model.

1.3 References

- Delgiudice, G. D., Moen, R. A., Singer, F. J., and Riggs, M. R. (2001). Winter nutritional restriction and simulated body condition of yellowstone elk and bison before and after the fires of 1988. *Wildlife Monographs*, 147:1–60.
- DeRuiter, S. L., Langrock, R., Skirbutas, T., Goldbogen, J. A., Calambokidis, J., Friedlaender, A. S., and Southall, B. L. (2017). A multivariate mixed hidden markov model for blue whale behaviour and responses to sound exposure. *The Annals of Applied Statistics*, 11:362–392.
- Hooten, M. B., Johnson, D. S., McClintock, B. T., and Morales, J. M. (2017). *Animal Movement: Statistical Models for Telemetry Data*. CRC Press, Boca Raton, FL.
- Hooten, M. B., Scharf, H. R., and Morales, J. M. (2019). Running on empty: recharge dynamics from animal movement data. *Ecology Letters*, 22(2):377–389.
- Leos-Barajas, V., Gangloff, E., Adam, T., Langrock, R., van Beest, F., Nabe-Nielsen, J., and Morales, J. M. (2017). Multi-scale modeling of animal movement and general behavior data using hidden markov models with hierarchical structures. *JABES*, 22:232–248.
- Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. (2004). Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, 85(9):2436–2445.
- Patterson, T. A., Basson, M., Bravington, M. V., and Gunn, J. S. (2009). Classifying movement behaviour in relation to environmental conditions using hidden markov models. *Journal of Animal Ecology*, 78(6):1113–1123.

- Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., and King, R. (2017). Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis*, 101(4):399–438.
- Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. (2008). State-space models of individual animal movement. *Trends in Ecology and Evolution*, 23(2):87–94.
- Wilmhurst, J. F., Fryxell, J. M., and Berman, C. M. (2000). The allometry of patch selection in ruminants. *Proceedings of the Royal Society of London, Series B*, 267:345–349.
- Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction using R*. Chapman & Hall, Boca Raton, FL, 2nd edition.

CHAPTER 2. MULTI-SCALE MODELING OF ANIMAL MOVEMENT AND GENERAL BEHAVIOR DATA USING HIDDEN MARKOV MODELS WITH HIERARCHICAL STRUCTURES

A paper accepted by *Journal of Agricultural, Biological and Environmental Statistics*

Vianey Leos-Barajas^{1*}, Eric J. Gangloff¹, Timo Adam², Roland Langrock², Floris M. van Beest³,
Jacob Nabe-Nielsen³ and Juan M. Morales⁴

¹Iowa State University, USA

²Bielefeld University, Germany

³Aarhus University, Denmark

⁴INIBIOMA-CONICET, Argentina

2.1 Abstract

Hidden Markov models (HMMs) are commonly used to model animal movement data and infer aspects of animal behavior. An HMM assumes that each data point from a time series of observations stems from one of N possible states. The states are loosely connected to behavioral modes that manifest themselves at the temporal resolution at which observations are made. Due to advances in tag technology and tracking with digital videorecordings, data can be collected at increasingly fine temporal resolutions. Yet, inferences at time scales cruder than those at which data are collected, and which correspond to larger-scale behavioral processes, are not yet answered via HMMs. We include additional hierarchical structures to the basic HMM framework, incorporating multiple Markov chains at various time scales. The hierarchically structured HMMs allow for behavioral inferences at multiple time scales and can also serve as a means to avoid coarsening data. Our proposed framework is one of the first that models animal behavior simultaneously at multiple time scales, opening new possibilities in the area of animal movement and behavior modeling. We illustrate the application of hierarchically structured HMMs in two real-data examples: (i) vertical

movements of harbor porpoises observed in the field, and (ii) garter snake movement data collected as part of an experimental design.

2.2 Introduction

Hidden Markov models (HMMs) and related state-switching models are prevalent in the field of animal movement modeling, where they provide a flexible framework to infer aspects of animal behavior from various types of movement data (Morales et al., 2004; Patterson et al., 2009, 2017; Langrock et al., 2012, 2014). They are very natural models for time series data related to animal movement, as they account for the serial dependence typically observed and allow each observation to be (loosely) connected to distinct underlying behavioral modes. A basic HMM for movement data consists of two stochastic processes: an observed movement process and an underlying state process, the latter of which can be related to distinct behavioral modes, at least in the sense of serving as a proxy of the actual behavioral process (Patterson et al., 2009; Langrock et al., 2012). Applications of HMMs to movement data often focus on investigating the effect of individual and environmental covariates on state occupancy, and thus ultimately on the dynamics of the variation in behavioral modes in response to internal and external drivers.

Generally, movement data are analyzed such that the observation process is assumed to stem from a single (behavioral) state process. It may however be the case that there are two (connected) behavioral processes that occur at distinct time scales. For instance, so-called hierarchical HMMs have been used to process data on handwriting in order to distinguish between distinct letters but also to recognize a word, defined as a sequence of written letters (Fine et al., 1998). However, these versatile extensions of HMMs have not yet been applied to movement data, even in light of the intuitive idea that distinct behaviors manifest themselves at different time scales (hereafter referred to as multi-scale behaviors). A motivating example to have in mind is a central-place forager such as the southern elephant seal. These animals exhibit large-scale migration movements (from land colonies to either the sea ice zone around Antarctica or into open-ocean pelagic zones, and back), but also movement patterns where much more frequent changes take place between behavioral

modes, e.g. “foraging” and “resting” modes (Biuw et al., 2007; Hindell et al., 2016; Michelot et al., 2017). The modeling framework we propose regards such data as stemming from two behavioral processes, which operate on different time scales: the first process determines the behavioral mode at the cruder time scale (e.g. whether or not an elephant seal is performing a migratory trip, and also what kind of migratory trip), while the second process, at the finer time scale, determines the behavioral mode *nested within* the large-scale mode (e.g. whether an elephant seal is resting or foraging, given it is close to the sea ice zone, or whether it is traveling or foraging, given it is on a migratory trip).

For multi-scale modeling of animal movement data, we propose an extension to the standard HMM that allows for a hierarchical state process, where two (or more) different Markov chains, operating at different time scales, will be tied together. To illustrate the application of hierarchically structured HMMs in a real-data setting, we model vertical movements of a harbor porpoise (*Phocoena phocoena*) throughout its natural habitat in the northeastern part of the North Sea. While the data were collected at a dive-by-dive resolution, the aim here is to infer dive patterns at two different temporal scales: an hourly scale to infer the general behavioral mode (e.g. resting or traveling), which may persist for a large number of consecutive dives, and a fine-scale process to infer more nuanced state transitions at a dive-by-dive resolution given the general behavioral mode. As a second real-data example, we model baby garter snake (*Thamnophis elegans*) movement data produced in a controlled experimental design context. This experiment includes tracking individuals over two segments of a behavioral trial repeated three times, resulting in six discrete time series produced per individual. The hierarchically structured HMM here has two Markov chains, where one Markov chain models the transitions among three types of movements (distance traveled in 1/2 s) and the second Markov chain models transitions across six time series produced per snake. This subset of individuals served as the control group for the larger experiment and therefore did not receive any additional experimental stimuli. Thus, we use the second Markov chain here to investigate personality and repeatability in their movement patterns. That is, we attempt to answer

if individual garter snakes differ in their general movement strategies or if they have tendencies to exhibit the same general movement pattern across multiple time series.

A conceptual challenge with HMMs, and in fact any discrete-time models for behavioral data, is that the temporal resolution of the observations being analyzed (e.g. hourly, daily, etc.) determines what kind of behaviors may be inferred at all. Strictly speaking, this is not a problem arising from the model applied, but rather from the sampling protocol, i.e. the data. For instance, Towner et al. (2016) processed white shark location data, collected every five minutes, into distance traveled and turning angle and subsequently connected each bivariate observation to “area-restricted search” and “transiting” behavior. Were the shark’s location observed once per day, we would not be able to infer the same behaviors because switches between these behavioral modes occur at a much finer temporal scale. The hierarchically structured HMMs will not solve the conceptual challenges associated with data processing or data collection required to infer multi-scale behaviors. However, it does offer new opportunities in the analysis of animal movement data, allowing for identification of general behavioral patterns that are a composition of fine-scale observations and inferences to be made at multiple time scales.

2.3 Hidden Markov Models with Hierarchical Structures

In Section 2.3.1 we first detail the basic HMM framework in order to introduce the necessary notation that will be used throughout the paper. In Section 2.3.2, we introduce the hierarchical model formulation, distinguishing between two types of latent states, production states and internal states, which occur at distinct time scales.

2.3.1 Basic HMM Framework

A basic HMM is composed of two stochastic processes: an observable state-dependent process $\{Y_t\}_{t=1}^T$ and an unobservable state process $\{S_t\}_{t=1}^T$ taking on a finite number of states. Here we call the state a *production state* (as it produces an observation), in order to differentiate it from other forms of the latent states which we introduce in Section 2.3.2. As is general practice, we assume

a first-order Markov process at the production state level, such that the distribution of S_t , the production state at time t , is completely determined by the previous state S_{t-1} . We further assume Y_t , $t = 1, \dots, T$, to be conditionally independent of past and future observations and production states, given the production state S_t , such that the production states effectively select from which of finitely many possible distributions each observation is drawn. Due to the Markov property, the evolution of the production states over time is governed by the transition probability matrix (t.p.m.), $\boldsymbol{\Gamma} = (\gamma_{ij})$, where $\gamma_{ij} = \Pr(S_t = j | S_{t-1} = i)$ for $i, j = 1, \dots, N$, with N denoting the number of production states. The initial distribution, $\boldsymbol{\delta}$, is a vector of probabilities with entries $\delta_i = \Pr(S_1 = i)$, of the first observation y_1 belonging to one of the N production states. It is common to assume the initial distribution to be the stationary distribution, defined as the solution to $\boldsymbol{\Gamma}\boldsymbol{\delta} = \boldsymbol{\Gamma}$. However, $\boldsymbol{\delta}$ can also be estimated. In order to ensure identifiability when estimating the entries of the t.p.m., we map the entries of each row onto the real line with the use of the multinomial logit link and set the diagonal entries of the matrix as the reference categories:

$$\gamma_{ij} = \frac{\exp(\eta_{ij})}{\sum_{k=1}^N \exp(\eta_{ik})}, \quad \text{where} \quad \eta_{ij} = \begin{cases} \beta^{(ij)} & \text{if } i \neq j; \\ 0 & \text{otherwise.} \end{cases}$$

We similarly use a multinomial logit link transformation for the initial distribution, if estimated rather than assumed to be the stationary distribution. The state-dependent distributions for Y_t will be represented in terms of probability density or mass functions $f(y_t | S_t = i) = f_i(y_t)$; $i = 1, \dots, N$. If the observations are multivariate, in which case we write $\mathbf{Y}_t = (Y_{1t}, \dots, Y_{Rt})$, we can either formulate a joint distribution $f_i(\mathbf{y}_t)$ or assume contemporaneous conditional independence by allowing the joint distribution to be represented as a product of marginal densities, $f_i(\mathbf{y}_t) = f_i^1(y_{1t})f_i^2(y_{2t}) \cdots f_i^R(y_{Rt})$. While parametric families are usually chosen for the f_i , such as a Gaussian or gamma distribution, we can also estimate the distribution nonparametrically by expressing it as a linear combination of a large number of basis functions (Langrock et al., 2015). The likelihood of an individual time series can be expressed concisely as a matrix product,

$$L_p(y_1, \dots, y_T) = \boldsymbol{\delta}^\top \mathbf{P}(y_1) \prod_{t=2}^T \boldsymbol{\Gamma} \mathbf{P}(y_t) \mathbf{1}, \quad (2.1)$$

where $\mathbf{P}(y_t) = \text{diag}(f_1(y_t), \dots, f_N(y_t))$, $\boldsymbol{\delta}$ is the initial distribution (a column vector), and $\mathbf{1}$ is a column vector of ones. Calculation of the matrix product given above the computational cost of which notably is only linear in T is equivalent to applying the forward algorithm, which is an efficient recursive scheme for calculating the likelihood of an HMM (Zucchini et al., 2016).

2.3.2 Extension to allow for hierarchical structures

The framework for the basic HMM accounts for switches at the production state level. In a movement modeling analysis, the production states are generally thought to be proxies for behavior occurring at the time scale at which the data were collected (or processed). However, as outlined in the elephant seal example in the introduction, production states alone may not be sufficient to encompass complex multi-scale behavioral processes. More specifically, there may be crude-scale behavioral processes (e.g. migration) that manifest themselves as a *sequence* of production states (e.g. resting or foraging) and associated observations. Intuitively, we would then connect a behavior occurring at a cruder time scale to one of K *internal states*, such that each internal state generates a distinct HMM, with the corresponding N *production states* producing the actual observations.

Akin to the basic HMM framework, we can think of a fine-scale sequence of observations, $\mathbf{y}_m = (y_{1,m}, \dots, y_{T,m})$ with one such sequence for each $m = 1, \dots, M$ to be produced by a sequence of production states, $S_{1,m}, \dots, S_{T,m}$, during a given time frame (namely the m^{th} of M time frames). In typical analyses of telemetry data, this component of the model would correspond to behaviors such as resting or foraging, represented by the production states, and the associated observations (e.g. step lengths and turning angles). However, in addition we now assume that the way in which the sequence of production states is generated depends on which of K possible internal states is active during the current (m^{th}) time frame. The length of the sequence of production states produced by the k^{th} internal state can be dictated by the data collection process or imposed by the analysis. The corresponding K -state internal state process, $\{H_m\}_{m=1}^M$, is such that H_m serves as a proxy for a behavior occurring at a cruder time scale, namely throughout the m^{th} time frame.

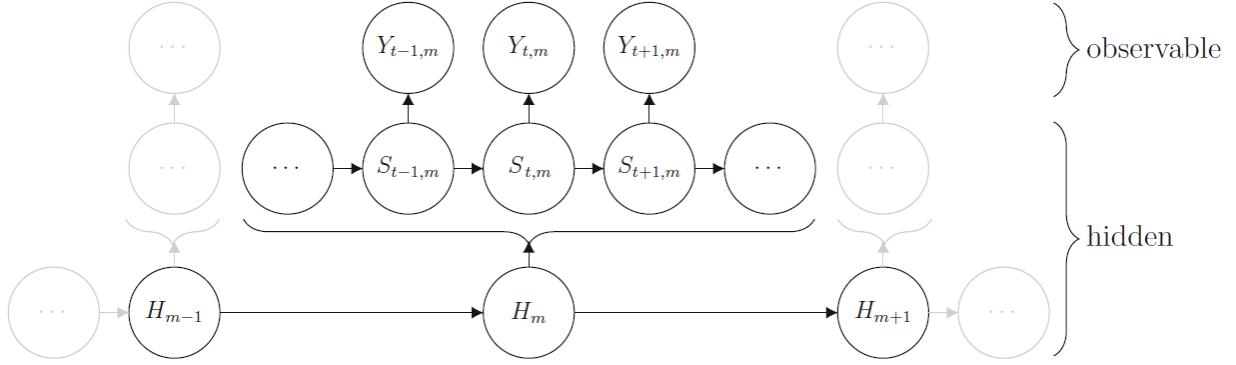


Figure 2.1 Dependence structure in hierarchically structured HMMs.

For example, in the motivating elephant seal example, the internal states could indicate whether or not a seal is on a migratory trip during the m^{th} time frame, for $m = 1, \dots, M$.

With such a hierarchical model formulation, we account for differences observed across the M time frames \mathbf{y}_m , $m = 1, \dots, M$, by connecting each with one of K crude-scale behavioral processes while still modeling the transitions among production states at the time scale at which the data were collected. Supposing that there are multiple time frames per individual, we can model the manner in which an animal switches among the K internal states (behavioral processes). We assume a first-order Markov process at the time frame level, i.e. $\Pr(H_m | H_{m-1}, \dots, H_1) = \Pr(H_m | H_{m-1})$, such that the m^{th} internal state is conditionally independent of all other internal states given the internal state at the $(m - 1)^{\text{th}}$ time point. The $K \times K$ t.p.m. for the internal states $\{H_m\}_{m=1}^M$ examines persistence in the internal states, as well as the manner in which an animal will switch among them. Figure 2.1 displays the dependence structure of hierarchically structured HMMs with two Markov chains, one at the level of the production states, $S_{t,m}$, and the other at the level of the internal states, H_m .

We represent the hierarchical structure as a first-order HMM likelihood, which lends itself to the efficient evaluation known from basic HMMs. To state the likelihood of such a hierarchically

structured HMM, we first define

$$\mathbf{P}^{(I)}(\mathbf{y}_m) = \text{diag}(L_p(\mathbf{y}_m|H_m=1), \dots, L_p(\mathbf{y}_m|H_m=K)),$$

where the likelihoods $L_p(\mathbf{y}_m|H_m=k)$, $k=1, \dots, K$, have the form as given in (2.1), and can vary across k in terms of the production-level t.p.m. associated with the k -th internal state and potentially also the production state-dependent distributions. Then the likelihood for the hierarchically structured HMM is obtained as

$$L_h = \boldsymbol{\delta}^{(I)} \mathbf{P}^{(I)}(\mathbf{y}_1) \prod_{m=2}^M \boldsymbol{\Gamma}^{(I)} \mathbf{P}^{(I)}(\mathbf{y}_m) \mathbf{1},$$

where $\boldsymbol{\delta}^{(I)}$ denotes the vector of length K of initial probabilities for the internal states, and $\boldsymbol{\Gamma}^{(I)}$ denotes the $K \times K$ t.p.m. for the internal state process.

For ease of interpretation, in this work we will assume that the K internal states only vary across the production-level t.p.m.s. As the estimated production states are generally proxies for behaviors, allowing for only the t.p.m. to vary across the K HMMs leads to an interpretation of the K internal states (loosely connected to K behavioral processes) as distinct manners in which an animal will persist and switch among the production states (and hence behaviors). As long as the individual time series' likelihoods, L_p , can be evaluated in an efficient manner, we can evaluate the likelihood of the hierarchically structured HMM via the forward algorithm, and thus maximize it directly, since the general structure does not differ from that of the basic HMM. The Viterbi algorithm can be used for global state decoding, i.e. finding the sequence of the most likely internal and production states, respectively, given the observations.

2.4 Applications

2.4.1 Harbor Porpoises

2.4.1.1 The data

. To illustrate the application of hierarchically structured HMMs, we model vertical movements of a harbor porpoise (*Phocoena phocoena*) throughout its natural habitat in the northeastern part

of the North Sea. From a time-depth recorder (LAT1800ST, Lotek, Ontario, Canada), we obtained observations of the dive depth every second. Assuming a “dive” to be any vertical movement deeper than two meters below the surface, we used the R package `diveMove` (Luque, 2007) to process the raw data into measures of the dive duration, the maximum depth and the dive wigginess (as represented by the absolute vertical distance covered at the bottom of each dive) to characterize the porpoise’s vertical movements at a dive-by-dive resolution. Previous applications of HMMs, though not hierarchically structured HMMs, with dive-by-dive data of marine mammals have been presented in Hart et al. (2010) and DeRuiter et al. (2016). Overall, we consider 275 hours of observations, comprising 7,585 dives in total (hence, about 28 dives per hour).

2.4.1.2 Model formulation and model fitting

. Behavioral modes of marine mammals, e.g. resting, foraging and traveling, do not necessarily manifest themselves at a dive-by-dive resolution. For example, foraging behavior typically coincides with a large proportion of extensive, wiggly dive sequences. However, foraging sequences may be interspersed by short periods of resting behavior (shallow and smooth dives) even though the dominant behavioral mode may still be foraging. Such patterns are especially likely to occur in harbor porpoise dive data, a species that needs to feed almost continuously to meet energy requirements (Wisniewska et al., 2016). In these cases, hierarchically structured HMMs have strong potential to infer the movement strategies adopted over time, by modeling the transitions between distinct dive patterns (as represented by multiple HMMs) rather than modeling dive-by-dive observations using a single HMM. Thus, to draw a more detailed picture of the behavioral dynamics at multiple time scales, we use hierarchical HMMs, where a crude-scale K -state Markov chain selects which of K fine-scale HMMs describes the dive pattern observed at any point in time. Intuitively, the crude-scale process describes the general behavioral mode (e.g. resting or traveling) which may persist for a large number of consecutive dives while the fine-scale process captures more nuanced state transitions at the dive-by-dive level, given the general behavioral mode.

In terms of the crude time scale, we segmented the time series into hourly intervals and allowed each segment to be connected to one of $K = 2$ HMMs with $N = 3$ (dive-by-dive level) states each. This somewhat arbitrary time scale was chosen based on exploratory analysis of the data set, which suggested that a certain dive pattern is typically adopted for several hours before switching to another one (c.f. Figure 2.3). As comprehensively discussed in Pohle et al. (2017), model selection criteria such as Akaike’s Information Criterion (AIC) or the Bayesian Information Criterion (BIC) typically tend to favor models with larger number of states than are biologically sensible. This is indeed a well-known and notorious problem in applications of HMMs to ecological data (see also Langrock et al. (2015), DeRuiter et al. (2016), Li and Bolker (2017)). Thus, following Pohle et al. (2017), instead of relying on formal model selection procedures, the number of states was chosen pragmatically, with particular emphasis on model parsimony and biological intuition.

The state-dependent distributions were kept the same across the two dive-level HMMs, which were instead allowed to differ only by the t.p.m.s. This assumption implies that any of the three types of dives — as generated by the three different production states — could in principle occur in both crude-level behavioral modes, but will not occur equally often, on average, due to the different Markov chains active at the dive-by-dive level. The initial state distributions, both for the internal and for the production state process, were assumed to be the stationary distributions of the respective Markov chains. We assumed gamma distributions for each of the three dive variables (dive duration, maximum depth and dive wigginess), with an additional point mass on zero in case of dive wigginess to account for the zeros observed. We assumed contemporaneous conditional independence, i.e. for any given dive, the three variables observed are conditionally independent given the production state active at the time of the dive. These assumptions could in fact be relaxed if deemed necessary. However, for this case study we decided that in order to illustrate the key concepts, it would be best to focus on a relatively simple yet biologically informative model structure.

We computed the likelihood in C++ using the forward algorithm (Zucchini et al., 2016) and used the R function `nlm` (Team, 2019) to obtain maximum likelihood estimates via direct numerical likelihood maximization, which took about 15 minutes (on a 3.6 GHz Intel Core i7-4790 CPU).

2.4.1.3 Fitted state-dependent distributions

. The fitted (dive-level) state-dependent distributions displayed in Figure 2.2 suggest three distinct dive types: State 1 captures the shortest (lasting less than 25 seconds), shallowest (less than 10 meters deep) and smoothest (less than 8 meters absolute vertical distance covered) dives with small variance. State 2 captures moderately long (10-60 seconds), moderately deep (5-25 meters) and moderately wiggly (5-30 meters) dives with moderate variance. State 3 captures the longest (40-180 seconds), deepest (10-80 meters) and wiggliest (10-80 meters) dives with high variance.

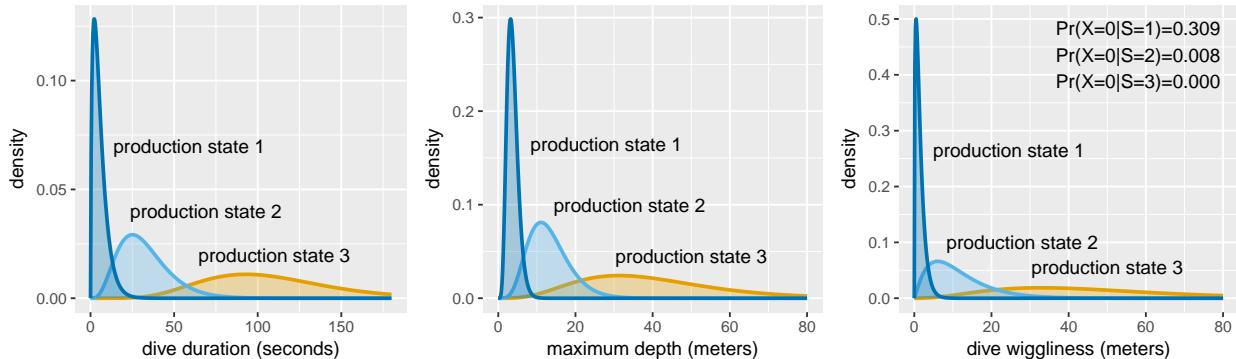


Figure 2.2 Fitted state-dependent distributions for the dive duration, the maximum depth and the dive wigginess, the latter together with the estimated point mass on zero.

In the next section, we discuss the ($K = 2$) distinct dive-level switching patterns among the ($N = 3$) states discussed here, as well as the crude-level process that selects which of the dive-level switching patterns is active in any given hour.

2.4.1.4 Estimated transition probability matrices

. The t.p.m. of the crude-level Markov chain, which selects among the dive-level HMMs, and the t.p.m.s of those two dive-level HMMs, which describe the switching between different types of dives, were estimated as follows:

- crude level:

$$\hat{\Gamma}^{(I)} = \begin{pmatrix} 0.789 & 0.211 \\ 0.219 & 0.781 \end{pmatrix}$$

- dive level:

$$\hat{\Gamma}_1 = \begin{pmatrix} 0.406 & 0.443 & 0.150 \\ 0.240 & 0.600 & 0.159 \\ 0.196 & 0.366 & 0.437 \end{pmatrix} \text{ and } \hat{\Gamma}_2 = \begin{pmatrix} 0.277 & 0.153 & 0.570 \\ 0.124 & 0.248 & 0.628 \\ 0.057 & 0.087 & 0.856 \end{pmatrix}$$

The corresponding stationary distributions are $(0.509, 0.491)$, $(0.277, 0.506, 0.217)$ and $(0.083, 0.110, 0.807)$, respectively. The former of these three stationary distributions implies that, according to the fitted model, in the long run, approximately half of the observations were generated by each of the two HMMs. Furthermore, according to the estimated t.p.m. $\hat{\Gamma}^{(I)}$, there is fairly strong persistence in the crude-level states, indicating that the porpoise typically remains in any given internal state for several hours before switching to the other internal state. This is also confirmed by Figure 2.3, which displays the first 25% of the decoded observations. In particular, Figure 2.3 shows that there are bouts of several hours where production states 1 and 2 are dominant (yet still interspersed with occasional dives generated by production state 3), but also such where production state 3 is dominant. Bouts of the former type are assigned to internal state 1, while the latter are assigned to internal state 2. This again highlights the need to apply hierarchically structured HMMs, here effectively as a means to account for temporal heterogeneity in the state-switching pattern exhibited by the porpoise.

At the dive level, when the first HMM is active, then in the long run about 28%, 51% and 22% of the observations are generated in state 1, 2 and 3, respectively, whereas when the second HMM

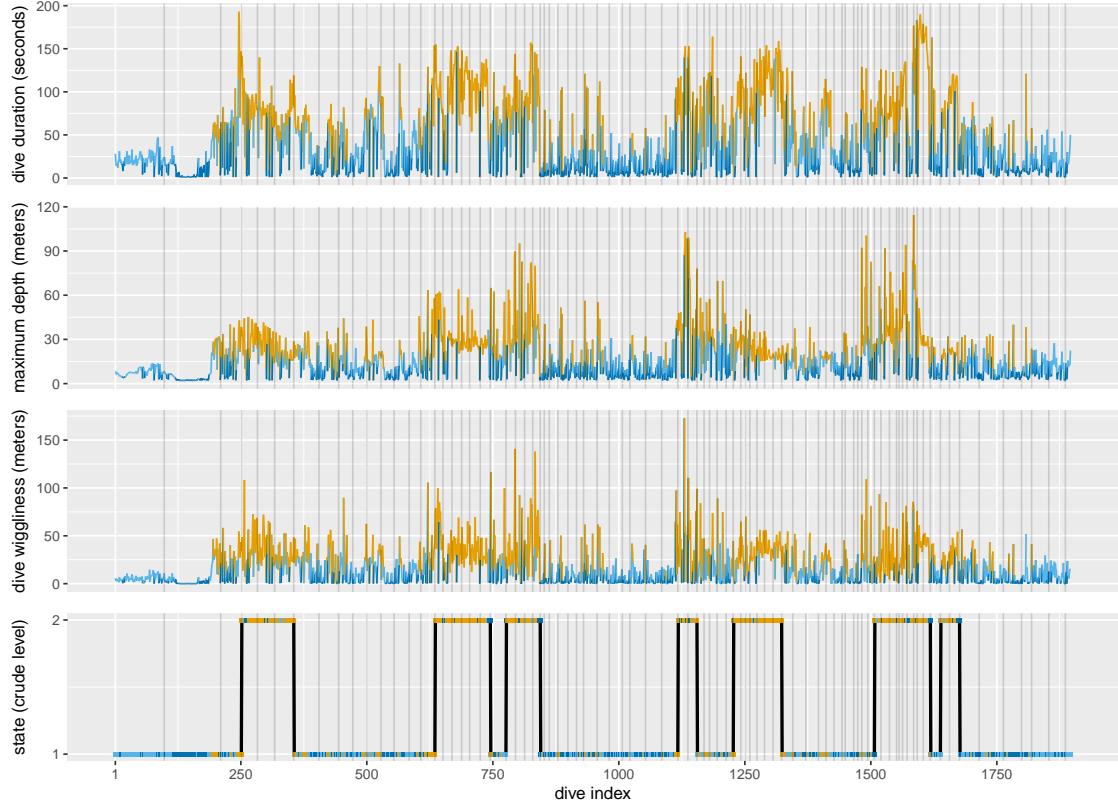


Figure 2.3 Exemplary sequence of the first 25% of the observations of the dive duration, the maximum depth and the dive wigginess. Hourly segments are indicated by vertical grey lines. The states were decoded using the Viterbi algorithm, where the colors refer to the production states.

is active, in the long run about 8%, 11% and 81% of the dives are generated in the respective states. Furthermore, if the first HMM is active, then switching takes place primarily between states 1 and 2, and additionally from state 3 to state 2. If the second HMM is active, then state 3 is dominant, with fairly strong persistence and lots of switches from states 1 and 2 to state 3.

Concluding remarks. The second HMM is indicative of foraging behavior, particularly due to the extensive wigginess during dives, which often indicates prey-chasing. The interpretation of the first HMM, which involves a large proportion of relatively short, shallow and smooth dives, could indicate a resting and/or a traveling behavior. Indeed, traveling from one area to another while remaining close to the water surface is likely the most efficient strategy. However, a more detailed

interpretation of the first HMM would require inclusion of other variables such as the step length, which may prove useful to distinguish between resting and traveling.

2.4.2 Garter Snakes

2.4.2.1 The data

. We model the movements of 19 juvenile garter snakes (*Thamnophis elegans*) in repeated trials that are a subset from a larger experiment quantifying behaviors in the offspring of wild-caught females across experimental treatments, manuscript in preparation. Using EthoVision XT 8.5 (Noldus Information Technology, Wageningen, The Netherlands), we extracted movement data for each of the snakes across two segments in each of three trials (six time series total per individual). In brief, snakes were placed in a novel test arena (circular enclosure with diameter of 24.5 cm) for 60 s. Snakes then received an additional stimulus in the arena and were observed for an additional 60 s. Each 60 s segment was videorecorded and we disregarded the first and last 5 s to eliminate behaviors elicited by the experimenter at the beginning or end of the trial, resulting in analyzed segments each lasting 50 s. This resulted in a total of six recorded segments for each snake. The individuals included here represent the control group, which was not exposed to any additional stimuli in the test arena during the first two trials and was exposed to a novel object during the third trial (that is, between tracks 5 and 6).

2.4.2.2 Model formulation and model fitting

. The snakes displayed a variety of general movement strategies, from the extreme of remaining motionless to moving rapidly around the test arena for the duration of the trial. We calculated the distance moved within 1/2 s and subsequently applied a square root transformation to deal with extreme values present. We assumed that each observed distance conditional on one of three production states was generated by a state-dependent gamma density. Further, to investigate habituation and behavioral plasticity over the course of the six time series per snake, we assumed that each time series was generated by one of three internal state-dependent HMMs. The complete

hierarchically structured HMM fitted to the observed distances was composed of three production states, kept the same across the internal states, and three internal states. In this manner, we investigated whether there was persistence at the internal state level, i.e. if the garter snakes tended to repeat the same general movement patterns across time series or switch strategies. As in the porpoise example, we also used the R function `n1m` (Team, 2019) to obtain maximum likelihood estimates via direct numerical likelihood maximization, which took about 8 minutes (on a 3.6 GHz Intel Core i7-4790 CPU).

2.4.2.3 Fitted state-dependent distributions

. The fitted state-dependent gamma distributions for the three production states, shown in Figure 2.4, correspond to three general types of movement strategies: motionless (or nearly so), slow exploratory, and rapid escape, which the video recordings demonstrate. The estimated average distance traveled in production states 1–3 are: 0.0148, 0.459 and 1.891 cm per 1/2 s, respectively. The largest amount of variability in observed step lengths corresponds to production state 3, with a standard deviation of 0.487 cm^{1/2}.

2.4.2.4 Estimated transition probability matrices and initial state distributions

. The t.p.m.s of the crude-level and of the production-level Markov chains, respectively, were estimated as follows:

- crude level:

$$\hat{\mathbf{F}}^{(I)} = \begin{pmatrix} 0.166 & 0.578 & 0.256 \\ 0.680 & 0.226 & 0.095 \\ 0.157 & 0.208 & 0.635 \end{pmatrix}, \quad \hat{\boldsymbol{\delta}}^{(I)} = \begin{pmatrix} 0.903 \\ 0.072 \\ 0.025 \end{pmatrix}$$

- movement level:

$$\hat{\mathbf{F}}_1 = \begin{pmatrix} 0.947 & 0.047 & 0.006 \\ 0.018 & 0.919 & 0.063 \\ \sim 0 & 0.244 & 0.756 \end{pmatrix}, \quad \hat{\boldsymbol{\delta}}_1 = \begin{pmatrix} 0.413 \\ 0.103 \\ 0.484 \end{pmatrix}$$

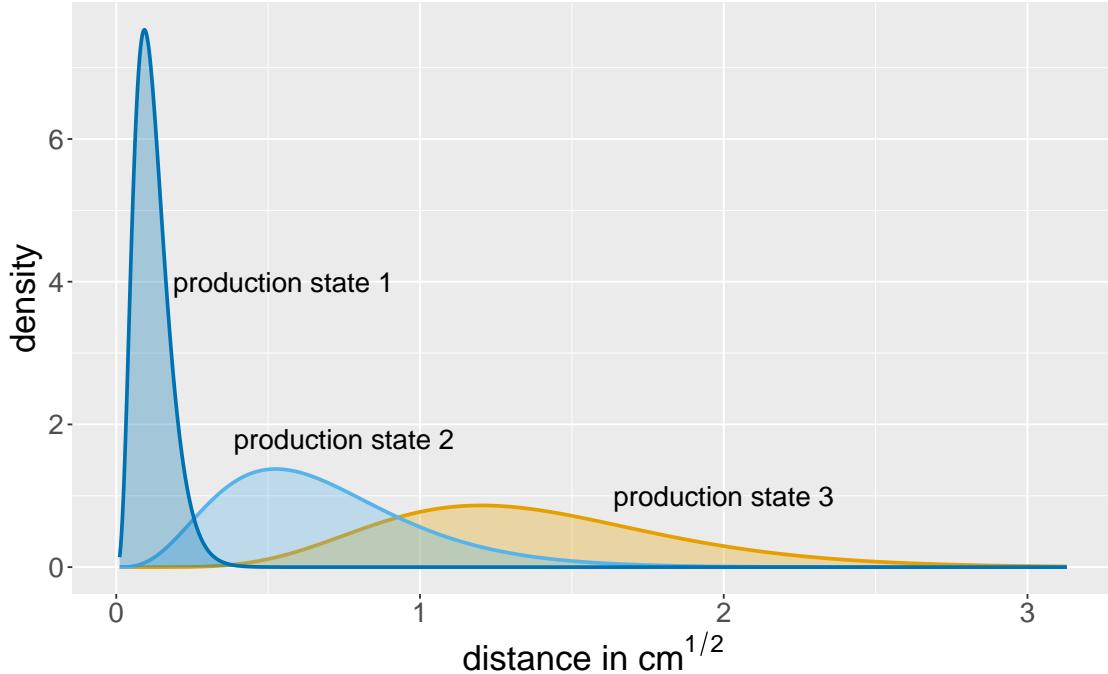


Figure 2.4 Fitted state-dependent distributions for distance traveled.

$$\hat{\Gamma}_2 = \begin{pmatrix} 0.806 & 0.144 & 0.050 \\ 0.019 & 0.657 & 0.324 \\ \sim 0 & 0.185 & 0.815 \end{pmatrix}, \quad \hat{\delta}_2 = \begin{pmatrix} 0.087 \\ 0.024 \\ 0.889 \end{pmatrix}$$

$$\hat{\Gamma}_3 = \begin{pmatrix} 0.994 & 0.006 & \sim 0 \\ 0.003 & 0.997 & \sim 0 \\ \sim 0 & 0.018 & 0.982 \end{pmatrix}, \quad \hat{\delta}_3 = \begin{pmatrix} 0.315 \\ 0.442 \\ 0.243 \end{pmatrix}$$

The three estimated t.p.m.s at the movement level, corresponding to the three internal states, can generally be interpreted as representing three different levels of behavioral flexibility. When the first internal state is active, characterized by $\hat{\Gamma}_1$, individuals are showing more persistence in the motionless and slow exploratory behavioral states overall, and are likely to transition from the rapid escape state to the exploratory state. When the second internal state is active, $\hat{\Gamma}_2$ demonstrates that individuals are switching regularly between behavioral states. When the third internal state is active, $\hat{\Gamma}_3$ demonstrates that individuals seldom transition between states. Thus, the three internal

states reflect a continuum of behavioral flexibility within a short, but ecologically relevant time scale in the context of a potential predation event (< 1 min).

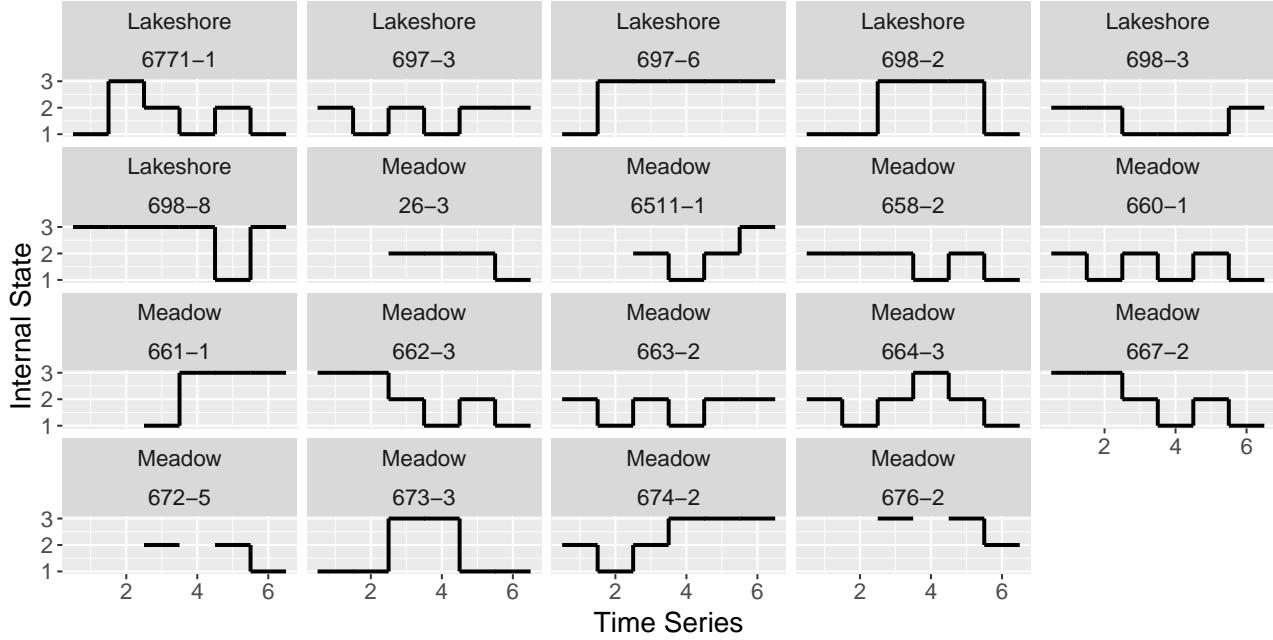


Figure 2.5 Internal state decodings by garter snake. Each garter snake is from one of two ecotypes: Lakeshore or Meadow.

At the crude level, the t.p.m. ($\hat{\Gamma}^{(I)}$) indicates that individuals are readily switching among the three movement level HMMs across trials and in fact are more likely to switch between the movement HMMs representing the greatest behavioral flexibility ($\hat{\Gamma}_1$ and $\hat{\Gamma}_2$). At the crude time scale, we observe persistence in the movement-level HMM describing snakes that are behaviorally inflexible in their movements ($\hat{\Gamma}_3$). Overall, these results indicate that, at the broader time scale, many individuals are readily altering their level of behavioral flexibility while some individuals remain persistent in their behaviors both within and across trials (i.e. at both the movement and crude levels).

2.4.2.5 Concluding remarks

HMMs are not yet commonly applied to animal movement data from experimental designs, even though they typically produce multiple time series per individual. Introducing multiple Markov chains in the HMM formulation lends itself to characterizing the consistency of individual behaviors and variation among individuals at different time scales. We show that individuals employing multiple movement strategies in a narrow time frame are more likely to switch between strategies at a crude time scale, while individuals consistent in their behaviors at the movement time scale are also consistent at the crude time scale. Furthermore, these patterns are independent of the behavioral strategies exhibited: individuals may consistently remain in any one of the three behavioral states.

2.5 Discussion

HMMs have proven to be useful statistical tools for modeling animal movement data, providing a framework to infer drivers of variation in movement patterns, and thus behavior. The basic HMM, however, has so far been used to infer aspects of animal behavior only when a single data point can be thought to stem from one of N possible (production) states, which are loosely connected to behavioral modes that manifest themselves at the temporal resolution at which observations are made. Yet, thanks to advances in tag technology and battery life, as well as software capable of tracking recorded images, data can be collected at finer temporal resolutions and over longer periods of time. Inferences at time scales cruder than those at which data are collected, and which correspond to larger-scale behavioral processes, are not yet answered via HMMs. We provide a corresponding extension to incorporate multiple Markov chains in an HMM, allowing for multi-scale behavioral inferences. The extension is straightforward in the sense that likelihood inference via application of the forward algorithm is essentially analogous as in the case of basic HMMs. The hierarchically structured HMMs can also be used to avoid coarsening data, such as acceleration data that can be collected many times per second (Leos-Barajas et al., 2016). As this is, as of yet, an area of movement ecology that has received little attention, our proposed framework is one of

the first that models animal behavior simultaneously at multiple time scales. Since different types of behavioral modes occur at different temporal and spatial scales, statistical models that account for this, such as the hierarchically structured HMMs proposed in this work, may contribute to providing a more nuanced picture of animal behavior.

In this manuscript, we did not discuss how to implement model selection and model checking for hierarchically structured HMMs. In principle, since we are fitting the models using maximum likelihood, model selection could be conducted using standard information criteria. However, while conceptually this is completely straightforward, in practice this procedure is notoriously error-prone already for basic HMMs, due to the strong tendency of information criteria to favor models with many more states than are biologically reasonable (Langrock et al., 2015; Pohle et al., 2017; Li and Bolker, 2017). Given the additional state process, this issue will be exacerbated within hierarchically structured HMMs as presented in this work, since the number of states both for the production process and for the internal state process needs to be chosen. We cannot currently offer a satisfactory solution to this problem, except by saying that biological *a priori* expert knowledge ought to be taken into account. For general advice regarding the issue of model selection in HMMs, see Pohle et al. (2017). For model checking, possible avenues are (i) simulation-based model assessment and (ii) analyses of pseudo-residuals. Regarding (i), the fundamental concept is the idea that the fitted model should generate data similar to the observed data in all important aspects. Quantification of aspects of the data patterns should reflect key behaviors believed to be important to the problem. Pseudo-residuals, as discussed for example in Patterson et al. (2009), Langrock et al. (2012) and in Zucchini et al. (2016), can be calculated also for hierarchically structured HMMs, most easily by conditioning on Viterbi-decoded internal states, hence calculating the pseudo-residuals at the production level, given the (fixed) most likely internal state sequence. Both model selection and model checking needs to be explored further before these models may become a tool that is routinely applied in the analysis of animal behavior data.

Using *ad hoc* choices of the exact model formulations (yet such that are grounded in biological theory), in Section 2.4 we demonstrated how the hierarchically structured HMMs, applied to move-

ment data collected on harbor porpoises and garter snakes, respectively, provided new insights into the behavior of these species. However, a hierarchically structured HMM not only allows for new inferences to be made from movement studies — it can also be applied to the study of behavior in general. Being able to characterize persistence of movement patterns at multiple time scales allows us to learn about personality, individual specialization, and cognition, among other things. Numerous studies across a wide range of taxa have shown that individual animals behave differently from other individuals and that these differences are maintained through time (Sih et al., 2004; Réale et al., 2007; Sih et al., 2015; Roche et al., 2016; David et al., 2016). These observations have given rise to the burgeoning field of animal personality which explores the ecology and evolutionary significance of such behavioral differences among individuals. Such studies have included a variety of behavioral measures but have only recently incorporated models of movement as a behavioral trait (Schliehe-Dieks et al., 2012; McKellar et al., 2015; Spiegel et al., 2017). Importantly, the animal personality framework has recently incorporated an understanding of how individuals differ in their behavioral plasticity (Mathot and Dingemanse, 2015; Stamps, 2016), which requires more specific theoretical models as well as more sophisticated statistical approaches (Dingemanse and Dochtermann, 2013; Kleun and Brommer, 2013; Japyassú and Malange, 2014). Thus, the field is attempting to address two fundamental questions: (1) how do behaviors differ among individuals and (2) how do individual behaviors change over time or context? Addressing these questions therefore requires analysis at two levels: (1) to identify and categorize behavioral states (production states) and (2) to identify patterns of changes in behavioral states (internal states). In the HMM framework, the internal states may reflect general movement patterns associated with endogenous behavioral plasticity (Stamps, 2016) or personality which allows for further examination of persistence or switching among movement patterns at the cruder time scale.

The addition of multiple Markov chains in the HMM framework to conduct multi-scale behavioral inferences necessitates the selection of the temporal resolution at two time scales: the observation level and the level of the individual time series. The selected temporal resolution at the level of the internal states will need to be tied to the specific biological question of interest.

There may be a natural manner in which the data are segmented that produces time series of unequal length. However, this need not be an issue as long as each time series is reflective of some general behavioral process irrespective of the length of the time series. Formulating the hierarchically structured HMM, in terms of selecting the number of production states and internal states, will need to be done in a pragmatic fashion in order to balance model complexity with biological intuition. Due to the HMM's inherent flexibility, the internal states may be formulated in a few manners, e.g. a single HMM (such as has been described in Section 2), assuming a distribution of HMMs, or allowing for longer state dwell times via the hidden semi-Markov model (Langrock et al., 2012), in order to account for unexplained variability in the state processes. In particular, as the number of production states, N , increases, so will the number of ways in which two HMM's t.p.m.s Γ_i and Γ_j can differ. To account for all of these possibilities may require a large number of internal states, if each internal state is assumed to only correspond to one t.p.m. for the HMM.

Adding hierarchical structures to the HMM opens new possibilities for modeling multi-scale behaviors and provides an avenue to study animal personality and general behavior from movement studies. In this manner, environmental covariates can also be included to understand their effects on state occupancy and dynamics of variation in behavioral modes at broader time scales than that at which the data are collected. Further, this framework may be adapted for simultaneous modeling of multiple animal behavior data streams collected at distinct temporal resolutions. The internal states can be adapted to generate a sequence of fine-scale observations as well as one observation from a distinct data stream.

2.6 References

- Biuw, M., Boehme, L., Guinet, C., Hindell, M., Costa, D., Charrassin, J. B., Roquet, F., Bailleul, F., Meredith, M., Thorpe, S., Tremblay, Y., McDonald, B., Park, Y. H., Rintoul, S. R., Bindoff, N., Goebel, M., Crocker, D., Lovell, P., Nicholson, J., Monks, F., and Fedak, M. A. (2007). Variations in behavior and condition of a southern ocean top predator in relation to in situ oceanographic conditions. *Proceedings of the National Academy of Sciences*, 104:13705–13710.
- David, M., Dall, S. R. X., and Bshary, R. (2016). Unravelling the philosophies underlying ‘animal personality studies: A brief re-appraisal of the field. *Ethology*, 122:1–9.

- DeRuiter, S. L., Langrock, R., Skirbutas, T., Goldbogen, J. A., Calambokidis, J., Friedlander, A. S., and Southall, B. L. (2016). A multivariate mixed hidden markov model for blue whale behaviour and responses to sound exposure. *Annals of Applied Statistics*, 11:362–392.
- Dingemanse, N. J. and Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: mixed-effect modelling approaches. *Journal of Animal Ecology*, 82:39–54.
- Fine, S., Singer, Y., and Tishby, N. (1998). The hierarchical hidden markov model: Analysis and applications. *Machine Learning*, 32:41–62.
- Hart, T., Mann, R., Coulson, T., Pettorelli, N., and Trathan, P. N. (2010). Behavioural switching in a central place forager: patterns of diving behaviour in the macaroni penguin (*eudyptes chrysophous*). *Marine Biology*, 157:1543–1553.
- Hindell, M., McMahon, C. R., Bester, M. N., Boehme, L., Costa, D., Fedak, M. A., Guinet, C., Herraiz-Borreguero, L., Harcourt, R. G., Huckstadt, L., Kovacs, K. M., Lydersen, C., McIntyre, T., Muelbert, M., Patterson, T. A., Roquet, F., Williams, G., and Charrassin, J. B. (2016). Circumpolar habitat use in the southern elephant seal: implications for foraging success and population trajectories. *Ecosphere*, 7:e01213.
- Japyassú, H. F. and Malange, J. (2014). Plasticity, stereotypy, intra-individual variability and personality: handle with care. *Behavioural Processes*, 109:40–47.
- Kleun, E. and Brommer, J. E. (2013). Context-specific repeatability of personality traits in a wild bird: A reaction-norm perspective. *Behavioral Ecology*, 24:650–658.
- Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., and Morales, J. M. (2012). Flexible and practical modeling of animal telemetry data: hidden markov models and extensions. *Ecology*, 93:2336–2342.
- Langrock, R., Kneib, T., Sohn, A., and DeRuiter, S. L. (2015). Nonparametric inference in hidden markov models using p-splines. *Biometrics*, 71:520–528.
- Langrock, R., Marques, T. A., Baird, R. W., and Thomas, L. (2014). Modeling the diving behavior of whales: a latent-variable approach with feedback and semi-markovian components. *JABES*, 19:82–100.
- Leos-Barajas, V., Photopoulou, T., Langrock, R., Patterson, T. A., Murgatroyd, M., Watanabe, Y. Y., and Papastamatiou, Y. P. (2016). Analysis of accelerometer data using hidden markov models. *Methods in Ecology and Evolution*, 8:161–173.
- Li, M. and Bolker, B. M. (2017). Incorporating periodic variability in hidden markov models for animal movement. *Movement Ecology*.

- Luque, S. P. (2007). Diving behaviour analysis in r. an introduction to the divemove package. *R News*, 7:8–14.
- Mathot, K. J. and Dingemanse, N. J. (2015). *Plasticity and Personality. In Integrative Organismal Biology*, pages 55–69. John Wiley & Sons, Hoboken NJ.
- McKellar, A. E., Langrock, R., Walters, J. R., and Kesler, D. C. (2015). Using mixed hidden markov models to examine behavioural states in a cooperatively breeding bird. *Behavioral Ecology*, 26:148–157.
- Michelot, T., Langrock, R., Bestley, S., Jonsen, I. D., Photopoulou, T., and Patterson, T. A. (2017). Estimation and simulation of foraging trips in land-based marine predators. *Ecology*, 98:1932–1944.
- Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. (2004). Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, 85(9):2436–2445.
- Patterson, T. A., Basson, M., Bravington, M. V., and Gunn, J. S. (2009). Classifying movement behaviour in relation to environmental conditions using hidden markov models. *Journal of Animal Ecology*, 78(6):1113–1123.
- Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., and King, R. (2017). Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis*, 101(4):399–438.
- Pohle, J., Langrock, R., van Beest, F. M., and Schmidt, N. M. (2017). Selecting the number of states in hidden markov models – pitfalls, practical challenges and pragmatic solutions illustrated using animal movement. *JABES*, 22(3):270–293.
- Réale, D., Reader, S. M., Sol, D., McDougall, P. T., and Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, 82:291–318.
- Roche, D. G., Careau, V., and Binning, S. A. (2016). Demystifying animal ‘personality’ (or not): why individual variation matters to experimental biologists. *Journal of Experimental Biology*, 219:3832–3843.
- Schliehe-Dieks, S., Kappeler, P. M., and Langrock, R. (2012). On the application of mixed hidden markov models to multiple behavioural time series. *Interface Focus*, 2:180–189.
- Sih, A., Bell, A., and Johnson, J. C. (2004). Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology and Evolution*, 19:372–378.

- Sih, A., Mathot, K. J., Moir'on, M., Montiglio, P. O., Wolf, M., and Dingemanse, N. J. (2015). Animal personality and state-behaviour feedbacks: a review and guide for empiricists. *Trends in Ecology and Evolution*, 30:50–60.
- Spiegel, O., Leu, S. T., Bull, C. M., and Sih, A. (2017). What's your move? movement as a link between personality and spatial dynamics in animal populations. *Ecology Letters*, 20:3–18.
- Stamps, J. A. (2016). Individual differences in behavioural plasticities. *Biological Reviews*, 91:534–567.
- Team, R. C. (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.
- Towner, A. V., Leos-Barajas, V., Langrock, R., Schick, R. S., Smale, M. J., Kaschke, T., Jewell, O. J. D., and Papastamatiou, Y. P. (2016). Sex-specific and individual preferences for hunting strategies in white sharks. *Functional Ecology*, 30(8):1397–1407.
- Wisniewska, D. M., Johnson, M., Teilmann, J., nate, L. R.-D., Shearer, J., Sveegaard, S., Miller, L. A., Siebert, U., and Madsen, P. T. (2016). Ultra-high foraging rates of harbor porpoises make them vulnerable to anthropogenic disturbance. *Current Biology*, 26:1441–1446.
- Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction using R*. Chapman & Hall, Boca Raton, FL, 2nd edition.

CHAPTER 3. HIDDEN MARKOV MODELS WITH MULTI-SCALE STATE PROCESSES IN MOVEMENT ECOLOGY

Vianey Leos-Barajas^{1*}, Mark S. Kaiser¹, Kim Holland², Yannis P. Papastamatiou³

¹Department of Statistics, Iowa State University, USA

²Department of Biology, University of Hawai'i at Mānoa, USA

³Department of Biological Sciences, Florida International University, USA

3.1 Abstract

Fine-scale movement data can be collected remotely on a wide range of animals and reveal detailed movement patterns. Patterns may consist of differentiating between small and large signals, but can also manifest themselves at temporal scales larger than at which the data were collected and may be viewed as a composition of the fine-scale discernible movements. Hidden Markov models (HMMs) are a common class of time series models applied to animal movement data. Their ability to connect an observation process to an underlying state process, generally serving as a proxy for a finite set of animal behaviors of interest, matches the intuition that the observed movements stem from an underlying (unobserved) behavioral process. We can further extend the HMM framework to consist of multiple state processes to reflect that different behaviors are identified by different compositions of the observed movement processes. We refer to this extension as a multi-scale HMM whereby one state process is connected to the underlying behaviors that generate the movements at the temporal scale at which the data are processed and another is connected to a larger-scale behavioral process, defined as a composition of fine-scale behavioral states. We present two formulations of the multi-scale HMM applied to tiger shark depth data. In one application, absolute changes in depth at a 10-min scale are taken to stem from one of two fine-scale behavioral processes while sequences of movements across days are taken to stem from one of two larger scale behavioral processes. In another, we construct four fine-scale states,

reflective of small and large ascents/descents at a 10-min scale and connect small ascents/descents to a single larger scale behavioral state and larger variations in ascents/descents to another. The multi-scale HMM framework extends the types of behaviors that can be identified and allows for potential environmental drivers of behavior to be incorporated into the larger scale behavioral or fine-scale behavioral process. As it can be expressed as a particular formulation of a basic HMM, all model fitting and validation techniques in the HMM toolbox apply.

3.2 Introduction

Much of the analysis of individual animal movement is driven by the desire to connect observed movement metrics to classes of behavioral ‘states’. General state-space models provide a mechanistic manner in which to connect the observed movements to an underlying behavioral process and have been widely used to make inferences about the drivers of animal behavior (Morales et al., 2004; Patterson et al., 2008; Hooten et al., 2017; Patterson et al., 2017). Although of much use, the manner in which these models have been implemented to date does not allow for identification of complex behavioral patterns, that is, sequences of discernible actions/movements that provide information about the presence of larger scale behavioral patterns. We develop and illustrate multi-scale models to address this objective. Our multi-scale approach constitutes an extension of basic hidden Markov models (HMMs).

Hidden Markov models (HMMs) are popular time series models used widely in the field of animal movement (Zucchini et al., 2016). A basic HMM assumes that the observation at time t , Y_t , is generated by an underlying state, S_t . In the context of animal movement, the observations $\{Y_t\}_{t=1}^T$ can be the geoposition of an animal, a measure of acceleration, a complete dive or any metric that measures the activity of an animal. The state process $\{S_t\}_{t=1}^T$ serves as a proxy for a finite collection of behaviors of interest, such that the observation Y_t is a result of the animal exhibiting one of the (unobserved) behaviors of interest, S_t . We present a multi-scale HMM that connects fine-scale actions/movements to a set of fine-scale discernible states (denoted *production* states) and then connects a sequence of fine-scale states to larger scale states, which represent behavioral

patterns of interest. These are referred to as the *internal* states of the overall process because they are connected with the functional intent of an individual animal. For instance, recording a white shark's location every 5 min may inform fine-scale behaviors like traveling or area-restricted search (ARS) behavior. However, examining the composition of traveling and ARS behaviors can inform the type of overall hunting strategy the white shark implemented before a prey attempt (Towner et al., 2016). In this example, the fine-scale movements connected to traveling or ARS reflect the production states, while the composition of fine-scale behaviors inform the internal state, i.e. the type of hunting strategy implemented by the white shark. In this paper, we present two approaches to a multi-scale HMM construction depending on whether a sequence of observations produced by one internal state is fixed, or is a random variable, expanding upon the model details presented in Leos-Barajas et al. (2017) and Pirotta et al. (2018).

In Section 2, we describe the basic HMM as well as circumstances under which it is identifiable. Using the results in Section 2, Section 3 covers formulation of the multi-scale HMM and conditions under which it is identifiable. In Section 4 we discuss analysis of the models developed in a Bayesian framework using dynamic Hamiltonian Monte Carlo, as implemented in the software **Stan**. In Section 5, we analyze data on tiger shark depths using a multi-scale HMM under both fixed and random internal state duration. Section 6 contains concluding remarks.

3.3 Hidden Markov Models

A basic discrete-time HMM is a stochastic time series model composed of a state process, denoted by unobservable random variables $\{S\}_{t=1}^T$, and a state-dependent observation process, denoted by observable random variables $\{Y_t\}_{t=1}^T$. At each point in time, we assume the state process takes on one of $n \in \{1, \dots, N\}$ values, for known $N \in \mathbb{N}$. The sequence of observations $\{Y_t\}_{t=1}^T$ are taken to be conditionally independent given the underlying state sequence, and generated according to state-dependent distributions written as $f(Y_t|S_t = n) = f_n(y_t)$, for $n \in \{1, \dots, N\}$. The evolution of the states over time is assumed to follow a first-order Markov property and is governed by an $N \times N$ transition probability matrix (t.p.m.), Γ . The entries of Γ are the

conditional probabilities $\gamma_{ij} = \Pr(S_t = j | S_{t-1} = i)$ for $i, j = 1, \dots, N$. Initially, we assume that $\boldsymbol{\Gamma}$ is constant across time. The initial state distribution, $\boldsymbol{\nu}$, is a vector of probabilities with entries, $\nu_i = \Pr(S_1 = i); i = 1, \dots, N$.

For one time series, the likelihood defined by a HMM can be expressed as a matrix product. Let $\mathbf{P}(y_t)$ be an $N \times N$ diagonal matrix with entries $P_{nn}(y_t) = f_n(y_t)$ and let $\mathbf{1}$ denote a column vector of ones of size N . The likelihood may then be written as,

$$\mathcal{L} = \boldsymbol{\nu}^\top \mathbf{P}(y_1) \prod_{t=2}^T \boldsymbol{\Gamma} \mathbf{P}(y_t) \mathbf{1}, \quad (3.1)$$

Identifiability

We begin with identifiability as discussed in Teicher (1963), Cappé et al. (2005), Holzmann et al. (2006), Holzmann et al. (2015) and G. Alexandrovich (2016). In this context, for parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ such that $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, identifiability implies that joint marginal densities for a collection of random variables are distinct, i.e. $p(\mathbf{y}|\boldsymbol{\theta}_1) \neq p(\mathbf{y}|\boldsymbol{\theta}_2)$. G. Alexandrovich (2016) show that identifiability of a basic HMM holds when (i) the state-dependent densities f_n are distinct and (ii) the t.p.m. $\boldsymbol{\Gamma}$ is ergodic and has full rank. The initial distribution $\boldsymbol{\nu}$ may be taken to be the stationary distribution, estimated or taken to be a vector with non-zero probabilities that reflect prior information. If assumptions (i) and (ii) hold for any two HMMs with the same joint distribution $p(\mathbf{y}|\boldsymbol{\theta})$, for sequential observations $\{Y\}_{t=1}^T$, then the parameters coincide (up to some permutation). Related to the preceding notion of identifiability, implementation of a HMM involves issues of parameter estimation and labeling. In particular, in a Bayesian framework, this may involve joint posterior distributions that have multiple modes. Assigning non-exchangeable priors to parameters of the state-dependent distributions, $\{f_n\}_{n=1}^N$, may alleviate or reduce difficulties encountered when this occurs (Betancourt, 2017b).

3.4 Multi-Scale State Processes

In this section we extend the basic HMM structure by allowing multi-scale state processes. In particular, we consider two state processes evolving at different temporal resolutions, the *production*

state process and the *internal* state process. As mentioned previously, the production state process reflects a set of fine-scale behavioral states connected to the fine-scale actions/movements of the animal. Each observation is taken to have been produced by one of the production states. The internal states, however, reflect a larger scale behavioral state of interest that is defined as a composition of production states. For example, Pirotta et al. (2018) constructed an HMM that modeled a fine-scale process evolving at a resolution of 10 min while also identifying larger-scale movement phases in the movements of fulmars. The fine-scale process was able to identify periods of transitory behavior vs area-restricted search behavior, while the larger scale process identified periods in which the fulmar was heading out to sea, heading toward the colony or heading towards the nearest boat.

We present two general cases for application of multi-scale models in movement ecology, (i) the observations are segmented such that each internal state is assumed to emit a pre-specified number of production states, $K \in \mathbb{N}$, or (ii) the number of production states corresponding to internal states are random and vary according to some distribution defined on the positive integers.

3.4.1 Equal-length internal state duration times

We first present details for a multi-scale HMM where each internal state is assumed to produce a segment of K production states, and by extension a sequence of K observations. To emphasize the multi-scale modeling of the time series, we map the usual observed process time index t to a set of indices $\{m, k\}$, for $m = 1, \dots, M$ and $k = 1, \dots, K$ as demonstrated in Table 3.1. The production state process is indexed in the same manner.

Table 3.1 Mapping indices from the original temporal scale to indexing via segments.

Original Indexing	Mapping	Segment Indexing	Segment
Y_1, Y_2, \dots, Y_K		$Y_{1,1}, Y_{1,2}, \dots, Y_{1,K}$	1
$Y_{K+1}, Y_{K+2}, \dots, Y_{2K}$	$m = \lceil t/K \rceil$	$Y_{2,1}, Y_{2,2}, \dots, Y_{2,K}$	2
\vdots	$k = t - K \cdot (m - 1)$	\vdots	\vdots
$Y_{T-K+1}, Y_{T-K+2}, \dots, Y_T$		$Y_{M,1}, Y_{M,2}, \dots, Y_{M,K}$	M

For a time series of length T , we first partition the observations into M non-overlapping segments of length K . We use the indices $m \in \{1, \dots, M\}$ and $k \in \{1, \dots, K\}$ to denote the k^{th} observation in the m^{th} segment so that what was originally indexed as Y_t is now indexed as $Y_{m,k}$ with $k = t - K \cdot (m - 1)$ and $m = \lceil t/K \rceil$.

A multi-scale HMM describes the evolution of three stochastic processes: (i) an internal state process, (ii) a production state process and (iii) a state-dependent observation process. Given the indexing described in Table 3.1, we describe the manner in which the three processes evolve across and/or within segments. The internal state process is denoted by random variables $\{H_m\}_{m=1}^M$, assumed to evolve across segments, such that H_m takes on values $v \in \{1, \dots, V\}$, for a known $V \in \mathbb{N}$. We assume that H_m evolves across segments according to a $V \times V$ internal state t.p.m., Ω , and follows a first-order Markov property such that the entries of Ω are $\{\delta_{i,j}\} = \Pr(H_m = j | H_{m-1} = i)$ for $i, j = 1, \dots, V$.

For $H_m = v$, we define a sequence of production states $\{S_{m,k}\}_{k=1}^K$, such that at each time $S_{m,k}$ takes on one of $n_v \in \{1, \dots, N_v\}$ values. As such, we allow the number of production states, N_v , to depend on the internal state active during the m^{th} segment. In this manner we have the flexibility to describe different larger scale behaviors of interest by varying number of fine-scale actions/movements, if needed. Within each segment, the production state processes evolve according to internal state-dependent t.p.m.s, Γ_v , $v = 1, \dots, V$. We again assume a first-order Markov property for each of the production state processes, conditional on the internal state H_m . The observation process, $\{Y_{m,k}\}_{k=1}^K$, is taken to be conditionally independent given the underlying production state process $\{S_{m,k}\}_{k=1}^K$ and internal state H_m , and further generated according to the states-dependent distribution, $f(y_{m,k}|S_{m,k} = n_v, H_m = v)$, hereafter denoted $f_{v,n_v}(y_{m,k})$. In Section 3.4.3 we expand upon the possible specifications of the states-dependent distributions and its implications in the analysis of animal movement.

Given the multi-scale HMM framework as has been presented thus far, we can estimate not only the production state process that generated the fine-scale movements/actions we observe (what is typically done via HMMs) but also gain insight into how the composition of production

states informs the larger scale behavior of an animal. Although presented separately, we can also construct a single transition matrix that encompasses both the fine-scale and larger-scale processes of interest which has the advantage of presenting the multi-scale HMM as a specially structured HMM and allowing for the well-known HMM toolbox, from model fitting techniques to model checking, to be implemented. As such, we formulate a single t.p.m. $\Delta^{m,k}$ to describe the evolution of the internal state process across segments and the production state process within segments. Let $\mathbf{0}_{i,j}$ be an $N_i \times N_j$ matrix with zero entries, for $i, j = 1, \dots, V$, and the matrix $\boldsymbol{\Pi}_k$ denote a $N_v \times N_v$ matrix with identical row entries, $\boldsymbol{\nu}_v$, the initial distribution for the v^{th} internal state. We can then express $\Delta^{m,k}$ in the following manner,

$$\Delta^{m,k} = \begin{cases} \begin{pmatrix} \boldsymbol{\Gamma}_1 & \mathbf{0}_{1,2} & \cdots & \mathbf{0}_{1,v} \\ \mathbf{0}_{2,1} & \boldsymbol{\Gamma}_2 & \cdots & \mathbf{0}_{2,v} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{V,1} & \mathbf{0}_{V,2} & \cdots & \boldsymbol{\Gamma}_V \end{pmatrix} & \text{for } k \in \{2, \dots, K\}, m \in \{1, \dots, M\} \\ \begin{pmatrix} \delta_{1,1}\boldsymbol{\Gamma}_1 & \delta_{1,2}\boldsymbol{\Pi}_2 & \cdots & \delta_{1,K}\boldsymbol{\Pi}_V \\ \delta_{2,1}\boldsymbol{\Pi}_1 & \delta_{2,2}\boldsymbol{\Gamma}_2 & \cdots & \delta_{2,K}\boldsymbol{\Pi}_V \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{V,1}\boldsymbol{\Pi}_1 & \delta_{V,2}\boldsymbol{\Pi}_2 & \cdots & \delta_{V,V}\boldsymbol{\Gamma}_V \end{pmatrix} & \text{for } k = 1, m \in \{2, \dots, M\} \end{cases} \quad (3.2)$$

The transition matrix $\Delta^{m,k}$ allows for switches across internal states from $y_{m,K}$ to $y_{m+1,1}$, but within the m^{th} segment, the observations are generated according to the dynamics of a single internal state. In particular, if there is no change in internal state across segments, the process does not reinitiate at the initial state distribution, $\boldsymbol{\nu}_v$, but continues to evolve according to the dynamics of $\boldsymbol{\Gamma}_v$.

Likelihood

We denote the initial internal state distribution as $\boldsymbol{\nu}^*$ and the initial distribution of the combined internal state and production state process as $\boldsymbol{\nu}^0$, with entries $\nu_{v,n}^0 = \nu_v^* \cdot \nu_{v,n}$. The matrix of production state-dependent densities is given as $\mathbf{Q}(\mathbf{y}_m) = \text{diag}(f_{1,1}(y_{m,k}), \dots, f_{V,N_V}(y_{m,k}))$, the t.p.m. is given by $\Delta^{m,k}$ and $\mathbf{1}$ denotes a column vector of ones of size $\sum_{k=1}^V N_v$. Given these components, for one time series, we write the likelihood of the multi-scale HMM as,

$$L = (\boldsymbol{\nu}^0)^\top \mathbf{Q}(y_{1,1}) \prod_{k=2}^K \Delta^{1,k} \mathbf{Q}(y_{1,k}) \left[\prod_{m=2}^M \prod_{k=1}^K \Delta^{m,k} \mathbf{Q}(y_{m,k}) \right] \mathbf{1} \quad (3.3)$$

Identifiability

Identifiability of the multi-scale HMM relies on the same conditions as a basic HMM. Within each internal state, the set of production state-dependent densities, $\{f_{v,n}\}_{n=1}^{N_v}$ must be distinct and the overall t.p.m., $\Delta^{m,k}$, must be ergodic and have full rank. Further, for the set of parameters describing the V internal state processes, generally denoted as $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_V$, we must have that at least one entry $\theta_{i,w} \neq \theta_{j,w}$, for $i,j \in \{1, \dots, V\}$ and $w \in \{1, \dots\}$.

3.4.2 Variable-length internal state duration times

Rather than process the time series into segments of length K , thus fixing the points in time at which transitions across internal states can occur, we can allow transitions to occur at varied points in time. In particular, fixing the number of observations, K that are produced by an internal state at a time, as done in the previous section, is a special case of the model we present in this section.

One manner to allow for variable state-duration times is have the process evolve in the same manner as a basic HMM. At every point in time, we allow for a possible internal state-switch. As such, we return to the usual index t to denote time across the observation and both production and internal state processes (i.e. Y_t, S_t, H_t). However, there are still two state processes to consider at each point in time, and continue to use $f_{n_v,v}(y_t)$ to denote the states-dependent distribution for y_t .

As before, we can combine the production and internal state process to construct an overall t.p.m., Δ , that governs the evolution of both state processes over time as follows,

$$\Delta = \begin{pmatrix} \delta_{1,1}\Gamma^1 & \delta_{1,2}\Pi_2 & \cdots & \delta_{1,K}\Pi_V \\ \delta_{2,1}\Pi_1 & \delta_{2,2}\Gamma^2 & \cdots & \delta_{2,K}\Pi_V \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{V,1}\Pi_1 & \delta_{V,2}\Pi_2 & \cdots & \delta_{V,V}\Gamma^V \end{pmatrix} \quad (3.4)$$

The t.p.m. of equation (3.4) is in fact no different than that of the t.p.m. in equation (3.2) at times for which the internal state transitions are allowed (at $k = 1, m \in \{1, \dots, M\}$). Details for a process of interest at the internal state level that must meet a minimal amount of time that it must be exhibited before switches can occur are provided in the Appendix.

Likelihood

We can write the likelihood similarly to a basic HMM as follows,

$$L = (\nu_0)^\top \mathbf{Q}(y_1) \left[\prod_{t=2}^T \Delta \mathbf{Q}(y_t) \right] \mathbf{1} \quad (3.5)$$

Identifiability

The identifiability of a multi-scale HMM allowing for internal state switches at any point in time is the same as in Section 3.4.1.

3.4.3 State-Dependent Distributions

There are a variety of manners in which the states-dependent distributions can be defined in order to appropriately reflect the biological processes of interest.

1. The state-dependent distributions are equal across internal states, i.e. $\forall v \in V, f_{v,n_v}(\cdot) = f_{n_v}(\cdot)$ and $N_1 = \dots = N_V$

2. There are $R \leq N_v$ state-dependent distributions equal across internal states, i.e. $f_{v,n_v}(\cdot) = f_{n_v}(\cdot)$ for $n_v \in \{1, \dots, R\}$, $R \in \mathbb{N}$
3. All state-dependent distributions vary across internal states, $\forall v \in V, f_{v,n_v}(\cdot) \neq f_{n_v}(\cdot)$

The assumption given in 1. indicates that overall there are N biologically relevant fine-scale states of interest, yet the order and frequency in which they occur defines different internal state processes, i.e. $\Gamma_1 \neq \dots \neq \Gamma_V$. This construction may be particularly useful when using the model formulation presented in Section 3.4.1 as the data can not usually be processed into fixed-length segments that exclude some behaviors of interest in some segments and not in others. It further provides a consistent representation of the behaviors of interest across internal states and an avenue to make inferences about the drivers under different environmental conditions. We demonstrate this approach in Section 3.6.1.

Under assumptions 2. and 3., some production states only occur in one or more, but not all, internal states. While we can arbitrarily increase flexibility and capture a larger variety of patterns through this formulation, determining the production and internal state construction using domain expertise provides a manner in which to validate its utility and necessity. We present an example in Section 3.6.2 in which we assume the structure of 3. but use domain expertise to guide the number of production and internal states, as well as the manner in which we construct the states-dependent distributions in order to capture the larger scale behaviors of interest.

3.5 Inference

In Sections 3.3 and 3.4, the likelihood of all models can be written as a matrix product as we can structure the t.p.m. of the multi-scale HMMs presented here as in basic HMMs. The models are presented in this manner intentionally as none of the likelihoods are functions of the unknown states, **S** and **H**. Zucchini et al. (2016) detail how to evaluate the likelihood efficiently using the so-called *forward algorithm*. As long as there are no other discrete random variables present, exact inference for the parameters can be conducted in a Bayesian framework using dynamic Hamiltonian Monte

Carlo, which is implemented in **Stan** (Carpenter et al., 2017; Leos-Barajas and Michelot, 2018). Given the draws from the posterior distribution, posterior draws of the states can be obtained through the *forward-backward* algorithm, detailed in Frühwirth-Schnatter (2006).

Issues of label-switching within each internal state can be controlled by ordering of the production state means, or another parameter of interest. In particular, **Stan** has built-in functions that make it easier for the user to specify an ordering. However, an ordering at the internal state level to deter label-switching across internal states largely depends on the model formulation. A further attempt to deal with label-switching through specification of non-exchangeable priors may not work very well in cases of the multi-scale HMMs as these models require a lot of data, leading to the likelihood masking much of the influence of the prior (Betancourt, 2017b). Other issues when conducting inference for these models are directly related to problems associated with discrete mixture models - the posterior distribution is likely to be highly multimodal (a common problem in HMMs) and there can be issues related to estimability of the mixture components. As the true number of components is itself unknown, we may encounter issues with label degeneracy and poor posterior sampling due to a misspecified model formulation. The benefit of using **Stan**, and in particular Hamiltonian Monte Carlo, is that model specification and sampling performance, whether good or bad, manifests itself into various diagnostics. For a conceptual overview of HMC, see Betancourt (2017a).

3.6 Diving With a Tiger Shark

Tiger shark data were collected in Oahu, Hawai'i. The shark's depth was recorded in 0.5 m intervals every 2 s over a 23 day period, from March 9 to March 31, 2009. On some days, the tiger shark inhabited varied ranges of depth levels and performed many dives whereas other days it remained relatively constant in the water column and dove less. For sharks, movement in the water column is not easily segmented into types of dives. However, dives, or simply ascending or descending, can be an important part of a shark's behavior.

3.6.1 A day in the life

We extracted a depth position every ten minutes from the available data record and sequentially computed the absolute change in depth position, $y_t^* = |d_t - d_{t-1}|$, for the tiger shark in order to understand how the depths inhabited by the shark differ across days. On some days the tiger shark tends to remain in the same part of the water column for long periods of time while on others, it tends to move up and down more frequently throughout the day. We processed the data as described in Section 3.4.1 to produce $M = 144$ observations per day, across $K = 21$ days, excluding the first and last days for which only partial data records were available.

We fit a multi-scale HMM to account for switches in $\mathbf{y}_m = 2\mathbf{y}_m^*$ (as depth is recorded in 0.5 increments) at the ten-minute time scale as well as to account for how general behavior changes across days. We constructed a HMM with $V = 2$ internal states and $N_1 = N_2 = 3$ production states, common to the internal states. The $V = 2$ internal states are distinguished by the state-switching dynamics, $\{\boldsymbol{\Gamma}^v\}_{v=1}^2$, and we assume that the production state-dependent densities, f_n follow negative binomial distributions. That is, $f_{v,n}(y_{k,m}) \sim \text{Neg Binom}(\mu_n, \phi_n)$. We use the parametrization that is defined by a mean, μ_n , and dispersion parameter, ϕ_n , with $E(y_{k,m}|S_{k,m} = n) = \mu_n$ and $Var(y_{k,m}|S_{k,m} = n) = \mu_n + \mu_n^2/\phi_n$. The first observation stemming from the v^{th} internal state is assumed to be generated by the corresponding production state stationary distribution, $\boldsymbol{\nu}_v$. We take $\{\boldsymbol{\Gamma}^v(t)\}$ to be a function of time through a multinomial logistic regression transformation of each row in order to investigate how activity levels vary across the time of day. We kept the entries of the internal state t.p.m. $\boldsymbol{\Omega}$ fixed across segments.

We evaluated the likelihood using the forward algorithm in order to integrate over the unobserved latent states, which allows for use of a dynamic Hamiltonian Monte Carlo algorithm implemented in the software **Stan** and obtain draws from the joint posterior distribution of the multi-scale HMM directly. To account for potential label-switching, we imposed an ordering constraint on the production state-dependent parameters $\mu_1 < \mu_2 < \mu_3$. We assigned non-exchangeable priors for the means, $\mu_1 \sim t_3^+(5, 5)$, $\mu_2 \sim t_3^+(10, 10)$, $\mu_3 \sim t_3^+(30, 10)$, and exchangeable priors for the dispersion parameters $\phi \sim t_3^+(0, 1)$. For the entries of the t.p.m., we assigned $N(0, 1)$ priors to the coefficients

associated with time of day and $N(-2, 1)$ priors for the intercept terms. Three chains were run for 2,000 iterations, with half of the iterations serving as the warm-up period, leading to a total of 3,000 posterior draws for each parameter and effective sample sizes provided in the Appendix. For each parameter we obtained estimated \hat{R} values of one and no issues with respect to the HMC specific diagnostics (Betancourt, 2017a). We implemented the forward-backward algorithm to locally decode the sequence of production states, $S_{k,m}$, and internal states, H_k , and identify the likely states that produced the data. The production state-dependent densities are shown in Figure 3.1. The estimated production state dependent means and dispersion parameters, along with 95% credible intervals, for the internal and production state process and other parameters are provided in Table 3.2

Table 3.2 Means and 95% credible intervals of production state-dependent distribution parameter estimates.

Production State-Dependent Distributions					
μ_1	1.45	(1.17, 1.81)	ϕ_1	2.20	(1.34, 3.76)
μ_2	11.57	(10.01, 13.44)	ϕ_2	1.54	(1.23, 2.00)
μ_3	41.72	(38.60, 45.31)	ϕ_3	1.37	(1.19, 1.60)

The internal state process estimates demonstrate some dependence over time, as the diagonal entries are above 0.5.

$$\Omega = \begin{pmatrix} \mathbf{0.84}(0.66, 0.98) & \mathbf{0.16}(0.02, 0.34) \\ \mathbf{0.23}(0.07, 0.40) & \mathbf{0.77}(0.60, 0.93) \end{pmatrix} \quad (3.6)$$

As there are only 21 days used in the analysis, we naturally see wide 95% credible intervals at the internal state process level.

Using the forward-backward algorithm to produce estimates of the state probabilities, we evaluated the internal state probabilities across the 21 days, shown in Figure 3.2.

Posterior predictive checks along with residuals of the fitted model are presented in the Appendix.

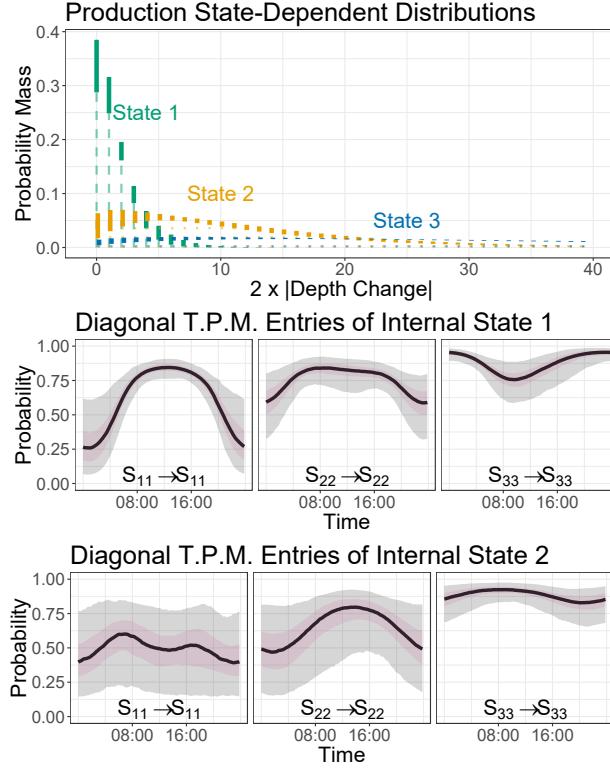


Figure 3.1 Comparing activity levels across days reflective of internal state 1 or 2. Top row: the local state probabilities of production state 1 given internal state 1 for day 20 and given internal state 2 for day 4. The black line denotes the median probabilities, the purple shaded area spans the 25th and 75th quantile, while the gray spans the 2.5th and 97.5th quantile. Bottom row: The value of 2 times the absolute change in depth colored by the median local state probability of production state 1 for day 20 and day 4.

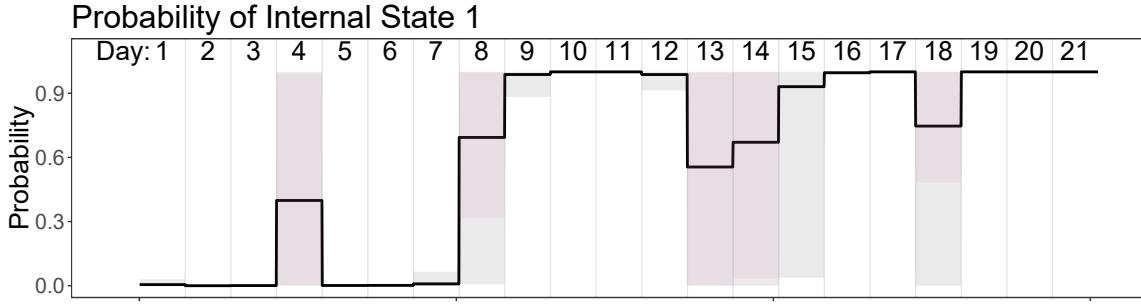


Figure 3.2 The internal state probability, $P(H_k = 1)$, across the 21 days. The black line denotes the mean probabilities, the purple shaded area spans the 25th and 75th quantile, while the gray spans the 2.5th and 97.5th quantile.

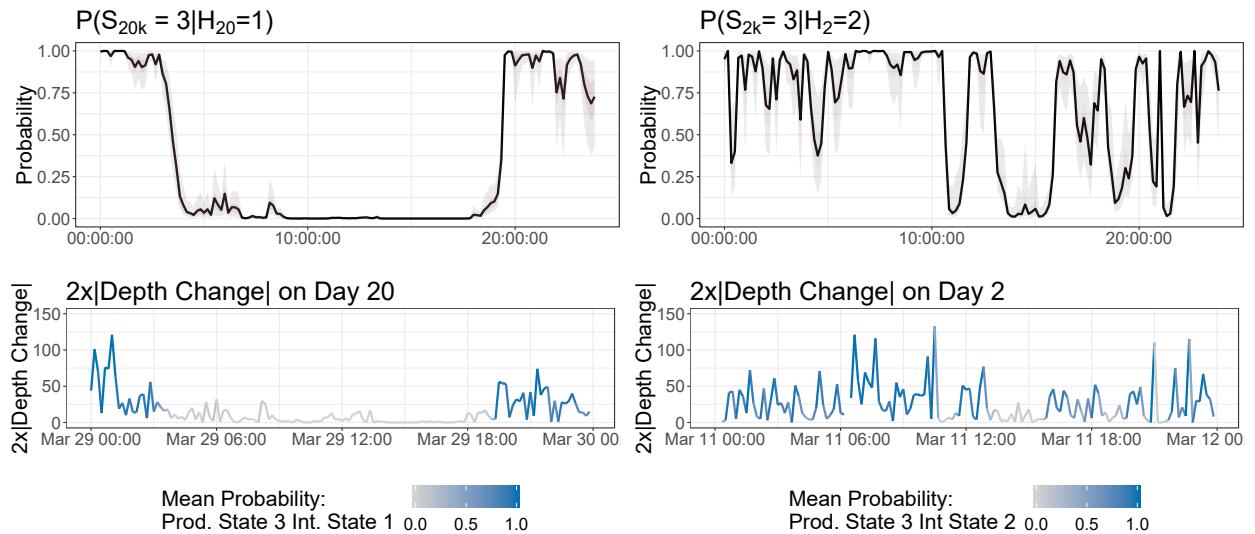


Figure 3.3 Comparing activity levels across days reflective of internal state 1 or 2. Top row: the local state probabilities of production state 1 given internal state 1 for day 20 and given internal state 2 for day 4. The black line denotes the median probabilities, the purple shaded area spans the 25th and 75th quantile, while the gray spans the 2.5th and 97.5th quantile. Bottom row: The value of 2 times the absolute change in depth colored by the median local state probability of production state 1 for day 20 and day 4.

3.6.2 Up and Down

In the previous section we modeled ranges of distance traveled by the tiger shark in the water column during 10-minute non-overlapping windows. As we explicitly chose to model diving patterns on a daily scale, the choice of segmenting the data by day inherently imposes some structure on when transitions can occur and the length of the process of interest (144 observations produced on a daily scale). Relaxing when transitions across internal states can occur provides much more flexibility, yet at a cost in that it further increases the difficulty in selection of the number of states and expands the types of processes that can be captured.

In this section, we model the ascent and descent in the water column of the tiger shark. As in the previous section, we sub-sample the raw data set to obtain an observation every 10 minutes and take pairwise differences of sequential observations to determine whether the shark was generally ascending, descending or stayed relatively constant, defined by random variables $y_t = 2 \cdot (d_t - d_{t-1})$. We multiply by two as descents are measured in 0.5 m. The resulting observations are integer-valued and either negative, positive or zero. Our primary objective in this section is to distinguish periods when the shark is actively moving around the water column, defined by a composition of relatively large ascents/descents, vs staying relatively constant, defined by small ascents/descents and/or no change in depth.

We first specify four production states to reflect the fine-scale movements of interest: small ascents, small descents, large ascents and large descents. Then, in accordance with our aims, we construct two internal states, such that internal state 1 is only composed of production states that correspond to small ascents and descents. Internal state 2 is only composed of the production states that correspond to relatively larger ascents and descents. Formally, we define a multi-scale HMM with two internal states, $V = 2$, and two internal-state dependent production states $N_1 = N_2 = 2$. We assign the states-dependent distributions a negative binomial distribution, $f_{i,j}(sgn(y_t) \cdot y_t) \sim \text{Neg Bin}(\mu_{i,j}, \phi_{i,j})$, using the parametrization that is defined by a mean, $\mu_{i,j}$, and dispersion parameter, $\phi_{i,j}$, with $E(sgn(y_t) \cdot y_t | S_{t,n}, H_t) = \mu_{t,n}$ and $\text{Var}(sgn(y_t) \cdot y_t | S_{t,n}, H_t) =$

$\mu_{t,n} + \mu_{t,n}^2/\phi_{t,n}$. Further, the production state that corresponds to large ascents is represented by a mixture of two negative binomial distributions.

In order to fit the model according to our aims, we specify restrictions on the means of the state-dependent distributions as well as specify whether $f_{i,j}$ models the non-negative or non-positive integers. Let $f_{1,1}$ and $f_{2,1}$ denote the production state-dependent distributions corresponding to ascents (incl. zero), with $f_{1,2}$ and $f_{2,2}$ corresponding descents (incl. zero). Then, we specify that $\mu_{1,1} < \mu_{2,1}$ and $\mu_{1,2} < \mu_{2,2}$. We conduct full Bayesian inference using **Stan** using three chains with 2000 iterations each, half of which correspond to the warm-up period, for a total of 3000 joint posterior draws. Estimates of the state-dependent parameters are given in Table 3.3.

Table 3.3 Marginal means and 95% credible intervals for the production state-dependent distributions for each internal state.

Sign	Internal State 1			Internal State 2		
+	$\mu_{1,1}$	5.51	(4.83, 6.29)	$\mu_{2,1}$	33.16	(31.20, 35.17)
	$\phi_{1,1}$	0.78	(0.65, 0.83)	$\phi_{2,1}$	1.13	(1.02, 1.25)
-	$\mu_{1,2}$	5.74	(4.94, 6.59)	$\mu_{2,2}$	36.15	(33.89, 36.95)
	$\phi_{1,2}$	0.69	(0.57, 0.81)	$\phi_{2,2}$	0.99	(0.90, 1.09)

The internal state process estimates are

$$\boldsymbol{\Omega} = \begin{pmatrix} \mathbf{0.96}(0.95, 0.98) & \mathbf{0.04}(0.02, 0.05) \\ \mathbf{0.02}(0.01, 0.03) & \mathbf{0.98}(0.97, 0.99) \end{pmatrix} \quad (3.7)$$

For internal state 1, the production state process is estimated as

$$\boldsymbol{\Gamma}_1 = \begin{pmatrix} \mathbf{0.47}(0.43, 0.52) & \mathbf{0.53}(0.57, 0.48) \\ \mathbf{0.53}(0.49, 0.58) & \mathbf{0.47}(0.42, 0.51) \end{pmatrix} \quad (3.8)$$

For internal state 2, the production state process is estimated as

$$\boldsymbol{\Gamma}_1 = \begin{pmatrix} \mathbf{0.45}(0.42, 0.48) & \mathbf{0.55}(0.52, 0.58) \\ \mathbf{0.60}(0.56, 0.63) & \mathbf{0.40}(0.37, 0.44) \end{pmatrix} \quad (3.9)$$

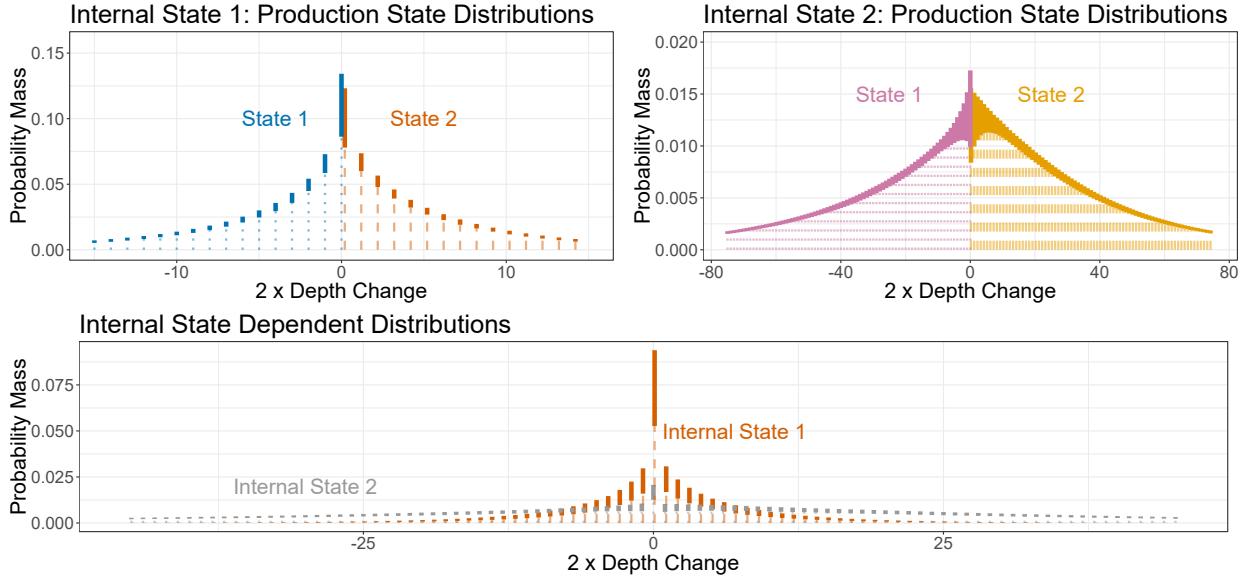


Figure 3.4 Top row: Point-wise credible intervals for the production state-dependent distributions for each internal state. Bottom row: Internal state-dependent distributions constructed as a weighted sum of the production state-dependent distributions.

Figure 3.4 demonstrates that the multi-scale HMM formulation is able to capture small ascents and descents into one internal state, on average moving $\pm 2\text{-}3\text{m}$ in the water column every 10 minutes. There is also rapid state-switching between the estimated production states, providing insight into how the shark moves in the water column to stay at a relative constant depth. Other movements are relegated into the other internal state, which captures relatively larger movements, on average $\pm 17\text{-}18\text{m}$. In Figure 3.5 we see that the model formulation is able to distinguish when the shark is remaining at a relatively constant depth level and when it is not. We further see in Figure 3.5 that the internal state probabilities, while not always close to 0 or 1, do not have much variability in their estimates, in contrast with the internal state estimates of the previous section.

While we allow for switches to occur at any point in time, the formulation used in this section is greatly assisted by the restriction placed on the means indicating that the biological process of interest must adhere to certain conditions. We capture switches across all combinations of internal and production states, yet are now poised to conduct inference at the internal state level (general

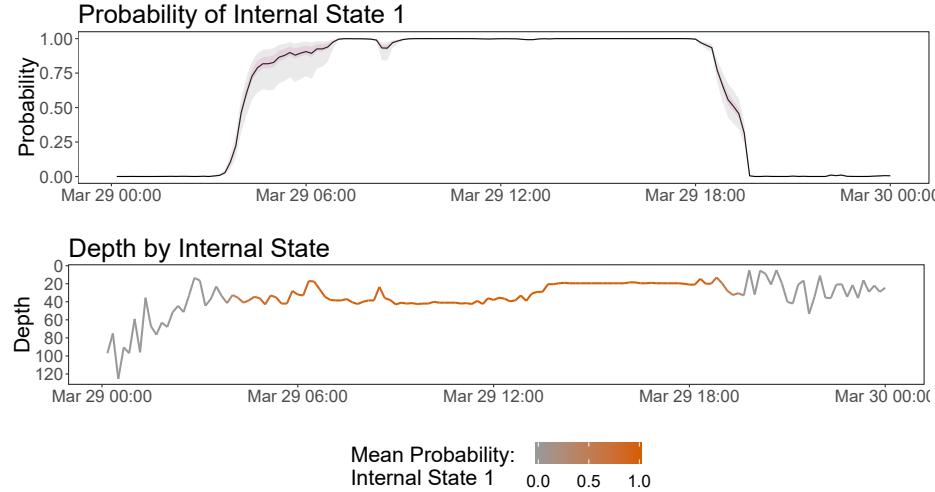


Figure 3.5 Top row: Mean probability of internal state 1 along with point-wise credible intervals. Bottom row: A subset of the dive profile of the tiger shark colored by the mean probability of internal state 1.

activity across the water columns) and/or at the production state level (identifying drivers of types of ascents and/or descents).

Through further model checking, via residuals and posterior predictive checks provided in the Appendix, we can see that the multi-scale HMM, while capturing the structure of the marginal distribution does not capture the autocorrelation structure completely. The assumption of a geometric distribution for the internal state-dwell time may be too unrealistic when considering its biological interpretation. The inability of the model presented here to capture the autocorrelation structure well enough may be alleviated by introducing covariates into the t.p.m., as in Section 3.6.1, or assuming another state-dwell time distribution for the internal state process (although this falls outside of the scope of this paper).

3.7 Discussion

We have expanded upon the description of the multi-scale HMMs described in Leos-Barajas et al. (2017) and Pirotta et al. (2018) and discussed conditions under which they are identifiable. The expanded HMM framework allows for identification of a broader class of patterns formed by

multiple fine-scale discernible actions/movements. In particular, the manner in which we can define these patterns, with some possibilities outlined in Section 3.4, is quite flexible. We demonstrate how these models can be applied using tiger shark depth data and how processing of the data can lend itself to application of the multi-scale HMM using either the formulation presented in Section 3.4.1 or 3.4.2.

In Section 3.6, we showed that the tiger shark dive patterns differed across days through the use of multi-scale HMM framework presented in Section 3.4.1. While there were only 21 complete days of data available, we demonstrate how inference about general behaviors at the daily timescale can be made by capturing the fine-scale patterns across 10-minute period. For application of the multi-scale HMM without specifying the switches across internal states, we simplified the problem by splitting movements across the water column into either non-negative (ascent) or non-positive (descent) values (with zero included in both) and were able to fit a simplistic multi-scale HMM that captures the larger scale behaviors of interest – minimal movements across the water column vs large movements across the water column in either direction.

Applications of these models to animal movement data in general allows one to capture more complex patterns than those captured by general state-space and hidden Markov models. However, applications of multi-scale HMMs, in particular to make inferences about biological systems, is still in its infancy with more investigation needed to fully understand how best to coordinate data processing and model formulation. Generally, because the multi-scale HMMs presented here are specially structured HMMs, model evaluation can be done in the same manner. The well-known problem of selecting the number of states for an HMM is exacerbated in this construction, as now we need to select the number of production and internal states, though general guidelines provided in Pohle et al. (2017) can be used along with domain expertise.

While we allow for variable duration in the internal state process in Section 3.4.2, we do so by assuming a geometric distribution. This has the advantage that we are no longer required to segment the time series a priori and yet it still adheres to the same framework as the basic HMM. In general, this approach is useful for capturing patterns that differ in the lengths of times in which

they are exhibited yet are still connected to the same general process. However, for identification of processes in which we want to connect a sequence of events to a larger scale process, assuming another integer-valued dwell-time distribution, e.g. Poisson or negative binomial, may better reflect the biological realities of the system in question. Regardless, any instance in which the internal state produces segments of varied lengths will require that the internal state duration be explicitly modeled.

3.8 References

- Betancourt, M. (2017a). A conceptual understanding to hamiltonian monte carlo. *arXiv:1701.02434*.
- Betancourt, M. (2017b). Identifying bayesian mixture models. Retrieved from https://betanalpha.github.io/assets/case_studies/identifying_mixture_models.html.
- Cappé, O., Moulines, E., and Ryden, T. (2005). *Inference in Hidden Markov Models*. Springer-Verlag, New York.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76.
- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov-Switching Models*. Springer-Verlag, New York.
- G. Alexandrovich, H. Holzmann, A. L. (2016). Nonparametric identification and maximum likelihood estimation for hidden markov models. *Biometrika*, 103:423–434.
- Holzmann, H., Munk, A., Suster, M., and Zucchini, W. (2006). Hidden markov models for circular and linear-circular time series. *Environ Ecol Stat*, 13:325–347.
- Holzmann, H., Munk, A., Suster, M., and Zucchini, W. (2015). Hidden markov models with state-dependent mixtures: minimal representation, model testing and applications to clustering. *Stat Comput*, 25:1185–1200.
- Hooten, M. B., Johnson, D. S., McClintock, B. T., and Morales, J. M. (2017). *Animal Movement: Statistical Models for Telemetry Data*. CRC Press, Boca Raton, FL.
- Leos-Barajas, V., Gangloff, E., Adam, T., Langrock, R., van Beest, F., Nabe-Nielsen, J., and Morales, J. M. (2017). Multi-scale modeling of animal movement and general behavior data using hidden markov models with hierarchical structures. *JABES*, 22:232–248.

- Leos-Barajas, V. and Michelot, T. (2018). An introduction to animal movement modeling using hidden markov models using stan for bayesian inference. *arXiv:1806.10639*.
- Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. (2004). Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, 85(9):2436–2445.
- Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., and King, R. (2017). Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *ASTA Advances in Statistical Analysis*, 101(4):399–438.
- Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. (2008). State-space models of individual animal movement. *Trends in Ecology and Evolution*, 23(2):87–94.
- Pirotta, E., Edwards, E. W. J., New, L., and Thompson, P. M. (2018). Central place foragers and moving stimuli: A hiddenstate model to discriminate the processes affecting movement. *Journal of Animal Ecology*, 22(3):270–293.
- Pohle, J., Langrock, R., van Beest, F. M., and Schmidt, N. M. (2017). Selecting the number of states in hidden markov models – pitfalls, practical challenges and pragmatic solutions illustrated using animal movement. *JABES*, 22(3):270–293.
- Teicher, H. (1963). Identifiability of finite mixtures. *Annals of Mathematical Statistics*, 34:1265–1269.
- Towner, A. V., Leos-Barajas, V., Langrock, R., Schick, R. S., Smale, M. J., Kaschke, T., Jewell, O. J. D., and Papastamatiou, Y. P. (2016). Sex-specific and individual preferences for hunting strategies in white sharks. *Functional Ecology*, 30(8):1397–1407.
- Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction using R*. Chapman & Hall, Boca Raton, FL, 2nd edition.

3.9 Appendix

3.9.1 Tuning parameter for internal state duration

For any N -state HMM, the amount of time that a given state is active before a switch occurs is referred to as the state-dwell time, denoted by the collection of random variables $\{D_n\}_{n=1}^N$, with the subscript n denoting the state-dwell time of the n^{th} state. In a basic HMM, D_n necessarily follows a geometric distribution. By extension, the framework that allows for switches across internal states at any point in time has state-dwell times, $\{D_v\}_{v=1}^V$, that also follow geometric distributions.

A geometric distribution at the internal state level can allow for rapid state-switching as the mode is one, implying the most frequent duration that an internal state is active is one time step. However, if the process of interest at the internal state level has a minimal amount of time that it must be exhibited, the model formulation can be altered by introducing a tuning parameter, x_v . The tuning parameter reflects the desire to only allow switches across internal states after a minimum number of observations, $\{x_v\}_{v=1}^V$, are produced. Let $w(d_v)$ denote the internal state-dwell time distribution for the v^{th} internal state, then with the additional restriction, the implied internal state-dwell time distribution is assumed to follow a truncated geometric distribution,

$$D_v \sim \frac{w(d_v)}{\sum_{d_v=x_v}^{\infty} w(d_v)} I(d_v \geq x_v) \quad (3.10)$$

Because the multi-scale HMM is flexible enough to capture an array of patterns, setting a tuning parameter x_v adds some structure to the modeling framework. Given a truncated geometric dwell-time distribution for each $\{D_v\}_{v=1}^V$ with minimum internal state length, $\{x_v\}_{v=1}^V$, we can construct a t.p.m. that models the evolution across the expanded production and internal state process at the observed time t , as in the basic HMM. For example, supposing $V = 2$ and given $x_1 = x_2 = 5$,

we can construct a $\left(\sum_{v=1}^V x_v N_v\right) \times \left(\sum_{v=1}^V x_v N_v\right)$ t.p.m. as follows,

$$\Delta = \left(\begin{array}{cc|cc|cc|cc|cc|cc} \mathbf{0}_{1,1} & \boldsymbol{\Gamma}^1 & \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} \\ \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \boldsymbol{\Gamma}^1 & \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} \\ \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \boldsymbol{\Gamma}^1 & \mathbf{0}_{1,1} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} \\ \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \boldsymbol{\Gamma}^1 & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} \\ \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \mathbf{0}_{1,1} & \delta_{1,1}\boldsymbol{\Gamma}^1 & \delta_{1,2}\boldsymbol{\Pi}^2 & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} \\ \hline \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,2} & \boldsymbol{\Gamma}^2 & \mathbf{0}_{2,2} & \mathbf{0}_{2,2} & \mathbf{0}_{2,2} \\ \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,2} & \mathbf{0}_{2,2} & \boldsymbol{\Gamma}^2 & \mathbf{0}_{2,2} & \mathbf{0}_{2,2} \\ \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,2} & \mathbf{0}_{2,2} & \mathbf{0}_{2,2} & \boldsymbol{\Gamma}^2 & \mathbf{0}_{2,2} \\ \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,2} & \mathbf{0}_{2,2} & \mathbf{0}_{2,2} & \mathbf{0}_{2,2} & \boldsymbol{\Gamma}^2 \\ \delta_{2,1}\boldsymbol{\Pi}^1 & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,1} & \mathbf{0}_{2,2} & \mathbf{0}_{2,2} & \mathbf{0}_{2,2} & \mathbf{0}_{2,2} & \delta_{2,2}\boldsymbol{\Gamma}^2 \end{array} \right) \quad (3.11)$$

As before, the entries $\delta_{i,j}$, for $i, j = 1, 2$, reflect the probability of remaining in or switching across internal states. The $\mathbf{0}_{i,j}$ entries denote matrices with zero entries of dimension $N_i \times N_j$, for $i, j = 1, \dots, V$.

3.9.2 Section 3.6.1 Tables and Plots

3.9.2.1 Posterior predictive checks

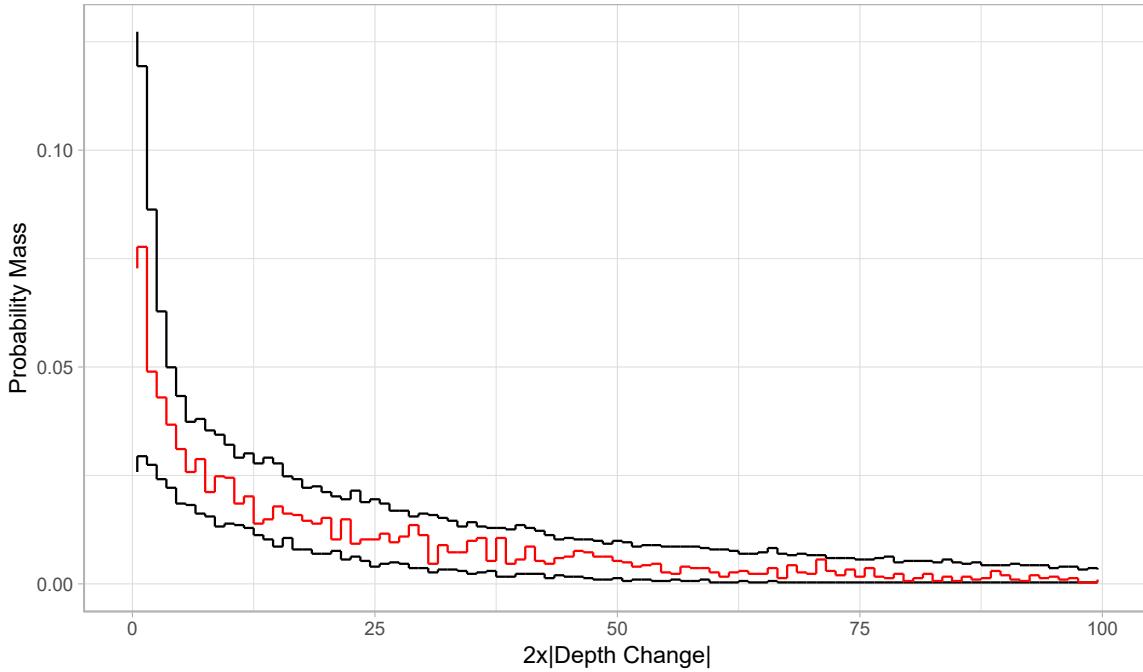


Figure 3.6 Histogram of observed values (in red) along with 95% pointwise posterior predictive intervals for observations $y = 0, 1, 2, \dots, 100$.

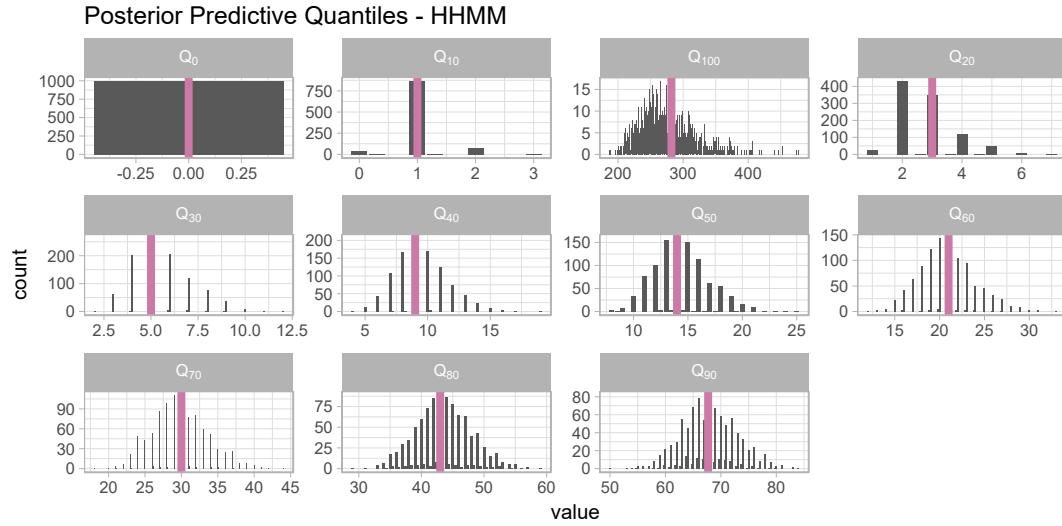


Figure 3.7 Distributions of the quantiles (in black) alongside the quantile of the observations (in purple).

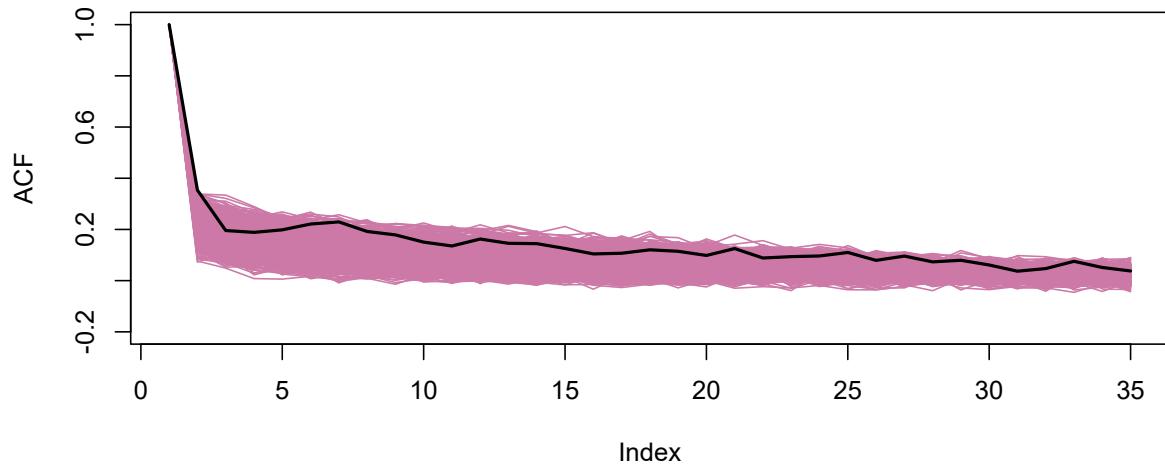


Figure 3.8 Autocorrelation structure of posterior predictive distributions (in purple) alongside the autocorrelation structure of the observation process (in black).

3.9.2.2 Residuals

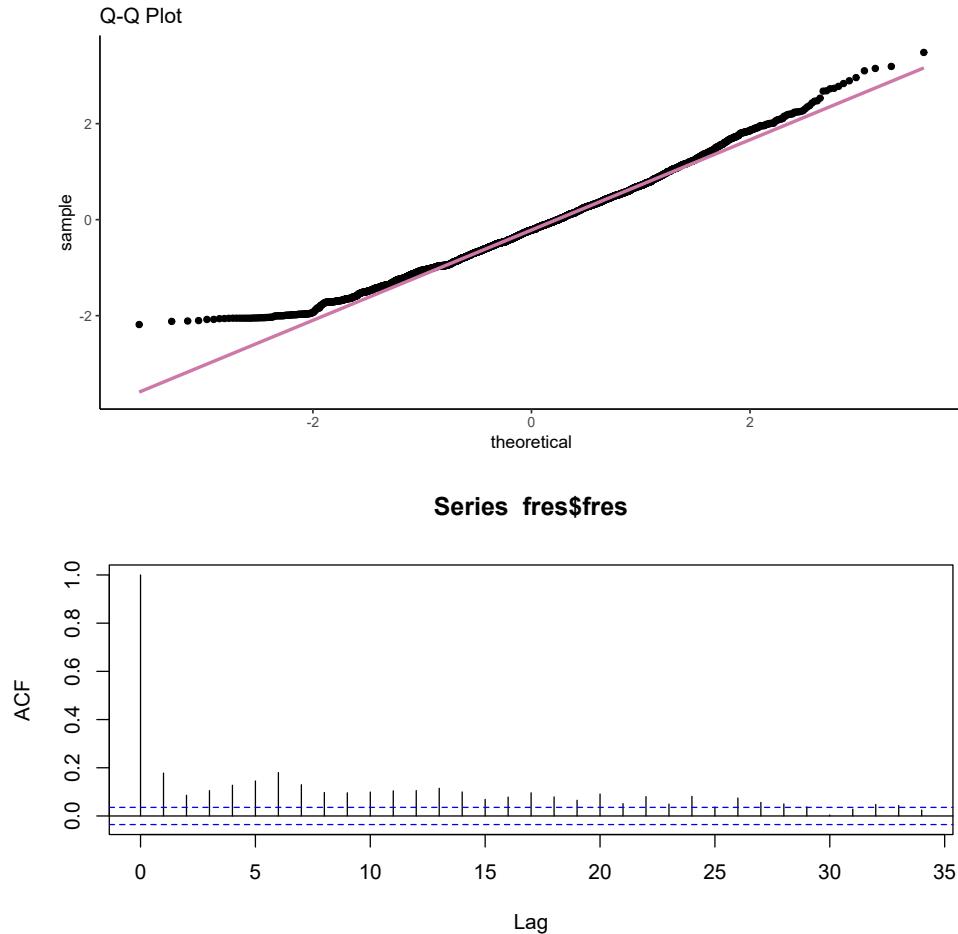


Figure 3.9 Top: QQ-plot of the residuals evaluated at the estimated marginal means of the parameters. Bottom: Autocorrelation structure of the residuals evaluated at the estimated marginal means of the parameters.

3.9.2.3 Summary of posterior draws

Table 3.4 Results of HMC posterior draws for model in Section 3.6.1. Values for the marginal means and chosen quantiles are reported, along with the effective number of samples, n_{eff} , and value of \hat{R} .

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
alpha[1,1]	0.84	0.00	0.08	0.66	0.79	0.85	0.90	0.98	1048.07	1.00
alpha[1,2]	0.16	0.00	0.08	0.02	0.10	0.15	0.21	0.34	1048.07	1.00
alpha[2,1]	0.23	0.00	0.08	0.07	0.17	0.23	0.29	0.40	2957.99	1.00
alpha[2,2]	0.77	0.00	0.08	0.60	0.71	0.77	0.83	0.93	2957.99	1.00
beta[1,1]	-1.22	0.01	0.35	-1.95	-1.45	-1.21	-0.98	-0.55	1206.10	1.00
beta[1,2]	0.06	0.01	0.26	-0.48	-0.10	0.06	0.24	0.55	2268.03	1.00
beta[1,3]	-0.53	0.01	0.51	-0.46	0.20	0.51	0.86	1.54	1151.51	1.00
beta[2,1]	-2.31	0.03	0.64	-3.89	-2.66	-2.19	-1.85	-1.36	620.47	1.01
beta[2,2]	0.78	0.01	0.49	-0.18	0.47	0.78	1.08	1.78	2525.20	1.00
beta[2,3]	2.74	0.04	0.97	0.58	2.15	2.86	3.42	4.31	703.29	1.01
beta[3,1]	-2.00	0.01	0.37	-2.76	-2.24	-1.99	-1.75	-1.28	970.28	1.00
beta[3,2]	-0.14	0.01	0.32	-0.80	-0.34	-0.14	0.07	0.46	2046.47	1.00
beta[3,3]	-0.26	0.01	0.41	-1.09	-0.54	-0.25	0.01	0.52	1201.35	1.00
beta[4,1]	-2.36	0.01	0.39	-3.25	-2.58	-2.32	-2.09	-1.71	829.39	1.01
beta[4,2]	-0.19	0.01	0.36	-0.93	-0.42	-0.18	0.04	0.49	1431.07	1.00
beta[4,3]	1.74	0.02	0.63	0.57	1.34	1.70	2.12	3.07	811.93	1.00
beta[5,1]	-3.57	0.01	0.51	-4.65	-3.89	-3.54	-3.22	-2.62	1359.61	1.01
beta[5,2]	1.26	0.02	0.68	-0.21	0.83	1.28	1.73	2.50	1150.07	1.00
beta[5,3]	-1.18	0.01	0.49	-2.09	-1.50	-1.19	-0.88	-0.19	2254.02	1.00
beta[6,1]	-2.59	0.01	0.43	-3.56	-2.86	-2.56	-2.30	-1.81	1026.65	1.00
beta[6,2]	-0.02	0.02	0.53	-0.96	-0.36	-0.06	0.30	1.12	1069.61	1.00
beta[6,3]	-0.71	0.01	0.46	-1.62	-1.01	-0.71	-0.42	0.22	1796.62	1.00
beta2[1,1]	-1.01	0.02	0.86	-2.70	-1.56	-1.02	-0.43	0.67	1562.44	1.00
beta2[1,2]	-0.14	0.02	0.83	-1.73	-0.68	-0.15	0.38	1.60	2312.33	1.00
beta2[1,3]	-0.64	0.04	1.03	-2.55	-1.37	-0.66	0.04	1.38	583.49	1.01
beta2[2,1]	-1.19	0.02	0.86	-3.09	-1.67	-1.15	-0.60	0.33	1304.89	1.00
beta2[2,2]	-0.27	0.02	0.89	-2.07	-0.85	-0.24	0.34	1.40	1832.74	1.00
beta2[2,3]	0.89	0.04	0.95	-1.04	0.30	0.87	1.52	2.81	660.73	1.01
beta2[3,1]	-2.78	0.02	0.71	-4.22	-3.24	-2.76	-2.32	-1.45	1763.31	1.00
beta2[3,2]	0.15	0.03	0.90	-1.56	-0.47	0.16	0.77	1.89	990.43	1.00
beta2[3,3]	-0.24	0.02	0.78	-1.79	-0.76	-0.25	0.26	1.29	2090.26	1.00
beta2[4,1]	-0.81	0.02	0.57	-1.90	-1.18	-0.82	-0.42	0.31	713.10	1.01
beta2[4,2]	0.40	0.01	0.52	-0.66	0.06	0.39	0.74	1.41	1775.48	1.00
beta2[4,3]	0.80	0.02	0.56	-0.30	0.42	0.80	1.17	1.89	1271.92	1.00
beta2[5,1]	-3.72	0.01	0.52	-4.87	-4.05	-3.69	-3.37	-2.82	1633.27	1.00
beta2[5,2]	-0.31	0.03	0.75	-1.71	-0.84	-0.35	0.19	1.24	507.77	1.01
beta2[5,3]	0.19	0.02	0.60	-1.00	-0.23	0.19	0.59	1.40	1202.18	1.00
beta2[6,1]	-2.44	0.02	0.52	-3.57	-2.75	-2.40	-2.07	-1.54	1101.24	1.00
beta2[6,2]	-0.44	0.01	0.49	-1.41	-0.75	-0.44	-0.13	0.55	2009.17	1.00
beta2[6,3]	0.38	0.01	0.51	-0.64	0.05	0.39	0.71	1.39	2191.91	1.00
init1[1]	0.22	0.00	0.17	0.01	0.08	0.17	0.31	0.63	4681.56	1.00
init1[2]	0.39	0.01	0.25	0.02	0.17	0.36	0.57	0.87	2429.59	1.00
init1[3]	0.40	0.01	0.25	0.02	0.18	0.38	0.59	0.89	2439.87	1.00
init2[1]	0.34	0.01	0.24	0.01	0.14	0.30	0.50	0.84	1468.20	1.00
init2[2]	0.28	0.00	0.21	0.01	0.10	0.23	0.42	0.78	2554.55	1.00
init2[3]	0.38	0.01	0.25	0.02	0.17	0.36	0.57	0.89	964.03	1.00
mu[1]	1.45	0.00	0.16	1.17	1.34	1.44	1.55	1.81	1251.65	1.00
mu[2]	11.57	0.03	0.85	10.02	11.00	11.51	12.10	13.44	1061.01	1.01
mu[3]	41.72	0.05	1.72	38.60	40.55	41.62	42.79	45.31	1293.69	1.00
d[1]	2.20	0.01	0.63	1.34	1.77	2.07	2.48	3.76	1809.75	1.00
d[2]	1.54	0.01	0.20	1.23	1.40	1.51	1.65	2.00	871.07	1.01
d[3]	1.37	0.00	0.10	1.19	1.29	1.36	1.43	1.60	1113.17	1.00
lp_	-11582.90	0.20	5.55	-11594.75	-11586.45	-11582.58	-11578.98	-11572.97	736.28	1.00

3.9.3 Section 3.6.2 Tables and Plots

3.9.3.1 Posterior predictive checks

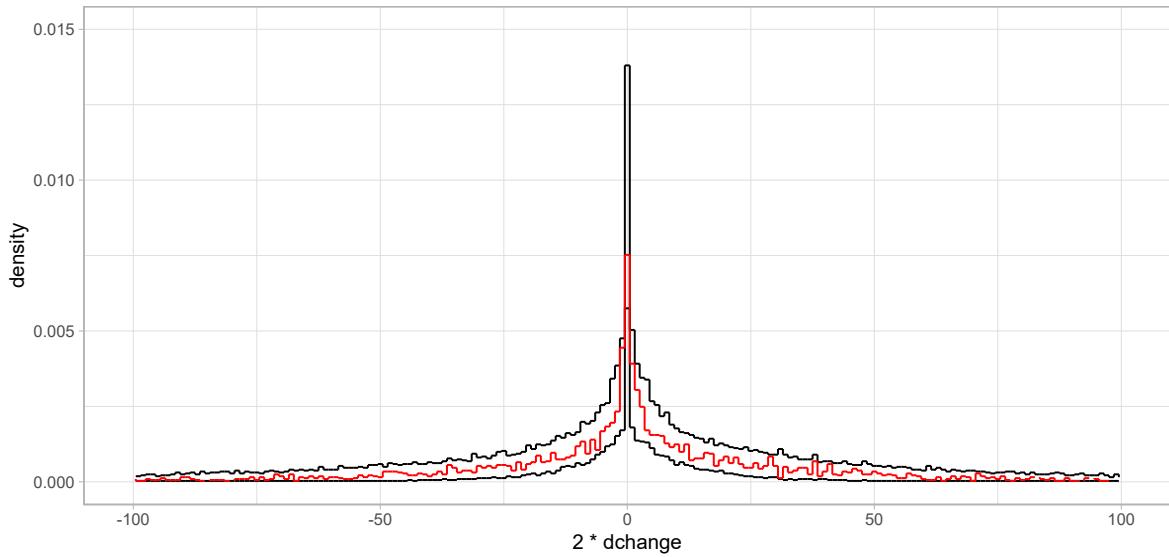


Figure 3.10 Histogram of observed values (in red) along with 95% pointwise posterior predictive intervals for observations $y = -100, -99, \dots, 99, 100$.

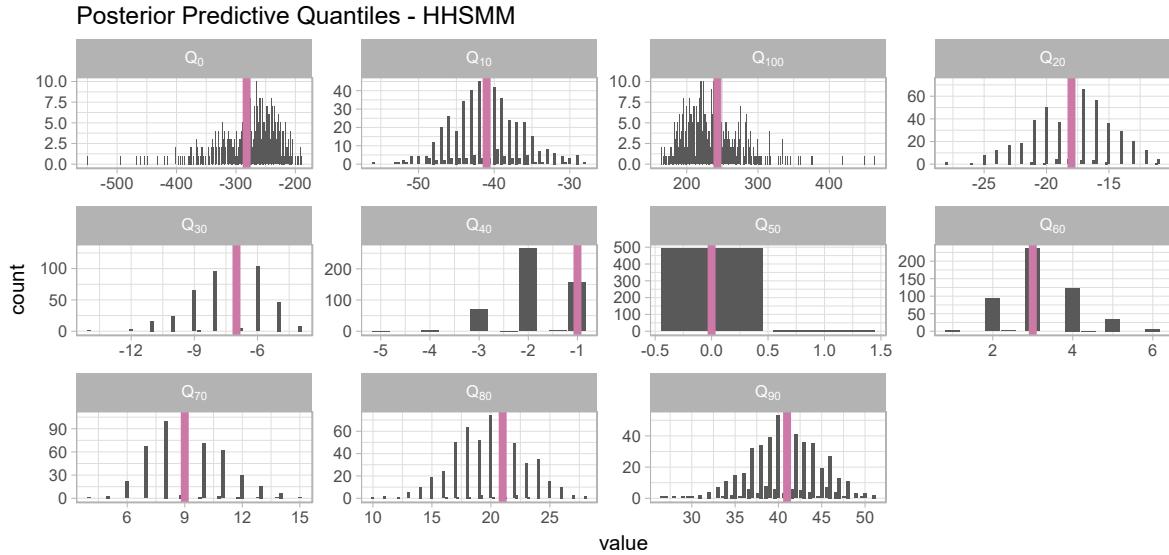


Figure 3.11 Distributions of the quantiles (in black) alongside the quantile of the observations (in purple).

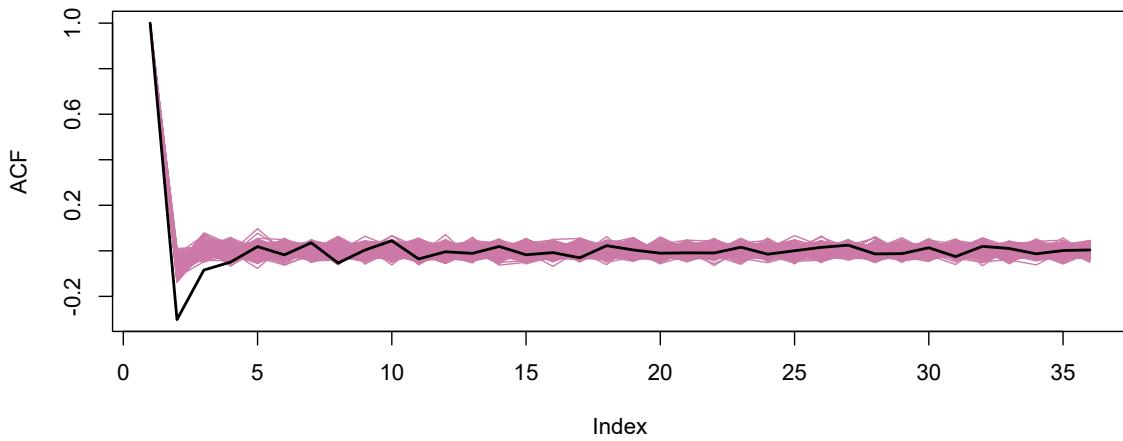


Figure 3.12 Autocorrelation structure of posterior predictive distributions (in purple) alongside the autocorrelation structure of the observation process (in black).

3.9.3.2 Residuals

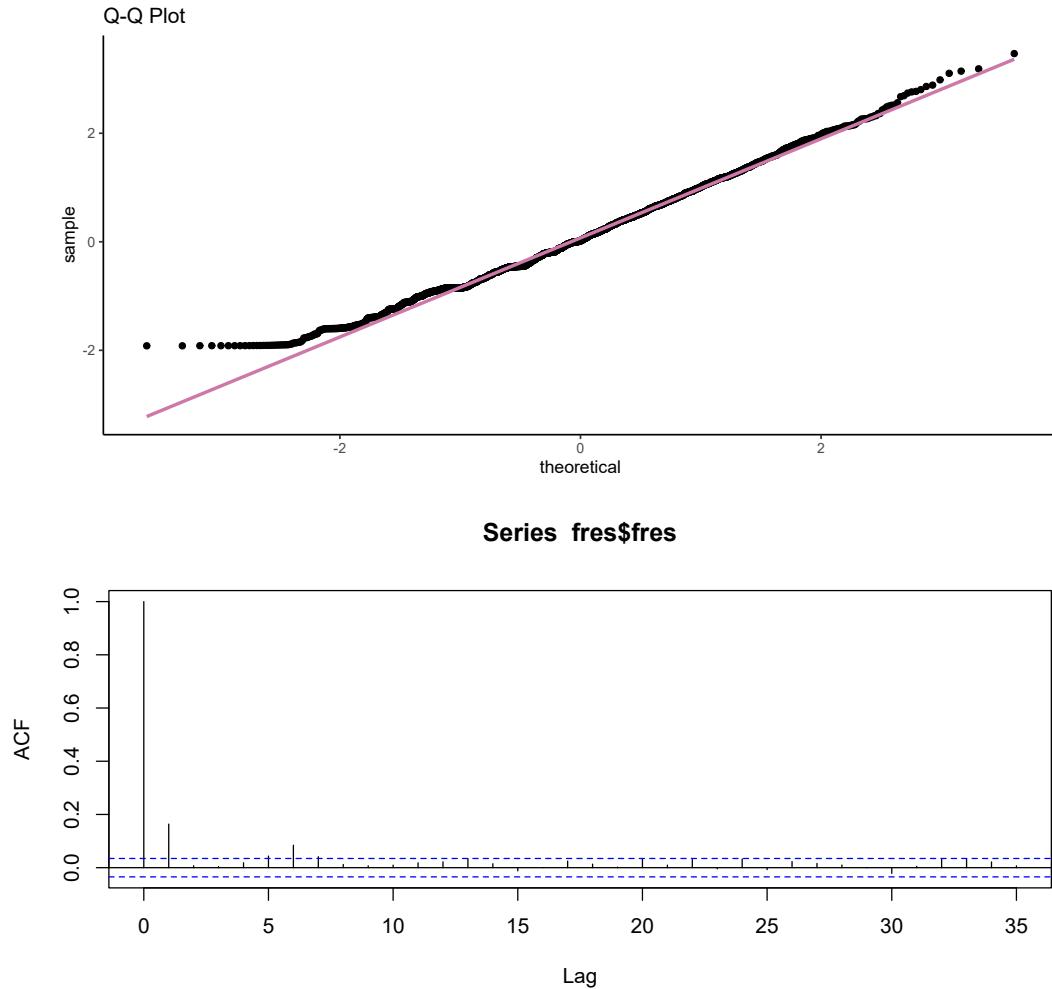


Figure 3.13 Top: QQ-plot of the residuals evaluated at the estimated marginal means of the parameters. Bottom: Autocorrelation structure of the residuals evaluated at the estimated marginal means of the parameters.

3.9.3.3 Summary of posterior draws

Table 3.5 Results of HMC posterior draws for model in Section 3.6.2. Values for the marginal means and chosen quantiles are reported, along with the effective number of samples, n_{eff} , and value of \hat{R} .

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
alpha[1,1]	0.96	0.00	0.01	0.95	0.96	0.96	0.97	0.98	4114.80	1.00
alpha[1,2]	0.04	0.00	0.01	0.02	0.03	0.04	0.04	0.05	4114.80	1.00
alpha[2,1]	0.02	0.00	0.00	0.01	0.02	0.02	0.02	0.03	4097.35	1.00
alpha[2,2]	0.98	0.00	0.00	0.97	0.98	0.98	0.98	0.99	4097.35	1.00
theta1[1,1]	0.47	0.00	0.02	0.43	0.46	0.47	0.49	0.52	5105.53	1.00
theta1[1,2]	0.53	0.00	0.02	0.48	0.51	0.53	0.54	0.57	5105.53	1.00
theta1[2,1]	0.53	0.00	0.02	0.49	0.52	0.53	0.55	0.58	5834.08	1.00
theta1[2,2]	0.47	0.00	0.02	0.42	0.45	0.47	0.48	0.51	5834.08	1.00
theta2[1,1]	0.45	0.00	0.02	0.42	0.44	0.45	0.46	0.48	8049.52	1.00
theta2[1,2]	0.55	0.00	0.02	0.52	0.54	0.55	0.56	0.58	8049.52	1.00
theta2[2,1]	0.60	0.00	0.02	0.56	0.59	0.60	0.61	0.63	6502.24	1.00
theta2[2,2]	0.40	0.00	0.02	0.37	0.39	0.40	0.41	0.44	6502.24	1.00
mus1[1]	5.51	0.00	0.38	4.83	5.25	5.49	5.76	6.29	6065.22	1.00
mus1[2]	33.16	0.01	1.02	31.20	32.45	33.15	33.84	35.17	5627.30	1.00
mus2[1]	5.74	0.01	0.42	4.94	5.45	5.72	6.01	6.59	4927.73	1.00
mus2[2]	36.15	0.01	1.21	33.89	35.31	36.08	36.95	38.61	6761.96	1.00
d1[1]	0.78	0.00	0.07	0.65	0.73	0.78	0.83	0.93	4121.19	1.00
d1[2]	0.69	0.00	0.06	0.57	0.64	0.68	0.73	0.81	4436.68	1.00
d2[1]	1.13	0.00	0.06	1.02	1.09	1.13	1.17	1.25	5252.41	1.00
d2[2]	0.99	0.00	0.05	0.90	0.95	0.99	1.02	1.09	5488.39	1.00
statdist[1,1]	0.50	0.00	0.02	0.47	0.49	0.50	0.51	0.53	4201.66	1.00
statdist[1,2]	0.50	0.00	0.02	0.47	0.49	0.50	0.51	0.53	4201.66	1.00
statdist[2,1]	0.52	0.00	0.01	0.50	0.52	0.52	0.53	0.54	6815.92	1.00
statdist[2,2]	0.48	0.00	0.01	0.46	0.47	0.48	0.48	0.50	6815.92	1.00
statalp[1]	0.34	0.00	0.05	0.25	0.31	0.34	0.38	0.45	6210.92	1.00
statalp[2]	0.66	0.00	0.05	0.55	0.62	0.66	0.69	0.75	6210.92	1.00
lp_	-15070.99	0.07	2.60	-15076.83	-15072.61	-15070.71	-15069.08	-15066.85	1311.82	1.00

CHAPTER 4. INCORPORATING BODY CONDITION INTO THE ANALYSIS OF ANIMAL MOVEMENT

Vianey Leos-Barajas¹, Mark S. Kaiser¹, Agustina di Virgilio², Juan M. Morales²

¹Department of Statistics, Iowa State University, USA

²INIBIOMA-CONICET, Argentina

4.1 Abstract

A long-sought goal in ecology is to connect movement with population dynamics. For many species and especially for ungulates, there is a known link between condition (e.g. fat reserves) and the probability of survival and reproduction. Assuming a particular genetic makeup and physiology, condition reflects the history of behavioral decisions, including movement and habitat use. However, the condition of an animal can also have a direct implication on the types of movements that it performs and the habitats that it visits. Movement data for ungulates are typically collected at a fine temporal scale, e.g. a position recorded by a GPS device every five or ten minutes. However, fat reserves cannot be measured remotely and must be done manually. This in turn creates a mismatch in the temporal scale at which the two data streams are observed, i.e. every five minutes for movement vs approximately once a month for condition. Further, the temporal mismatch leads to various challenges when jointly modeling the two processes.

For the movement model, we use discrete-time, finite-state hidden Markov models (HMMs) with the positional data of the sheep serving as the observation process and the underlying state process serving as a proxy for behaviors of interest. To incorporate condition as a potential covariate affecting the movement, and thus behavioral, process, we make use of the physiological equations that describe the evolution of body fat in Merino sheep in order to predict daily values of the condition process. The physiological equations are expressed as a function of the states inferred by

HMM, as well as the distance that the sheep travels. Altogether, we present a general modeling framework the describes the interaction between condition and movement, using Merino sheep as a case study.

4.2 Introduction

An animal's movements are driven, in part, by biological necessity such as the need to forage and rest. Underlying the desire to fulfill a biological need is the actual ability to move in the manner required to achieve the task. For instance, a highly migratory species may travel for days on end, yet they can only do so if they are in good enough condition. In most cases, however, models applied to the analysis of animal movement do not incorporate an animal's condition as a factor that explains the movement, and thus behavioral, process.

Most animal movement data is collected with the intention of gaining insight into the underlying fine-scale behavioral process. The observed data, whether from GPS or accelerometers, can be thought to stem from the animal exhibiting a set of N behaviors of interest. An animal that forages may exhibit area-restricted search patterns, while traveling involves traversing longer distances. Movement data is frequently modeled via hidden Markov models, or other types of state-space models, as they relate the observations to underlying latent processes that can serve as proxies for an animal's behavior (Patterson et al., 2017; Hooten et al., 2017). If the animal's condition is known and can be quantified, then incorporating condition as a driver of the underlying behavioral process is possible (Patterson et al., 2009, 2017; Pirotta et al., 2019) . However, in the approaches developed to date, movements do not feed back into the varying condition process. Further, frequent, direct observation of the condition of an animal is unlikely to be feasible. Some condition processes evolve quickly relative to the behavioral processes occurring at a fine temporal scale, such as hunger, but other processes, like body fat, evolve more slowly. Yet in many animals, such as marine mammals and ungulates, condition is related to individual fitness and can be an important factor in understanding how the behavioral process evolves over time.

For ungulates, data on condition may be collected on an infrequent basis, and matching the condition process to the fine-scale movement process is challenging due to the temporal mismatch. For example, in a current experiment with Merino sheep, body fat which is a proxy for condition can be measured approximately every 30 days, while locations are recorded every 5 min. In this and similar scenarios, it is unclear how to incorporate infrequent condition observations into a model of animal movement with finer temporal resolution. For situations in which the physiological dynamics have been well studied, such as the evolution of body fat in Merino sheep, there exist theoretical equations that describe the manner in which some condition processes evolve, which makes it possible to produce predictions of the condition process at the temporal scale needed for inclusion as a covariate in the movement model.

In this paper, we (i) use the infrequent observations of condition to provide predicted values of condition on a daily scale and (ii) construct a model for the fine-scale movement process with predicted condition included as a potential driver. For demonstration purposes we focus on Merino sheep. In Section 2, we describe the movement and condition data. In Section 3 we first introduce the movement and physiological condition models separately, and then combine them into one overall model. In Section 4 we detail how to fit the model and in Section 5 provide methods to assess the adequacy of model and predicted quantities. In Section 6 we simulate condition data at a monthly time scale and fit the model. Section 7 contains concluding remarks and future directions.

4.3 Data Structure

We construct a general framework to model two types of data sources, positional data and physiological data. The movement data, specifically positional data, are taken to be recorded at a fine temporal scale, such as every five minutes, whereas the physiological data are taken to be recorded on a more course scale, such as once a month. Both assumptions are in line with the type of data that are typically collected in movement ecology.

GPS data are commonly collected on a fine temporal scale as the devices needed are small, have long battery lives and provide positions of the animal with small to negligible measurement

error. The position of the sheep is reported every five minutes for multiple months at a time. In order to connect the positions of the animal to behaviors of interests, we transform the positional data into step lengths and turning angles (Morales et al., 2004). Step lengths are calculated as the shortest distance between two consecutive positions, while turning angles are calculated using three consecutive positions and reported in the interval $(-\pi, \pi]$.

In this application, we use percentage of body fat (in kg) in Merino sheep as a proxy for the sheep's overall body condition. We suppose that the measure of body fat percentage is collected once a month, approximately every 30 days.

4.4 Model

To construct a joint model for movement and condition, movement of the animal enters the condition process through a transfer function, while the condition of the animal is included in the movement process as a covariate. Given that movement is observed at a finer temporal scale than is condition, we denote observations of condition over time by $\{g_t\}_{t=1}^T$ and observations of movement observed in the interval $[t, t + 1)$ by $\{d_{t,k}\}_{k=1}^K$ for distance and $\{a_{t,k}\}_{k=1}^K$, for turning angle, for $K \in \mathbb{N}$.

4.4.1 Movement

We model movement using a hidden Markov model (HMM) with $N = 3$ states. An HMM is a stochastic time series model composed of a state-dependent observation process, $\mathbf{y}_{t,k} = (d_{t,k}, a_{t,k})$, and an underlying state process, $\{S_{t,k} : k = 1, \dots, K_t; t = 1, \dots, T\}$. In particular, the HMM is completely defined by three components: the state-dependent densities, a transition probability matrix (t.p.m.) and the initial state distribution (Zucchini et al., 2016).

4.4.1.1 State-Dependent Distributions

A basic HMM applied to a univariate time series takes the observations to be conditionally independent given the states. For our 3-state HMM, we go further by assuming contemporaneous conditional independence for $d_{t,k}$ and $a_{t,k}$ given the state $S_{t,k}$, rather than specify a joint state-dependent distribution for these two components of movement. The state-dependent distributions for $d_{t,k}$ and $a_{t,k}$ are given as follows,

$$f_d(d_{t,k}|S_{t,k} = n) \sim \text{gamma}(\mu_n, \nu_n)$$

$$f_a(a_{t,k}|S_{t,k} = n) \sim \text{von Mises}(\psi_n, \kappa_n)$$

We express the gamma distribution in terms of its mean parametrization, such that $\{\mu_n\}_{n=1}^N$ and $\{\nu_n\}_{n=1}^N$ denote the mean and standard deviation of the N state-dependent distributions, respectively. For now we assume that the step lengths are strictly positive although in practice, a point mass can be placed on zero in the state-dependent step length distributions if the sheep does not move at all during some five-minute intervals. The support of the von Mises distribution is assumed to be the interval $(-\pi, \pi]$.

4.4.1.2 Transition Probability Matrix and Initial State Distribution

For the HMM, we assume a first-order Markov property for the evolution of the state sequence, so that $S_{t,k}$ depends on other states only through $S_{t-1,k}$. We further allow the condition of an animal during the t^{th} interval, g_t , to affect the manner in which the states are generated, by incorporating g_t as a covariate in models of the entries of the t.p.m. We do assume that condition is constant within each interval $[t, t+1)$. We denote the t.p.m. during the t^{th} period as $\Gamma^t(g_t)$, which has entries $\gamma_{ij}^t(g_t) = \Pr(S_{t,k} = i | S_{t-1,k} = j, g_t)$ for $i, j = \{1, 2, 3\}$. One manner in which to incorporate the covariate value is to use a multinomial logit link. We map the entries of the t.p.m. onto the real line and introduce covariates, g_t , in the following manner,

$$\gamma_{ij}^t(g_t) = \frac{\exp(\rho_{ij}^t)}{\sum_{j=1}^3 \exp(\rho_{ij}^t)}, \quad \text{where} \quad \rho_{ij}^t = \begin{cases} \tau_0^{(ij)} + \tau_1^{(ij)} g_t & \text{if } i \neq j; \\ 0 & \text{otherwise.} \end{cases}$$

With this formulation, we can estimate the effect that the condition g_t may have on the probabilities that sheep transition between traveling, foraging or resting behaviors in the interval $[t, t + 1]$.

To initiate the state process, we must further estimate or specify the initial distribution $\boldsymbol{\delta}$, which is a vector of probabilities with entries $\delta_i = \Pr(S_{1,1} = i)$, for $i = 1, \dots, N$. One way to estimate the initial distribution is to have it be the stationary distribution, (i.e. the solution to $\boldsymbol{\delta} = \boldsymbol{\delta}\Gamma$). Further, we can also compute the marginal stationary distribution, and $\boldsymbol{\delta} = \boldsymbol{\delta}\Gamma(g_t)$, so that the assumption of stationarity is conditioned on the value of the covariate g_t .

4.4.2 Body Condition (% of fat)

The percentage of body fat for Merino sheep is measured sparsely throughout the year which recaptures the sheep due to a lack of available devices that can measure this quantity remotely. Ideally, the percentage of body fat would be collected at a finer temporal scale, such as once per day. Assuming that the condition of the sheep is reasonably constant throughout a day, we could then include the daily condition values into the movement model in order to understand what the effect of condition is on the fine-scale movement process. However, over a period of T days, condition is not likely to be observed every day, but only on a subset of days. The observed condition process is denoted as $\{g_{P_u} : u = 1, \dots, U; P_1 < P_2 < \dots < P_U\}$. We take $P_1 = 1$, denoting that the process begins on the first day in which we can observe condition, while $P_U \leq T$. For ruminant ungulates such as the Merino sheep, daily changes in body fat can be expressed via a set of deterministic equations that are particular to their physiology. We use these equations to predict the body fat percentages on the days when this quantity is not observed.

The value of g_t , percentage of body fat on the t^{th} day, is a ratio of the underlying body fat (in kg), b_t , and overall body mass (in kg), w_t . Body mass is assumed to be a sum of b_t and lean mass (in kg), m_t , giving $w_t = b_t + m_t$. In order to describe the evolution of the dynamics of g_t , we first need to understand how the sheep gains or loses energy, and subsequently converts these into gains or losses in body fat and lean mass. Of interest is then the energy balance, E_t which is a function of three components: energetic intake, I_t , movement (locomotion) costs, L_t , and daily maintenance

costs, C_t . Energetic intake further depends on identifying periods in which the sheep is foraging, while locomotion cost is a function of the total distance traveled, \mathbf{d}_t , scaled to body mass.

4.4.2.1 Energy Intake from foraging

For grazing dynamics associated with the sheep exhibiting foraging behavior, which we can obtain from the movement model, we denote energy intake as I , expressed in MJ*min⁻¹. The value of I is a function of forage biomass, V (g*m⁻²), forage quality, Q (MJ*g⁻¹), grass biomass at which intake is half maximum, b (g*m⁻²), the maximum instantaneous cropping rate, R_{max} , in g*min⁻¹ and the maximum amount of time spent feeding in minutes, t_{max} . Given these quantities, I is defined by Wilmhurst et al. (2000), as

$$I = Q * \frac{R_{max} * V}{b + V} * t_{max}$$

For Merino sheep, we let $Q = 0.0152$ MJ*g⁻¹, $b=30.8$ g*m⁻², $R_{max} = 7.02$ g*min⁻¹, $V=100$ g*m⁻² and $t_{max} = 5$ minutes.

4.4.2.2 Movement costs

For movement, general locomotion costs, denoted L_t and expressed in MJ*km⁻¹, for each kilometer traveled depends on body mass, W_t , as $L_t = 0.01243 * W_t^{0.66}$. This does not depend on the specific type of behavior exhibited, but only the entire distance traveled by the sheep. The distance traveled is estimated through GPS recordings, with the total distance (in km), $\sum_{k=1}^K d_{t,k}$, represented as a sum of distance traveled between five-minute positions during the t^{th} day. Although the assumption of a straight line distance traveled between consecutive positions leads to some underestimation of total distance traveled, the fine scale of recorded movement data (5 min) should render such underestimation negligible relative to other sources of uncertainty in the problem.

4.4.2.3 Energetic Balance

Overall, sheep increase energy levels according to their foraging intake. Energy levels decrease due to locomotion cost and due to daily maintenance $C_t = 0.445*W_t^{0.75}$ MJ per day. The energy balance, $E_t(\mathbf{S}_{t-1}, \mathbf{d}_{t-1})$, is expressed as

$$E_t(\mathbf{S}_{t-1}, \mathbf{d}_{t-1}) = \left[\sum_{k=1}^K \mathbf{I}(S_{t-1,k} = \text{foraging}) \right] I - \left[\sum_{k=1}^K d_{t-1,k} \right] L_{t-1} - \frac{K}{288} C_{t-1} \quad (4.1)$$

4.4.2.4 Body Fat and Lean Muscle Mass

The amount of fat metabolized by an animal depends on E_t . If E_t is positive, the extra energy is used to generate protein and fat. From the total extra energy, 80% goes to fat with a conversion of 38.12 MJ/kg and 20% goes to protein with a conversion of 22.64MJ/kg (Robbins, 1993). If E_t is negative, fat and protein are consumed. The evolution of body fat, b_t , is then,

$$\begin{aligned} b_t &= b_{t-1} + f_b(\mathbf{S}_{t-1}, \mathbf{d}_{t-1}) \\ f_b(\mathbf{S}_{t-1}, \mathbf{d}_{t-1}) &= \frac{.8 * E_t}{38.12}, \end{aligned} \quad (4.2)$$

and the evolution of muscle mass, m_t , is given by,

$$\begin{aligned} m_t &= m_{t-1} + f_m(\mathbf{S}_{t-1}, \mathbf{d}_{t-1}) \\ f_m(\mathbf{S}_{t-1}, \mathbf{d}_{t-1}) &= \frac{.2 * E_t}{22.64}. \end{aligned} \quad (4.3)$$

4.4.2.5 Body Fat Percentage

We use equations (4.1), (4.2) and (4.3) to obtain predictions of g_t on days that the value is not observed. The value of g_t depends on the amount of predicted body fat at time t , b_t , and the predicted weight of the sheep, w_t , with body fat mass typically ranging from $0.1W_t$ to $0.3W_t$ (Moen et al., 1997; Delgiudice et al., 2001). The predicted body fat percentage for the t^{th} day is estimated as,

$$g_t = b_t / (b_t + m_t). \quad (4.4)$$

Equation (4.1) gives the energetic balance E_t as a function of known foraging events. For free-roaming animals, foraging events are unlikely to be directly observed and must typically be inferred

from their observed movements patterns (Morales et al., 2004; Hooten et al., 2017). In the following section, we replace the known foraging events in equation (4.1) with probabilities that the animal's movements are associated with foraging behavior.

4.4.3 Joint model

As noted in Section 3.2, the model for condition in equation (4.1) relies on the ability to detect whether the sheep is foraging. As the states (behaviors) of the movement process must be predicted, the component $\sum_{k=1}^K I(S_{t,k} = \text{foraging})$ in equation (1) is predicted using $\widehat{\Pr}(S_{t,k} = \text{foraging}|\cdot)$. Functions of this quantity, which include E_t , b_t and m_t , may be expressed as linear transformations of $\sum_{k=1}^K \widehat{\Pr}(S_{t,k} = \text{foraging}|\cdot)$. Figure 4.1 displays the dependence structure of the joint processes $\{Y_t, S_t, g_t\}$.

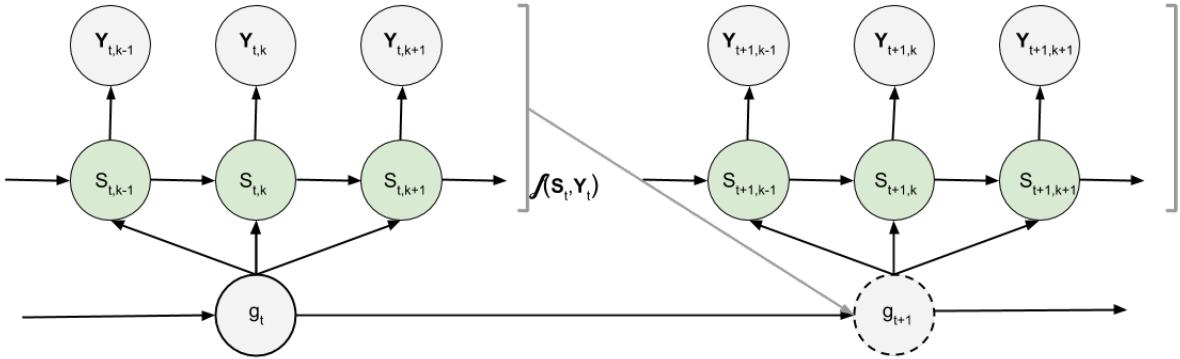


Figure 4.1 Diagram of the joint model for condition and behavior, with the dashed circle representing a predicted value of the condition process.

Given two observations of condition observed P days apart, g_t and g_{t+P} our aim is to predict the values of condition that are not observed, $\{g_{t+1}, \dots, g_{t+P-1}\}$, using the equations presented in Section 4.4.2. One manner to produce predictions is to begin with b_t and m_t and update the predictions forward in time. We update the predictions for body fat and lean muscle mass using equations (2) and (3) and then take $g_{t+1} = b_{t+1}/(b_{t+1} + m_{t+1})$.

To construct the likelihood of the joint model, we first assume that on day t , the state $S_{t,1}$ is generated according to the initial state distribution $\delta(g_t)$ and is independent of the state $S_{t-1,K}$. Let $f(\mathbf{y}_{t,k}|S_{t,k} = n) = f_d(d_{t,k}|S_{t,k} = n)f_a(a_{t,k}|S_{t,k} = n)$ and $\mathbf{B}(\mathbf{y}_{t,k})$ be an $N \times N$ diagonal matrix with entries $B_{nn}(\mathbf{y}_{t,k}) = f(\mathbf{y}_{t,k}|S_{t,k} = n)$ for $n = 1, \dots, N$. Using the state probabilities $\Pr(S_{t,k} = n|\mathbf{y}_t) = \phi_{t,k}(n)$, the predicted value of E_t , expressed in equation (4.1), is given by replacing the term $\sum_{k=1}^K I(S_{t-1,k} = \text{foraging})$ with $\sum_{k=1}^K \phi_{t,k}(\text{foraging})$ and subsequently used to produce predictions of g_t . For a single time series, given the sequence of observed and predicted values of condition, the likelihood of the HMM can be expressed a matrix product,

$$\mathcal{L} = \prod_{t=1}^T \delta(g_t)^\top \mathbf{B}(\mathbf{y}_{t,1}) \left[\prod_{k=1}^K \Gamma(g_t) \mathbf{B}(\mathbf{y}_{t,k}) \right] \mathbf{1} \quad (4.5)$$

with $\mathbf{1} = (1, \dots, 1)$ denoting a vector of length N . In particular, we can express the likelihood as a product of likelihoods for the observation process on day t , by letting

$\mathcal{L}_t = \delta(g_t)^\top \mathbf{B}(\mathbf{y}_{t,1}) \left[\prod_{k=1}^K \Gamma(g_t) \mathbf{B}(\mathbf{y}_{t,k}) \right] \mathbf{1}$ so that $\mathcal{L} = \prod_{t=1}^T \mathcal{L}_t$. As equation (4.5) requires the state probabilities, we use both the *forward* and *backward* algorithm to sequentially evaluate the likelihood in Section 4.5 (Zucchini et al., 2016).

4.5 Fitting the Joint Model

4.5.1 Likelihood Evaluation

Beginning with day 1, assuming we observe the body fat percentage, g_1 , we evaluate the likelihood, \mathcal{L}_1 , via the forward algorithm for the sequence of observations, $\{\mathbf{y}_{1,k}\}_{k=1}^K$. We define the sequence of *forward variables*, $\{\alpha_{1,k}\}_{k=1}^K$, starting at time $t = 1, k = 1$,

$$\alpha_{1,1} = \delta(g_1) \mathbf{B}(\mathbf{y}_{1,1}), \quad \text{where } \alpha_{1,1}(n) = \Pr(S_{1,1} = n, \mathbf{y}_{1,1})$$

At time $t = 1, k$, we have,

$$\alpha_{1,k} = \alpha_{1,k-1} \Gamma \mathbf{P}(\mathbf{y}_{1,k}), \quad \text{where } \alpha_{1,k}(n) = \Pr(S_{1,k} = n, \mathbf{y}_{1,1}, \dots, \mathbf{y}_{1,k})$$

The likelihood, \mathcal{L}_1 , is obtained by summing over $\alpha_{1,K}$,

$$\mathcal{L}_1 = \sum_{n=1}^N \alpha_{1,K}(n) = \boldsymbol{\alpha}_{1,K} \mathbf{1}.$$

where $\mathbf{1}$ is a $N \times 1$ vector of ones.

Suppose that the value of g_2 is not observed and must be predicted in order to evaluate the likelihood of day 2, \mathcal{L}_2 . The predicted value of g_2 is a function of the state sequence, \mathbf{S}_1 , that produced the observations \mathbf{y}_1 . In the HMM literature, inferring the underlying state sequence is called *state decoding*. Here we employ *local* state decoding, which provides estimates of $\Pr(S_{1,k} = n | \mathbf{y}_1)$, for $n \in \{1, \dots, N\}$. For state decoding of the observation process on day 1, we make use of the *forward-backward algorithm*, which uses the forward variables $\{\boldsymbol{\alpha}_{1,k}\}_{k=1}^K$ and the backward variables, $\{\boldsymbol{\beta}_{1,k}\}_{k=1}^K$.

To compute the backward variables, we begin at $t = 1, k = K$, such that $\boldsymbol{\beta}_{K,1} = \mathbf{1}$, for $\mathbf{1} = (1, 1, 1)$. At $t = 1, k \neq K$, we have,

$$\boldsymbol{\beta}_{1,k} = \mathbf{P}(\mathbf{y}_{1,k}) \boldsymbol{\beta}_{1,k+1}, \quad \text{where } \beta_{1,k}(n) = \Pr(S_{1,k} = n | \mathbf{y}_{1,k+1}, \dots, \mathbf{y}_{1,K})$$

Given $\{\boldsymbol{\alpha}_{1,k}\}_{k=1}^K$ and $\{\boldsymbol{\beta}_{1,k}\}_{k=1}^K$ we obtain $\Pr(S_{1,k} = n | \mathbf{y}_1)$ as follows,

$$\Pr(S_{1,k} = n | \mathbf{y}_1) = \frac{\alpha_{1,k}(n) \beta_{1,k}(n)}{\mathcal{L}_1}$$

Assuming that state 2 always reflects the ‘foraging’ state , we use $\Pr(S_{1,k} = 2 | \mathbf{y}_1)$, to predict the value of g_2 by first computing the expected value of E_2 as,

$$\mathbb{E}(E_2) = \left[\sum_{k=1}^K \Pr(S_{1,k} = 2 | \mathbf{y}_1) \right] I - \left[\sum_{k=1}^K d_{1,k} \right] L_1 - \frac{K}{288} C_1$$

Subsequently, following the equations to update body fat and lean muscle mass, we produce a prediction of g_2 for inclusion in the HMM. Given the prediction for g_2 , we can compute \mathcal{L}_2 using the forward algorithm. To compute the likelihood of the joint process, taking into account the observed conditions $\{g_{P_u}\}_{u=1}^U$, we present a sequential approach,

1. Starting at $P_1 = 1$,

- (a) evaluate \mathcal{L}_1 using the forward algorithm.
 - (b) Produce prediction for g_2 through local decoding of the state process, \mathbf{S}_1 .
 - (c) Then, for $j = 2, \dots, P_2 - 1$,
 - Evaluate \mathcal{L}_j using the forward algorithm
 - For $j \neq P_2 - 1$, produce prediction for g_{j+1} through local decoding of the state process, \mathbf{S}_j .
2. For $u \in \{2, \dots, U - 1\}$, repeat the process in Step 1 starting with g_{P_u} and produce predictions of the condition process up to $g_{P_{u+1}-1}$ to obtain $\{\mathcal{L}_{P_u}, \dots, \mathcal{L}_{P_{u+1}-1}\}$.
 3. For $u = U$, we repeat the procedure outlined in Step 1, starting at g_{P_U} to produce predictions up to g_T (unless $P_U = T$) and obtain $\{\mathcal{L}_{P_U}, \dots, \mathcal{L}_T\}$.
 4. Given the individual daily likelihoods, the likelihood of the joint model is a product,

$$\mathcal{L} = \prod_{t=1}^T \mathcal{L}_t$$

4.5.2 Bayesian inference

We fit the model in a Bayesian framework and obtain samples from the joint posterior distribution, $p(\boldsymbol{\theta}|\mathbf{y}) \propto \mathcal{L}(\boldsymbol{\theta}; \mathbf{y})\pi(\boldsymbol{\theta})$, using Markov chain Monte Carlo. As all of the parameters of interest are continuous, we specifically use the dynamic Hamiltonian Monte Carlo algorithm implemented in the software **Stan** to conduct inference (Betancourt, 2017a; Carpenter et al., 2017). To combat the issue of label-switching in HMMs, we impose an ordering on the means of the step length state-dependent distributions, such that $\mu_1 < \mu_2 < \mu_3$, as well as assign non-exchangeable priors (Betancourt, 2017b). For the mean, ψ and concentration parameters, κ , of the turning angle distribution, we first map the parameters from the constrained space to the unconstrained space. As ψ is constrained between $[-\pi, \pi]$, yet values close to the boundaries reflect similar distributions, the MCMC chains can not efficiently explore the parameter space in the natural representation.

We thus use

$$x_n = \kappa_n \cos(\psi_n)$$

$$p_n = \kappa_n \sin(\psi_n)$$

for $n = 1, \dots, N$, as the point defined by (x_n, p_n) is unconstrained in \mathbb{R}^2 . Full specification of the prior distributions are outlined in Section 4.6. We follow the approach detailed in Leos-Barajas and Michelot (2018) to fit the HMM in the R package **rstan** (Stan Development Team, 2018).

Given the fitted model, we obtain draws from the posterior predictive distribution, $p(\mathbf{y}_{rep}|\mathbf{y})$. We compare the step length autocorrelation structure of the posterior predictive draws to the observed autocorrelation structure, and do similarly for the marginal distributions of step lengths and turning angles.

4.6 Simulation

We present a simulation of the joint model for condition and behavior with movement data collected every 5-minutes and a value of condition calculated each day. To reflect the reality that condition may not be able to be measured on a daily basis, we fit four different models, shown in Table 4.1.

Table 4.1 Models fitted to the simulated data.

Model	Condition Assumption
M1	No condition information
M2	Condition is observed every 30 days, condition assumed constant across 30 days until new observation is recorded
M3	Condition is observed every 30 days, daily condition predicted
M4	Condition is observed on a daily basis

Model 4 demands the most manual effort, yet is taken to be ideal scenario in this simulation as presented thus far for the Merino sheep. Model 3 aims to approximate the results of Model 1 by predicting body condition on a daily basis when it is not observed. Model 2 demonstrates how the lack of observing or predicting the condition on a daily basis affects parameters estimates as

compared to the results of Model 4. Model 1 demonstrates the type of inference one would make if not accounting for the effect of condition on movement. For application to the Merino sheep, observed in Patagonia, the most realistic scenario is that condition is observed approximately every 30 days.

For the movement process, the entries of the t.p.m. are a function of the condition process as described in Section 4.4.1. The values chosen for this simulation reflect the general idea that a sheep in poor condition makes an attempt to increase its time foraging while a sheep with a high index of body fat does not need to forage as often. We assume that the resting dynamics are consistent across levels of condition. For our simulation, the value of condition on the first day is taken to be $g_1 = 25$. For subsequent days, we update the value of condition depending on the sequence of states and distances traveled simulated by the HMM as described in Section 4.4.2, until a new observation of condition is available. We take state 1 to reflect resting behavior, State 2 is connected to foraging behavior and State 3 is connected with traveling behavior.

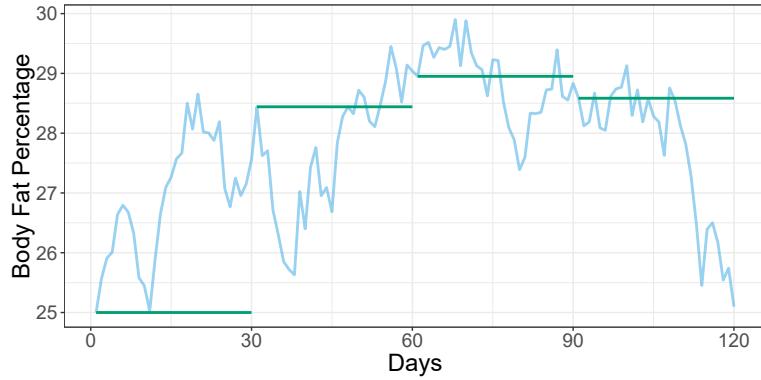


Figure 4.2 Blue line: simulated condition data. Green segments: processed condition data for Model 3.

The prior distributions for each of the four models is given in Table 4.2.

The prior distributions for the state-dependent step length distribution means are centered at the true values of the simulated data process. We further reinforce the ordering of the state-dependent step-length distribution means in the model specification, $\mu_1 < \mu_2 < \mu_3$, as mentioned in Section 4.5.2. Each of the models were fit using the R package `rstan`. We ran three chains each

Table 4.2 Prior distributions for the parameters associated with the HMM across all four models.

Parameter	Prior Distribution
μ	$\mu_1 \sim N^+(.01, .05); \mu_2 \sim N^+ (.1, .1); \mu_3 \sim N^+ (1, .1)$
σ	$\sigma_j \sim N^+(0, 1)$
\mathbf{x}	$x_1 \sim N(-0.5, 1); x_2 \sim N(2, 2)$
\mathbf{p}	$p_j \sim N(0, 0.05)$
τ	$\tau_0^{i,1} \sim N(-3, 1); \tau_1^{i,2} \sim N(0, 1)$

sampling 500 observations, of which the first 250 were used for the warmup phase. In total, we obtain 750 draws from the joint posterior distribution, $p(\boldsymbol{\theta}|\mathbf{y})$. Tables of the HMC results including effective sample size and estimated \hat{R} are provided in the Appendix.

4.6.1 Movement Model Results

The state-dependent parameter estimates are consistent across the four models, as shown in the Appendix. The estimation of the transition probability matrix entries across the four models is displayed in Figure 4.3.

The t.p.m. results for Model 3 and Model 4 are consistent with one another, demonstrating that in this context we are able to replicate the dynamics of the movement process via simulation of the covariates, rather than knowing the true covariate values. This result is due in part to the accuracy in which we can predict the underlying body fat percentage, discussed in Section 4.6.2. Posterior predictive checks for the four models are given in the Appendix.

4.6.2 Predicted Condition Results

Given the known body fat percentage, we compare the true body fat percentage with the predicted body fat percentage provided by the results of Model 3.

In the top panel of Figure 4.4, we see that the predictions are quite close to the true body fat percentage values. Although not easily observed in the figure, there is more variability around the estimate of the body fat percentage over time as the uncertainty continues to compound before

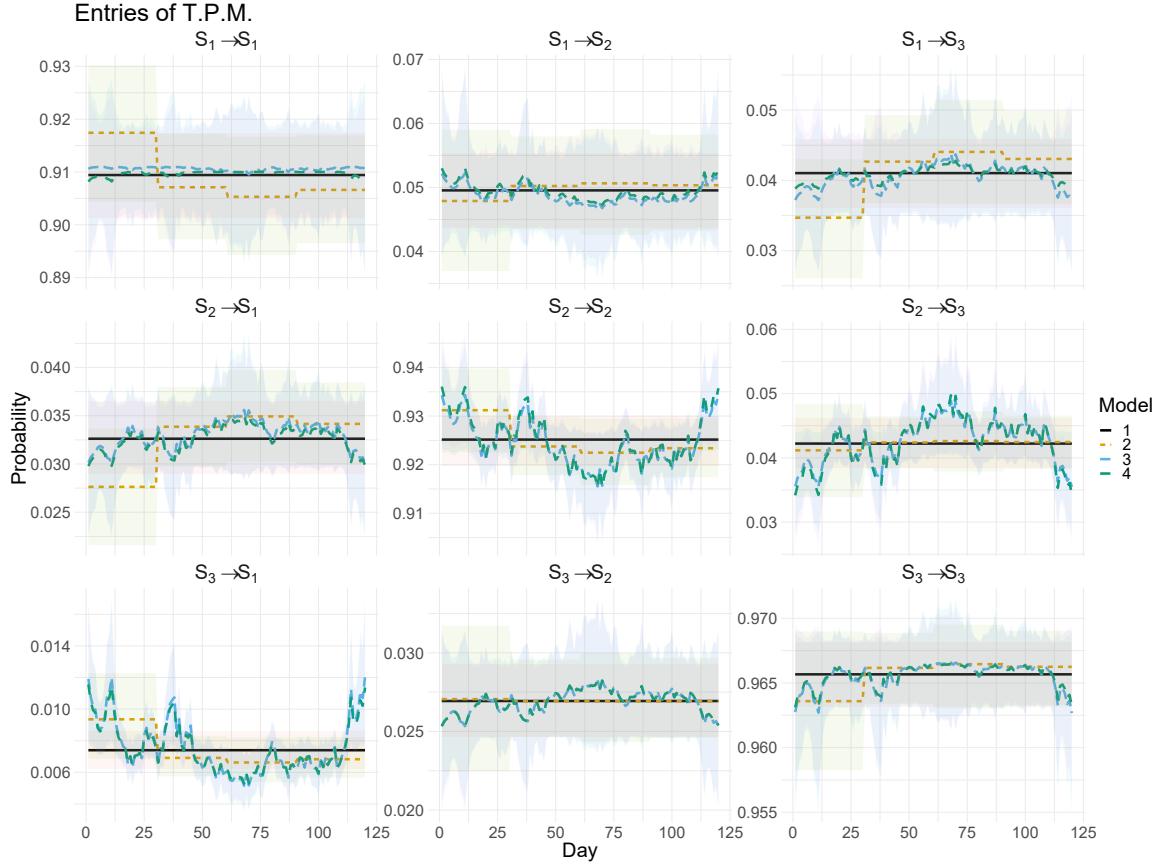


Figure 4.3 Mean t.p.m. under the four possible model constructions, along with 95% credible intervals.

reinitiating the prediction process at an observed body fat percentage. The error and uncertainty around prediction of the body fat percentage in the bottom panel of is displayed in Figure 4.4. While the model formulation has led to a consistent underestimation of the body fat percentage up to approximately 0.10-0.15 by the 29th day of prediction, the values are close enough to replicate the movement dynamics of interest.

4.7 Discussion

The aim of this paper has been to lay the foundation for a modeling framework that allows for interaction of the body condition and its movement dynamics when the condition is sparsely observed and predicted via a set of physiological equations specific to the species of interest. Our

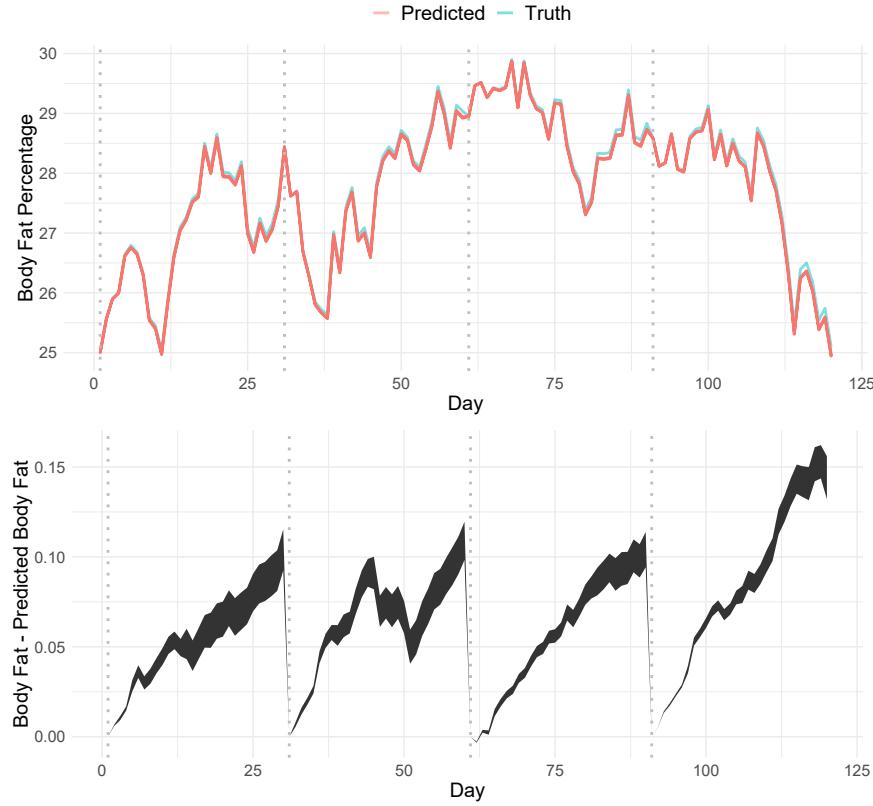


Figure 4.4 Top: Predicted body fat percentages compared to the true body fat percentage, taken to be a value between 0 and 100, along with 95% credible intervals. Bottom: 95% credible intervals of distribution of differences between true body fat percentage and predicted body fat percentage. Dashed lines: Days in which body fat percentage is directly observed.

modeling framework combines two processes of Merino sheep evolving at distinct temporal scales: body fat percentage (linked to body condition) and movement (linked to behavior). We demonstrate that with use of the physiological equations, we are able to predict the body fat percentage at times when it is not observed in order to include as a covariate in the movement model.

An advantage of predicting the body fat percentage, aside from inclusion in the movement model, is that we are able to observe how condition evolved for the animal over the course of the period of interest. Not only can we build models for the movement process, as given by the HMM in this paper, but we now have data for the condition process which opens the possibilities to answering a greater class of research questions.

The framework presented in this paper is simplistic as overall condition is not the only factor that affects the fine-scale movement process. Moving forward, covariates such as time of day, landscape features, and other factors can be directly included into the HMM framework, along with condition. Further, the condition of the animal also depends on the quality of the food it ingests and other factors not currently included. There is also no guarantee that positional data alone can provide enough detail to identify actual foraging behavior. Other movement data, such as accelerometer data, can provide a more accurate picture of the foraging activity. It would also be beneficial to know how long the animal foraged as well as ruminated. However, this basic framework allows for clear extensions to improve the biological interpretation of the movement process and allows future work to be taken in many directions.

4.8 References

- Betancourt, M. (2017a). A conceptual understanding to hamiltonian monte carlo. *arXiv:1701.02434*.
- Betancourt, M. (2017b). Identifying bayesian mixture models. Retrieved from https://betanalpha.github.io/assets/case_studies/identifying_mixture_models.html.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76.
- Delgiudice, G. D., Moen, R. A., Singer, F. J., and Riggs, M. R. (2001). Winter nutritional restriction and simulated body condition of yellowstone elk and bison before and after the fires of 1988. *Wildlife Monographs*, 147:1–60.
- Hooten, M. B., Johnson, D. S., McClintock, B. T., and Morales, J. M. (2017). *Animal Movement: Statistical Models for Telemetry Data*. CRC Press, Boca Raton, FL.
- Leos-Barajas, V. and Michelot, T. (2018). An introduction to animal movement modeling using hidden markov models using stan for bayesian inference. *arXiv:1806.10639*.
- Moen, R., Pastor, J., and Cohen, Y. (1997). A spatially explicit model of moose foraging and energetics. *Ecology*, 78:505–521.
- Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. (2004). Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, 85(9):2436–2445.

- Patterson, T. A., Basson, M., Bravington, M. V., and Gunn, J. S. (2009). Classifying movement behaviour in relation to environmental conditions using hidden markov models. *Journal of Animal Ecology*, 78(6):1113–1123.
- Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., and King, R. (2017). Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis*, 101(4):399–438.
- Pirotta, E., Schwarz, L. K., Costa, D. P., Robinson, P. W., and New, L. (2019). Modeling the functional link between movement, feeding activity, and condition in a marine predator. *Behavioral Ecology*, 30:434–445.
- Robbins, C. T. (1993). *Wildlife feeding and nutrition*. Academic Press, San Diego, second edition.
- Stan Development Team (2018). RStan: the R interface to Stan. R package version 2.18.2.
- Wilmshurst, J. F., Fryxell, J. M., and Berman, C. M. (2000). The allometry of patch selection in ruminants. *Proceedings of the Royal Society of London, Series B*, 267:345–349.
- Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction using R*. Chapman & Hall, Boca Raton, FL, 2nd edition.

4.9 Appendix

4.9.1 Posterior Predictive Checks

Model 1

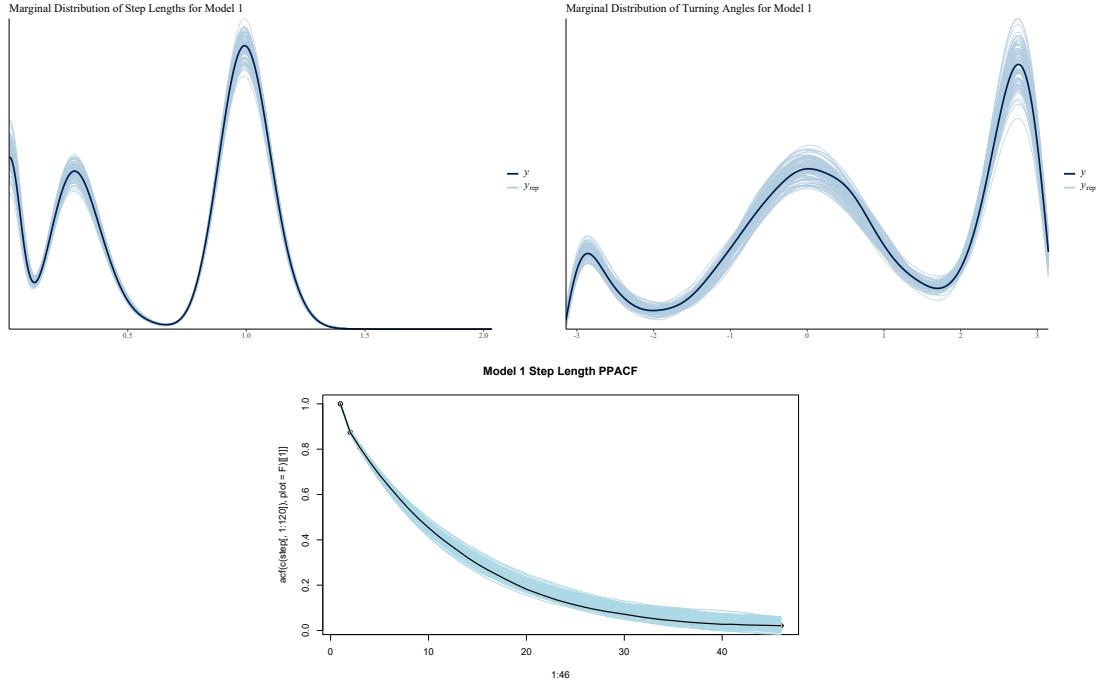


Figure 4.5 Top: Posterior predictive step length distributions for model 1 (in light blue) and the marginal distribution of the observed data (in black). Middle: Posterior predictive turning angle distributions for model 1 (in light blue) and the marginal distribution of the observed data (in black). Bottom: Posterior predictive autocorrelation function for model 1 (in light blue) and the observed autocorrelation function of the observed data (in black).

Model 2

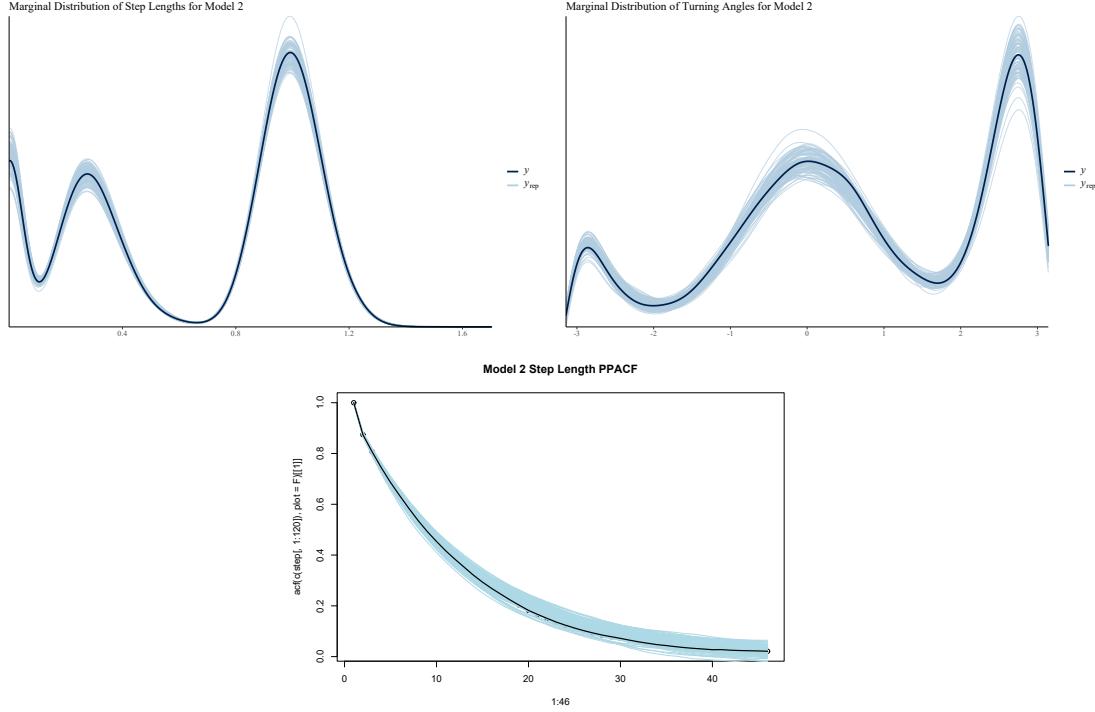


Figure 4.6 Top: Posterior predictive step length distributions for model 2 (in light blue) and the marginal distribution of the observed data (in black). Middle: Posterior predictive turning angle distributions for model 2 (in light blue) and the marginal distribution of the observed data (in black). Bottom: Posterior predictive autocorrelation function for model 2 (in light blue) and the observed autocorrelation function of the observed data (in black).

Model 3

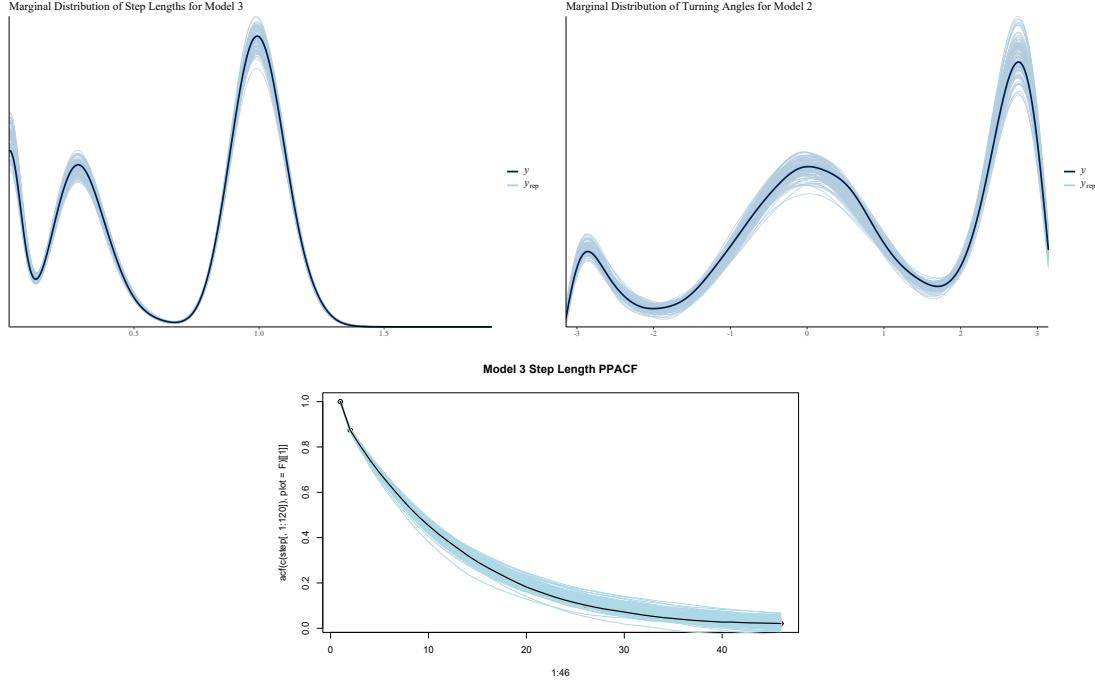


Figure 4.7 Top: Posterior predictive step length distributions for model 3 (in light blue) and the marginal distribution of the observed data (in black). Middle: Posterior predictive turning angle distributions for model 3 (in light blue) and the marginal distribution of the observed data (in black). Bottom: Posterior predictive autocorrelation function for model 3 (in light blue) and the observed autocorrelation function of the observed data (in black).

Model 4

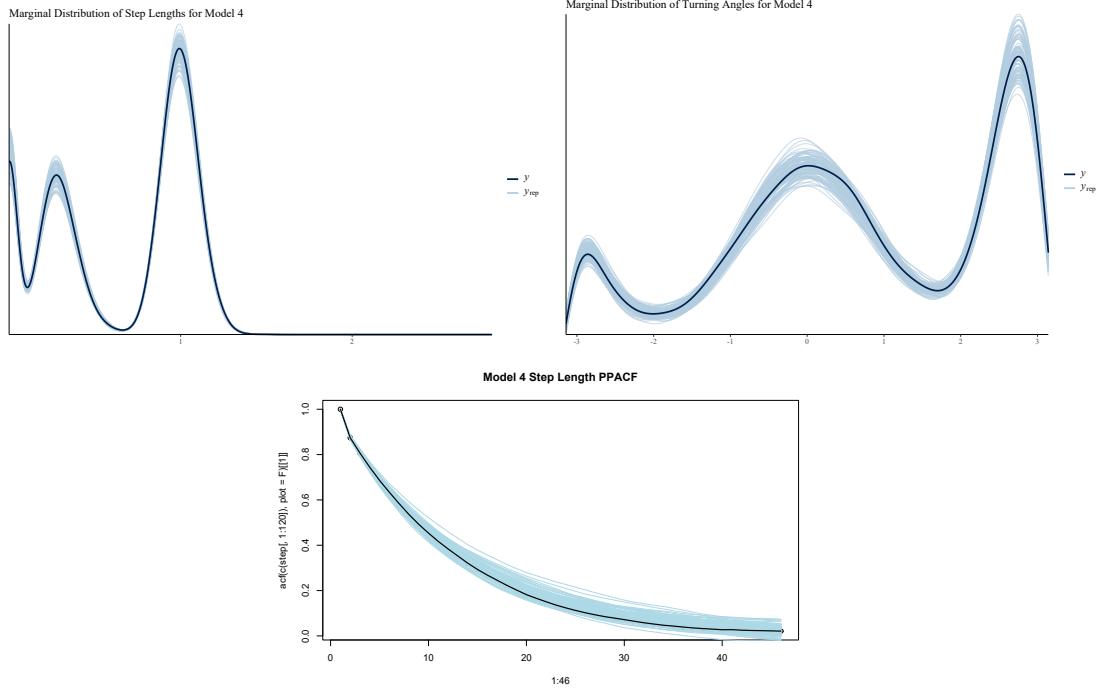


Figure 4.8 Top: Posterior predictive step length distributions for model 4 (in light blue) and the marginal distribution of the observed data (in black). Middle: Posterior predictive turning angle distributions for model 4 (in light blue) and the marginal distribution of the observed data (in black). Bottom: Posterior predictive autocorrelation function for model 4 (in light blue) and the observed autocorrelation function of the observed data (in black).

4.9.2 HMC Posterior Draws

Table 4.3 HMC Posterior Draws for Model 1. Values for the marginal means and chosen quantiles are reported, along with the effective number of samples, n_{eff} , and value of \hat{R} .

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
mu[1]	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	377.27	1.00
mu[2]	0.30	0.00	0.00	0.30	0.30	0.30	0.30	0.30	1168.94	1.00
mu[3]	1.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	828.97	1.00
sigma[1]	0.05	0.00	0.00	0.04	0.05	0.05	0.05	0.06	363.42	1.00
sigma[2]	0.10	0.00	0.00	0.10	0.10	0.10	0.10	0.10	1748.94	1.00
sigma[3]	0.10	0.00	0.00	0.10	0.10	0.10	0.10	0.10	1655.18	1.00
xangle[1]	-8.84	0.01	0.16	-9.16	-8.95	-8.84	-8.74	-8.54	382.49	1.00
xangle[2]	-0.99	0.00	0.02	-1.03	-1.00	-0.99	-0.98	-0.96	1157.76	1.00
xangle[3]	1.00	0.00	0.01	0.98	0.99	1.00	1.01	1.02	1073.49	1.00
yangle[1]	3.11	0.00	0.07	2.97	3.06	3.12	3.17	3.25	409.37	1.00
yangle[2]	0.13	0.00	0.01	0.10	0.12	0.13	0.14	0.16	1142.42	1.00
yangle[3]	-0.01	0.00	0.01	-0.03	-0.01	-0.00	0.00	0.02	916.26	1.00
theta[1,1]	0.91	0.00	0.00	0.90	0.91	0.91	0.91	0.92	569.96	1.00
theta[1,2]	0.05	0.00	0.00	0.04	0.05	0.05	0.05	0.06	704.54	1.00
theta[1,3]	0.04	0.00	0.00	0.04	0.04	0.04	0.04	0.05	888.35	1.00
theta[2,1]	0.03	0.00	0.00	0.03	0.03	0.03	0.03	0.04	590.26	1.01
theta[2,2]	0.93	0.00	0.00	0.92	0.92	0.93	0.93	0.93	674.52	1.00
theta[2,3]	0.04	0.00	0.00	0.04	0.04	0.04	0.04	0.05	950.42	1.00
theta[3,1]	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	640.78	1.00
theta[3,2]	0.03	0.00	0.00	0.02	0.03	0.03	0.03	0.03	987.44	1.00
theta[3,3]	0.97	0.00	0.00	0.96	0.96	0.97	0.97	0.97	994.39	1.00
shape[1]	0.04	0.00	0.00	0.04	0.04	0.04	0.04	0.04	739.18	1.00
shape[2]	8.98	0.00	0.13	8.75	8.89	8.99	9.07	9.22	1276.52	1.00
shape[3]	100.19	0.02	1.03	98.24	99.47	100.21	100.93	102.12	1728.26	1.00
rate[1]	4.09	0.02	0.35	3.40	3.86	4.08	4.32	4.78	366.51	1.00
rate[2]	29.79	0.01	0.45	28.94	29.49	29.78	30.10	30.65	1555.21	1.00
rate[3]	100.25	0.02	1.03	98.31	99.52	100.26	100.99	102.14	1698.80	1.00
loc[1]	2.80	0.00	0.00	2.79	2.80	2.80	2.81	2.81	1248.16	1.00
loc[2]	3.01	0.00	0.01	2.98	3.00	3.01	3.02	3.04	1187.95	1.00
loc[3]	-0.01	0.00	0.01	-0.03	-0.01	-0.00	0.00	0.02	912.60	1.00
kappa[1]	9.37	0.01	0.17	9.05	9.26	9.37	9.49	9.71	368.75	1.00
kappa[2]	1.00	0.00	0.02	0.97	0.99	1.00	1.01	1.04	1085.28	1.00
kappa[3]	1.00	0.00	0.01	0.98	0.99	1.00	1.01	1.02	1078.21	1.00
lp_	86124.37	0.17	3.02	86117.65	86122.57	86124.76	86126.55	86129.18	299.61	1.00

Table 4.4 HMC Posterior Draws for Model 2. Values for the marginal means and chosen quantiles are reported, along with the effective number of samples, n_{eff} , and value of \hat{R} .

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
mu[1]	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	550.76	1.00
mu[2]	0.30	0.00	0.00	0.30	0.30	0.30	0.30	0.30	901.83	1.00
mu[3]	1.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	833.56	1.00
sigma[1]	0.05	0.00	0.00	0.04	0.05	0.05	0.05	0.06	539.89	1.00
sigma[2]	0.10	0.00	0.00	0.10	0.10	0.10	0.10	0.10	1160.49	1.00
sigma[3]	0.10	0.00	0.00	0.10	0.10	0.10	0.10	0.10	1283.41	1.00
xangle[1]	-8.85	0.01	0.16	-9.16	-8.96	-8.85	-8.74	-8.56	789.04	1.00
xangle[2]	-0.99	0.00	0.02	-1.02	-1.00	-0.99	-0.98	-0.96	1030.21	1.00
xangle[3]	1.00	0.00	0.01	0.97	0.99	1.00	1.01	1.02	1316.39	1.00
yangle[1]	3.12	0.00	0.07	2.99	3.07	3.12	3.16	3.25	733.29	1.00
yangle[2]	0.13	0.00	0.01	0.10	0.12	0.13	0.14	0.16	1097.43	1.00
yangle[3]	-0.00	0.00	0.01	-0.02	-0.01	-0.00	0.00	0.02	1284.37	1.00
tau[1,1]	-2.96	0.01	0.12	-3.22	-3.04	-2.96	-2.88	-2.73	591.36	1.00
tau[1,2]	0.09	0.01	0.20	-0.28	-0.05	0.09	0.22	0.50	667.55	1.00
tau[2,1]	-3.28	0.01	0.14	-3.56	-3.38	-3.27	-3.18	-3.05	450.66	1.00
tau[2,2]	0.33	0.01	0.21	-0.08	0.18	0.32	0.46	0.73	445.70	1.00
tau[3,1]	-3.52	0.01	0.12	-3.77	-3.60	-3.52	-3.45	-3.31	465.22	1.00
tau[3,2]	0.31	0.01	0.18	0.01	0.20	0.30	0.43	0.68	482.61	1.00
tau[4,1]	-3.12	0.00	0.09	-3.32	-3.18	-3.12	-3.06	-2.95	558.70	1.00
tau[4,2]	0.06	0.01	0.15	-0.22	-0.04	0.06	0.16	0.34	474.60	1.00
tau[5,1]	-4.65	0.01	0.15	-4.95	-4.74	-4.65	-4.55	-4.36	699.39	1.00
tau[5,2]	-0.43	0.01	0.24	-0.88	-0.61	-0.45	-0.28	0.06	673.35	1.00
tau[6,1]	-3.58	0.00	0.09	-3.76	-3.64	-3.57	-3.51	-3.41	372.68	1.00
tau[6,2]	-0.01	0.01	0.15	-0.27	-0.11	-0.02	0.10	0.29	540.26	1.00
shape[1]	0.04	0.00	0.00	0.04	0.04	0.04	0.04	0.04	817.10	1.00
shape[2]	8.98	0.00	0.12	8.75	8.90	8.98	9.07	9.22	1043.06	1.00
shape[3]	100.23	0.03	1.00	98.26	99.55	100.26	100.90	102.13	1267.65	1.00
rate[1]	4.09	0.02	0.34	3.44	3.85	4.09	4.31	4.79	521.17	1.00
rate[2]	29.77	0.01	0.40	29.01	29.49	29.76	30.05	30.57	1116.43	1.00
rate[3]	100.29	0.03	1.00	98.32	99.59	100.32	100.97	102.18	1273.88	1.00
loc[1]	2.80	0.00	0.00	2.79	2.80	2.80	2.81	2.81	702.35	1.00
loc[2]	3.01	0.00	0.01	2.98	3.00	3.01	3.02	3.04	1087.51	1.00
loc[3]	-0.00	0.00	0.01	-0.03	-0.01	-0.00	0.00	0.02	1288.14	1.00
kappa[1]	9.38	0.01	0.16	9.07	9.27	9.38	9.50	9.70	804.94	1.00
kappa[2]	1.00	0.00	0.02	0.97	0.99	1.00	1.01	1.03	1048.93	1.00
kappa[3]	1.00	0.00	0.01	0.97	0.99	1.00	1.01	1.02	1316.52	1.00
lp_	86145.09	0.19	3.46	86137.81	86142.82	86145.46	86147.69	86150.71	320.02	1.00

Table 4.5 HMC Posterior Draws for Model 3. Values for the marginal means and chosen quantiles are reported, along with the effective number of samples, n_{eff} , and value of \hat{R} .

	mean	se.mean	sd	2.5%	25%	50%	75%	97.5%	n.eff	Rhat
mu[1]	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	443.25	1.01
mu[2]	0.30	0.00	0.00	0.30	0.30	0.30	0.30	0.30	1017.73	1.00
mu[3]	1.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	830.48	1.00
sigma[1]	0.05	0.00	0.00	0.04	0.05	0.05	0.05	0.06	447.24	1.01
sigma[2]	0.10	0.00	0.00	0.10	0.10	0.10	0.10	0.10	1354.08	1.00
sigma[3]	0.10	0.00	0.00	0.10	0.10	0.10	0.10	0.10	1862.89	1.00
xangle[1]	-8.80	0.01	0.17	-9.16	-8.91	-8.80	-8.70	-8.44	625.67	1.00
xangle[2]	-1.00	0.00	0.02	-1.03	-1.01	-1.00	-0.99	-0.97	938.66	1.00
xangle[3]	1.00	0.00	0.01	0.98	0.99	1.00	1.01	1.02	1015.53	1.00
yangle[1]	3.10	0.00	0.08	2.95	3.05	3.10	3.15	3.25	582.55	1.00
yangle[2]	0.13	0.00	0.01	0.10	0.12	0.13	0.14	0.16	1050.05	1.00
yangle[3]	-0.01	0.00	0.01	-0.03	-0.01	-0.01	0.00	0.02	970.34	1.00
tau[1,1]	-2.89	0.00	0.08	-3.04	-2.94	-2.88	-2.84	-2.74	688.31	1.00
tau[1,2]	-0.03	0.00	0.09	-0.20	-0.09	-0.03	0.03	0.14	754.24	1.00
tau[2,1]	-3.09	0.00	0.09	-3.27	-3.15	-3.09	-3.03	-2.92	677.90	1.00
tau[2,2]	0.01	0.00	0.09	-0.18	-0.06	0.01	0.07	0.19	711.33	1.00
tau[3,1]	-3.32	0.00	0.07	-3.46	-3.37	-3.33	-3.28	-3.19	1036.34	1.00
tau[3,2]	0.01	0.00	0.09	-0.18	-0.05	0.01	0.07	0.17	917.51	1.00
tau[4,1]	-3.10	0.00	0.07	-3.23	-3.15	-3.11	-3.06	-2.97	691.93	1.01
tau[4,2]	0.06	0.00	0.09	-0.12	-0.01	0.06	0.12	0.23	737.67	1.00
tau[5,1]	-4.73	0.00	0.10	-4.92	-4.79	-4.73	-4.67	-4.54	594.73	1.00
tau[5,2]	-0.12	0.00	0.09	-0.30	-0.19	-0.12	-0.06	0.05	838.57	1.00
tau[6,1]	-3.57	0.00	0.07	-3.69	-3.61	-3.57	-3.52	-3.43	567.24	1.00
tau[6,2]	-0.00	0.00	0.09	-0.18	-0.06	-0.00	0.05	0.15	508.54	1.00
shape[1]	0.04	0.00	0.00	0.04	0.04	0.04	0.04	0.04	822.55	1.00
shape[2]	8.99	0.00	0.13	8.76	8.91	9.00	9.08	9.25	1356.91	1.00
shape[3]	100.19	0.03	1.07	98.15	99.53	100.19	100.85	102.43	1755.62	1.00
rate[1]	4.08	0.02	0.32	3.52	3.86	4.06	4.30	4.72	455.91	1.01
rate[2]	29.78	0.01	0.44	28.94	29.48	29.78	30.09	30.65	1488.97	1.00
rate[3]	100.25	0.03	1.07	98.21	99.59	100.24	100.89	102.56	1802.72	1.00
loc[1]	2.80	0.00	0.00	2.79	2.80	2.80	2.81	2.81	1066.11	1.00
loc[2]	3.01	0.00	0.01	2.98	3.00	3.01	3.02	3.04	1069.50	1.00
loc[3]	-0.01	0.00	0.01	-0.03	-0.01	-0.01	0.00	0.02	968.14	1.00
kappa[1]	9.33	0.01	0.18	8.95	9.22	9.33	9.45	9.70	603.23	1.00
kappa[2]	1.01	0.00	0.02	0.97	1.00	1.01	1.02	1.04	924.49	1.00
kappa[3]	1.00	0.00	0.01	0.98	0.99	1.00	1.01	1.02	1019.12	1.00
lp__	84108.77	0.19	3.38	84101.40	84106.80	84109.18	84110.97	84114.60	331.27	1.00

Table 4.6 HMC Posterior Draws for Model 4. Values for the marginal means and chosen quantiles are reported, along with the effective number of samples, n_{eff} , and value of \hat{R} .

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
mu[1]	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	522.82	1.00
mu[2]	0.30	0.00	0.00	0.30	0.30	0.30	0.30	0.30	834.55	1.01
mu[3]	1.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00	646.76	1.00
sigma[1]	0.05	0.00	0.00	0.04	0.05	0.05	0.05	0.06	518.37	1.00
sigma[2]	0.10	0.00	0.00	0.10	0.10	0.10	0.10	0.10	1390.35	1.00
sigma[3]	0.10	0.00	0.00	0.10	0.10	0.10	0.10	0.10	1254.14	1.00
xangle[1]	-8.85	0.01	0.16	-9.15	-8.96	-8.85	-8.73	-8.54	664.88	1.00
xangle[2]	-0.99	0.00	0.02	-1.03	-1.00	-0.99	-0.98	-0.96	1262.31	1.00
xangle[3]	1.00	0.00	0.01	0.98	0.99	1.00	1.01	1.02	1107.80	1.00
yangle[1]	3.12	0.00	0.07	2.99	3.07	3.12	3.17	3.25	750.11	1.00
yangle[2]	0.13	0.00	0.01	0.11	0.12	0.13	0.14	0.16	1473.34	1.00
yangle[3]	-0.01	0.00	0.01	-0.03	-0.01	-0.01	0.00	0.02	2057.65	1.00
tau[1,1]	-2.85	0.01	0.15	-3.15	-2.95	-2.85	-2.75	-2.57	758.83	1.00
tau[1,2]	-0.12	0.01	0.25	-0.57	-0.30	-0.12	0.06	0.39	823.32	1.00
tau[2,1]	-3.17	0.01	0.17	-3.53	-3.27	-3.16	-3.06	-2.84	617.01	1.00
tau[2,2]	0.10	0.01	0.28	-0.44	-0.08	0.09	0.30	0.69	541.33	1.00
tau[3,1]	-3.46	0.01	0.14	-3.72	-3.54	-3.46	-3.37	-3.20	491.13	1.00
tau[3,2]	0.19	0.01	0.22	-0.25	0.04	0.20	0.33	0.65	553.17	1.00
tau[4,1]	-3.32	0.01	0.12	-3.58	-3.40	-3.31	-3.24	-3.09	492.44	1.00
tau[4,2]	0.41	0.01	0.20	0.02	0.27	0.41	0.54	0.83	501.56	1.00
tau[5,1]	-4.44	0.01	0.19	-4.81	-4.57	-4.44	-4.31	-4.05	694.92	1.00
tau[5,2]	-0.80	0.01	0.33	-1.43	-1.01	-0.80	-0.59	-0.16	650.73	1.00
tau[6,1]	-3.64	0.01	0.12	-3.89	-3.72	-3.64	-3.57	-3.42	544.71	1.01
tau[6,2]	0.11	0.01	0.19	-0.26	-0.02	0.11	0.23	0.48	519.17	1.01
shape[1]	0.04	0.00	0.00	0.04	0.04	0.04	0.04	0.04	693.61	1.00
shape[2]	8.98	0.00	0.13	8.74	8.89	8.98	9.08	9.22	1066.62	1.00
shape[3]	100.15	0.03	1.01	98.20	99.48	100.17	100.83	102.10	1182.95	1.00
rate[1]	4.07	0.01	0.33	3.41	3.84	4.07	4.29	4.74	541.28	1.00
rate[2]	29.78	0.01	0.44	28.94	29.45	29.78	30.12	30.62	1239.01	1.00
rate[3]	100.21	0.03	1.01	98.31	99.51	100.22	100.87	102.13	1224.39	1.00
loc[1]	2.80	0.00	0.01	2.79	2.80	2.80	2.81	2.81	945.92	1.00
loc[2]	3.01	0.00	0.01	2.98	3.00	3.01	3.02	3.04	1497.23	1.00
loc[3]	-0.01	0.00	0.01	-0.03	-0.01	-0.01	0.00	0.02	2081.04	1.00
kappa[1]	9.38	0.01	0.17	9.06	9.26	9.39	9.50	9.70	653.62	1.00
kappa[2]	1.00	0.00	0.02	0.97	0.99	1.00	1.01	1.03	1240.25	1.00
kappa[3]	1.00	0.00	0.01	0.98	0.99	1.00	1.01	1.02	1108.26	1.00
lp_	86146.97	0.18	3.41	86139.88	86144.91	86147.23	86149.40	86153.04	342.31	1.01

CHAPTER 5. FUTURE WORK SUMMARY AND DISCUSSION

There are two general motivations underlying this work, to extend the HMM framework to capture a larger variety of biological processes, done in chapters 2 and 3, and to incorporate physiological processes that are sparsely observed into the analysis of animal movement, done in chapter 4. As both approaches presented in this dissertation are not common techniques applied in the analysis of animal movement, there is much work left to be done to develop and refine their implementation.

5.1 Multi-scale animal behavior

Extensive efforts have been made to develop methodological approaches that assist in connecting the observed movement process to underlying behaviors (Zucchini et al., 2016; Patterson et al., 2017; Hooten et al., 2017). Part of the challenge in the collection of data is to determine how the temporal scale at which the data is collected connects to the animal's behavior. Intuitively, different behaviors will manifest themselves at different temporal scales.

HMMs are an appealing tool in the analysis of animal movement data as they provide a simplistic representation of how the movement process is generated according to an underlying state process (serving as a proxy for behaviors of interest). Part of the appeal is also due to the algorithms available to evaluate the likelihood efficiently (the *forward* and *forward-backward* algorithm) and the relative ease with which they can be fit (Zucchini et al., 2016). As the multi-scale HMMs presented in chapters 2 and 3 can be written in the general form of a basic HMM, we are able to retain the advantages that a basic HMM provides (via model fitting and evaluation) but its structure allows for identification of a larger variety of behaviors than before.

Issues related to applying HMMs to animal movement data are exacerbated in the multi-scale HMM framework. Connecting the temporal resolution at which the data are collected to behaviors

of interest still presents a challenge, although the multi-scale HMM now allows for behaviors to be identified via compositions of fine-scale movement patterns. However, as HMMs are stochastic processes that can capture a large variety of patterns, application of multi-scale HMMs without domain expertise and additional structure included in the model formulation may result in highly multi-modal likelihoods and, subsequently, posterior distributions. Selecting the number of internal states and the number of production states within an internal state should be done in a manner that matches domain expertise, to some extent (Pohle et al., 2017).

Overall, this is a first step to matching the biological intuition that different behaviors manifest themselves at not only different temporal scales, but can be represented as various compositions of fine-scale discernible movements.

5.2 Physiology and Movement

Chapter 4 develops methodology that incorporates condition of an animal into the analysis of animal movement using physiological equations. The motivation behind this work is to be able to predict condition for inclusion in an HMM applied to animal movement data. As this is among the first approaches that tackles this problem, the goal has not been to provide a complete framework that applies to all systems, but rather one that is flexible, straightforward and provides clear opportunities to expand and customize.

The HMM portion of this chapter only includes body fat percentage as a potential driver of the fine scale movement process. In a full analysis, many other drivers would be included, like time of day, proximity to different patches of food or season. Drivers can be included in the transition probability matrix but also in the state-dependent process. Importantly, any customization of the HMM to account for random effects or environmental drivers of behavior does not change the manner in which predictions for condition are produced.

Future work in this area will incorporate more structure into the HMM applied to animal movement data and also explore manners in which to incorporate other physiological processes like hunger and fatigue as drivers of behavior, similar to the work presented in Hooten et al. (2019).

5.3 References

- Hooten, M. B., Johnson, D. S., McClintock, B. T., and Morales, J. M. (2017). *Animal Movement: Statistical Models for Telemetry Data*. CRC Press, Boca Raton, FL.
- Hooten, M. B., Scharf, H. R., and Morales, J. M. (2019). Running on empty: recharge dynamics from animal movement data. *Ecology Letters*, 22(2):377–389.
- Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., and King, R. (2017). Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis*, 101(4):399–438.
- Pohle, J., Langrock, R., van Beest, F. M., and Schmidt, N. M. (2017). Selecting the number of states in hidden markov models – pitfalls, practical challenges and pragmatic solutions illustrated using animal movement. *JABES*, 22(3):270–293.
- Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction using R*. Chapman & Hall, Boca Raton, FL, 2nd edition.