Incorporating Temporal Heterogeneity in Hidden Markov Models For Animal Movement

INCORPORATING TEMPORAL HETEROGENEITY IN HIDDEN MARKOV MODELS FOR ANIMAL MOVEMENT

BY

MICHAEL LI , Hon.B.Sc.

A THESIS
SUBMITTED TO THE DEPARTMENT OF MATHEMATICS & STATISTICS
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

Master of Applied Science (2015)
(Mathematics & Statistics)

McMaster University
Hamilton, Ontario, Canada

TITLE:	Incorporating Temporal Heterogeneity in Hidden Markov Models For Animal Movement
AUTHOR:	Michael Li Hon.B.Sc.
SUPERVISOR:	Dr. Benjamin M. Bolker
NUMBER OF PAGES:	ix, 59

Abstract

Clustering time-series data into discrete groups can improve prediction as well as providing insight into the nature of underlying, unobservable states of the system. However, temporal heterogeneity and autocorrelation (persistence) in group occupancy can obscure such signals. We use latent-state and hidden Markov models (HMMs), two standard clustering techniques, to model high-resolution hourly movement data from Florida panthers. Allowing for temporal heterogeneity in transition probabilities, a straightforward but rarely explored model extension, resolves previous HMM modeling issues and clarifies the behavioural patterns of panthers.

Acknowledgements

First, I would like my deepest gratitude to my supervisor Dr. Benjamin Bolker; without him I could never be who I am today. Second, I would like to thank Dr. Jonathan Dushoff and all the members of Bolker-Dushoff lab for giving me a wonderful environment to work and collaborate with people who share great passion for mathematical/statistical biology. Next, I would like to thank McMaster University, my Statistics professors and my classmates for the wonderful experience of my Master's program.I would also like to thank Madelon van de Kerk and all the people involve in the panther research for providing me with the data for my thesis. Lastly, to a special friend Ms. Yanling Jin; thank you for being there and motivation support throughout my whole master's life, it was truly the happiest time of my life.

Contents

List of Figures

Chapter 1

Introduction

1.1 Background

The Florida panther (*Puma concolor coryi*) is a subspecies of puma, an endangered species that was once found throughout the southeastern United States. The remaining population of Florida panthers inhabits a small part of south Florida, roughly 5% of its original inhabited range (van de Kerk *et al.*, 2015a). The population of Florida panthers had experienced severe pressure from human development expansions resulting in habitat loss. Furthermore, in the 1800s the Florida panthers became primary hunting targets for settlers, which led to the presumed extinction of the population by the early 1950s. The Florida panther was rediscovered in 1973 by a survey team organized by the National Geographic Society and immediately added to the US endangered species list. From that point on, research programs have been initiated to help recover Florida panthers.

The first panther research program was initiated in 1981 by state and federal wildlife agencies to study and learn more about the panthers and restore their population. The panther population continued to decrease after humans stopped hunting them due to low reproduction/ inbreeding rates , diseases and road kills (Coffin, 2007). Biologists have tracked the panthers using high frequency geographic positioning system (GPS) ratio collars to provide morphological and biomedical treatments and examinations. The population was recovered after a genetic restoration in 1995, where eight female pumas (*P.c. stanleyana*) were transfered from Texas to reverse the inbreeding depression and increase population numbers (Johnson *et al.*, 2010).

1.2 Animal Movement

Animal movements is one of many important research topics in panther research. In animal movement, ecologists are interested in factors that determine behavioural sequences of movement patterns (i.e. when animals move in a certain pattern, and under what circumstances they switch to alternative movement pattern). Understanding animal movement can help with conservation development such as road and development planning, as well as better estimates of demography so humans can avoid interfering with their lives and vice versa. In recent years, animal ecologists have proposed the multiphasic movement framework, which considers combinations of internal states and external (environmental) factors in an attempt to capture the underlying movement patterns (Firle *et al.*, 1998; Nathan *et al.*, 2008) Hidden Markov models (HMMs) are able both to group observed movements into behavioural states and to characterize movement behaviour within the states. Hidden Markov models have been explored in the animal movement literature (Langrock *et al.*, 2012; Franke *et al.*, 2006; Patterson, 2009; McKellar *et al.*, 2014) and previously in Florida panthers research (van de Kerk *et al.*, 2015b).

1.3 Movement Data

In animal ecology, movement data are frequently used to study animal movement patterns. Movement data are often in the form of longitudinal designs (Lanza *et al.*, 2004), where they conceptually targets three elements (Collins, 2006):

1. Theoretical model of change: Can we use linear, quadratic, periodic, or other types of models to capture the characteristic shape of change in the data?
2. Temporal design: Can the temporal design (timing, frequency and spacing of observations) capture the changes of interest in the data?
3. Statistical model of change: Can we validate the goodness of fit of the models used to fit the data?

Raw, unprocessed movement data are usually coordinates roughly indicating the animal's position at a given time. The most common way to locate animals is radio tracking, by attaching a transmitter to the animal; e.g. with ankle or neck collars. Movement data can be further processed into other useful forms such as the distance and angle between consecutive positions.

1.4 Florida Panther Data

GPS collars were attached to 18 Florida Panthers in 2005-2012 by the Florida Fish and Wildlife and Conservation Commission (FWC). The GPS collars were programmed to collect location coordinate data at fixed intervals. The time interval between consecutive coordinates may vary depending on battery life and equipment technology. Data resolution had improved over the years as technology advances, time intervals between consecutive coordinates used to be twice a week and now, hourly data. For our study, we used the subset of panthers with a significant amount of hourly-interval data. The hourly GPS coordinates were later decomposed into step lengths (in meters) and turning angles (in radians). In addition, animal-level covariates such as sex, age, reproductive status and covariates that vary within animals (i.e. times, current habitat type, etc) are also present in the dataset. The data is available at the University of Florida Institutional Repository (http://ufdc.ufl.edu//IR00004241/00001). Due location sensitivity of GPS locations of Florida Panthers, only processed step length and turning angle data are available to the public.

1.5 Objective

In a previous Florida panther study (van de Kerk *et al.*, 2015b), hidden semi-Markov models (HSMM), an extension of HMM that permits explicit modeling of dwell times (ie. time spent in a specific state using alternative distributions such as Poisson and negative binomial instead of the default geometric distribution) (Langrock *et al.*, 2012) were used to reveal multiphasic movement of Florida panthers. The Florida panthers were modeled independently with 2- to 6-state time-homogeneous HSMM models and Bayesian information criteria (BIC) were used to evaluate goodness of fit between models with different numbers of states. BIC values of these models decrease as the number of states increases from three to six states, suggesting the six states model was favoured statistically. On the other hand, it was hard to interpret the six states biologically. Despite the flexibility to model dwell time, it did not affect the results (negative binomial parameter close to 1 for dwell time implying is it geometrically distributed), thus, in this thesis, we will use HMM instead of HSMM.

The hidden Markov model structure is very sensitive to the number of hidden states. The number of movement states directly affects the complexity of the movement model. In behavioural ecology, most HMMs are often limited to two (Schliehe-Diecks *et al.*, 2012; McKellar *et al.*, 2014; Langrock *et al.*, 2012; Fryxell *et al.*, 2008) and sometimes (in a few examples) three states (Dean *et al.*, 2012; Morales *et al.*, 2004; Franke *et al.*, 2006; van de Kerk *et al.*, 2015b). Models with more than three states are difficult to interpret biologically, especially since the behavioural states cannot be directly observed. Thus, a number of obvious, yet unexplored questions arise that are applicable in many hidden Markov applications are:

1. Are models with large numbers (>3) of states biologically interpretable and statistically defensible?
2. Do omitted covariates in HMMs lead to overprediction of the number of states?
3. Can we find a flexible framework for exploring these issues?

In Figures 1.1, 1.2, 1.3, multiple peaks in the step length distribution exist in population, individual, and individual hourly level. The previous study of Florida panthers using hidden semi Markov model (HSMM), showed BIC decreases as the number of hidden states increase to 5 to 6, especially for cats with the most data (van de Kerk *et al.*, 2015b), but we cannot really identify 5 or 6 distinct peaks in Figure 1.1, 1.2, 1.3.

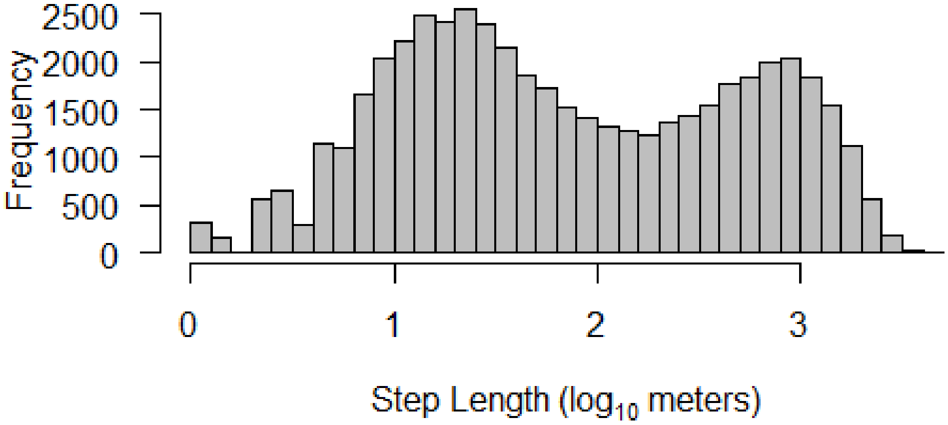

Figure 1.1: Histogram of step length in $\log_{10}$ meters at population level

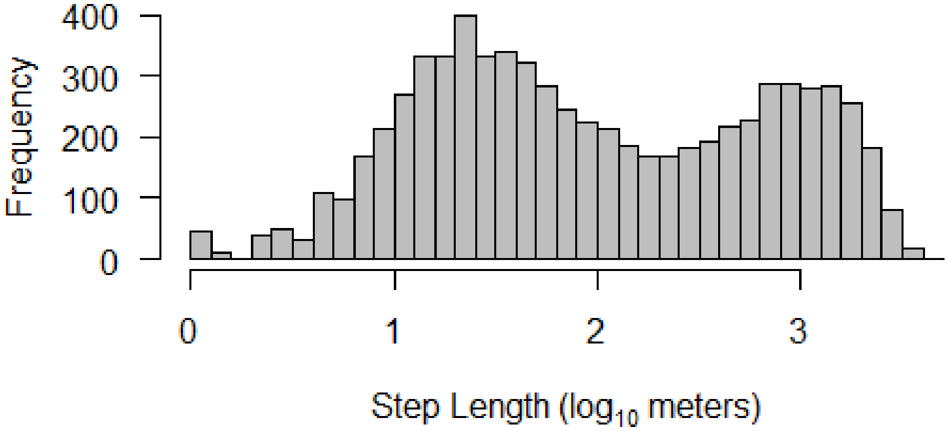

Figure 1.2: Histogram of step length in $\log_{10}$ meters at individual level (ID=130)

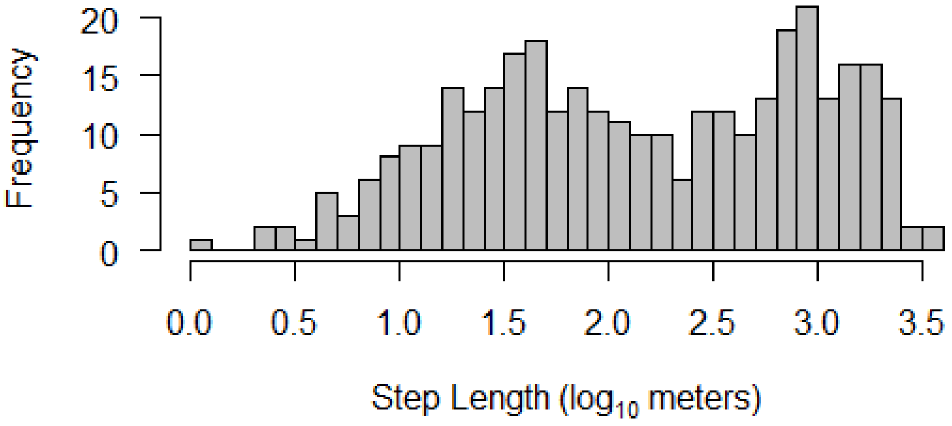

Figure 1.3: Histogram of step length in $\log_{10}$ meters at individual hourly level (ID=130 at 1:00)

In this thesis, we would like to answer the above questions using the Florida panther data by investigating cats with the most data (CatID 48,94,130,131 where each has more than 9000 observations):

1. Can modeling temporal heterogeneity reduce the estimated number of states to a statistically defensible and biologically interpretable level?
2. Can increasing structural complexity of HMM framework capture heterogeneity in behavioural transition patterns?
3. Can the model capture movement patterns over long periods of time and produce biologically reasonable movements?

1.6 Scope of the Thesis

Chapter 2 of this thesis introduces methods modeling longitudinal data, (i.e. finite mixture models, Markov models and finally HMM). Chapter 3 considers software to model HMMs, dimension reduction in animal movement data and then modeling and fitting the Florida panther data. Chapter 4 explores HMM extensions and modeling HMMs with temporal heterogeneity in transition probabilities. Chapter 5 explores different methods to evaluate goodness of fit and validate the model from model-based simulations. Finally, Chapter 6 conclusions this thesis and discusses future explorations.

Chapter 2

Mixture Models for Longitudinal Data

2.1 Finite Mixture Models

Mixture models are designed to accommodate a particular form of heterogeneity in the population; that is, the population may consist of clusters, each having a distinct distribution for the observed variable. Finite mixture models distributions consist of a finite number of component distributions, one for each cluster, and a mixing distribution that selects from these components.

Let $\mathbf{Y} = (Y_1, ..., Y_n)$ be an N-dimensional random variable and $\mathbf{y} = (y_1, ..., y_n)$ be an observation of Y. In the general case of M components, the probability density function for the random variable $\mathbf{Y}$ is given by :

$$p(y|\Theta) = \sum_{m=1}^{M} \lambda_m p_m(y|\theta_m), \tag{2.1}$$

where $p_m(y|\theta_m)$ is the probability density function of the m-th component, λ_m are the mixing proportions and $\mathbf{\Theta} = (\lambda_1, ..., \lambda_M, \theta_1, ..., \theta_M)$ is the set of parameters and satisfies

$$\lambda_m > 0, \;\; m = 1, ..., M \tag{2.2}$$

and

$$\sum_{m=1}^{M} \lambda_m = 1. \tag{2.3}$$

2.2 Markov Models

Although finite mixture models are a useful tool, they do not handle one of the key features of longitudinal data, sequential dependence. Markov models are useful for this scenario, allowing sequential dependence relationships in the data. Let $\mathbf{O} = (O_1, ...O_n)$ be a sequence of discrete random variables and $\mathbf{S} = \{s_1, ..., s_k\}$ be a set of k distinct states, satisfying the Markov property:

$$P(O_{t+1} = s_i|O_t, ..., O_1) = P(O_{t+1} = s_i|O_t), \;\; t = 1, ..., n \tag{2.4}$$

then it is called a Markov chain, where the past and the future are dependent only through the present.

The Markov property can be interpreted as specifying that conditioning on the whole process up to time t is equivalent to conditioning only on the most recent value at time t. This is called a first order Markov chain, where the state at the next time step only depends on the current state. The more general case of a k^{th} order Markov chain can be expressed as:

$$P(O_{t+1} = s_i|O_t, ..., O_1) = P(O_{t+1} = s_i|O_t, O_{t-1}, ..., O_{t-k+1}),\ \ k = 1, ..., n,\ \ t = 1, ..., n \tag{2.5}$$

where the state at the next time step depends on the most recent k states.

In general, a Markov model can be fully specified by two parts: an initial prior distribution and the transition probabilities between states. The initial prior distribution is the probability of the first observation starting from a certain state and is given by:

$$\psi_i = P(O_1 = s_i), \tag{2.6}$$

where

$$\psi_i > 0, \sum_{i=1}^{k} \psi_i = 1,\ \ i = 1, .., k. \tag{2.7}$$

The transition probabilities between states form a transition matrix:

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{1,1} & \cdots & \pi_{1,k} \\ \vdots & \ddots & \vdots \\ \pi_{k,1} & \cdots & \pi_{k,k}, \end{bmatrix} \tag{2.8}$$

where,

$$\pi_{ij} = P(O_{t+1} = s_j|O_t = s_i),\ \ \pi_{ij} \geq 0, \tag{2.9}$$

$$\sum_{j=1}^{k} \pi_{i,j} = 1, \ \ i, j = 1, ..., k \tag{2.10}$$

The initial prior and transition probability can be covariate dependent. Let $\mathbf{X} = \{x_1, ..., x_n\}$ be a set of covariates. Starting with the initial prior probability, we want to make $\mathbf{\Psi}$ a function of $\mathbf{X}$. In order to satisfy (2.7), we define the following multinomial logistic function:

$$\psi_i(x_t) = \frac{\exp(\eta_i(x_t))}{1 + \sum_{j=2}^{k} \exp(\eta_i(x_t))}, \ \ i = 2, ..., k, t = 1, ..., n. \tag{2.11}$$

$$\psi_1(x_t) = 1 - \sum_{i=2}^{k} \psi_i(x_t), \tag{2.12}$$

where $\eta_i(x_t)$ is a function of the covariates.

We can use the same method to parameterize the transition matrix (2.8):

$$\mathbf{\Pi}(x_t) = \begin{bmatrix} \pi_{1,1}(x_t) & \cdots & \pi_{1,k}(x_t) \\ \vdots & \ddots & \vdots \\ \pi_{k,1}(x_t) & \cdots & \pi_{k,k}(x_t), \end{bmatrix} \ t = 1, ..., n. \tag{2.13}$$

and using the same multinomial logistic approach:

$$\pi_{i,j}(x_t) = \frac{\exp(\eta_{i,j}(x_t))}{1 + \sum_{j=2}^{k} \exp(\eta_{i,j}(x_t))}, \ \ j = 2, ..., n, t = 1, ..., n, \tag{2.14}$$

$$\pi_{i,1}(x_t) = 1 - \sum_{j=2}^{n} \pi_{i,j}(x_t), \ \ i = 1, ..., n. \tag{2.15}$$

where (2.15) satisfies (2.10).

Given the initial prior distribution $\boldsymbol{\Psi}$ and transition matrix $\boldsymbol{\Pi}$, the probability of a state sequence $\mathbf{O} = \{O_1, ..., O_n\}$ can be calculated as the product of the two parts:

$$\begin{aligned} P(\mathbf{O}|\boldsymbol{\Psi}, \boldsymbol{\Pi}) &= P(O_1|\Psi)P(O_2|O_1)P(O_3|O_2) \times \cdots \times P(O_n|O_{n-1}) \\ &= P(O_1|\boldsymbol{\Psi}) \prod_{t=1}^{n-1} P(O_{t+1}|O_t) \end{aligned} \tag{2.16}$$

2.3 Hidden Markov Models

The Markov framework is useful in many applications, but Markov models are limited to fully observable system. The hidden Markov model, a special dependent mixture model, is an extension of the Markov model that incorporates mixture models (Rabiner, 1989; Zucchini and MacDonald, 2009). The hidden Markov framework is made more flexible by allowing any form of observable sequential data generated from an underlying unobserved Markov process in which the number of states and transition probabilities can be both known and unknown.

Let $\mathbf{Y} = \{y_1, ..., y_n\}$ be a sequence of either discrete or continuous observations and $\mathbf{Z} = \{z_1, ..., z_n\}$ be a sequence of discrete hidden states. An HMM can be characterized by a composition of an indirect Markov model and finite mixture model with the following specifications:

1. The number of hidden states in the model. In many applications, the hidden states often have some significant characteristics so that the observed data can be reasonably grouped into distinguishable clusters. For example, two-state models are often used in applications with dichotomous features. The hidden Markov framework is very sensitive to the number of states, and due to the

uncertainty, it is ideal to have minimal and sensible number of states by context. As in the previous section, we denote the individual states as $\mathbf{S} = \{s_1, ..., s_k\}$, where the number of distinct states is k.

2. The transition probability distribution $\mathbf{\Pi} = \{\pi_{i,j}\}$ where

$$\pi_{i,j} = P(z_{t+1} = s_j | z_t = s_i), \quad i, j = 1, ..., k \tag{2.17}$$

   Unlike (2.9) in the previous section, this transition process is unobservable.

3. The emission probability distribution:

$$P(y_t|z_t) = P(y_t|z_t = s_i),\ i = 1, ..., k, t = 1, ..., n. \tag{2.18}$$

   The observations of an HMM are generated by conditioning on the hidden state of the unobservable Markov process. This is similar to finite mixture model in section 2.1, but here it is conditioned upon hidden states rather than observable components.

4. The initial prior distribution $\mathbf{\Psi}$. This is the probability distribution of the initial state of the hidden Markov process

$$P(z_1) = P(z_1 = s_i) = \psi_i,\ i = 1, ..., k, \tag{2.19}$$

   with the usual constraints of non-negativity and summation to 1.

2.4 The Two Fundamental Problems for HMMs

Given an observable sequence $\mathbf{Y} = \{y_1, ..., y_n\}$ and a HMM model with defined initial prior distribution, state transition probabilities, and emission distribution defined, there are two fundamentals question of interest in many real-life HMM applications:

1. How do we efficiently compute the probability or likelihood of an observed sequence?
2. How do we estimate the HMM parameters to maximize the likelihood?

Answering the two fundamental problems can create a model to reproduce the characteristics of the observed data. The first problem can be used to evaluate the likelihood of the model given the observation sequence which can be used for model selection, whereas the second problem tunes the model to maximize the likelihood.

2.4.1 Problem 1: Evaluating HMM likelihoods

In the previous section, we defined all the requirements (2.17), (2.18) and (2.19) necessary to compute the probability of an observed sequence given an HMM. Given an observed sequence $\mathbf{Y} = \{y_1, ..., y_n\}$ and an HMM with the set of k distinct states $\mathbf{S} = \{s_1, ..., s_k\}$,the direct but inefficient method to compute the probability or likelihood $P(\mathbf{Y}|\mathbf{\Pi}, \mathbf{\Psi})$ is to first define a hidden sequence

$$\mathbf{Z} = \{z_1, ..., z_n\}, \;\; z_t \in S, t = 1, ..., n. \tag{2.20}$$

Using $\mathbf{Z}$, we can calculate the probability of the observed sequence $\mathbf{Y}$ by the product of (2.17), (2.18) and (2.19)

$$\begin{aligned} P(\mathbf{Y}|\mathbf{Z}, \mathbf{\Pi}, \mathbf{\Psi}) &= P(z_1|\mathbf{\Psi})P(y_1|z_1)P(z_2|z_1, \mathbf{\Pi})P(y_2|z_2) \times \cdots \times P(z_t|z_{t-1}, \mathbf{\Pi})P(y_t|z_t) \\ &= P(z_1|\mathbf{\Psi})P(y_1|z_1) \prod_{t=2}^{n} P(z_t|z_{t-1}, \mathbf{\Pi})P(y_t|z_t). \end{aligned} \tag{2.21}$$

There exist k^n possible sequences of $\mathbf{Z}$, and (2.14) is the probability of only one of these k^n. Thus, summing over all possible hidden sequence $\mathbf{Z}$'s will give the probability of the observed sequence

$$P(\mathbf{Y}|\mathbf{\Pi}, \mathbf{\Psi}) = \sum_{\forall \mathbf{Z}} P(\mathbf{Y}|\mathbf{Z}, \mathbf{\Pi}, \mathbf{\Psi}) \tag{2.22}$$

This direct method to calculate P($\mathbf{Y}$) is computationally inefficient. A more efficient calculation method is the *Forward Algorithm*, which calculates the probability of the observational subsequence $\mathbf{Y}_{1:t} = \{y_1, ..., y_t\}$ at a particular state at time t

$$\alpha_t(i) = P(\mathbf{Y}_{1:t}|z_t = s_i, \mathbf{\Pi}, \mathbf{\Psi}), \tag{2.23}$$

where $\alpha_t(i)$ is called the *forward variable*. $\alpha_t(i)$ can be solved by induction:

1. Base case:

$$\alpha_1(i) = P(z_1 = s_i|\mathbf{\Psi})P(y_1|z_1), \;\; i = 1, ..., k \tag{2.24}$$

2. Induction step:

$$\alpha_{t+1}(j) = \sum_{i=1}^{k} \alpha_t(i) P(z_{t+1} = s_j | z_t = s_i, \mathbf{\Pi}) P(y_{t+1} | z_{t+1}), \;\; j = 1, ..., k, t = 1, ..., n-1. \tag{2.25}$$

Summing up all the forward variables at time n will give the probability of the observed sequence

$$P(\mathbf{Y}|\mathbf{\Pi}, \mathbf{\Psi}) = P(\mathbf{Y}_{1:n}|\mathbf{\Pi}, \mathbf{\Psi}) = \sum_{i=1}^{k} \alpha_n(i). \tag{2.26}$$

2.4.2 Problem 2: Estimating HMM parameters:

As in the previous problem, there is no direct analytic way to estimate the parameters that maximizes the likelihood of the observed sequence. The *Baum-Welch algorithm*, a special expectation-maximization method developed by Baum's team, locally maximizes $P(\mathbf{Y}|\mathbf{\Pi}, \mathbf{\Psi})$ (Baum *et al.*, 1970; Baum, 1972).

Given the observed sequence $\mathbf{Y}$ and parameters $\mathbf{\Pi}, \mathbf{\Psi}$, we need to calculate $\gamma_t(i, j)$, the probability of being in state s_i at time t and s_j at time $t+1$ as the following:

$$\gamma_t(i, j) = P(z_t = s_i, z_{t+1} = s_j | \mathbf{Y}, \mathbf{\Pi}, \mathbf{\Psi}), \tag{2.27}$$

which we can further decompose by independence:

$$\gamma_t(i, j) = \frac{P(z_t = s_i | \mathbf{Y}_{1:t}, \mathbf{\Pi}, \mathbf{\Psi}) \pi_{i,j} P(z_{t+1} = s_j | \mathbf{Y}_{t+1:n}, \mathbf{\Pi}, \mathbf{\Psi})}{P(\mathbf{Y}|\mathbf{\Pi}, \mathbf{\Psi})}. \tag{2.28}$$

Notice that in the numerator of (2.28), the first probability can be rewritten as

$\alpha_t(i)$, using the forward variable defined in the previous section. The last probability in the numerator can be written in a similar manner with the backward variable $\beta_{t+1}(j)$ of the *backward algorithm*. Similar to the forward algorithm, the backward algorithm calculates the probability of the observational subsequence $\mathbf{Y_{t+1:n}} = \{y_{t+1}, ..., y_n\}$ at a particular state at time t

$$\beta_t(i) = P(\mathbf{Y}_{t+1:n}|z_t = s_i, \mathbf{\Pi}, \mathbf{\Psi}). \tag{2.29}$$

$\beta_t(i)$ can be solved by induction:

1. Base case:

$$\beta_n(i) = 1,\ \ i = 1, ..., k \tag{2.30}$$

2. Induction step:

$$\beta_t(i) = \sum_{j=1}^{k} P(z_{t+1} = s_j|z_t = s_i, \mathbf{\Pi})P(y_{t+1}|z_{t+1})\beta_{t+1}(j), i = 1, ..., k, t = 1, ..., n-1. \tag{2.31}$$

In the previous section, summing up all the forwards variables at time n will give the probability of the observed sequence $P(\mathbf{Y}|\mathbf{\Pi}, \mathbf{\Psi})$ (2.19). We can also calculate $P(\mathbf{Y}|\mathbf{\Pi}, \mathbf{\Psi})$ by the backward algorithm since we are using the full observed sequence by summing up all the backward variables at time 1:

$$P(\mathbf{Y}|\mathbf{\Pi}, \mathbf{\Psi}) = P(\mathbf{Y}_{1:n}|\mathbf{\Pi}, \mathbf{\Psi}) = \sum_{i}^{k} \beta_1(i). \tag{2.32}$$

Now that we have defined the backward variable, we can rewrite (2.21) as the following:

$$\gamma_t(i,j) = \frac{\alpha_t(i)\pi_{i,j}\beta_{t+1}(j)}{\sum_{i=1}^{k}\sum_{j=1}^{k}\alpha_t(i)\pi_{i,j}P(y_{t+1}|z_{t+1}=s_j)\beta_{t+1}(j)}. \tag{2.33}$$

1. E-step

   With the above equations, we can start the parameter estimation.

   $$\psi_i^* = \text{expected number of times in state } s_i \text{ at time } 1 = \sum_{j=1}^{k}\gamma_1(i,j). \tag{2.34}$$

   $$\begin{aligned}\pi_{i,j}^* &= \frac{\text{expected number of transitions from } s_i \text{ to } s_j}{\text{expected number of transitions from } s_i} \\ &= \frac{\sum_{t=1}^{n-1}\gamma_t(i,j)}{\sum_{t=1}^{n-1}\sum_{j=1}^{k}\gamma_t(i,j)}\end{aligned} \tag{2.35}$$

   $$\begin{aligned}P^*(O|z_t=s_j) &= \frac{\text{expected number of times in } s_j \text{ and observing } O}{\text{expected number of times in } s_j} \\ &= \frac{\sum_{t=1,\text{s.t } y_t=O}^{n}\sum_{j=1}^{k}\gamma_t(i,j)}{\sum_{t=1}^{n}\sum_{j=1}^{k}\gamma_t(i,j)}\end{aligned} \tag{2.36}$$

   Let $\mathbf{\Theta}$ be the set of all parameters in the HMM, $\mathbf{\Theta} = \{\mathbf{\Psi}, \mathbf{\Pi},$ and parameters of the emission probability distributions$\}$, then $\mathbf{\Theta}^*$ is the estimated parameters, and (2.34-2.36) are the estimation formulas. The E-step is calculating the expected value of the log likelihood of the HMM under the estimated parameter

$\mathbf{\Theta}^*$ by the auxiliary function

$$Q(\mathbf{\Theta}, \mathbf{\Theta}^*) = \sum_{Q} P(Q|\mathbf{Y}, \mathbf{\Theta})\log[P(\mathbf{Y}, Q|\mathbf{\Theta}^*] \tag{2.37}$$

2. M-step:

   The M-step of estimating the HMM parameters is to maximize the auxiliary function (2.37). Baum and his colleagues (Baum *et al.*, 1968, 1970) show that maximizing (2.37) will necessarily increase the likelihood:

$$\text{argmax}_{\mathbf{\Theta}^*}[Q(\mathbf{\Theta}, \mathbf{\Theta}^*)] \implies P(\mathbf{Y}|\mathbf{\Theta}^*) \geq P(\mathbf{Y}|\mathbf{\Theta}). \tag{2.38}$$

2.5 HMM: Discussion

In many HMM literatures and applications, there is one more fundamental question: how do we select the unobserved state sequence that best explains the observed sequence? We will not look into this problem in great detail because we are most interested in finding the model structure that can best simulate realizations resembling the original data, rather than knowing the best unobserved state sequence and understand patterns via simulations rather than finding the best unobserved state sequence given any model. Unlike problem 2, it is possible to find the best unobserved state sequence given the observed sequence of any HMM regardless of how well the HMM maximizes the likelihood. Even though the simulations of the HMM framework depend heavily on the unobserved Markov process, we want to focus our attention on the observable outcomes of simulations based on model.

The direct method of finding the best unobserved state is to find the state that gives the highest probability at the current time step:

$$\mathbf{Z}^* = \text{argmax}_{1\leq i\leq k} \sum_{j=1}^{k} \gamma_t(i,j), \;\; 1 \leq t \leq n, 1 \leq i \leq k \tag{2.39}$$

This approach may generate an impossible sequence because it does not account for the transitions in the Markov process; a more efficient approach that takes account of the Markov process is called *Viterbi Algorithm.* The Viterbi algorithm solves for the most probable state sequence by finding:

$$\delta_t(i) = \max_{\mathbf{Z}_{1:t-1}} P(\mathbf{Z}_{1:t-1}, z_t = s_i, \mathbf{Y}_{1:t}|\mathbf{\Theta}), \tag{2.40}$$

where $\delta_t(i)$ is the highest probability along a single path sequence up to time t. We can continue this process by induction :

1. Base case:

$$\delta_1(i) = \psi_i P(y_1 = z_1), \;\; i = 1, ..., k, \tag{2.41}$$

2. Induction step:

$$\delta_{t+1}(j) = [\text{argmax}_i \delta_t(i)\pi_{i,j}] P(y_{t+1}|z_{t+1}), \;\; j = 1, ..., k, t = 1, ..., n-1, \tag{2.42}$$

$$\mathbf{Z}^* = \text{argmax}_{1\leq i\leq k}(\delta_t(i)\pi_{i,j}), \;\; \text{for } i, j = 1, ..., k, t = 1, ..., n. \tag{2.43}$$

where $\mathbf{Z}^*$ records the most probable hidden sequence.

Chapter 3

Modeling Florida Panther Movements

3.1 Computational Software

HMM has been a hot topic for the past couple of decides for two reasons: better data and better computation power. Common statistical and computation software (SAS, R, SPSS, Matlab and etc) have the tools to handle very basic HMM frameworks but not complex HMMs. The statistical software that we are using in this thesis is R. There exist many packages in R that can model very basic HMM but with many limitations. For example, the *HMM* package (Himmelmann and www.linhi.com, 2010) can only model multinomial responses, and *mhsmm* package (O'Connell and Hjsgaard, 2011) can model multivariate GLM responses but not special distributions frequently used in animal movement models. Many researchers implement their own HMM codes oriented toward specific disciplines, such as, *move.hmm* (Augustine and Langrock, 2014), an R package specially coded for animal movement HMMs with special distributions such as the Weibull distribution for modeling step length distributions and the von Mises and wrapped Cauchy distributions for modeling turning angle distributions.

All of the package discussed above have one important limitation; they cannot model covariate-dependent transitions. Specifying the transition matrices governing the hidden process in HMM may require many parameters; hence many applications lack the data to parameterize complex covariate-dependent transmission models. The idea of covariate dependent transition is not new, but there are few software frameworks that allow modeling such complexity. The package we are using in this thesis is *depmixS4* (Visser and Speekenbrink, 2010), a dependent mixture model framework that can model covariate dependent transitions and allows the usuer to extend the set of response distributions by manually coding.

3.2 Dimension Reduction

There are numerous discussions in the animal movement literature on whether to model GPS coordinates directly (sequences of paired longitude and latitude positions) or to reduce the data to step lengths and turning angles. From a mathematical point of view, they are equivalent because converting to step lengths and turning angles is just a simple transformation where the dimensionality remains the same. One advantage of using step lengths and turning angles versus longitude and latitude positions is that we may able to ignore turning angles, thus reducing the dimensions of the problem. Step lengths and turning angles are negatively correlated; that is, longer step lengths tend to have less variation in turning angles (more directed), whereas short and medium step lengths have uniform variation in turning angle (i.e., more random movement). With this dimension reduction, not only we can reduce the number of parameters greatly, we also do not need to worry about the complexities of modeling a multivariate response.

3.3 Model Selection

We considered models ranging in complexity from 3 biologically interpretable states:

1. Resting mode, characterized by short step lengths
2. Moderately active mode, characterized by intermediate step lengths
3. Traveling mode, characterized by long step lengths

to 6 (the maximum number of states by BIC selection from the previous HSMM study (van de Kerk *et al.*, 2015b)) underlying movement states for FMM and HMM.

BIC was used for model selection because we are interested in exploring the data generation process by trying to determine the true number of underlying states. Furthermore, BIC will usually choose a less complex model when the number of observation is large. It is defined as

$$\mathrm{BIC} = -2\ln L + k\ln(n) \tag{3.1}$$

where $\ln L$ is the log-likelihood (Eq. 2.26), k is the number of free parameters and n is the number of observations. The BIC penalty term ($k\ln(n)$) is stricter than the $2k$ penalty term of the Akaike information criterion (AIC)

$$\mathrm{AIC} = -2\ln L + 2k, \tag{3.2}$$

as long as $n > \exp(2) = 7.38$.

In this thesis, we consider only log-normal models for the distribution of step lengths.

3.4 Modeling Florida Panther Data

3.4.1 FMM

We will start by modeling the data with FMM. This is a naive way to simply test how many clusters are in the data set, ignoring the serial dependency of the sequential data. The FMM is a special case of HMM; we can model it in the HMM framework by setting the transition matrix (2.8) with columns equal to the mixing proportions (2.2):

$$\mathbf{\Pi} = \begin{bmatrix} \lambda_1 & \cdots & \lambda_k \\ \vdots & \vdots & \vdots \\ \lambda_1 & \cdots & \lambda_k, \end{bmatrix} \tag{3.3}$$

This will satisfy all conditions of the hidden Markov transitions (2.8) because the mixing proportions sum up to 1 (2.3). The response distributions for step lengths used are log-normal distributions conditional upon the state the panther is in:

$$Y|Z = s_i \sim N(\mu_i, \sigma_i^2) \ \ i \in \{1, ..., 6\}, \tag{3.4}$$

where Y is the $\log_{10}$ transformed step lengths.

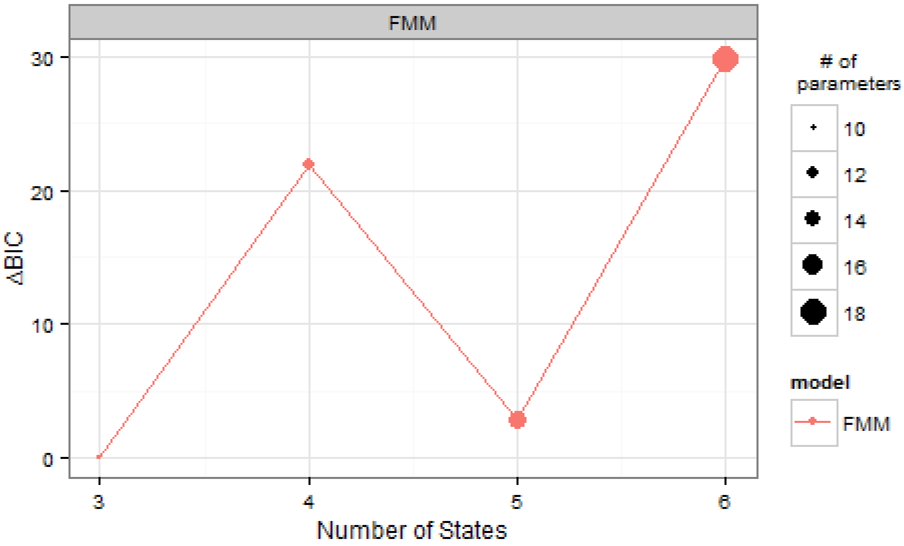

Figure 3.1: FMM BIC Comparison (ID = 130)

We standardize the BIC values of each type of model relative to the lowest BIC value within the same group.

$$\Delta\text{BIC} = \text{BIC} - \min(\text{BIC}) \tag{3.5}$$

Figure 3.1 suggests three underlying states in the for this individual Florida panther (ID=130). This result is encouraging because three-state models are biologically interpretable, but FMM does not account for serial dependency.

State	Mean $\log_{10}$ meters	Standard Deviation $\log_{10}$ meters	Mixing Proportion λ
1	1.47	0.51	0.65
2	2.73	0.27	0.20
3	3.16	0.17	0.15

Table 3.1: FMM parameter estimates for all three states (ID = 130)

3.4.2 HMM

In order to account for the serial dependency in the sequential data, we used the HMM to model the Florida panther data. Figure 3.2, we can see the relationship described in the previous study that as the number of state increase, the BIC decreases (van de Kerk *et al.*, 2015b). Furthermore, we are able to get qualitatively similar results when we reduce dimension by ignoring turning angles. Of course, the question still remains, what does the six different states mean? From a biological point of view, it is difficult to find interpretations for six distinct hidden states.

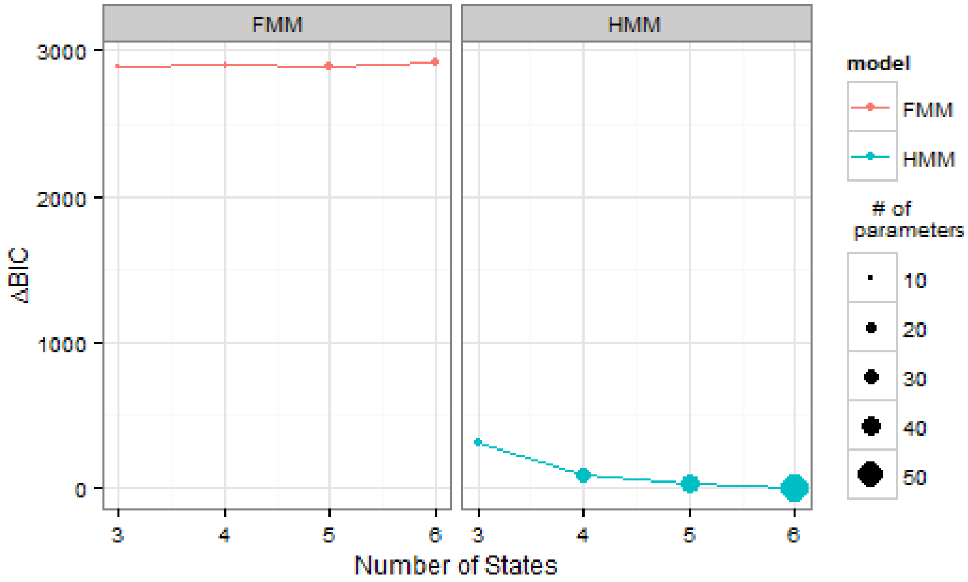

Figure 3.2: BIC Comparison Among States and Type (ID = 130). Note that BICs are now standardized relative to minimum HMM rather than minimum FMM values.

HMMs in general have much lower BICs compared to FMMs. As mentioned in the previous section, the FMM is a naive approach just to see how many clusters/ hidden states seem to appear the raw data, if we ignore serial dependency and treat it as cross-sectional data. HMM on the other hand accounts for the serial dependency of the longitudinal data, thus fit the data better via BIC. This Florida panther (ID=130) has 10286 observations and the six-state HMM has 54 parameters. At this point, a seven state HMM has a much higher BIC than six states. We have the flexibility to explore more complex HMM models and see if increasing the model complexity can reduce the number of estimated states via BIC.

State	Mean $\log_{10}$ meters	Standard Deviation $\log_{10}$ meters
1	0.81	0.39
2	1.22	0.24
3	1.72	0.25
4	2.41	0.28
5	2.94	0.18
6	3.27	0.12

Table 3.2: HMM parameter estimate for all six states (ID = 130)

$$\hat{\mathbf{\Pi}} = \begin{bmatrix} 0.36 & 0.30 & 0.14 & 0.10 & 0.08 & 0.02 \\ 0.12 & 0.57 & 0.17 & 0.08 & 0.05 & 0.01 \\ 0.06 & 0.16 & 0.59 & 0.13 & 0.05 & 0.01 \\ 0.06 & 0.14 & 0.16 & 0.35 & 0.24 & 0.05 \\ 0.01 & 0.02 & 0.05 & 0.13 & 0.35 & 0.44 \end{bmatrix}$$

Chapter 4

Variations of Hidden Markov Models

4.1 Temporal Heterogeneity

Like many species, Florida panthers have a diurnal cycle: specifically, they become less active during the day (as reflected in the data by shorter step lengths) and more active at night. The diurnal cycle exhibited by the Florida panthers is noticeable in Figure 4.1 and 4.2; on average, they tend to have shorter step lengths between the 8th hour and the 15th hour, longer step lengths between the 16th hour to 20th hour, and a mixture of step lengths at other times of the day.

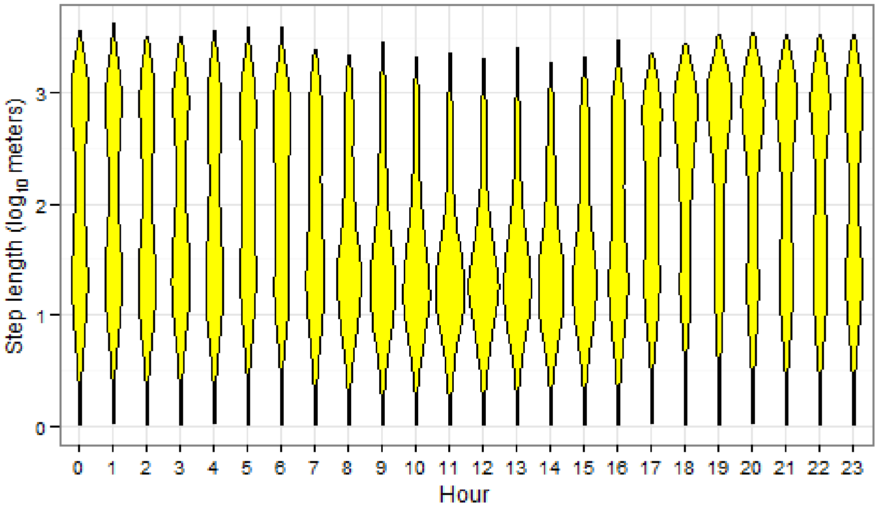

Figure 4.1: Violin Plot of Distance vs Time of Day For all Panthers

Of course, Figure 4.1 displays the marginal distribution of data at the population level, which includes panthers of different ages and sexes. But, even at the individual level Figure 4.2, we can still see a diurnal cycle, which is not accounted for by the homogeneous-time FMMs and HMMs discussed in the previous section or the homogeneous-time HSMMs used by van de Kerk et al (2014).

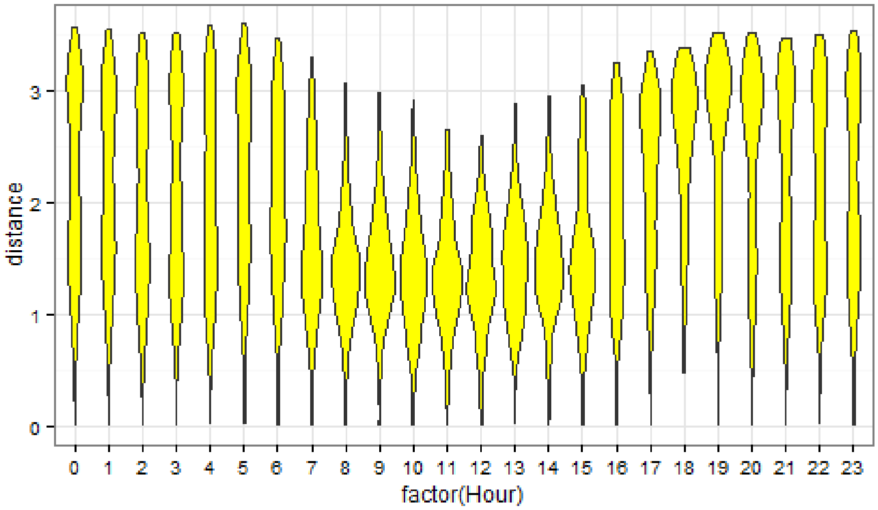

Figure 4.2: Violin Plot of Distance vs Time of Day (ID=130)

Referring back to the figures in section 1, which show marginal histograms across time, we see a sequence of bimodality observed in the data from the marginal distribution of step lengths across time at the population level (Figure 1.1) to Figure 4.1 broken down to time of day, or to Figures 1.2, 1.3, and 4.2 at the hourly and individual-specific level. Bimodality/ mixture of behaviours persists as we go to finer scales, indicating that mixture models are indeed required even once temporal variation is taken into account.

In the HMM framework, we can add temporal heterogeneity (i.e. allowing the model to depend on temporally varying covariates) in two places — the transitions or the emission probability. There are reasonable arguments for adding temporal heterogeneity in both places, but in this thesis, we will focus our attention on the unobservable behavioural change from the transition probabilities. It is more reasonable to think panthers move less during the day because they are more likely to rest, rather than panthers move less because it is day time.

4.2 HMM with Temporal Heterogeneity

In the Florida panther data, let $\mathbf{X}$ be a time of day covariate, i.e. a repeating sequence of 1 to 24. Using 2.13-2.15 described in section 2 on modeling covariate dependence using multinomial logistic, we modeled $\eta_{i,j}(x_t)$ in the following ways:

1. Hourly: A factor of all the hours,

$$\eta_{i,j}(x_t) = \sum_{m=1}^{24} c_{i,j,m} \mathrm{I}_m(x_t), \tag{4.1}$$

$$\mathrm{I}_m(x_t) = \begin{cases} 1 & \text{if } x_t = m \\ 0 & \text{if } x_t \neq m \end{cases} \tag{4.2}$$

2. Block:

$$\eta_{i,j}(x_t) = c_{i,j,0} + c_{i,j,1}\mathrm{I}_1(x_t) + c_{i,j,2}\mathrm{I}_2(x_t), \tag{4.3}$$

$$\mathrm{I}_1(x_t) = \begin{cases} 1 & \text{if } x_t \in \{7, 8, ..., 16\} \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

$$\mathrm{I}_2(x_t) = \begin{cases} 1 & \text{if } x_t \in \{17, 18, 19, 20\} \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

3. Quadratic functional transition:

$$\eta_{i,j}(x_t) = c_{i,j,0} + c_{i,j,1}\left(\frac{x_t}{24}\right) + c_{i,j,2}\left(\frac{x_t}{24}\right)^2 \tag{4.6}$$

4. Sinusoidal functional transition:

$$\eta_{i,j}(x_t) = c_{i,j,0} + c_{i,j,1}\cos\left(\frac{2\pi x_t}{24}\right) + c_{i,j,2}\sin\left(\frac{2\pi x_t}{24}\right), \tag{4.7}$$

where, $c_{i,j,m}, \;\; m = 1, ..., 24$ are parameters.

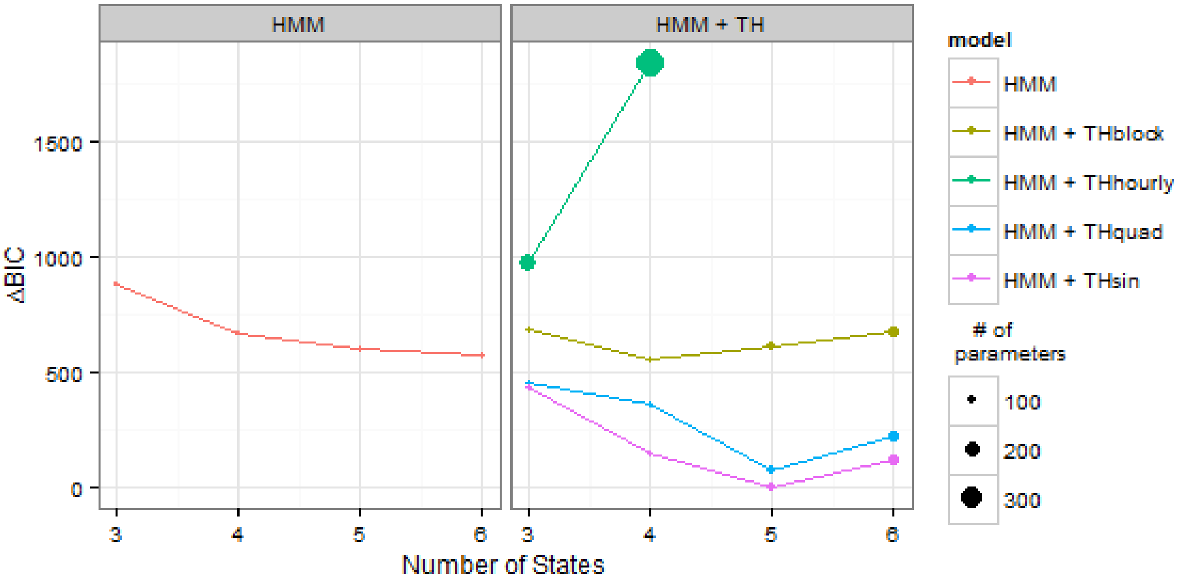

Figure 4.3: HMM BIC Comparison Among States

The 5-state HMM with sinusoidal transition is chosen via BIC to be the best model among all the other time-dependent transition HMMs. The BIC-optimal number of states decreases by one after accounting for temporal heterogeneity. Thus, adding serial dependence to the FMM (resulting in a time-homogeneous HMM) increases the number of BIC-optimal states from 3 to 6, but further extending the model by adding temporal heterogeneity in the transition probabilities (resulting in a time-heterogeneous HMM) decreases the optimal number of states from 6 to 5.

4.3 FMM with Temporal Heterogeneity

In the previous chapter, accounting for serial dependency using a underlying Markov process (HMM) increased the number of BIC-optimal states; however, the resulting HMMs fit better overall, with BIC values approximately 2500 units less than the corresponding FMMs. On the other hand, accounting for temporal heterogeneity transition in HMM gives a lower BIC than HMMs. So what happens when we account for temporal heterogeneity but not serial dependency?

Using the same multinomial logistic method described in section 2, we can define the mixing proportion λ_m in (2.1) as a function of covariate $\mathbf{X}$:

$$\lambda_m(x_t) = \frac{\exp(\eta_m(x_t))}{1 + \sum_{m=2}^{M} \exp(\eta_m(x_t))}, \;\; m = 2, ..., M, \tag{4.8}$$

$$\lambda_1(x_t) = 1 - \sum_{m=2}^{M} \lambda_m. \tag{4.9}$$

For simplicity, we only considered $\eta_m(x_t)$ a sinusoidal function:

$$\lambda_m(x_t) = c_{m,0} + c_{m,1}\mathrm{cos}\left(\frac{2\pi x_t}{24}\right) + c_{m,2}\mathrm{sin}\left(\frac{2\pi x_t}{24}\right), \tag{4.10}$$

where c_m are parameters.

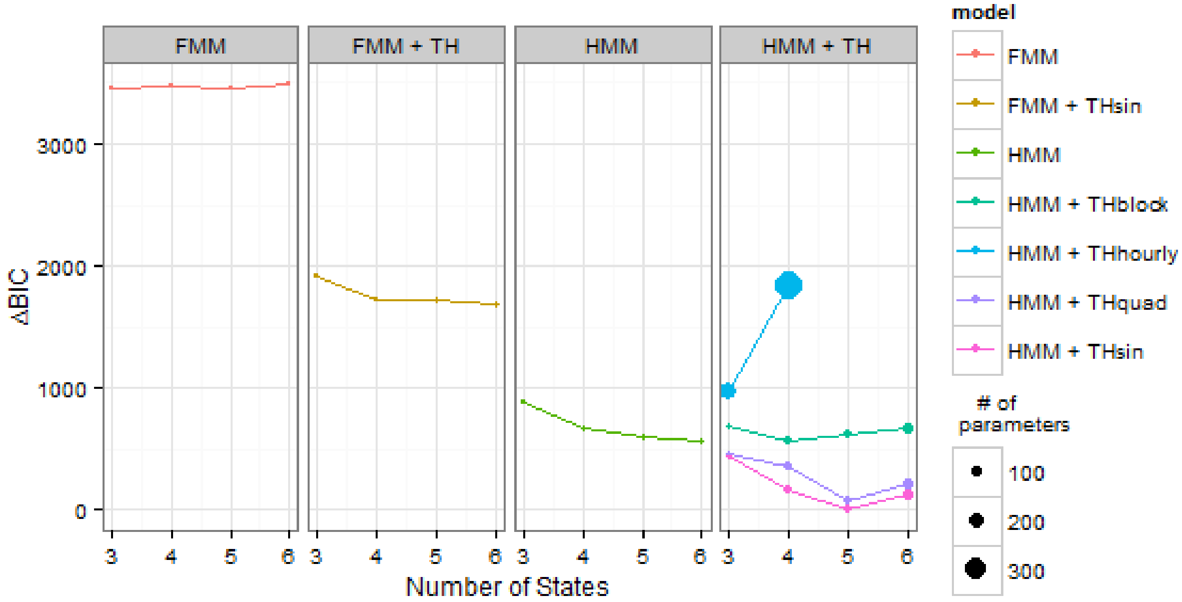

Figure 4.4: BIC Comparison of All Models

Comparing all the different type of models, the minimum FMM has 3 states as mentioned in the previous chapter. Models with either temporal heterogeneity (FMM with sinusoidal functional mixing proportion to account for temporal heterogeneity) or serial dependency (HMM), but not both, have minimum-BIC models with six states and both yield a lower BIC than FMM. Thus, increasing complexity in FMM may increase the estimated number of states. However, temporally heterogeneous HMMs fit best overall and suggest 5 underlying states. Thus, increasing complexity to account for serial dependency and temporal heterogeneity does in fact decrease the number of estimated states.

Chapter 5

Simulation Validation

5.1 Average Distance vs Time of Day

The ultimate goal in this thesis is to build a HMM capable of simulating data similar to the patterns exhibited by the Florida panthers. In van de Kerk et al (2015) and other animal research studies (Langrock *et al.*, 2014), researchers use the Viterbi algorithm describe at the end of chapter 2 to simulate the movement sequence. In this thesis, we will not use this method because we are not interested in finding a particular hidden sequence that best describes the observed sequence, but rather in simulating sequences based on the parameterized model. We would like simulations based on our models to produce patterns similar to Figure 4.2. For every panther, we simulated the same number of observations as the the observed data (14645, 10250, 10286, 9458 observations for cat 48, 94, 130, 131 respectively).

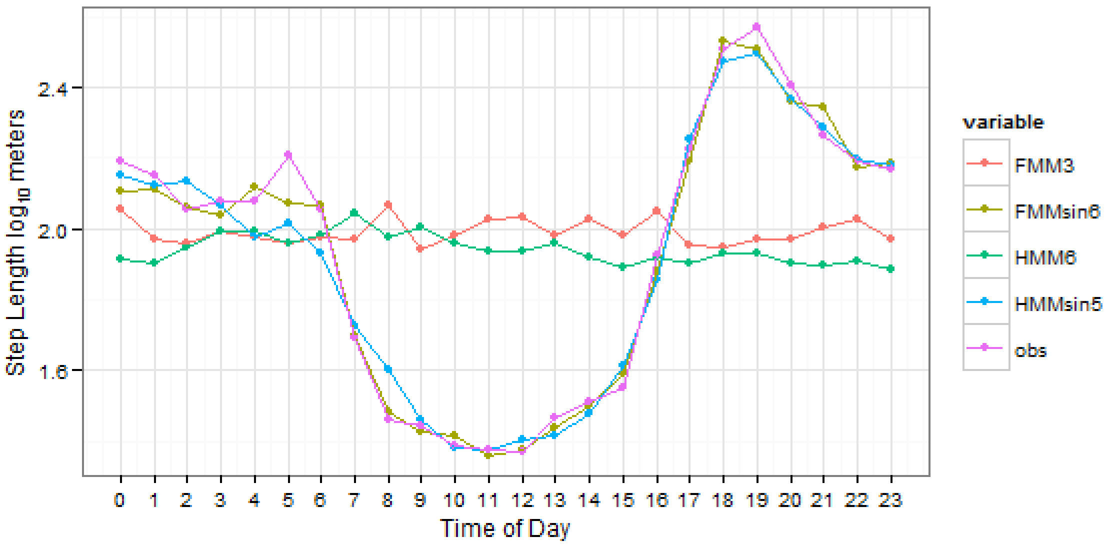

Figure 5.1: Average Step Length by Time of Day (ID = 130)

Models that assume temporal homogeneity (FMM and HMM) fail to capture the diurnal activity cycle exhibited by Florida panthers; the average step length is very similar at any given time of day for temporally homogeneous FMM and HMM (and would be more similar for a longer simulation). The average step length for FMM of large amount of observations is approximately the sum of the estimated mean times the mixing proportion:

$$\hat{\mu}_{\text{total}} = \sum_{m=1}^{M} \hat{\lambda}_m \hat{\mu}_m. \tag{5.1}$$

Similarly, for time-homogeneous HMM, long sequences of a Markov process converges to a stationary distribution such that the average step length is the sum of the relative occupancy times the mean of the state. Thus, for long sequences, the average step length given in any time of day simulated from FMM and HMM fails to capture the desired diurnal cycle.

Models incorporating temporal heterogeneity (FMM with sinusoidal temporal mixing proportion and HMM with sinusoidal temporal transitions) can capture the observed patterns in the hourly average step. Note that the simulated results from FMM with sinusoidally time-varying mixing proportions match the diurnal pattern of average step lengths reasonably well; nevertheless, its higher BIC ($\triangle$BIC ≈ 1500 units) indicates that the sinusoidally time-varying HMMs fits better overall.

5.2 Autocorrelation Function

In the previous section, we showed that models that account for temporal heterogeneity (FMM with sinusoidal temporal mixing proportion and HMM with sinusoidal temporal transitions) can capture the diurnal cycle exhibited by Florida panthers. We will consider another simulation validation approach to examine goodness of fit by autocorrelation plots. Autocorrelation is the correlation of a random variable with its own time-shifted sequence often used to examine serial correlation within data; it is a common tool in time series analysis. The autocorrelation function (ACF) is defined as follows

$$R_l = \frac{\mathrm{E}(Y_t - \mu_Y)(Y_{t+l} - \mu_Y)}{\sigma_Y^2} \tag{5.2}$$

where l is the lag or number of time steps in the sequence (Brockwell and Davis, 2006).

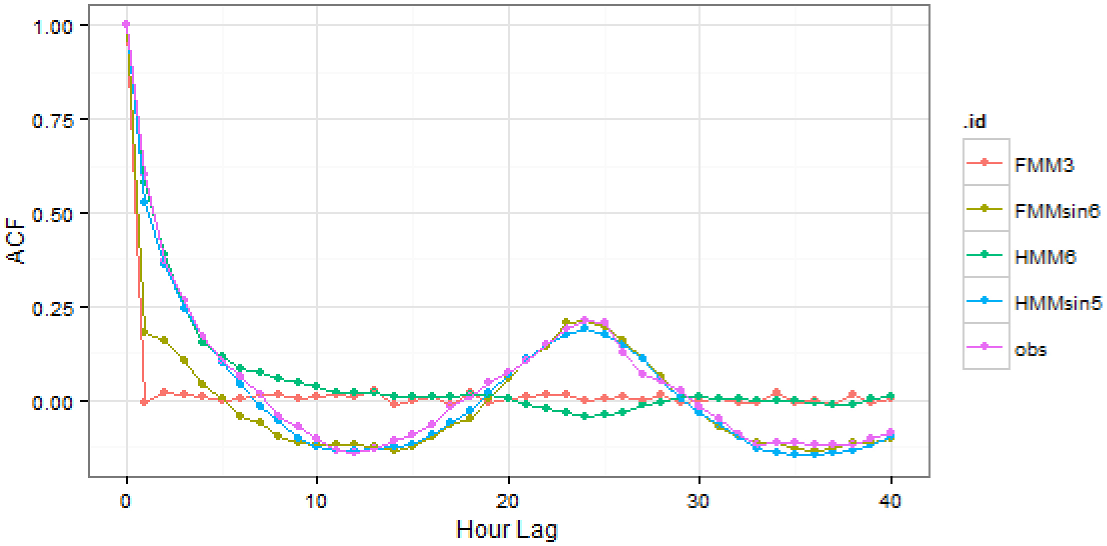

Figure 5.2: Autocorrelation Plot (ID = 130)

FMM has no autocorrelation at all. HMM without temporal heterogeneity can capture the first few lags of the ACF but is unable to capture lags beyond 7. FMM with temporal heterogeneity can capture lags beyond 12, but not short-term correlation. HMM with temporal heterogeneity can fully capture the autocorrelation pattern of the observed data.

5.3 Simulation Comments

There is no single way to select the best model that best describes the observed sequence. We used several reasonable approaches to compare these models: BIC, ACF plots and average step length vs time of day plots. While each metric has its own limitations, examining them together can give a good understanding of the model structure that best describes the observed sequence. It is hard to evaluate simulation via Viterbi algorithm because it is strongly influenced by the observational sequence itself; thus, it can simulate a strong diurnal pattern even from a model with no diurnal structure.

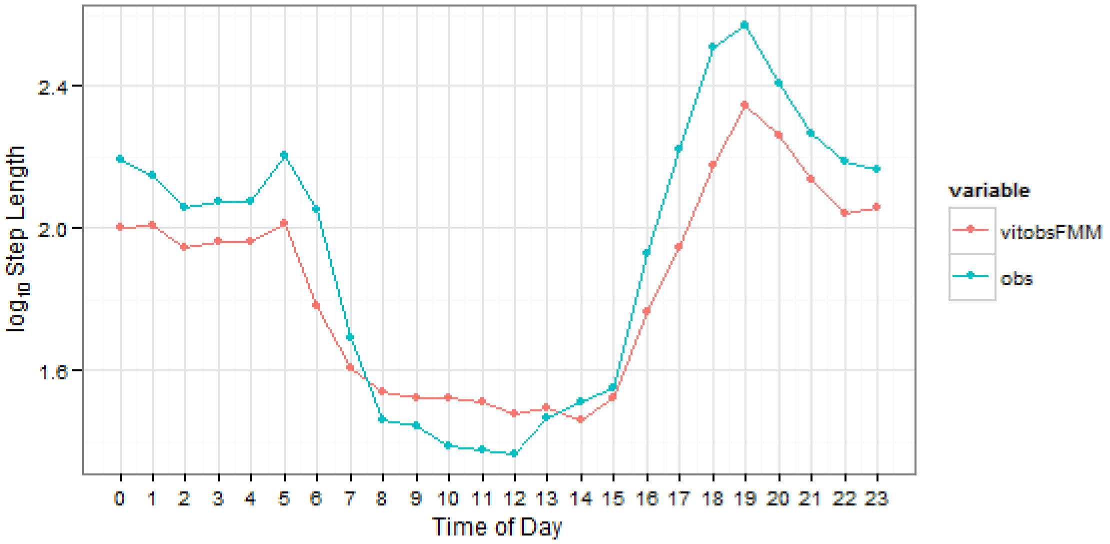

Figure 5.3: Viterbi Average Step Length by Time of Day

To illustrate this point, we used the three-state FMM (the worst fit with the highest BIC among all models), simulated via the Viterbi algorithm and compared it with the observed data (figure 5.3, 5.4). From the previous section, FMM models have approximately constant average step lengths regardless of the time of day — thus simulations from the model are unable to capture the diurnal cycle or to produce a realistic ACF pattern, yet the simulation via the Viterbi algorithm can capture the diurnal cycle and autocorrelation. Since the Viterbi algorithm is conditional upon the observed sequence, even FMM models can capture serial dependency and temporal heterogeneity through the information contained in the observed sequence.

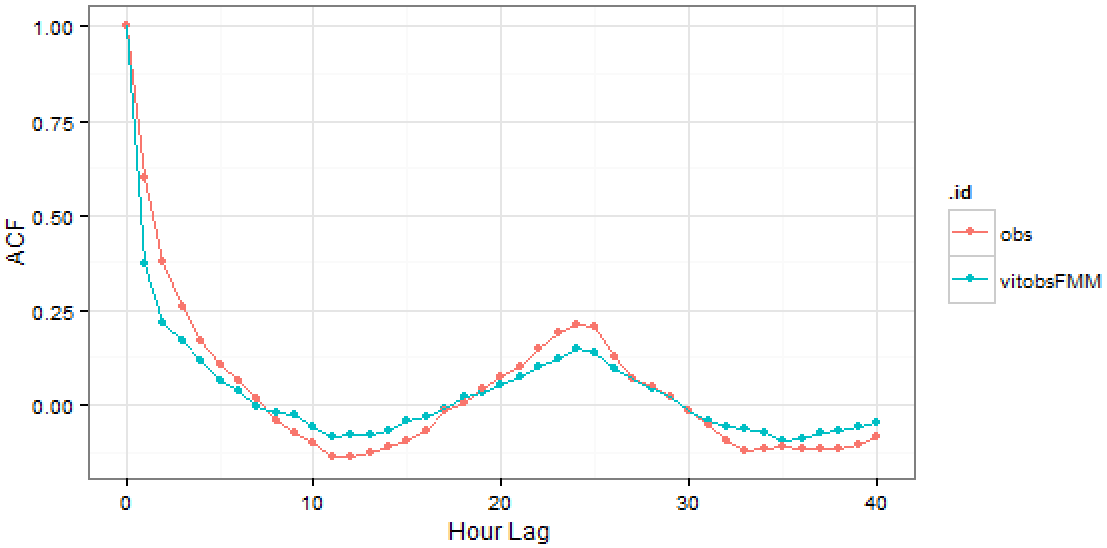

Figure 5.4: Viterbi ACF plot

Chapter 6

Conclusion

6.1 Multiple States

The biggest issue in analyzing mixture models with multiple underlying discrete states is interpreting them biologically. The states themselves have no intrinsic meaning (i.e, we do not know independently the state with the shortest step length really represents a "resting" state), though they can identify structure present in the data. We do not know if Florida panthers really move according to a particular, discrete set of hidden states, and even if they do, we don't know the true number nor can we observe them. Three distinct movement states have a useful biologically interpretable meaning for Florida panthers: Short step lengths correspond to resting states, intermediate step lengths correspond to foraging state and long step lengths correspond to a traveling state. A distinct long-step-length state is always identified, whereas short step lengths are often represented by multiple, nearly overlapping states. Thus, models with a small number of short step length states with higher variance in the emission (step length) distribution are easier to interpret biologically and logically than those with multiple discrete short step length states with low variance.

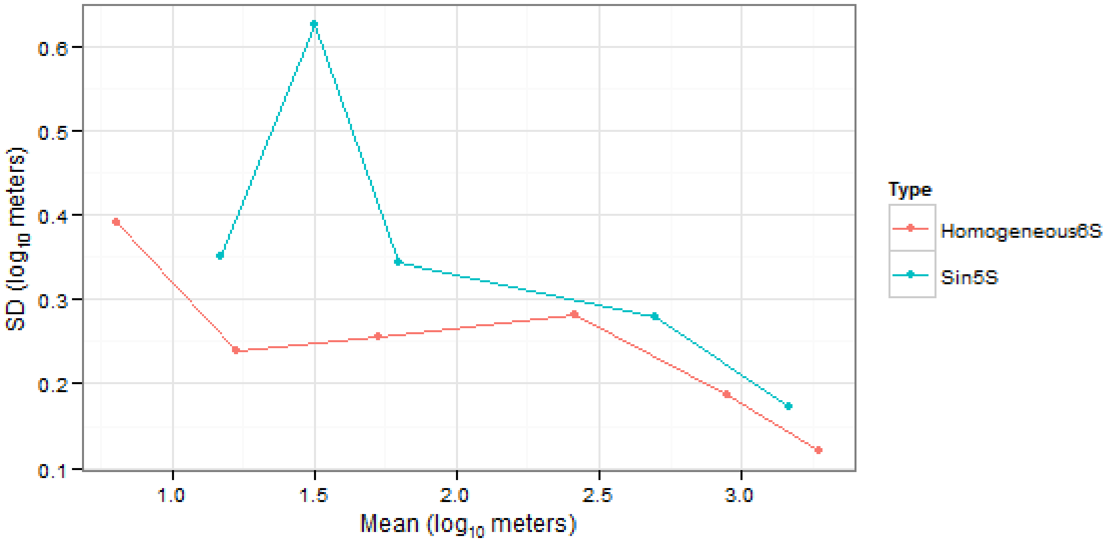

Figure 6.1: Mean vs. Standard deviation ($\log_{10}$ meters)

6.2 Discussion

Identifying behavioural states based on a set of observations is a common methodological problem in behavioural ecology; finding solutions that are both biologically and statistically interpretable and defensible is challenging. Models that suggest a higher number of states fit the data well but are hard to interpret biologically; models that are biologically interpretable are not statistically supported. While the HMM framework has been applied in many research areas, very few studies have considered modeling with more than 3 states. HMM is a simple, straight forward framework, but HMM extensions that add a small amount of complexity can complicate the model greatly and reveal deeper understanding. For example, van de Kerk *et al.* (2015b) used 2 to 6 movement states HSMM to model Florida panthers; Morales *et al.* (2004)

used HMM with transition probabilities depending on distances from different habitats to model elks; Schliehe-Diecks *et al.* (2012) used a model with time-dependent transitions but with a smaller resolution dataset and time intervals between successive points are 1 minute intervals, 40 minutes between 18:00 and 24:00 per nigh. Outside of ecology and animal behaviour, Raffa and Dubin (2015) used mixed effect transition and emission HMM in a smoking study.

We have considered a conceptually simple yet little-studied extension by which covariate independent HMM can overestimate the number of true states in covariate dependent HMM (in our case, time-dependent transitions). The motivation in researching this direction was driven by two factors: a large real life dataset that can freely, flexibly and capable of model more complex extensions of HMM and an interesting statistical and biological problem to understand hidden states in higher magnitude. HMM with more than three states are not biologically interpretable and are often discarded, but we cannot ignore the observation that increasing complexity (number of states) can increase goodness of fit statistically. Our analyses revealed that modeling temporal heterogeneity in HMM transition can reduce the number of states estimated in time homogeneous transitions HMM movement of Florida panthers and reduce the overall model complexity.

For some kinds of model predictions, the time-homogeneous transition model can perform as well as the time-inhomogeneous transition models. In particular, when predicting hidden states using the Viterbi algorithm; models with interesting additional structure can easily be overlooked. The Viterbi algorithm predicts the most likely sequence of movement states based on the observations, which allows the time-homogeneous transition model's prediction to incorporate the temporal heterogeneity from the observations. In general, the time-homogeneous transition model gives the broader scope of movement pattern whereas the time-inhomogeneous transitions models gives a more realistic movement structure. It is important to contrast a variety of metrics (i.e, ACF, simulated diurnal patterns, etc) to get a reliable understanding of how well models fit the data.

6.3 Generalization - Simulation

From the Florida panther analysis, we want answer the general question whether the HMMs that neglect covariate dependence (when the data really contain it) will overestimate the number of hidden states via BIC. We simulated data from a basic block transition model (described in section 4) with Gaussian and Poisson emissions. The covariate is a binary indicator variable that switches every 12 time steps.

1. Gaussian HMM:

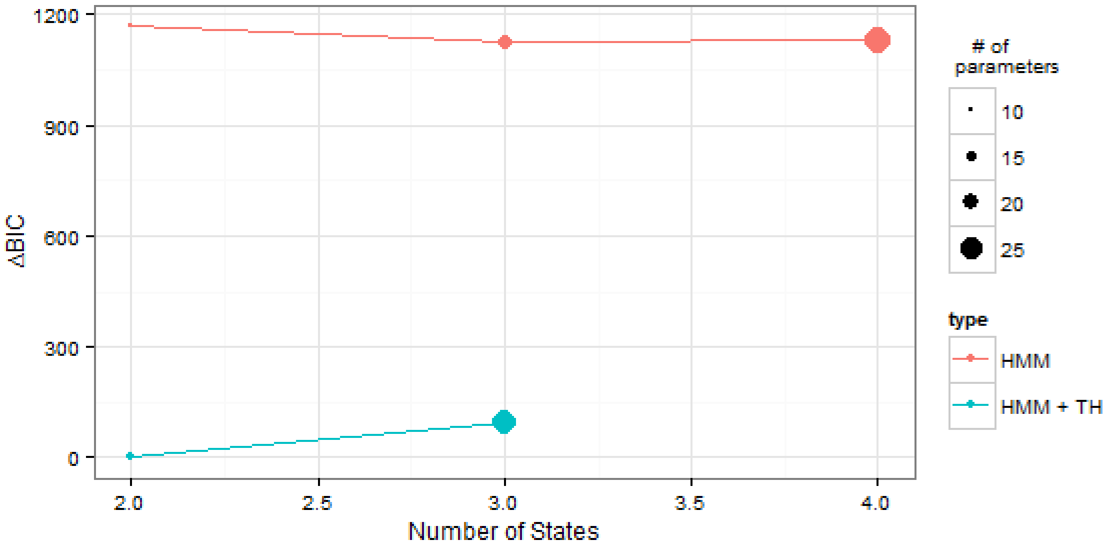

Figure 6.2: BIC Comparison of 2-state covariate dependent Gaussian HMM simulation

Model	Parameters	Parameter Value/Estimate
2-state HMM with Covariate (True Model)	$c_{1,2,0}$	-3
	$c_{1,2,1}$	4
	$c_{2,2,0}$	3
	$c_{2,2,1}$	-4
	μ_1	3
	μ_2	4.5
	σ_1	1
	σ_2	1
2-state HMM with Covariate (Fitted Model)	$\widehat{c_{1,2,0}}$	-2.785
	$\widehat{c_{1,2,1}}$	4.041
	$\widehat{c_{2,2,0}}$	2.89
	$\widehat{c_{2,2,1}}$	-3.86
	$\hat{\mu_1}$	2.98
	$\hat{\mu_2}$	4.5
	$\hat{\sigma_1}$	0.978
	$\hat{\sigma_2}$	1.00
2-state HMM without Covariate	$\hat{\mu_1}$	3.22
	$\hat{\mu_2}$	4.19
	$\hat{\sigma_1}$	1.12
	$\hat{\sigma_2}$	1.17
3-state HMM without Covariate	$\hat{\mu_1}$	3.13
	$\hat{\mu_2}$	3.61
	$\hat{\mu_3}$	4.68
	$\hat{\sigma_1}$	1.04
	$\hat{\sigma_2}$	1.15
	$\hat{\sigma_3}$	0.97

Table 6.1: Parameters and parameter estimates of Gaussian HMM

2. Poisson emissions:

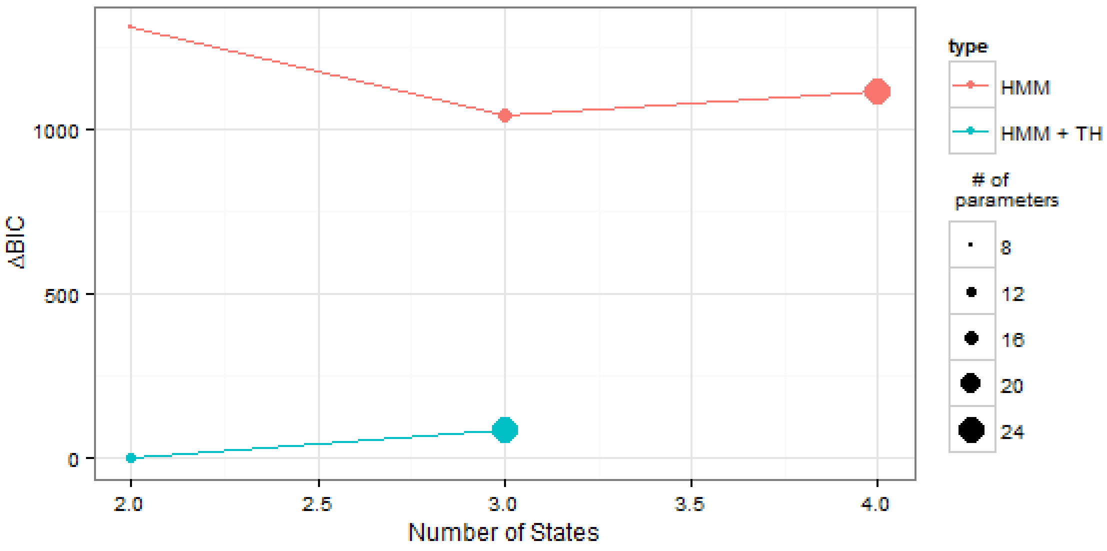

Figure 6.3: BIC Comparison of 2-state covariate dependent Poisson HMM simulation

Model	Parameters	Parameter Value/Estimate
2-state HMM with Covariate (True Model)	$c_{1,2,0}$	0
	$c_{1,2,1}$	0
	$c_{2,2,0}$	3
	$c_{2,2,1}$	-4
	μ_1	1
	μ_2	2
2-state HMM with Covariate (Fitted Model)	$\widehat{c_{1,2,0}}$	0.218
	$\widehat{c_{1,2,1}}$	-0.251
	$\widehat{c_{2,2,0}}$	2.16
	$\widehat{c_{2,2,1}}$	-2.84
	$\hat{\mu}_1$	0.99
	$\hat{\mu}_2$	2.00
2-state HMM without Covariate	$\hat{\mu}_1$	1.02
	$\hat{\mu}_2$	2.01
3-state HMM without Covariate	$\hat{\mu}_1$	1.00
	$\hat{\mu}_2$	1.99
	$\hat{\mu}_3$	2.01

Table 6.2: Parameters and parameter estimates of Poisson HMM

6.4 Future Research

Although time-inhomogeneous transition models offer an objective approach to (1) reduce the number of states to a biologically interpretable level, (2) capture the observational autocorrelation and temporal patterns, (3) reduce overall model complexity while improving goodness of fit, several challenges must be overcome in order to maximize its potential. The current model is univariate, based only on between-location step lengths and ignoring the corresponding turning angles. Predictions of movement trajectories in two dimensions requires us to characterize the distributions of both step length and turning angles. Furthermore, modeling temporal heterogeneity only reduces the number of BIC-optimal states by 1, resulting in models with 4 or 5 states. This suggests that there may be other factors such as habitat-dependent behaviour, seasonal-temporal heterogeneity, etc., that might also contribute to over predicting number of states. Alternatively, there might really be additional classes of behaviour that biologists have been overlooking. It would be interesting to look at other complexities such as habitant-dependent behaviour (influencing transition probability and/or emission probability); changes in behaviour over longer time scales (seasonal, animal's lifespan/ age); attraction/repulsion from features in the environment (road, con-specifics of the same or opposite sex), etc., that could eventually be added to the modeling framework. Despite the challenges, our study provides an important extension to HMM modeling that has rarely been applied before, as well as a real-life application that contributes to the field of animal movement ecology.

Appendix A

Florida Panther ID 131, 48, 94

Panther ID 131, 48, and 94 are the other three panthers with the most data available with 9458, 14645, 10250 observations respectively. There are two males (CatID 130, 131) and two females (CatID 48, 94) in our study; one female (CatID 94) produced a litter during the 24 months that she was GPS-tracked. The step length distribution of this individual female differs from that of the other three panthers. Our models did not account for presence of kittens as a factor, but it is a good direction to investigate in the future.

A.1 CatID 131

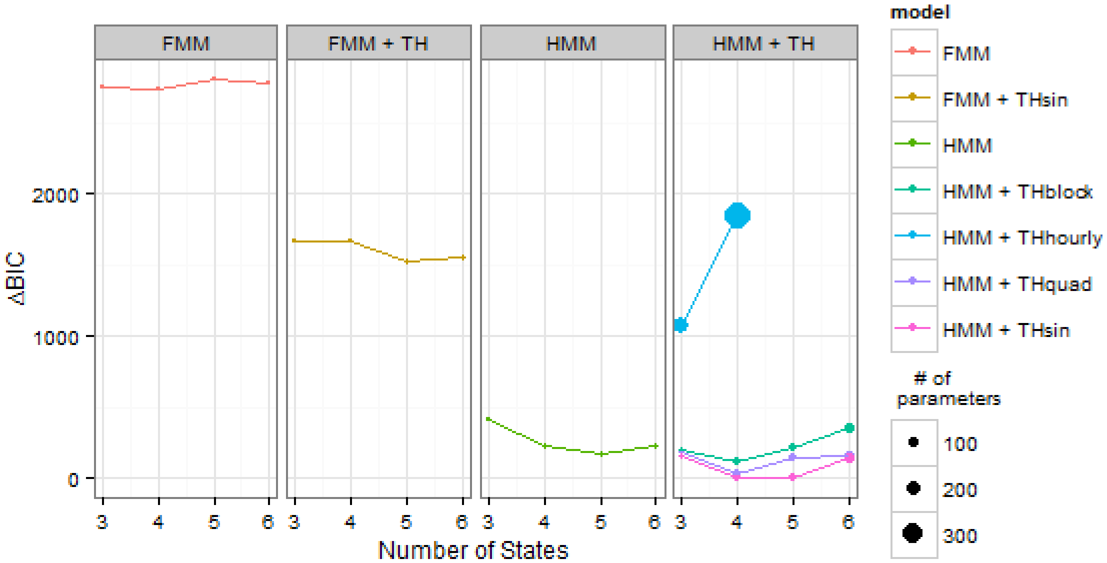

Figure A.1: BIC Comparison of All Models (ID=131)

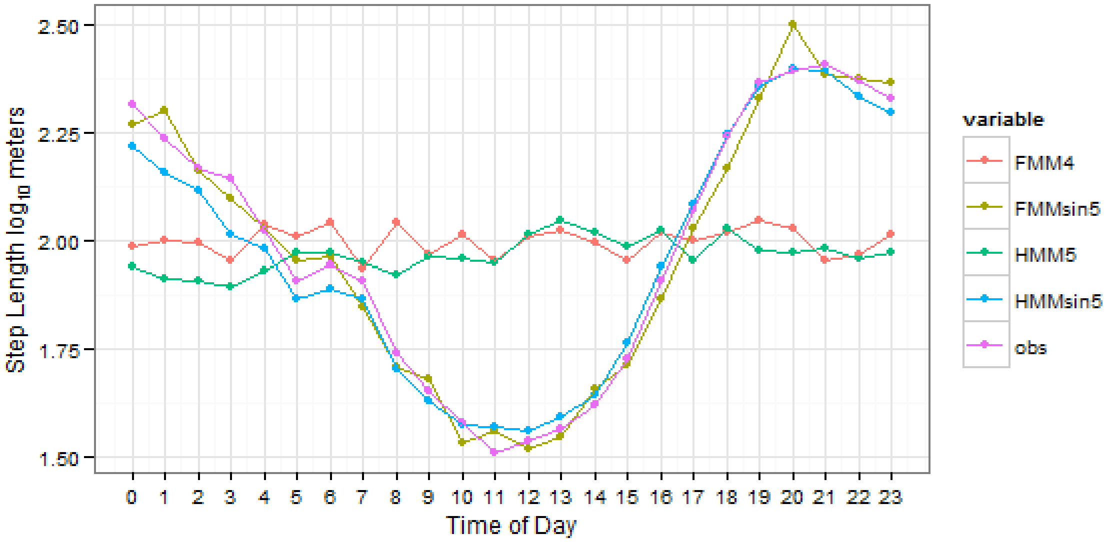

Figure A.2: Average Step Length by Time of Day (ID = 131)

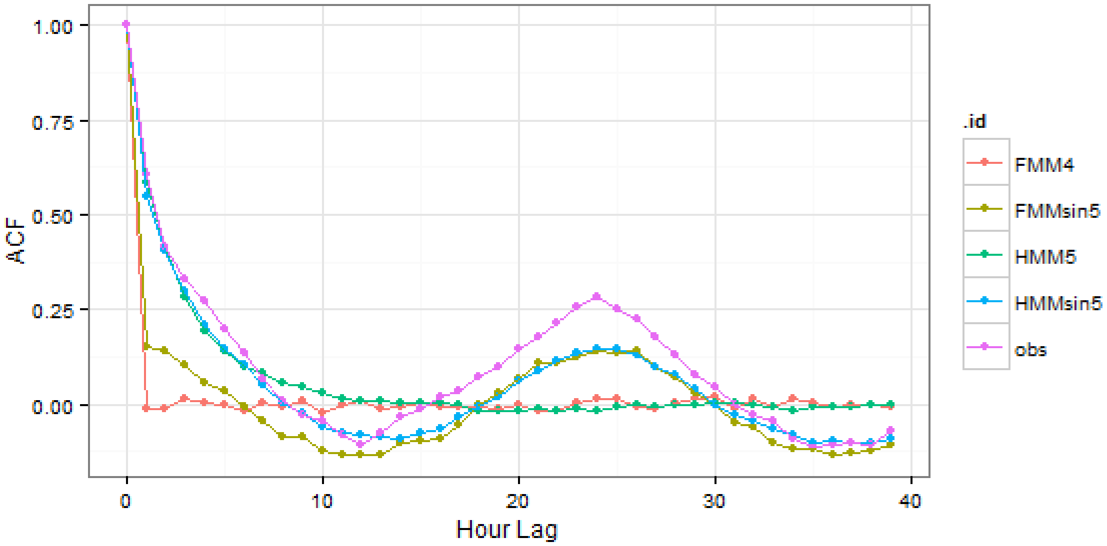

Figure A.3: Autocorrelation Plot (ID = 131)

A.2 CatID 48

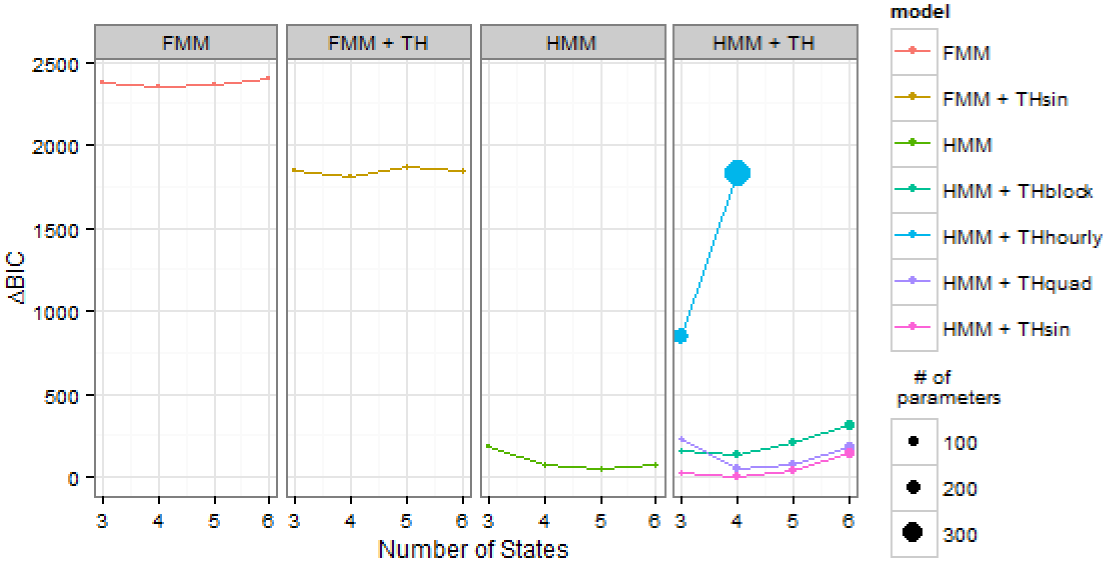

Figure A.4: BIC Comparison of All Models (ID=48)

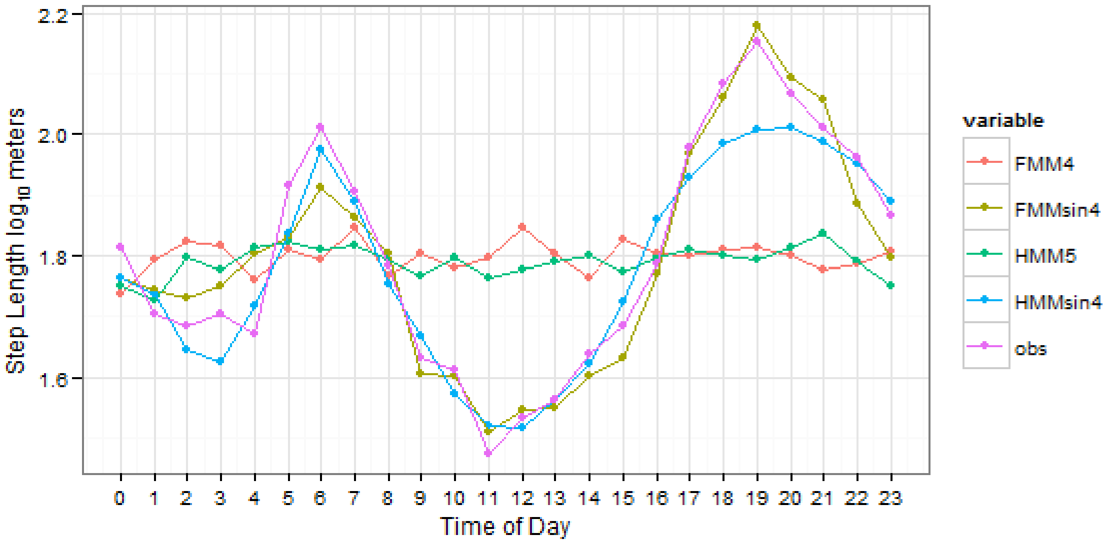

Figure A.5: Average Step Length by Time of Day (ID = 48)

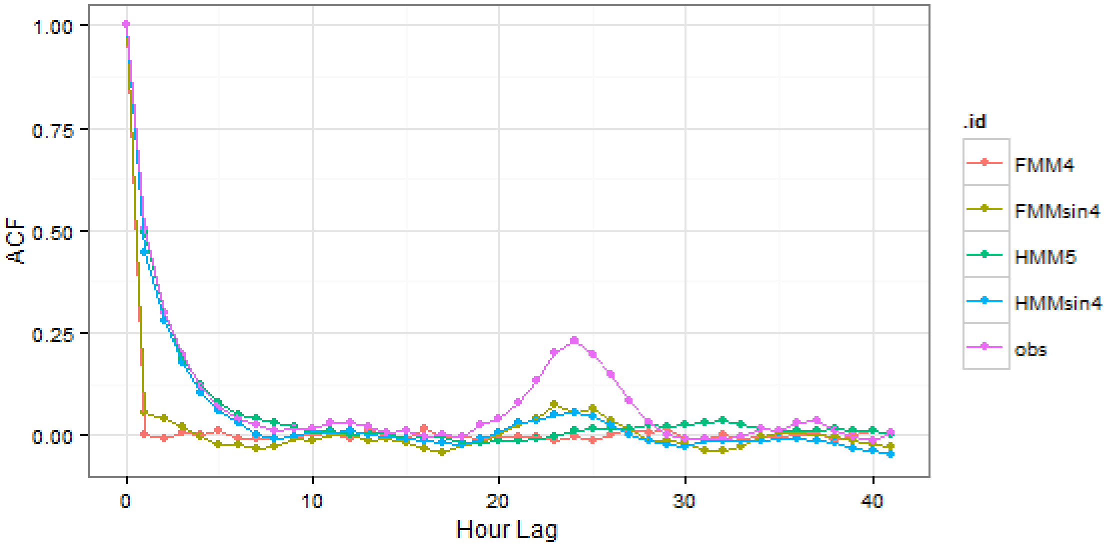

Figure A.6: Autocorrelation Plot (ID = 48)

A.3 CatID 94

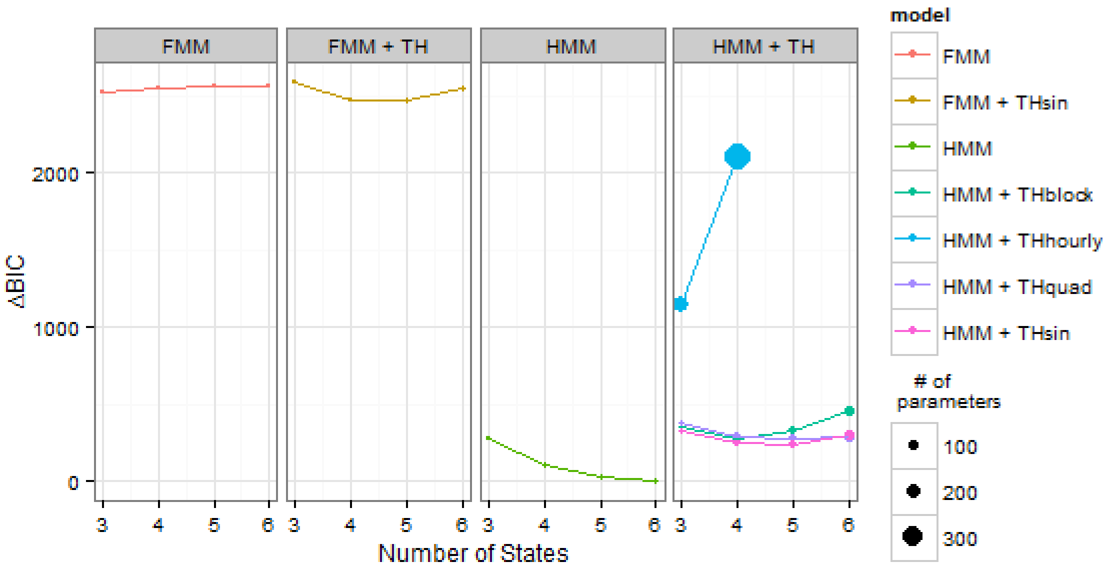

Figure A.7: BIC Comparison of All Models (ID=94)

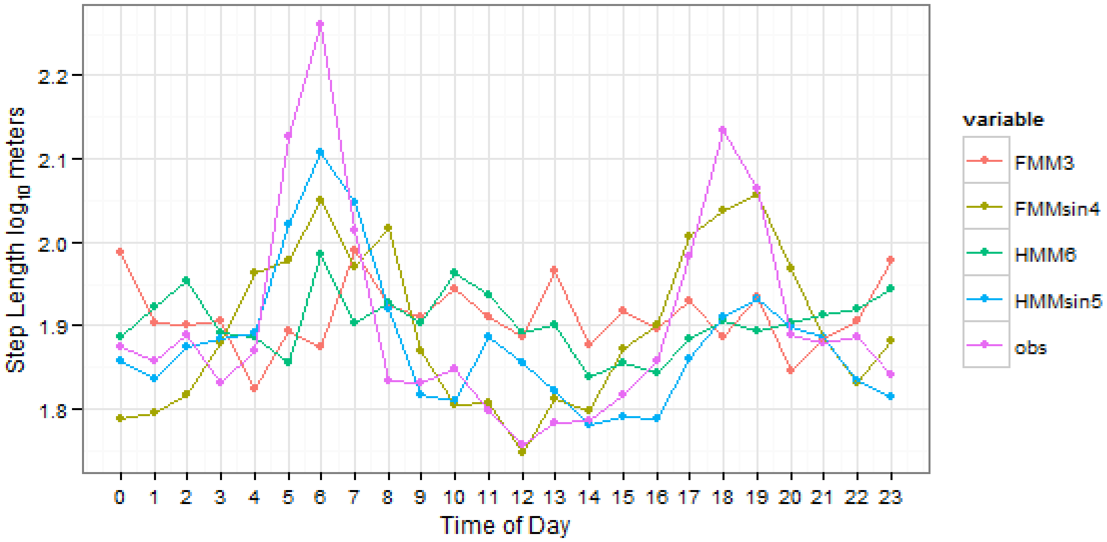

Figure A.8: Average Step Length by Time of Day (ID = 94)

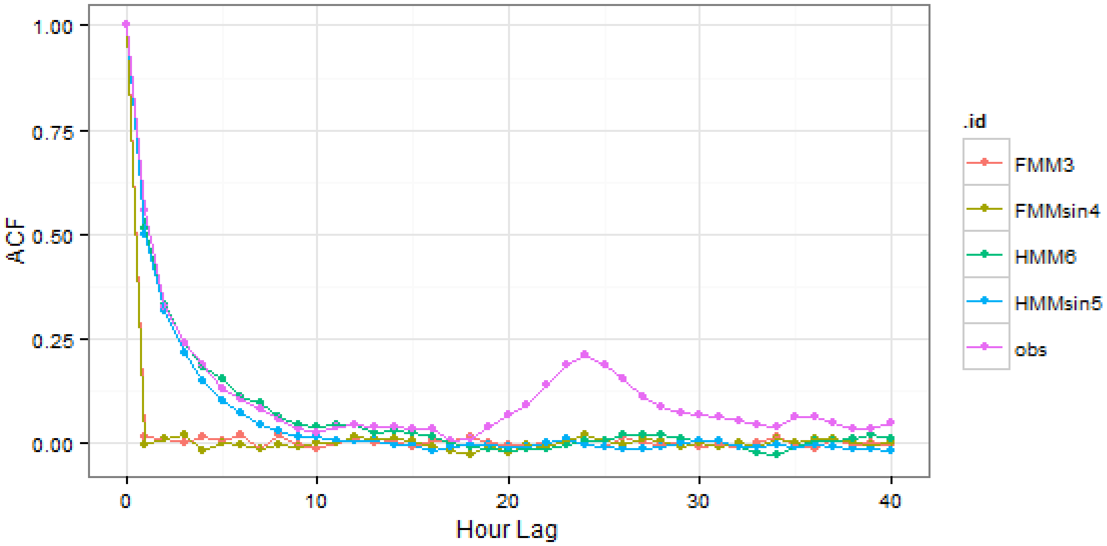

Figure A.9: Autocorrelation Plot (ID = 94)

Bibliography

Augustine, B. and Langrock, R. (2014). *move.HMM: Fit HMM and HSMM animal movement models.* R package version 1.0.1.

Baum, L. E. (1972). An equality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, **3**, 1–8.

Baum, L. E., Sell, G. R., *et al.* (1968). Growth transformations for functions on manifolds. *Pacific J. Math*, **27**(2), 211–227.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, pages 164–171.

Brockwell, P. J. and Davis, R. A. (2006). *Introduction to time series and forecasting.* Springer Science & Business Media.

Coffin, A. W. (2007). From roadkill to road ecology: A review of the ecological effects of roads. *Journal of transport Geography*, **15**(5), 396–406.

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annu. Rev. Psychol.*, **57**, 505–528.

Dean, B., Freeman, R., Kirk, H., Leonard, K., Phillips, R. A., Perrins, C. M., and Guilford, T. (2012). Behavioural mapping of a pelagic seabird: combining multiple sensors and a hidden Markov model reveals the distribution of at-sea behaviour. *Journal of The Royal Society Interface*, page rsif20120570.

Firle, S., Bommarco, R., Ekbom, B., and Natiello, M. (1998). The influence of movement and resting behavior on the range of three carabid beetles. *Ecology*, **79**(6), 2113–2122.

Franke, A., Caelli, T., Kuzyk, G., and Hudson, R. J. (2006). Prediction of wolf (Canis lupus) kill-sites using hidden Markov models. *Ecological Modelling*, **197**(12), 237–246.

Fryxell, J. M., Hazell, M., Brger, L., Dalziel, B. D., Haydon, D. T., Morales, J. M., McIntosh, T., and Rosatte, R. C. (2008). Multiple movement modes by large herbivores at multiple spatiotemporal scales. *Proceedings of the National Academy of Sciences*, **105**(49), 19114–19119.

Himmelmann, S. S. D. D. L. and www.linhi.com (2010). *HMM: HMM - Hidden Markov Models*. R package version 1.0.

Johnson, W. E., Onorato, D. P., Roelke, M. E., Land, E. D., Cunningham, M., Belden, R. C., McBride, R., Jansen, D., Lotz, M., Shindle, D., Howard, J., Wildt, D. E., Penfold, L. M., Hostetler, J. A., Oli, M. K., and O'Brien, S. J. (2010). Genetic Restoration of the Florida Panther. *Science*, **329**(5999), 1641–1645.

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**(11), 2336–2342.

Langrock, R., Hopcraft, J. G. C., Blackwell, P. G., Goodall, V., King, R., Niu, M., Patterson, T. A., Pedersen, M. W., Skarin, A., and Schick, R. S. (2014). Modelling group dynamic animal movement. *Methods in Ecology and Evolution*, **5**(2), 190–199.

Lanza, S. T., Donny, E. C., Collins, L. M., and Balster, R. L. (2004). Analyzing the acquisition of drug self-administration using growth curve models. *Drug and alcohol dependence*, **75**(1), 11–21.

McKellar, A. E., Langrock, R., Walters, J. R., and Kesler, D. C. (2014). Using mixed hidden Markov models to examine behavioral states in a cooperatively breeding bird. *Behavioral Ecology*, page aru171.

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.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., and Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, **105**(49), 19052–19059.

O'Connell, J. and Hjsgaard, S. (2011). Hidden semi Markov models for multiple observation sequences: The mhsmm package for R. *Journal of Statistical Software*, **39**(4), 1–22.

Patterson, T. A. (2009). Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology*, **78**(6), 1113–1123.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**(2), 257–286.

Raffa, J. D. and Dubin, J. A. (2015). Multivariate longitudinal data analysis with mixed effects hidden Markov models. *Biometrics*.

Schliehe-Diecks, S., Kappeler, P. M., and Langrock, R. (2012). On the application of mixed hidden Markov models to multiple behavioural time series. *Interface Focus*, **2**(2), 180–189.

van de Kerk, M., Onorato, D. P., and Oli, M. K. (2015a). The Florida panther: Past, present, and future. Technical report, University of Florida. `www.researchgate.net/publication/273968104_The_Florida_Panther_Past_Present_and_Future`.

van de Kerk, M., Onorato, D. P., Criffield, M. A., Bolker, B. M., Augustine, B. C., McKinley, S. A., and Oli, M. K. (2015b). Hidden semi-Markov models reveal multiphasic movement of the endangered florida panther. *Journal of Animal Ecology*, **84**(2), 576–585.

Visser, I. and Speekenbrink, M. (2010). depmixS4: An R package for hidden Markov models. *Journal of Statistical Software*, **36**(7), 1–21.

Zucchini, W. and MacDonald, I. L. (2009). *Hidden Markov Models for Time Series: An Introduction Using R*. CRC Press.