A NOVEL APPROACH AT UNCOVERING THE LATENT MOVEMENT STATES AND PREDICTING KILL-SITES OF LIONS USING HIDDEN MARKOV MODELS

Jenicca Poongavanan

Submitted in the fulfilment of the requirements for the degree of

BSc Mathematical Statistics

in the

Faculty of Science

at the

Nelson Mandela Metropolitan University

Supervisor: Dr V.L. Goodall
January 2017

ABSTRACT

Quantifying kill rates remains a key limitation to addressing many predator-prey questions. To date the most common approach to identify lion kill sites is done by identifying clusters of locations obtained using Global Positioning System (GPS) collars on predators. However, if clusters were determined by different decision rules, comparison across studies were not possible. This study investigated Hidden Markov Models (HMMs) as a predictive modelling technique. The Hidden Markov Model (HMM) is a stochastic model in which the system is modelled by a Markov process with “hidden” states. The states of the Markov chain can be interpreted as providing rough classifications of the animal behavioural dynamics. HMMs have been previously used in the analysis of wolves (*Canis lupus*), where aerial surveys were conducted to locate wolf-killed ungulate carcasses.

The objectives of this project were to evaluate whether HMMs can predict observer-confirmed kill sites from GPS lion relocation data and also provide additional insight into lion behaviour. GPS radio-collars have been used to obtain movement data of for six lions (*Panthera leo*) in the Kruger National Park, South Africa. HMMs were fitted for each lion and a set of states were predicted, from which the behaviours were inferred. This state sequence was used to identify potential kill sites, and the actual kill data were then used to confirm these sites. The observer-confirmed kill sites provide a unique opportunity to validate predicted kill sites from HMM modelling using GPS tracking data as it is seldom possible to validate the state predictions from an animal movement study.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to my supervisor Dr Victoria Lucy Goodall for always being there to guide, advise and help me for this project. I would also like to thank my parents for their love and support during my studies. My sincere regards to my friends especially Bracken Van Niekerk for her help, love and fun times when everything was going wrong. Additional thanks goes to the NMMU Statistical department for being the family away from home as well as to the National Research Foundation and NMMU for funding this project.

Table of Contents

ABSTRACT **2**

TABLE OF CONTENTS **4**

LIST OF FIGURES **6**

LIST OF TABLES **7**

CHAPTER 1 – INTRODUCTION **1**

CHAPTER 2- DATA SET **2**

2.1 Wildlife Tracking **2**

2.2 Study Area **2**

2.2.1 Study Species - Lion 5

CHAPTER 3 – LITERATURE REVIEW **6**

3.1 Model **6**

3.1.1 Deterministic Model 6

3.1.2 Stochastic Models 7

3.2 Background – Hidden Markov Models **8**

3.2.1 Hidden Markov Models used in animal movement research. 9

3.2.2 Extensions of the HMM used in animal movement research 10

3.2.3 Statistical Methods used to Identify kill sites 13

CHAPTER 4 – HIDDEN MARKOV MODEL **17**

4.1 Methodology **17**

4.1.1 Basic HMM 17

4.1.2 Moments Derivation 19

4.1.3 Parameter Estimation 20

4.1.4 Viterbi Algorithm and State Decoding 21

4.1.5 Model Selection and Model Checking 24

4.1.6 Missing Observation Data 25

CHAPTER 5 – HIDDEN MARKOV MODEL FOR ANIMAL MOVEMENT **26**

5.1 Analysis **26**

5.2 Results **27**

5.3 Conclusion **35**

CHAPTER 6 – KILL PREDICTIONS **36**

6.1 Introduction **36**

6.2 Methodology **36**
6.2.1 Training Model Analysis 36

6.3 Results **39**

6.4 Discussions **41**

CHAPTER 7 – CONCLUSION **43**

7.1 Summary **43**

7.2 Future Work **43**

REFERENCES **45**

APPENDICES **48**

List of Figures

Figure 2.1 - Study area with Mavumbye Male AU007 locations in the central region of the Kruger National Park, South Africa. 3

Figure 2.2 - Lion from Schotia Game Reserve (Photo - Jenicca Poongavanan) 5

Figure 4.1 - Directed graph of basic HMM (Langrock et al., 2012) 18

Figure 5.1 - Histogram of Thompson Female lion AM186 hourly displacement. 27

Figure 5.1 - Comparisons of Nsenami Female AM39/AM115 and Nsenami Male AM113/AM183 4-State HMM Activity State 31

Figure 5.2 - A comparison of the Three Female Lions at the KNP 32

Figure 5.3 - A comparison of the Three Male Lion HMM Activity state at the KNP. 33

Figure 5.4 - Activity State Comparison using HMM for Nsenami AM39/AM115 Female by Time of Day....34

Figure 5.5 - - Activity State Comparison using HMM for Nsenami Male by Time of Day....34

Figure 6.1 - Activity state of Thompson Female by Time of Day. 37

Figure 6.2 - Decision Rule analysis....39

Figure 6.1 - Nsenami Female Kill Prediction Analysis 40

Figure 6.2 - Proportion of kills to clusters 40

List of Tables

Table 2.1 - Details of GPS collars used in this study. 4

Table 2.2 - Details of clusters investigation. 5

Table 3.1 - Classification of Markov Processes (Ibe, 2013) 7

Table 3.2 - Shows the eight predictor variables they measured for each GPS clusters. 15

Table 5.1 - HMM outputs: Log-Likelihood and AIC 28

Table 5.2 - HMM parameter values and state dependent distribution means. 29

Table 6.1 - Different Decision rules applied to training model and results. 38

Table 6.2- Model predictions of missed kill. 42

Chapter 1 – Introduction

Over the last decade, ecology has developed a new sub-genre: ‘movement ecology, the study of how animals movements shape their overall ecology and the factors, both inherent and extrinsic, that influence movement’ (Patterson et al., 2016). Predation has a major influence on the dynamics of animal population and the ecosystem. Imbalances in the ratio of predator to prey can result in major regime changes and shifts in ecosystem (Ripple and Beschta, 2006). Calculation of predator kill-rates is important for gaining insight into predator-prey dynamics and management of predator-prey communities (Franke et al., 2006).

A practical problem is the adequate quantification of actual kill rates of predators, which requires locating all kill sites (Eberhardt et al., 2003), through continuous observation. Such methods are time-consuming, labor-intensive, or logistically unrealistic (Tambling et al., 2010). Lions are one of the most important predators, functionally and numerically, in many African systems (Kittle et al., 2016). Therefore it is of utmost important to develop and test models that increase the effectiveness of locating lion kill sites (Tambling et al., 2010). Chapter 2 of this project includes a review on different approaches used for locating lion kill sites in previous studies.

During recent years, the quality and quantity of animal movement data has improved and increased substantially. Advancements made in Global Positioning system (GPS) tracking collars for predators provides the potential to locate kill sites without the need for intensive monitoring (Webb et al., 2008). However, these approaches may require extensive field visitation to potential kill sites. Further, analytical methods that can discern characteristics of movement behaviors representing kill sites from GPS data is required (Webb et al., 2008). The objective of this project is to present an approach that uses the Hidden Markov Model (HMM) to identify observer-confirmed kill sites solely from the movement patterns of GPS collared lions and also provide an insight into lion’s behavior.

The initial part of the project consists of a section on the data sets that the project is based on. A sample of the data set is given. Chapter 3 is made up of the literature review where various application of the HMM is discussed. Chapter 4 consists of the origin and procedure of the HMM. All the algorithm coding and analyses are done using R statistical software (R Development Core Team, 2010) and the moveHMM package (Michelot et al., 2016). Based on the results, a summary and conclusions are presented with recommendations.

Chapter 2- Data Set

2.1 Wildlife Tracking

The study of animal ecology using telemetry tracking systems began in the late 1950s and early 1960s (LeMunyan et al., 1959) using collars or tags emitting very-high frequency (VHF) radio-signal pulses. Intensive monitoring of widely roaming animals with VHF systems, however, was costly, time consuming and often posed risk to personnel safety (Clark et al., 2006). Opportunities to document animal movement responses to changing environmental conditions over daily or longer time frames and with less danger have recently been facilitated by the advent of Global Positioning System (GPS) tracking (Owen-Smith, 2013).

Global Positioning System tracking has become a widely used remote sensing technique to monitor the movements of a wide range of species, both terrestrial and marine (Goodall, 2014). It includes the ability to determine the position on the surface of the earth with high precision and accuracy 24 hours a day, with position updates available in rapid succession (Cagnacci et al., 2010).

Although wildlife researchers made use of GPS devices for tracking animals, the technology has some disadvantages:

- Many GPS collars use the Global System for Mobile Communications (GSM), also used by cell phones, as a way to retrieve data from the collars. In order for GSM to work, the research must be conducted in an area that gets cellular phone service (Millspaugh and Marzluff, 2001).

- Being completely mobile devices, GPS tracking collars rely on battery power to function. The battery power of the GPS unit must support itself along with related electronic components which store data (Clark et al., 2006).

2.2 Study Area

The data were collected between April 2005 and May 2007 around the region of Satara rest camp (see figure below) in the central area of the Kruger National Park (KNP). The Kruger National Park is one of Africa's largest game reserves declared in 1898, with nearly two million

hectares of land that stretch for 352 kilometers from north to south along the Mozambique border (Gertenbach, 1983). It is the land of baobabs (*Adansonia digitate*), fever trees (Vachellia xanthophloea), knob thorns (*Acacia nigrescens*), marula (*Sclerocarya birrea*) and mopane trees (*Colophospermum mopane*) and the home of the Big Five : lions (*Panthera Leo*), leopards (*Panthera pardus*), rhinos (*Rhinocerotidae*), elephants (*Loxodonta Africana*) and buffalos (*Syncerus caffer*) along with the Little Five : buffalo weaver (*Dinemellia dinemelli*), elephant shrew (*Macroscelididae*), leopard tortoise (*Stigmochelys pardalis*), ant lion (*Myrmeleontidae)* and rhino beetle (*Dynastinae*) (Braack, 2006).

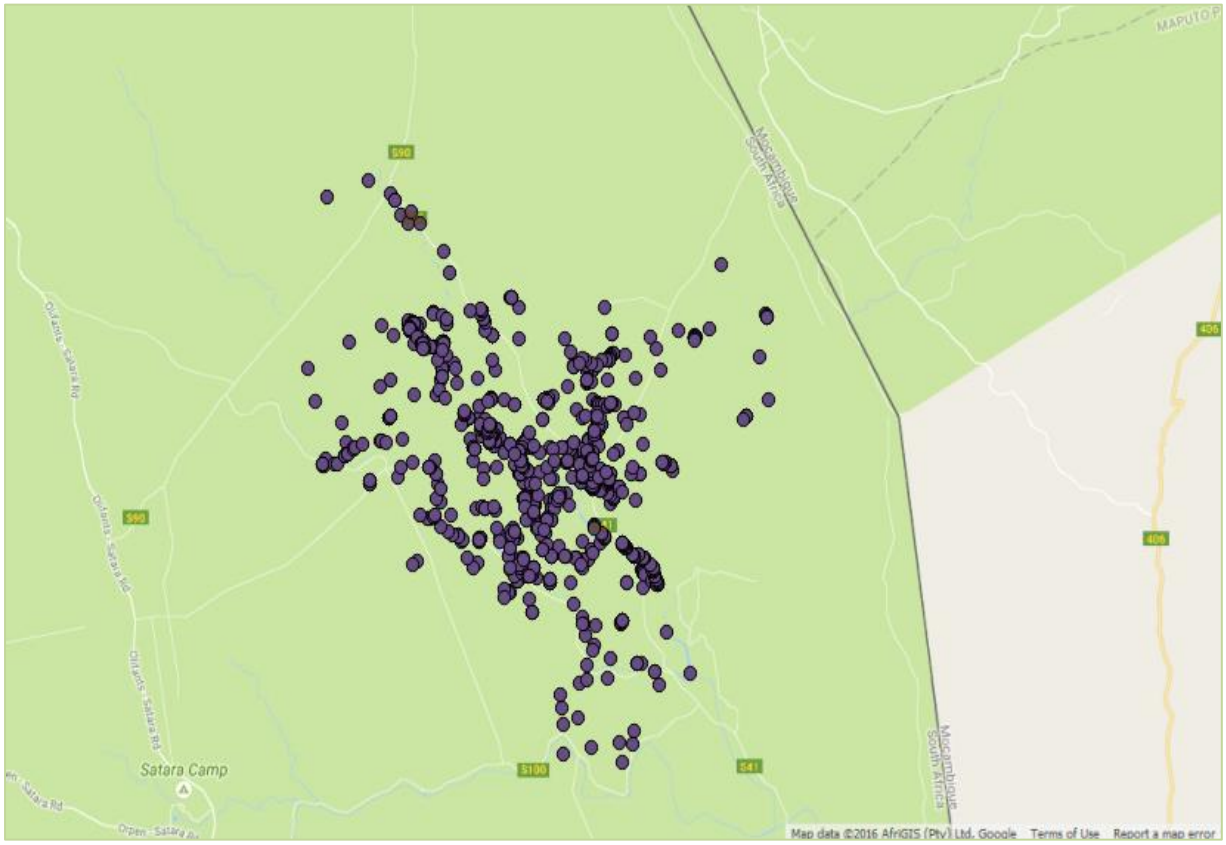


Figure 2.1 - Study area with Mavumbye Male AU007 locations in the central region of the Kruger National Park, South Africa.

The lion tracking data are used in this study to illustrate the application of the Hidden Markov Modelling approach. The study includes two types of data:

- GPS Tracking Data

Tracking data were collected using GPS collars placed on six lions and GPS coordinates were recorded once per hour for 24 hours per day or as the battery life diminished coordinates were

recorded once per hour between 18:00 and 06:00 hours and then again at 09:00, 12:00 and 15:00. Analysis of activity patterns were carried out on the hourly data (Tambling et al., 2010) covered in more details in the following chapter.

Table 2.1 - Details of GPS collars used in this study.

Lion	**Collar ID**	**Start Date**	**Last Date**	**Number of observations**
Thompson Female	AM186	24/01/2007	08/06/2007	1474
Nsenami Male	AM113/AM183	29/09/2005	21/01/2007	3070
Nsenami Female	AM39/AM115	29/09/2005	04/10/2006	4368
Mavumbye Male	AU007	23/08/2007	19/11/2007	1186
Matikitiki Male	AM139	22/11/2006	19/11/2007	900
First Female		20/04/2005	28/06/2005	966

- Observer-confirmed Kills

Using GPS location data, clusters where GPS positions showed that lions were stationary for more than two hours were identified as potential resting or kill sites. Those clusters were then investigated on foot and kill sites were identified by the presence of stomach contents and carcass remains such as bones, horns or teeth (Tambling et al., 2010). However, small prey items might have been entirely consumed by the lions making it impossible to identify kill sites (Tambling et al., 2012). Table 2.2 shows details of observer-confirmed kill data for the Thompson and Nsenami Female.

Table 2.2 - Details of clusters investigation.

Lion	Collar ID	Start Date	End Date	Number of Clusters investigated	Number of actual kills identified
Thompson Female	AM186	27/01/2007	01/05/2007	122	22
Nsenami Female	AM39/AM115	01/10/2005	06/09/2006	196	38

2.2.1 Study Species - Lion

Rightly called the 'king of the jungle', lions vary in color but typically sport light yellow-brown coats. Mature male lions are unique among big cats due to the thick brown and black manes that circle their necks and protect them while fighting (Stuart and Stuart, 2009). Lions tend to hunt mostly by night or in early mornings (Bygott, 1995). Previous research showed that they spend on average two to three hours a day socializing, grooming and defecating and are inactive for the rest of the time.

Fast Facts (males):

Height: 1.2m

Length: 1.5-2.4m

Weight: 150-227kg. In general, female lions are smaller than males.

Lifespan: 10-14 years

Top Speed: 50 mph (81 km/h) for short distances.

(Rudnai, 2012)

Figure 2.2 - Lion from Schotia Game Reserve (Photo - Jenicca Poongavanan)

Lions, in general, feed every two to three days and need on average between 5kg and 7kg of meat (Stuart and Stuart, 2009). Lions consume a wide variety of prey, from wildebeest, impala, zebra, giraffe, buffalo and warthogs to sometimes rhinos and hippos. They will also feed on smaller animals such as hares, birds and reptiles. Lions are also known to attack elephants when food is scarce (Nowak, 1999).

Chapter 3 – Literature Review

3.1 Model

There has been extensive research in the domain of animal movement ecology and its impact on the ecosystem (McGarigal et al., 2013). Depending on the type of data and the desired outcome in terms of ecological explanation, various analyses can be used (Goodall et al., 2014). This study investigated Hidden Markov Models (HMMs) as a predictive modelling technique. This literature review provides a summary of previous applications of the (HMM) and antecedent methods for locating kill sites.

Location coordinates are measured at discrete points in time, but the animals themselves follow a continuous path. How one estimates the continuous path from the discrete data depends upon how one models the animal's movement. Models of an animal's path in a plane may be either deterministic or stochastic.(Anderson-Sprecher and Ledolter, 1991)

3.1.1 Deterministic Model

The processes of this model are often described by differential equations, with a unique input leading to unique output for well-defined linear models and with multiple outputs possible for non-linear models (Nojima et al., 2013). If the deterministic model perfectly described the environmental system under consideration and there was no uncertainty or source of error, then given the value of the independent variable and the model parameters, one would be able to predict the value of the dependent variable exactly with no uncertainty (Ott, 1994). Deterministic models have been previously used to model the dynamics of an ecosystem (Érdi and Tóth, 1989). For example, climate change predictions come from very complex models that use deterministic but non-linear, physical relationships within the atmosphere and oceans (Evans, 2012). Although these models have been satisfactory in describing the main trends of ecological phenomena, deterministic models are not capable of reflecting the inherently random behaviors of the ecosystems (Érdi and Tóth, 1989). However, variability (stochasticity) plays a fundamental role in predator-prey population processes; it helps to explain observable population phenomena (Chesson, 1978).

3.1.2 Stochastic Models

A stochastic process is a random process evolving with time. More precisely, a stochastic process is a collection of random variables X_t indexed by time, T (Lawler, 2006).

Olivier Ibe (2013) states that a stochastic process $\{X(t), t \in T\}$ is called a first-order Markov process if for any $t_0 < t_1 < \cdots < t_n$, the conditional Cumulative Distribution Function (CDF) of $X(t_n)$ for given values of $X(t_0), X(t_1), \ldots, X(t_{n-1})$ depends only on $X(t_{n-1})$. That is,

$$P[X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}, X(t_{n-2}) = x_{n-2}, \ldots, X(t_0) = x_0]$$
$$= P[X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}]$$

This means that, given the present state of the process, the future state is independent of the past. This property is usually referred to as the *Markov property* (Ibe, 2013), hence the transition probability from one state to another depends solely on the current state (Goodall, 2014).

The common range of random variables, X_t's (the set of their possible values) is called the state space of the process. Markov processes are classified according to the nature of the time set T and the state space S (discrete or continuous) (Bosq and Nguyen, 2013). Thus, there are four basic types of Markov processes (Parzen, 1999):

Table 3.1 - Classification of Markov Processes (Ibe, 2013)

		State Space	
		Discrete	**Continuous**
Time	**Discrete**	Discrete-Time Markov Chains	Discrete-Time Markov Process
	Continuous	Continuous-Time Markov Chain	Continuous-Time Markov Process

Norris (1998) states that if X_0 is a vector which represents the initial state of a system, then there is a matrix M such that the state of the system after one iteration is given by the vector MX_0. Thus we get a chain of state vectors: $X_0, MX_0, M^2X_0, \ldots$ where the state of the system after n iterations is given by M^nX_0 ; such a chain is called a *Markov Chain* and the matrix M is called *Transition matrix* (Norris, 1998).

With huge progress made in tracking technology, there has been a substantial inflow of movement data leading to a demand for statistical tools that allow ecologists to analyze large amount of animal tracking data (Urbano et al., 2010). Thus the states of the Markov chain have been helpful in providing rough classifications of the behavioral dynamics of animals (Patterson et al., 2016)

The data used in this study consists of locations observed through time. These movements may be complex in terms of the underlying biology, but the behavior that we wish to estimate does not derive from the locations in space, it derives from changes in the way animals move (Jonsen et al., 2005). This involves, decomposing movement patterns into different behavioral states (Michelot et al., 2016). For the purpose of this study, a model that estimates unobservable states from these locations is required. Recently the State-Space Model (SSM) has gained popularity in movement ecology (Patterson et al., 2008).

State-Space Models have their origin in systems theory and famous early applications in the Apollo and Polaris aerospace programs (Hutchinson, 1984). SSMs are time-series models that allow unobserved states and biological parameters to be estimated from data observed with error (Jonsen et al., 2005). A modified version of the SSM with discrete hidden states known as the Hidden Markov Model (HMM), is used in this study and is explained in greater detail in the methodology section.

3.2 Background – Hidden Markov Models

HMMs are used in a variety of applications, they were first applied to speech recognition in the 1970s (Levinson et al., 1983). Applications of HMMs in Bioinformatics have been reported in Thompson (1983). Another area of application of HMMs is financial time series modelling, like the stock market (Rydén et al., 1998). Also, Le et al. (1992) have used HMMs to model time series of epileptic seizure counts. Similarly Leroux and Puterman (1992) applied HMMs to the pattern movement of a fetal lamb. HMMS have also been applied to hydro climatology by Zucchini and Guttorp (1991). HMMs have been used to model different communications environments (Chao and Yao, 1996) and; also used to model internet traffic by Costamagna et al. (2003). Hidden Markov Models (HMMs) have been used for the analysis of various animals; those are summarized in the next section.

3.2.1 Hidden Markov Models used in animal movement research.

Franke and al. (2004) have successfully used the HMM to analyze the movement of caribou *(Rangifer tarandus)* that reside year-round in central Alberta, Canada. The issue related to the study was to determine a set of interpretable states to the corresponding behaviours. They examined the possibility of three hidden states; bedding, feeding and relocating. They collected point locations every 15 minutes from 12 caribou using satellite tracking using Global Positioning System collars over a 10-day period. Distance between locations and turning angles were calculated between successive locations from Universal Transverse Mercator (UTM) coordinates. The distances were then clustered, resulting into four discrete observations: stationary, short, medium and long. The turning angles were also clustered in four classes: ahead, right, back and left.

The HMM allowed them to investigate the probabilities of remaining in or changing from a state to another and the times of the day associated with each of the states. Differences between the behaviour of individual animals were also apparent. They observed distinct behaviours based on the likelihood of them staying in their current state or changing state. Generally their results indicated that the caribou showed a propensity to forage for short periods, but would bed and relocate over long periods. They also found that HMMs constantly outperformed the auto-regressive time series model in terms of 'Percent correct' and 'Absolute Average Distance' between predicted and observed values. The authors did not include any covariates in the study. They found that it is apparent that the HMMs are capable of simulating the apparently different land-use strategies of each animal, which tends to support their interpretation of model parameters (Franke et al., 2004).

Another application of the HMM was in the analysis of Minke whales (*Balaenoptera acutorostrata*) by Christiansen et al (2011). In the study, they conducted individual follows of Minke whales from a land-based station and data on inter-breath intervals (IBI) were collected. IBI were estimated as the time elapsed between two consecutive surfacings. They found that two distinct states (dive types) were present within the surfacing data, which were defined as regular dives and deep dives. They assumed that the IBI to be the mixture of two normal distributions. From the transition probability matrix of the fitted model they were able to estimate the expected total length of stay in each state (dive type), from which the relative proportions that Minke whales spend in each dive type could be calculated. They found that Minke whales performed regular dives during 62% and deep dives during 38% of their time. They also investigated the effect of vessel presence (whale-watching boats) on Minke whales

dive types by including vessel presence as a covariate in the HMM and then comparing the relative proportions of the two dive types for the control and impact data. The authors found that the transition probability between dives types were significantly affected by the presence of whale-watching boats, which resulted in a change in relative proportions of the two dive types. They concluded that the HMM approach provided a way forward to understanding the diving behaviour of Minke whales and other species.

An application of the HMM in the fisheries field was the analysis of Atlantic cod (*Gadus morhua*) in the North Sea area by Pedersen et al. (2007). Data were collected using archival tags. Tags delivered highly accurate and detailed information of the immediate environment of the host animal. In this case, tags were programmed to record temperature and pressure every ten minutes. Data were retrieved from the tags after return through the commercial or recreational fisheries. In this study, the authors applied a direct Fokker-Planck-based methodology (FPM) using HMMs to the data. The model inferred two behavioural states, either high or low activity, estimated directly from the depth data. Their main aim was to obtain the most accurate reconstruction of the geographic movements of the cod. The movement was described in terms of an estimated probability distribution of the location of the fish and the Viterbi algorithm was used to predict the most likely route of the movement (Pedersen et al., 2008). The tidal patterns resulting from cod depth records were used in the calculations to reconstruct the migration pathway using a likelihood model.

3.2.2 Extensions of the HMM used in animal movement research

Kerk et al. (2015) studied the movement patterns of the Florida panther *(Puma concolor coryi)*, an endangered subspecies of puma that exhibits territorial behaviour. Their objective was to understand Florida panther movement patterns and factors influencing these patterns. Panthers were fitted with GPS collars and they made use of location data obtained on an hourly basis. The HMM assumes that the time spent in a specific state (dwell time) follows a geometric distribution (Langrock et al., 2012). In their study they used the Hidden Semi-Markov Model (HSMM), an extension of the HMM, that relaxes this assumption and permits explicit modelling of dwell times. Of 18 Florida panthers, data were sufficient for 13 panthers to permit HSMM modelling. They considered two to six underlying movement states for HSMM fitting, as well as alternative distributions for step length (log-normal, gamma and Weibull), turning angle (wrapped normal and wrapped Cauchy) and dwell times (Poisson and negative binomial). They determined the appropriate number of movement states for each panther

based on the Bayesian Information Criterion (BIC) values and biological interpretability of movement states.

The authors HSMM modelling revealed that distributions that consistently had the best fit were Weibull for step length, wrapped Cauchy for turning angle, and negative binomial for dwell time. Using these three distributions, they fitted a 3-state HSMM that were statistically well-supported and biologically interpretable for each panther. They identified the three states representing specific behavioural modes as: (1) a resting mode, characterized by short step lengths and turning angles around 180°; (2) a moderately active mode characterized by intermediate step lengths and variable turning angles and (3) a travelling mode characterized by long step lengths and turning angles around 0°. Moreover, for each panther the authors obtained state-specific estimates of parameters for the distributions used to model step lengths, turning angles and dwell times. They discovered that for most panthers the negative binomial shape parameters were close to 1, indicating that the dwell time distributions were close to geometric, and hence that a HMM would have sufficed for these individuals.

Next, the authors used the Viterbi algorithm to assign a specific state to every observed move segment of each panther and examined the distribution of these states during a diel period (a 24 hour period) and across different weeks of the year. They found that panthers were primarily in the resting mode during the day and in the travelling mode during the night. Their study also showed that females spend more time resting during mid-summer whereas males spent more time resting during the second half of the year. The estimated state occupancies also indicated that males spent the highest proportion of their time in the resting mode, whereas females divided their time more equally between travelling and resting modes. HSMM models and the Viterbi algorithm allowed the authors to resolve differences in movement patterns more precisely than was previously possible.

Another extension of the HMM, the mixed HMM was used in the analysis of red-cockaded woodpeckers (*Leuconotopicus borealis)* with the goal of identifying behavioural states by McKellar et al. (2015). A mixed HMM considers both covariates and random effects (MacKay, 2003). The inclusion of random effects offers an elegant and plausible way for modelling ‘personality’ (Schliehe-Diecks et al., 2012). They performed 440 observation sessions at 97 territories, where observers outfitted with a GPS would follow the birds and the geographic coordinates of their location were automatically recorded once each minute. In total their data

set consisted of 35 304 locations and included five covariates (group size and four environmental factors) in their model. They also included seasonal effects to account for differences in the step lengths distributions across seasons and random effects for accommodating heterogeneity across birds. They fitted a 2-state model to the data, where turning angles were modelled using a von Mises distribution and step lengths using a gamma distribution. Further, they fitted all nested models that included a subset of the covariates and used the Akaike information criterion (AIC) to select among the competing models. They found that a model including seasonality and covariates, had substantially more support than the simpler nested models. The authors identified two distinct behavioural states of woodpecker behaviour, termed resting and foraging. They also identified different movement patterns across seasons. The inclusion the covariates was found to substantially increase the likelihood of the fitted model. Their study showed that HMMs can be useful for describing the influence of multiple covariates on movement while accounting for heterogeneity in a social bird, an approach that can potentially be applied to a large range of species.

Mixed Hidden Markov Models are also used to model the feeding behaviour of 57 sub adult grey mouse lemurs (*Microcebus murinus*), small nocturnal primates endemic to Madagascar (Schliehe-Diecks et al., 2012). The authors collected feeding data of focal animals cumulatively for observation intervals of one minute. About 500 hours of focal observations were included in the analyses. They fitted a 2-state model that involved different motivational states, 'satiated' and 'hungry'. They considered a model that incorporated random effects and covariates (sex, body mass and time of night). For each of the covariates they conducted a likelihood ratio test of the simplified model against the full model. They also compared a model with no random effect and found that the AIC value selected the model that included random effects. Considering the transition probabilities in relation to the covariates, they found that mouse lemurs are more likely to switch between the two states at the beginning of the night than towards the end. Also the behaviour of the female mouse lemurs is more persistent, reflected by longer feeding and non-feeding periods. Moreover, they found that mouse lemurs with a high body mass stay longer satiated and exhibited shorter hungry periods. They concluded that HMMs and in particular those that incorporated covariates, can facilitate detecting differences in behaviour which are not directly obvious from the observations. Also by using a mixed HMM, they were able to identify how behavioural flexibility and personality differences interact and lead to differences in behavioural sequences.

3.2.3 Statistical Methods used to Identify kill sites

Various studies showed that HMMs can provide substantial insight into the movement behaviour of animals (Goodall, 2014). Following on from the caribou study, Franke et al. (2006) used HMMs to predict the locations of wolf kill sites from their movement patterns. In this study, they made use of an existing data set that utilized GPS technology to provide wolf movement data and twice daily aerial surveys were conducted to locate actual wolf kill sites. When a carcass was not apparent, wolf trails were backtracked until either a carcass was discovered or the search was abandoned. They used the actual kill locations to construct the 'known' state; in this case the states are no longer 'hidden'. While constructing the 'known' state, they identified four situations:

1. At kill - where wolves were stationary and close to a carcass
2. Bedded Away - where wolves were stationary and relatively far from a carcass
3. Relocating - when successive wolf locations where relatively far apart and
4. Missed Kill - where wolves were stationary and very far from a carcass.

This resulted in a state sequence describing the actual location of wolves relative to known carcass locations. They then developed a 3-state HMM for each pack of wolves to examine the possibility that at least one state corresponded to the behaviour associated with kill sites. All models were trained using distance between location, turning angle and travel rate per hour obtained from GPS coordinates. Information with regard to known carcass locations were not used to train the models (Franke et al., 2006). Distances between locations were classified into six classes based on distance moved. Where locations were not acquired; they divided the total distance moved by the number of time periods elapsed. The turning angles between successive movements were clustered into four categories depending on the direction of the move. Travel rates were stratified into two groups based on whether the travel rate was greater or less than the median wolf travel rate calculated in a different study. These distances, turning angle and travel rate categorical variables were then used as input for the HMM.

Applying commonly used measures such as the 'Percent Correct' and the 'Average Absolute Difference', the authors determined the degree of association between observation sequences predicted by the HMM and the training data. They determined the most likely behavioural 'state' sequence given the model and an observation sequence using the Viterbi procedure. They assumed that state 1 was consistent with bedding, state 2 with localized activity and state 3 with relocating behaviour. State 1 and 2 were collapsed into a single new state called 'Kill', state 3 remained on its own and was recoded 'No Kill'. The authors then compared the

‘known’ state sequence derived from the GPS data in conjunction with the carcass data to the HMM predicted sequence derived from the GPS wolf location data.

Models were able to accurately predict kill sites, with 22 out of 23 observed kill sites. There were 21 instances observed where packs were not associated with a carcass, their models agreed on 16 occasions; of the remaining five occasions, the models indicated that a kill existed in spite of the fact that none were recorded during aerial back-tracking. On three of these so-called ‘false positive’, models predicted the presence of a kill sites in the period between flights. They concluded that such observations were consistent with the consumption of smaller prey species, which were handled very quickly and were almost entirely consumed. However they also established that it was possible that these instances were actually sleeping bouts and no kill was present.

Webb et al. (2008) investigated a method that is effective for small ungulate prey species and minimizes field visits to potential kill sites. They presented a new approach that used a space-time clustering algorithm to identify clusters of wolf GPS locations and used multinomial logistic regression to model the probability of a cluster being a non-kill site, kill site of small or large bodied prey species based on wolf movement characteristics. They also investigated whether site-specific environmental covariates could improve predictive success. They assessed the accuracy of predicting kill sites using an independent set of GPS movement data and associated identified clusters. The technique successfully identified all kill sites of large bodied prey species, but 17% of kills of small bodied prey were not identified. In the end, the models were more successful at distinguishing kill sites of large bodied from non-kill sites than from small bodied prey species, which is consistent with results from Franke et al. (2006).

Lions (*Panthera Leo*) are of particular interest in predator-prey studies. Previous studies on methods for locating African lion kills using GPS movement data were carried out by Tambling et al. (2010). Their main aim was to develop and test models that increased the efficiency of locating lion kills from remotely accessed GPS data. They assessed performance of Generalized Linear Models (GLMs) versus Classification Trees (CT). During May 2005 and April 2007, 9 collars were deployed; four of which attempted fixes for 24 hours/day (hourly) and five of which attempted fixes for 16 hours/day (hourly). The combination of data sets using both collar schedules incorporated balanced sets of hourly readings at night (when lions are most active) across the entire data set but missed some data points due to GPS error during the day (Tambling et al., 2010).

From GPS coordinates they calculated distances between successive locations and missed GPS coordinates were treated as stationary locations. They then defined a GPS aggregation cluster where distance between successive locations was less than 100m for more than two consecutive recorded fixes. If a location was not recorded and distance to the following location was greater than 100m, a GPS cluster was not created. Clusters were then investigated on foot and they identified predation events at clusters from the presence of prey stomach contents, bones or horns and when possible they tried to identify prey species, age and sex (Tambling et al., 2010).

Table 3.2 - Shows the eight predictor variables they measured for each GPS clusters.

Variables	**Description**
Hours	Length of time lions spent at a cluster from the first point of the cluster until they left the cluster for the last time.
Return	A categorical variable describing a return visit to a cluster within 12 movement steps of leaving that cluster.
Ratio_24	Ratio of distance moved during the 24 hours prior to the cluster beginning against distance moved during the 24 hours following termination of the cluster.
Dist_1	Distance moved by lions during the first GPS interval of a cluster.
Dist_2	Distance moved by lions during the second GPS interval of a cluster.
Drain	Distance from the cluster to the nearest drainage line.
Road	Distance from the cluster to the nearest road.
Dark	A 5-valued categorical variable that accounted for the combined effect of the sun and the moon at the start of the cluster.

They used Generalized Linear Models (GLMs) to investigate variables with the probability of a binary response (kill = 1, no kill = 0) occurring at a GPS cluster. Due to possible time delays between cluster occurrence and cluster investigation that suggested declines in the ability to identify kills at cluster sites, models were developed based on data from clusters investigated during the first 4 weeks and first 16 weeks following a cluster occurrence. They used a stepwise $\propto$-to-enter approach for entering parameters into the model and parameters were also tested for collinearity. Models were assessed using the area under the curve (AUC) based on the receiver operator characteristic (ROC) curve, which is a plot of the proportion of true positives as a function of the proportion of false positives (Fawcett, 2006).

Classification Trees (CTs), which use a hierarchical splitting criterion that separates binary responses to predictor variables based on a set of rules were also used by the authors. The tree is constructed by repeatedly splitting the data, defined by a simple rule based on a single explanatory variable (De'ath and Fabricius, 2000). At each split the data are partitioned into two mutually exclusive groups, each of which is as homogeneous as possible and the splitting procedure is then applied to each group separately (Breiman et al., 1984). The objective is to partitione the response into homogenous groups, but also to keep the tree reasonably small (De'ath and Fabricius, 2000). In this study, data partitioning was restricted to a 2-fold *k* partitioning.

For kill site prediction they developed GLMs and CTs using cumulative monthly cluster data between March 2006 and April 2007 (training data set) and predicted the location of kill sites for clusters which were investigated during the following month (testing data set). For example, 'they used GLMs and CTs constructed using cluster data between March 2006 and September 2006 to predict the state of clusters investigated during October 2006' (Tambling et al., 2010). Finally, they used GLMs to evaluate the effectiveness of using the highest probability of finding a kill in the testing data set to order cluster investigation in the field. They then averaged the cumulative number of kills located with each additional cluster investigated and the cumulative number of kills found by searching clusters at random across the whole data set.

Out of 2 433 clusters they investigated 1 447 of them and found 234 kills. They investigated more clusters and found more kills during the first four weeks following cluster occurrence. They used standard GLMs for the development of predictive models. For both the 4- and 16-week data sets, four predictor variables were included in the final model: *Ratio_24, Hours, Dark* and *Dist_1*. Ratio of distance moved 24 hours before versus 24 hours after a cluster was the most important predictor variable in both GLM and CT analysis. The GLMs showed good discrimination with Akaike Information Criterion (AIC) values between 0.81 and 0.83. Area under the curve results for the CTs were lower than those for the GLMs. In all cases, GLMs outperformed CTs in their discriminative ability to separate kill from non-kill sites.

As a result, the objective of this study is to identify kill sites through characteristics of movement behaviours of lions by applying the Hidden Markov models.

Chapter 4 – Hidden Markov Model

Some authors in fact do not distinguish between Hidden Markov models (HMMs) and State (SSMs) (Cappé et al., 2009). However, the label HMM is usually used to indicate a model with a finite number of possible states, whereas in SSMs, the underlying state process typically involves an infinite number of states (McKellar et al., 2015). Nevertheless, in the literature on movement modelling, SSM approaches typically considers finite number of states (McKellar et al., 2015). In recent years Hidden Markov Models have gained popularity in animal movement studies as they provide a reasonably simple theoretical structure and the conclusions drawn from the model are easily adapted to interpret the results in terms of the behavior of an animal (Goodall, 2014). The HMM approach to movement modelling is part of the wider family of state switching models, which focus on the decomposition of the movement process into distinct underlying states. Such an approach conforms with the fact that an animal's movement is driven by switches in underlying behavior (Michelot et al., 2016). The standard HMM approach to model an individual animal's movement considers a bivariate time series comprising the step length and the turning angle at each time point (Michelot et al., 2016). Resulting observation sequences are then used as model input (Franke et al., 2004).

4.1 Methodology

This chapter describes the methodology used for the analysis in this project. For the rest of this chapter, the notation of Zucchini and MacDonald (2009) will be used.

4.1.1 Basic HMM

A Hidden Markov model is a doubly stochastic process in which an underlying stochastic process that is not observable can only be observed through another stochastic process that produces a sequence of observations called the 'state-dependent process' (Ibe, 2013, Zucchini and MacDonald, 2009). The Markov property states that when C_t is known, the distribution of X_t depends only on the current state C_t and not on previous states C_{t-1} (Zucchini and MacDonald, 2009).This structure is depicted by the graph below:

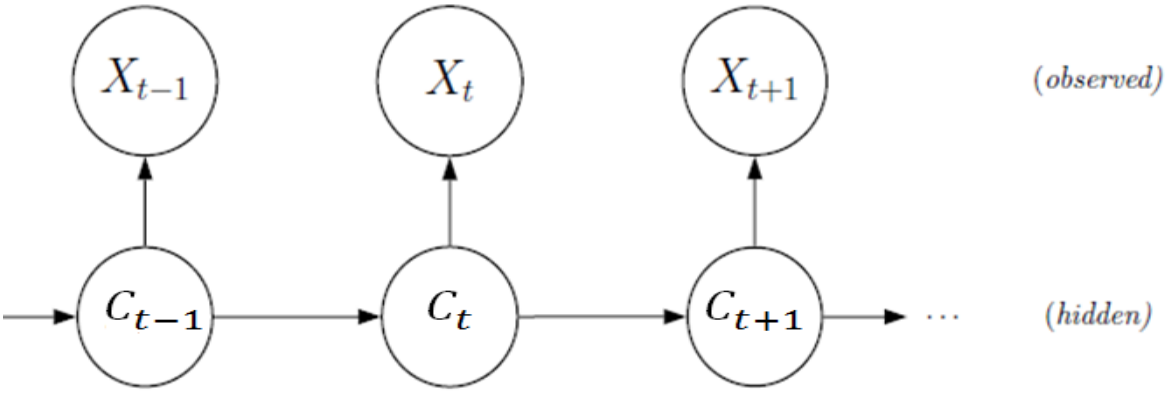


Figure 4.1 - Directed graph of basic HMM (Langrock et al., 2012)

A Markov chain is completely defined by its one-step transition probability matrix (Karlin, 2014). The one-step transition probability is the probability of transitioning from one state to another in a single step. The Markov chain is said to be time homogeneous if the transition probabilities from one state to another are independent of the time index (Nasseri, 2008).

Ibe (2013) display the state-transition probabilities as the entries of a $m \times m$ matrix $\mathbf{\Gamma}$.

$$\Gamma = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{bmatrix},$$

$\mathbf{\Gamma}$ is called the transition probability matrix, it is a stochastic matrix because the sum of entries in each row is equal to 1 (Ibe, 2013). For example, γ_{11} represents the probability of being in state 1 and remaining in state 1 and γ_{13} represents the probability of moving from state 1 to state 3. A Markov chain with transition probability matrix $\mathbf{\Gamma}$ is said to have stationary distribution $\boldsymbol{\delta}$ if $\boldsymbol{\delta\Gamma} = \boldsymbol{\delta}$ and $\boldsymbol{\delta 1}' = 1$. The first of these conditions expresses the stationarity and the second is the requirement that $\boldsymbol{\delta}$ is a probability distribution (Zucchini and MacDonald, 2009).

With $\boldsymbol{X}^{(t)}$ and $\boldsymbol{C}^{(t)}$ representing the histories from time 1 to time *t*, Zucchini and MacDonald (2009) represent the simplest HMM as:

$$\Pr(C_t | \boldsymbol{C}^{(t-1)}) = \Pr(C_t | \ C_{t-1}), t = 2,3, \ldots$$
$$\Pr(X_t | \boldsymbol{X}^{(t-1)}, \boldsymbol{C}^{(t)}) = \Pr(X_t | \ C_t), t \in \mathbb{N}.$$

Such a Hidden Markov model can be characterized by the distribution of $\{C_t\}$, the transition probability matrix of the Markov chain ($\boldsymbol{\Gamma}$) and the state dependent distributions defined by $p_i(x) = \Pr(X_t = x \mid C_t = i)$; which applies to both discrete- and continuous-valued observations (Goodall, 2014). p_i is defined as the *probability mass function* of X_t if the Markov chain is in state i at time t for the discrete case and the *probability density function* in the continuous case (Zucchini and MacDonald, 2009).

For this study, the distribution of the Markov chain is the same at any time t, thus only HMMs with stationary Markov chains are considered.

4.1.2 Moments Derivation

The expected value and variance of an m-state stationary HMM $\{X_t : t = 1,2, \ldots\}$ with transitional probability matrix $\boldsymbol{\Gamma}$, stationary distribution $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_m)$, $u_i(t) = \Pr(C_t = i)$ and state dependent distribution $p_i(x)$, then it follows (Goodall, 2014, Zucchini and MacDonald, 2009)

$$\mathrm{E}(X_t) = \sum_{i=1}^{m} \mathrm{E}(X_t \mid C_t = i) \Pr(C_t = i)$$
$$= \sum_{i=1}^{m} \delta_i \, \mathrm{E}(X_t | C_t = i)$$

Since the Markov chain is stationary and assuming that the mean and variance of the distribution p_i is given by μ_i and σ_i^2 respectively, then

$$\mathrm{E}(X_t) = \sum_{i=1}^{m} \delta_i \, \mathrm{E}(X_t | C_t = i)$$
$$= \sum_{i=1}^{m} \delta_i \, \mu_i$$
$$= \boldsymbol{\delta}\boldsymbol{\mu}'$$

Let Y_i denote the random variable with probability function $p_i(x)$. This implies $\mathrm{E}(Y_i) = \mu_i$ and $Var(Y_i) = \sigma_i^2$ (Goodall, 2014).

$$Var(Y_i) = \mathrm{E}((Y_i - \mu_i)^2) = \sigma_i^2$$
$$\sigma_i^2 = E(Y_i^2) - \mu_i^2$$
$$\mathrm{E}(Y_i^2) = \sigma_i^2 + \mu_i^2$$
$$\mathrm{E}(X_t^k) = \sum_{i=1}^{m} \delta_i \, \mathrm{E}(Y_i^k)$$
$$\mathrm{E}(X_t^2) = \sum_{i=1}^{m} \delta_i \, \mathrm{E}(Y_i^2)$$
$$= \sum_{i=1}^{m} \delta_i \left(\sigma_i^2 + \mu_i^2\right)$$
$$Var(X_t) = \sum_{i=1}^{m} \delta_i \left(\sigma_i^2 + \mu_i^2\right) - (\sum_{i=1}^{m} \delta_i \mu_i)^2$$
$$= \sum_{i=1}^{m} \delta_i \left(\sigma_i^2 + \mu_i^2\right) - (\boldsymbol{\delta\mu}')^2$$

4.1.3 Parameter Estimation

The estimation of the parameters of an HMM is usually performed by a maximum likelihood algorithm (Zucchini and MacDonald, 2009). In order to determine the parameter estimates for the m-state HMM, the likelihood function can be maximized (Goodall, 2014). Given an m-state HMM, where $\boldsymbol{\delta}$ is the initial distribution, transition probability matrix $\boldsymbol{\Gamma}$ and $\mathbf{P}(\mathrm{x})$ the $m \times m$ diagonal matrix with i^{th} diagonal element consisting of the state-dependent probability density $p_i(x)$, the likelihood function for a non-stationary HMM is given by (Goodall, 2014, Zucchini and MacDonald, 2009) :

$$L_t = \boldsymbol{\delta}\mathbf{P}(x_1)\boldsymbol{\Gamma}\mathbf{P}(x_2) \ldots \boldsymbol{\Gamma}\mathbf{P}(x_t)\mathbf{1}',$$

Where $\boldsymbol{\delta}$ and $\mathbf{1}'$ are row vectors. Goodall (2014) and Zucchini & MacDonald (2009) state that in the case where $\boldsymbol{\delta} = \boldsymbol{\delta\Gamma}$, the distribution of the Markov chain is stationary, then the likelihood function is defined as:

$$L_t = \boldsymbol{\delta}\mathbf{P}(x_1)\boldsymbol{\Gamma}\mathbf{P}(x_2) \ldots \boldsymbol{\Gamma}\mathbf{P}(x_t)\mathbf{1}'$$

A numerical maximization of the likelihood becomes feasible even for large T, as the computational cost of evaluating the likelihood function is linear in the number of observations, T (Patterson et al., 2016). Several problems arise when the likelihood is computed and maximized numerically such as numerical underflow, constraints on the parameters and multiple local maxima (Zucchini and MacDonald, 2009).

An alternative is given by the expectation-maximization (EM) algorithm, which uses maximum likelihood estimates derived from the forward-backward algorithm to estimate the unknown model parameters (Goodall, 2014). In the case of HMMs the EM algorithm is known as the Baum-Welch algorithm, designed to estimate the parameters of an HMM whose Markov chain is homogeneous but not necessarily stationary (Zucchini and MacDonald, 2009). But MacDonald (2014) states that there is no evident reasons to prefer it over direct numerical maximization of the likelihood.

4.1.4 Viterbi Algorithm and State Decoding

To analyze the process of behavioural state-switching, the HMM toolbox offers a variety of additional inferential techniques (Patterson et al., 2008). Given the HMM and the observations, information can be deduced about the states occupied by the underlying Markov chain, such an inference is known as *decoding* (Zucchini and MacDonald, 2009). There are two types of decoding which can be used to allocate each observation to a state (Goodall, 2014).

'Local decoding' of a state at time t refers to the determination of that state which is most likely at that time (Zucchini and MacDonald, 2009). The local decoding is essentially the conditional probability of being in state i at time t, given all observations (Goodall, 2014).

The distribution of state C_t given the observations $x^{(T)}$ can be found, using the forward and backward probabilities (Goodall, 2014). Hence Zucchini and MacDonald (2009) describe the conditional distribution of C_t given the observations can be obtained, for $i = 1,2, \dots, m$, as

$$\Pr(C_t = i \mid \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}) = \frac{\Pr(C_t = i, \boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)})}{\Pr(\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)})}$$
$$= \frac{\alpha_t(i)\beta_t(i)}{L_T}$$

The forward and backward probabilities are represented by $\alpha_t(i)$ and $\beta_t(i)$ respectively and L_T the likelihood. Further, for $t = 1,2,\dots,T$, the authors define the forward probabilities with $\boldsymbol{\delta}$ denoting the initial distribution of the Markov chain as:

$$\boldsymbol{\alpha}_t = \boldsymbol{\delta}\,\mathbf{P}(x_1)\,\boldsymbol{\Gamma}\,\mathbf{P}(x_2)\dots\boldsymbol{\Gamma}\,\mathbf{P}(x_t) = \boldsymbol{\delta}\,\mathbf{P}(x_1)\prod_{s=2}^{t}\boldsymbol{\Gamma}\,\mathbf{P}(x_s)\,,$$

For $= 1,2,\dots,T \;\; and \;\; j = 1,2,\dots,m$,

$$\boldsymbol{\alpha_t}(j) = \Pr(\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}, C_t = j)$$

The backward probabilities $\boldsymbol{\beta_t}$ which, for $t = 1,2,\dots,T$, is defined by

$$\boldsymbol{\beta}_t' = \boldsymbol{\Gamma}\mathbf{P}(x_{t+1})\boldsymbol{\Gamma}\mathbf{P}(x_{t+2})\dots\boldsymbol{\Gamma}\mathbf{P}(x_T)\mathbf{1}' = \left(\prod_{s=t+1}^{T}\boldsymbol{\Gamma}\mathbf{P}(x_s)\right)\mathbf{1}'\,,$$

For $t = 1,2,\dots,T-1 \; and \; i = 1,2,\dots,m$,

$$\beta_t(i) = \Pr(X_{t+1} = x_{t+1}, X_{t+2} = x_{t+2}, \dots, X_T = x_t \mid C_t = i)\,,$$

provided that $\Pr(C_t = i) > 0$. In a more compact notation:

$$\beta_t(i) = \Pr(\boldsymbol{X}_{t+1}^{T} = \boldsymbol{x}_{t+1}^{T} \mid C_t = i)\,,$$

Where $\boldsymbol{X}_a^b$ denotes the vector $(X_a, X_{a+1}, \dots, X_b)$.

The above equations show that the forward probabilities are joint probabilities, while the backward probabilities are conditional probabilities of the observations $x_{t+1}, x_{t+2}, \dots, x_T$, given that Markov chain is in state i at time t (Goodall, 2014).

'Global decoding' refers to the determination of the most likely sequence of states (Zucchini and MacDonald, 2009). It is the conditional probability of the sequence of states given the sequence of observations (Goodall, 2014). Instead of maximizing the conditional probabilities of the observations given the state, global decoding seeks that sequence of states $c_1, c_2, \dots, c_T$ which maximizes the conditional probability $\Pr\left(\boldsymbol{C}^{(T)} = \boldsymbol{c}^{(T)} \mid X^{(T)} = \boldsymbol{x}^{(T)}\right)$; or equivalently, the joint probability: $\Pr\left(\boldsymbol{C}^{(T)}, \boldsymbol{X}^{(T)}\right) = \delta_{c_1} \prod_{t=2}^{T} \gamma_{c_{t-1},c_t} \prod_{t=1}^{T} p_{c_t}(x_t)$ (Zucchini and MacDonald, 2009).

In order to determine the most likely sequence of states without maximizing over all possible sequences of states, the Viterbi algorithm, a dynamic programming algorithm which applies the concept of global decoding is used (Goodall, 2014).

Following the notation of Zucchini and MacDonald (2009) the Viterbi algorithm is defined by:

$$\xi_{1i} = \Pr(C_1 = i, X_1 = x_1) = \delta_i p_i(x_1),$$

And, for $t = 2,3, \dots, T$,

$$\xi_{ti} = \max_{c_1, c_2, \dots, c_{t-1}} \Pr(\boldsymbol{C}^{(t-1)} = \boldsymbol{c}^{(t-1)}, C_t = i, \boldsymbol{X}^{(t)} = x^{(t)})$$

The probabilities ξ_{tj} satisfy the recursion below for $t = 2,3, \dots, T \; and \; j = 1,2, \dots, m$:

$$\xi_{tj} = (\max_i(\xi_{t-1,i}\gamma_{ij}))\, p_j(x_t)$$

It thus provides an efficient means to calculate the $T \times m$ matrix of values ξ_{tj} and the required number of calculations is linear in T. The optimal sequence of states $i_1, i_2, \dots, i_T$ can then be determined recursively from

$$i_T = \underset{i=1,\dots,m}{\operatorname{argmax}}\, \xi_{Ti}$$

And, for $t = T-1, T-2, \dots, 1$, from

$$i_T = \underset{i=1,\dots,m}{\operatorname{argmax}}(\xi_{Ti}\, \gamma_{i,i_{t+1}})$$

The Viterbi algorithm is relevant to both stationary and nonstationary underlying Markov chains; there is no necessity to assume that the initial distribution $\boldsymbol{\delta}$ is the stationary distribution (Zucchini and MacDonald, 2009). The results from local and global decoding will usually be similar although not exactly the same, as in both instances, the allocated state is the state or sequence of state that maximizes the conditional probabilities (Goodall, 2014).

4.1.5 Model Selection and Model Checking

Model selection for HMMs is necessary to identify the number of states required in the model (Goodall, 2014). Zucchini and MacDonald (2009) stipulate that increasing the number of states always improves the fit of the model as judged by the likelihood. But in the case of HMMs, Goodall (2014) specifies that even more parameters are estimated for each increase in the number of states due to additional parameters in the transition probability matrix. The author also states that for a stationary HMM, the stationary distribution parameters are not estimated as these are determined directly from the transition probability matrix.

Two information criteria were considered when selecting the number of states to include in the models. Criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used. Zucchini and MacDonald (2009) defines these as follows:

$$AIC = -2\log L + 2p$$
$$BIC = -2\log L + p\log T$$

where $\log L$ is the log-likelihood of the fitted model and p denotes the number of parameters of the model. The first term is a measure of fit, which decreases as the number of states increases. The second term is a penalty term and increases as the number of states increases and T is the number of observations. Compared to AIC, the penalty term of BIC has more weight and thus favors models with fewer parameters (Zucchini and MacDonald, 2009).

These criteria provide a decision rule for selecting the best of several fitted models; however they do not guarantee that the selected model is indeed appropriate. Therefore, one still has to assess the fit of the chosen model. Zucchini and MacDonald (2009) found it useful to compare the autocorrelation functions (ACF) of the fitted HMMs with the ACF of the observed time series. The ACF of the model and the sample should correspond closely, particularly for the shorter lags, which would indicate that the fitted model had a similar dependence structure between successive observations as was present for the observed data (Goodall, 2014).

In general, residuals are a popular tool for assessing the fit of the model, where the expected observations are compared with the actual observations (Goodall, 2014). Under ideal circumstances, residuals are independently and identically distributed but this is not the case when dealing with HMMs which renders the residual analysis slightly more complicated (Bulla et al., 2015). One solution to this problem is the use of so-called pseudo-residuals, which are

based on the concept of p-values (Bulla et al., 2015). Zucchini and MacDonald (2009) describe two versions of pseudo-residuals, one based on the uniform distribution and the other on the normal distribution. The pseudo-residuals can be used for checking the goodness-of-fit of the model as well as for detecting outlier observations (Goodall, 2014). Both versions of pseudo-residuals contain information on how extreme and how rare the observations are (Bulla et al., 2015).

4.1.6 Missing Observation Data

When analyzing animal movement data, missing observations are a serious challenge. There are two main causes for missing data; missed GPS locations and changes in the frequency of observations. In the case of hidden Markov time series models, however, the adjustments needed to calculate the likelihood if data are missing turns out to be a simple one (Zucchini and MacDonald, 2009). In order to maximize the likelihood using direct numerical maximization, the computation of the observed-data likelihood can be modified so that the likelihood can be maximized as normal using an optimizer function such as nlm in R (Goodall, 2014). To compute the likelihood, one has to replace the respective diagonal matrices containing the state-dependent probabilities of the missing observations by a unit matrix (Bulla et al., 2015). Following the notation of Zucchini and MacDonald (2009), the likelihood can be written as, assuming observation 3, 5 and 6 are missing:

$$L_T^{-(3,5,6)} = \boldsymbol{\delta}\mathbf{P}(x_1)\boldsymbol{\Gamma}\mathbf{P}(x_2)\boldsymbol{\Gamma}^2\mathbf{P}(x_4)\boldsymbol{\Gamma}^3\mathbf{P}(x_7)\ldots\boldsymbol{\Gamma}\mathbf{P}(x_T)\mathbf{1}'$$

The fact that, even in the case of missing observations, the likelihood of an HMM can be easily computed is essential in the derivation of conditional distributions (Zucchini and MacDonald, 2009).

Chapter 5 – Hidden Markov Model for Animal Movement

5.1 Analysis

The HMM approach utilizes inputs such as step lengths and turning angle usually from either easting-northing coordinates or longitude-latitude values (Michelot et al., 2016). In this study, only step lengths were used as input to the HMM. Missing observations were included in the dataset in order to construct a complete time series dataset and coded as such to denote the occasions when a location was not obtained. Some of the lions' collars were replaced due to battery life of the collars. The data for these lions from before and after replacement are merged to form one dataset per lion.

The analysis was carried out using hourly data. In some cases due to battery life, GPS coordinates were recorded once per hour between 18:00 and 06:00 and then again at 09:00, 12:00 and 15:00 hours. In order to not disrupt the time series a movement rate was used. A movement rate was calculated by dividing distance by time travelled between two successive location points.

Two-, three- and four-state HMMs were fitted to the lion data described in section 2.2. Log-normal models were selected as the best performing distribution and therefore chosen as the underlying distribution for the HMM analysis. The four-state model was the most complex model used as a more complex model becomes computationally more intensive to fit and difficult to interpret biologically. Figure 5.1 illustrates the distinct displacement pattern observed from the Thompson lioness where there is a large number of short movements and fewer longer movements.

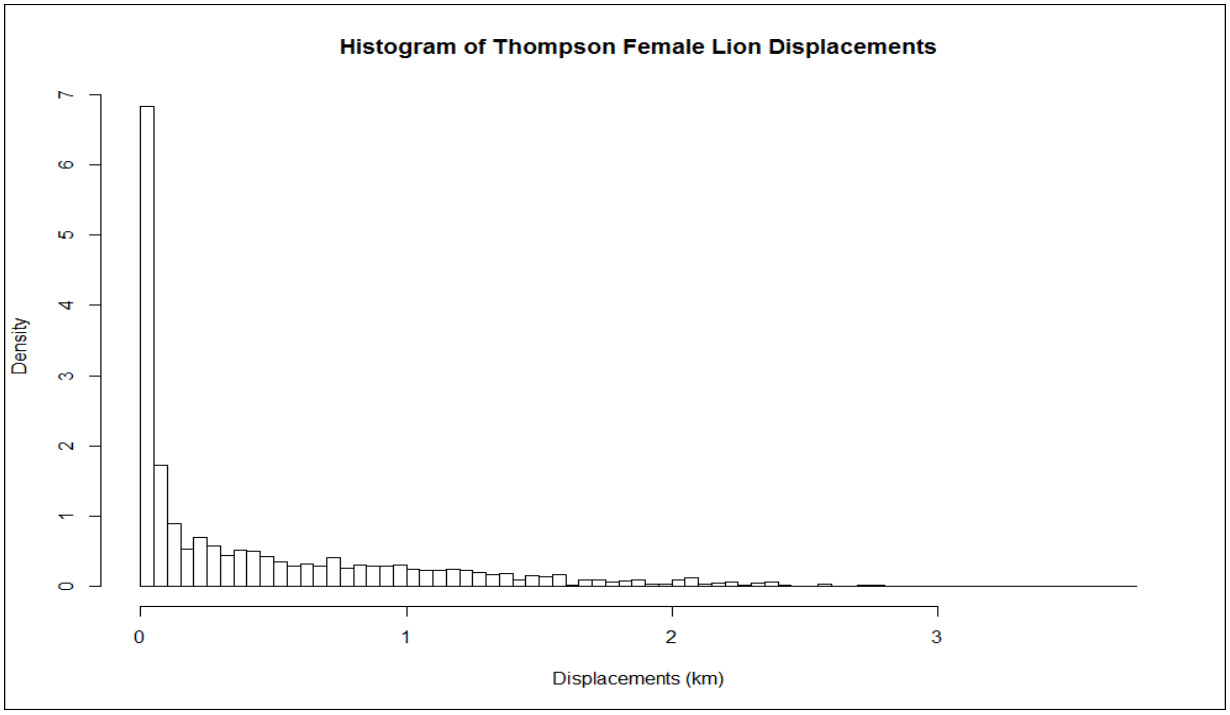


Figure 5.1 - Histogram of Thompson Female lion AM186 hourly displacement.

The parameter estimates were obtained using the R statistical software (R Development Core Team, 2010) and the moveHMM package (Michelot et al., 2016). The AIC was used to determine which HMM provided the best fit to the data.

5.2 Results

Two-, three- and four-state log-normal HMMs were fitted and the results are shown in Table 5.1. The AIC values selected a four-state model as the model that "best" describes the movement of all lions, except for 'First Female', whose AIC value indicated a simpler three-state model as the most appropriate.

Four-state models will be considered for these analysis purposes as they were predominately selected by the AIC and make the most ecological sense regarding the lion movement pattern and behaviour with the exception of the First Female where a three-state model was selected as suggested by the AIC. From the biological perspective, four-states were possible to interpret in terms of known behaviours of these animals. The four possible states corresponded to a 'resting/eating', 'local movement', 'active movement' and 'relocation'.

Table 5.1 - HMM outputs: Log-Likelihood and AIC

Name	Collar ID	Number of observations	Model	Maximum Log-Lik	AIC
First Female	-	966	2-state	1323.263	-2632.57
			3-state	1343.415	-2658.83
			4-state	1350.172	-2654.343
Nsenami Female	AM39/AM115	4368	2-state	5626.463	-11238.93
			3-state	5792.462	-11556.92
			4-state	5834.587	-11623.17
Thompson Female	AM186	1474	2-state	187.6806	-361.3611
			3-state	233.7085	-439.4171
			4-state	251.6422	-457.2844
Matikitiki	AM139	900	2-state	1179.667	-2345.33
			3-state	1205.007	-2382.014
			4-state	1234.75	-2423.588
Mavumbye	AU007	1186	2-state	1748.116	-3482.231
			3-state	1806.77	-3585.54
			4-state	1818.051	-3590.102
Nsenami Male	AM113/AM183	3070	2-state	1204.283	-2394.566
			3-state	1324	-2620
			4-state	1369.274	-2692.549

From Table 5.2 based on the stationary distribution $\boldsymbol{\delta}$, the Nsenami and Mavumbye male lions spent the most time in state one, 35% and 45% respectively, being clearly inactive with the states two, three and four becoming increasingly active except for Matikitiki. The expected value for the fourth state for the Matikitiki Male is much lower than for the fourth state of the other lions. This state represents very different movement compared to the state four for the other lions. The lionesses spent less time in the resting state compared to the male lions. The shape and scale (μ and σ) parameter values are relatively consistent for all the lions, but there are distinct differences in the expected value of the distribution.

Table 5.2 - HMM parameter values and state dependent distribution means.

Lions	Collar Id	δ	μ	σ	Γ_1	Γ_2	Γ_3	Γ_4	Exp Value
First Female	_	0.596	-4.655	0.956	0.751	0.196	0.050	_	0.015
		0.293	-1.659	0.963	0.437	0.462	0.100	_	0.303
		0.111	-0.154	0.502	0.168	0.364	0.468	_	0.972
		_	_	_	_	_	_	_	_
Nsenami Female	AM39/ AM115	0.175	-5.460	0.947	0.705	0.170	0.112	0.014	0.007
		0.427	-4.200	0.843	0.066	0.740	0.133	0.061	0.021
		0.221	-1.667	0.921	0.103	0.254	0.429	0.213	0.288
		0.177	-0.277	0.567	0.004	0.143	0.280	0.573	0.890
Thompson Female	AM186	0.180	-4.628	0.909	0.582	0.105	0.245	0.068	0.015
		0.214	-3.533	0.731	0.061	0.604	0.212	0.123	0.038
		0.340	-1.227	0.879	0.135	0.153	0.471	0.241	0.431
		0.266	0.102	0.479	0.061	0.052	0.340	0.547	1.243
Matikitiki Male	AM139	0.145	-6.225	1.090	0.607	0.000	0.315	0.078	0.004
		0.201	-4.790	1.371	0.115	0.102	0.061	0.722	0.021
		0.181	-3.138	1.004	0.144	0.148	0.680	0.028	0.072
		0.473	-1.722	0.720	0.017	0.325	0.000	0.659	0.232
Mavumbye Male	AU007	0.459	-5.102	0.909	0.771	0.111	0.098	0.021	0.009
		0.252	-3.806	1.291	0.243	0.512	0.138	0.106	0.051
		0.157	-1.253	0.638	0.279	0.293	0.244	0.184	0.350
		0.131	0.218	0.455	0.000	0.201	0.299	0.500	1.379
Nsenami Male	AM113/ AM183	0.351	-4.946	1.052	0.629	0.205	0.158	0.008	0.012
		0.273	-2.438	1.142	0.305	0.417	0.182	0.095	0.168
		0.255	-0.447	0.656	0.182	0.229	0.427	0.162	0.794
		0.120	0.674	0.375	0.004	0.240	0.342	0.415	2.106

The transition probability matrix is represented by the Γ's in Table 5.2. The diagonal probabilities of remaining in the same state are the highest amongst all lions except for the Matikitiki Male and Mavumbye Male lion. This is interpreted as the same state persisting for more than one time step at the time resolution (hourly). The probability of transitioning from the resting state to the most active fourth state is low for all lions. When in a resting state the lions prefer to remain in the same state or move to an active movement state (for example 0.315 for Matikitiki Male). When the lions are in a local movement state, they are most likely to remain in the same state or move to a resting state, except for Matikitiki Male lion where probability of remaining in the second state was 0.102. From an active state the lions are most likely to remain in the same state or move into a local state except for the Mavumbye Male lion, with the probability of remaining the same state being 0.244. When the lions are in a relocating state they are most likely to remain in the same state (for example 0.659 for Matikitiki Male lion) or move to the third state for the exception of Matikitiki Male with a probability of almost 0 of moving into the third state.

The expected movement distances for each state are shown in the last column of Table 5.2.
The expected movement distance is consistent among all lions around 10m in the resting state. In the second state the expected movement distance is around 50m except for the Nsenami Male lion which is 168m. The expected movement distance among the lioness in the third state is consistent around 345m whereas the expected movement distance among the male lions are inconsistent and are around 72m, 350m and 794m respectively, where it can be assumed that male lions range more widely. In the fourth state, among the lionesses the expected movement distance is consistent around 1km except for the First Female where a fourth state was not considered. As for the male lions the expected movement distance is around 1.5km except for the Matikitiki Male lion with an expected value of 232m.

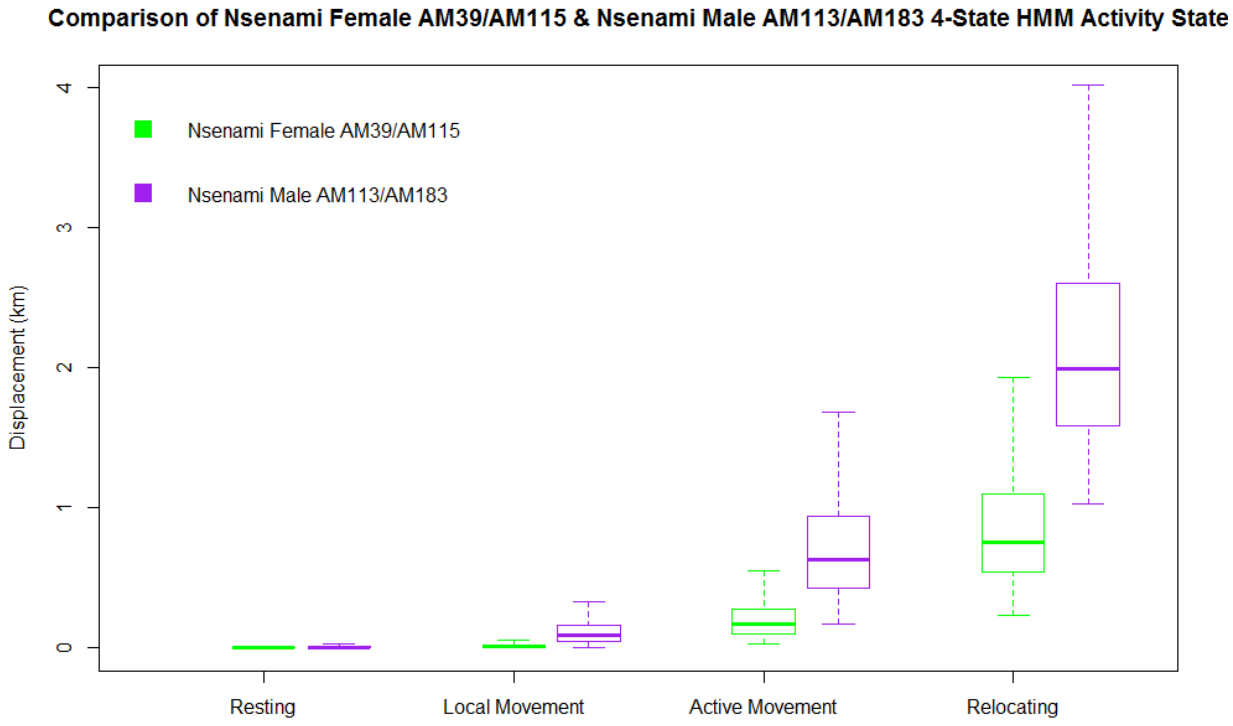


Figure 5.1 - Comparisons of Nsenami Female AM39/AM115 and Nsenami Male AM113/AM183 4-State HMM Activity State

Figure 5.2 compares the activity state of the Nsenami Female and Nsenami Male lion. Both lions had a similar resting state. The Nsenami Male had clearly a wider local, active and relocating state the Nsenami Female. The second and third state of the lioness is relatively similar to the Nsenami Male third and fourth state. The Nsenami Male had a much longer relocating state. It can be assumed that the Nsenami Male demonstrated a more patrolling behaviour than the Nsenami Female which goes in hand with the biological interpretation of a male lion's behaviour. Apart from the resting state, there is a distinct difference in the behaviour of the two lions, despite being associated with the same pride.

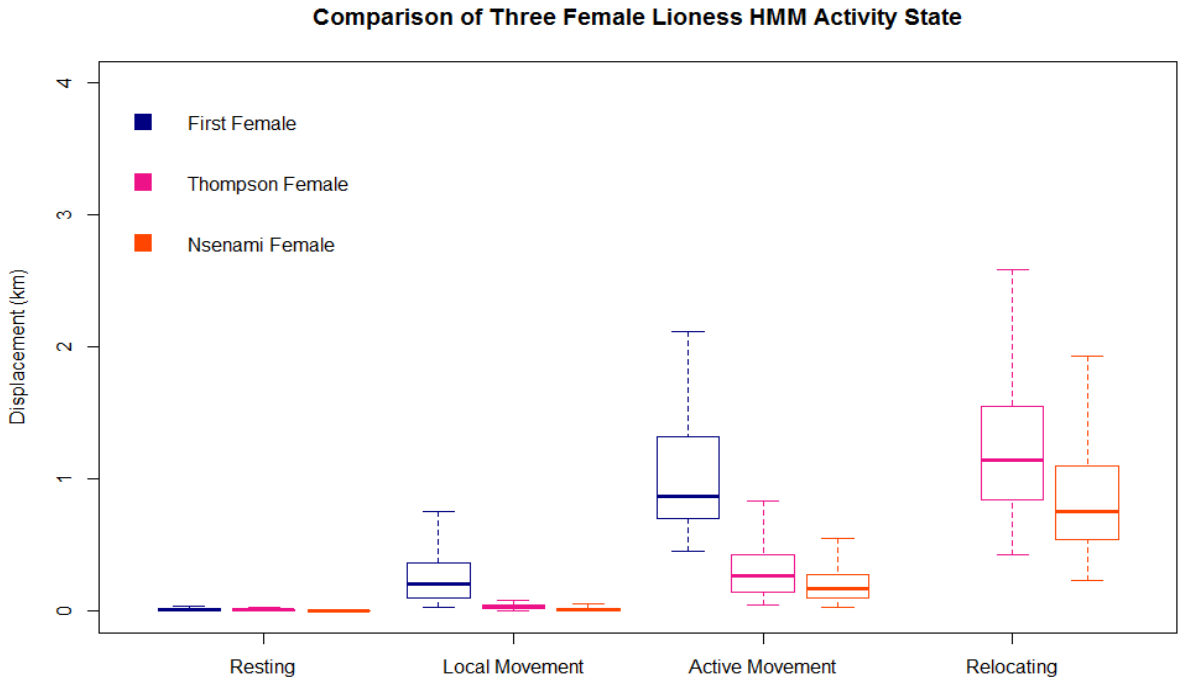


Figure 5.2 - A comparison of the Three Female Lions at the KNP

Figure 5.3 differentiates the activity state between the three lionesses: First, Thompson and Nsenami Females. A three-state HMM was fitted in the case of the First Female. The three lionesses had almost identical resting state. The First Female lion second and third state was relatively similar to the third and fourth state of the Thompson and Nsenami Female lions. The Thompson and Nsenami Females have a comparably similar behaviour except for a slightly wider fourth state in the case of Thompson Female lion.

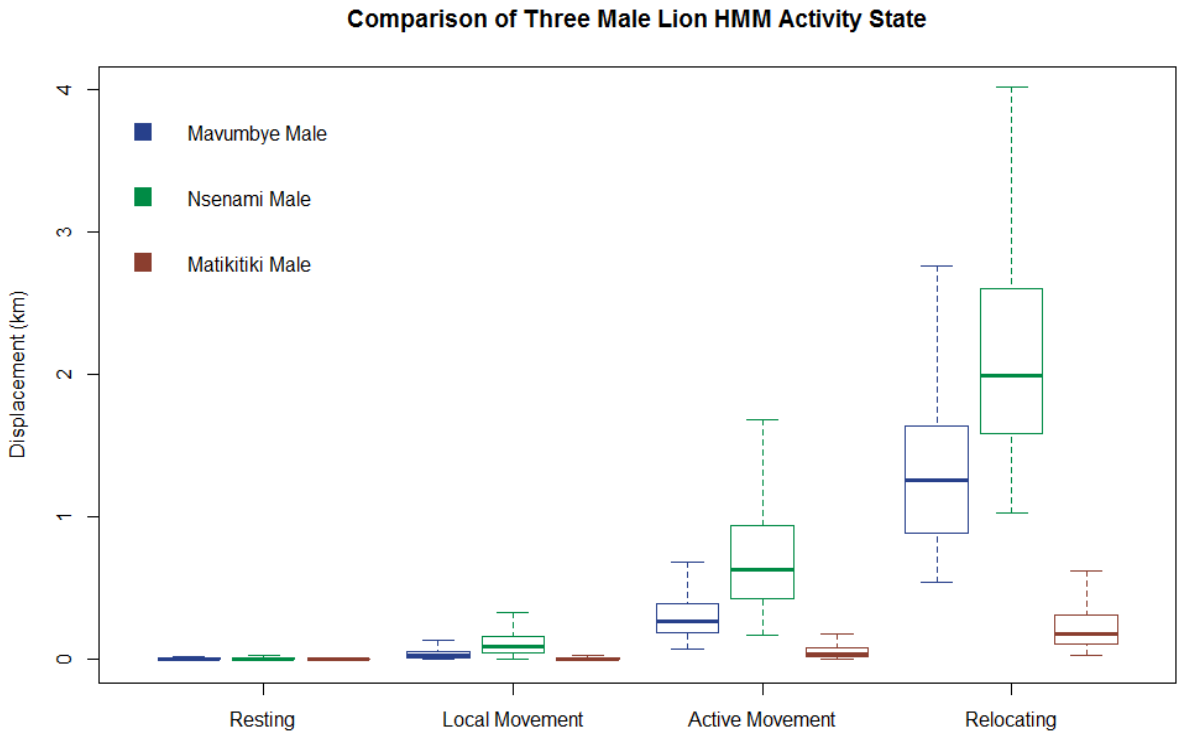


Figure 5.3 - A comparison of the Three Male Lion HMM Activity state at the KNP.

Figure 5.4 distinguishes the activity state of the three male lions: Mavumbye, Nsenami and Matikitiki Male. The male lions had relatively consistent resting state amongst individuals and female lions in Figure 5.2. Each male lion had distinctive activity states, with wider movements in the case of Nsenami Male. Mavumbye Male third state was relatively similar to Matikitiki Male fourth state. Matikitiki seems to have a very short range. Figure 5.4 shows very distinctive patterns of behaviour amongst the male lions.

Figure 5.5 and 5.6 depicts the activity states by time of day of the Nsenami Female and Nsenami Male lion. Quite different patterns are observed between the two. In Figure 5.5 the relocating state is dominant predawn, followed by an increase in local movement and resting states. These activity states ease off during the heat of the day and then pick up again later into the afternoon when it is cooler. Lions usually hunt at night which explains the high level of activity during the night. In Figure 5.6, the Nsenami Male displays a dominant resting state on a more diel cycle. The activity states become subservient between 9am and 3pm. Also the Nsenami Female shows a higher level of activity at night than the Nsenami Male lion which can be expected as it is usually the female lions of a pride that do majority of the hunting (Funston et al., 2001).

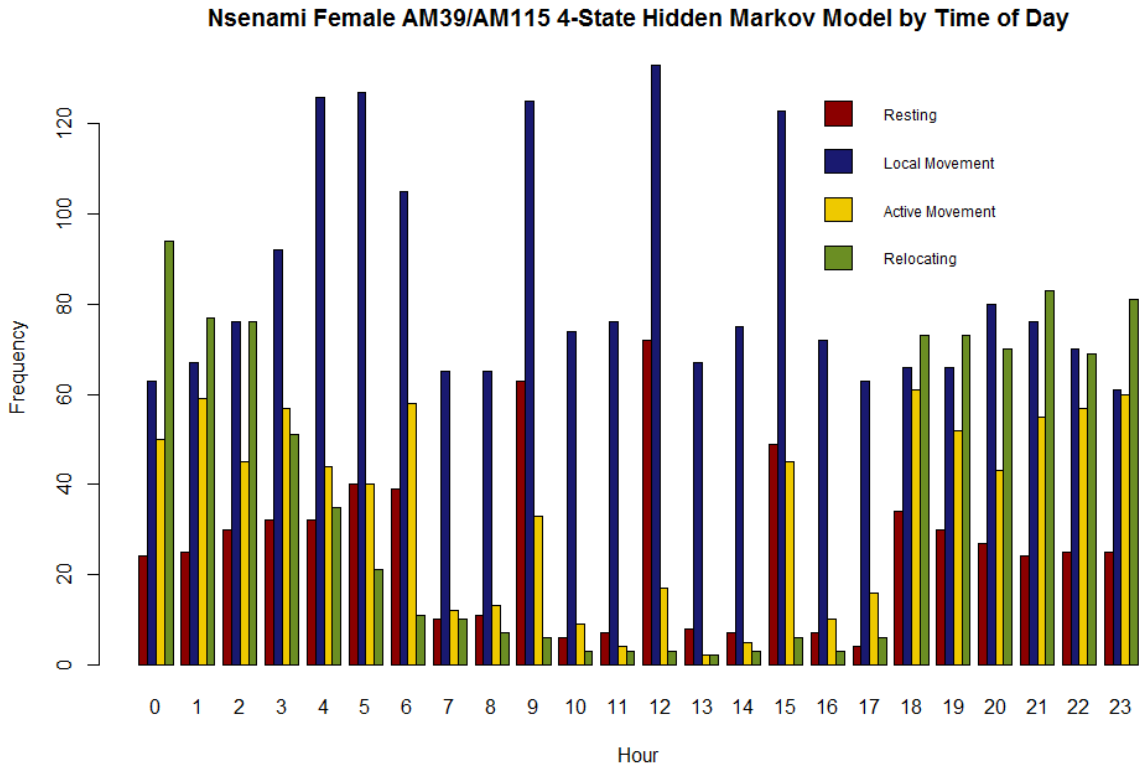


Figure 5.4 - Activity State Comparison using HMM for Nsenami AM39/AM115 Female by Time of Day.

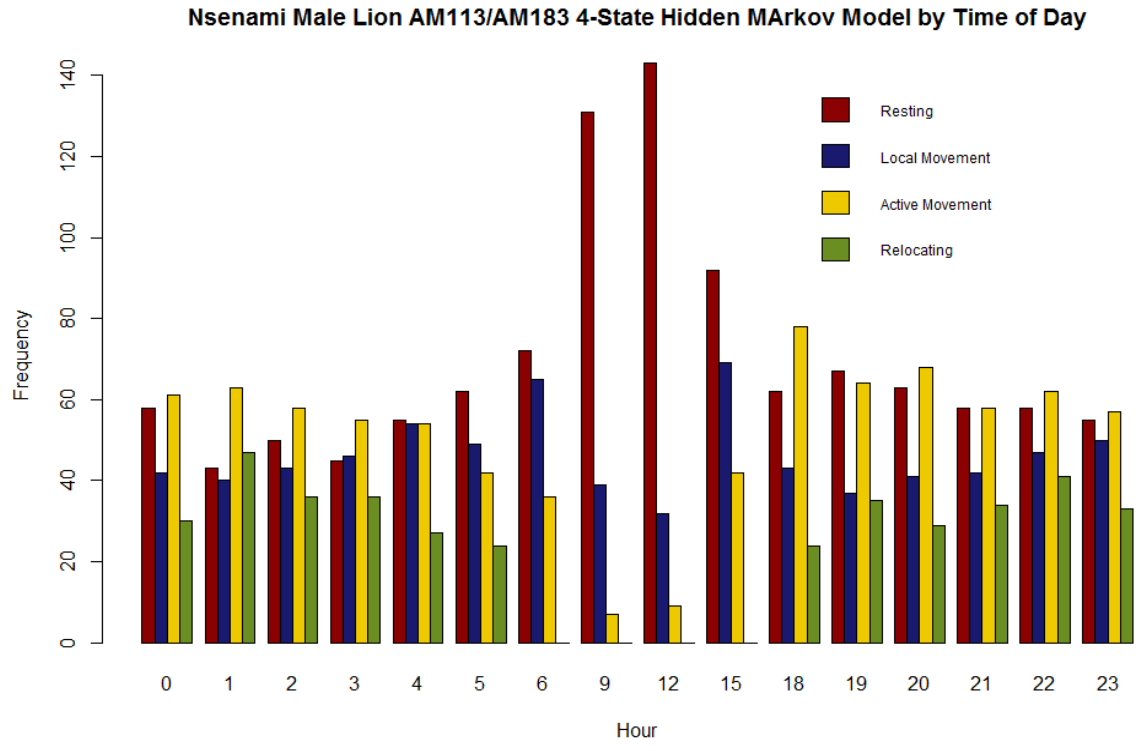


Figure 5.5 - - Activity State Comparison using HMM for Nsenami Male by Time of Day.

Figure 5.5 shows that the Nsenami Male is more active at night which is anticipated as lion usually hunt during the night. A study carried out on the ungulate movement patterns (Goodall, 2014) at the Kruger National Park showed very different patterns compared to the lions in this study. Ungulates tend to be more active during the day as expected as the ungulates would want to avoid risk of predation during the night. Moreover Van Niekerk (2015) analysed eland movement data in two regions in the Eastern Cape, it was observed that elands were more active during the day than at night. Compared to lions, ungulates tend to rest mostly at night and lions during the day.

5.3 Conclusion

Among the variety of models that have been used to analyze animal tracking data, Hidden Markov Models have stood out in recent years because of their appealing combination of model flexibility, clear interpretability and computational tractability (Michelot et al., 2016). Using distance between locations or movement rate to behavioural states, it was shown that the HMM can be used to describe and predict the complexities of animal movement. Model parameters enabled inference about behavioural states and transitions to the most likely behavioural state. It was found that each lion had a very distinctive pattern of behaviour although they generally have a dominant resting state. One significant difference between the lions and lionesses was that male lions tend to have a longer relocation movement.

Chapter 6 – Kill Predictions

6.1 Introduction

Predation rate and kill rate are both fundamental statistics for understanding predation (Vucetich et al., 2011). It is also crucial for managing African wildlife to attenuate human-wildlife conflict (Tambling et al., 2010). Unfortunately, evaluating kill rates of large carnivores has been logistically difficult (Franke et al., 2006). Lions are one of the most important predators, functionally and numerically, in many African systems (Kittle et al., 2016). In African lions, several techniques such as opportunistic carcass location, stomach content and scat analysis have been used to assess factors that affect both hunting behaviour and success (Tambling et al., 2010).

The advent of Global Positioning System provides researchers with the potential to estimate kill rates without the need of rigorous continuous observation. The objective of this project was to present a new approach that uses HMMs to identify observer-confirmed kill sites from movement patterns of GPS collared lions. In this study, an existing data set that utilized GPS technology provided lion movement data combined with confirmed kill-sites was used. Chapter 5 of the study provided insight into lions' behaviour in terms of bout lengths and transition. This section of the project aimed at evaluating whether HMMs could predict observer-confirmed kill sites from GPS relocation data and specific decision rules.

6.2 Methodology

Generally, the methodology involved using the HMM approach and a set of decision rules to analyse the state sequence of a lion close to a kill.

6.2.1 Training Model Analysis

The decision rules were built on the observer-confirmed kill data obtained from Tambling et al. (2010) study. Only two sets of observer-confirmed kill data were available for the Thompson Female kills and Nsenami Female. The Thompson Female was used as the training model. The first step in constructing the rules consisted of determining the state sequence of the training model using the Viterbi Algorithm from the moveHMM package (Michelot et al., 2016). The Viterbi algorithm is used to find the sequence of states which maximizes the probability of the observation sequence given the model (Goodall, 2014).

From Chapter Four, it was inferred that state one corresponded to resting, state two to local movements, state three to active movements and state four to relocating behaviour. Figure 6.1 shows the most likely states for hourly movements the Thompson female lion. It can be observed that this lioness has a dominant second state.

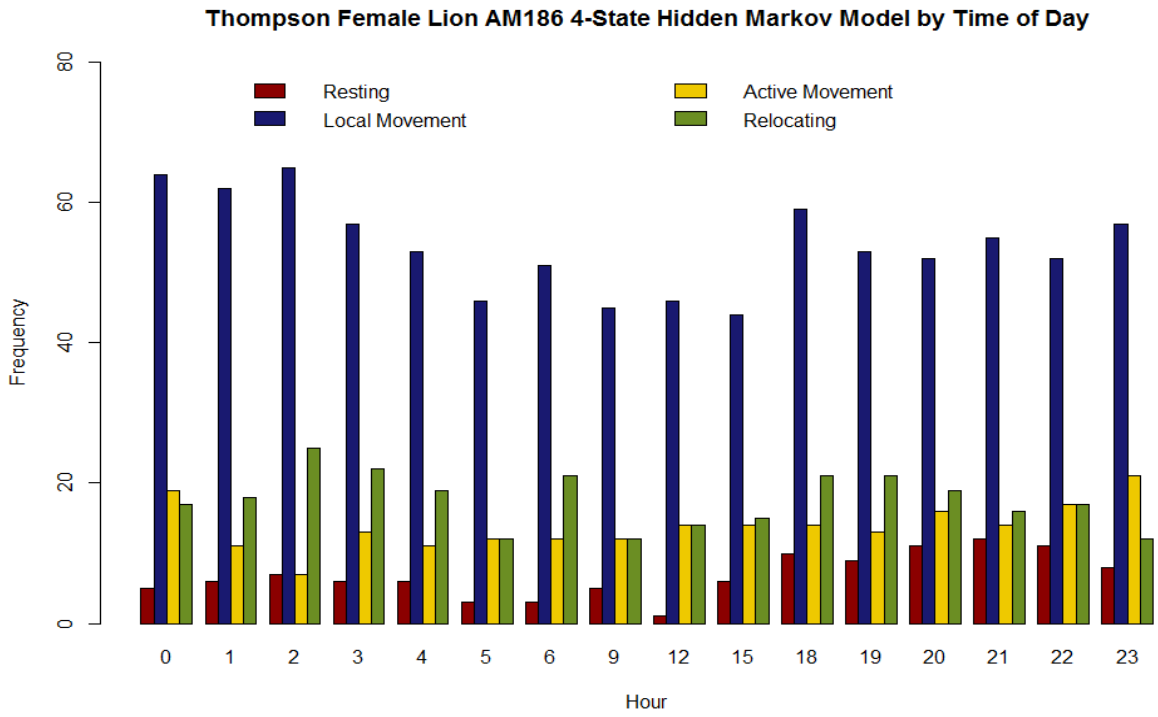


Figure 6.1 - Activity state of Thompson Female by Time of Day.

Afterwards the actual kills from the observer-confirmed data were matched to the state sequence of the lioness. A moving average of the states before and after a kill was computed, where the states values ranged from one to four. A three-, four- and five-state moving average was found. For example, in the case of a three-state moving average the sum of the three states before a kill and the sum of the three states after a kill was calculated. Across the three set of moving averages, most of the actual kills occurred where the average score before a kill ranged from 6 to 9 and the average score after a kill ranged from 5 to 11. Different sets of rules were constructed from those ranges. For example, in the case of a three-state moving average a kill could be suspected when the score of the states would sum up to be greater or equal to 6 and the score of the next three states would be less or equal to 11 . Table 6.1 shows the different ranges applied across the different moving averages to the training model and the results observed.

Table 6.1 - Different Decision rules applied to training model and results.

Moving Average	**Rule**	**Predictions**	**Comments**
3-States Moving Average	Sum before ≥ 6.5 Sum after ≤ 10.5	794 (out of 1474 observations) kills were predicted. 19 kills (86%) were correctly predicted out of the 22 actual kills.	The rule predicted a kill site before the clusters were investigated and a kill was found.
	Sum before ≥ 7.5 Sum after ≤ 9.5	506 (out of 1474 observation) kills were predicted by the model. 17 kills (77%) were correctly predicted out of the 22 actual kills.	3 kills were predicted close to the place and time where the kill actually happened - 2 impalas & buffalo.
	Sum before ≥ 8.5 Sum after ≤ 7.5	208 (out of 1474 observation) kills were predicted by the model. 17 kills (63%) were correctly predicted out of the 22 actual kills.	2 kills were predicted close to the place and time where the kill actually happened – Buffalo & Zebra.
4-States Moving Average	Sum before ≥ 6.5 Sum after ≤ 10.5	539 (out of 1474 observation) kills were predicted. 15 kills (68%) were correctly predicted out of the 22 actual kills.	3 kills were predicted close to the place and time where the kill actually happened - Impala, buffalo and zebra.
	Sum before ≥ 7.5 Sum after ≤ 9.5	373 (out of 1474 observation) kills were predicted. 12 kills (54.5%) were correctly predicted out of the 22 actual kills.	The model predicted kills when the kills actually happened.
	Sum before ≥ 8.5 Sum after ≤ 7.5	93 (out of 1474 observation) kills were predicted. 4 kills (18%) were correctly predicted out of the 22 actual kills.	The model predicted kills close to when the kills actually happened. – Wildebeests
5-States Moving Average	Sum before ≥ 6.5 Sum after ≤ 10.5	266 (out of 1474 observation) kills were predicted. 10 kills (45%) were correctly predicted out of the 22 actual kills.	The model predicted kills close to when the kills actually happened.
	Sum before ≥ 7.5 Sum after ≤ 9.5	149 (out of 1474 observation) kills were predicted. 6 kills (27%) were correctly predicted out of the 22 actual kills.	The model predicted kills close when the kills actually happened.
	Sum before ≥ 8.5 Sum after ≤ 7.5	78 (out of 1474 observations) kills were predicted. 4 kills (18%) were correctly predicted out of the 22 actual kills.	

The ‘best’ decision rule with 86% successful kill prediction was the case of a three-state moving average where the score before a kill was set to be greater than 6.5 and the score after a kill to be 10.5. In this instance, if the score before was greater than 6.5 and the score after was less than 10.5, the model would predict a kill. Figure 6.2 shows the result of the best model under the column ‘Rule-1’ where 19 out of the 22 confirmed kills were predicted correctly.

Obs ID	Lion	Collar	Date	Time	states	Species	score befor	score after	Observer-Confirmed	RULE - 1
36	F5	AM186	28/01/2007	09:01:20	1	Impala	13	14	kill	kill
80	F5	AM186	01/02/2007	01:04:26	1	Wildebeest	19	5	kill	kill
150	F5	AM186	07/02/2007	03:02:37	2	Buffalo	19	14	kill	kill
183	F5	AM186	09/02/2007	21:04:53	3	Impala	16	18	kill	no Kill
193	F5	AM186	10/02/2007	21:01:17	1	Zebra	20	11	kill	kill
208	F5	AM186	11/02/2007	23:05:23	2	Wildebeest	19	11	kill	kill
276	F5	AM186	18/02/2007	01:04:14	1	Impala	19	13	kill	kill
319	F5	AM186	22/02/2007	05:09:44	2	Giraffe	17	10	kill	kill
382	F5	AM186	27/02/2007	15:01:20	1	Kudu	18	5	kill	kill
435	F5	AM186	05/03/2007	12:01:20	2	zebra	12	11	kill	kill
525	F5	AM186	14/03/2007	09:12:47	2	zebra	15	15	kill	kill
570	F5	AM186	18/03/2007	09:11:43	1	wildebeest	8	11	kill	no Kill
640	F5	AM186	24/03/2007	06:12:22	2	X	19	16	kill	kill
677	F5	AM186	27/03/2007	02:06:19	1	Impala	9	16	kill	no Kill
771	F5	AM186	03/04/2007	15:01:58	1	Kudu	18	5	kill	kill
797	F5	AM186	05/04/2007	06:09:14	2	Waterbuck	14	13	kill	kill
849	F5	AM186	09/04/2007	20:12:59	2	X	16	10	kill	kill
897	F5	AM186	13/04/2007	19:04:43	1	Wildebeest	11	7	kill	kill
950	F5	AM186	17/04/2007	22:01:31	1	Wildebeest	18	9	kill	kill
997	F5	AM186	21/04/2007	22:02:00	2	Wildebeest	14	10	kill	kill
1055	F5	AM186	26/04/2007	21:06:20	1	Wildebeest	17	7	kill	kill
1099	F5	AM186	30/04/2007	06:10:08	2	Waterbuck	13	10	kill	kill

Figure 6.2 - Decision Rule analysis

6.3 Results

Based on the training model, a 3-state moving average with the score before a potential kill being greater than or equal to 6.5 and the score after a potential kill being less than or equal to 10.5 was considered as being the ‘best’ decision rule. In this section the ‘best’ decision rule is applied to the Nsenami Female lion. Figure 6.3 shows the procedure followed and the result observed. The model predicted 1865 kills out of 4366 observations. Out of 38 confirmed kills, the model successfully predicted 33 kills, 87% success rate respectively. Figure 6.4 shows the proportion of predicted kills and observer-confirmed kills compared to the number of clusters investigated by the Tambling et al. (2010) study. They investigated 196 clusters and have been able to confirm 38 kills.

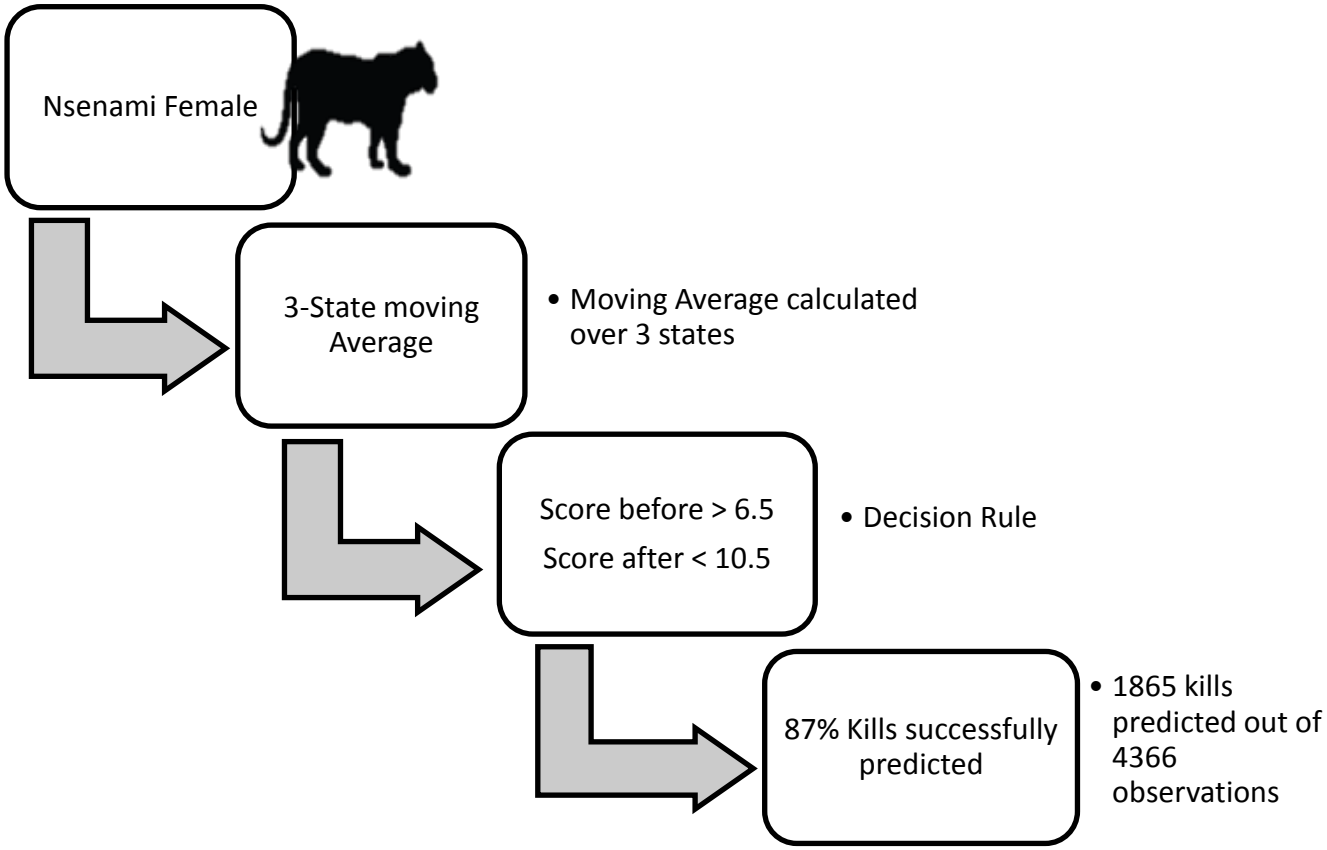


Figure 6.1 - Nsenami Female Kill Prediction Analysis

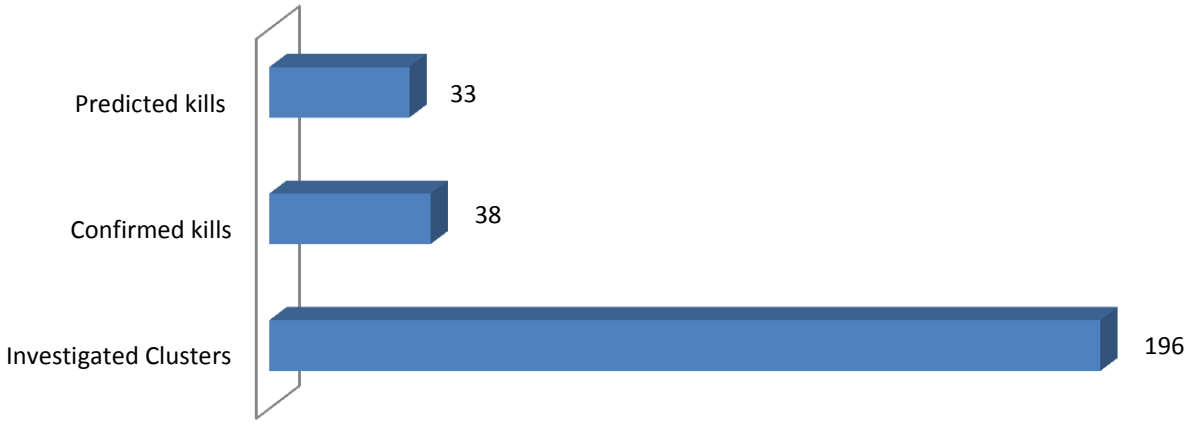


Figure 6.2 - Proportion of kills to clusters

6.4 Discussions

Although continuous observation remains the best technique to study predator-prey relationships, the HMM together with the decision rules were effective at identifying potential kill sites from lion GPS data. In the case of the Nsenami Female, there are three occasions where the model predicted kills a couple of hours before clusters were checked and an actual kill was found. Table 6.2 shows an example where the model predicted a kill two hours prior to a kill was found. It can be assumed that the kill might have happened when the model predicted so but clusters were only investigated hours later.

Moreover stomach contents location is often considered the most appropriate diagnostic of the actual kill site because lions regularly move carcasses from the kill site whereas stomach contents are almost always deposited within a few meters of the actual kill (Davies et al., 2016). It was observed from previous studies that often collared lions would move away and then returned to a cluster (Tambling et al., 2010). In instance, there are occasions where the models predicted a kill but an actual kill was not found which can be assumed that the lions might have moved the carcass.

Previous lion hunting studies carried out by Tambling et al. (2010) investigated 1447 out of 2433 clusters and 234 kills were found. Although visiting all clusters in the field to determine whether they are kill sites will provide the most accurate measure of kill rate, this requires extensive field effort and associated cost. Using the HMM approach and a set of decision rules, 87% of the known kills found from Tambling et al. (2010) study were successfully predicted with no field work needed. The HMM analysis and state sequence pattern is a novel approach in the study of predicting lion kill sites. HMMs have only been used, as far as I am aware for kill-site prediction in the case of wolfs (Franke et al., 2006).

When resources for field inspections are available but limited, the HMM approach can help guide these efforts. For example, if only 51 clusters were investigated out of 121 in the case of Thompson Female, 100% of the kills would have been detected. Alternatively, predicting the probability of a kill allows prioritization of sites to visit. In the end, the HMMs were successful at predicting kill sites from non-kill sites. Inferring kill-sites from frequent relocation data combined with HMM analysis may contribute toward the alleviating cost and danger of field observation by identifying priority sites to investigate.

Table 6.2- Model predictions of missed kill.

	Lion	Collar	Covera	Date	Time	Long	Lat	states.1		Rule	Checked	Specie
1418	F4	AM39	No	30/11/2005	07:33:10	369315	7304099	4		kill		
1419	F4	AM39	No	30/11/2005	08:33:10	370018	7303598	2		kill		
1420	F4	AM39	Yes	30/11/2005	09:33:10	370007	7303591	2		kill		
1421	F4	AM39	Yes	**30/11/2005**	**10:33:38**	369987	7303580	2		**kill**		
1422	F4	AM39	No	30/11/2005	11:33:48	369981	7303581	2		no Kill		
1423	F4	AM39	Yes	**30/11/2005**	**12:33:38**	369979	7303579	2		**no Kill**	**kill**	Giraffe
1424	F4	AM39	Yes	30/11/2005	13:33:11	370082	7303588	2		no Kill		
3020	F4	AM115	No	17/06/2006	20:02:50	363629	7309690	4		no Kill		
3021	F4	AM115	No	17/06/2006	21:02:49	363921	7310645	4		no Kill		
3022	F4	AM115	No	17/06/2006	22:10:18	364301	7310922	4		no Kill		
3023	F4	AM115	No	17/06/2006	23:03:19	364754	7310496	4		no Kill		
3024	F4	AM115	No	**18/06/2006**	**00:01:08**	365291	7310328	4		**no Kill**	**kill**	**impala**
3025	F4	AM115	No	**18/06/2006**	**01:01:23**	366210	7310203	4		**kill**		
3026	F4	AM115	No	18/06/2006	03:01:49	365998	7309599	4		kill		
3027	F4	AM115	No	18/06/2006	04:08:45	366104	7308469	3		kill		
3028	F4	AM115	No	18/06/2006	12:00:55	366206	7307231	1		kill		
3029	F4	AM115	No	18/06/2006	19:03:19	366196	7307260	1		no Kill		
3374	F4	AM115	No	19/07/2006	06:01:02	366591	7303850	3		kill		
3375	F4	AM115	No	19/07/2006	09:10:20	366909	7302926	1		kill		
3376	F4	AM115	No	19/07/2006	12:01:14	366899	7302920	1		no Kill		
3377	F4	AM115	No	19/07/2006	15:02:37	366908	7302910	3		no Kill		
3378	F4	AM115	No	**19/07/2006**	**18:06:41**	367348	7302971	2		**no Kill**	**kill**	impala
3379	F4	AM115	No	19/07/2006	19:08:32	367350	7302961	2		no Kill		
3380	F4	AM115	No	**19/07/2006**	**21:01:07**	367358	7302980	3		**kill**		
3381	F4	AM115	No	19/07/2006	22:02:20	367370	7303119	4		kill		
3382	F4	AM115	No	19/07/2006	23:01:19	368171	7302874	4		kill		
3383	F4	AM115	No	20/07/2006	01:00:31	369565	7302219	2		kill	checked but no kill	
3384	F4	AM115	No	20/07/2006	02:10:53	369567	7302229	3		kill		

Chapter 7 – Conclusion

7.1 Summary

The study has shown that HMMs can be applied to animal movement data to provide relevant results with biological interpretation in terms of the movement behaviour of the lion. From this project, the log-normal distribution has been shown to be the most suitable distribution. A movement rate was used in cases where GPS coordinates were collected on an irregular basis. Model parameters enabled inference about behavioural states and transitions to the most likely behavioural state. Four-state HMMs were fitted to the lions with the exception of the First Female where a three-state model was fitted. The transition probability matrix from the Markov Chain provided additional information about the probabilities of moving from one state to the next.

The most dominant state being state one consistent with the resting state, lions indeed do not do much during daytime. It has also been observed that lions in the same pride have very distinct behavioural patterns. Male lions had relatively longer movement than lionesses which can be associated to a patrolling behaviour of the male lion. A useful output of the HMM is the predicted behavioural state from the Viterbi algorithm.

Moreover, the study aimed at investigating the ability of the HMM to predict lion kills in the central area of the KNP. The HMM is a novel approach never used in the study of lions' behaviour and kills predictions previously. Using the HMM approach and a set of decision rules, the model successfully predicted lion kills. On some occasions the model predicted kills very close to the time and location where the kill actually happened.

7.2 Future Work

The result of this study provides insight into the behaviour of lions and the ability to predict kills from lions' movement data and the HMM approach. The dialogue between statisticians and ecologists has intensified over recent decades. Animal movement is complex and irregular and thus very hard to model. The HMM can be applied to other ecological studies involving other predators to answer different ecological questions

Further work includes:

1. Inclusion of covariates (time and ecological)
2. Investigating the behaviour of lions in different regions.

"Lions are the epitome of all that is wild and wonderful about African savannas. Without lions savannas are a tame version of their former self"

Dr. Paul Funston

References

ANDERSON-SPRECHER, R. & LEDOLTER, J. 1991. State-space analysis of wildlife telemetry data. *Journal of the American Statistical Association,* 86**,** 596-602.

BOSQ, D. & NGUYEN, H. T. 2013. *A course in stochastic processes: stochastic models and statistical inference*, Springer Science & Business Media.

BRAACK, L. E. 2006. *Kruger National Park*, New Holland Publishers.

BREIMAN, L., FRIEDMAN, J., STONE, C. J. & OLSHEN, R. A. 1984. *Classification and regression trees*, CRC press.

BULLA, J., LAGONA, F., MARUOTTI, A. & PICONE, M. 2015. Environmental conditions in semi-enclosed basins: A dynamic latent class approach for mixed-type multivariate variables. *Journal de la Société Française de Statistique,* 156**,** 114-137.

BYGOTT, J. 1995. JP Hanby JD Bygott Ten Population Changes in Lions and Other Predators. *Serengeti: dynamics of an ecosystem***,** 249.

CAGNACCI, F., BOITANI, L., POWELL, R. A. & BOYCE, M. S. 2010. Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 365**,** 2157-2162.

CAPPÉ, O., MOULINES, E. & RYDÉN, T. Inference in hidden markov models. Proceedings of EUSFLAT Conference, 2009. 14-16.

CHAO, C.-C. & YAO, Y.-L. 1996. Hidden Markov models for the burst error statistics of Viterbi decoding. *IEEE transactions on communications,* 44**,** 1620-1622.

CHESSON, P. 1978. Predator-prey theory and variability. *Annual Review of Ecology and Systematics,* 9**,** 323-347.

CLARK, P. E., JOHNSON, D. E., KNIEP, M. A., JERMANN, P., HUTTASH, B., WOOD, A., JOHNSON, M., MCGILLIVAN, C. & TITUS, K. 2006. An advanced, low-cost, GPS-based animal tracking system. *Rangeland Ecology & Management,* 59**,** 334-340.

COSTAMAGNA, E., FAVALLI, L. & TARANTOLA, F. Modeling and analysis of aggregate and single stream internet traffic. Global Telecommunications Conference, 2003. GLOBECOM'03. IEEE, 2003. IEEE, 3830-3834.

DAVIES, A. B., TAMBLING, C. J., KERLEY, G. I. & ASNER, G. P. 2016. Effects of Vegetation Structure on the Location of Lion Kill Sites in African Thicket. *PloS one,* 11**,** e0149098.

DE'ATH, G. & FABRICIUS, K. E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology,* 81**,** 3178-3192.

EBERHARDT, L., GARROTT, R., SMITH, D., WHITE, P. & PETERSON, R. 2003. Assessing the impact of wolves on ungulate prey. *Ecological Applications,* 13**,** 776-783.

ÉRDI, P. & TÓTH, J. 1989. *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models*, Manchester University Press.

EVANS, M. R. 2012. Modelling ecological systems in a changing world. *Phil. Trans. R. Soc. B,* 367**,** 181-190.

FAWCETT, T. 2006. An introduction to ROC analysis. *Pattern recognition letters,* 27**,** 861-874.

FRANKE, A., CAELLI, T. & HUDSON, R. J. 2004. Analysis of movements and behavior of caribou (Rangifer tarandus) using hidden Markov models. *Ecological Modelling,* 173**,** 259-270.

FRANKE, A., CAELLI, T., KUZYK, G. & HUDSON, R. J. 2006. Prediction of wolf (Canis lupus) kill-sites using hidden Markov models. *Ecological modelling,* 197**,** 237-246.

FUNSTON, P., MILLS, M. & BIGGS, H. 2001. Factors affecting the hunting success of male and female lions in the Kruger National Park. *Journal of Zoology,* 253**,** 419-431.

GERTENBACH, W. D. 1983. Landscapes of the Kruger National Park. *Koedoe,* 26**,** 9-121.

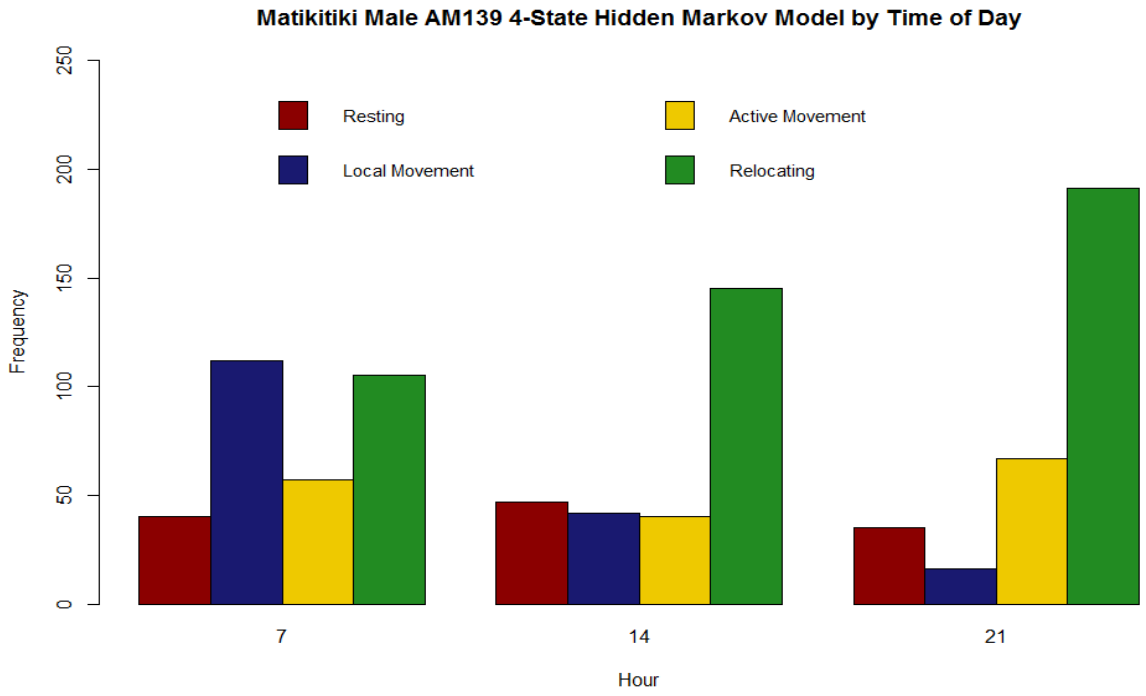
GOODALL, V., FATTI, P. & OWEN-SMITH, N. 2014. Analysing GPS Tracking Data of Animal Movements in the Kruger National Park.

GOODALL, V. L. 2014. *Statistical approaches towards analysing ungulate movement patterns in the Kruger National Park.*

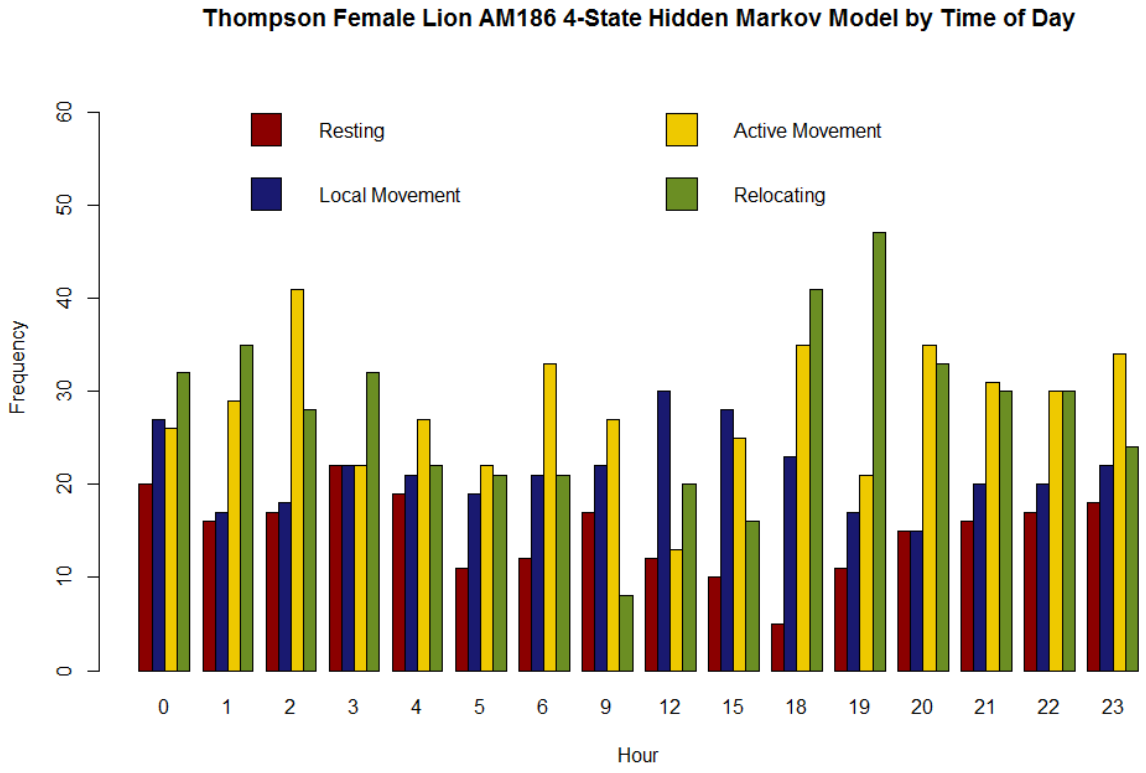
HUTCHINSON, C. E. 1984. The Kalman filter applied to aerospace and electronic systems. *IEEE transactions on aerospace and electronic systems*, 500-504.
IBE, O. 2013. *Markov processes for stochastic modeling*, Newnes.
JONSEN, I. D., FLEMMING, J. M. & MYERS, R. A. 2005. Robust state–space modeling of animal movement data. *Ecology,* 86, 2874-2880.
KARLIN, S. 2014. *A first course in stochastic processes*, Academic press.
KERK, M., ONORATO, D. P., CRIFFIELD, M. A., BOLKER, B. M., AUGUSTINE, B. C., MCKINLEY, S. A. & OLI, M. K. 2015. Hidden semi-Markov models reveal multiphasic movement of the endangered Florida panther. *Journal of Animal Ecology,* 84, 576-585.
KITTLE, A. M., BUKOMBE, J. K., SINCLAIR, A. R., MDUMA, S. A. & FRYXELL, J. M. 2016. Landscape-level movement patterns by lions in western Serengeti: comparing the influence of inter-specific competitors, habitat attributes and prey availability. *Mov Ecol,* 4, 17.
LANGROCK, R., KING, R., MATTHIOPOULOS, J., THOMAS, L., FORTIN, D. & MORALES, J. M. 2012. Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology,* 93, 2336-2342.
LAWLER, G. F. 2006. *Introduction to stochastic processes*, CRC Press.
LE, N. D., LEROUX, B. G., PUTERMAN, M. L. & ALBERT, P. S. 1992. Exact likelihood evaluation in a Markov mixture model for time series of seizure counts. *Biometrics*, 317-323.
LEMUNYAN, C. D., WHITE, W., NYBERG, E. & CHRISTIAN, J. J. 1959. Design of a miniature radio transmitter for use in animal studies. *The Journal of Wildlife Management,* 23, 107-110.
LEROUX, B. G. & PUTERMAN, M. L. 1992. Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics*, 545-558.
LEVINSON, S. E., RABINER, L. R. & SONDHI, M. M. 1983. An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *The Bell System Technical Journal,* 62, 1035-1074.
MACDONALD, I. L. 2014. Numerical Maximisation of Likelihood: A Neglected Alternative to EM? *International Statistical Review,* 82, 296-308.
MACKAY, R. J. 2003. *Hidden Markov models: Multiple processes and model selection.* University of British Columbia.
MCGARIGAL, K., CUSHMAN, S. A. & STAFFORD, S. 2013. *Multivariate statistics for wildlife and ecology research*, Springer Science & Business Media.
MCKELLAR, A. E., LANGROCK, R., WALTERS, J. R. & KESLER, D. C. 2015. Using mixed hidden Markov models to examine behavioral states in a cooperatively breeding bird. *Behavioral Ecology,* 26, 148-157.
MICHELOT, T., LANGROCK, R. & PATTERSON, T. A. 2016. moveHMM: An R package for the statistical modelling of animal movement data using hidden Markov models. *Methods in Ecology and Evolution*.
MILLSPAUGH, J. & MARZLUFF, J. M. 2001. *Radio tracking and animal populations*, Academic Press.
NASSERI, S. 2008. Application of an improved transition probability matrix based crack rating prediction methodology in Forida's highway network.
NOJIMA, Y., HAN, X.-H., TANIGUCHI, K. & CHEN, Y.-W. 6th International Conference on Biomedical Engineering and Informatics (BMEI). 2013 Tianxiang Zhen Hangzhou Normal University No.16 Xuelin St., Xiasha Higher Education Zone Hangzhou, China.
NORRIS, J. R. 1998. *Markov chains*, Cambridge university press.
NOWAK, R. M. 1999. *Walker's Mammals of the World*, JHU Press.
OTT, W. R. 1994. *Environmental statistics and data analysis*, CRC Press.
OWEN-SMITH, N. 2013. Daily movement responses by African savanna ungulates as an indicator of seasonal and annual food stress. *Wildlife Research,* 40, 232-240.
PARZEN, E. 1999. *Stochastic processes*, SIAM.
PATTERSON, T. A., PARTON, A., LANGROCK, R., BLACKWELL, P. G., THOMAS, L. & KING, R. 2016. Statistical modelling of animal movement: a myopic review and a

discussion of good practice. *arXiv preprint arXiv:1603.07511*.
PATTERSON, T. A., THOMAS, L., WILCOX, C., OVASKAINEN, O. & MATTHIOPOULOS, J. 2008. State–space models of individual animal movement. *Trends in ecology & evolution,* 23**,** 87-94.
PEDERSEN, M. W., RIGHTON, D., THYGESEN, U. H., ANDERSEN, K. H. & MADSEN, H. 2008. Geolocation of North Sea cod (Gadus morhua) using hidden Markov models and behavioural switching. *Canadian Journal of Fisheries and Aquatic Sciences,* 65**,** 2367-2377.
R DEVELOPMENT CORE TEAM 2010. R: A language and environment for statistical computing Vienna, Austria: R Foundation for statistical computing
RIPPLE, W. J. & BESCHTA, R. L. 2006. Linking a cougar decline, trophic cascade, and catastrophic regime shift in Zion National Park. *Biological Conservation,* 133**,** 397-408.
RUDNAI, J. A. 2012. *The social life of the lion: a study of the behaviour of wild lions (Panthera leo massaica [Newmann]) in the Nairobi National Park, Kenya*, Springer Science & Business Media.
RYDÉN, T., TERÄSVIRTA, T. & ÅSBRINK, S. 1998. Stylized facts of daily return series and the hidden Markov model. *Journal of applied econometrics***,** 217-244.
SCHLIEHE-DIECKS, S., KAPPELER, P. & LANGROCK, R. 2012. On the application of mixed hidden Markov models to multiple behavioural time series. *Interface Focus***,** rsfs20110077.
STUART, C. & STUART, T. 2009. *Dangerous Creatures of Africa*, Struik Nature.
TAMBLING, C. J., CAMERON, E. Z., TOIT, J. T. & GETZ, W. M. 2010. Methods for locating African lion kills using global positioning system movement data. *The Journal of Wildlife Management,* 74**,** 549-556.
TAMBLING, C. J., LAURENCE, S., BELLAN, S., CAMERON, E. Z., DU TOIT, J. T. & GETZ, W. M. 2012. Estimating carnivoran diets using a combination of carcass observations and scats from GPS clusters. *Journal of Zoology,* 286**,** 102-109.
THOMPSON, E. A. 1983. Optimal sampling for pedigree analysis: parameter estimation and genotypic uncertainty. *Theoretical population biology,* 24**,** 39-58.
URBANO, F., CAGNACCI, F., CALENGE, C., DETTKI, H., CAMERON, A. & NETELER, M. 2010. Wildlife tracking data management: a new vision. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 365**,** 2177-2185.
VANNIEKERK, B. 2015. The application of Independent Mixture Models and Hidden Markov Models to eland movement data in two regions in the Eastern Cape. . Nelson Mandela Metropolitan University.
VUCETICH, J. A., HEBBLEWHITE, M., SMITH, D. W. & PETERSON, R. O. 2011. Predicting prey population dynamics from kill rate, predation rate and predator–prey ratios in three wolf-ungulate systems. *Journal of Animal Ecology,* 80**,** 1236-1245.
WEBB, N. F., HEBBLEWHITE, M. & MERRILL, E. H. 2008. Statistical methods for identifying wolf kill sites using global positioning system locations. *The Journal of Wildlife Management,* 72**,** 798-807.
ZUCCHINI, W. & GUTTORP, P. 1991. A Hidden Markov Model for Space-Time Precipitation. *Water Resources Research,* 27**,** 1917-1923.
ZUCCHINI, W. & MACDONALD, I. L. 2009. *Hidden Markov models for time series: an introduction using R*, CRC press.

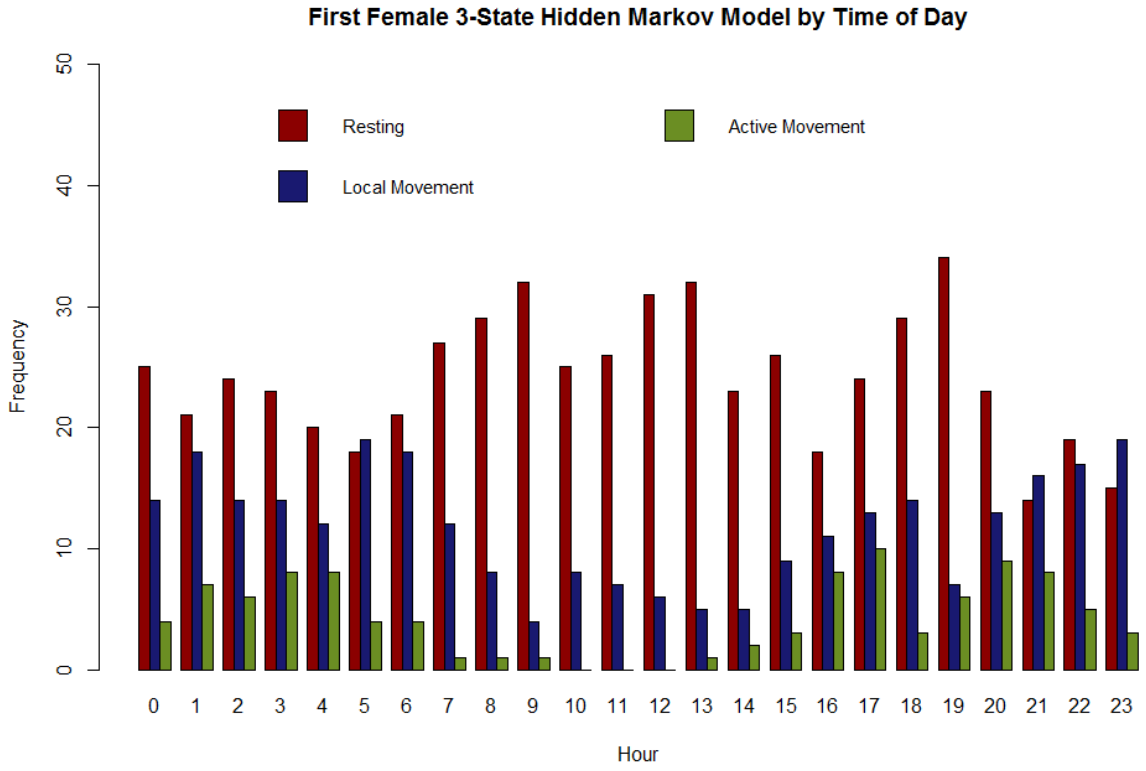
Appendices



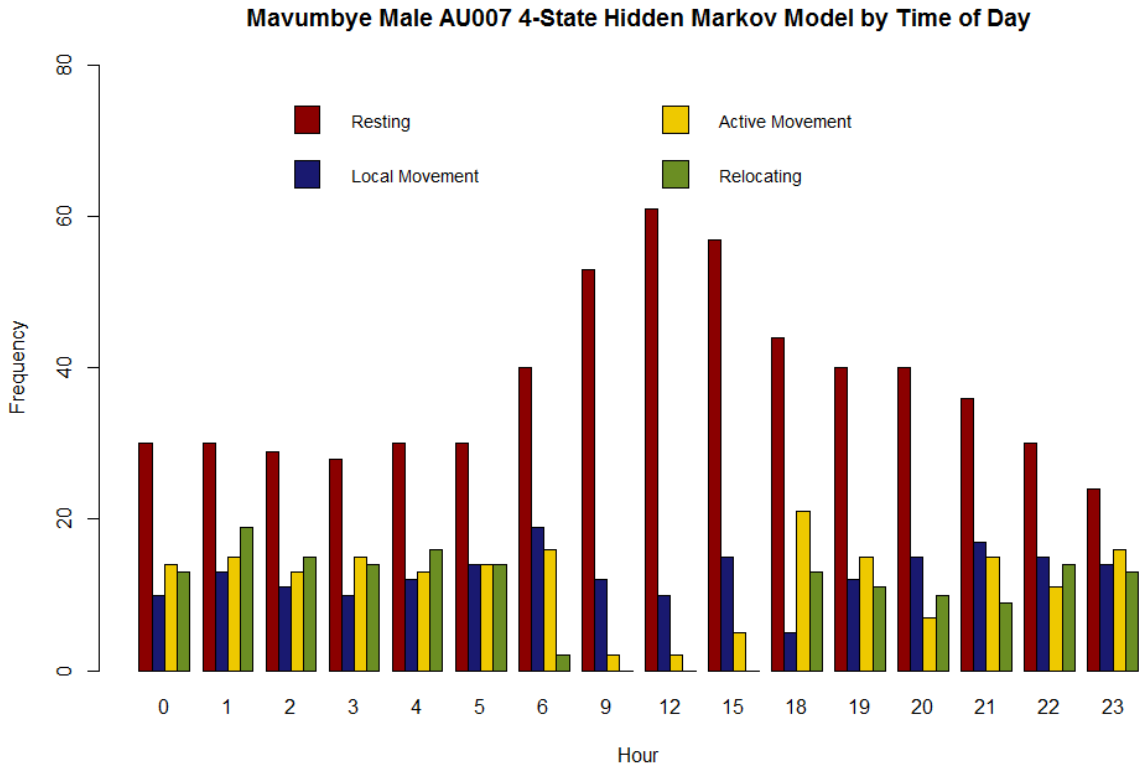
Appendix A: Activity state comparison using HMM for Matikitiki Male AM139 by Time of the Day



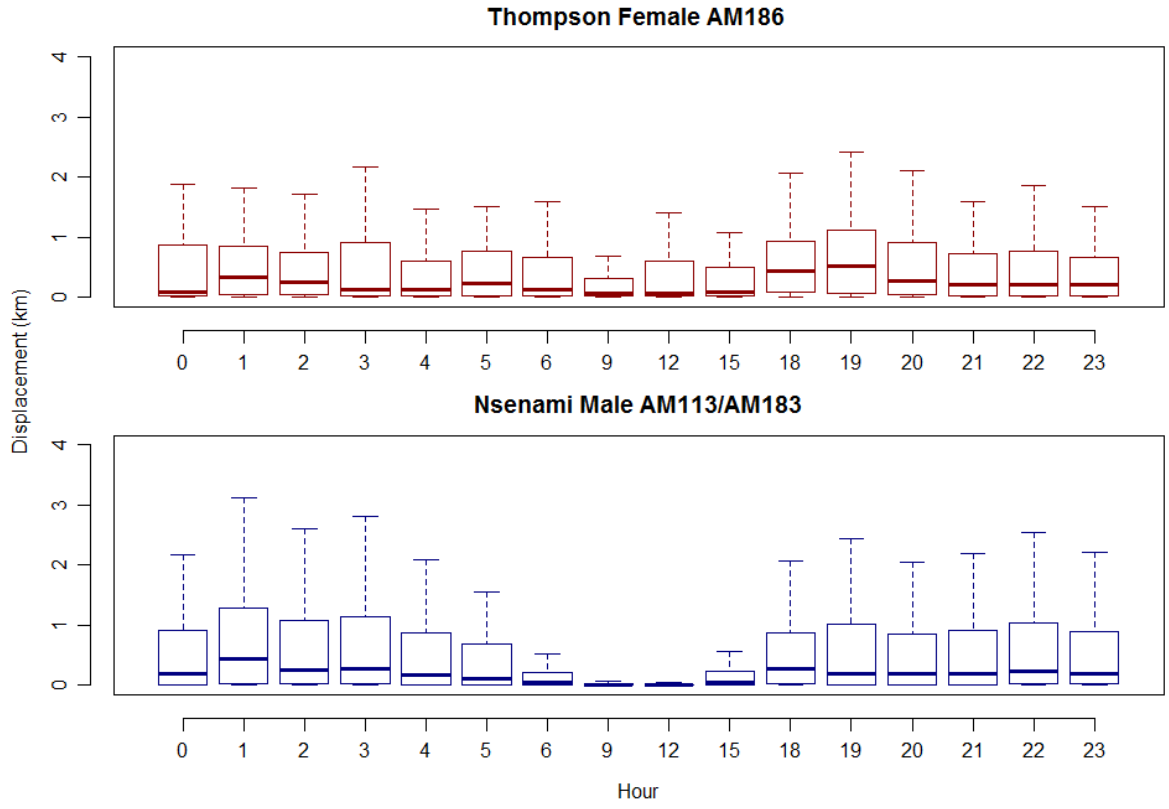
Appendix B: Activity State Comparison using HMM for Nsenami Female AM39/AM115 by Time of Day



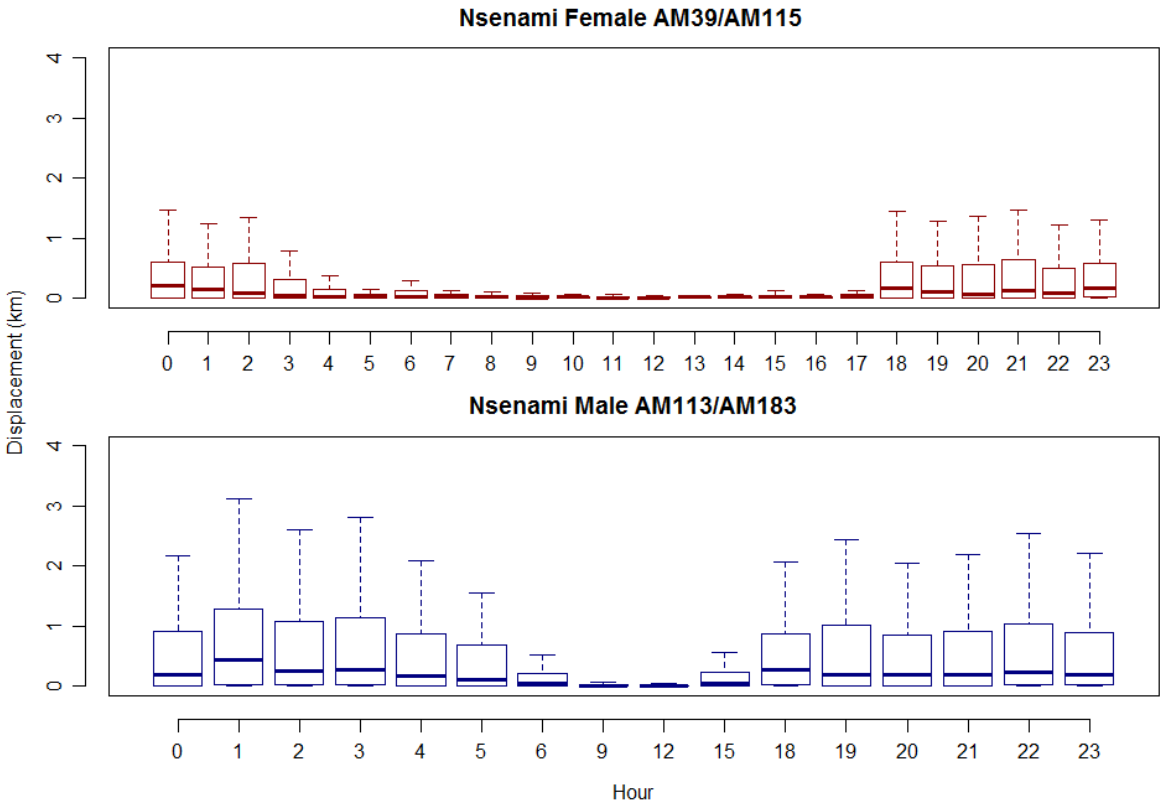
Appendix C: Activity State Comparison using HMM for First Female by Time of Day



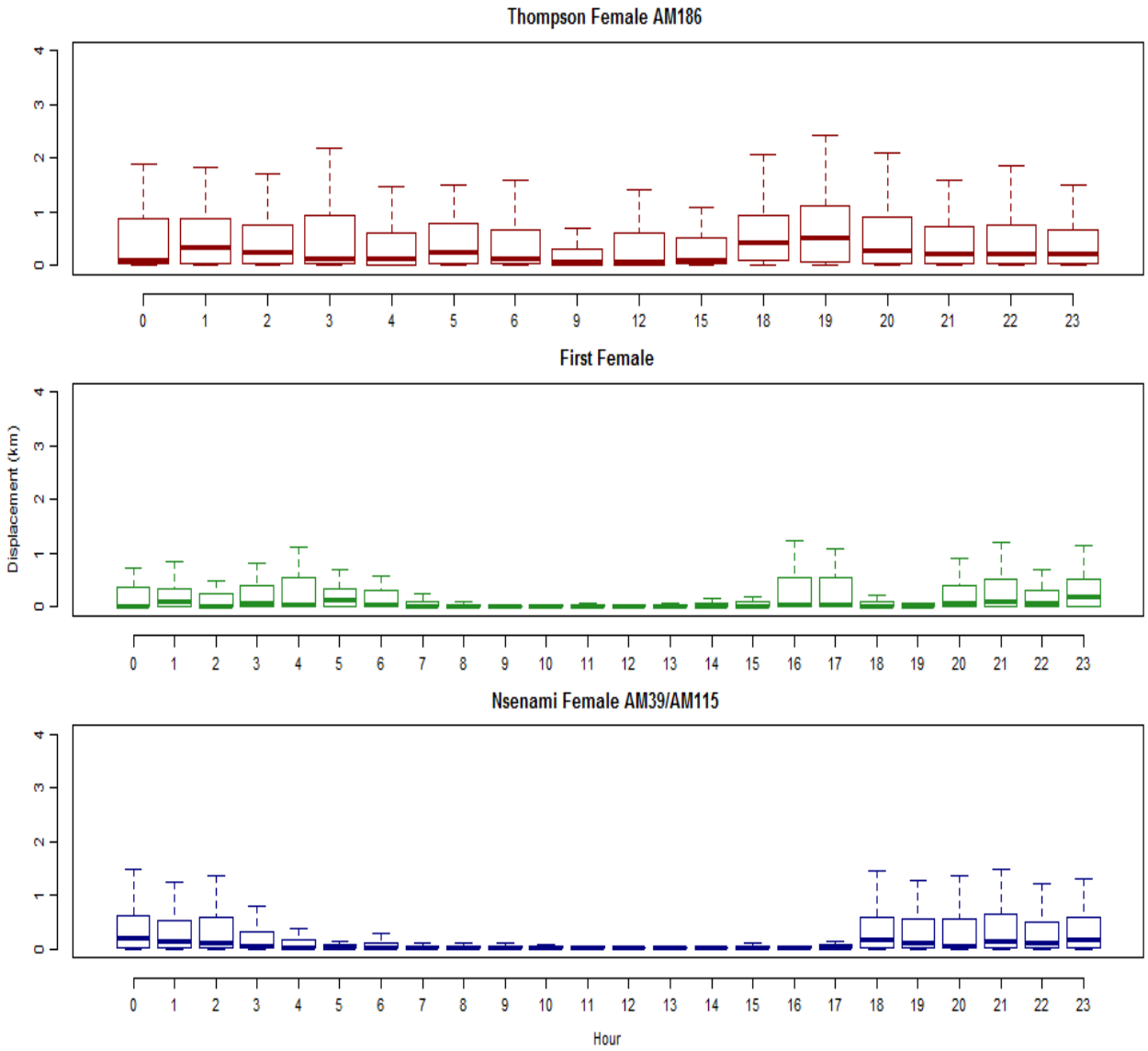
Appendix D: Activity State Comparison using HMM for Mavumbye Male by Time of Day



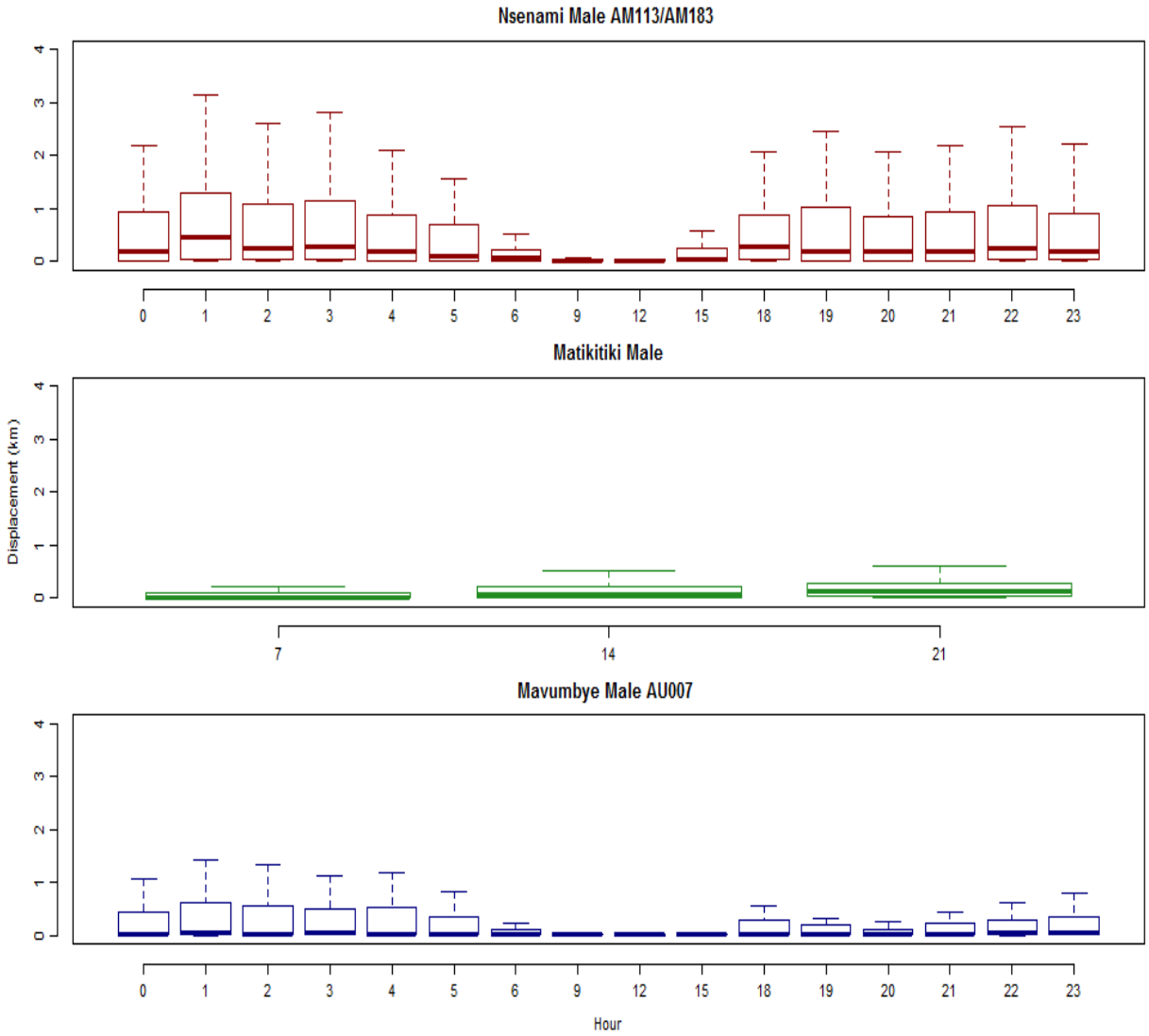
Appendix E: Comparison of Thompson Female and Nsenami Male Lion movement by Time of Day



Appendix F: Comparison of Nsenami Female and Nsenami Male Lion Movement by Time of Day



Appendix G: Comparison of Thompson, First and Nsenami Female Movement by Time of Day



Appendix H: Comparison of Nsenami, Matikitiki and Mavumbye Male Lion Movement by Time of Day