STATISTICAL APPROACHES TOWARDS ANALYSING UNGULATE MOVEMENT PATTERNS IN THE KRUGER NATIONAL PARK

Victoria Lucy Goodall

A Thesis submitted to the Faculty of Science, University of the Witwatersrand, in fulfilment of the degree of Doctor of Philosophy.

Johannesburg, 2014.

DECLARATION

I declare that this thesis is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

_____

_____ day of _____ 2014 in Johannesburg.

ABSTRACT

In this thesis I investigate the application of various statistical approaches towards analysing time series data collected using GPS collars placed on three ungulate species in the same region of the Kruger National Park, South Africa. Animal movement tracking is a rapidly advancing area of ecological research and large datasets are being collected with GPS locations of the animal, with shorter periods between successive locations. A statistical challenge is to segment the movement paths into groups which correspond to different behavioural activities. The aim of my study was to investigate and compare alternative statistical approaches for analysing GPS data and to establish the best statistical framework for interpreting these large herbivore movements. The focus was on which methods are the most appropriate for these animals and the comparison of the movement patterns across species and season.

Independent Mixture, Hidden Markov and Bayesian State-Space Models were used to analyse the hourly and daily movements of sable antelope, buffalo and zebra. Mixture Models provide a basic clustering technique to segment the movement paths and identify different underlying groups within the data assumed to correspond to different behavioural states. Posterior probabilities of group membership are used to allocate movements between successive locations to different states. This method ignores the dependence between successive movements. Hidden Markov models (HMMs) use a time series technique and include a dependency between successive observations via a Markov process. Extensions to the HMMs were applied to allow for the inclusion of seasonal covariates and irregular time gaps between successive observations caused by missing locations. A Bayesian state-space model fits a random walk using MCMC methods. The results were very similar to the HMMs but were more challenging to fit and required much more processing time.

In the absence of informative prior information, the Bayesian method does not provide any improvement on the HMMs. The HMMs perform slightly better in terms of state allocation accuracy than the Mixture Models. However Mixture Models perform acceptably if only a straightforward clustering of the observations is required. However, if a more robust method is required, the HMMs are relatively easy to fit and extend, allow for investigation of the state switching probabilities and are recommended as the best method for analysing this type of data.

ACKNOWLEDGEMENTS

I would like to express my immense gratitude to my PhD supervisors Prof. Paul Fatti and Prof. Norman Owen-Smith. It has been a privilege to work with you both and despite being a "long-distance" student, you have given me tremendous support, guidance and encouragement throughout this PhD journey. Thank you!

I would also like to extend special thanks to my former MSc supervisor, mentor and Head of Department Prof. Sarah Radloff from Rhodes University. You have always been an amazing role model and teacher and without your support and encouragement I would never have got this far. Also from Rhodes University, I would like to thank Dr Lizanne Raubenheimer and Jeremy Baxter for their friendship and support over the years. I would like to thank Prof. Iain MacDonald from the University of Cape Town with whom I spent many hours discussing the Hidden Markov models and who was always willing to assist me with my research in this field. Thanks also to Prof. Walter Zucchini who also provided some valuable insight and assistance. Thanks must also go to the external examiners who provided valuable comments on this thesis.

Special thanks also to a number of people who have provided guidance, ideas and support through this study over the years:

My colleagues at Synovate South Africa: Dr Jannie Hofmeyr and Paul Holtzmann; and the South African Environmental Observation Network: Dr Nicky Allsopp and Johan Pauw.

Colleagues from the University of the Witwatersrand staff: Prof. Peter Fridjhon, Mark Dowdeswell, Prof. Gordon Kas, Dr Joe Chirima, Linda McRae-Samuel and Maud Manuel.

Others: Prof. Wayne Getz, Dr Steven Bellan, Dr Leo Polansky, Dr Rico Holdo, Dr Roland Langorck, Dr Sam Ferreira and Dr Juan Morales.

Then on a personal note, I would like to thank friends who have supported me over the years of this PhD. Pamela Jayne Stretton Palmer, Gail Halforty and Riette Vorster - thank you for your friendship and understanding when the times were tough and for pouring that glass of wine when it was needed. To Ernita van Wyk, it has been invaluable going through this PhD process together and having someone who understands the ups and downs. Hang in there, your PhD will soon be done! To my fiancé, Peter Allnutt, thank you for your love and support during this process and putting up with my nuttiness along the way. And then finally to my family, Heather and Eric Goodall, the most special thank you for everything you have given me throughout my life. You instilled a love for the bush in me, which has been an inspiration during this study. You have always loved me, supported all my academic endeavours and supported me every step of the way. And of course, thank you for providing the “coat” catalyst for setting out on this study.

TABLE OF CONTENTS

LIST OF FIGURES

TABLE OF TABLES

1 CHAPTER ONE – INTRODUCTION

1.1. General Introduction

Movement ecology has been defined as an area of research which studies "all types of movement involving all organisms" (Nathan 2008). The movement of any organism is a fundamental part of its biology and life (Nathan et al. 2008) and can range from tiny movements of very small organisms to long range movements such as the albatross or the famous mass-migration in the Serengeti. Understanding these movements in terms of how, why, when and where the animals move will be important in the long-term conservation effort for these animals.

Our ability to track the movements of animals has advanced rapidly in recent years. It can take the form of actually observing the animal, radio tracking and now more recently using GPS technology to obtain regular location coordinates for the tracked animal. The use of GPS tracking has provided a huge improvement in the tracking of an animal. No longer does a human observer need to get within a short distance of the animal as with observation and radio tracking, and the GPS units can record regularly spaced observations throughout the day and the night. As technology has improved, the range of species which can be tracked has increased as GPS tracking units have reduced in size and the time intervals between locations has decreased as battery power has increased (Rahimi, Owen-Smith 2007). This means that it is now possible to track animals over longer periods of time, spanning the seasons and multiple years. Other remote sensing techniques are now being used in conjunction with the GPS tracking data. For example activity collars which can sense whether the animal's head is up or down, internal temperature recorders placed under the skin of the animal to record the internal body temperature, NDVI (Normalised Difference Vegetation Index)

measurements which monitor the greenness of the vegetation, and time depth recorder loggers used to estimate the drift rate of seals as a measure of the body condition of the animal (Dragon et al. 2012) among many others.

However, this wealth of tracking data brings with it associated challenges. As the time intervals between successive locations decrease, the strength of the autocorrelation will increase both in space and time. Locations from the same individual are unlikely to be independent and are more likely to have a complex autocorrelation structure which creates a complex time-series of movement behaviours (Eckert et al. 2008, Fauchald, Tveraa 2003, Fauchald, Tveraa 2006, Jonsen, Flemming & Myers 2005, Morales et al. 2004, Friar et al. 2005, Sibert et al. 2006). Very large datasets can be recorded which provides some challenges in terms of storage and processing of the data, although with modern computers, this is far less of a problem. Numerous tracking studies are being conducted on a vast array of species around the world, and a major challenge is to determine the most appropriate statistical framework for analysing these data. Various branches of analysis exist for these types of data, including home range analysis and the investigation of the underlying behavioural states which are driving the observed movements. The ability to analyse animal tracking data has lagged behind ability to collect these types of data (Jonsen, Myers & Flemming 2003).

1.2 Aims and Objectives

The overall aim of this project was to establish the best statistical framework (or frameworks) for analysing and interpreting datasets obtained by GPS tracking on the movements and ecological relationships of large mammalian herbivores, using data from the Kruger National Park, South Africa. In order to achieve this aim, the objectives were to understand the following aspects of the movements of sable antelope (*Hippotragus niger*), buffalo (*Syncerus caffer*) and plains zebra (*Equus quagga*):

1. To compare and contrast the suitability of different statistical modelling approaches for analysing these data.
2. To differentiate the movement patterns of the three ungulate species.
3. For each species, to distinguish movement within foraging patches from movement paths between these patches.
4. To establish how these movement patterns within and between patches vary over the seasonal cycle.
5. To reveal unusual movement patterns that might indicate stressful conditions.
6. To customize and improve, if possible, the statistical models and techniques in order to be more accurate, according to appropriate statistical criteria, in describing the movement patterns of the three species.

Common time-series techniques do not adequately explain the movement patterns due to the complexity of the data and the non-constant and possibly non-Gaussian estimation errors (Jonsen, Myers & James 2006). This has meant that more complex models are being applied to ecological data in order to adequately explain the movement patterns.

1.3 Research Outline

In Chapter 2 the background of the study is presented, along with a description of the species studied and the regions in which they live. The analyses in this thesis are based on the linear displacements between successive GPS locations, also known as the 'step-lengths'. This is a commonly used metric in the analysis of animal movement data. It is often used in conjunction with the angle between the GPS locations, known as the 'turning angle'. The effect of including the turning angle in the analyses is investigated in Section 6.3.2, although for the majority of the analyses only the step-length was used. This is similar to the analyses presented by Forester et al. (2007) and Owen-Smith, Goodall & Fatti (2012). This

step-length is used as the input to fit models which allow inferences to be made about the underlying groupings within the movement data assumed to correspond to the behavioural state of the animal. A common complication with any remote sensing data collection technique is accounting for the observation error. In Chapter 2, the error associated with GPS observations is examined for collars similar to those used to collect the data for this study. Basic summaries and exploratory data analyses of the study data are also presented.

Although the lack of independence of the successive movements is acknowledged, this assumption is ignored in Chapter 3 and Independent Mixture Models are fitted to the observed displacements. This method provides a simple clustering technique to segment the observed observations into discrete groups based on the latent state. Each state is assumed to correspond to a different movement type which is inferred to represent the behavioural state of the animal during the time period of the observation. This analysis is introduced by first examining where breakpoints occur in the distribution of the linear displacements. Issues of the goodness-of-fit and model selection for the independent mixture models are investigated, as well as a simulation study to access the accuracy of the model selection and parameter estimates. Models are then adapted in order to fit seasonal models to investigate the ability of the approach to pick up changing behaviour over the seasonal cycle. Different distributions (Log-normal and offset Gamma) are used to model the displacement data which allows for different interpretation of the results in terms of how the time periods corresponding to the linear displacements are allocated to the different behavioural states using either maximum likelihood or proportionate allocation.

Since the assumption of independence was violated in Chapter 3, Chapter 4 investigates the strength and the patterns of autocorrelation between successive movements across the three species. These patterns differ by species and by season, but are strong across all of the herds and hence should be accounted for in

the modelling process. This leads into Chapter 5 which uses Dependent Mixture Models, also known as Hidden Markov models to fit the observed movements. The general process is very similar to the Mixture Models in that the observed movements are clustered and the latent states are assumed to correspond to the behaviour of the animal. However, the autocorrelation is now accounted for in the model using a Markov process and the transitions from one state to another are now formally modelled via the state transition matrix. Once again, the goodness-of-fit and model selection are investigated as well as the different fitting approaches using either direct numerical maximization or the Expectation-Maximization (EM) algorithm. When the observed data are modelled as a time series, missed locations need to be included in the modelling process and the fitting methods need to be modified to cope with these missing data. The Hidden Markov models are fitted using the displacement between hourly observations, and also the 24 hour displacements anchored at various times during the day and night.

The basic Hidden Markov models are then extended to account for seasonality and irregular time gaps between successive GPS locations. Hidden Markov models seem to be gaining in popularity in their application to animal movement data and have been fitted to data from a variety of species. Chapter 6 applies some novel extensions to the basic model which appear useful for their application to these types of data which often span the seasonal cycle and may be obtained at irregular locations to extend the battery life of the tracking device or focus on movements at certain time of the day.

Due to the advancement of computing power and the accessibility of programing software to fit models using a Bayesian approach, Bayesian state-space models are a common technique applied to animal tracking data. The basic framework is similar to that of the Mixture and Hidden Markov models and is investigated for the three species in Chapter 7. Non-informative prior distributions are used for the

analyses, so the results are expected to be very similar to those obtained for the frequentist fitting approach.

Chapter 8 is a comparison chapter where the three fitting techniques are compared in terms of the inferences obtained from the models, the accuracy of the parameter estimates and the time taken to fit the models. All three techniques have their own strengths and weaknesses for their application to these movement data, and these are compared and their suitability discussed for movement analyses. All the findings from this study of these statistical techniques and their application to movement data from three large herbivore species are summarized in a short conclusion in Chapter 9.

1.4 Contribution of the Present Study

Three statistical methods were selected for this application to animal movement data obtained using GPS tracking collars. These methods were selected as they have been used in other movement research, are reasonably simple analyses and provide output that is simple to interpret in terms of the behaviour of the animal. Throughout the work conducted in this study, the desired statistical outcome was to identify which methods are the most effective, why they are preferred and to identify what the differences are between the models. The choice of model can have an impact on the output and the subsequent inferences and these need to be understood at the outset of a movement study. Each of the models investigated are based on slightly different assumptions and vary in complexity. From the biological perspective, it is important to find statistical models which do not only provide a theoretical fit to the data, but they also need to provide easily interpretable output which is realistic in terms of the biology of the species. These results and inferences are then intended to be integrated with other ecological studies which make use of the movement data to answer much wider ranging ecological questions.

The need for this study has come from the fact that the ability to collect animal movement data has outpaced the ability to statistically analyse the data. GPS tracking data are expensive to collect, and basic analyses do not do justice to the data. The output from these analyses are needed to link into more complex ecological studies of the animal as GPS tracking is opening up new areas of study into the behaviour and ecology of the tracked animal. The need for a better understanding of movement processes is acknowledged, particularly in the light of global environmental changes and issues such as disease outbreaks and biological invasions (Nathan 2008). Statistical modelling approaches need to be advanced and adapted in order to cater for these developing areas of science. Four data challenges were identified by Nathan et al. (2008). These were a) splitting the movement path into a string of unique units which is now made easier through the use of GPS locations which can be used to calculate the step-lengths and turning angles making up the movement path; b) classifying these step-lengths and turning angles into movement phases and in c) these movement phases need to be decomposed into the behaviours which generate them. For example, the behaviour of running could be associated with a movement phase of escaping from a predator, and walking with feeding as a searching for food phase. The fourth challenge identified by Nathan et al. (2008) is to integrate the movement data and analysis within an environmental context using metrics and techniques such as environmental covariates, remote sensing and GIS to enhance the analyses. All four of these challenges require statistical analyses and this study focuses on techniques and ways of improving the analyses to achieve the outcomes of the second and third challenges. The state-space approach has been suggested as the likely method to gain in popularity as the resolution of movement tracking data and the ability to collect landscape data improves with advances in technology (Getz, Saltz 2008).

This study can be considered a “second generation” movement study, since a number of movement analyses have already been published using a wide variety of techniques. Each of the methods used in this study have been used in other

movement research. However, very few studies have compared and contrasted the results from a variety of methods in order to understand the impact that the choice of analysis can have on the results of the study. The study also adapts and customizes one of the methods in order to make it more suitable for this specific application to movement data.

2 CHAPTER TWO – UNGULATE MOVEMENT IN THE KRUGER NATIONAL PARK

2.1 Introduction

GPS tracking is becoming a widely used remote sensing technique to monitor the movements of a wide range of species, both terrestrial and marine. Data for this study were collected using GPS collars placed on sable antelope, buffalo and zebra herds in the Kruger National Park. This chapter provides background information on the species studied, the regions they live in and some exploratory analyses of their movements. The errors associated with GPS devices are a challenge for analysing these types of data, and the magnitude of the errors associated with these movement data are investigated.

2.2 Background

The Kruger National Park (KNP) was proclaimed in 1898. It is situated in the north-eastern corner of South Africa in an area known as the Lowveld. It is approximately 19 685 square kilometres in extent, almost the size of Wales. It is bordered by the Lebombo mountain range in the east which also forms the South African border with Mozambique. It stretches between the Crocodile River in the south and the Limpopo River in the north. The KNP is home to around 507 bird, 114 reptile and 147 mammal species.

Figure 2.1 - Map of the Kruger National Park with the Punda Maria (top left) and Pretoriuskop (bottom left) regions circled to indicate the areas where the data were collected. (Downloaded 6th May 2011 - http://www.sanparks.org/images/parks/kruger/maps/originals/knp_map.JPG)

Remote sensing is becoming a common technique for monitoring many ecological phenomena. As technology is improving, the scope of remote sensing is widening and new innovative techniques to monitor, measure and track things are becoming

available. GPS tracking of animals is becoming a far more accessible technique used to monitor the movement and behaviour of animals in the wild. GPS locations are now more accurate as more satellites are available to connect to in order to obtain a location fix, and the technology within handheld/small GPS devices is improving. VHF tracking of animals has been used in the past to study the movement and behaviour of animals in the wild. However, this tracking is ad hoc and dependent on being sufficiently close to the animal. GPS collars now allow for regular locations of the animal to be recorded, no matter what terrain the animal is in. They also are able to obtain night time locations which were difficult to obtain in the past, especially for areas with nocturnal predators. It also means that it is possible to track animals which are human-shy and the observer does not influence the behaviour of the study animal. However, as the tracking technology improves, it enables the collection of many more location estimates with far shorter time intervals between them. Recently more GPS devices have been used in animal tracking (Graves, Waller 2006). This increased observed time series of locations provides challenges for the analysis of these data. Regular locations will have strong autocorrelation, at both the spatial and temporal scale. Many techniques used in the past for the analysis of animal movements did not take into account the serial correlation between successive observations. Many of the studies using GPS tracking data have been based in the northern hemisphere, with very little published on the movements of southern African herbivores. A statistical framework and baseline results from fitted models were required for southern African ungulates, which can then be compared to future studies of similar species or similar regions.

The rare antelope populations in the KNP have declined dramatically in the past decades (Ogutu, Owen-Smith 2003). Sable antelope numbers dropped from 2240 in 1987 to around 550 in 1995 (Chirima 2009). The populations of roan antelope (*Hippotragus equinus*) and tssessbe (*Damaliscus lunatus*) also declined alarmingly. These declines were particularly evident in the far northern areas of the Park, where buffalo and zebra seemed to be thriving in the same areas. These

two species have similar resource requirements to the rare antelope. The Northern Plains Programme was started to investigate and monitor the changes that were happening in the north of the Park which was the preferred habitat of rare species such as roan antelope, tssessbe, reedbuck (*Redunca arundinum*) and eland (*Taurotragus oryx*) (Joubert 2007a). One of the suggested causes of these population declines was the expanded surface water availability due to artificial water points (Chirima 2009). This water availability opened up areas to high-density species such as zebra and wildebeest (*Connochaetes taurinus*) that were now able to use these areas for foraging without having to migrate in search of water (Joubert 2007a). A number of artificial water points have been closed in these regions however the populations of these low-density antelope have not recovered (Grant, van der Walt 2000). Further research and understanding was required of the ecology and behavioural patterns of these rare antelope species. A research programme was undertaken by the Centre for African Ecology at the University of the Witwatersrand which aimed to understand the ecology of these rare species. GPS collars were placed on sable antelope as well as buffalo and zebra which were hypothesized to be competing with the sable for resources. A number of studies have been done using these GPS data (Cain, Owen-Smith & Macandza 2012, Owen-Smith, Goodall & Fatti 2012, Macandza, Owen-Smith & Cain 2012a, Macandza, Owen-Smith & Cain 2012b, Owen-Smith 2013) to investigate various ecological hypotheses. However, a gap existed to apply and extend statistical modelling techniques which could be used to provide a baseline model for the movements of these species.

Rapid advancement in technology has now made satellite tracking of the movements of animals possible without human interference. GPS/GSM collars were placed on sable antelope, buffalo and zebra in the far northern part of the KNP while three sable herds were tracked in the south-western region. The buffalo and zebra are high-density species and were known to utilize the same areas as the sable.

2.2.1 Study Species

Sable antelope are large ungulates which can weigh approximately 230kgs. The bulls are shiny black with a white underbelly. Cows and younger bulls are a lighter brownish colour. A distinctive characteristic of the sable antelope are the large horns which sweep backwards in a very pronounced curve.

Figure 2.2 - Sable antelope (Photo - Jacqueline Viljoen)

Sable antelope are normally found in the wetter regions of Africa, however the KNP is the southern extremity of their distribution (Chirima 2009). Breeding herds consist of females and juveniles. Often females will remain with the herd into which they were born, while young males will be chased out by the territorial bulls (Apps 1992).

Buffalo are large ruminant grazers, with distinctive horns. They live in large herds which can number more than 1000 animals. Buffalo herds can live in a variety of different habitats and in the KNP they prefer to utilize areas close to water if forage is available (Redfern et al. 2003). They are said to do most of their grazing at night, in the early morning and late evening (Apps 1992). Typical of ruminant

behaviour they will switch between ruminating and foraging throughout the day, and these states can be influenced by temperature and resource availability (Polansky et al. 2010). Due to their large size and the large groups that they live in, buffalo can have a dramatic impact on the vegetation, particularly when food is limited in an area (Joubert 2007b).

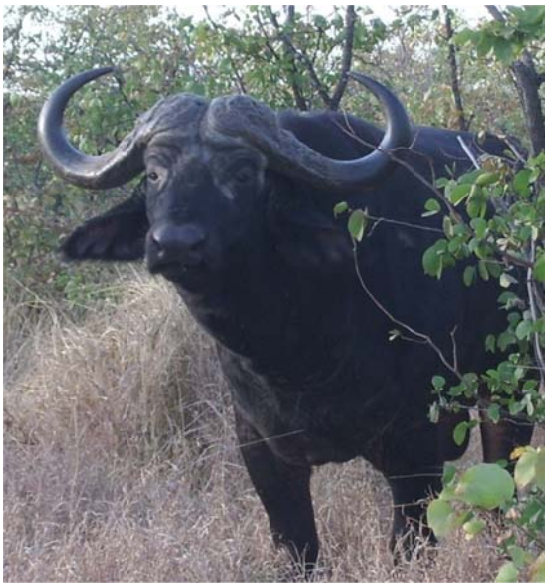

Figure 2.3 - Buffalo (Photo - Victoria Goodall)

Zebra are plentiful within the KNP, with only impala (*Aepyceros melampus*) being more numerous. Individuals show great variation in colouring and patterns of stripes. Zebra are often seen together with wildebeest as they both favour short grass (Apps 1992). They tend to live in groups of around 10 animals.

Figure 2.4 - Zebra (Photo - Victoria Goodall)

The distribution of sable in the KNP is smaller than that of the buffalo and zebra due to their lower numbers. However the buffalo and zebra will mostly utilize the same areas as the sable. How they utilize the areas relative to one another can be examined through these GPS tracking data.

GPS/GSM collars (Supplier: Africa Wildlife Tracking, http//www.awt.co.za) were placed on females from sable, buffalo and zebra herds in the north-western region near the Punda Maria restcamp. These herds are referred to as the Punda Maria herds. Three sable herds were tracked in the south-western region of the KNP, near the Pretoriuskop restcamp. These herds utilize separate regions which do not overlap. These herds are referred to as the Numbi, Phabeni and Shitlave herds, see Figure 2.5.

Figure 2.5 - Map of the Kruger National Park showing restcamps and animal locations near Punda Maria (top left) and Pretoriuskop (bottom left)

The vegetation in the Punda Maria region is characterized by *Colophospermum mopane*/*Acacia nigrescens* savanna and *Colophospermum mopane* forest. The vegetation in the Pretoriuskop region includes 'Lowveld Sour Bushveld of Pretoriuskop' and '*Combretum collinum*/*Combretum zeyheri* Woodland' (Gertenbach 1983). The Punda Maria region receives lower annual rainfall compared to the Pretoriuskop area, which is also at a higher altitude (Joubert 2007b).

2.2.2 GPS/GSM Tracking

GPS/GSM collars are able to store location data while the animal is outside of cell phone communication range and all these stored location coordinates will be downloaded to a website (www.yrless.co.za) once the animal comes back within range. The collars are costly and when added to the cost of capturing the animal, it is an expensive technique for obtaining data. For studies such as this one, the

game capture process requires veterinarians, game capture experts, helicopter pilots and many other support crews to facilitate the capture. The capture process itself is stressful and dangerous for the animal. Even once the capture process is complete, the collar installed and the antidote drug administered, the animal still needs to recover from the drugs, avoid predators and meet up again with their herd. GPS technology has enabled researchers to record the sequential movement patterns of animals and compile datasets with many observations over a relatively short time period (Rahimi, Owen-Smith 2007). This advancement means that datasets are much larger and there is strong autocorrelation between the observations. These are some of the new challenges which need to be addressed by the statistical models applied to animal movement data using GPS.

Figure 2.6 - Sable antelope with a GPS/GSM tracking collar (Photo - Jacqueline Viljoen)

2.3 Challenges Associated with GPS Tracking

For every location fix obtained, the unit records the date and time of the observation and the longitude and latitude coordinates of the location point on the earth's surface. The longitude and latitude can be used to calculate the straight-line displacement between two locations. The 'Great Circle Distance' can be used. It is given by (http://en.wikipedia.org/wiki/Great-circle_distance):

displacement = *r**(acos(sin(lat1)*sin(lat2)+cos(lat1)*cos(lat2)*cos(lon1-lon2)))

where lat1,lon1 = latitude and longitude at time 1

and lat2, lon2 = latitude and longitude at time 2

and *r* is the radius of the earth.

More recent GPS collars also record the PDOP (Positional Dilution of Precision) values which are a measure of the position of the usable satellites which estimated the location. It is a measure of the accuracy of the location estimate, and indicates the magnitude of the error which is likely to be associated with the observation. The R package 'adehabitat' (Calenge 2006) uses the function `ltraj` to calculate the displacement between successive locations using UTM (Universal Transverse Mercator) coordinates. This method can be used by converting the longitude and latitude coordinates to UTM coordinates and applying the `ltraj` function to determine the displacement distance and the turning angle between successive locations. Both this and the Great Circle method provided almost identical results and can be used to prepare the input data for all of the models.

2.3.1 GPS Error

As with all remote sensing techniques, there is a possibility of erroneous location estimates being obtained. One of the challenges associated with the analysis of GPS data is that there is an inherent error in all GPS location estimates, as is the case with any measurement data. The accuracy of GPS estimates using handheld

devices similar to those used for the collars is improving. However every location estimate will have some error associated with it. This can be dependent on the type of GPS unit and the number of satellites that it has been able to connect to when obtaining the location estimate. Some tracking systems such as ARGOS (often used for marine and terrestrial animal tracking in the USA) provide a quality class for each location estimate based on the accuracy of the estimate of the location (Cushman, Chase & Griffin 2005). ARGOS systems use Doppler based positioning associated with larger error but provide the only automated reporting data available in North America (Cagnacci et al. 2010, Tomkiewicz et al. 2010). Some clearly erroneous observations are occasionally recorded. In this study these can be seen when the location points occur outside the boundaries of the KNP or where the animal appears to have moved an excessive distance in the time period, and then returns close to the original location in the next movement. Displacements of more than 10km in one hour were manually removed from all the datasets.

Another source of error within the data is when locations are not obtained so the location estimate is missing. This is caused by the GPS unit being unable to connect with sufficient satellites to obtain a location estimate within the time allowed. This time is set by the manufacturer, and provides a trade-off between a missed location and the battery drain of continuously searching for satellites. This could be influenced by the terrain and vegetation cover where the animal is (Ott, van Aarde 2010) and whether the animal is lying down at the time when the collar attempts to take a reading. This has the potential to introduce biases into the observed data, as the missing data are not random and could actually be informative. Changes in the frequency of observation also result in missed locations when the data are consolidated into a set of regularly spaced observations. This is a feature of the data used in this study since the timing of the observations was altered during the study. At the outset, the collars were programmed to obtain locations every 6 hours, this was changed for certain periods to obtain intensive hourly observations which could then be used in

conjunction with fieldwork in the areas used by the animal. The collars were originally set to record locations only every 6 hours in order to save battery power and to extend the period of time over which the animal could be tracked over a number of seasons. As many statistical time series analysis techniques require regularly spaced observations, datasets need to be created with regular time intervals. For a time series of the hourly locations, these periods of six hour estimates will show as missing data. Although some studies of animal movement use various interpolation techniques to estimate the missing locations, the aim in this study was to rather let the model take into account the missing data rather than estimating the locations. This is particularly important as the bouts of missing data due to the changes in the observation frequency would make estimating the locations unreliable. Models applied to these data either need to incorporate the missing data in the model fitting process or allow for irregularly spaced observations, which required customization of models.

In order to investigate how much error is associated with each obtained location estimate, several collars were left stationary in trees for variable periods to obtain a series of observations from exactly the same point. These data were presented in the supplementary material S2 in Owen-Smith, Goodall & Fatti (2012). These collars had been used for tracking the movements of lion or wildebeest in the central region of the KNP. Location estimates were obtained over the course of a few days or weeks during which the collar was not moved. Unfortunately no high precision differential GPS was used to obtain a true (more precise) location estimate, so the 'true' location remains unknown. The average longitude and latitude coordinates were used as an estimate of the true location and the distance from each observed location to the central point was calculated using the Great Circle formula. Static locations were obtained for four different collars with sample size n, represented by the codes AG035 (n = 167), AG116 (n = 513), AG196 (n = 178) and AG335 (n = 74). The PDOP values were recorded for these 4 collars, but were not included in the investigation of the location accuracy since these data were not available for the sable, buffalo and zebra collars. The observed

location points and histograms of the displacements from the average location are shown in Figure 2.7 to Figure 2.10.

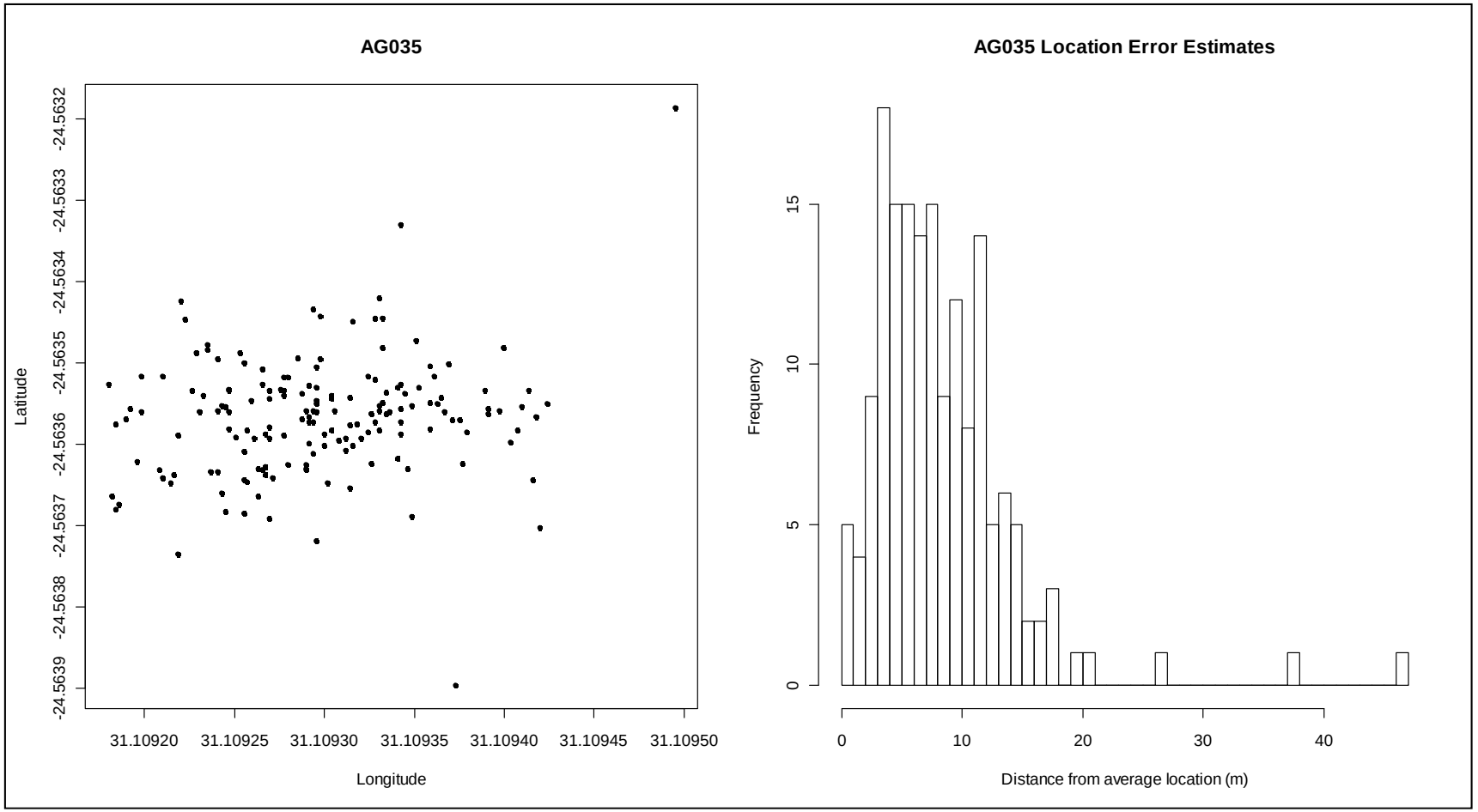

Figure 2.7 - Error estimates from collar AG035

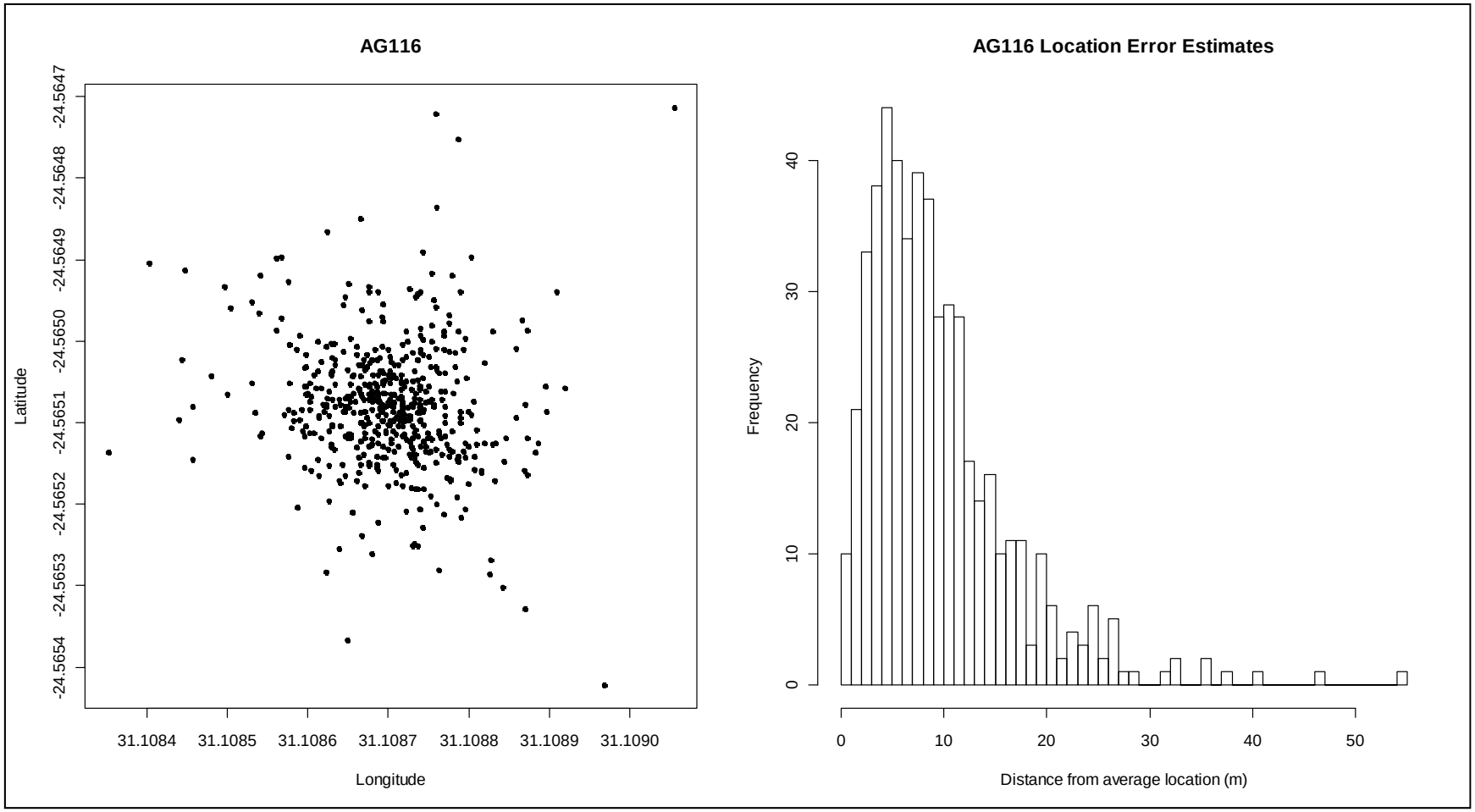

Figure 2.8 - Error estimates from collar AG0116

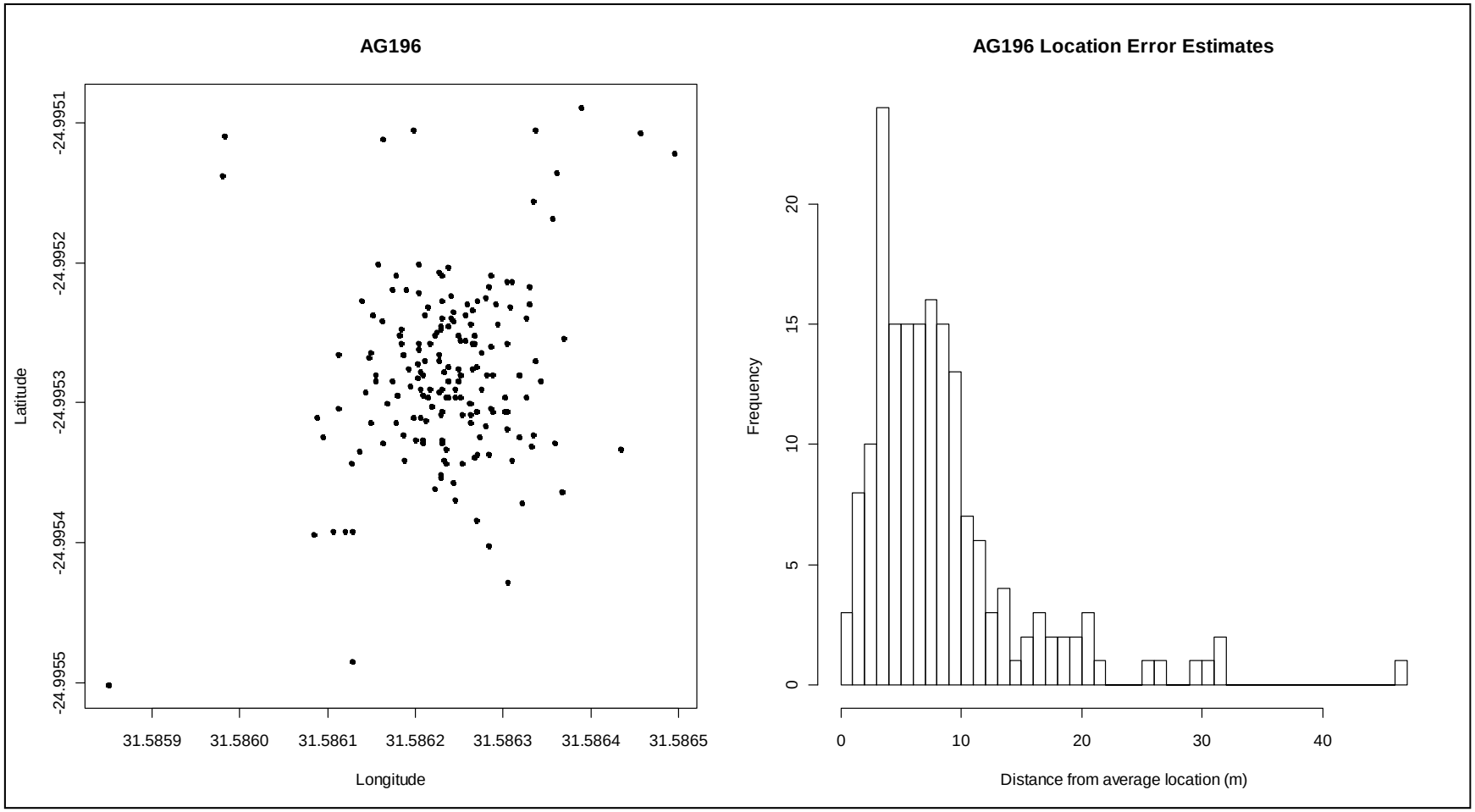

Figure 2.9 - Error estimates from collar AG196

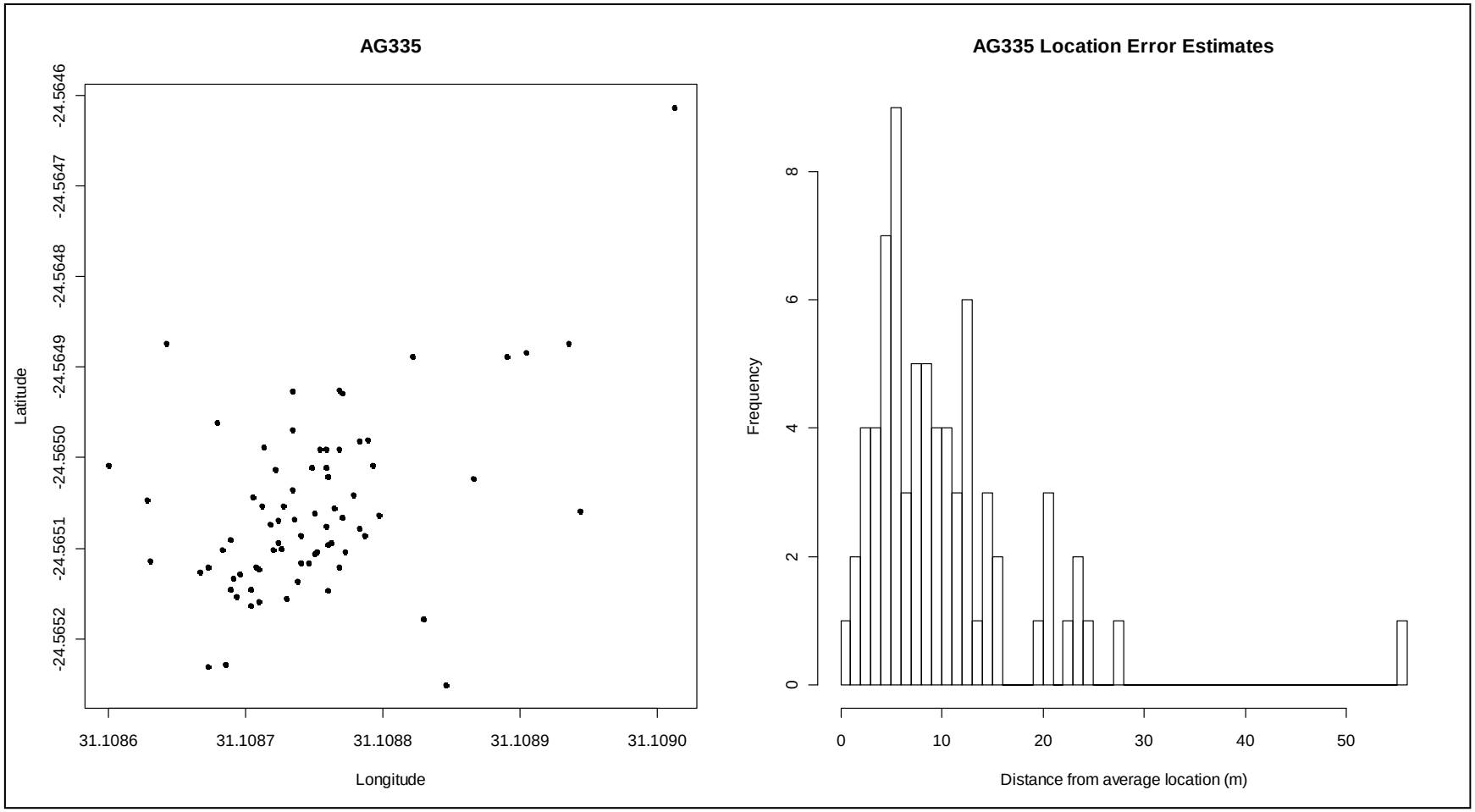

Figure 2.10 - Error estimates from collar AG335

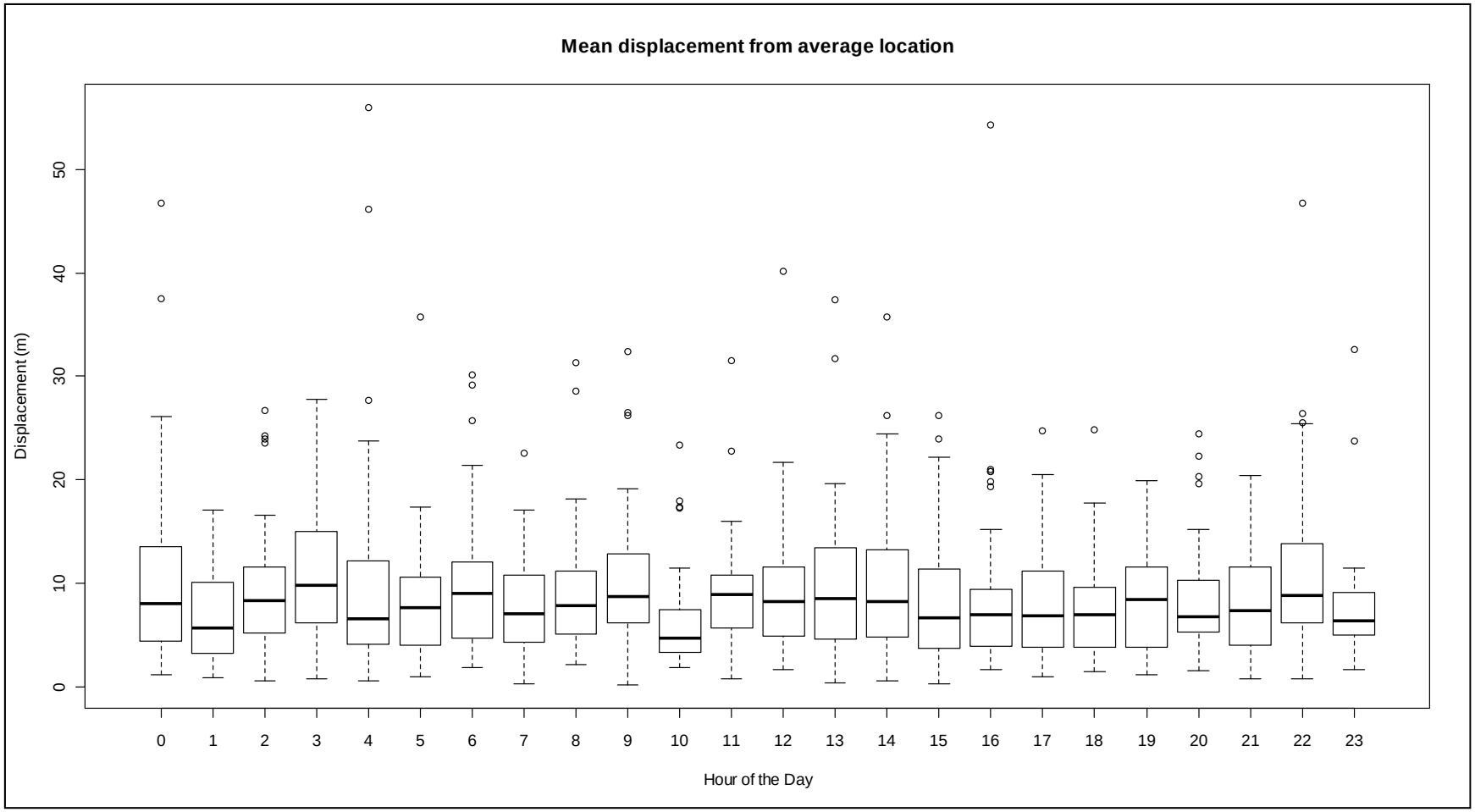

Figure 2.11 - Mean displacement from the average location by time of day - Lion collars

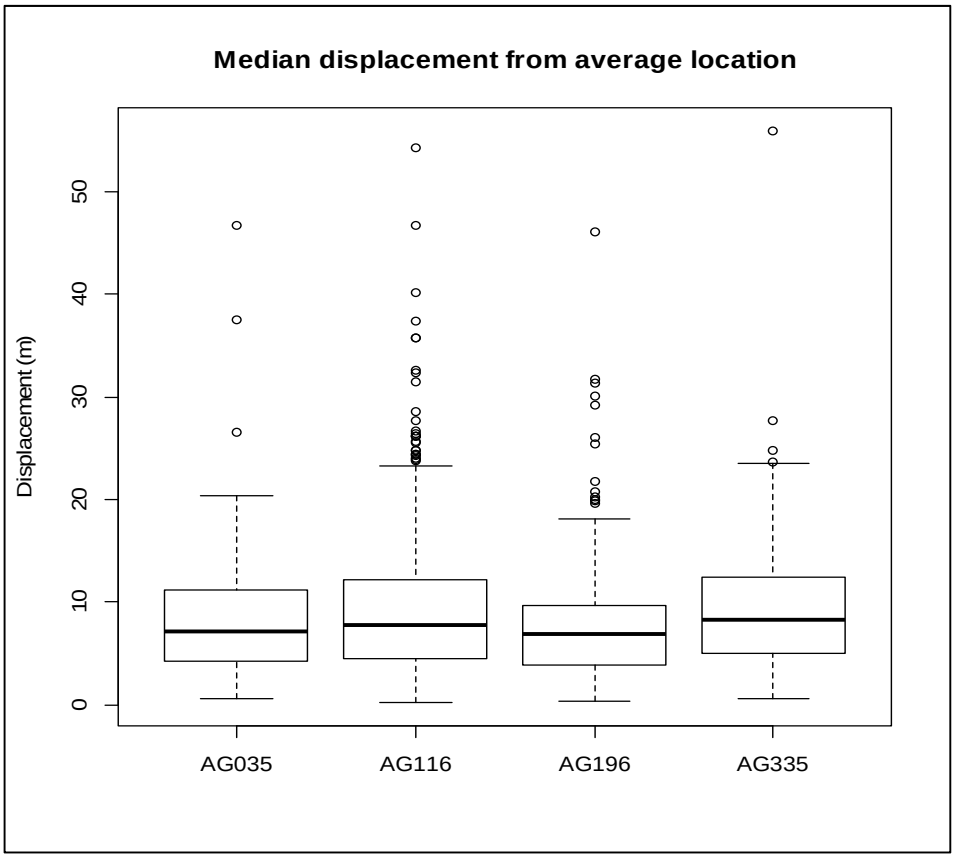

Figure 2.12 - Boxplot showing median displacement from average location - Lion collars

The mean displacement from the average location ranges from 8.22m to 10.19m for the four stationary collars. From Figure 2.12, it seems that all 4 collars have similar error medians as well and Figure 2.11 does not show any particular pattern in the error displacement given the time of day at which the location estimate was obtained. While this analysis shows the spread of estimated locations and error distances for the observed locations, the average location is not necessarily the true location and in fact it will be biased by the outlier observations. One outlier observation will shift the average location in its direction and hence bias the estimate of the average location. Outliers can be seen in the corners and along the boundary of the plots in Figure 2.7 to Figure 2.10. In order to obtain more conclusive results about the true measurement error of the collar, highly accurate differential GPS coordinates for the exact location are needed. Also the test collar would need to be placed in a variety of locations and cover types to determine the

impact of other variables on the accuracy of the measurements. Errors will be part of any measurement including remotely sensed locations. For displacements with an hour interval between them for a large mobile ungulate, over the course of a study errors of approximately 10m will be irrelevant. However, if a study was done using very short intervals between observations with much shorter movements, the errors would be more influential and might need to be modelled during the analysis.

Figure 2.13 - Lioness with tracking collar in the Addo Elephant National Park (Photo - Victoria Goodall)

The errors associated with each observation mean that each displacement between successive locations will either be slightly under- or over-estimated. However, it is assumed that these errors would be randomly distributed around the true location and independent of one another. The error distribution of the displacement distances would be the same, whether the observed displacement was large or small.

The output from the Great Circle formula or other method of calculating displacement between location coordinates provides displacement estimates with many decimal places, due to the trigonometric functions used in the calculations. Although these decimal places imply greater precision in each location estimate than would be true due to the precision of the GPS device, a decision was made not to round off or round-up the displacement distances. The aim of this study is to model the movement of the animal, any rounding or truncating of the input data has the danger of biasing the output data. The observed displacement is a continuous variable, and rounding of the displacement has a binning effect which discretizes the observations. If a continuous probability function is used to model these data, the effect of the binning will influence the model output. Models were tested using various rounding options, but the most consistent results were obtained using the unrounded input. This method is also more replicable with future studies and does not require justification of the choice of rounding off. In some cases identical location estimates were obtained for successive location estimates which results in a displacement distance of exactly zero. It is thought that these identical locations occur when the GPS unit is unable to connect with sufficient satellites to obtain an accurate location and the previous location is used to estimate the current location. Since most distributions used to model the displacement distances (e.g. Gamma, Weibull etc.) are only valid for positive displacements, these exact zeros are problematic. Given the inherent error associated with each observation, it is very unlikely that identical observations at successive intervals are true zeros and the animal has not moved at all during the hour. These zero displacements were replaced with a value of 1m for the modelling process. Very few cases of exact zero displacements occurred. Some datasets did not have any identical locations for successive observations.

While it might seem like a trivial issue of how to round off observations, GPS tracking data with short inter-observation intervals are characterized by a number of very small observations. Based on whether these are rounded up or rounded off, the shape of the distribution of observations will be affected. Histograms of

the observations less than 20m for the Punda Maria sable are shown in Figure 2.14. These histograms show how the decision about how to round the observation data can change the distribution pattern for the smallest observations. Although this might seem trivial for measurement data which includes error anyway, the fit of any model in this region would be influenced by the decision of how to round the data. To avoid making an arbitrary decision, it was decided that it is better to fit all models to the data using the raw observations with all the decimal places, even though it is acknowledged that the accuracy of the GPS is not at the level of the decimal places obtained.

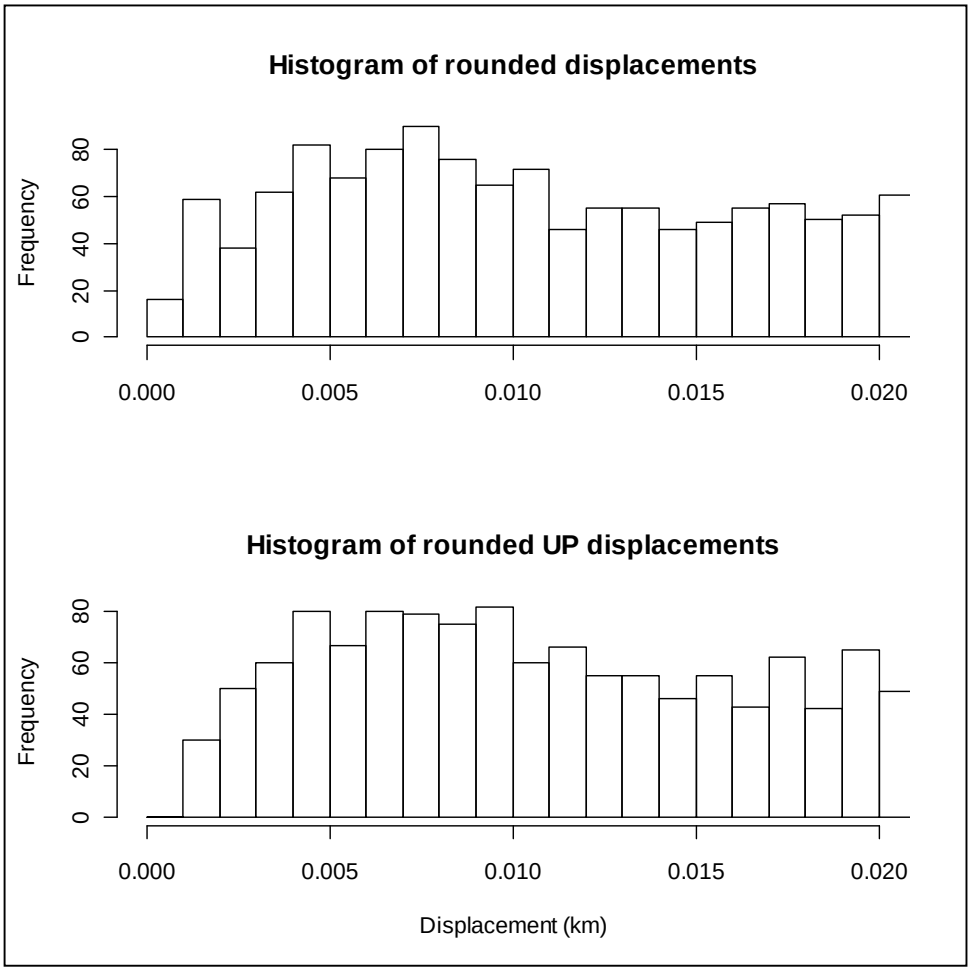

Figure 2.14 - Histogram showing the different distribution pattern of observations for rounded off and rounded up displacements - Punda Maria Sable

2.4 Daily Analysis

GPS tracking data can be scaled up to calculate displacements over longer time intervals to investigate patterns at different temporal scales. For the analyses in this study, models were fitted to the hourly displacements as well as to daily displacements, calculated as the displacement between the current location and the one 24 hours before. Daily displacements were anchored at 8pm, 2am, 8am and 2pm to coincide with the timing of the 6 hour location recordings to maximise the available data. These times were chosen to coincide with various activities for the animal. For example, the 8pm location would represent the location at the end of the afternoon foraging session. This would allow investigation of how these species exploit foraging areas, without being confused by movements to water if they occurred during that day. The analysis of the hourly data was able to investigate the behavioural patterns and activity states of the animals. However the daily displacement allows investigation of the patches utilized by the animal at the time of day the locations were anchored (either 2am, 8am, 2pm or 8pm) and identification of when the animals shifted their locations from one patch to another.

2.5 Kruger National Park – Sable antelope, buffalo and zebra

In June 2006 GPS/GSM collars were placed on adult female sable usually in the same herd, two adult female buffalo in separate herds which often joined up, and four zebra herds in the Punda Maria region. Three adult female sable were also collared in three different herds in different areas of the Pretoriuskop region. Details of the collars, date range and number of observations for each animal are shown in Table 2.1.

Table 2.1 - Details of GSM/GPS collars used in the Kruger National Park study

Species	CollarID	Location	Start Date	Last Date	Number of observations
Sable	AM143	Punda Maria	22/05/2006	18/12/2007	379
Sable	AM146	Punda Maria	23/05/2006	Died 11/02/2007	905
Sable	AM149	Punda Maria	23/05/2006	15/07/2007	547
Sable	AM279	Punda Maria	18/06/2007	09/01/2009	3 499
Sable	AM140	Numbi	24/05/2006	Removed 18/08/2007	1 073
Sable	AM285	Numbi	18/08/2007	27/03/2008	4 730
Sable	AM148	Phabeni	24/05/2006	Removed 18/08/2007	1 336
Sable	AM284	Phabeni	18/08/2007	21/03/2009	12 749
Sable	AM151	Shitlave	23/05/2006	Removed 18/08/2007	926
Sable	AM286	Shitlave	18/08/2007	29/09/2008	8 745
Buffalo	AM150	Punda Maria	23/05/2006	11/03/2007	1 063
Buffalo	AM152	Punda Maria	23/05/2006	29/06/2009	14 670
Zebra	AM141	Punda Maria	23/05/2006	29/09/2007	999
Zebra	AM142	Punda Maria	23/05/2006	06/11/2008	10 679
Zebra	AM277	Punda Maria	18/06/2007	06/06/2010	22 223
Zebra	AM280	Punda Maria	18/06/2007	08/09/2008	4 508

The movements obtained for the collared animal are assumed to represent the movements of the herd since the females form the nucleus of the herd structure. Some of the collars were replaced when the batteries died and hence they have more than one collar code given in Table 2.1. The two buffalo collars were initially deployed on two separate herds, each numbering about two hundred animals, however these herds then joined up. Rainfall was slightly above average in the 2005/6 season and slightly below average in the 2006/7 season (Cain, Owen-Smith & Macandza 2012).

2.5.1 Seasonal Analysis

For all the analyses in this thesis, seasonal models are run with data for the year split into three blocks to run separate seasonal analyses to account for changing behaviour as a result of seasonal conditions. April, May, June and July represent the 'early dry' season; August, September, October and November are considered the 'late dry' season which then transitions to the beginning of the rains and the 'wet' season from December through to March. This split was chosen to take into account the lagged effect of rainfall on food and water availability and used the movement rates as a guide for the split (Owen-Smith, Goodall & Fatti 2012). An alternative approach would be to use rainfall as a covariate within the models. However these data could be unreliable due to the isolated nature of rainfall events in this region. Instead an indicator of "lagged rainfall" could be used as a covariate, such as the NDVI which is a measure of the greenness of the vegetation.

2.5.2 Exploratory Data Analysis

An initial exploratory data analysis was conducted for each herd to see where the animals were moving within their home ranges. Figure 2.15 to Figure 2.18 shows the locations recorded for each herd. The boxplots in Figure 2.19 to Figure 2.23 show the displacements of each of the herds by time of day. There is a clear daily cycle for all of the herds, with much shorter moves during the night, and peaks in the displacement distance in the time after dawn and before dusk. The morning movement peak is slightly shorter for the Punda Maria sable herd than the afternoon peak. This is in contrast to the buffalo herd where the afternoon movements tend to be longer than the morning ones. There is a very clear rest period for the buffalo between 12pm through to 2pm (see Figure 2.19). The Pretoriuskop sable herds tend to move throughout the day, but do also have the early morning and late afternoon peaks. This is similar to the Punda Maria sable which did not show the obvious resting phase of the buffalo herd. The Phabeni herd moves less than the other two Pretoriuskop herds (see Figure 2.20 and Figure

2.21). The movement of the zebra herds, shown in Figure 2.22 and Figure 2.23, show very clear periods of rest during the night and up until 6am.

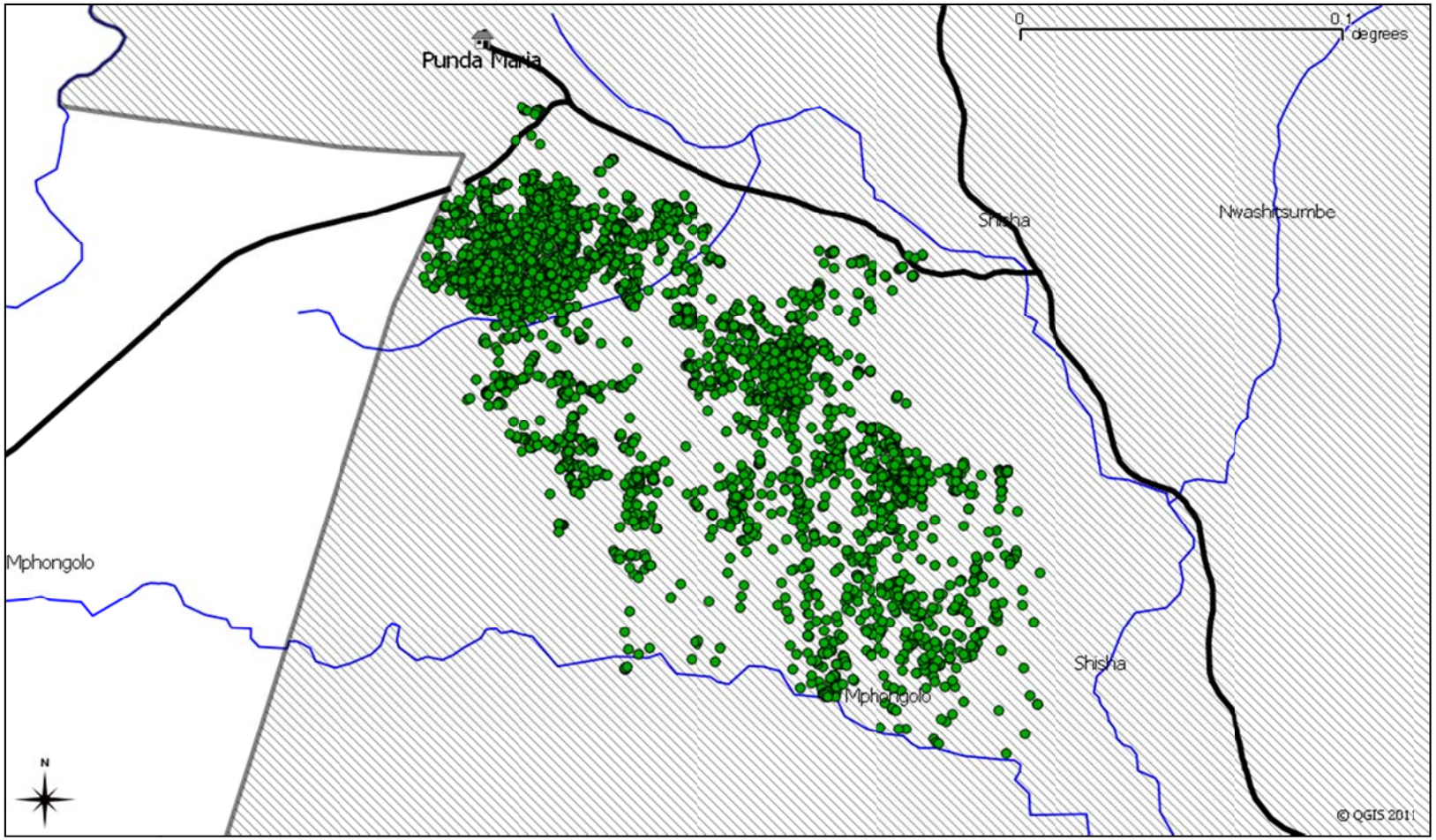

Figure 2.15 - Punda Maria Sable locations

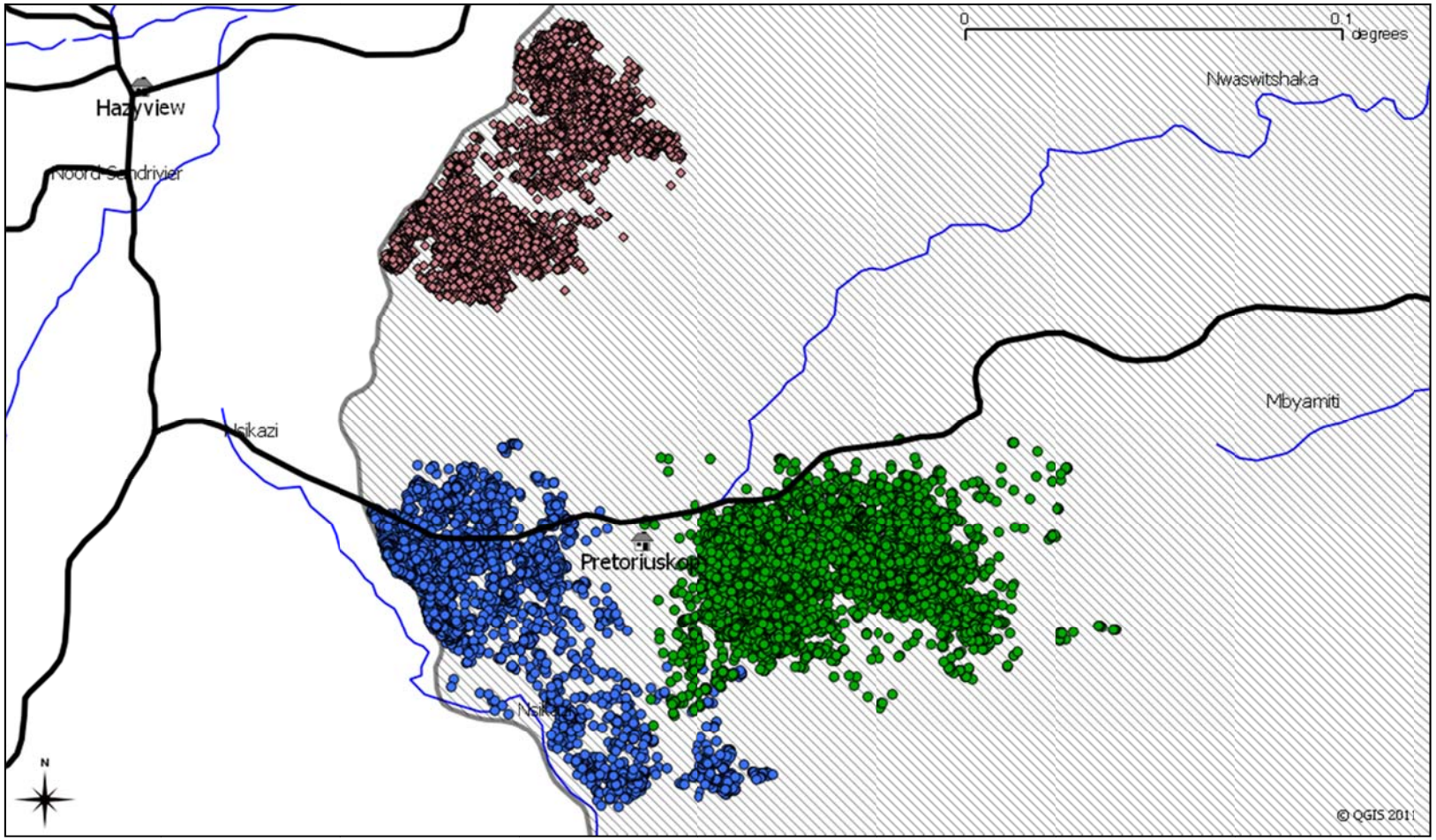

Figure 2.16 - Pretoriuskop Sable locations showing the Phabeni herd (pink), Numbi herd (blue) and Shitlave herd (green)

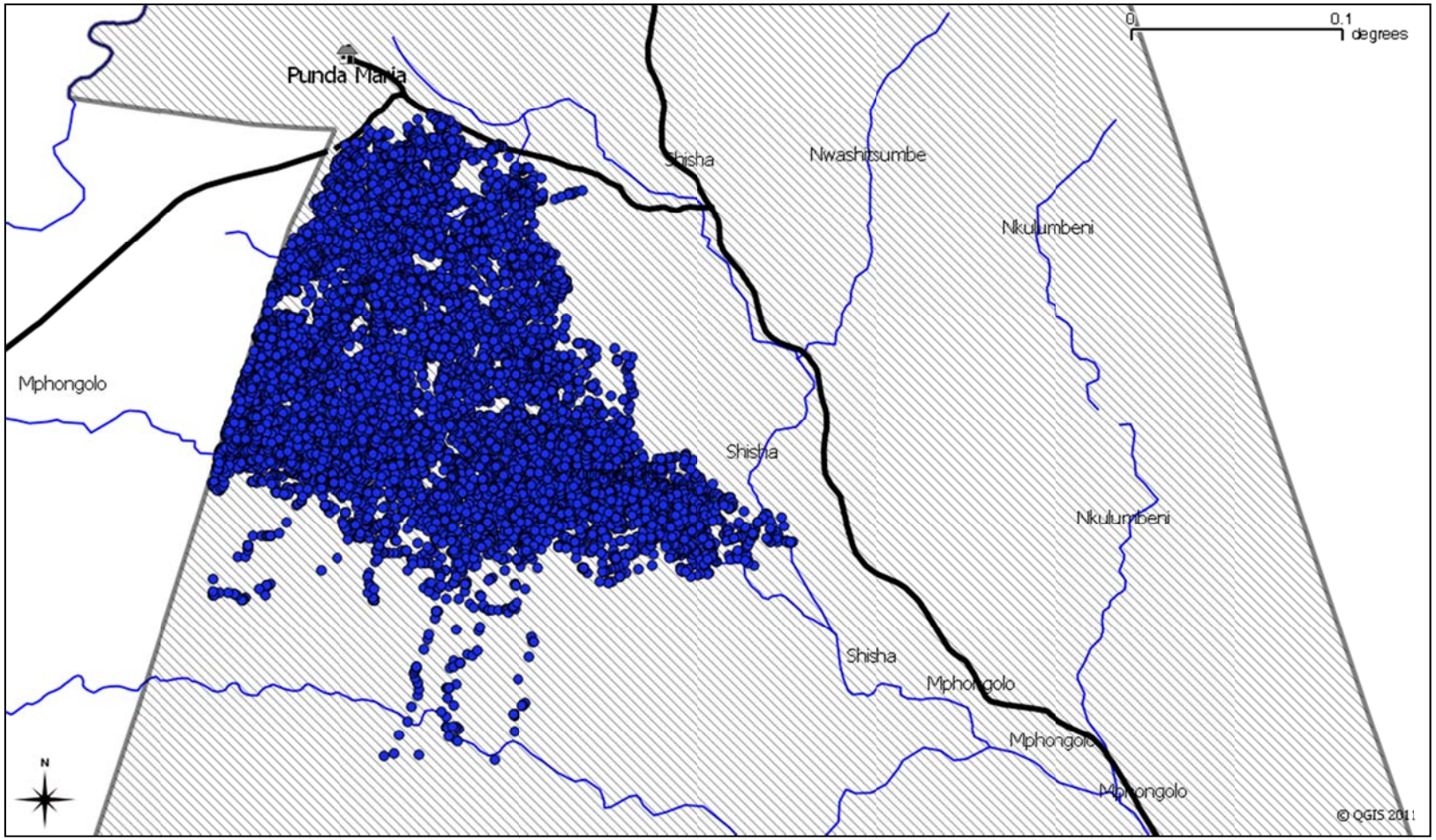

Figure 2.17 - Punda Maria Buffalo locations

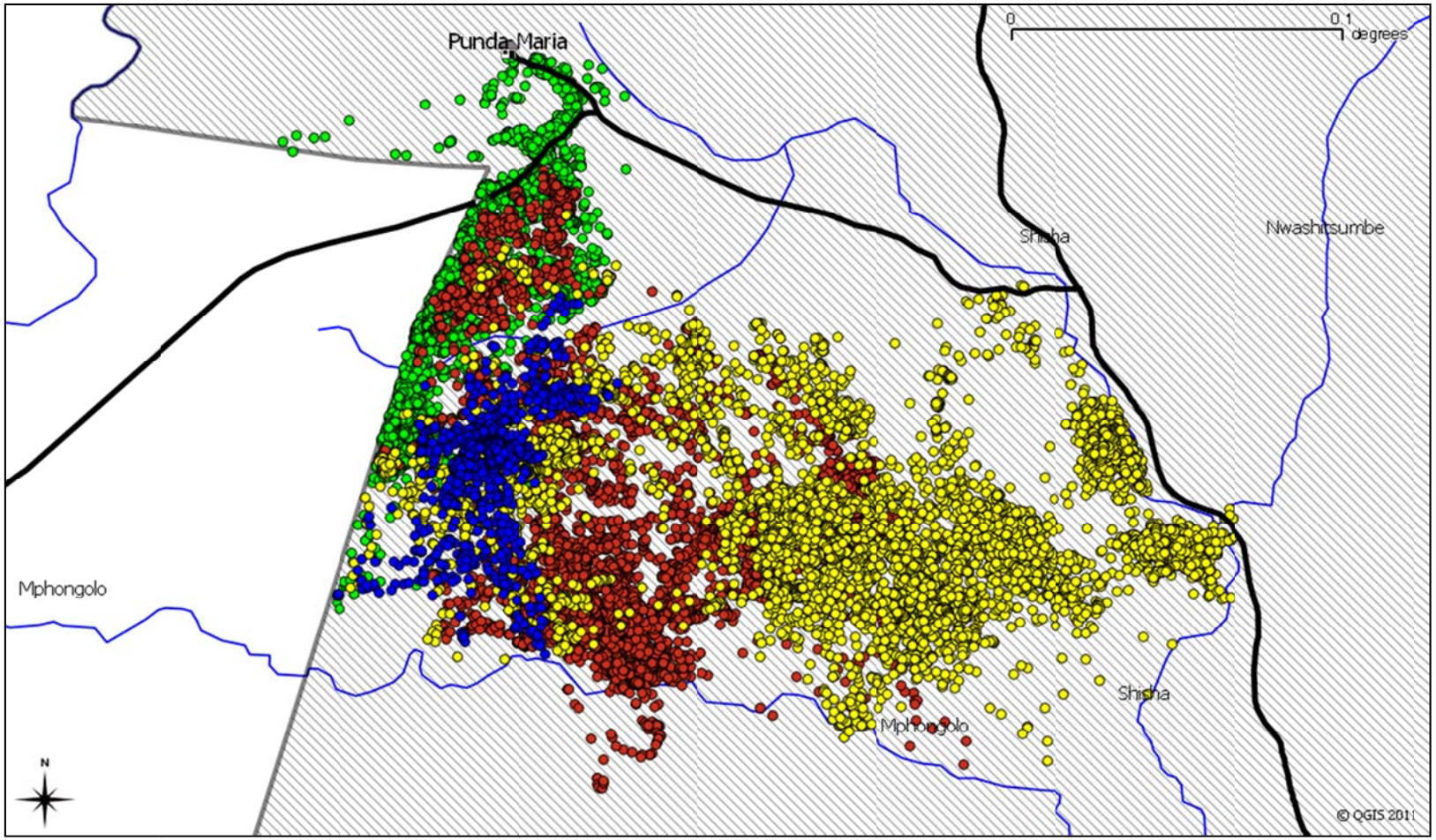

Figure 2.18 - Punda Maria Zebra locations with Zebra141 (blue), Zebra142 (red), Zebra277 (yellow) and Zebra280 (green)

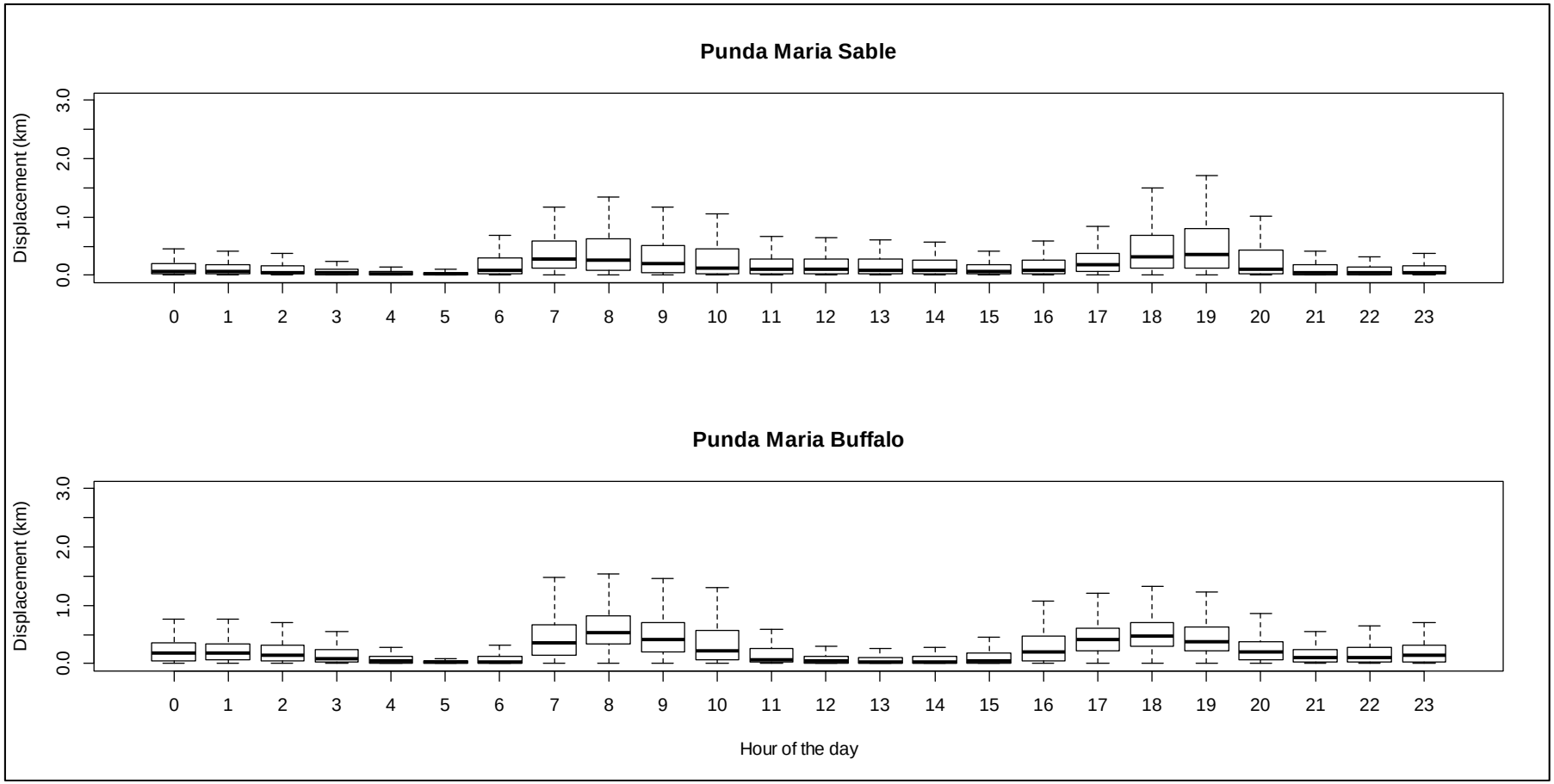

Figure 2.19 - Comparison of Punda Maria Sable and Buffalo movement by time of day

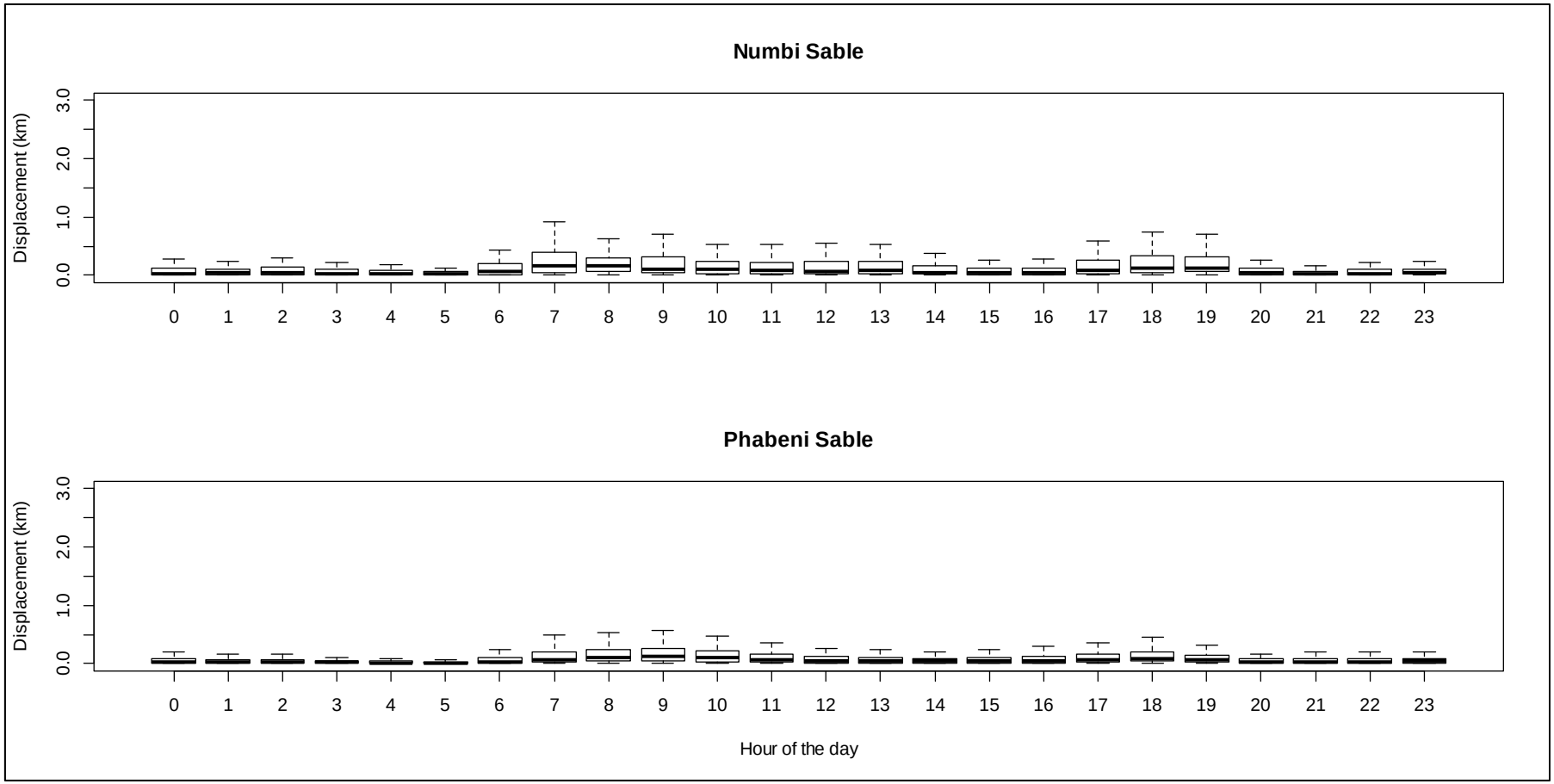

Figure 2.20 - Comparison of Pretoriuskop Numbi and Phabeni Sable herd's movement by time of day

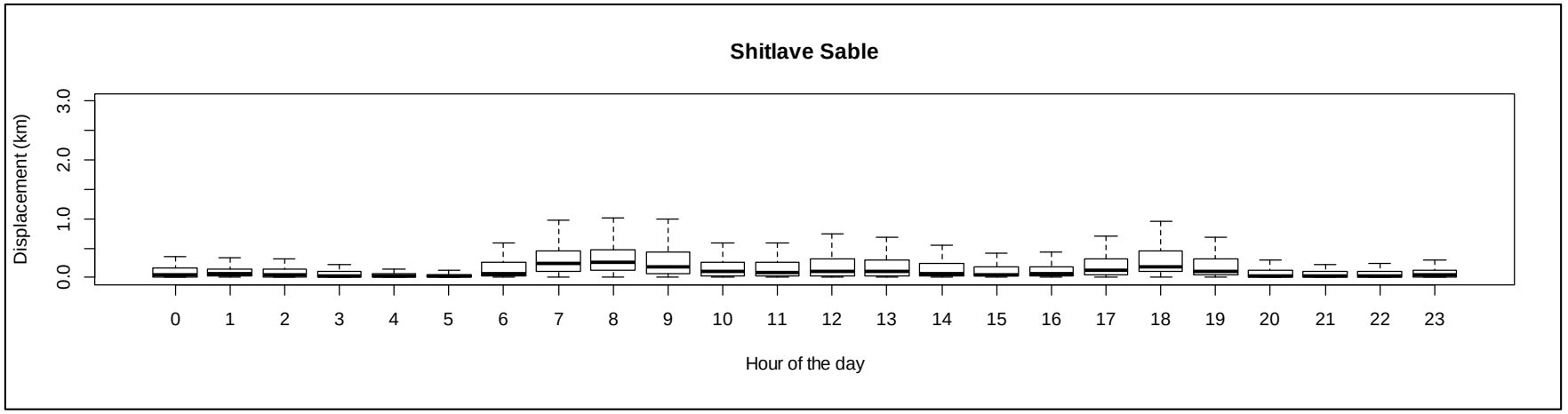

Figure 2.21 - Pretoriuskop Shitlave Sable herd's movement by time of day

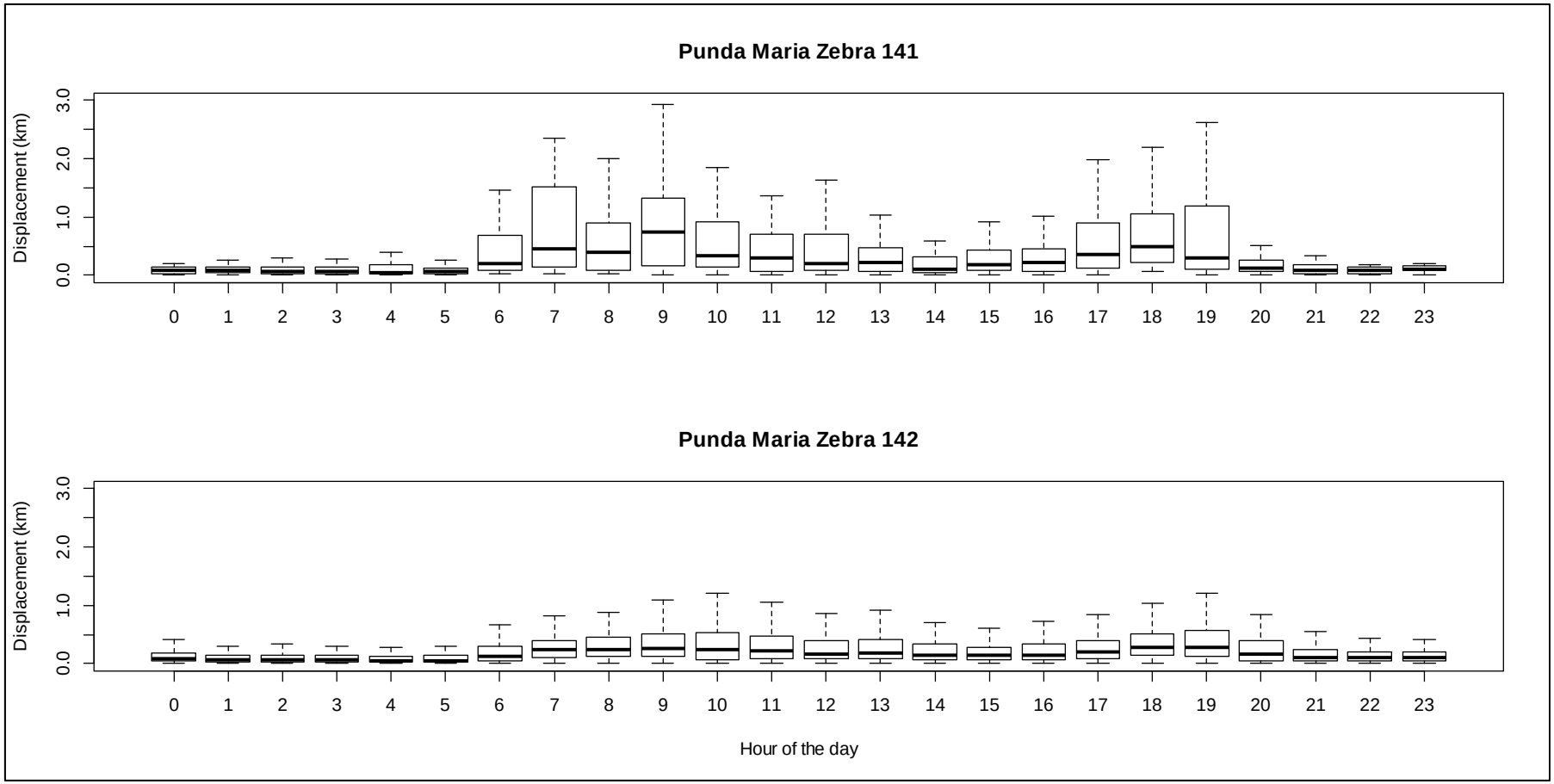

Figure 2.22 - Comparison of Punda Maria Zebra AM141 and AM142 herd's movement by time of day

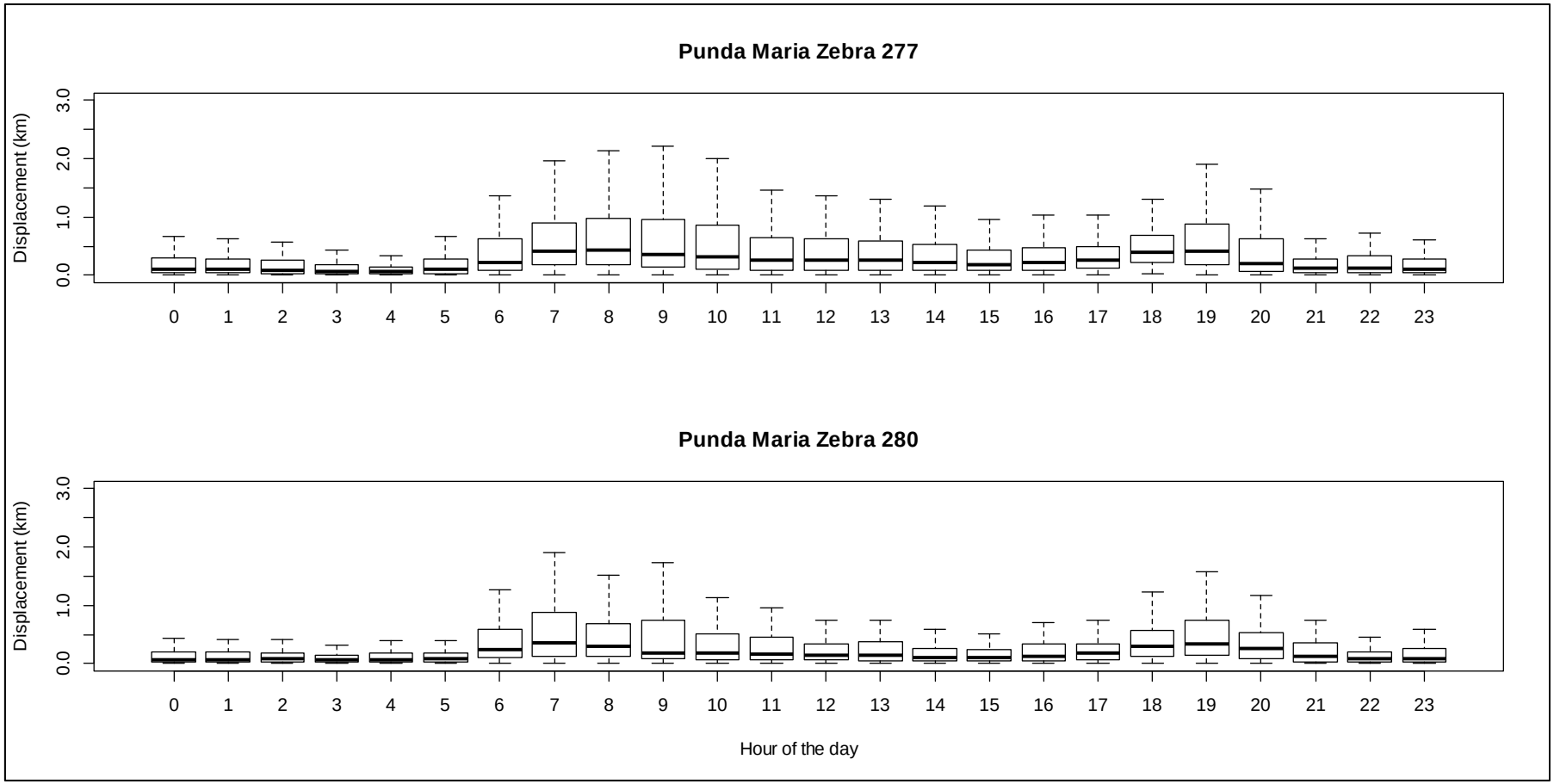

Figure 2.23 - Comparison of Punda Maria Zebra AM277 and AM280 herd's movement by time of day

Figure 2.24 shows the changing longitude and latitude coordinates over time for the Punda Maria sable herd's hourly movements. For the latitude coordinates, there are periods with extreme movements southward. These movements occur during the late dry season and correspond to the locations obtained near to the Mpongolo River shown in Figure 2.15. These Latitude/Longitude plots are useful to see how the animal remains within one area for a while, and then shifts to another region indicated by a sharp change in latitude, longitude or both simultaneously. The pattern for the buffalo movement is different, shown in Figure 2.25, with less obvious patches indicated by the longitude or latitude not remaining consistent for a period of time. It seems the buffalo are not utilizing small patches, but rather moving throughout the landscape.

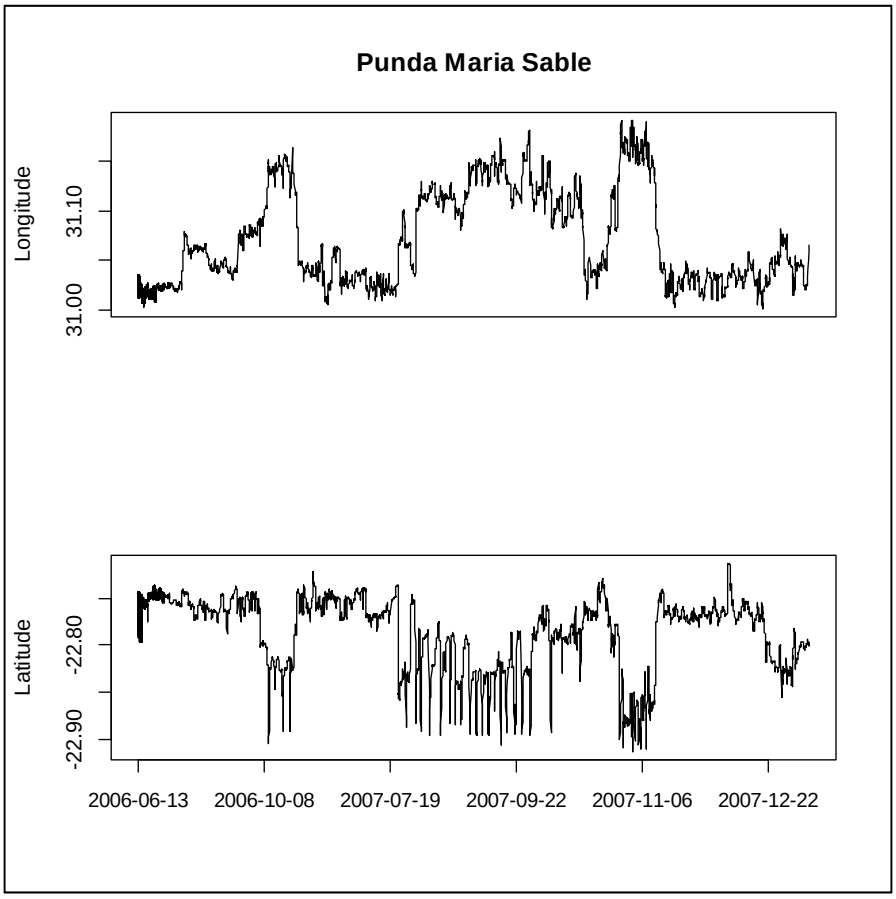

Figure 2.24 - Punda Maria Sable - Longitude and Latitude over time

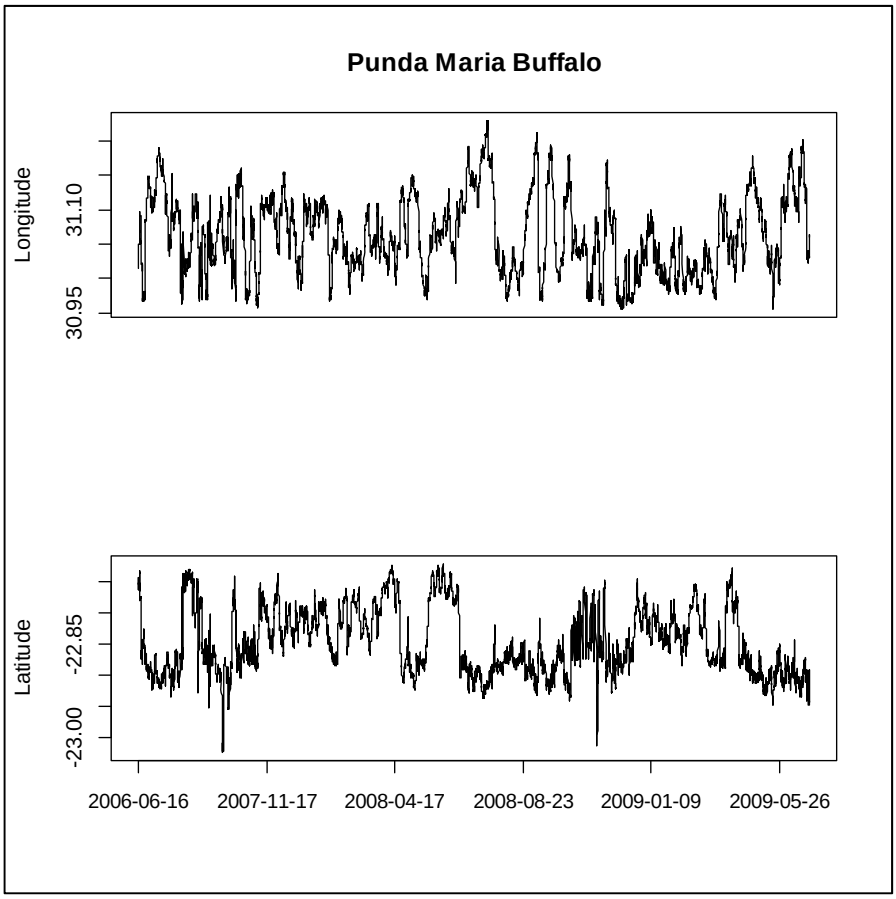

Figure 2.25 - Punda Maria Buffalo - Longitude and Latitude over time

This seems to be confirmed with the scatterplot view of their locations in Figure 2.17. In a comparison of the movement patterns across the seasons, Figure 2.28 shows how the buffalo are moving similar distances across all the seasons, while the sable in the same region are moving longer distances during the Late Dry season compared to the other seasons.

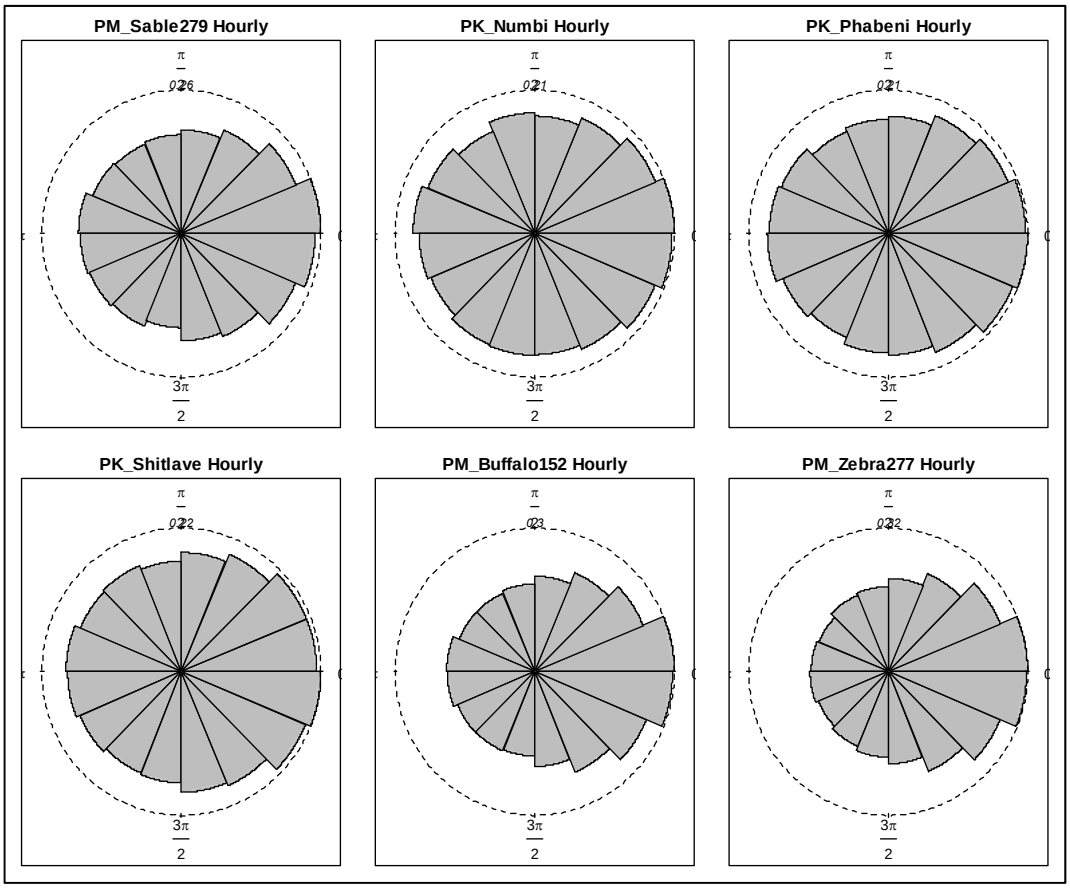

Figure 2.26 - Windrose plots of hourly turning angles

A common metric used in the analysis of animal movement data is the turning angle between successive locations. This shows whether the animal persists walking in the same direction or reverses direction. Figure 2.26 shows windrose plots created using the package '`circular`' (Agostinelli, Lund 2011) in R using hourly data. The plots show the density of directional observations. For the hourly data, there is a slight persistence in the forward directions for every animal. These forward movements tend to be the longer movements.

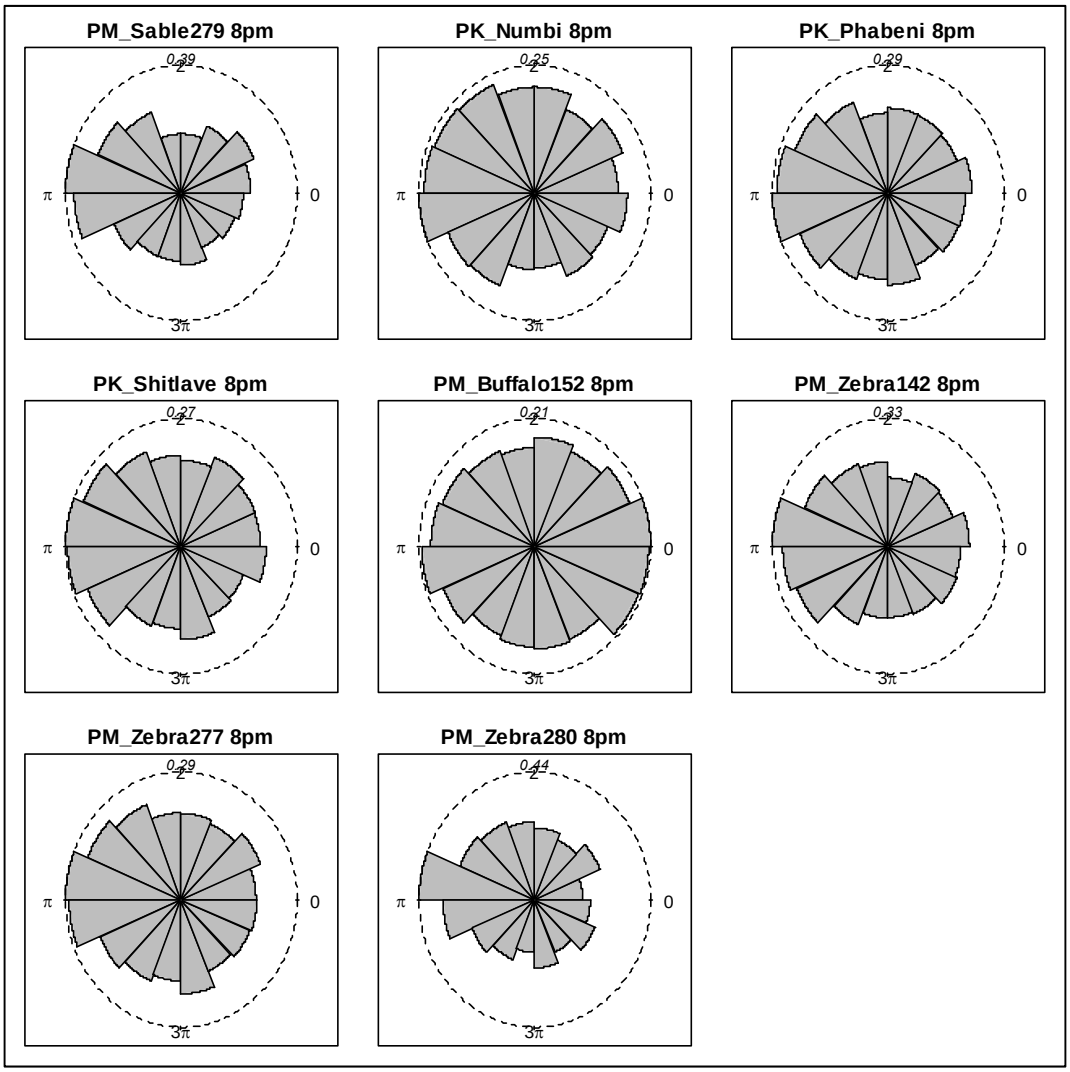

Figure 2.27 – Windrose plots of daily turning angles (8pm)

A very different picture is seen if the angles between 24 hour locations are investigated, the plots are shown in Figure 2.27 for 8pm displacements. There are more direction reversals between the successive location points. This is likely to represent periods when the animal remains in similar regions for resting or foraging, depending on the time of day of the location. The turning angle is not used as an input for the analyses in the following chapters, only the linear displacement between successive observations is used. Section 6.3.2 in the Hidden Markov Model Extensions chapter investigates the impact of including this metric in the modelling process.

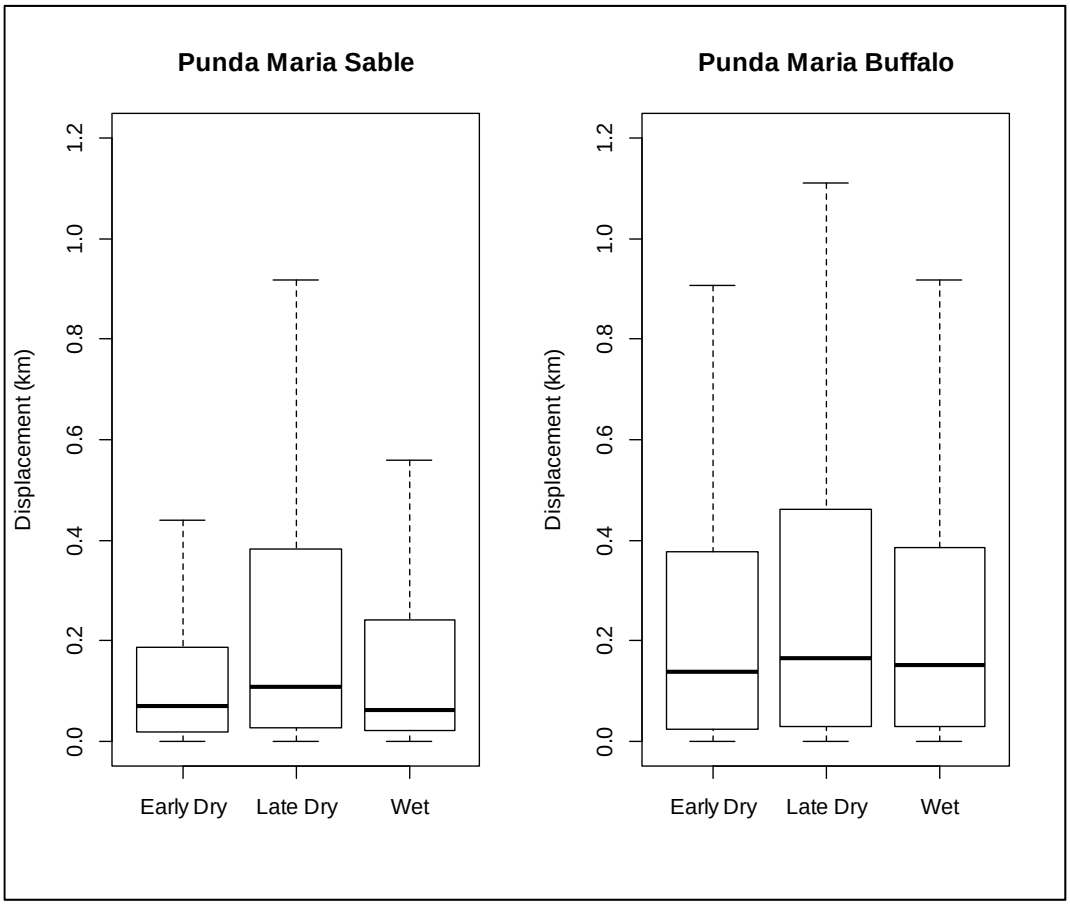

Figure 2.28 - Comparison of the distance moved by season - Punda Maria Sable and Buffalo

2.6 Summary

Advances in GPS tracking technology are opening up new areas of research into the patterns of space use by animals. However, in some ways the ability to collect data with shorter inter-location intervals over longer periods of time, is outperforming the ability to analyse the data. This is due to the lack of suitable models and the technical complexities of fitting such models to the data (McClintock et al. 2012). While there is a degree of error associated with all GPS locations, the aim of modelling the movement is to make inferences about the behaviour of the animal based on observations which include an error component that cannot be teased apart from the observation process. Analysis of when location estimates could not be obtained has the potential to provide some

information about what the animal might have been doing at the time of observation (e.g. lying down) or where it might have been (e.g. dense cover).

Depending on the temporal scale of the analysis and the species being tracked, decisions about how to round the observation data can influence the shape of the distribution of the observed data. To avoid arbitrary decisions about how to round the data which could bias the model input, a decision was made to always use the raw displacement distances, even if the decimal places of the calculated displacement imply a greater precision than that obtainable from the GPS unit. The errors are assumed to be random and independent.

Exploratory data analysis can provide some interesting insight into how animals are utilizing the areas in which they live, and how the movements change through the seasonal cycles. Scatterplots of the locations along with environmental features such as rivers and water points can enable identification of journeys that have been made to water, which for the Punda Maria sable can be seen in particular during the Late Dry season.

In response to this recently available data, statistical methods are being investigated and improved upon in order to better analyse the movement data and explain the results ecologically. Common time-series analysis does not adequately explain the movement patterns due to the complexity of the data and the non-constant and possibly non-Gaussian estimation errors (Jonsen, Myers & James 2006). This has meant that more complex models are being applied to ecological data in order to accurately explain the movement patterns. With the technology advancing rapidly, it is now the statistician's challenge to investigate and adapt existing models, and possibly develop new models, for analysing these data in order to gain insights into the ecological processes underlying it. Many different statistical techniques have been used in the analysis of movement data. Each

analysis has its own advantages and disadvantages. Three methods are focused on in this thesis, all of which have been used to predict the behavioural state of the animal during the observation intervals. Each method has its pros and cons and these are reviewed in each analysis chapter.

3 CHAPTER THREE – DETERMINING ANIMAL ACTIVITY MODES USING INDEPENDENT MIXTURE MODELS

3.1 Introduction

The objective of this study is to analyse the movement patterns of three different ungulate species and infer the behavioural states of the animals. These results can be used to compare the behavioural patterns of the three species and differentiate between them. The behavioural states allow different areas of utilisation to be defined, while movements between and within foraging areas are expected to vary over the seasonal cycle.

Successive location records obtained using GPS devices with suitably small time periods between locations can be used to identify various behavioural states of the animal which take into account the time series nature of the data (Franke, Caelli & Hudson 2004, Franke et al. 2006, Pedersen et al. 2011b). For large ungulates, the most common activity states are resting and foraging, and bouts of these states tend to last for 2-3 hours (Owen-Smith 2002). The Hidden Markov Models used by Franke et al. (2004 & 2006) are also known as dependent mixture models, where the observation is generated by a distribution which depends on the state of an underlying and unobserved Markov process (Zucchini, MacDonald 2009). The marginal distribution of the Hidden Markov model is a finite mixture distribution while a Markov chain is the underlying 'parameter process' (Zucchini, MacDonald 2009). Therefore the activity state of the animal is inferred from its hourly displacement distances.

In a cluster analysis context, mixture models, both dependent and independent, are extremely useful in modelling the heterogeneity between clusters and identifying an underlying group structure within the observed data (McLachlan, Peel 2000). Independent mixture models are used in order to gain an initial understanding of the underlying groupings within the movement data. These groups are assumed to correspond to the behavioural states of the animal. The independent mixture models do not take into account the time series nature of the observed data and the possible dependence between observations. Despite the fact that the sequential structure within the data is being ignored, it is expected that these models will be useful for interpreting GPS locations of various animals in terms of the activities performed at these locations by partitioning the movement rates into activity states. An investigation of the patterns of correlation within the observed data is done in Chapter 4 and then a similar state allocation analysis is performed using dependent mixture models in Chapter 5.

3.2 Break Point Analysis

A method for partitioning movement data into distinct movement modes was introduced in the early 1980s (Slater, Lester 1982, Sibly, Nott & Fletcher 1990, Tolkamp, Kyriazakis 1999, Johnson et al. 2002). The aim was to partition the distribution of observed movement rates through finding breakpoints and assuming that these regions correspond to different behavioural states of the animal. When plotted in a histogram, the log-frequencies of observations are expected to show regions of linearly declining segments. These segments correspond to exponentially declining regions of the raw observations. The breakpoints are assumed to correspond to transition regions between predominant movement modes. However, the picture obtained in the histogram is dependent on the size of the bins into which the observed data is allocated relative to the amount of data available over different ranges. The authors used a method of non-linear curve fitting to the movement rate data of woodland Caribou (*Rangifer tarandus caribou*) to segment their data into intrapatch, interpatch and migratory

movements (Johnson et al. 2002). This process allowed for the identification of what they called scale criteria, which are essentially the breakpoints between the different rates of movement identified from the data. These breakpoints provide information about the displacement regions where the predominant activity states, inferred from the movement rates, transition from one behaviour to another.

Following the ideas presented in Johnson et al. (2002), a segmented regression is used to fit different linear functions over the displacement distances, with the log frequency of the movement displacements as the response variable. Two important considerations of segmented regression are how many segments to break the line into and where to position the breakpoints. The package '`segmented`' was used in R, which iteratively fits linear models to the data in order to determine the optimal breakpoints and the estimated slope and intercept parameters for each linear segment. The user specifies the number of segments required in the model. The segmented (or broken-line) relationship is defined by the slope of the line segment and the break-point where the linear relationship changes. An analysis of variance is then used to compare the fit of models with increasing complexity. If a significant difference is found between the two models, then the more complex model is considered the better model for describing the data.

3.3 Finite Mixture Models

Mixture models have been used in a wide variety of applications including astronomy, biology, genetics, medicine, economics and marketing among others (McLachlan, Peel 2000). They are suitable for modelling distributions which are multimodal since their definition makes them able to cope with "unobserved heterogeneity within the population" (Zucchini, MacDonald 2009). With the advance in computing power, mixture models have become a popular modelling approach.

A mixture distribution is a statistical distribution which can be expressed as a combination of simpler component distributions (Everitt, Hand 1981). The mixture models assume that the observations are drawn independently from a variety of different categories or states, and that each state has its own distribution (Bolker 2008, Zucchini, MacDonald 2009) with its corresponding proportional distribution. The density function of a random variable X with finite mixture distribution can then be defined as:

$$f(x) = \sum_{i=1}^{m} \delta_i p_i(x, \theta_i)$$

where θ_i is the vector of parameters of the i-th of the m component distributions whose probability density functions are given by $p_i(x,\theta_i)$ and $\delta_1,...,\delta_m$ are the proportions of each component in the model, known as the mixing parameters. Since the mixing parameter δ_i is the proportion of the distribution represented by the i-th component, the δ_i sum to one. If X_i denotes the random variable corresponding to the i-th component with probability density function $p_i(x,\theta_i)$, then the expected value of the mixture is given by: $E(X) = \sum_{i=1}^{m} \delta_i E(X_i)$. The k-th moment about the origin is a linear combination of the k-th moments of the components but the variance is not a linear combination of the variances of the component distributions (Zucchini, MacDonald 2009). The expected value and variance of a mixture model are given by:

$$E(X) = \sum_{i=1}^{m} \delta_i E(X_i) \text{ and } E(X^k) = \sum_{i=1}^{m} \delta_i E(X_i^k)$$

$$Var(X) = E(X^2) - (E(X))^2$$

$$= \sum_{i=1}^{m} \delta_i E(X_i^2) - (\sum_{i=1}^{m} \delta_i E(X_i))^2$$

For example, if X is a random variable which is distributed as two-state (m = 2) mixture distribution with δ_1 and δ_2 the mixing parameters, μ_1 and μ_2 the expectations, and σ_1^2 and σ_2^2 the variances of the component distributions, then:

$$E(X) = \delta_1\mu_1 + \delta_2\mu_2$$

$$E(X^2) = \sum_{i=1}^{m} \delta_i E(X_i^2)$$

$$= \delta_1 E(X_1^2) + \delta_2 E(X_2^2)$$

$$= \delta_1(\sigma_1^2 + \mu_1^2) + \delta_2(\sigma_2^2 + \mu_2^2)$$

$$Var(X) = \delta_1(\sigma_1^2 + \mu_1^2) + \delta_2(\sigma_2^2 + \mu_2^2) - (\delta_1\mu_1 + \delta_2\mu_2)^2$$

$$= \delta_1\sigma_1^2 + \delta_1\mu_1^2 + \delta_2\sigma_2^2 + \delta_2\mu_2^2 - (\delta_1^2\mu_1^2 + 2\delta_1\delta_2\mu_1\mu_2 + \delta_2^2\mu_2^2)$$

$$= \delta_1\sigma_1^2 + \delta_2\sigma_2^2 + (\delta_1 - \delta_1^2)\mu_1^2 + (\delta_2 - \delta_2^2)\mu_2^2 - 2\delta_1\delta_2\mu_1\mu_2$$

$$= \delta_1\sigma_1^2 + \delta_2\sigma_2^2 + \delta_1(1-\delta_1)\mu_1^2 + \delta_2(1-\delta_2)\mu_2^2 - 2\delta_1\delta_2\mu_1\mu_2$$

$$= \delta_1\sigma_1^2 + \delta_2\sigma_2^2 + \delta_1\delta_2\mu_1^2 - 2\delta_1\delta_2\mu_1\mu_2 + \delta_1\delta_2\mu_2^2$$

$$= \delta_1 Var(X_1) + \delta_2 Var(X_2) + \delta_1\delta_2(\mu_1 - \mu_2)^2$$

There is often a natural ordering of the components, for example according to the size of their means such that $(\mu_1) < (\mu_2) < \ldots < (\mu_m)$ (McLachlan, Peel 2000).

3.3.1 Estimation of Mixture Distributions

Given a random sample of observations $x_j = 1,\ldots,n$ of X, a number of different techniques may be used to estimate the parameters of the mixture distributions. These include maximum likelihood, Bayesian estimation, inversion and error maximization techniques such as Kabir's intervals, Bartholomew's intervals, method of moments, moment generating functions and other difference methods (Everitt, Hand 1981). Since the development of the EM algorithm, the method of maximum likelihood has been the most commonly used parameter estimation technique (McLachlan, Peel 2000). However many statistical computer software programs now include optimizer functions which can use direct numerical maximization to find the maximum likelihood instead of needing to perform the iterative EM algorithm. Due to its ease of application, the method of maximum likelihood using direct numerical maximization was used to estimate the mixture model parameters for this study. Under general conditions, the estimators obtained using maximum likelihood estimation are consistent (converges in probability to the true parameter values (Wackerly, Mendenhall & Scheaffer 1996)) and asymptotically normally distributed (Everitt, Hand 1981).

The likelihood function is the joint distribution of the observations, as a function of the parameters. Using the notation in MacDonald & Zucchini (2009), under the assumption that the observations are independent, the likelihood function for a mixture model is given by:

$$L(\theta_1,...,\theta_m,\delta_1,...\delta_m \mid x_1,...,x_n) = \prod_{j=1}^{n}\sum_{i=1}^{m}\delta_i p_i(x_j,\theta_i)$$

.

Using the example of the 2-state mixture model discussed above, the likelihood function for this mixture is given by:

$$L(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\delta_1 \mid x_1,...,x_n) = \prod_{j=1}^{n}\delta_1 p_1(x_j,\boldsymbol{\theta}_1) + (1-\delta_1)p_2(x_j,\boldsymbol{\theta}_2)$$

This likelihood function cannot, in general, be explicitly solved for the unknown parameters $\theta_i,\delta_i,$ for $i = 1,...,m$ in the mixture model and it can be unbounded in which case the global maximum does not exist (Everitt, Hand 1981). In most cases, if the boundary values of the parameters are avoided, a local maximum can be found (McLachlan, Peel 2000, Titterington, Smith & Makov 1985). The likelihood can only be unbounded for continuous observations since for discrete observations the likelihood is a probability and hence ranges between zero and one (Zucchini, MacDonald 2009). The case of an unbounded likelihood can occur for a mixture of normal distributions when the mean of one distribution is equal to one of the observations and the variance for that distribution tends to zero (McLachlan, Peel 2000), in which case the likelihood function tends to infinity. This can be solved by placing restrictions on the variances of the distributions within the mixture or by selecting the largest local maximum (Everitt, Hand 1981). The possibility of local maxima being found instead of the global maximum needs to be kept in mind and models should be checked using a variety of initial starting values.

3.3.2 Parameter Estimates

Once the optimal parameter estimates have been obtained using one of the methods discussed in the above section, it is possible to calculate standard errors and confidence intervals around these estimates for each parameter. The standard error provides a measurement of the standard deviation of the estimated parameter. A few different techniques can be used to estimate the standard errors including bootstrapping, the Jack-knife or the use of the asymptotic variance-covariance matrix of the estimators of the parameters. However, this can run into problems when the parameters are on the boundary of the parameter space (Zucchini, MacDonald 2009). A more computationally intensive method, but one that is more accessible since the advance in computing power, is the parametric bootstrap. The process involved is detailed in Zucchini & MacDonald (2009) as follows:

1. Fit the model, and obtain the estimated model parameters $\hat{\boldsymbol{\theta}}$.
2. Generate a 'bootstrap sample' of observations from the fitted model, using parameters $\hat{\boldsymbol{\theta}}$. The length of the bootstrap sample should be the same as the length of the observed data series.
3. Estimate the parameters $\hat{\boldsymbol{\theta}}^*$ for each bootstrap sample.
4. Repeat steps 2 and 3 K times, where K is large and record all of the parameter estimates $\hat{\boldsymbol{\theta}}^*$.
5. The standard deviation of each element $\hat{\boldsymbol{\theta}}_j^*$ is an estimate of the standard error for the j-th parameter estimate.
6. Confidence intervals for each parameter can be estimated using the quantile method. For example the 5% and 95% quantiles provide a 90% confidence interval for each parameter.

The width of the bootstrap confidence intervals can be used to investigate the stability of the model parameter estimates. If the parameter estimate intervals overlap, it could indicate that the two states, represented by the overlapping parameters are not in fact distinct and the model is perhaps over-fitted.

3.3.3 Goodness-of-fit and Model Selection

Assessing the fit of a mixture model to the data is problematic, particularly for multivariate data (McLachlan, Peel 2000). In the univariate case, the density curve for the fitted mixture can be compared with a histogram of the observed data. It is also possible to compare the empirical distribution function with the fitted mixture cumulative distribution (McLachlan, Peel 2000). The Kolmogorov-Smirnov test can be used to test whether the sample distribution arises from a hypothesized distribution by using a comparison of their cumulative distributions (Crawley 2009); however this is not recommended when the parameters of the null distribution have been estimated from the data.

In many cases the number of components *m* to include in the model is unknown, unless it is based upon *a priori* information about the system, for example if known externally existing groupings exist with the dataset (McLachlan, Peel 2000). The component densities are often assumed to come from the same parametric family (McLachlan, Peel 2000), although is not necessarily the case and they can be discrete or continuous distributions (Zucchini, MacDonald 2009). The choice of the number of components can be informed by knowledge of the field of application and informal graphical techniques, such as histograms of the observations. More formal hypothesis testing using the likelihood ratio test can also be used (Everitt, Hand 1981). The Likelihood Ratio Test Statistic is defined as: $-2\log(\lambda) = 2\{\log L(\hat{\boldsymbol{\theta}}_1) - \log L(\hat{\boldsymbol{\theta}}_2)\}$ where $\hat{\boldsymbol{\theta}}_i$ are the maximum likelihood estimates of the unknown model parameters under two nested competing models. However, for mixture models the regularity conditions for the Likelihood Ratio Test Statistic to be asymptotically χ^2 – distributed with the degrees of freedom equal to the difference in number of the model parameters, do not hold (McLachlan, Jones 1988, McLachlan, Peel 2000). Following on from these regularity conditions, the Akaike and Bayesian Information Criteria are also affected. However, they are still often used to determine the number of components to include in a mixture model (McLachlan, Peel 2000). McLachlan

(1987) suggested a method of bootstrapping to determining the *P*-values for the Likelihood Ratio Test Statistic (McLachlan 1987).

The fit of the model based on the likelihood will always improve with additional states and parameters due to an improved flexibility of the model allowing for a closer fit to the data (Clarke 2007). Information criteria such as the Akaike and Bayesian Information Criteria (AIC and BIC) are used to determine how many components are required for the mixture distribution. The information criteria are defined as: $AIC = -2\log L + 2p$ where L is the log-likelihood and p is the number of parameters; and $BIC = -2\log L + p\log T$ where T is the number of observations (Zucchini 2000). The penalty term of BIC has more weight than AIC and hence the BIC more often selects the simpler model with fewer parameters (Quinn, Keough 2002). The likelihood tests are not strictly applicable to compare the models when the models are non-nested (Clarke 2007). However, in practice the AIC and BIC, which are penalized likelihoods, are accepted as being applicable to non-nested models (Vuong 1989, Clarke 2007). The Gamma and log-normal distributions are not nested distributions, meaning that one distribution does not simplify to the simpler distribution if some of the parameters are fixed. In contrast, a Weibull distribution collapses to an exponential distribution if the Weibull shape parameter is fixed at one.

As discussed in McLachlan & Peel (2000), many different information criteria exist and can be used for model selection. These include Bias Correction of the Log Likelihood, AIC, Cross-Validation-Based Information Criterion, Minimum Information Ratio Criterion, Informational Complexity Criterion, Laplace's Method of Approximation, BIC, Laplace-Metropolis Criterion, Laplace-Empirical Criterion (LEC), Reversible Jump Method, Minimum Message Length Principle, Classification Likelihood Criterion, Normalized Entropy Criterion and Integrated Classification Likelihood (ICL) Criterion. McLachlan & Peel (2000) compared the performance of seven of these criteria and found that the ICL, and its large

cluster size approximation ICL-BIC and LEC were the only criteria to correctly select the true number of components in all simulated cases. In their comparison, the AIC tended to select models with too many components.

There is no one statistic or test that will provide a conclusive answer of what the value of *m*, the number of components, should be. According to Everitt & Hand (1981) the decision about what distribution to use and how many components are needed in the mixture should be based on sound theoretical reasons based on knowledge of the area of application of the model.

3.4 Simulation

As discussed above in the application of mixture models for this analysis, one of the unknown features is the number of components to include in the model. In order to investigate the effect of sample size and the consistency of the model selection parameters, datasets were simulated representing mixtures of two, three and four log-normal distributions. Datasets of size 100, 1000, 10 000 and 100 000 were simulated for each of the 2, 3 and 4 state models. For each of the simulated datasets, mixture models were estimated using maximum likelihood with reasonable starting parameters. The results of the simulation are shown in Table 3.1. The maximum log-likelihood value is italicized, while the lowest AIC and BIC values are highlighted in blue. The models selected by AIC or BIC which incorrectly predicted the true number of components in the model are shown in red text. As expected, the maximum likelihood value was always greatest for the more complex four-state model, since the likelihood will always increase as more parameters are added to the model.

For a genuine two-state model, the BIC always selected the correct model. The AIC selected the correct model, except in the case of the dataset with 1000 observations where the AIC would have selected a more complex model (3-state).

For a true 3-state model, the BIC incorrectly selected the simpler 2-state model for the 100 and 1000 observation datasets. The AIC selected the more complex 4-state model for the smallest dataset (100 observations), and the correct 3-state model for all other datasets. The true 4-state model was correctly selected by the AIC for the 10 000 and 100 000 observation datasets, whereas AIC selected the simpler 3-state models for the smaller datasets. The BIC incorrectly selected a two-state model for the dataset with the smallest number of observations and a three-state model for the 1000 and 10 000 observations. It only correctly selected a four-state model for the dataset of size 100 000.

Table 3.1 - Results of model simulation

2-state Log-normal model																	
Model	**Sample size**	**p1**	**p2**	**p3**	**p4**	λ_1	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3	μ_4	σ_4	***Log- like***	***AIC***	***BIC***
True Model		0.60	0.40				-3.50	1.00	-1.00	0.80							
2-state	100	0.57	0.43				-3.73	0.77	-0.91	0.69					-78.7	-147.5	-134.5
3-state	100	0.42	0.33	0.25			-3.97	0.63	-1.98	1.21	-0.71	0.48			-79.6	-143.1	-122.3
4-state	100	0.31	0.13	0.23	0.33		-4.30	0.41	-3.23	0.24	-0.70	0.47	-1.80	1.14	***-81.8***	-141.6	-112.9
2-state	1000	0.64	0.36				-3.42	1.06	-0.90	0.73					-717.4	-1424.8	-1400.3
3-state	1000	0.05	0.57	0.38			-5.12	0.42	-3.36	0.91	-0.95	0.75			-721.1	-1426.2	-1386.9
4-state	1000	0.06	0.36	0.43	0.16		-5.19	0.40	-3.78	0.66	-1.04	0.79	-2.72	0.45	***-722.3***	-1422.5	-1368.6
2-state	10000	0.58	0.42				-3.54	0.97	-1.02	0.81					-6877.1	-13744.3	-13708.2
3-state	10000	0.60	0.05	0.35			-3.51	0.99	-1.50	0.50	-0.90	0.78			-6877.9	-13739.9	-13682.2
4-state	10000	0.57	0.04	0.11	0.28		-3.52	0.97	-2.65	1.54	-1.48	0.57	-0.76	0.73	***-6878.6***	-13735.2	-13655.9
2-state	100000	0.60	0.40				-3.50	1.00	-1.01	0.81					-69245.8	-138481.7	-138434.1
3-state	100000	0.59	0.10	0.31			-3.52	0.99	-1.59	0.76	-0.86	0.77			-69248.2	-138480.4	-138404.3
4-state	100000	0.39	0.25	0.21	0.15		-3.66	0.93	-2.91	1.22	-1.19	0.75	-0.64	0.73	***-69249.1***	-138476.1	-138371.5

3-state Log-normal model																	
Model	**Sample size**	**p1**	**p2**	**p3**	**p4**	λ_1	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3	μ_4	σ_4	***Log- like***	***AIC***	***BIC***
True Model		0.50	0.40	0.10			-3.50	1.00	-1.00	0.80	0.50	0.50					
2-state	100	0.81	0.19				-2.61	1.28	0.16	0.65					-23.9	-37.7	-24.7
3-state	100	0.18	0.60	0.22			-4.19	0.62	-2.28	0.95	0.08	0.68			-24.9	-33.7	-12.9
4-state	100	0.17	0.64	0.14	0.05		-4.19	0.62	-2.22	1.00	-0.05	0.41	1.03	0.06	***-30.8***	-39.7	-11.0
2-state	1000	0.49	0.51				-3.53	0.91	-0.66	0.95					-163.5	-317.0	-292.4
3-state	1000	0.50	0.42	0.08			-3.50	0.92	-0.85	0.78	0.61	0.42			-170.2	-324.3	-285.1
4-state	1000	0.46	0.04	0.41	0.08		-3.46	0.99	-3.64	0.37	-0.83	0.78	0.61	0.41	***-171.0***	-319.9	-265.9
2-state	10000	0.52	0.48				-3.42	1.04	-0.68	0.96					-1915.6	-3821.2	-3785.2
3-state	10000	0.52	0.37	0.11			-3.42	1.03	-0.98	0.77	0.44	0.52			-1959.3	-3902.5	-3844.9
4-state	10000	0.42	0.06	0.42	0.10		-3.38	0.87	-4.63	0.82	-1.03	0.83	0.45	0.51	***-1960.1***	-3898.2	-3818.9
2-state	100000	0.49	0.51				-3.51	1.00	-0.75	1.00					-19137.8	-38265.6	-38218.0
3-state	100000	0.49	0.42	0.09			-3.51	1.00	-0.98	0.84	0.53	0.48			-19605.1	-39194.1	-39118.0
4-state	100000	0.43	0.10	0.38	0.09		-3.57	0.97	-2.52	1.29	-0.95	0.80	0.52	0.49	***-19605.6***	-39189.2	-39084.5

4-state Log-normal model																	
Model	**Sample size**	**p1**	**p2**	**p3**	**p4**	λ_1	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3	μ_4	σ_4	***Log- like***	***AIC***	***BIC***
True Model		0.40	0.30	0.20	0.10		-3.50	1.00	-1.00	0.80	-0.50	0.50	0.50	0.20			
2-state	100	0.43	0.57				-3.63	1.03	-0.53	0.86					3.8	17.6	30.7
3-state	100	0.46	0.26	0.28			-3.55	1.06	-1.14	0.34	0.19	0.50			-0.5	15.1	35.9
4-state	100	0.32	0.08	0.40	0.20		-3.43	0.51	-5.24	0.47	-1.08	0.59	0.40	0.39	***-2.8***	16.5	45.1
2-state	1000	0.41	0.59				-3.45	1.03	-0.49	0.78					117.7	245.4	269.9
3-state	1000	0.41	0.50	0.09			-3.47	1.02	-0.67	0.70	0.53	0.19			96.1	208.2	247.5
4-state	1000	0.40	0.06	0.46	0.09		-3.52	0.98	-1.74	0.39	-0.57	0.64	0.54	0.20	***94.7***	211.5	265.5
2-state	10000	0.43	0.57				-3.38	1.07	-0.49	0.76					906.7	1823.4	1859.4
3-state	10000	0.43	0.48	0.09			-3.39	1.07	-0.67	0.67	0.52	0.19			667.3	1350.5	1408.2
4-state	10000	0.18	0.31	0.43	0.09		-3.85	0.87	-2.75	1.25	-0.62	0.63	0.52	0.19	***663.4***	1348.7	1428.0
2-state	100000	0.44	0.56				-3.34	1.08	-0.49	0.76					8495.0	16999.9	17047.5
3-state	100000	0.44	0.47	0.09			-3.36	1.07	-0.67	0.68	0.50	0.19			6176.6	12369.2	12445.3
4-state	100000	0.20	0.29	0.41	0.09		-3.81	0.89	-2.62	1.24	-0.62	0.63	0.50	0.19	***6137.3***	12296.5	12401.1

These results show that the model selection seems to be more accurate when the number of observations is larger. As the true model complexity increased, the worse the model selection accuracy became, especially for the smaller datasets. In terms of the consistency of model parameter estimates, it was noticeable that the model parameter estimates improved as the number of observations available increased. However, for the more complicated genuine four-state models, the choice of starting parameters influenced how well the true parameters were estimated. In nearly all cases, when more than 10 000 observations were available, the correct model was selected, although the parameter estimates could be quite different to the true model parameter estimates. As the number of states within the true model or the number of fitted states increases, so does the sensitivity to the choice of the starting parameters. With more fitted states, the likelihood surface is more complex which increases the risk of finding a local maximum. Particularly when fitting complex models to the observed data, a number of different starting parameters should be tested to ensure that a global maximum is found.

From these simulation results, the AIC is more often the correct predictor of the best model being the one with the same number of components as the true model. This is contrary to the simulation results presented in McLachlan & Peel (2000), although their simulated datasets ranged in size between 200 and 625. Based on the results obtained in this log-normal mixture simulation exercise, the AIC will be used as the model selection criterion for the models run using the animal movement data. However the BIC will also be reported, as occasions when the two criteria would select different models indicates that there is some uncertainty about how many components are present in the population.

3.5 State Allocation

A mixture model is often applied as an analysis technique to uncover an underlying grouping within the data (McLachlan, Peel 2000). In this case the assumption is that the observations are from a mixture of groups in various proportions. Once the model has been run, a probabilistic clustering of each observation to one of the *m* 'states' is done using the estimated posterior probabilities of the state membership (McLachlan, Basford 1988). The observation is allocated to the state with the maximum estimated posterior probability where the δ_i are taken as the prior probabilities. The posterior probabilities, τ_i, are given by:

$$\tau_i(x_j) = \frac{\delta_i p_i(x_j, \theta_i)}{\sum_{k=1}^{m} \delta_k p_k(x_j, \theta_i)}.$$

This method is known as the 'Mixture Likelihood' clustering method (McLachlan, Peel 2000). This method for clustering the observations can provide unreliable clustering, however this is mostly a problem when the sample size is very low (McLachlan, Basford 1988). In this study, the sample sizes for all of the herds are large ($n \geq 999$), with mostly more than 7000 observations per dataset.

For this study investigating the hourly displacements of large ungulates, it is assumed that the different states would correspond to the animal resting, foraging and travelling. The resting phase is a stationary activity, but the animal might make very short movements during this period. The foraging phase is characterized by tortuous movements searching for food and then stationary bouts while feeding. The movements searching for food will be shorter when there is more available food in the area. The animal will then make longer, more directed journeys towards a water source, other feeding areas, or preferred areas of shelter which will be considered a 'travelling' state. There is also the potential for a component distribution to pick up a mixed activity state where movement and

foraging are mixed together, or movement and resting. Then there is also the potential for mixed states, where the animal feeds and makes shorter relocating movements which do not persist as long as an hour. This state will be labelled as 'mixed movement'. The interpretation of the activity states is subjective and is informed by the expected value and variance of the state-dependent distribution, as well as an understanding of the animal's behaviour.

It is usually not possible to observe the underlying state directly, so confirmation that the model is predicting the correct state is impossible. The state allocation is therefore a prediction only. The posterior probabilities are an indication of the confidence of the state prediction. Where two states have similar posterior probabilities with the maximum only slightly greater than the next highest, it is reasonable to assume that the true underlying state could be either of the two states. While in the case where the maximum posterior probability is much greater than the other state's probabilities, it is more likely that the correct state has been predicted.

In order to determine the expected accuracy of the state allocation procedure, a series of 100 000 states were generated in a proportion similar to the mixing parameters. Observations were then generated based on the 'known' state corresponding to state-dependent parameters similar to those fitted to the animal movement data. The accuracy of the state allocation was found by comparing the fitted state with the 'known' state from the mixture generation. 3- and 4-state models were tested, and the accuracy of the state allocation was always around 80%. This accuracy level would be higher if the state distributions were more dissimilar. However for the models applied to the animal movement, there is an overlap between the distributions which means that the state allocation is not perfect. However, an 80% accuracy level is still reasonably good for inferring the behavioural state of the animal based on the GPS locations only. And in most cases, if the state is incorrectly predicted, it is most likely that it would be

allocated to one of the neighbouring states, which corresponds to a reasonably similar behaviour. It is very unlikely that a true 'relocating' observation is incorrectly allocated to the resting state and vice versa. The more different the individual distributions are from one another, the more accurate the state allocation is.

Since the true underlying state for the animal movement data is unknown, bar-plots of the state prediction by time of day provide a visual method for assessing the effectiveness of the state predictions. They allow comparison of the peaks in behavioural states with the expected behaviour of the animals. These species would be expected to show a peak in foraging after sunrise and before sunset. The animals would be expected to rest during the heat of the day and the longer relocation movements would be most likely during the day, as movement at night is dangerous due to the risk of predation from nocturnal predators. The results of the mixture model analyses can also be compared to field studies which have quantified the amount of time that these species have spent performing the various activities indicated by the states. While the results for different animals of the same species being observed in different areas are likely to differ slightly, the observed results should at least be reasonably close to the behaviour predicted by the model and hence relatively consistent for the species.

This Mixture Likelihood allocation of every observation into one state is a drawback when applied to animal movement data. It is not always realistic to expect that an animal would spend a complete hour performing the behaviour corresponding to one state only. Shorter time intervals between locations readings would make it more likely for the activity bouts to span the complete interval, however more frequent location recordings come at the expense of battery life of the GPS unit. At the time of this study's data collection, a collar with a suitable weight to be carried by a large ungulate had a minimum observation interval of one hour if data collection was required to span a year and hence the full seasonal

cycle. For shorter periods of observation, it would be possible to record locations more frequently. Although it is not expected that the animal will remain in one state for the complete hour, it is expected that the dominant activity state for that hour will be predicted by the model. Although the allocation to a specific state using the maximum posterior probability implies that there is no behavioural overlap during the observation interval, the true behaviour of the animal may include some periods of resting or moving while foraging or some foraging while in a moving mode. The states identified could be expanded to recognize these combinations of states. The basic assumption of the mixture model is that the population is made up of a mixture of distinct distributions, and the 'Mixture Likelihood' clustering then classifies every observation to one state.

However, in the context of animal movement, it could be assumed that each observation is in fact a mixture of different states itself, particularly when the time interval between location readings is longer than the expected length of the bouts for each activity state. To account for this in this study, the posterior probabilities are used as a proxy for the proportion of time during the observation interval spent in each of the activity states. From the posterior probabilities for the fitted model, it is possible to determine the proportion of time allocated to each activity state for each observed movement distance based on the posterior probability proportion. In fact, any possible movement distance can be separated into the contribution from each activity state based on the fitted model. The shape of the distributions will affect the proportions for each state, and these can be investigated for their ability to make biological sense in terms of what activities are likely to have contributed to the observed movements. This proportionate state allocation is used to investigate the performance of the various distributions for appropriately modelling the recorded movement of the three species. This method violates the assumption that the model is based on, that the observed movement comes from a single behavioural state. However, it is an attempt to answer the important ecological question of how the various behavioural states could contribute to an observed movement.

3.6 Model Interpretation

The mixing parameters correspond approximately to the overall proportion of time spent in each state, while the distribution state-dependent means indicate the expected movement distance within each state. In other words, the mixing parameters indicate the distribution of these movement states for the herd (Titterington, Smith & Makov 1985). The proportion of time spent in each state and the expected distances can be compared for the three species in order to investigate the different movement patterns of the species and for species in different regions. Comparisons of the activity state during each hour of the day and a spatial investigation of the regions used for each activity can provide further insight into the behaviour of these three species.

3.7 Analysis

3.7.1 Model Estimation

The method of maximum likelihood was used to determine the optimal parameters for each model. Direct numerical maximization was used in preference to the EM algorithm due to the ease and speed of applying optimizer functions within statistical computing software. Two-, three- and four-state mixtures of Exponential, Gamma, Weibull and Log-normal distributions were modelled. These continuous distributions were selected as they are all valid for positive observations only and they are skew distributions with extended right tails. This pattern of displacements was seen for all of the herds and species, illustrated for the Punda Maria sable herd in Figure 3.1. The Weibull distribution has been used in other animal movement studies to model the displacement distances (Haydon et al. 2008, Morales et al. 2004, Morales et al. 2005).

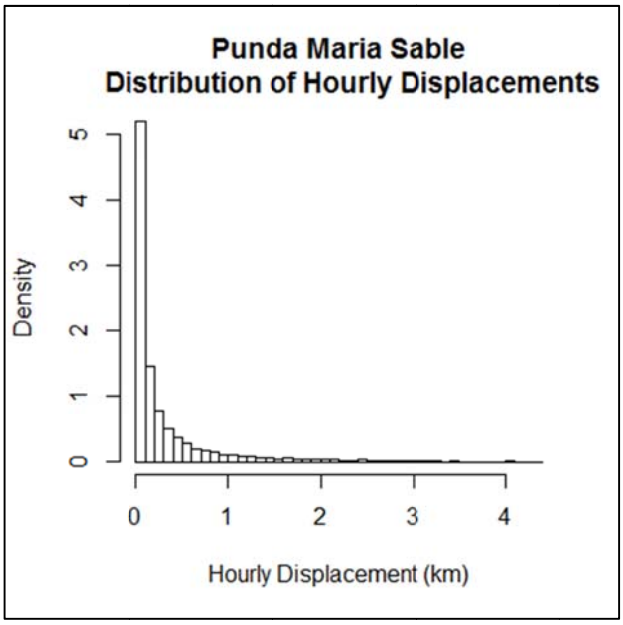

Figure 3.1 - Histogram of Punda Maria sable hourly displacements

The parameter estimates were obtained by numerical maximization of the logarithm of the likelihood function. The optimization was run in R (R Development Core Team 2010), using the `nlm` function which is an unconstrained optimizer using a Newton type algorithm. The R code used was based upon the code provided in Zucchini & MacDonald (2009) adapted for mixture models of combinations of Exponential, Gamma, Weibull and Log-normal distributions. Since `nlm` is an unconstrained optimizer, it was necessary to ensure that only valid parameter estimates were obtained (Zucchini, MacDonald 2009). A constrained optimizer such as `constrOptim` in R could have been used, however in some cases these constrained optimizers are known to be slow (Zucchini, MacDonald 2009). Parameters were transformed before the optimization was run in order to constrain them according to $\sum_{i=1}^{m} \delta_i = 1,\ \delta_i > 0$ (for $i = 1,\ldots,m$) and other distribution-specific parameter constraints. For non-negative parameters, this transformation involved optimizing the unconstrained natural logarithm of the parameter and then transforming back using the exponential function to get the true parameter estimate for variables (Zucchini, MacDonald

2009). For example, the σ_i parameter value within the log-normal distribution is transformed by defining $\eta_i = \log(\sigma_i)$. The likelihood is then maximised with respect to η_i. The constrained optimal parameter can then be obtained by transforming back $\hat{\sigma}_i = e^{\hat{\eta}_i}$. A logit transformation was used to constrain the mixing parameters to sum to one.

3.7.2 Distributions

The log-normal distribution has a density function given by: $f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{\frac{-(\log x - \mu)^2}{2\sigma^2}}$ for $x > 0$. The expected value is given by $E(X) = e^{(\mu + \frac{\sigma^2}{2})}$ and the variance by $Var(X) = e^{2\mu+\sigma^2} \cdot (e^{\sigma^2} - 1)$. Using the notation defined for a mixture model, a two-state log-normal mixture is defined as the weighted sum of the component distributions where $\delta_2 = 1 - \delta_1$ given by:

$$p(x) = \delta_1 \frac{1}{\sqrt{2\pi}\sigma_1 x} e^{\frac{-(\log x - \mu_1)^2}{2\sigma_1^2}} + \delta_2 \frac{1}{\sqrt{2\pi}\sigma_2 x} e^{\frac{-(\log x - \mu_2)^2}{2\sigma_2^2}}$$

The density of the gamma distribution is given by: $f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}}$ for $x > 0$ with α known as the shape parameter and β the scale parameter. The expected value of the distribution is equal to $E(x) = \alpha\beta$ and variance $Var(x) = \alpha\beta^2$. Both the log-normal and gamma distributions can rise sharply near zero which makes them suitable for the analysis of the animal movements, where many very short movements are expected during the periods where the animal is resting and then a spread of longer movements while the animal is active.

3.7.3 Model Offsets

Given that the posterior probability of state membership can be used as a proxy for the contribution of each state to the observed movement distance, the very sharp rise in the left tails of the gamma and Weibull distributions makes them unsuitable for determining the activity states of the animals. For the gamma and Weibull mixtures, very small movements will still have a proportion of time allocated to the active movement states due to the steep rise in the distributions corresponding to these states. To counteract this, an offset model was considered, where offsets are included for the second, third and fourth states. This means that these states only contribute above a certain threshold value. These threshold values were based on the breakpoints obtained from the segmented regression analysis of the log-transformed observation frequencies. The offset values were chosen to be slightly smaller than the break points obtained from the segmented regression, so that a gradual transitioning in for the state could be achieved. The decision of how much to reduce the breakpoints by is a subjective one. The log-normal distributions did not need to include an offset values as the distributions have a more extended left tail.

In the offset case, the distributions forming the mixture are known as 'three-parameter' Gamma or Weibull distributions. The three-parameters refer to the two distributional parameters and the third is the offset value. For the gamma distribution, the probability distribution function for an offset model is defined as follows, with the offset θ known as the shift or location parameter and $x > \theta$:

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}(x-\theta)^{\alpha-1}e^{\frac{-(x-\theta)}{\beta}}$$

.

The offset Weibull distribution is similarly defined for $x > \theta$ by:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x-\theta}{\beta}\right)^{\alpha-1} e^{-\alpha\left(\frac{x-\theta}{\beta}\right)}$$

3.7.4 Seasonal Models

The climatic conditions in the Kruger National Park vary considerably throughout the seasonal cycle. Summers are characterized by hot days, warm nights and this is when the majority of the annual rainfall occurs. The winters are very dry with warm days and cool nights. The veld conditions and water availability fluctuate through the year, and this will influence the behavioural patterns of the animals. Relocation distances are expected to be longer during the dry season when less food is available and journeys to remaining water sources become necessary (Cain, Owen-Smith & Macandza 2012). It is possible that a single mixture model fitted to the full dataset, potentially spanning a number of seasons, will obscure the changes in behaviour that occur through the seasonal cycle. Mixture models are run for each season separately using a subset of the data and then compared to the single model fitted for the full data series. The consistency of the estimated parameters for the seasonal model relative to the aggregate model will indicate whether the models are sensitive to changing behavioural patterns across the different seasons.

Following the approach of Owen-Smith, Goodall & Fatti (2012) months are allocated to seasons such that April to July represents the 'Early Dry' season; August to November the 'Late Dry' season and December to March are the 'Wet' season. This seasonal split was informed by the mean movement rates for each month (Owen-Smith, Goodall & Fatti 2012). However, the environmental conditions experienced by the animals will still vary through these seasons each year based on the timing and amount of rainfall, the temperature and other climatic factors. It is expected that in very dry years, especially when the first summer rains only fall very late, the animals will have to make increasingly longer journeys to find food and water.

3.8 Results

3.8.1 Break Point Analysis

The R package '`segmented`' (Muggeo 2003, Muggeo 2008) was used to run segmented regression models on the log-frequencies of the observed displacements, binned into 1m categories. The segmented relationship is defined by the slope of the linear functions and the breakpoints where a change in the slope of the linear function occurs. The results for models were run using two or three breakpoints which correspond to 3- and 4-segment models respectively. The breakpoints are shown in Table 3.2. These breakpoints will be used later in the mixture model analysis, as offsets applied to Gamma or Weibull distributions.

Table 3.2 - Segmented regression - Optimal breakpoints corresponding to distance moved in kilometres per hour

		3-state		4-state		
Species	**Region**	***BP*1**	***BP*2**	***BP*1**	***BP*2**	***BP*3**
Sable	Punda Maria	0.188	0.637	0.079	0.257	0.698
Sable	Numbi	0.195	0.560	0.078	0.258	0.591
Sable	Phabeni	0.193	0.587	0.058	0.270	0.615
Sable	Shitlave	0.169	0.598	0.069	0.281	0.617
Buffalo	Punda Maria	0.064	1.048	0.044	0.084	1.050
Zebra 141	Punda Maria	0.150	0.331	0.156	0.294	0.639
Zebra 142	Punda Maria	0.196	0.722	0.182	0.571	0.944
Zebra 277	Punda Maria	0.370	0.971	0.137	0.515	1.314
Zebra 280	Punda Maria	0.283	0.634	0.163	0.385	0.808

Assuming that the states implied by the 4-state (3 breakpoints) model correspond to the animal resting, foraging, mixed movement and relocation, the breakpoints imply the distances at which a transition from one predominant state to another is

expected. The sable from all herds have similar breakpoints, with a resting state up till about 80m, foraging up to between 250m and 280m, and the travelling behaviour starting from around 500m-600m. The breakpoints identified for the buffalo are quite different, with the first two breakpoints at 40m and 80m respectively. The transition to relocating movements begins at more than 1km. This seems to suggest that a large number of very small displacements are observed with a very static resting phase and then a slightly more active state with movements up to 80m. The breakpoints identified for the zebra herds are very different compared to the other two species. The first state has much longer movements with the breakpoints identified between 140m and 180m. The second breakpoint is between 300m and 500m for the four herds and the final transition happens between 700m and 1.4km. The shortest breakpoints were identified for the Zebra141 herd which had the fewest observations. The second and third breakpoints are similar to those obtained for the buffalo and sable. An analysis of variance was run to compare models of increasing complexity (increasing number of breakpoints) for each herd. In all cases, the most complex model (three breakpoints) was selected as the 'best' model. While this shows that the data could support more complex models with more breakpoints, these would be difficult to interpret ecologically and distinctions between behavioural states would be difficult.

This breakpoint analysis using segmented regression models provides a crude method for allocating observations to different behavioural states. The models are useful in identifying the region where one activity state is likely to transition into another.

3.8.2 Mixture Models

Initially mixture models were run using combinations of exponential, log-normal, Weibull and Gamma distributions. The log-normal, Gamma or Weibull models provided a consistently better fit to the observed data than the exponential models.

Mixtures of all Gamma or all Weibull distributions failed to converge for the more complex 3- or 4-state models, unless an exponential or log-normal distribution was used for the first component. Alternatively an offset Gamma or Weibull distribution was able to converge and provide a good fit to the data, using offsets informed by the segmented regression analysis. The focus for the rest of this chapter is on mixture models using log-normal distributions or offset Gamma or Weibull distributions, the reasons for which will be discussed later in this chapter.

The maximum log likelihood, AIC and BIC values for the log-normal mixture models are shown in Table 3.3, fitted for 2-, 3- and 4-states. The minimum AIC and BIC values are highlighted in green. For all three Pretoriuskop herds, both the AIC and BIC values select a three-state model as the best model to describe the movement. The information criteria for the Punda Maria Sable, Zebra 142, Zebra 277 and Zebra 280 differ, with the AIC selecting 4-state models and BIC selecting the less complex 3-state models. All of these datasets are larger than 4500 observations. In the simulation exercise discussed above, the AIC tended to be the more reliable measure for selecting the correct number of states from the simulated data for the large datasets, so for interpretation purposes for these herds, the 4-state model will be used. The smallest dataset (T=999) for Zebra 141 also had differing models selected by AIC (3-state) and BIC (2-state). Due to the smaller sample size, the less complex 2-state model will be used for the interpretation of this herd's movement. The buffalo movement, with the largest number of observations (T=15733), had a 4-state model selected by AIC but only a 2-state model selected by the BIC. For the simulation study, this type of split only occurred for the very small datasets (T=100) when the true model was a 3-state model. However, due to the very large sample size for the buffalo dataset, 4-state model selected by the AIC will be used for model interpretation.

The negative log-likelihood value for Zebra277 is the only positive value, while all the other herds have negative values. The reason for this is that Zebra277 has a higher proportion of observations over about 260m which is the transition region for the density contributions for an observation to the overall log-likelihood function to switch from negative to positive. This is based on the density from the mixture of fitted log-normal distributions for the Zebra277 herd which is negative for observations below 260m and positive above. The Zebra280 herd has 33% of observations over 260m while in comparison the Zebra277 has 42% observations greater than this. The larger proportion of observations above 260m for the Zebra277 means that its log-likelihood value is positive unlike every other herds. Across the whole time-series, the average movement for the Zebra277 herd (0.4km) is longer than the Zebra280 herd (0.3km) but the maximum distance movement is not that different, 4.5km compared to 4.1km.

Table 3.3 - Log-normal mixture models - Maximum Likelihood, AIC and BIC

Species	Region	Observations	Distribution	Model	-LogLik	AIC	BIC
Sable	Punda Maria	5330	Log-normal	2-state	-2781.6	-5553.2	-5520.294
				3-state	-2802.979	-5589.958	-5537.31
				4-state	**-2814.189**	-5606.378	-5533.986
Sable	Numbi	5803	Log-normal	2-state	-6010.7	-12011.4	-11978.07
				3-state	-6028.979	-12041.96	-11988.63
				4-state	**-6031.164**	-12040.33	-11967
Sable	Phabeni	14085	Log-normal	2-state	-20060.59	-40111.19	-40073.43
				3-state	-20086.11	-40156.21	-40095.79
				4-state	**-20089.1**	-40156.2	-40073.12
Sable	Shitlave	9671	Log-normal	2-state	-8362.515	-16715.03	-16679.14
				3-state	-8386.831	-16757.66	-16700.25
				4-state	**-8388.687**	-16755.37	-16676.43
Buffalo	Punda Maria	15733	Log-normal	2-state	-6308.061	-12606.12	-12567.81
				3-state	-6316.775	-12617.55	-12556.24
				4-state	**-6335.018**	-12648.04	-12563.74

Zebra 141	Punda Maria	999	Log-normal	2-state	-74.45057	-138.9011	-114.3674
				3-state	-78.79625	-141.5925	-102.3385
				4-state	**-81.0878**	-140.1756	-86.2013
Zebra 142	Punda Maria	7747	Log-normal	2-state	-3023.796	-6037.593	-6002.817
				3-state	-3037.619	-6059.238	-6003.597
				4-state	**-3042.687**	-6063.375	-5986.869
Zebra 277	Punda Maria	7409	Log-normal	2-state	140.52194	291.0439	325.5961
				3-state	95.98198	207.964	263.2476
				4-state	**87.0043**	196.0086	272.0236
Zebra 280	Punda Maria	4508	Log-normal	2-state	-1119.227	-2228.455	-2196.386
				3-state	-1155.141	-2294.283	-2242.974
				4-state	**-1161.629**	-2301.258	-2230.708

Table 3.4 - Maximum Likelihood Parameter estimates for the 'best' model

Species	**Region**	$p1$	$p2$	$p3$	$p4$	$\mu 1$	$\mu 2$	$\mu 3$	$\mu 4$	$\sigma 1$	$\sigma 2$	$\sigma 3$	$\sigma 4$
Sable	Punda Maria	0.52	0.32	0.13	0.02	-3.67	-1.64	-0.26	0.78	1.17	0.82	0.58	0.26
Sable	Numbi	0.57	0.33	0.10		-3.82	-1.90	-0.50		1.09	0.77	0.53	
Sable	Phabeni	0.65	0.20	0.16		-3.97	-2.40	-1.20		1.19	0.58	0.71	
Sable	Shitlave	0.56	0.33	0.11		-3.73	-1.70	-0.43		1.18	0.73	0.53	
Buffalo	Punda Maria	0.42	0.34	0.22	0.02	-3.89	-1.48	-0.59	0.60	1.21	0.72	0.54	0.28
Zebra 141	Punda Maria	0.84	0.16			-2.22	0.11			1.26	0.50		
Zebra 142	Punda Maria	0.28	0.50	0.07	0.14	-3.34	-1.99	-1.14	-0.33	1.11	0.88	0.35	0.55
Zebra 277	Punda Maria	0.32	0.38	0.22	0.08	-3.15	-1.75	-0.57	0.41	1.28	0.86	0.62	0.41
Zebra 280	Punda Maria	0.25	0.50	0.18	0.07	-3.63	-2.07	-0.74	0.34	1.12	0.89	0.62	0.38

Table 3.5 - State-dependent distribution means for the 'best' models

Species	Region	Mean 1	Mean 2	Mean 3	Mean 4
Sable	Punda Maria	0.05	0.27	0.91	2.25
Sable	Numbi	0.04	0.20	0.70	
Sable	Phabeni	0.04	0.11	0.39	
Sable	Shitlave	0.05	0.24	0.75	
Buffalo	Punda Maria	0.04	0.30	0.64	1.89
Zebra 141	Punda Maria	0.24	1.26		
Zebra 142	Punda Maria	0.07	0.20	0.34	0.84
Zebra 277	Punda Maria	0.10	0.25	0.68	1.64
Zebra 280	Punda Maria	0.05	0.19	0.58	1.51

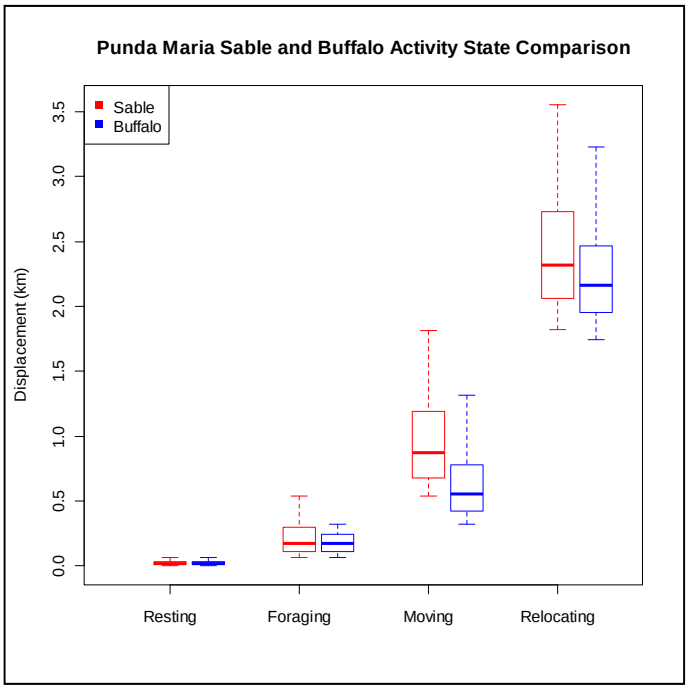

Figure 3.2 - Activity State Comparison - Punda Maria sable and buffalo

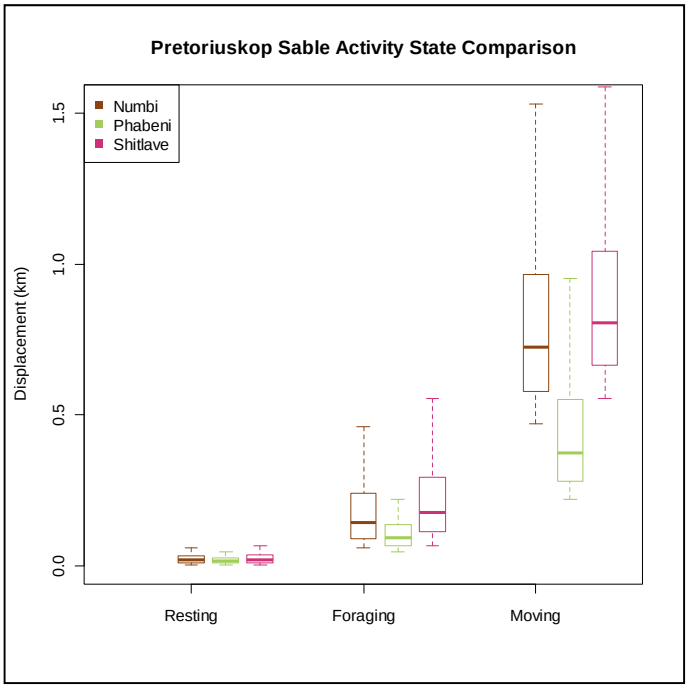

Figure 3.3 - Activity State Comparison - Pretoriuskop sable

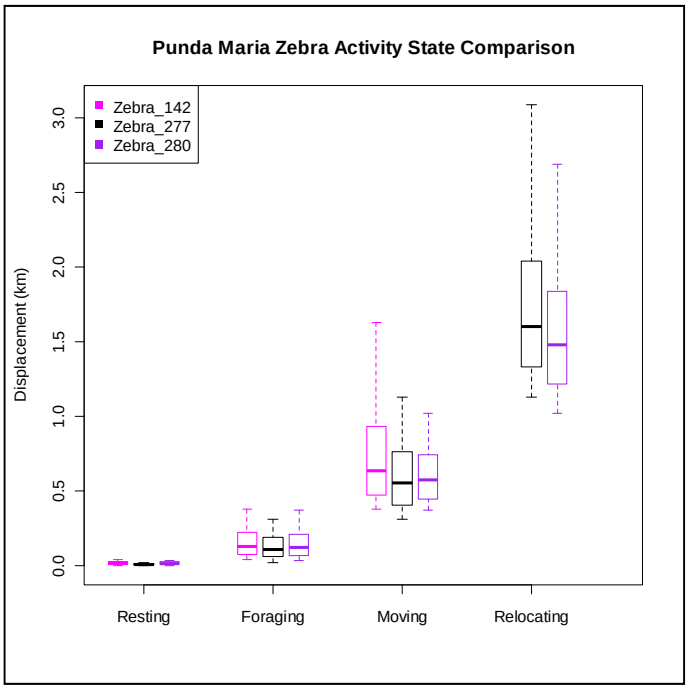

Figure 3.4 - Activity State Comparison - Punda Maria zebra

The maximum likelihood parameter estimates for the 'best' model are shown in Table 3.4. For the four sable herds, the parameter estimates are reasonably consistent. The expected movement distance, shown in Table 3.5, for the resting state is consistently around 50m, while the mean movement within the foraging state is around 200m, although less for the Phabeni herd. The Numbi and Shitlave herds have a mean for their mixed movement state of around 700m, while the Phabeni herd moved slower and has a mean movement distance of only 400m within the movement state. The Punda Maria sable and buffalo have similar means for their resting and foraging states, although the models find shorter movement distances for the mixed movement and travelling states for the buffalo. However the mixing parameters indicate that the buffalo have more time allocated to the third state than the sable antelope in the same region. The three zebra herds, for which a 4-state model was selected as the best, have similar expected movement distances for the first two behavioural states, while the Zebra142 herd's expected movement distance in the travelling state is 840m in comparison to the other two herds over 1.5km. Further comparative plots of the movements

allocated to the respective states based on the fitted models for the herds and species are shown in Figure 3.2 to Figure 3.4. These plots show that the sable and buffalo are moving similar distances in the first and second states, but the sable are making much longer moves in the third and fourth states. The Phabeni sable herd has noticeably shorter movements across all three states, with the Numbi and Shitlave herds similar to one another. The third state for Zebra142 did not have any observations allocated to it which means that only three states are plotted in Figure 3.4 for this herd. This is due to the very low mixing parameter (0.07) value for this state which means that it is never the most likely state for the observed displacements. This suggests that a 3-state model could be the better model for this herd even though the 4-state model was selected by the AIC. The mixing parameters for the foraging state for all of the zebra herds are far higher than for the buffalo and sable. This is consistent with the non-ruminant grazers who need to spend more time feeding to fulfil their nutritional requirements (Janis 1976). The mixing parameters for the two active movement states ranges between 10% and 22% of the time for the buffalo and sable which is consistent with the amount of time normally spent moving between feeding sites for grazers (Owen-Smith 2002).

These expected movement distances discussed above are shorter than the segmented regression breakpoints, since they are mean values rather than upper transition points. The regression provides a fixed transition point, while the mixture model allows for fading in of the higher state. This suggests that there is actually reasonable consistency in the regions identified by both models where a transition from one activity state to another occurs.

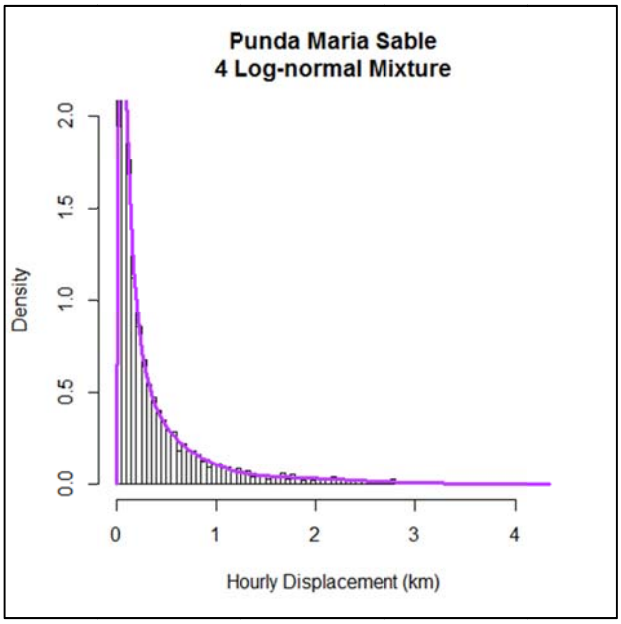

Figure 3.5 - Punda Maria Sable 4-state Log-normal mixture model

The goodness-of-fit of the model is assessed using Figure 3.5 and Figure 3.6. Figure 3.5 shows a histogram of the observed data, with the fitted mixture density which appears to fit the observed data well. Figure 3.6 shows a very close correspondence between the empirical distribution function with the fitted cumulative mixture distribution. These plots provide a reasonable assessment of the goodness-of-fit of the model to the observed data and show that the 4-state log-normal mixture does appear to provide a good fit to the observed hourly displacements for the Punda Maria sable herd.

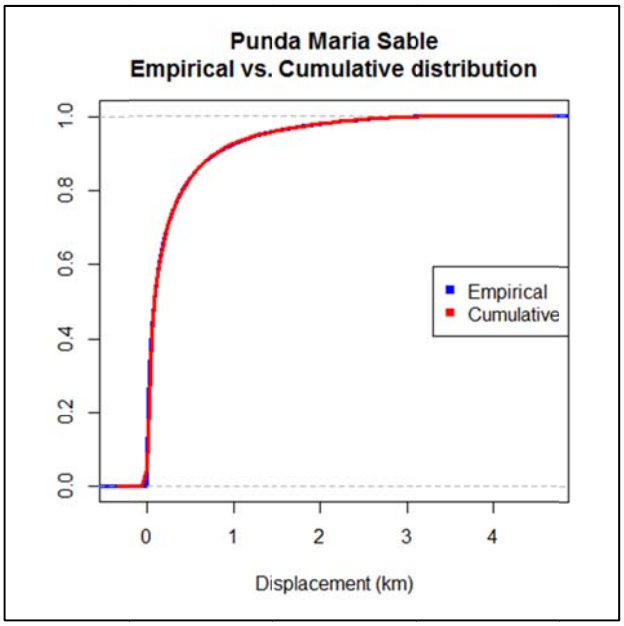

Figure 3.6 - Punda Maria sable - Empirical vs. Cumulative distribution plot - 4-state log-normal mixture model

3.8.3 Parameter Estimates

The parametric bootstrap was used to determine the standard error of each parameter estimate and 90% confidence intervals were calculated around each estimate. The bootstrap estimates were calculated using 500 bootstrap samples, which is much larger than the suggested 50 to 100 replications (Efron, Tibshirani 1993). For long time-series, the bootstrap parameter estimation can take time to compute since the length of the bootstrap sample is the same as the length of the observed time series and the mixture model parameter estimation for the 500 bootstrap samples is time consuming, particularly for the more complex 4-state models.

The results for the Punda Maria Sable herd are shown in Table 3.6. The maximum likelihood estimate (MLE) calculated from the observed data during the model fitting process is almost always the same value as the mean of the bootstrap parameter estimates from the bootstrap sample (labelled Bootstrap Mean). The standard error for each of the parameter estimates range between 0.01 and 0.18

indicating that the maximum likelihood estimators have a reasonably small standard deviation compared to their mean values and should provide precise estimates of the parameters. It is noticeable that the 90% confidence intervals for the mixing parameters and the state-dependent parameters for the two- and three-state models do not overlap at all. However for the 4-state model, there is a slight overlap for the mixing parameters intervals which are quite wide. However there is no overlap in the confidence intervals for the μ_i's which suggests that the states are well defined even for the more complex 4-state model. However, as discussed in Zucchini and MacDonald (2009), the parametric bootstrap estimate of the distribution of the estimators is derived under the assumption that the fitted model is indeed the correct model. The true model will never be known with 100% accuracy, except in the case of simulated data. The same results are shown for the Punda Maria buffalo herd in Table 3.7. The standard errors are also very small for nearly all of the parameters, only the μ_2 parameter has a larger standard error for both the 3- and 4-state models (0.34 and 0.21 respectively). As with the sable herd, the 90% confidence intervals for the mixing parameters do overlap, while there is no overlap for the μ_i's which again shows that the states are well defined for the 4-state model for buffalo movement.

Table 3.6 - Bootstrap Parameter estimates – Punda Maria Sable for 2-, 3- and 4-state mixture models

	delta1	delta2	mu1	mu2	sigma1	sigma2
MLE	0.63	0.37	-3.36	-0.88	1.30	0.98
90% CI Lower Limit	0.57	0.31	-3.52	-1.03	1.22	0.91
90% CI Upper Limit	0.69	0.43	-3.19	-0.73	1.38	1.05
Bootstrap Mean	0.63	0.37	-3.36	-0.88	1.30	0.98
Bootstrap Standard Error	0.04	0.04	0.10	0.09	0.05	0.04

	delta1	delta2	delta3	mu1	mu2	mu3	sigma1	sigma2	sigma3
MLE	0.56	0.37	0.06	-3.54	-1.26	0.33	1.22	0.91	0.48
90% CI Lower Limit	0.50	0.24	0.04	-3.72	-1.47	0.05	1.14	0.65	0.36
90% CI Upper Limit	0.65	0.44	0.13	-3.33	-1.11	0.45	1.31	1.01	0.61
Bootstrap Mean	0.57	0.35	0.08	-3.52	-1.29	0.29	1.23	0.86	0.49
Bootstrap Standard Error	0.05	0.06	0.03	0.12	0.11	0.13	0.05	0.11	0.08

	delta1	delta2	delta3	delta4	mu1	mu2	mu3	mu4	sigma1	sigma2	sigma3	sigma4
MLE	0.52	0.32	0.13	0.02	-3.67	-1.64	-0.26	0.78	1.17	0.82	0.58	0.26
90% CI Lower Limit	0.43	0.16	0.06	0.01	-3.89	-1.96	-0.58	0.61	1.07	0.54	0.33	0.17
90% CI Upper Limit	0.61	0.42	0.24	0.05	-3.41	-1.38	-0.10	0.86	1.28	0.97	0.69	0.35
Bootstrap Mean	0.52	0.31	0.14	0.03	-3.66	-1.67	-0.32	0.75	1.18	0.78	0.54	0.26
Bootstrap Standard Error	0.06	0.08	0.05	0.01	0.16	0.18	0.14	0.08	0.07	0.13	0.11	0.06

Table 3.7 - Bootstrap Parameter estimates – Punda Maria Buffalo for 2-, 3- and 4-state mixture models

	delta1	delta2	mu1	mu2	sigma1	sigma2
MLE	0.45	0.55	-3.80	-1.03	1.25	0.80
90% CI Lower Limit	0.43	0.54	-3.86	-1.06	1.22	0.78
90% CI Upper Limit	0.46	0.57	-3.73	-1.00	1.29	0.82
Bootstrap Mean	0.45	0.55	-3.79	-1.03	1.26	0.80
Bootstrap Standard Error	0.01	0.01	0.04	0.02	0.02	0.01

	delta1	delta2	delta3	mu1	mu2	mu3	sigma1	sigma2	sigma3
MLE	0.42	0.08	0.50	-3.92	-2.13	-0.93	1.20	0.59	0.75
90% CI Lower Limit	0.38	0.02	0.22	-4.06	-2.46	-1.01	1.13	0.32	0.67
90% CI Upper Limit	0.45	0.36	0.55	-3.81	-1.37	-0.64	1.25	0.89	0.78
Bootstrap Mean	0.41	0.13	0.46	-3.93	-2.04	-0.88	1.19	0.60	0.73
Bootstrap Standard Error	0.03	0.11	0.10	0.10	0.34	0.12	0.04	0.19	0.04

	delta1	delta2	delta3	delta4	mu1	mu2	mu3	mu4	sigma1	sigma2	sigma3	sigma4
MLE	0.42	0.34	0.22	0.02	-3.89	-1.48	-0.59	0.60	1.21	0.72	0.54	0.28
90% CI Lower Limit	0.40	0.12	0.12	0.01	-3.99	-1.98	-0.84	0.44	1.16	0.49	0.42	0.20
90% CI Upper Limit	0.45	0.44	0.43	0.03	-3.79	-1.29	-0.51	0.71	1.26	0.80	0.65	0.35
Bootstrap Mean	0.43	0.30	0.25	0.02	-3.89	-1.56	-0.64	0.59	1.21	0.68	0.54	0.27
Bootstrap Standard Error	0.02	0.10	0.10	0.01	0.06	0.21	0.10	0.08	0.03	0.10	0.07	0.05

3.8.4 State Allocation

The initial state allocation was done by assigning the observation to the state with the maximum posterior probability using the 'Mixture Likelihood' clustering method. In order to judge the effectiveness of this method of allocation, a plot of the state allocation by time of day for the Punda Maria sable is shown in Figure 3.7. It shows a peak in resting between 3am and 5am, followed by a peak in foraging from 7am to 9am. There is a second peak in the foraging state around 5pm-7pm. The mixed movement and travelling states are more apparent during the morning hours, with another increase around 6pm-8pm. There is almost no travelling late at night and resting becomes the dominant state again from 8pm. This representation is slightly blurred when using the aggregate data since the sunrise and sunset times vary throughout the year.

These state allocations seem consistent with the typical behaviour of ungulates, with two foraging peaks during the hours just after sunrise and just before sunset and very little movement at night when predators are active. If journeys to water or some other resource are taken, these are likely to begin during the mid-morning which corresponds with the results obtained using the maximum posterior probability method of clustering the observations. This graph provides evidence that the results obtained using this modelling technique, are in fact making biological sense. Also the log-normal mixture model appears to provide a suitable method for determining the activity states of an animal based on the displacement distances between successive GPS location recordings, using the method of 'Mixture Likelihood' clustering.

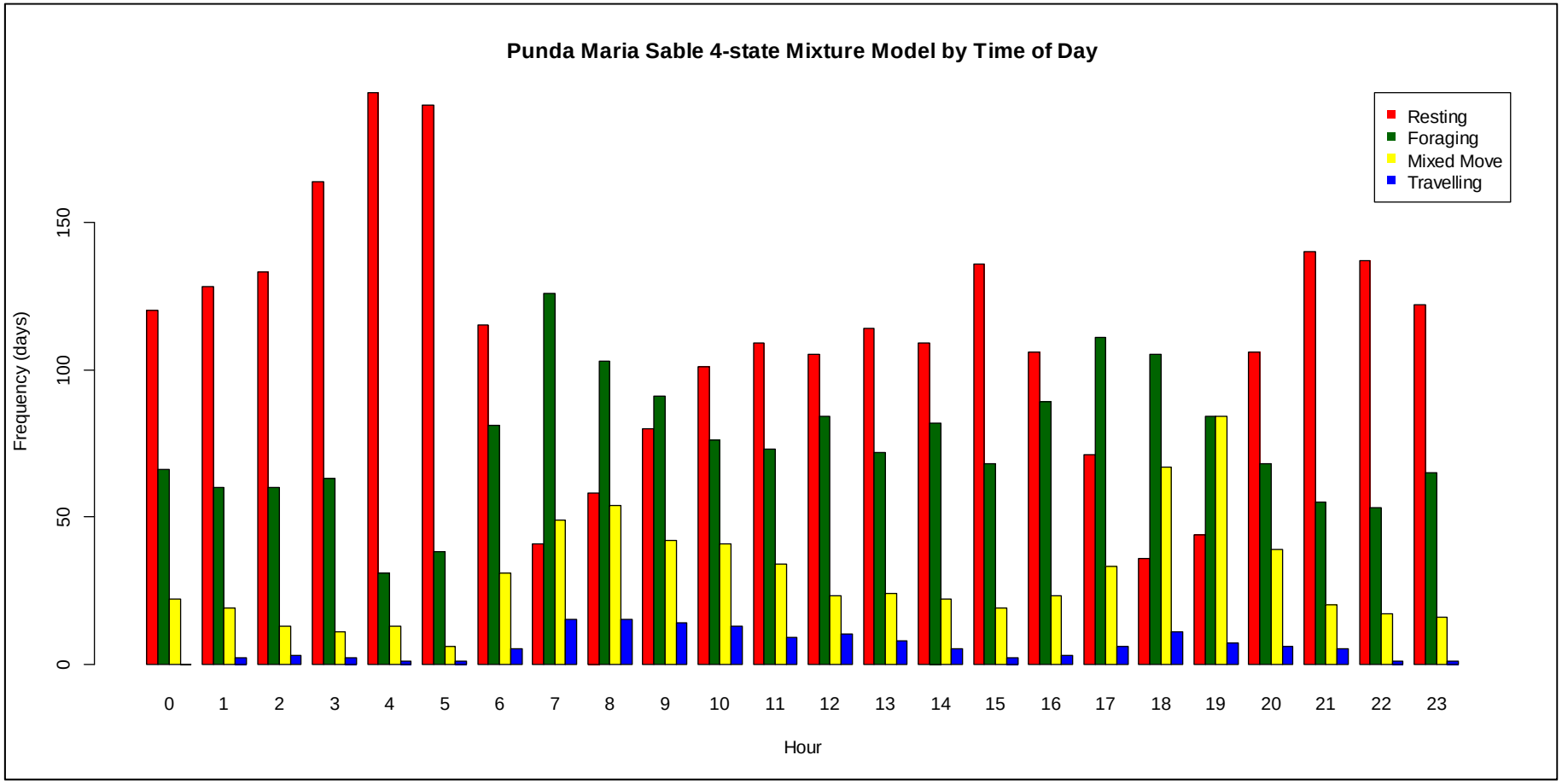

Figure 3.7 - Maximum posterior probability clustering by time of day - Punda Maria sable – 4-state Log-normal mixture

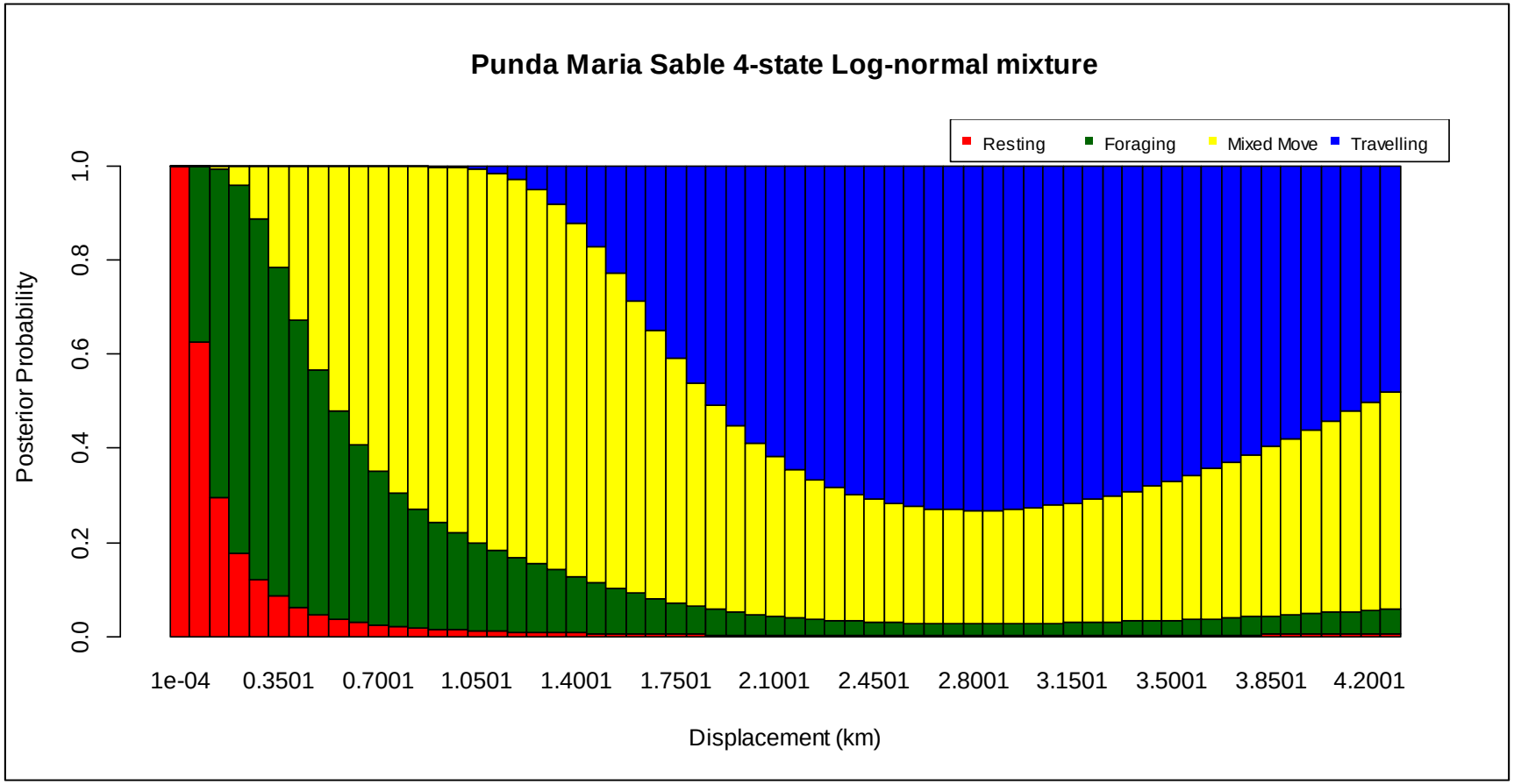

Figure 3.8 - Posterior probability of state membership – Punda Maria sable 4-state log-normal mixture

However, as discussed before, it is not realistic to expect an animal to remain within one behavioural state for the entire duration of the interval between observations, in this case one hour. A second state allocation was done, using the proportions of the posterior probabilities associated with each observation as a proxy for the amount of time spent in the corresponding activity state during the hour. Figure 3.8 shows the posterior probabilities for state membership for the model fitted to the Punda Maria sable. This plot illustrates the major drawback of the model when using a mixture of four log-normal distributions using the state allocation proportions. The extended right tails of the distributions coupled with the larger prior probabilities for the first, second and third distributions and the very small prior probability for the fourth state, means that the posterior probabilities actually increase again for very large observed values for the second and third states. This is clearly not a realistic situation, since a movement of for example 4km, cannot have a greater contribution from the mixed movement activity, than an observation of 2.7km. The highest state, in this case the fourth state, which is assumed to correspond to the animal travelling or relocating, should increase in contribution as the observed value increases, not decrease towards the maximum observation value as shown in Figure 3.8. From this, it is concluded that the mixture of log-normal distributions provides the closest fit to the observed data and provides a realistic picture of the behaviour of the animal using the maximum posterior probability method for clustering the observations. However, this mixture does not perform well when using the proportion of posterior probabilities as a proxy for the contribution of each state to the observed movement distance during one hour.

3.8.5 Offset Model

If the posterior probabilities are required to determine the contribution of each state to the observed distance, an alternative distribution to the log-normal is needed. The Weibull and Gamma distributions do not have as fat a tail as the log-normal, which means that the contribution of previous state will not extend as far to the right. However neither the Gamma nor the Weibull mixture provided a

good fit to the observed data due to the poor separation between the Gamma distributions which manifested in poor convergence of the maximization. In terms of the state contribution this means that very small observed movements would have some proportion allocated to the active movement states which is unreasonable. In fact, in many cases, it was not possible to obtain convergence when maximising the likelihood of a mixture of Gamma distributions.

Instead, an offset Gamma or Weibull distribution (also known as the three-parameter distribution) was used for the second and higher components within the mixture. Both the Gamma and Weibull mixture models were run using offsets informed by the breakpoints obtained from the segmented regression analysis of the log-frequencies of observations. The offset value was fixed and chosen to be slightly lower than the segmented regression breakpoints, to allow the left tail of the distribution to 'fade in'. This however is a reasonably subjective decision about the exact value of the breakpoint, and it will influence the model depending on the choice made. Using the breakpoints identified from the segmented regression as a guide for the mixture model offsets is at least useful as an objective way of informing the decision about the magnitude of the offset value. The model selection criteria, AIC and BIC, always selected the log-normal model in preference to the Gamma or Weibull models, with or without the offset (if convergence was possible without the offset). However when a state allocation of each observation into the contribution of each state is required, then the offset model provides more biologically meaningful results, since the fourth state does not decline in terms of contribution for increasing observed values as shown in Figure 3.9. The Gamma and Weibull offset mixtures were compared using the AIC and BIC values, and the offset Gamma mixture was always selected in preference to the offset Weibull mixture.

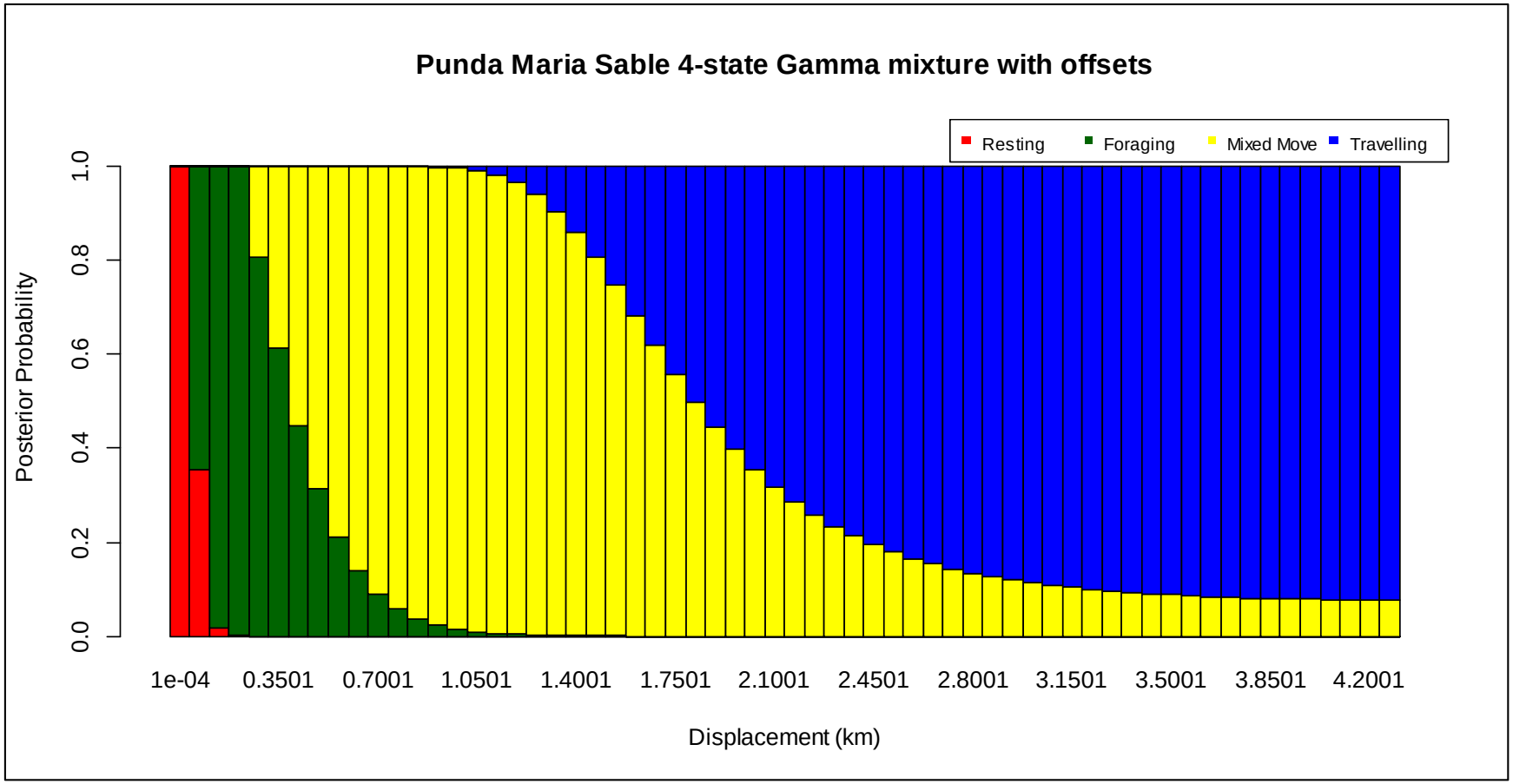

Figure 3.9 - Posterior probability of state membership – Punda Maria sable - Gamma model with offsets

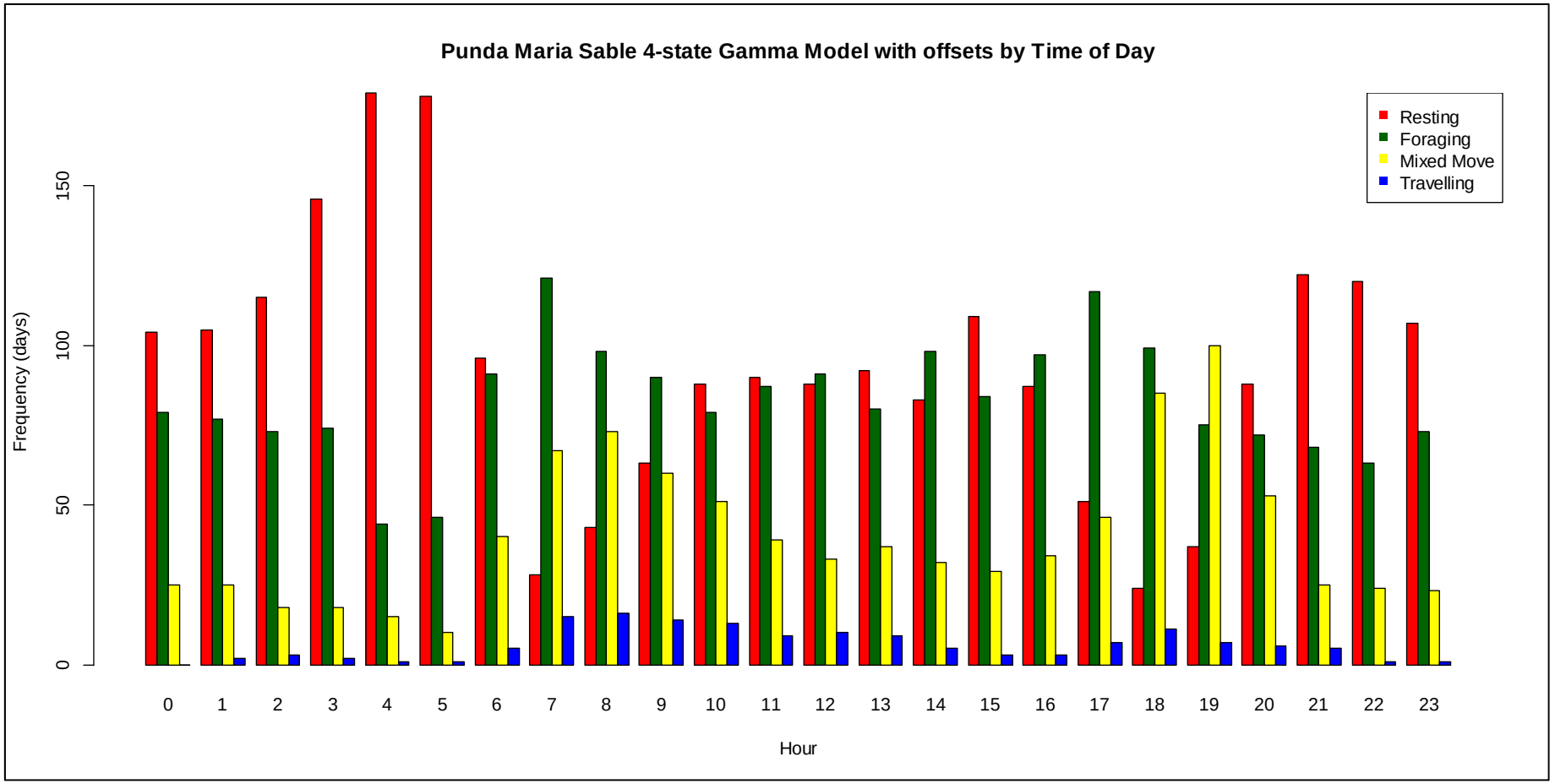

Figure 3.10 - Maximum posterior probability clustering by time of day - Punda Maria sable – Gamma mixture with offsets

Table 3.8 – Fixed offset, fitted parameter and expected distribution values for the Punda Maria sable 4-state Gamma Offset Model

Offset 1	0.05
Offset 2	0.24
Offset 3	0.65
p**1**	0.44
p**2**	0.35
p**3**	0.18
p**4**	0.03
α**1**	1.53
α**2**	1.26
α**3**	1.60
α**4**	6.94
β**1**	0.02
β**2**	0.11
β**3**	0.36
β**4**	0.24
Expected Value 1	0.03
Expected Value 2	0.14
Expected Value 3	0.58
Expected Value 4	1.67

For the Punda Maria sable herd, the 4-state Gamma with Offset Model was selected as the 'best' model using the AIC and BIC. The offsets used and the fitted mixing and distribution parameters are shown in Table 3.8. The expected values from the fitted distribution show that the states are associated with slightly shorter distances than the log-normal models but the behaviour inferred from the models would still correspond to the animal in a 'resting', 'foraging', 'mixed movement' and 'travelling' phase. The state allocations run for the Gamma model with offset using the maximum posterior clustering method are plotted by time of day in Figure 3.10. The pattern of allocation is very similar to that obtained using the log-normal mixture model which was shown in Figure 3.7.

However, the difference between the log-normal and the Gamma offset mixture is evident when plotting the posterior probability proportions of the membership of each component state, Figure 3.9. This state allocation is far more realistic if the

proportion of posterior probabilities is used as a proxy for the contribution of each state to the movement of the animal during the course of the observation interval. The longer movements have increasingly large proportions of the allocation going to the more active movement component. The foraging state has almost no proportion of time allocated to it for movements greater than 1.3km/h, while the mixed movement state declines in proportion gradually towards the maximum observed movement for the Punda Maria sable. The most active relocation state becomes the dominant state from around 1.8km/h. The above analyses show that despite the information criteria favouring the log-normal mixture and the maximum posterior probability clustering providing a very similar allocation of states, the posterior probability proportion allocation provides a far more biologically defendable scenario based on the Gamma model with offsets. This posterior probability proportion allocation method assumes that each observation is itself a mixture of underlying and unobserved states, compared with the usual mixture model assumption that the population is composed of a number of groups and each observation belongs to one of these unobserved groups.

It is acknowledged that this method of using the posterior probabilities as a proxy for the proportion of each state contributing to the observed displacement is not theoretically justifiable. However for the Mixture Model approach used here, there is no simple way to allocate the contributions of the different states to the observed movement. This will always be an issue when the time between locations is longer than the time the animal might spend in a particular state. This method is an attempt at answering the ecological question of how the animal behaved during the time between locations, and how different behavioural states could make up the total observed displacement between successive locations. Although this method is not ideal, it does provide an ecologically reasonable allocation of the observed displacement into a number of states. A further complication is that an observed displacement could be as a result of a number of different combinations of the various states. This is an important area for future study.

3.8.6 Seasonal Models

For each herd, the observed movement data was split into seasonal subsets representing the Wet, Early Dry and Late Dry seasons. The sample sizes within each seasonal subset are shown in Table 3.9. Data up until June 2008 were included for Zebra142 and Zebra277 only, to be comparable with the other Zebra herds.

Table 3.9 – Herd Sample sizes per season

Species	Region	Wet	Early Dry	Late Dry
Sable	Punda Maria	757	981	3592
Sable	Numbi	2562	336	2905
Sable	Phabeni	4947	3649	5489
Sable	Shitlave	2780	3085	3806
Buffalo	Punda Maria	4861	5269	5603
Zebra141	Punda Maria	0	273	726
Zebra142	Punda Maria	2751	2769	2227
Zebra277	Punda Maria	2640	2381	2388
Zebra 280	Punda Maria	2021	290	2197

The likelihood values, model selection information criteria and convergence for the Punda Maria Sable herd models are shown in Table 3.10. These models were run using the simpler log-normal mixture models, which do not require the model offsets as discussed for the Gamma mixtures above. In some cases it was not possible to obtain model parameter estimates due to the models not converging, for example the Wet season 4-state model. This suggests that there are insufficient underlying groups within the data to fit the required model. The lowest AIC and BIC values are highlighted in Table 3.10 in green, red and yellow for the Wet, Early Dry and Late Dry seasons respectively. Using AIC to determine the 'best model', the results show that a 2-state model is selected as the best model for the

wet season data, while a 4-state model provides the best fit for the Early and Late Dry seasons. This appears to be a reasonable result since it is likely during the Wet season that the animals do not have to make such long journeys to water due to surface water availability after the rains and the availability of more forage. The BIC however finds the simpler models as the best fit for the seasonal data. These results show that the overall best seasonal fit to the data might actually be a combination of the models with a different number of components for the different seasonal blocks.

Table 3.10 - Punda Maria Sable Seasonal Log-normal Mixture Model Fit

Species	Region	States	Season	Converge	-LogLik	AIC	BIC
Sable	Punda Maria	2-state	Wet	Yes	-622.61	-1235.21	-1212.07
			Early Dry	Yes	-986.45	-1962.91	-1938.47
			Late Dry	Yes	-1265.79	-2521.57	-2490.64
Sable	Punda Maria	3-state	Wet	Yes	-623.40	-1230.81	-1193.77
			Early Dry	Yes	-996.02	-1976.05	-1936.94
			Late Dry	Yes	-1291.75	-2567.50	-2518.01
Sable	Punda Maria	4-state	Wet	No			
			Early Dry	Yes	-999.06	-1976.13	-1922.36
			Late Dry	Yes	-1301.77	-2581.54	-2513.49

The model parameter estimates and standard errors for the "best" models selected by AIC are shown in Table 3.11. The Late Dry season and the aggregate 'Year data' (the model run for all data combined as in the previous sections) parameter estimates are very similar but this is expected since the majority of the year data is actually from the Late dry season. The Early Dry season data contains a different pattern of movement with the foraging state being split into a slow and then a faster foraging mode. The fourth state for the Early Dry season corresponds to the mixed movement (third) state of the Year model, with an expected displacement

distance of 0.9km. No relocation states are found during the Early Dry season. The Wet season has a clear resting state and then the second state encompasses some foraging and movement with an expected displacement of 660m. The results are consistent with the expectation that the animals are moving less during the wet season when conditions are good with plenty of water and food available. The late dry season is a time of stressful conditions on the animals and they are moving a lot more than at other times of the year, searching for food and water. The standard errors for the full data model are always smaller than for the seasonal models due to the larger volume of data available to fit the model. In all cases the confidence intervals around the parameter estimates do not overlap which implies that the states are well defined, even for the models run on the seasonal subsets of data.

Table 3.11 - Maximum Likelihood Parameter estimates for 'best' seasonal model and state-dependent means – Punda Maria sable

Season	***p*1**	***p*2**	***p*3**	***p*4**	***μ*1**	***μ*2**	***μ*3**	***μ*4**	***σ*1**	***σ*2**	***σ*3**	***σ*4**	**Mean 1**	**Mean 2**	**Mean 3**	**Mean 4**
Year data	0.52	0.32	0.13	0.02	-3.67	-1.64	-0.26	0.78	1.17	0.82	0.58	0.26	0.05	0.27	0.91	2.25
Std Error	0.06	0.08	0.05	0.01	0.16	0.18	0.14	0.08	0.07	0.13	0.11	0.06				
Wet	0.83	0.17			-3.15	-0.62			1.46	0.64			0.12	0.66		
Std Error	0.06	0.06			0.16	0.16			0.09	0.12						
Early Dry	0.50	0.25	0.21	0.04	-3.98	-2.26	-1.19	0.00	0.99	0.47	0.54	0.21	0.03	0.12	0.35	1.03
Std Error	0.06	0.14	0.12	0.01	0.18	0.23	0.31	0.09	0.09	0.19	0.18	0.06				
Late Dry	0.46	0.36	0.14	0.04	-3.73	-1.61	-0.27	0.76	1.12	0.86	0.56	0.27	0.04	0.29	0.90	2.21
Std Error	0.07	0.10	0.07	0.01	0.19	0.22	0.17	0.07	0.07	0.16	0.12	0.05				

3.8.7 Seasonal comparison

In order to determine whether the seasonal model split improves the fit to the observed data, a likelihood ratio test is performed. In this case the null hypothesis is that the model distributions for the seasonal models are the same, which implies that a seasonal model is not needed since the annual model would explain the movement adequately since the distributions do not change in the different seasons. Model fit has been compared using the full year's data, or running separate models for the individual seasons. This shows whether the model fit can be significantly improved using the additional parameters required for the seasonally fit model. However, it is not a formal comparison of whether the behavioural patterns actually differ between the seasons. In order to formally test this, a likelihood ratio test is done to compare the parameters of a single model fitted to the data from both seasons, and the models fitted separately to data from the dry season and the wet season. The null hypothesis is that the distribution of the data for the wet season is the same as the distribution for the dry season. The likelihood ratio test statistic is given by:

$$G^2 = -2\log\left(\frac{max\{\text{likelihood}|\text{H}_0\}}{max\{\text{unrestricted likelihood}\}}\right)$$

$$= -2\left(max\{\text{log-likelihood reduced}\} - max\{\text{log-likelihood full}\}\right)$$

where the likelihood under the alternative hypothesis is the joint likelihood of the wet season and dry season. $G^2 \sim \chi^2(df2 - df1)$ where df1 and df2 are the number of parameters in the reduced and full models respectively.

To demonstrate this testing process, the buffalo time series was used as it spanned more than a year. In order to highlight the potential difference in the movement, the movement was split into 3 months from the early and late dry seasons to represent winter movement (June, July and August) and 3 months from the wet season to represent summer movement (December, January and February). This split is arbitrary and could be as wide or as narrow as the researcher wanted, or

could have been consistent with the models run using the 3-season split. Data used for this comparison was selected to be from winter 2008 and summer 2008/9, due to fewer GPS locations being missed during this period. 2-, 3- and 4-state log-normal Mixture Models were fitted to the data separately for the two seasons and then one set of models using the combined data as the input. The AIC statistic was used to determine the "best" model for each dataset.

3.8.7.1 Punda Maria Buffalo

3-state log-normal mixture models were found to provide the best fit for the summer, winter and the combined dataset.

$$
\begin{aligned}
G^2 &= -2\left(\text{combinedLL} - \left(\text{winterLL} + \text{summerLL}\right)\right) \\
&= -2\left(-1768.09 - \left(-1217.231 + (-572.5689)\right)\right) \\
&= 43.42034 \\
G^2 &\sim \chi^2\left(df2 - df1\right) \\
G^2 &\sim \chi^2\left(13 - 8\right) \\
G^2 &\sim \chi^2\left(5\right)
\end{aligned}
$$

At a 5% level of significance $G^2 = 43.42 > \chi^2(5) = 11.0705$, therefore the null hypothesis of no difference in movement is rejected, and it is concluded that there is a highly significant difference in the movement of the buffalo herd between summer and winter ($p < 0.001$). This is an example of how to test for significant seasonal differences and could be repeated for all of the herds.

3.8.8 Herd Comparison

One of the questions which is likely to be posed from an ecological point of view about animal movement data used to infer behavioural patterns, is whether two herds or two species are spending significantly different amounts of time in the same behavioural state. The results from this type of hypothesis can be used to identify when one herd or species is under more stress than another, or just when the patterns for that species are significantly different from another. This can be

tested for each of the parameters from the fitted model using a test for the difference in population means for independent samples. A 2-sample *t*-test could be used which assumes the variances are equal and unknown. However for animal tracking studies, the datasets are always large (much greater than 30 observations) which means the approximate large sample *Z*-test can be used. This test has the added advantage that the variances do not have to be equal. For large samples (n_1 and $n_2 \geq 30$), the test statistic is given by:

$$Z = \frac{\left(\bar{X}_1 - \bar{X}_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \quad \text{when } \sigma_1^2 \text{ and } \sigma_2^2 \text{ are known}$$

or

$$Z = \frac{\left(\bar{X}_1 - \bar{X}_2\right)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1) \quad \text{when } \sigma_1^2 \text{ and } \sigma_2^2 \text{ are unknown}$$

For the fitted Mixture models, $\bar{X}_i$ are the estimated parameter values and σ_i^2 and s_i^2 are the variance and standard error of the parameter estimates obtained using the parametric bootstrap procedure.

In order to test the similarity or difference between multiple parameters, a one-way analysis of variance F-test is used, using the ratio of the 'Between Group Mean Square' and the 'Within Group Mean Square'. The test statistic is given by:

$$BG = \frac{\sum_{i=1}^{K} n_i \left(\hat{\delta}_i - \overline{\delta}\right)^2}{K-1} \qquad \text{where } \overline{\delta} = \frac{\sum_{i=1}^{K} n_i \hat{\delta}_i}{\sum_{i=1}^{K} n_i}$$

$$WG = \frac{\sum_{i=1}^{K} \left((n_i - 1) SE\left(\hat{\delta}_i\right) \right)^2}{\left(\sum_{i=1}^{K} n_i \right) - K}$$

$$F = \frac{BG}{WG}$$

$$F \sim F\left(K-1, N-K\right) \qquad \text{where } N = \sum_{i=1}^{K} n_i$$

In this case the δ_i represents any parameter of interest. This test will provide information on whether any of the parameters are significantly higher or lower than any of the other parameters. It will not identify which parameters are different, this would be determined using multiple comparison tests as discussed above for the difference in population means.

Pretoriuskop Sable

In the Pretoriuskop region, the Numbi, Phabeni and Shitlave sable herds live in relative close proximity to one another and will experience reasonably similar climatic and habitat conditions. The inference drawn from the mixing parameters is in terms of the amount of time that the animal spends within a particular activity state. The amount of time that an animal spends resting will be indicative of the stress levels of the animal and how benign the conditions are in which it is living. The test for the difference in populations means is used to test whether the mean time spent resting is significantly different from another herd, i.e. whether the δ parameters are significantly different for the separate herds. This could be done for any state.

Using the full dataset, the Numbi and Shitlave sable herds had very similar mixing parameters (Table 3.4). However, to test this formally, a null hypothesis that the two herds spend the same amount of time in the resting state is tested. This corresponds to:

H_0: $\delta_1 S = \delta_1 N$

H_1: $\delta_1 S \neq \delta_1 N$

where $\delta_1 S$ denotes the fitted mixing parameter for the first state for the Shitlave herd and $\delta_1 N$ for the Numbi herd. No evidence was found at the 95% confidence level to suggest that these herds spent a different amount of time in the resting state

(Z_{obs} = -0.056 > -$z_{0.025}$ = -1.960)

and hence the null hypothesis is not rejected. In contrast for the Shitlave and Phabeni sable herds, the null hypothesis is rejected

(Z_{obs} = -1.982 < -$z_{0.025}$ =-1.960)

with the inference that the Phabeni herd spends a significantly longer time in the resting state than the Shitlave herd. Similar tests could be carried out for every estimated parameter, using the parameter estimate and the standard error of the estimate.

In order to compare the resting behaviour of all 3 sable herds in the Pretoriuskop region, the null hypothesis is given by:

H_0: $\delta_1 N = \delta_1 P = \delta_1 S$.

Using the between and within group mean squares, we fail to reject the null hypothesis and conclude that at the 95% level of significance, that the amount of time spent resting by all 3 herds is the same.

F_{obs} = 1.919 < F(2, 29 556) = 2.996.

This is in contrast to the pairwise tests between the herds where the Phabeni herd was found to spend a longer amount of time in the resting state. This implies that although the Phabeni herd is found to rest more the Shitlave herd, when the movements are compared across all three herds including the Numbi herd, there are no significant differences in the movement across the three herds. This apparent contradiction comes from the multiple comparison tests, in this situation the ANOVA is the more appropriate test to compare the movement of the herds.

3.9 Conclusion

In the analysis of GPS movement data using mixture models, the component distributions describe the movement modes for the animals, and these can be interpreted as different behavioural or activity states. One of the challenges of animal movement analysis is to link the statistical patterns obtained from the analysis to the behavioural characteristics of the animal, in order to gain insight into the underlying ecological processes (Nathan 2008). The mixture models have been used to uncover the underlying and unobserved behaviour of the animal. The modelling technique provides a straightforward process for taking the locations of the animals and inferring the behaviour associated with the movement patterns. There is always a danger that the mixture models tend to be over-interpreted (Zucchini, MacDonald 2009). It is used here as a clustering technique, and there is always the associated uncertainty about cluster membership and the interpretation of each cluster.

One of the main drawbacks of the mixture model technique for the analysis of GPS animal movement data is that it does not take into account the time series nature of the input data and ignores the serial correlation between successive observations. While there is a strong serial correlation in the displacements between successive location points (shown in Chapter 4), large ungulates tend to

only remain in a particular behavioural state for several hours which only covers a few successive location observations.

Despite the fact that this serial autocorrelation within the data is ignored, the mixture models have been able to uncover patterns within the data. Clear peaks in foraging behaviour were shown for all three species after sunrise and before sunset. The three species studied in the Punda Maria area live in the same region of the Park, but utilize the area differently and display different patterns in the amount of time spent in each behavioural state.

Models run for the separate seasons are able to pick up the seasonal variation in behaviour. The method of mixture modelling for animal movement is flexible enough to be applied to seasonal data and pickup seasonal variation but remains consistent with the results from the annual model. Using the state allocation data, it is possible to conduct a spatial analysis of the areas utilized by the animals for the various behavioural states. This extension of the mixture model classification analysis is an important one which can inform habitat and resource utilisation studies. This type of analysis could also be supplemented by direct observation of the animals if possible which would reveal the actual activity state. Activity collars are also available which record whether the animal's head is up or down at the time the location reading is taken. If their head is up, the animal is likely to be moving and when it is down, it is likely to be feeding. This technology could start revealing the underlying hidden states, and the mixture model's predicted state allocation could be checked for accuracy against the observed states especially with shorter inter-observation intervals. These methods can also be modified and adapted to investigate the activities of the animals in slightly different ways (Owen-Smith, Goodall & Fatti 2012).

The mixture of log-normal distributions fitted the observed data for these species the best, however this might change for different species or different time periods between successive locations. However, depending on the assumptions about the state allocation, a mixture of Gamma models with offsets was preferable when the contribution of each state to a single observation is required. The failure to take into account the time series nature of GPS tracking data needs to be addressed. Further work using Hidden Markov Models, also known as dependent mixture models, is shown in Chapter 5. A comparison of the results of these two methods will indicate how influential the decision is to ignore the dependence within the data for the mixture model analysis.

4 TIME SERIES ANALYSIS INTRODUCTION AND CORRELATION STUDY

4.1 Introduction

One of the commonly acknowledged challenges of animal tracking analyses is the serial correlation between successive observations. Observations are generally not independent in time or space (Boyce et al. 2010). One of the solutions when dealing with serially correlated data is to subsample from the original dataset. If observations are selected with sufficiently large time intervals between them, there will be far less serial correlation in the dataset used for the analysis, and therefore statistical methods which assume independence between observations are applicable. However, given the costs and risks associated with the GPS tracking collars, throwing away data is not desirable and is a waste of information. In fact the pattern of correlation within the data can actually provide valuable information about the behavioural patterns of the animal being studied. One of the challenges for the analysis of GPS tracking data is to utilize time series techniques which take into account the dependency between observations, and actually use the dependency to add strength to the inferences and interpretations of the model output. Time series techniques used in the analysis of animal movement data include correlated random walks (Morales et al. 2004, Bergman, Schaefer & Luttich 2000, Haydon et al. 2008, Mårell, Ball & Hofgaard 2002), Hidden Markov models (Franke, Caelli & Hudson 2004, Franke et al. 2006, Patterson et al. 2009, Pedersen et al. 2008, Hart et al. 2010), hierarchical Bayesian state-space models (Eckert et al. 2008, Jonsen, Myers & James 2006), Ornstein-Uhlenbeck method (Brillinger et al. 2004, Preisler et al. 2004) and Lévy walks (Atkinson et al. 2002, Mårell, Ball & Hofgaard 2002, Viswanathan et al. 1996, Viswanathan et al. 1999) among others.

Animal movements can be correlated both spatially and temporally. Spatial autocorrelation implies that the next location will be related to the current location. This can be studied through the presence or absence of an animal in a pixel of a landscape (Boyce et al. 2010). Temporal autocorrelation will apply to time series measures such as the step length and turning angles between successive location estimates, where successive movement metrics are likely to be related to one another through time.

4.2 Literature

Surprisingly few authors have actually used the patterns of correlation within the data to gain insight into the movement and behavioural patterns of the species being studied. Polansky et al. (2010) used Fourier and wavelet transforms to investigate patterns within animal's movement (Polansky et al. 2010). They found that the spectral signatures given by these techniques were useful for characterizing temporal dependency in the movement data. They were able to link the temporal patterns of the movement of lion (*Panthera leo*) in the KNP to the moonlight brightness, something that is known to influence the hunting success of lion (Funston, Mills & Biggs 2001). They ran similar analyses for the movement of buffalo in 6 different herds in the KNP. They found that even when the movement could be described as an uncorrelated random walk, there was a synchrony in the movements when the animals came within one kilometre of one another (Polansky et al. 2010).

In a similar study, Wittemyer et al. (2008) also used the Fourier and wavelet analyses to study the patterns of autocorrelation within elephant movement data. They examined the impact of social rank, predation risk and season on the patterns of correlation. They found that the correlation was weaker during the wet seasons, and that the patterns were more consistent for higher ranking individuals (Wittemyer et al. 2008).

Cushman et al. (2005) studied autocorrelation patterns of elephant (*Loxodonta africana*) movement in Botswana, to investigate seasonal variability in the scale and pattern of the movement (Cushman, Chase & Griffin 2005). They used Mantel correlograms as they felt these were a suitable analysis technique for movement data as they do not assume a fixed elliptical home range and have a significance test. They used sliding windows over the data to examine the autocorrelation as animal movement data is not expected to be stationary over long time series (Cushman, Chase & Griffin 2005). Their analyses showed that the autocorrelation pattern of the three elephant herds were very similar, and varied across the season, in particular being influenced by rainfall events. They note that the shape of the correlogram can provide insight into the type of behaviour of the animal during the analysis period, namely periodic, directional or random (Cushman, Chase & Griffin 2005). They mention that the patterns of autocorrelation within animal movement data should not be considered a problem, but rather an opportunity to gain insight into the patterns of the movement and how it varies with time. This will lead to a better understanding of the movement ecology of the species.

Boyce et al. (2010) studied the autocorrelation patterns of wolves (*Canis lupus*), cougars (*Puma concolor*), grizzly bears (*Ursus arctos*) and elk (*Cervus elaphus*) in Canada using the step lengths derived from GPS location coordinates. They found that the pattern of autocorrelation often differed by season. The patterns for predators tended to decay towards random movement in wilderness areas, although showed strong autocorrelation patterns in areas disturbed by humans. The elk movement showed a strong daily pattern as a result of crepuscular activity (activity around dawn and dusk). Through the analysis of the autocorrelation structure within the movement data, the authors were able to use this very simple analysis technique to gain insight into the behaviour of these species, how the patterns differed between individuals of the same species, how the species differed and also the influence of human disturbances on the behavioural patterns (Boyce et al. 2010).

Of these animal movement studies which investigate the autocorrelation structure of the data, two use Fourier and wavelet transforms to study the patterns. These analyses require complete time series data which means that any missing location needs to be estimated and included in the dataset. In this thesis, interpolation to estimate missing locations has been avoided and hence these methods are not applicable for this study. The other studies have used graphical techniques to compare the patterns of the autocorrelation function across species and season, which will be applied to the Kruger National Park data. This study aims to investigate whether patterns in autocorrelation exist for these three species and whether they persist across the seasons. Few studies have looked at a formal test to compare patterns of autocorrelation, in this study the herds are compared using a simple correlation coefficient. Further work is required to find a more comprehensive method to formally compare the patterns of autocorrelation across the seasons and across species or individuals.

4.3 Correlation

The autocorrelation function (ACF) is a commonly used tool in the investigation of temporal dependency within a time series. The ACF describes the linear relationship between X_t and X_{t+h} for different lags h. This means it looks at the similarity of observations as a function of the time separating them. If the series is stationary, the autocorrelation depends only on the difference between t and h, and not on t itself.

For observations in a time series $x_1, x_2, \ldots, x_n$, the sample mean is:

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$$

and the sample autocorrelation function at lag h is given by:

$$\hat{\rho}(h) = \frac{\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

ρ ranges between -1 and 1 where 1 indicates a perfect linear relationship and -1 a perfect negative relationship. The autocorrelation provides a scale-free measure of the strength of the statistical dependence within the data. For a stationary time series, the autocorrelation will converge to zero as h increases once any regular trend and seasonality has been removed.

Polansky et al. (2010) note that one of the drawbacks of the ACF in the analysis of animal movement data is that it will be sinusoidal for data with cyclic switches between behavioural modes. They found that Fourier analysis was better suited to identifying dominant frequency patterns in movement data (Polansky et al. 2010). One of the advantages of using the ACF to investigate the dependency within animal movement data is that missing data can occur within the data and the ACF can be calculated despite the missing observations. Fourier analysis as used by Polansky et al. (2010) cannot cope with missing data. For an animal movement study where some locations' fixes have been missed (see section 2.3.1), these locations have to be estimated in order to run a Fourier analysis. Polansky et al. (2010) estimated the longitudinal and latitudinal coordinates using a Kalman smoother obtained from a state-space model. Linear interpolation between missing values could also be used.

The purpose of this KNP study is to use the animal movement data to differentiate the behavioural modes of the animals and to contrast the behaviour of the three species. A similar approach is used as Boyce et al. (2010), where the ACF is plotted for all the species and compared across region and species using a visual comparison of the plots.

4.4 Methods

One of the acknowledged problems with the analysis of the KNP tracking data using independent mixture models (Chapter 3) is that the technique ignores the serial dependency within the data. Independence of the observations is one of the key assumptions for the mixture model but for the purposes of that chapter this assumption was completely ignored and not investigated. Successive step lengths are likely to be autocorrelated as an animal is likely to continue moving in a similar way to how it is currently moving (Boyce et al. 2010), particularly when the time intervals between successive locations are short.

Unlike the independent mixture model analysis where missed locations could be removed from the dataset, time series analyses of animal movement data needs to take into account when observations were missed. The hourly observations used in Chapter 3 were formatted into a time series dataset with equal hourly intervals between each observation. If the location coordinates for a particular fix were not successfully obtained, this observation was recorded in the time series dataset as missing. If one location is missed, this results in two missing displacements since these cannot be calculated.

R software (R Development Core Team 2011) was used to run all of the autocorrelations. Missing observations were accounted for in the analysis using the `na.action=na.pass` argument in the `acf` function in R. Autocorrelations were calculated up to a lag of 120 hours (five days) provided that sufficient data were available. Due to the fact that animal movement data are unlikely to be stationary over long periods of time (Cushman, Chase & Griffin 2005), the autocorrelations were also calculated separately for each season (see Section 2.5.1 for explanation of the seasonal split).

In order to formally compare the autocorrelation patterns of the species and herds, a Pearson correlation coefficient was calculated between the vectors of autocorrelation up to lag 24 for pairwise comparisons of the daily correlation pattern between all herds. This was repeated for vectors of length 120 (five days). For the one day test, this investigates how similar the autocorrelation pattern and hence the movements over a day are for all of the herds. The five day correlation will provide information on how similarly these patterns persist over time. It is expected that the correlation coefficients will all be quite high since all the species studies are ungulates that follow quite a similar daily cycle.

4.5 Results

The Punda Maria sable show a strong daily cycle in their movement (Figure 4.1), particularly from 48 hours with very little decay after five days (lag 120). There are weaker positive rounded peaks around 12 hours, which is consistent with the autocorrelation patterns for Elk in Canada (Boyce et al. 2010). Boyce et al. (2010) attributed this to the changing sunrise and sunset hours which change through the year. In the seasonal plots of the autocorrelation these rounded peaks are not as obvious (Figure 4.1) which agrees with their findings. The rounded peaks indicate periods of similar movement, rather than a clear sharp peak where activity is very consistent from day to day in this case. These weaker peaks and troughs suggest behavioural switching during the course of the day as the type of movement changes. The patterns are not consistent from year to year when run seasonally, which is likely to be caused by changing climatic conditions from year to year. 2007 was a below average rainfall year, and the movement is somewhat more regular in this year compared to 2006 which was above average rainfall. This suggests that behaviour is more consistent for this sable herd during tough years when they have to regularly search for food and water.

The Numbi sable herd has a similar pattern of movement autocorrelation to the Punda Maria sable herd (Figure 4.2), with positive autocorrelation peaks showing a clear daily (24 hour) cycle, although the rounded 12-hour peaks do not occur for the Numbi sable. Seasonal patterns are highly irregular, with the Early and Late Dry season in 2007 being the only seasons with significant correlations. The Phabeni herd has a far more regular ACF (Figure 4.3), with very clear 24 hour peaks which have not decayed much after five days. As with the other sable herds, the pattern is far less obvious when split across the seasons. The Shitlave herd has a very similar autocorrelation pattern to the Phabeni herd when plotted for all the data (Figure 4.4). The patterns are irregular in 2006, but in 2007 the 24 hour cycle is evident except in the wet season.

The Punda Maria buffalo herd has a very distinctive and different pattern of autocorrelation (Figure 4.5). There are very clear, positive strong autocorrelations at 24 hours, for the daily cycle. This does not decay much up to five days. Between these very strong peaks are significant double humps to the autocorrelation pattern. These also are as strong after five days. While the period of the cycle during a 24 hour period varies across the seasons, there is always a very regular and strong cyclic pattern to the data. Unlike the sable who in some seasons had very little pattern to the movement autocorrelation, there is always a pattern to the buffalo movement, which suggests that they vary their behaviour through the seasons far less than the sable, and always move in a cyclic manner, no matter what conditions they experience. Polansky et al. (2010) found that the buffalo they studied showed two or three relatively active periods during the day which seems to be consistent with the autocorrelation peaks during the day, preceded and followed by the negative peaks which would be associated with resting. Buffalo are ruminants who will spend periods of the day resting to “chew the cud”. The patterns seen in Figure 4.5 seem to be consistent with this active and then resting behaviour.

All 4 zebra herds showed very different patterns of movement autocorrelation (Figure 4.6 to Figure 4.9), with only Zebra142 and Zebra277 showing clear daily cycles. The patterns vary quite considerably through the seasons indicating changing movement strategies with different conditions, and no clear pattern across all the animals. Zebra141 does not show a cyclic pattern, but this data series was much shorter than the other zebra herds.

The sable herds showed more irregular movement during the wet season, when food and water will be more easily accessible. Zebra142 also showed this irregularity during the wet season compared to the dry season movements; however this was not true for Zebra277. Nearly all the animals show a 24 hour daily cycle to their movement, as expected. The buffalo show very regular patterns through the year, which is consistent with ruminant behaviour. It is expected that this consistent pattern of autocorrelation would hold for other buffalo due to the nature of their behaviour, however, the zebra herds show quite different behavioural patterns across the four herds.

The results for the herd correlation comparison are shown in Table 4.1 and Table 4.2 for the one day and five day pairwise comparisons respectively. Only the top diagonal of part of the table is shown since the lower half would be identical. The results for Zebra141 are excluded from the discussion below due to the small sample size for this herd which resulted in slightly unusual correlation patterns. For the correlation coefficients up to lag 24, the highest values are for the Pretoriuskop sable herds with one another, as well as the zebra herds with one another. The weakest correlations are for Buffalo152 with both the sable and the zebra herds. The pattern is similar for the results up to lag 120. The highest correlations are for the Pretoriuskop sable herds with one another and between Zebra142 and Zebra277. The lowest correlations were between Zebra142 and Sable279 and then between the buffalo herd with zebra and Shitlave sable herds.

As expected, all of the correlations were high across all of the herds, and the comparison using the Pearson correlations does not really differentiate the species and herds. However the results obtained are consistent with the idea that the buffalo pattern is far more regular than for any of the other herds. Species with multiple collars in similar regions did tend to have higher correlations with one another.

4.6 Discussion

The autocorrelation patterns have allowed a comparison of the species and regions and suggested that behavioural switches might be occurring in these herds. Very different patterns of autocorrelation are present during the different seasons and years. This suggests that the movement patterns are certainly influenced by the environmental conditions. Harsh conditions during the late dry season are not just forcing the animals to change the distance that they are walking, but also the patterns of how they move. The necessity of trekking to water during the late dry seasons seems to force the animals to move in a more cyclic pattern, whereas in the wet seasons less obvious patterns can be seen in the autocorrelations of the movements. This suggests that random movements are more common during the wet season when more forage and water are available.

Interesting differences can be seen between species in the same region. The Phabeni and Numbi herds, although located close to one another geographically, have for example very different patterns for the wet season in 2007/8. The Shitlave and Phabeni herd have far more similar patterns of autocorrelation. The autocorrelation pattern for the buffalo herd appears to be very distinctive, and future studies of GPS buffalo locations would be able to investigate whether this is a regular pattern for this species. Of all the herds studied here, the buffalo is the only one with a large number of individuals making up the herd (±400 animals). It is likely that larger herds will have more synchronized and consistent movement

patterns, this is also likely to be associated with ruminant animals. The patterns of the zebra were much more inconsistent with two of the herds showing a daily cycle and two others being far more inconsistent through time. For all herds, the patterns varied through the seasons and years.

For some species, the presence of cubs or calves could influence the autocorrelation patterns seen in their movement. Boyce et al. (2010) identified a female grizzly bear with a cub that had a very strong diel cycle which was very different to the patterns of other bears tracked in their study. They concluded that the female bear could have changed her behaviour to avoid human and male bear contact due to the presence of a cub. Calves and foals of buffalo and zebra respectively will join the herd soon after birth and hence their presence is unlikely to affect the movement patterns of the mother. Sable antelope however leave their calf in a hiding place for a few weeks, returning periodically to feed the calf (Apps 1992). It is possible that the mother's movement during these times of returning to suckle a calf could produce unusual autocorrelation patterns in the movement data.

During most of the seasons for all of the species, there is strong autocorrelation within the observation data, showing a daily cycle. This is expected for GPS tracking data with short intervals between observations. It is also provides clear evidence for the fact that time series analyses of the data are needed to take into account the dependency within the data. The mixture model analysis presented in Chapter 3, although producing realistic results which can be interpreted ecologically, is ignoring this dependency. The strong autocorrelation within the observation data violates one of the key assumptions of the independent mixture models.

4.7 Conclusion

Very few studies have investigated the patterns of autocorrelation within animal movement telemetry data. With the advances in technology and battery power, it is now possible to collect more data with much shorter inter-observation intervals. This means that there will be stronger autocorrelation between successive observations which provides an opportunity for studying these patterns of association within the data. A study of the autocorrelation within animal movement data can provide insight into the behavioural patterns of the animal and assist with interpretations of the movement. It also provides an opportunity to link the autocorrelation patterns to environmental covariates. These autocorrelation patterns can be studied in conjunction with other time series analyses of the movement data, and should become an integral part of movement ecology studies. As the number of studies of various species increases, it will be possible to compare the autocorrelation patterns across numerous herds of the same species, and investigate whether a "signature" pattern exists for that species. The strong cyclic patterns within the data provide evidence for the necessity of using time series methods to analyse these types of data. While the independent Mixture Models in the previous chapter provided very realistic results and are reasonably simple to model, the natural extension to take into account this strong autocorrelation is the dependent Mixture Model, known as the Hidden Markov Model.

Table 4.1 - Pearson correlation coefficients for pairwise comparison of autocorrelation patterns up to lag 24

	PK_Numbi	PK_Phabeni	PK_Shitlave	PM_Buffalo152	PM_Zebra141	PM_Zebra142	PM_Zebra277	PM_Zebra280
PM_Sable279	0.969	0.943	0.945	0.926	0.981	0.968	0.973	0.974
PK_Numbi		0.990	0.994	0.945	0.927	0.983	0.979	0.986
PK_Phabeni			0.993	0.955	0.897	0.964	0.962	0.980
PK_Shitlave				0.939	0.901	0.975	0.969	0.978
PM_Buffalo152					0.885	0.917	0.944	0.959
PM_Zebra141						0.949	0.953	0.947
PM_Zebra142							0.991	0.979
PM_Zebra277								0.990

Table 4.2 - Pearson correlation coefficients for pairwise comparison of autocorrelation patterns up to lag 120

	PK_Numbi	PK_Phabeni	PK_Shitlave	PM_Buffalo152	PM_Zebra141	PM_Zebra142	PM_Zebra277	PM_Zebra280
PM_Sable279	0.953	0.935	0.933	0.859	0.823	0.823	0.938	0.927
PK_Numbi		0.968	0.973	0.858	0.774	0.942	0.936	0.935
PK_Phabeni			0.977	0.894	0.780	0.930	0.940	0.938
PK_Shitlave				0.845	0.768	0.959	0.949	0.925
PM_Buffalo152					0.705	0.764	0.828	0.837
PM_Zebra141						0.800	0.883	0.850
PM_Zebra142							0.961	0.899
PM_Zebra277								0.956

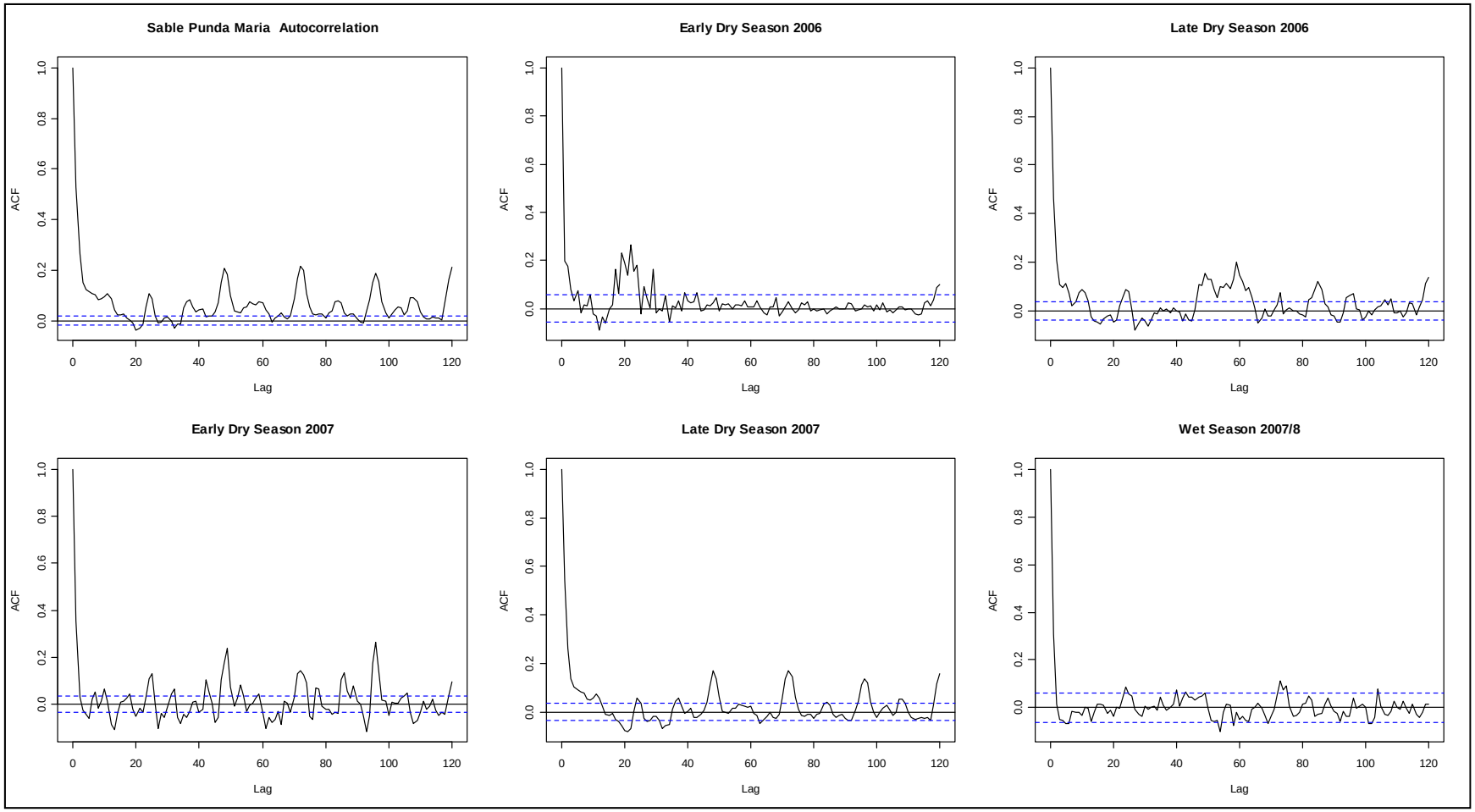

Figure 4.1 – Sable Punda Maria – Autocorrelation

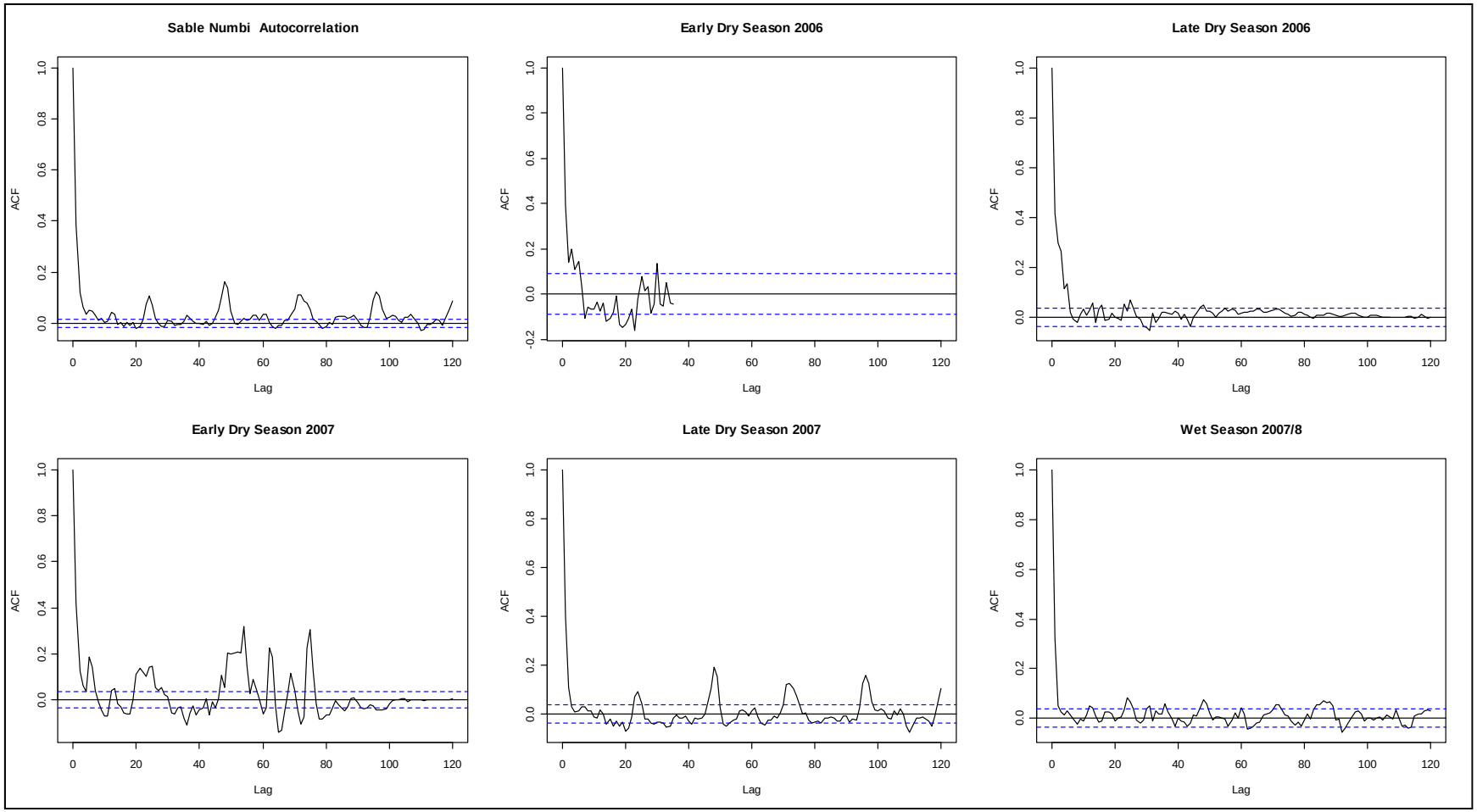

Figure 4.2 – Sable Numbi – Autocorrelation

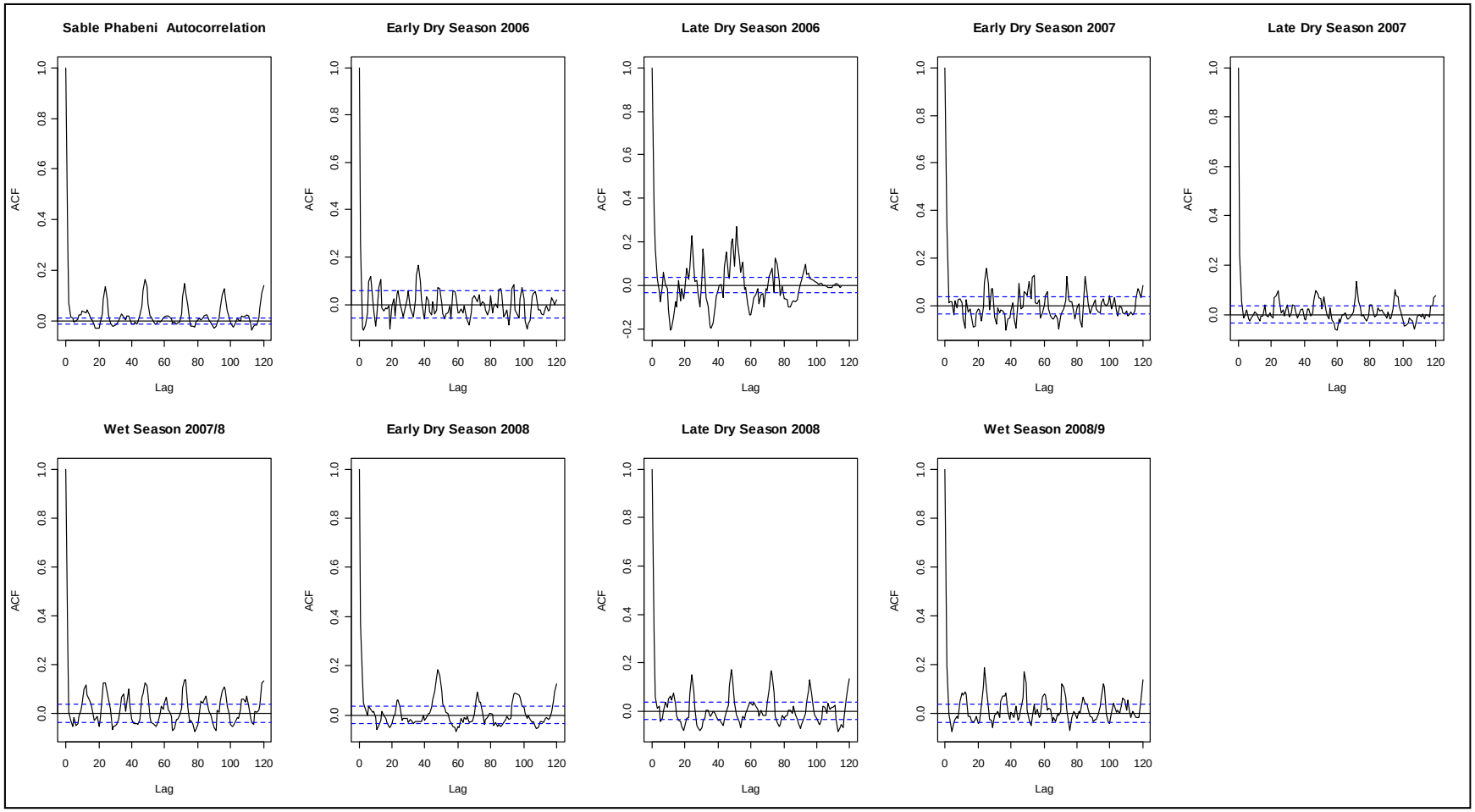

Figure 4.3 – Sable Phabeni – Autocorrelation

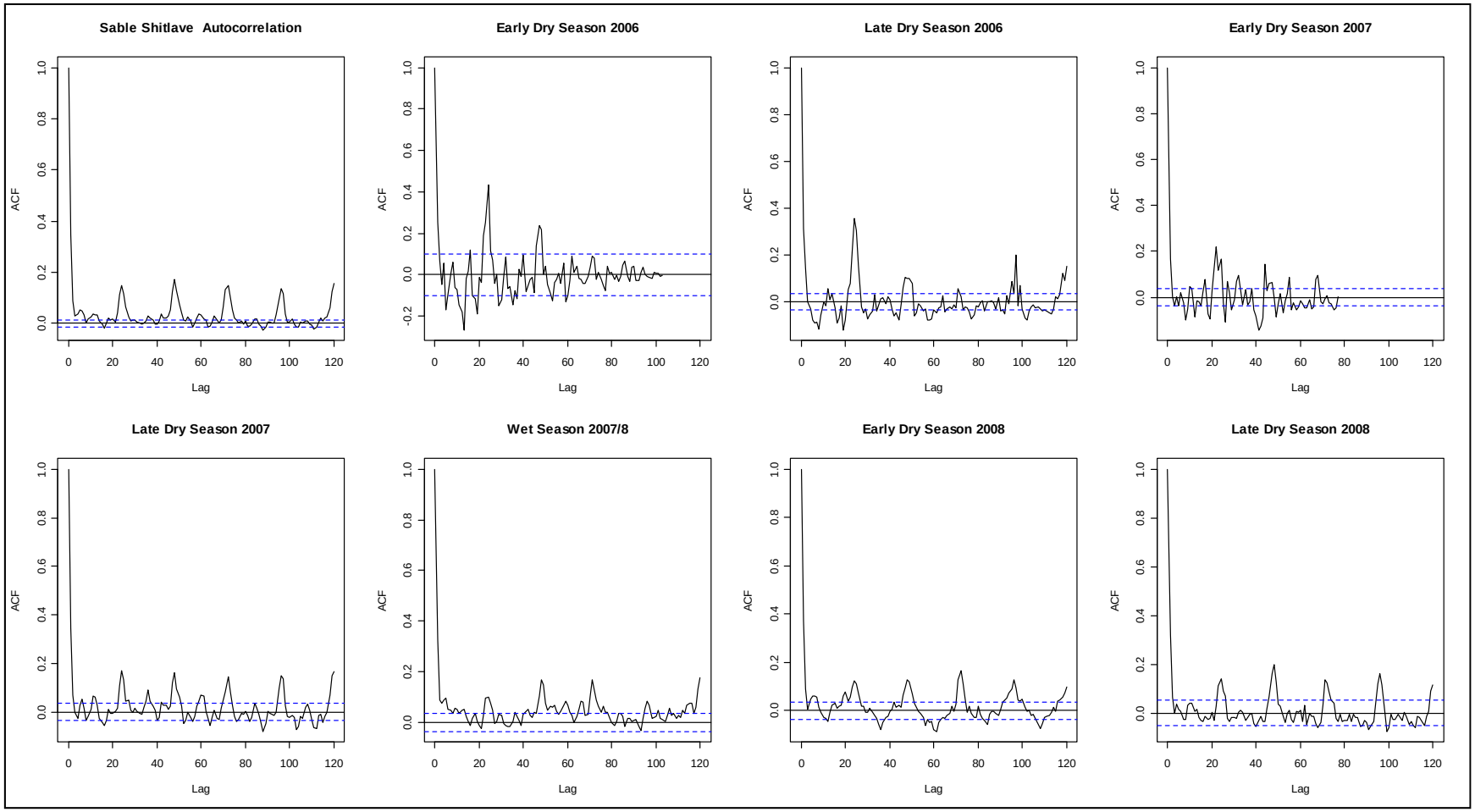

Figure 4.4 – Sable Shitlave – Autocorrelation

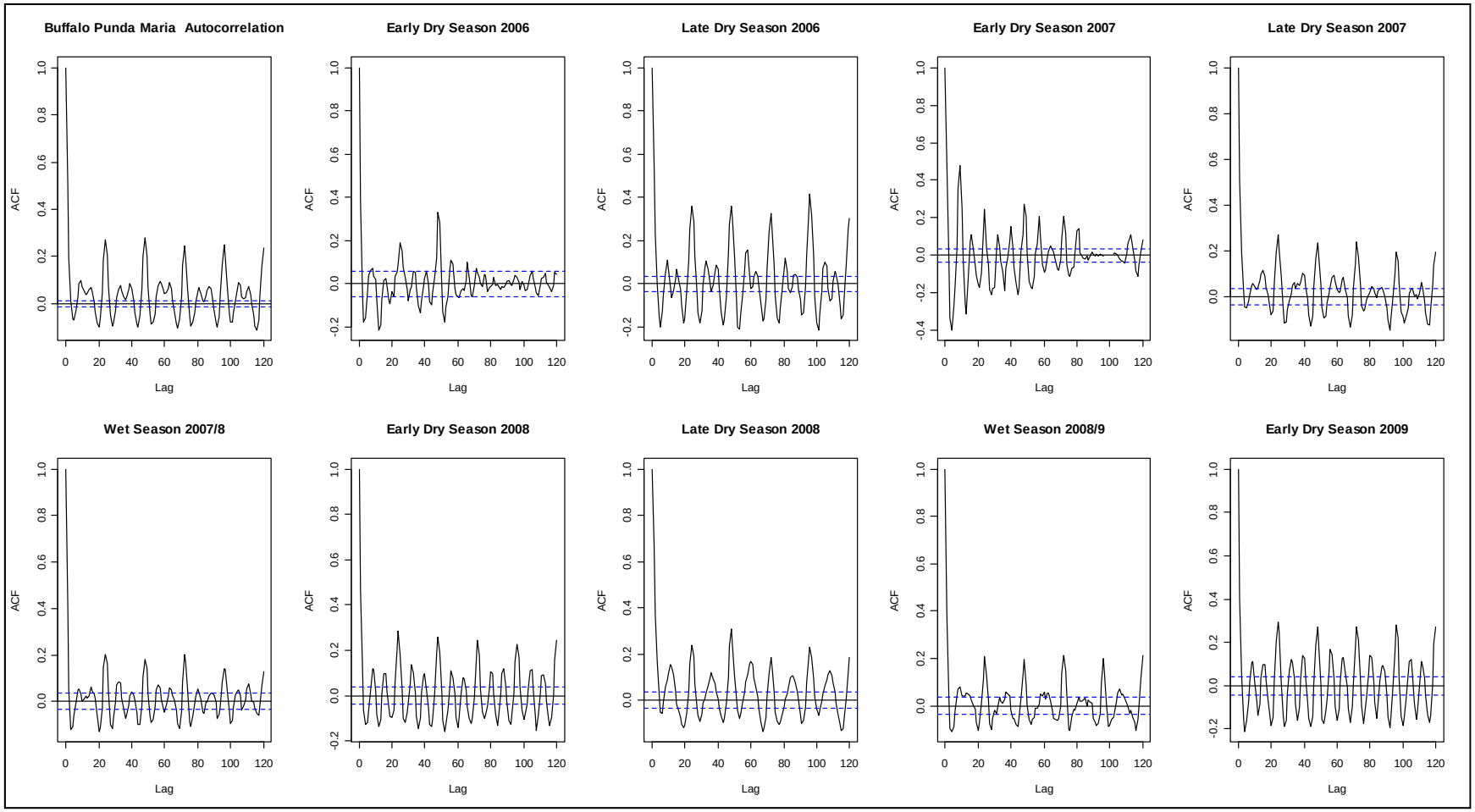

Figure 4.5 – Buffalo Punda Maria – Autocorrelation

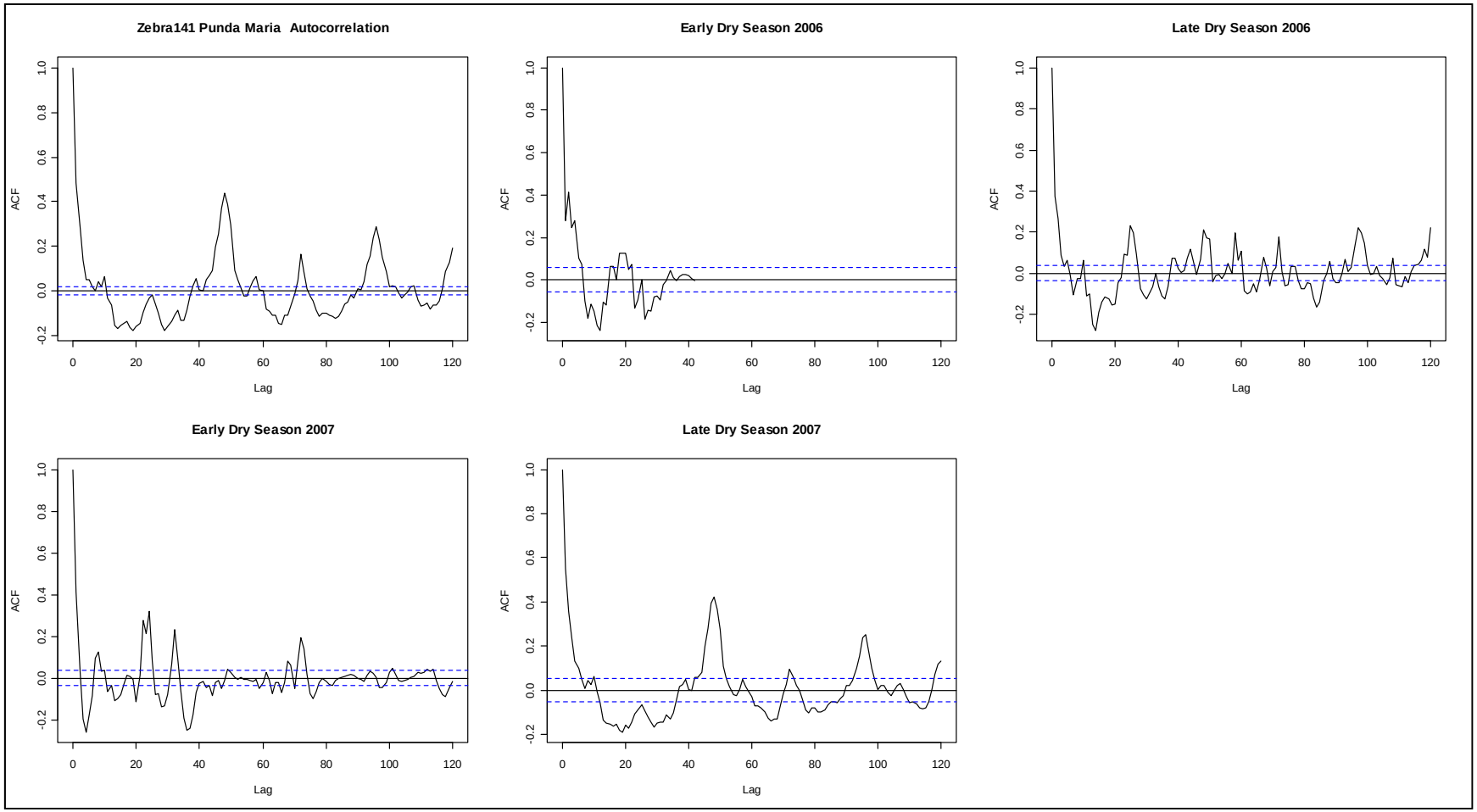

Figure 4.6 – Zebra141 Punda Maria – Autocorrelation

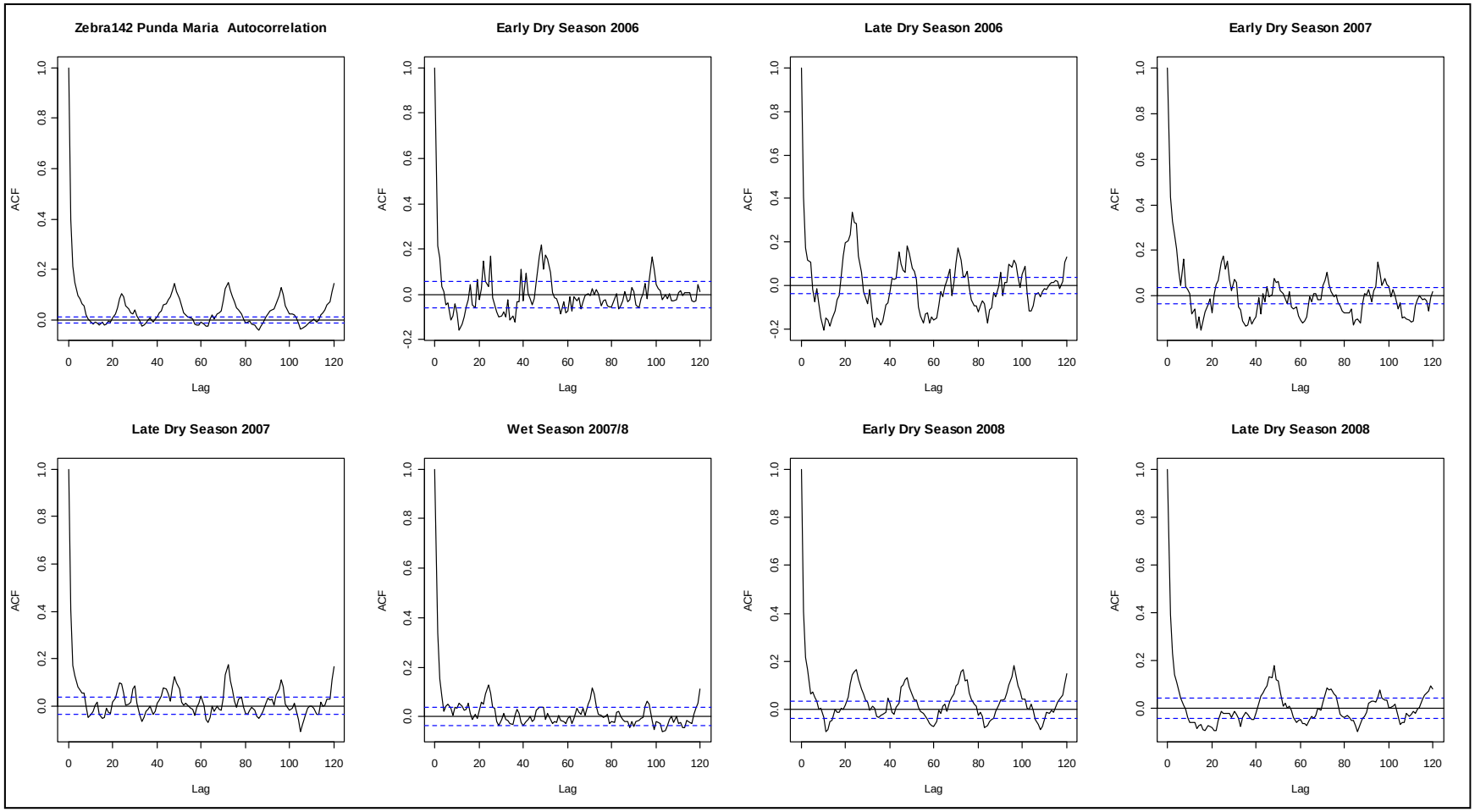

Figure 4.7 – Zebra142 Punda Maria – Autocorrelation

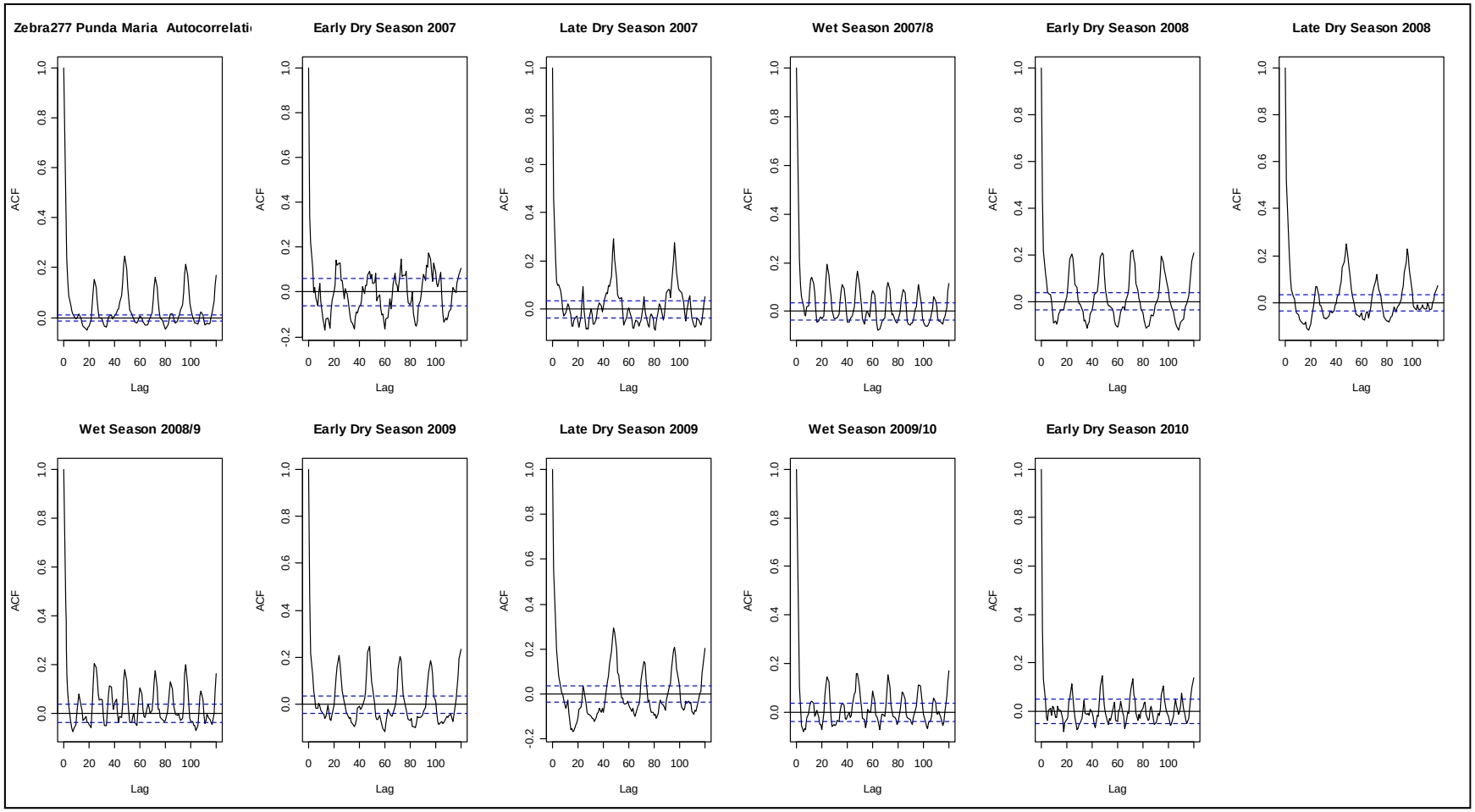

Figure 4.8 – Zebra277 Punda Maria – Autocorrelation

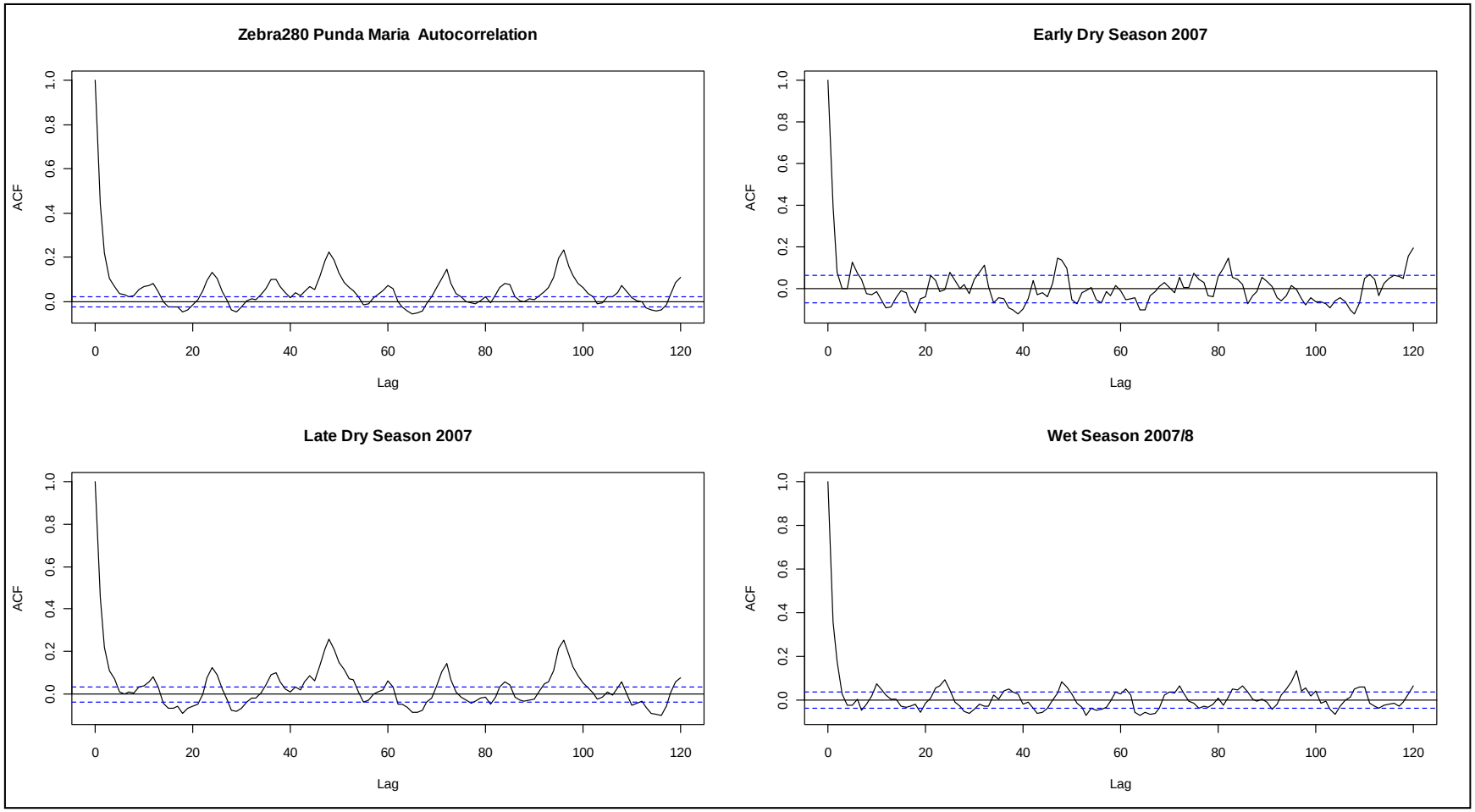

Figure 4.9 – Zebra280 Punda Maria – Autocorrelation

5 HIDDEN MARKOV MODELS FOR ANIMAL MOVEMENT

5.1 Introduction

In Chapter 4, it was shown that there is a strong serial dependence within the movement data for sable antelope, buffalo and zebra herds studied in the Kruger National Park. This serial correlation is present in the hourly data as well as the daily data, although the daily data does not show as strong a cycle as the hourly data. This is not unexpected for movement data with short inter-observation intervals, since the way in which the animal is moving now is likely to be related to the way it was moving in the previous time period, depending on the characteristic periodicity of the states. The correlation can be considered on two levels, the activity correlation between hourly intervals or a spatial correlation from day to day between the regions used by the animals. In the analysis of the behavioural states of the animal using independent mixture models in Chapter 3, this dependence between successive observations is ignored. In order to statistically account for the autocorrelation, time series techniques are required.

The underlying assumption of both the mixture models and Hidden Markov models is that the observed data comes from a population with unobserved groupings. Each observation is associated with one of these groups. For the time series version, the dependence within the model is achieved by assuming that the groups are associated with one another via a Markov process. In fact, the marginal distribution of a hidden Markov model is a mixture distribution. As with the mixture models, the objective of the analysis, is to infer the behavioural state of the animal based on the observed GPS tracking data at hourly intervals, and also to investigate the characteristics of the daily displacements. Behavioural states inferred from the independent mixture models are expected to be similar to those

determined from the dependent version of the analysis, although the probability of transitioning from one state to another through time is now formally incorporated into the model. In this way, the mechanism for changing behaviour through time is accounted for through the state transition matrix, and the time series nature of the data is acknowledged, instead of ignoring the known serial dependence as was done for the independent mixture models.

Hidden Markov models are becoming popular in the analysis of animal movement, as they provide a reasonably simple theoretical structure. The inferences that can be drawn from these models are easily adapted to interpret the results in terms of the behaviour of an animal. Hidden Markov models can be used to provide simple and meaningful summaries of the data. The states and transition probabilities can be linked to what is going on in reality through making inferences about the biological interpretation of the unknown states and can allow for comparison between groups or individuals. "The states cannot be observed, and are designed to account for overdispersion and serial correlation in the observed series" (Langrock et al. 2012).

5.2 Hidden Markov Models

Beichelt and Fatti (2002) define a Markov chain as follows. Let $\{X_0, X_1, \ldots\}$ be a sequence of random variables and $\mathbf{Z} = \{0, \pm1, \pm2, \ldots\}$ be the union of the sets of their realisations. Then $\{X_0, X_1, \ldots\}$ is called a discrete-time Markov chain with state space $\mathbf{Z}$ if

$$P(X_{t+1} = i_{t+1} | X_t = i_t, \ldots, X_1 = i_1, X_0 = i_0) = P(X_{t+1} = i_{t+1} | X_t = i_t)$$

holds for any $t = 1, 2, \ldots$ and any $i_0, i_1, \ldots, i_{t+1}$ with $i_k \in \mathbf{Z}$ (Beichelt, Fatti 2002).

This implies that the future state of a discrete Markov Chain depends only on its present state, and not the sequence of previous states, hence the transition probability from one state to another depends solely on the current state. MacDonald and Zucchini (1997) define a *hidden Markov model* (HMM) as an underlying and unobserved sequence of states following a Markov chain with finite state space. In other words, a hidden Markov model is a type of *dependent* mixture model (Zucchini, MacDonald 2009, Leroux, Puterman 1992) which takes into account the time dimension of the observations. In fact, many of the properties of mixture models are used to prove the properties of hidden Markov models (Leroux, Puterman 1992). The hidden Markov model itself is not a Markov process, except under certain conditions (Zucchini, MacDonald 2009). For some hidden Markov models, it can be proved that the probability of being in a state does depend on the historical sequence of states and not just the previous state, which violates the Markov assumption. As one example Zucchini and MacDonald (2009) show that a two-state Bernoulli HMM is not a Markov process. Spreij (2001) investigated the conditions under which a hidden Markov model is a Markov process itself, where they proved that a hidden Markov chain is itself Markov if and only if certain conditions hold for all the initial distributions of the transition probability matrix (Spreij 2001).

The probability distribution of an observation at any time is determined only by the current state of that Markov chain. For the rest of this chapter, the HMM notation of Zucchini and MacDonald (2009) will be used. A hidden Markov process $\{X_t\}$ comprises two probabilistic mechanisms, firstly an unobserved Markov chain $\{C_t\}$ on m states which the authors call the 'parameter process'. The 'state dependent process' $\{X_t: t = 1,2,\ldots\}$ is defined such that the distribution of X_t depends only on the current state C_t and not on the previous sequence of observations or states. This process depends only on the current observation and not the previous states or observations which satisfies the Markov process. This model can be defined simply as (Zucchini, MacDonald 2009):

$$P(C_t|\mathbf{C}^{(t-1)}) = P(C_t|C_{t-1}),\ t = 2,3,\ldots$$

$$P(X_t|\mathbf{X}^{(t-1)},\mathbf{C}^{(t)}) = P(X_t|C_t),\ t = 1, 2, 3,\ldots$$

where $\mathbf{X}^{(t)}$ and $\mathbf{C}^{(t)}$ represent the history from $t = 1,\ldots,T$.

Hence such a hidden Markov model can be characterised by the distribution of $\{C_t\}$, the transition probability matrix of the Markov chain ($\mathbf{\Gamma}$), and the state dependent distributions defined by $p_i(x) = P(X_t = x \mid C_t = i)$. This equation applies to discrete and continuous distributions of any parametric family (Ephraim, Merhav 2002) and it could comprise distributions of the same or different families. $p_i(x)$ is 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.

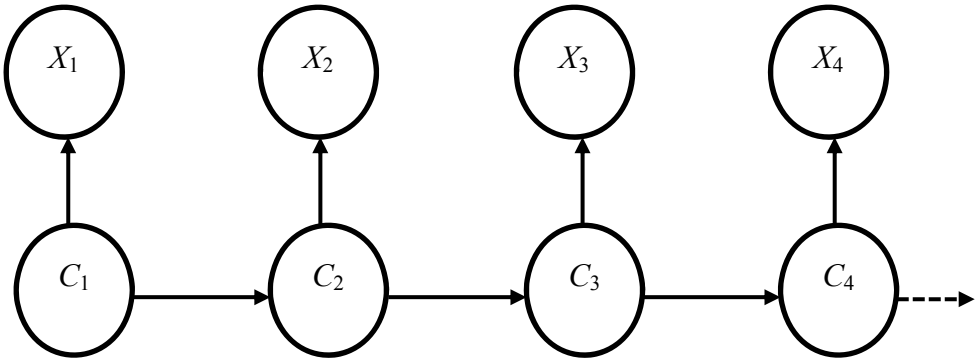

Figure 5.1 – Graphical model of a basic HMM (Zucchini, MacDonald 2009)

The structure of an HMM is shown in Figure 5.1 with only one possible route from each state to the next and the observations only depending on the current state and each state has its own distribution (Zucchini, MacDonald 2009). The number of states of the HMM is known as the *order* of the model (Ephraim, Merhav 2002). A Markov chain is said to have a transition probability matrix $\mathbf{\Gamma}$, where $\mathbf{\Gamma}$ is a square matrix of probabilities with each row summing to one:

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

$\boldsymbol{\Gamma}$ has a stationary distribution $\boldsymbol{\delta}$ if and only if $\boldsymbol{\delta}\boldsymbol{\Gamma} = \boldsymbol{\delta}$ and $\boldsymbol{\delta}\mathbf{1}' = 1$ (Zucchini, MacDonald 2009) where $\mathbf{1}'$ is a row vector of ones. The first equation provides the requirement for stationarity of the Markov chain and the second ensures that $\boldsymbol{\delta}$ is a probability distribution. If the Markov chain is irreducible, then the chain has a unique stationary distribution (Leroux, Puterman 1992). However if the chain is not irreducible, then an infinite number of distinct stationary distributions exist (Leroux, Puterman 1992). For a stationary Markov chain, the state probabilities at $t = 1$ are the same as at $t = 0$, and are still the same after T steps (Beichelt, Fatti 2002). Hence for a stationary Markov chain, the state probabilities are time independent. A hidden Markov model is stationary and ergodic if the Markov chain is stationary, irreducible and aperiodic. For definitions of irreducible, ergodic and aperiodic, see for example Beichelt & Fatti (2002) page 146, 294 and 150 respectively. A Markov chain is said to be homogenous if the transition probabilities $P(C_{s+t} = j \mid C_s = i)$ do not depend on s (Zucchini, MacDonald 2009).

Changes in the dependence structure which allows for feedback can be made if the HMM is extended (Zucchini, Raubenheimer & MacDonald 2008). The feedback mechanism allows the next state at time $t + 1$ to be influenced by an external state as well as the previous state of the Markov chain via the transition probabilities. The graphical model of the feedback loop used in Zucchini et al. (2008) is shown in Figure 5.2 below.

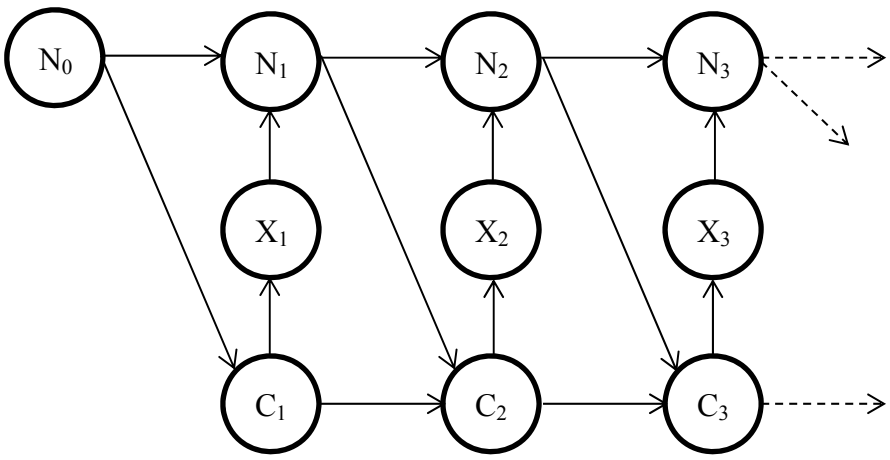

Figure 5.2 - Graphical model of an HMM with feedback loop (Zucchini, Raubenheimer & MacDonald 2008)

The unobserved state $\{C_t\}$ is driven by the process $\{N_t\}$ which in their example corresponds to the nutrient level of an animal. The process N_t is influenced by $N_{t\text{-}1}$ and X_t as can be seen in the graphical model. This model specification implies that the transition probabilities are given by:

$$\gamma_{ij}(n_t) = P(C_{t+1} = j \mid C_t = i, N_t = n_t), \qquad t = 1, 2, ..., T-1$$

The feedback process influences the unknown state and is a function of the observations. This function can be dependent on unknown parameters in which case the process is unobservable. In the application of HMMs to animal movement, the feedback process could correspond to the habitat type that the animal was moving through, the proximity to other species or other ecologically relevant features. The flexibility of HMMs to be extended in this way is one of the advantages of using them as a modelling technique for complex systems.

For a stationary hidden Markov model, the assumption is that the transition probabilities remain constant over time. This assumption would not be valid if there was a trend in the environmental conditions or features influencing the behaviour (MacDonald, Raubenheimer 1995). The stationary distribution is an estimate of the amount of time on average that the process remains within each state (Langrock et al. 2012). If there is any time trend within the observation data, the assumption of stationarity is invalid and a non-stationary Markov chain should be used.

5.3 Moment Derivation

For an m-state stationary HMM $\{X_t: t = 1,2,\ldots,\}$ with transition probability matrix given by $\mathbf{\Gamma}$, stationary distribution $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_m)$, $u_i(t) = \mathrm{P}(C_t = i)$ and state dependent distribution $p_i(x)$ defined as above, then assume that the mean and

variance of the distribution p_i is given by μ_i and σ_i^2 respectively. The expected value and variance of the HMM are given by:

$$E(X_t) = \sum_{i=1}^{m} E(X_t \mid C_t = i) \mid P(C_t = i)$$

$$= \sum_{i=1}^{m} u_t(t) E(X_t \mid C_t = i)$$

Since the Markov chain is stationary

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

$$= \sum_{i=1}^{m} \delta_i \mu_i$$

$$= \boldsymbol{\delta\mu}'$$

where $\boldsymbol{\delta\mu}'$ denotes the vector product. Let Y_i denote the random variable with probability function $p_i(x)$. This implies $E(Y_i)=\mu_i$ and $Var(Y_i)=\sigma_i^2$.

$$Var(Y_i) = E\left(\left(Y_i - \mu_i\right)^2\right) = \sigma_i^2$$

$$\sigma_i^2 = E(Y_i^2) - \mu_i^2$$

$$E(Y_i^2) = \sigma_i^2 + \mu_i^2$$

$$E(X_t^k) = \sum_{i=1}^{m} \delta_i E(Y_i^k)$$

$$E(X_t^2) = \sum_{i=1}^{m} \delta_i E(Y_i^2)$$

$$= \sum_{i=1}^{m} \delta_i (\sigma_i^2 + \mu_i^2)$$

$$Var(X_t) = \sum_{i=1}^{m} \delta_i (\sigma_i^2 + \mu_i^2) - (\sum_{i=1}^{m} \delta_i \mu_i)^2$$

$$= \sum_{i=1}^{m} \delta_i (\sigma_i^2 + \mu_i^2) - (\boldsymbol{\delta\mu}')^2$$

For example, if $m = 2$, then the variance of the 2-state stationary hidden Markov model is given by:

$$Var(X_t) = \sum_{i=1}^{m} \delta_i(\sigma_i^2 + \mu_i^2) - (\sum_{i=1}^{m} \delta_i \mu_i)^2$$

$$= \delta_1(\sigma_1^2 + \mu_1^2) + \delta_2(\sigma_2^2 + \mu_2^2) - (\delta_1\mu_1 + \delta_2\mu_2)^2$$

$$= \delta_1\sigma_1^2 + \delta_2\sigma_2^2 + \delta_1\mu_1^2 + \delta_2\mu_2^2 - \delta_1^2\mu_1^2 - 2\delta_1\delta_2\mu_1\mu_2 - \delta_2^2\mu_2^2$$

$$= \delta_1\sigma_1^2 + \delta_2\sigma_2^2 + \delta_1\mu_1^2(1-\delta_1) + \delta_2\mu_2^2(1-\delta_2) - 2\delta_1\delta_2\mu_1\mu_2$$

$$= \delta_1\sigma_1^2 + \delta_2\sigma_2^2 + \delta_1\delta_2(\mu_1^2 - 2\mu_1\mu_2 + \mu_2^2)$$

$$= \delta_1\sigma_1^2 + \delta_2\sigma_2^2 + \delta_1\delta_2(\mu_1 - \mu_2)^2$$

Let $\mathbf{M} = \text{diag}(\mu_1, \ldots, \mu_m)$ and $\gamma_{ij}(k) = (\mathbf{\Gamma}^k)_{ij}$ for $k \in \mathbb{N}$, then

$$E(X_t X_{t+k}) = \sum_{i,j=1}^{m} E((X_t X_{t+k}) \mid C_t = i, C_{t+k} = j)\delta_i\gamma_{ij}(k)$$

$$= \sum_{i,j=1}^{m} E(X_t \mid C_t = i)E(X_{t+k} \mid C_{t+k} = j)\delta_i\gamma_{ij}(k)$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{m} \delta_i E(X_t \mid C_t = i)\gamma_{ij}(k)E(X_{t+k} \mid C_{t+k} = j)$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{m} \delta_i\mu_i\gamma_{ij}(k)\mu_j$$

$$= \boldsymbol{\delta}\mathbf{M}\mathbf{\Gamma}^{\mathbf{k}}\boldsymbol{\mu}'$$

Using this result, the autocorrelation function (ACF) for a hidden Markov model can be computed as follows:

$$\rho_k = Corr(X_t, X_{t+k})$$

$$= \frac{Cov(X_t, X_{t+k})}{\sigma_t \sigma_{t+k}}$$

$$= \frac{E(X_t X_{t+k}) - E(X_t)E(X_{t+k})}{\sigma_t \sigma_{t+k}}$$

Since the Markov chain is stationary

$$= \frac{\boldsymbol{\delta}\mathbf{M}\boldsymbol{\Gamma}^k\boldsymbol{\mu} - (\boldsymbol{\delta}\boldsymbol{\mu}')^2}{Var(X_t)}$$

The states of a hidden Markov model can be permuted, however there is often a natural ordering of parameters, for example $(\mu_1) < (\mu_2) < \ldots < (\mu_m)$ (Leroux, Puterman 1992).

5.4 Parameter Estimation

The likelihood function can be maximised in order to estimate the parameter estimates for the *m*-state HMM. Similar to the mixture models, the maximum likelihood parameters are consistent and asymptotically normal (Ephraim, Merhav 2002). The likelihood, L_T, is defined as the likelihood of obtaining the observations $x_1, x_2, x_3, \ldots, x_T$. Given an *m*-state HMM, with initial distribution $\boldsymbol{\delta}$, transition probability matrix $\boldsymbol{\Gamma}$ and state-dependent probability functions $p_i(x)$, then using the notation of Zucchini & MacDonald (2009), the likelihood function for a non-stationary HMM is given by:

$L_T = \boldsymbol{\delta}\mathbf{P}(x_1)\boldsymbol{\Gamma}\mathbf{P}(x_2)\,\boldsymbol{\Gamma}\mathbf{P}(x_3)\ldots\boldsymbol{\Gamma}\mathbf{P}(x_T)\mathbf{1}'$.

where $\mathbf{P}(x) = \text{diag}(p_i(x))$, the diagonal matrix of the state-dependent probability functions, and $\boldsymbol{\delta}$ and $\mathbf{1}'$ are row vectors. If $\boldsymbol{\delta}$ is the *stationary* distribution of the Markov chain, then

$L_T = \boldsymbol{\delta}\boldsymbol{\Gamma}\mathbf{P}(x_1)\boldsymbol{\Gamma}\mathbf{P}(x_2)\,\boldsymbol{\Gamma}\mathbf{P}(x_3)\ldots\boldsymbol{\Gamma}\mathbf{P}(x_T)\mathbf{1}'$.

The likelihood function for the stationary HMM involves an extra term $\boldsymbol{\Gamma}$, but may actually be simpler to compute (Zucchini, MacDonald 2009). These equations can be transformed using what is referred to as the *forward probabilities* ($\boldsymbol{\alpha}_t$) such that:

$\boldsymbol{\alpha}_0 = \boldsymbol{\delta}$

$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1}\boldsymbol{\Gamma}\mathbf{P}(x_t)$ for t = 1, 2,…T;

$L_T = \boldsymbol{\alpha}_T\mathbf{1}'$.

The elements of $\boldsymbol{\alpha}_t$ are defined as the *forward probabilities*, with

$$\begin{aligned}\boldsymbol{\alpha}_t &= \boldsymbol{\delta}\mathbf{P}(x_1)\boldsymbol{\Gamma}\mathbf{P}(x_2)\boldsymbol{\Gamma}\mathbf{P}(x_3)...\boldsymbol{\Gamma}\mathbf{P}(x_t)\\ &= \boldsymbol{\delta}\mathbf{P}(x_1)\prod_{s=2}^{t}\boldsymbol{\Gamma}\mathbf{P}(x_s)\end{aligned}$$

while the *backward probabilities* $\boldsymbol{\beta}_t$ for t=1,2,…,T are defined by:

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

$\beta_t(i)$ is defined as the conditional probability $P(X_{t+1} = x_{t+1}, X_{t+2} = x_{t+2}, ..., X_T = x_T \mid C_t = i)$. The backward probabilities are used within the Estimation Maximization (EM) algorithm discussed in Section 5.4.3. From the above equations: $\alpha_t(j)\beta_t(j) = P(\boldsymbol{X}^{(T)} = \boldsymbol{x}^{(T)}, C_t = j)$.

In order to estimate the parameters in each state, the transition probabilities and the state allocation, the likelihood L_T above can be maximised either using direct numerical maximization or the EM algorithm. Software programs such as R (R Development Core Team 2011), now have built-in optimization functions such as `nlm`, `optim` and `constrOptim` which enable parameters to be estimated directly. Order Tm^2 calculations are required for the computation of the likelihood

function using the forward probabilities. This is in comparison with the order Tm^T calculations required for the direct calculation of the likelihood in the form $L_T = \sum_{c_1,\ldots,c_T=1} \prod_{k=1}^{T} \gamma_{c_{k-1}c_k} P(X_k \mid C_k)$ (Ephraim, Merhav 2002, Zucchini, MacDonald 2009). Using the forward recursion makes the likelihood function far simpler to compute and means that it can be maximised directly (Zucchini, MacDonald 2009).

Various R packages exist which can be used to estimate hidden Markov models, including `msm` (Jackson 2011), `HiddenMarkov` (Harte 2011) and `HMM` (Scientific Software Development - Dr. Lin Himmelmann and www.linhi.com 2010). However using a package to run the model can be restrictive if the package is not completely flexible. For example, the package `HiddenMarkov` (Harte 2011) does not run if there are missing observational data, which is common with telemetry data. Also extensions to the model, for example to include the seasonality, can be tricky when using the R packages.

5.4.1 Direct Numerical Maximization

With the power of modern computing and built-in optimizer functions within statistical software programs, it is possible to evaluate the likelihood function and maximise it relative to the model parameters. Issues that complicate this process include numerical underflow (the error due to the absolute value of the computed quantity being smaller than the precision of the computing device), parameter constraints and multiple local maxima in the likelihood function (Zucchini, MacDonald 2009). The problem of numerical underflow can be resolved in most cases by computing the logarithm of the likelihood function using a scaling of the vector of forward probabilities and it can easily be extended for a non-stationary Markov chain (Zucchini, MacDonald 2009). As with the maximum likelihood estimation of the mixture models discussed in Chapter 3, the distributional model parameters need to be constrained to valid values if an unconstrained optimizer

function is used to maximise the likelihood. As with the mixing parameters, the elements of each row of $\mathbf{\Gamma}$ and $\boldsymbol{\delta}$ have to sum to one, and some distributional parameters (such as the σ value as defined for the log-normal distribution) are only valid for positive values. If an unconstrained optimizer such as `nlm` is used, then the optimization needs to be run for transformed functions of the original values which are unconstrained, and then untransformed back to obtain the constrained parameter estimates. This is the same process that was described for mixture models in Section 3.7.1. In order to avoid finding local maxima rather than the global maximum, it is recommended to use a range of starting values to confirm that the global maximum is found for each optimization (Zucchini, MacDonald 2009). Similar to the independent mixture models, it is also possible for the hidden Markov models to have an unbounded likelihood for certain parameter combinations. This is still only a problem for continuous distributions since for discrete distributions, the likelihood is a probability and ranges between zero and one. An advantage of the direct maximization over the EM algorithm, which is described later in this section, is that specification of the derivatives is not needed. Also the EM algorithm is reported to be much slower than direct maximization (Zucchini, MacDonald 2009).

5.4.2 Standard Errors

The process defined in the mixture model chapter to determine the bootstrap standard errors can be extended to determine the standard errors for the parameter estimates of a hidden Markov model. This was described on page 57 of Section 3.3.2. Standard errors can be found for each of the model parameters, including the state-dependent parameters and the transition probabilities. If the model is stationary, standard errors can also be found for the stationary distribution parameters obtained from the transition probability matrices. The width of the confidence intervals around the bootstrap mean provides an indication of the stability of the model in terms of consistent parameter estimation. Also if the intervals around each parameter estimate do not overlap with the other parameter's intervals, it provides some evidence that the model is not over fitted

and that the states are actually different from one another. The standard errors can also be estimated using the Hessian of the log-likelihood at the maximum, however, this can be problematic if some of the parameters are close to the boundary of their parameter space (Zucchini, MacDonald 2009). The bootstrap process is relatively easy to code, although the calculations are computationally intensive and time consuming (Zucchini, MacDonald 2009) when a very large number of simulations and multi-state models are used. Indeed, the bootstrap parameter estimates obtained for this study were found to be computationally time consuming for the independent mixture models, and are even more so for the HMMs with the same number of states, due to the additional parameter estimation that is required.

5.4.3 EM Algorithm

The EM algorithm allows for estimation of the unknown model parameters using maximum likelihood estimates derived from the forward-backward algorithm (MacDonald, Zucchini 1997). In the context of HMMs, the EM algorithm is often called the Baum Welch algorithm and it estimates the model parameters for a hidden Markov model where the Markov chain is homogenous but not necessarily stationary. A modification to the algorithm is needed if the assumption of stationarity is made (Zucchini, MacDonald 2009). The EM algorithm is an iterative algorithm used for maximum likelihood estimation (Zucchini, MacDonald 2009). It is a two-step process which involves computing the conditional expectations of the unknown states given the observations, and the current estimates of the parameters (known as the E-step) and then a maximising step with respect to the model parameters (known as the M-step). These two steps are repeated until a convergence criterion has been reached (Ephraim, Merhav 2002, Zucchini, MacDonald 2009).

The forward probabilities are defined by a joint probability as follows:

For $t = 1, 2, ..., T$ and $j = 1, 2, ..., m$

$\alpha_t(j) = P(\mathbf{X}^{(t)} = \mathbf{x}^{(t)}, C_t = j)$

Proof: From the above definition of a hidden Markov model, $\boldsymbol{\alpha}_t$ is given by:

$\boldsymbol{\alpha}_t = \boldsymbol{\delta}\mathbf{P}(x_1)\boldsymbol{\Gamma}\mathbf{P}(x_2)...\boldsymbol{\Gamma}\mathbf{P}(x_t)$

$= \boldsymbol{\delta}\mathbf{P}(x_1)\prod_{s=2}^{t}\boldsymbol{\Gamma}\mathbf{P}(x_s)$

Therefore for t = 1,2,...,T-1

$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t\boldsymbol{\Gamma}\mathbf{P}(x_{t+1})$

$= \left(\sum_{i=1}^{m}\alpha_t(i)\gamma_{ij}\right)p_j(x_{t+1})$

Since $\boldsymbol{\alpha}_1 = \boldsymbol{\delta}\mathbf{P}(x_1)$

$\boldsymbol{\alpha}_1(j) = \delta_j p_j(x_1)$

$= P(C_1 = j)P(X_1 = x_1 \mid C_1 = j)$

$= P(X_1 = x_1, C_1 = j)$

This proves the statement for the case where $t = 1$.

If the statement holds for some $t \in \mathbb{N}$, then it also holds for $t + 1$. Before this can be proved, a further result is needed.

For the graphical model shown in Figure 5.1, the joint distribution of the set of random variables V_i in a directed graphical model is given by:

$$P(V_1, V_2, ..., V_n) = \prod_{i=1}^{n} P(V_i \mid \text{pa}(V_i))$$

where pa(V_i) denotes all the parents of V_i in the set $V_1, V_2, ..., V_n$

Therefore, the joint distribution of $\mathbf{X}^{(t)}$ and $\mathbf{C}^{(t)}$ is given by:

$$\text{P}(\mathbf{X}^{(t)}, \mathbf{C}^{(t)}) = \text{P}(\text{C}_1) \prod_{k=2}^{t} P(C_k \mid C_{k-1}) \prod_{k=1}^{t} P(X_k \mid C_k) \quad \text{.....(a)}$$

$$\therefore \text{P}(\mathbf{X}^{(t+1)}, \mathbf{C}^{(t+1)}) = \text{P}(\text{C}_{t+1}|\text{C}_t)\text{P}(\text{X}_{t+1}|\text{C}_{t+1})\text{P}(\mathbf{X}^{(t)}, \mathbf{C}^{(t)}) \quad \text{..... (A)}$$

Summing over $\mathbf{C}^{(t-1)}$ gives

$$\text{P}(\mathbf{X}^{(t+1)}, C_t, C_{t+1}) = \text{P}(\mathbf{X}^{(t)}, C_t)\text{P}(\text{C}_{t+1}|\text{C}_t)\text{P}(\text{X}_{t+1}|\text{C}_{t+1}) \quad \text{..... (2)}$$

Therefore, returning to the proof of the forward probabilities

$$\alpha_{t+1}(j) = \sum_{i=1}^{m} \alpha_t(i)\gamma_{ij} p_j(x_{t+1})$$

$$= \sum_{i=1}^{m} \underbrace{P(\mathbf{X}^{(t)} = \mathbf{x}^{(t)}, C_t = i)}_{\alpha_t} \underbrace{P(C_{t+1} = j \mid C_t = i)}_{\gamma_{ij}} \underbrace{P(X_{t+1} = x_{t+1} \mid C_{t+1} = j)}_{p_j(x_{t+1})}$$

From equation (2) above this leads to

$$= \sum_{i=1}^{m} P(\mathbf{X}^{(t+1)} = \mathbf{x}^{(t+1)}, C_t = i, C_{t+1} = j)$$

$$= P(\mathbf{X}^{(t+1)} = \mathbf{x}^{(t+1)}, C_{t+1} = j)$$

$\therefore$ Since it holds for $t+1$, by induction it also holds for t.

$\therefore$ For $t = 1, 2, ..., T$ and $j = 1, 2, ..., m$

$$\alpha_t(j) = P(\mathbf{X}^{(t)} = \mathbf{x}^{(t)}, C_t = j)$$

This shows that by definition the forward probability is the joint probability

$\mathrm{P}(X_1 = x_1,\ X_2 = x_2, \ldots, X_t = x_t, C_t = j)$.

In order to show that the backward probabilities are indeed probabilities, the following results (3) and (4) are needed.

For $t = 0, 1, \ldots, T\text{-}1$

$$P(\mathbf{X}_{t+1}^{T}, \mathbf{C}_{t+1}^{T}) = P(X_{t+1} \mid C_{t+1}) \left(P(C_{t+1}) \prod_{k=t+2}^{T} P(C_k \mid C_{k-1}) \prod_{k=t+2}^{T} P(X_k \mid C_k) \right)$$
$$= P(X_{t+1} \mid C_{t+1}) P(\mathbf{X}_{t+2}^{T}, \mathbf{C}_{t+2}^{T})$$

Summing over C_{t+2}^{T} and diving by $P(C_{t+1})$ provides the following:

$$P(\mathbf{X}_{t+1}^{T} \mid C_{t+1}) = P(X_{t+1} \mid C_{t+1}) P(\mathbf{X}_{t+2}^{T} \mid C_{t+1}) \quad \text{.....(3)}$$

$$P(\mathrm{X}_{t+1}^{T} \mid C_t, C_{t+1}) = \frac{1}{P(C_t, C_{t+1})} \sum_{C_{t+2}^{T}} P(\mathrm{X}_{t+1}^{T}, \mathrm{C}_t^{T})$$

which from the joint distribution above reduces to

$$\sum_{C_{t+2}^{T}} \prod_{k=t+2}^{T} P(C_k \mid C_{k-1}) \prod_{k=t+1}^{T} P(X_k \mid C_k)$$

And $P(\mathrm{X}_{t+1}^{T} \mid C_{t+1}) = \frac{1}{P(C_{t+1})} \sum_{C_{t+2}^{T}} P(\mathrm{X}_{t+1}^{T}, \mathrm{C}_{t+1}^{T})$

$$= \frac{1}{P(C_{t+1})} \sum_{C_{t+2}^{T}} P(C_{t+1}) \prod_{k+2}^{T} P(C_k \mid C_{k-1}) \prod_{k+1}^{T} P(X_k \mid C_k)$$

$$= \sum_{C_{t+2}^{T}} \prod_{k+2}^{T} P(C_k \mid C_{k-1}) \prod_{k+1}^{T} P(X_k \mid C_k)$$

Which is the same expression as above. Therefore for $t = 1, 2, ..., T$ -1

$$P(X_{t+1}^{T} \mid C_{t+1}) = P(X_{t=1}^{T} \mid C_t, C_{t+1}) \quad \text{.....(4)}$$

Backward probabilities

From the definition of $\boldsymbol{\beta}'_t$ as given above:

$$\boldsymbol{\beta}_t^{'} = \boldsymbol{\Gamma} \mathrm{P}(x_{t+1}) \boldsymbol{\Gamma} \mathrm{P}(x_{t+2}) ... \boldsymbol{\Gamma} \mathrm{P}(x_T) \mathbf{1}'$$

$$= \boldsymbol{\Gamma} \mathrm{P}(x_{t+1}) \boldsymbol{\beta}_{t+1}^{'} \qquad \text{for } t = 1, 2, ..., T-1$$

$\beta(i)$ is a conditional probability of the observations being x_{t+1}, $x_{t+2,\ldots}$,x_T given that the Markov chain is in state i at time t.

For $t = 1, 2, ..., T-1$ and $i = 1, 2, ..., m$

$$\beta_t(i) = P(X_{t+1} = x_{t+1}, X_{t+2} = x_{t+2}, ..., X_T = x_T \mid C_t = i)$$

provided that $P(C_t = i) > 0$. This can be written as :

$$\beta_t(i) = P(\mathbf{X}_{t+1}^T = \mathbf{x}_{t+1}^T \mid C_t = i) \qquad \text{.....(B)}$$

where X_a^b denotes the vector $(X_a, X_{a+1}, ..., X_b)$

Proof: Using the results (3) and (4) above, the proof of the backward probabilities follows by induction. In order to establish the above result (B) for $t = T - 1$, note that:

$$\beta'_{T-1} = \Gamma \mathbf{P}(x_T)\mathbf{1}'$$

$$\beta_{T-1}(i) = \sum_j P(C_T = j \mid C_{T-1} = i) P(X_T = x_T \mid C_t = j) \qquad \text{.....(5)}$$

From (4)

$$P(C_T \mid C_{T-1}) P(X_T \mid C_T) = P(C_T \mid C_{T-1}) P(X_T \mid C_{T-1}, C_T)$$

$$= \frac{P(C_T \mid C_{T-1})}{P(C_{T-1})} \frac{P(X_T, C_{T-1}, C_T)}{P(C_T \mid C_{T-1})}$$

$$= \frac{P(X_T, C_{T-1}, C_T)}{P(C_{T-1})} \qquad \text{.....(6)}$$

Substituting (6) into (5)

$$\beta_{T-1}(i) = \frac{\sum_j P(X_T = x_T, C_{T-1} = i, C_T = j)}{P(C_{T-1} = i)}$$

$$= \frac{P(X_T = x_T, C_{T-1} = i)}{P(C_{T-1} = i)}$$

$$= P(X_T = x_T \mid C_{T-1} = i)$$

which proves (B) for $t = T$.

In order to prove that $t+1$ implies validity for t, note that the recursion for β_t, and the induction hypothesis shows that :

$$\beta_t(i) = \sum_j \gamma_{ij} P(X_{t+1} = x_{t+1} \mid C_{t+1} = j) P(\mathbf{X}_{t+2}^T = \mathbf{x}_{t+2}^T \mid C_{t+1} = j) \quad \text{.....(7)}$$

But from (3) and (4)

$$P(X_{t+1} \mid C_{t+1}) P(\mathbf{X}_{t+2}^T \mid C_{t+1}) = P(\mathbf{X}_{t+1}^T \mid C_{t+1})$$

$$= P(\mathbf{X}_{t+1}^T \mid C_{t+1}, C_t)$$

Substituted into (7)

$$\beta_t(i) = \sum_j P(C_{t+1} = j \mid C_t = i) P(\mathbf{X}_{t+1}^T \mid C_t = i, C_{t+1} = j)$$

$$= \frac{\sum_j P(\mathbf{X}_{t+1}^T = \mathbf{x}_{t+1}^T, C_t = i, C_{t+1} = j)}{P(C_t = i)}$$

$$= \frac{P(\mathbf{X}_{t+1}^T = \mathbf{x}_{t+1}^T, C_t = i)}{P(C_t = i)}$$

$$\therefore\ \beta_t(i) = P(\mathbf{X}_{t+1}^{T} = \mathbf{x}_{t+1}^{T} \mid C_t = i)$$

which is the required conditional probability.

If there are missing observation data at random, the empty product is replaced with an identity matrix (Zucchini, MacDonald 2009). If $t = T$, then $\boldsymbol{\beta}_T = \mathbf{1}$. It is shown by Zucchini & MacDonald (2009) that:

$$\beta_t(i) = P(X_{t+1} = x_{t+1}, X_{t+2} = x_{t+2}, ..., X_T = x_T \mid C_t = i)$$

The above equations reinforce that the forward probabilities are joint probabilities, while the backward probability is the conditional probability of the observations x_{t+1}, x_{t+2},…,x_T given that the Markov chain is in state i at time t (Zucchini, MacDonald 2009). The following result shows how the forward and backward probabilities are used during the EM algorithm, and also for local decoding, which determines the most likely state at time t given all the observations:

For $t = 1, 2, ..., T$:

$$P(C_t = j \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}) = \frac{\alpha_t(j)\beta_t(j)}{L_T}$$

and for $t = 2, ..., T$:

$$P(C_{t-1} = j, C_t = k \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}) = \frac{\alpha_{t-1}(j)\gamma_{jk}p_k(x_t)\beta_t(k)}{L_T}$$

It is said that the EM Algorithm is typically used when "some of the data are missing, and exploits the fact that the complete-data log-likelihood may be straightforward to maximise even if the likelihood of the observed data is not" (Zucchini, MacDonald 2009). In the context of HMMs, the 'missing data' usually refers to the unobserved sequence of states of the Markov chain. However, there appears to be little or no published literature dealing with the EM algorithm when the *observed* data are missing. This will be discussed below in Section 5.6. The 'complete-data log-likelihood' mentioned above refers to the log-likelihood of the

parameters of interest $\boldsymbol{\theta}$, based on the observed and the missing data (Zucchini, MacDonald 2009). The EM algorithm can be described as follows: (Little, Rubin 2002, Zucchini, MacDonald 2009).

E-step: The conditional expectations of the missing data (latent states) given the observations and the current estimate of $\boldsymbol{\theta}$ are calculated. This equates to computing the conditional expectations of the functions of the missing data that appear in the complete-data log-likelihood.

M-step: Maximise the complete-data log-likelihood with respect to $\boldsymbol{\theta}$, with the functions of the missing data replaced by their conditional expectations calculated in the E-step. This is done by updating the estimates of $\boldsymbol{\theta}$.

These steps are repeated until some convergence criterion is met. The solution obtained for $\boldsymbol{\theta}$ is a stationary point of the likelihood function although it can have converged to a local maximum rather than the global maximum.

The EM algorithm can be applied to non-stationary HMMs as follows, using the notation of Zucchini & MacDonald (2009):

The sequence of states of the Markov chain can be represented as a binomial variable such that $u_i(t) = 1$ if and only if $c_t = j$, $(t = 1,2,\ldots,T)$

and $v_{jk}(t) = 1$ if and only if $c_{t-1} = j$ and $c_t = k$, $(t = 2,3,\ldots,T)$.

The complete-data log-likelihood of the HMM, is then the log-likelihood of the observations $x_1, x_2,\ldots, x_T$ and the latent variable data $c_1, c_2,\ldots,c_T$ given by:

$$\log(P(\mathbf{x}^{(T)},\mathbf{c}^{(T)})) = \log(\delta_{c_1}\prod_{t=2}^{T}\gamma_{c_{t-1},c_t}\prod_{t=1}^{T}p_{c_t}(x_t))$$

$$= \log\delta_{c_1} + \sum_{t=2}^{T}\log\gamma_{c_{t-1},c_t} + \sum_{t=1}^{T}\log p_{c_t}(x_t)$$

$$= \sum_{j=1}^{m}u_j(1)\log\delta_j + \sum_{j=1}^{m}\sum_{k=1}^{m}\left(\sum_{t=2}^{T}v_{jk}(t)\right)\log\gamma_{jk} + \sum_{j=1}^{m}\sum_{t=1}^{T}u_j(t)\log p_j(x_t) \qquad \text{..... (8)}$$

During the E-step, the $u_i(t)$ and $v_{jk}(t)$ from (8) above are replaced by their conditional expectations given the observations and the parameter estimates $\boldsymbol{\theta}$ at that iteration of the algorithm.

$$\hat{u}_j(t) = P(C_t = j \mid \mathbf{x}^{(T)}) = \frac{\alpha_t(j)\beta_t(j)}{L_T}$$

and

$$\hat{v}_{jk}(t) = P(C_{t-1} = j, C_t = k \mid \mathbf{x}^{(T)}) = \frac{\alpha_{t-1}(j)\gamma_{jk}p_k(x_t)\beta_t(k)}{L_T}$$

These two equations are used to replace $u_i(t)$ and $v_{jk}(t)$ in (8) above and the complete-data log-likelihood is then maximised with respect to the initial distribution $\boldsymbol{\delta}$, the transition probability matrix $\boldsymbol{\Gamma}$ and the state-dependent distribution parameters. Since each of these values only occurs separately in the three terms of (8) above, this means that three separate maximizations can be performed. The maximization of the third term involving the state-dependent distribution parameters can be complicated. For some distributions such as the Poisson and Normal, a closed-form solution is available, however, for some like the Gamma distribution, direct numerical maximization is needed (Zucchini, MacDonald 2009). It would then probably be easier to follow the route of using direct numerical maximization to estimate the HMM rather than attempting the more time consuming and complicated EM algorithm.

For stationary hidden Markov models, analytic maximization is needed during the 'M-step' of the EM algorithm, which is a disadvantage of this method of parameter estimation for stationary HMMs (Zucchini, MacDonald 2009). If analytic maximization is needed anyway, then it is computationally simpler to estimate the maximum likelihood during direct numerical maximization as described above. However, for non-stationary models, the EM algorithm is reasonably easy to apply.

Parameter estimates obtained using direct numerical maximization and the EM algorithm should provide very similar maximum likelihood parameter estimates, but they will not be identical. It is expected that the stationary model will always have a slightly lower likelihood value since constraining the initial distribution to be the stationary distribution has to decrease the maximum likelihood (Zucchini, MacDonald 2009). Results obtained using the different software packages can also differ slightly, and comparison of a variety of estimation methods can be informative (Zucchini, MacDonald 2009). However, all the estimates using slightly different methods of maximization should provide very similar results. Models run using EM and direct numerical maximization to model the same dataset are seldom compared (Zucchini, MacDonald 2009), but will be compared for the analysis of the ungulate movement data used in this study. Only non-stationary HMMs are fitted using the EM algorithm since direct numerical maximization would be necessary anyway in the M-Step of the algorithm. Due to the ease of applying functions such as `nlm` in R, direct numerical maximization does seem to be an advantageous method and has been reported to produce maximum likelihood estimates quicker than the EM algorithm (Zucchini, MacDonald 2009). The comparison will be discussed further in Section 5.14.2.

5.4.4 Viterbi Algorithm and Local Decoding

Often one of the desired outcomes of applying Hidden Markov models to an observed data series is to determine the underlying groupings within the population. There are two types of decoding which can be used to allocate each observation to a state. Local decoding refers to the most likely state for an observation at the time it was observed, while global decoding refers to the most likely *sequence* of states (Zucchini, MacDonald 2009). The local decoding is essentially the conditional probability of being in state i at time t, given all observations. Global decoding is the conditional probability of the sequence of states given the sequence of observations. In each case the allocated state is the state (or sequence of states) that maximises the conditional probabilities described above.

Local decoding is also known as the maximum *a posteriori* (MAP) symbol decision rule (Leroux, Puterman 1992). For each time period $t \in \{1,...,T\}$, the distribution of state C_t given the observations $\mathbf{x}^{(T)}$ can be found, using the forward and backward probabilities such that:

$$P(C_t = i \mid \mathbf{X}^{(T)} = \mathbf{x}^{(T)}) = \frac{\alpha_t(i)\beta_t(i)}{L_T}$$ (Zucchini, MacDonald 2009)

The conditional distribution of an observation given all other observations within the data series is given by:

$$P(X_t = x_t \mid \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}) = \sum_{i=1}^{m} w_i(t)\, p_i(x)$$

with $\mathbf{X}^{(-t)}$ and $w_i(t)$ defined below.

The notation $\mathbf{X}^{(-t)}$ represents the observations at all times *other* than at time t. The conditional probability is actually a mixture of the m state-dependent probability distributions where $w_i(t) = \frac{d_i(t)}{\sum_{j=1}^{m} d_j(t)}$ and $d_i(t)$ is the product of the i-th entry of the vector $\boldsymbol{\alpha}_{t-1}\boldsymbol{\Gamma}$ and the i-th entry of the vector $\boldsymbol{\beta}_t$.

The Viterbi algorithm is used to find the sequence of states which maximises the probability of the observation sequence given the model (Tucker, Anand 2005). This is the concept of global decoding introduced above. The algorithm determines the sequence of states $c_1, c_2,\ldots,c_T$ which maximises the conditional probability: $P(\boldsymbol{C}^{(T)} = \mathbf{c}^{(T)} \mid \boldsymbol{X}^{(T)} = \mathbf{x}^{(T)})$ where $\boldsymbol{X}^{(T)}$ and $\boldsymbol{C}^{(T)}$ represent the full sequence of observations and states (Zucchini, MacDonald 2009). This is equivalent to maximizing the joint probability $P(\mathbf{C}^{(T)}, \mathbf{X}^{(T)}) = \delta_{c_1} \prod_{t=2}^{T} \gamma_{c_{t-1},c_t} \prod_{t=1}^{T} p_{c_t}(x_t)$

The Viterbi algorithm is an example of a dynamic programing algorithm which makes it possible to determine the most likely sequence of states without

maximising over all possible sequences of states, which would not be feasible except for a very small number of states since m^T calculations would be required (Zucchini, MacDonald 2009, Cappé, Moulines & Rydén 2005).

The Viterbi algorithm is implemented by defining:

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

and for $t = 2,3,...,T$,

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

The probabilities ξ_{tj} satisfy the recursion below for t = 2, 3, …,T and j = 1, 2, ..,m: $\xi_{tj} = \left(\max_i(\xi_{t-1,i}\gamma_{ij})\right) p_j(x_t)$

Proof:

By definition, for $t = 2,3,...,T$

$$\xi_{tj} = \max_{c_1, c_2, \ldots c_{t-1}} P(\mathbf{C}^{(t-1)} = \mathbf{c}^{(t-1)}, C_t = j, \mathbf{X}^{(t)} = \mathbf{x}^{(t)})$$

The objective is to maximise $P(\mathbf{C}^{(t)}, \mathbf{X}^{(t)})$ which is the joint probability of the sequence of states and the sequence of observations. In Equation (A) on page 144, it was shown that:

$$\mathrm{P}(\mathbf{X}^{(t+1)}, \mathbf{C}^{(t+1)}) = \mathrm{P}(\mathrm{C}_{t+1}|\mathrm{C}_t)\mathrm{P}(\mathrm{X}_{t+1}|\mathrm{C}_{t+1})\mathrm{P}(\mathbf{X}^{(t)}, \mathbf{C}^{(t)})$$

If t is replaced with t -1, then

$$\mathrm{P}(\mathbf{X}^{(t)},\mathbf{C}^{(t)}) \;=\; \mathrm{P}(\mathrm{C}_t|\mathrm{C}_{t-1})\mathrm{P}(\mathrm{X}_t|\mathrm{C}_t)\mathrm{P}(\mathbf{X}^{(t-1)},\mathbf{C}^{(t-1)})$$

$$\mathrm{P}(\mathbf{C}^{(t)},\mathbf{X}^{(t)}) = \mathrm{P}(\mathbf{C}^{(t-1)}=\mathbf{c}^{(t-1)},\mathbf{X}^{(t-1)}=\mathbf{x}^{(t-1)})\gamma_{c_{t-1},j}\,p_j(x_t)$$

The maximum is then given by:

$$p_j(x_t)\max_{c_{t-1}}\left(\gamma_{c_{t-1},j}\max_{c_1,c_2,\ldots c_{t-2}}\mathrm{P}(\mathbf{C}^{(t-1)}=\mathbf{c}^{(t-1)},\mathbf{X}^{(t-1)}=\mathbf{x}^{(t-1)})\right)$$

$$\Rightarrow \xi_{tj} = p_j(x_t)\max_{c_{t-1}}\left(\gamma_{c_{t-1},j}\xi_{t-1,c_{t-1}}\right)$$

if c_{t-1} is replaced by i

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

It is then possible to calculate the T x m matrix of values for ξ_{tj}, and the required number of calculations is linear in T. The optimal sequence of states (i_1, i_2,…, i_T) is then obtained recursively such that:

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

and for $t = T$ -1, T - 2,...,1, from

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

It is possible to maximise a sum of the logarithms of the probabilities rather than the product of the probabilities in order to reduce the risk of numerical underflow (Harte 2011). The Viterbi algorithm can be used when the underlying Markov process is stationary or non-stationary meaning that the initial distribution $\boldsymbol{\delta}$ does not have to be assumed to be the stationary distribution (Zucchini, MacDonald 2009).

The results from local and global decoding will usually be similar although not exactly the same (Zucchini, MacDonald 2009). Being able to maximise the conditional probability for the entire sequence of states is an advantage of using the time series analysis in comparison to the independent mixture models. The performance of the Viterbi algorithm can be checked by simulating a data series from an actual HMM, fitting an HMM model to the simulated data series and then using the Viterbi algorithm to decode the simulated observations and comparing the predicted states with the sequence of known (simulated) states (Zucchini, MacDonald 2009). Zucchini & MacDonald (2009) recommend using this simulation as a method for quantifying the accuracy of the Viterbi path, particularly in cases when the state allocation is the desired outcome of the analysis. For this animal movement analysis, the known HMM to simulate from is specified using parameters similar to those estimated for the observed movement data. The accuracy of this simulated state allocation provides an estimate of the accuracy of the state allocation for the fitted model.

5.5 Model Selection and Goodness-of-Fit

Model selection for HMMs is needed to determine how many states are required in the model. When using the method of maximum likelihood, as the number of states in the model increases, the likelihood of the model will also increase and the apparent fit of the model improves. However, as the number of states increases, so does the number of parameters to be estimated. This was also discussed for the mixture models where the number of mixing parameters and state-dependent distribution parameters increased with the increase in the number of states in the model. However for HMMs, even more parameters are estimated for each increase in the number of states due to the additional parameters in the transition probability matrix. For a stationary HMM, the stationary distribution parameters are not estimated as these are determined directly from the transition probability matrix. A further challenge is to determine which state-dependent probability distributions provide the best fit to the data. Criteria such as the

Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) penalize the likelihood improvement due to the additional parameters in the model. By extending the HMM, it is possible to restrict the number of additional parameters by making assumptions about the state-dependent distributions or the transition probability matrix (Zucchini, MacDonald 2009). An example is to use a circular distribution, such as von Mises, to model average wind direction rather than a categorical HMM based on the wind direction categorized into the 16 compass points (Zucchini, MacDonald 2009). Zucchini & MacDonald (2009) describe a method for fitting stochastic volatility models which placed restrictions on the transition probability matrix which meant that the number of parameters to be estimated did not increase dramatically for very large number of states.

As discussed in the mixture model chapter, the Akaike Information criterion is defined as: AIC = $-2\log L + 2p$ where L is the log-likelihood of the fitted model and p is the number of parameters. The Bayesian Information criterion is given by: BIC = $-2\log L + p \log T$ with L and p as defined for the AIC and T is the number of observations. The BIC tends to select the simpler models (fewer parameters) than the AIC. The AIC and BIC values for the HMMs can be used to compare the fitted HMMs with different numbers of states, or models with different families of state-dependent distributions. These criteria can also be used to compare the HMM to equivalent independent mixture models, despite the starting assumptions of these models being different. As with the mixture models, although the likelihood ratio test is not applicable for comparing models with non-nested state-dependent distributions, the AIC and BIC which are penalized likelihoods are accepted as being applicable to non-nested models (Vuong 1989, Clarke 2007). These model selection criteria tend to select the simpler model with fewer states when the sample sizes are small, however datasets obtained using GPS tracking are generally large so this should not influence the choice of model.

It is also possible to compare the autocorrelation functions (ACF) of the fitted HMMs with the ACF of the observed time series (Zucchini, MacDonald 2009). 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. The ACF for an HMM model was defined in Section 5.3 above.

A further measure of the goodness-of-fit of the model is derived from the concept of the residuals from regression models, where the expected observations are compared with the actual observation. A similar concept is used to calculate a measure called a pseudo-residual to investigate the fit of the fitted HMM (Zucchini, MacDonald 2009). The pseudo-residual is meant to perform the role of residuals in regression models. Zucchini & MacDonald (2009) define two types of pseudo-residual, 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.

The uniform pseudo-residual is the probability of obtaining an observation less than or equal to x_t, under the fitted model, where x_t is an observation from a continuous distribution X_t. This is shown as: $u_t = P(X_t \leq x_t)$. This implies that u_t is the observation x_t transformed by its distribution function under the model (Zucchini, MacDonald 2009). If the model is true, then the pseudo-residual given by u_t has a uniform distribution between zero and one. This transformation of the observations allows for comparison of the pseudo-residuals for each observation, even though each observation might have its own distribution, but the pseudo-residuals are identically distributed U(0,1) (Zucchini, MacDonald 2009). A qq-plot of the expected against sample quantiles is useful for assessing whether the uniform residuals do in fact follow a uniform distribution, if not, the model is not valid. However, the uniform pseudo-residuals are not particularly useful for the identification of outliers, since it is difficult to distinguish values lying close to the

extremes at zero and one. However, using the distribution function of the standard Normal distribution, it is possible to create normal pseudo-residuals (Zucchini, MacDonald 2009). The normal pseudo-residual then measures the deviation from the median which provides an easier measure to assess the fit of the model and the presence of extreme values, using familiar techniques associated with the Normal distribution. If the distribution function of the standard Normal is defined as Φ, then $Z = \Phi^{-1}(F(X))$ where X is a random variable with distribution function F. If the fitted model is the true model, then the normal pseudo-residuals are distributed as standard Normal (Zucchini, MacDonald 2009). In this case the model fit can be more accurately assessed using a histogram of the pseudo-residuals, a qq-plot or classical tests for normality. The normal pseudo-residual z_t is defined as: $z_t = \Phi^{-1}(u_t) = \Phi^{-1}(F_{X_t}(x_t))$, where $F_{Xt}(x_t)$ is the cumulative distribution function of the U(0,1) distribution. The normal pseudo-residual is more useful in the identification of outliers since its absolute value increases with increasing deviation from the median, and these outliers can be more easily identified using the normal scale. The above description of pseudo-residuals is only valid for continuous distributions. For discrete distributions, the pseudo-residuals are defined as intervals rather than points. The definition for the lower and upper bounds of the uniform (u^-_t, u^+_t) and normal pseudo-residuals (z^-_t, z^+_t) for a discrete distribution are given by:

$$[u_t^-, u_t^+] = [F_{X_t}(x_t^-), F_{X_t}(x_t)]$$

$$[z_t^-, z_t^+] = [\Phi^{-1}(F_{X_t}(x_t^-)), \Phi^{-1}(F_{X_t}(x_t))]$$

where x_t^- is the greatest realization possible that is strictly less than x_t. For the discrete case where the observed count is close to the boundary of the domain, the pseudo-residuals will not provide a good indication of the fit of the model (Harte 2011). In order to use the pseudo-residuals for discrete distributions in a qq-plot to validate the model, then a "mid-pseudo-residual" is calculated. This is defined as:

$$z_t^m = \Phi^{-1}\left(\frac{u_t^- + u_t^+}{2}\right).$$

These mid-pseudo-residuals can then be used to check for normality.

In order to apply the concept of pseudo-residuals to HMMs, an ordinary pseudo-residual is defined as the normal pseudo-residual calculated from the conditional distribution of X_t, given $\boldsymbol{X}^{(-t)}$. For continuous distributions, the ordinary pseudo-residual is defined as:

$$z_t = \Phi^{-1}\left(P\left(X_t \le x_t \mid \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}\right)\right)$$

$$= \Phi^{-1}\left(\frac{P\left(X_t \le x_t, \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}\right)}{P(\mathbf{X}^{(-t)} = \mathbf{x}^{(-t)})}\right)$$

This can be calculated using the forward and backward probabilities (Harte 2011). If the model is correct then $z_t \sim N(0,1)$.

5.6 Missing observation data

Missing observation data is a serious challenge in the analysis of animal movement data. There are two main causes for missing data within the datasets used in this study. This was discussed in detail in section 2.3.1. Ideally the missing data should be non-informative and random. In the case of changes in the frequency of observation which creates chunks of missing data, this is not random. It is possible to test for the randomness of the missing data in a movement dataset using the R package 'AdehabitatLT' (Calenge 2006). However if the missingness of the data is non-informative, one of the *reported* advantages of using HMMs for the analysis of animal tracking data is that it can be extended to deal with these missing data. However, there appears to be little or no discussion in the published literature dealing with the case where missing *observation* data existed as well as the latent states. In most cases the discussion

about missing data refers to the unobserved latent states as being unknown and considered 'missing', even though the latent states are always unknown.

In order to maximise the likelihood using direct numerical maximization, it is possible to modify the calculation of the observed-data likelihood so that the likelihood can be maximised as normal using an optimizer function such as `nlm` in R. When an observation is missing, the state-dependent probability of obtaining an observation x is replaced by '1' – the probability of obtaining *any* observation ($p_j(x) = 1$ if x is missing). This corresponds to replacing the diagonal matrices $\mathbf{P}(x_t)$ with the identity matrix in the likelihood function. The likelihood can then be written as follows, assuming that observation 3, 5 and 6 are missing (Zucchini, MacDonald 2009):

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

This simple modification of the observed-data likelihood and maximization using an available optimizing function is easy to code and implement using direct numerical maximization. However, if the maximization is performed using the EM algorithm, changes to the definition of the complete-data likelihood are needed.

For the EM algorithm, there seems to be little or no published literature which discusses the application of the algorithm in the case of latent variables and genuinely missing data. However the EM algorithm can be applied to maximise the likelihood in the presence of uninformative missing data. The modification of the observed data likelihood is simple. The 'complete-data' is made up of the sequence of observations and the latent variables, while the 'incomplete' data is the sequence of observations only (Ephraim, Merhav 2002). If x^A is the set of existing observations, and the latent variables $\boldsymbol{c}^{(T)}$, the complete-data likelihood is given by:

$$\log(P(\boldsymbol{x}^A, \boldsymbol{c}^{(T)})) = \log\left(\delta_{c1} \prod_{t=2}^{T} \gamma_{c_{t-1}, c_t} \prod_{t \in A} p_{c_t}(x_t) \right)$$

This is the same as the definition given for the complete-data likelihood given above in the EM Algorithm section, except for the observations x_t, $t \in A$ (the existing observations) rather than $t = \{1,2,\ldots T\}$ (the complete time series). The zero-one random variables $u_j(t)$ and $v_{jk}(t)$ are defined as above and they are replaced during the E-step of the algorithm by their conditional expectations:

$$\hat{u}_j(t) = E(u_j(t) \mid \boldsymbol{x}^A) = P(C_t = j \mid \boldsymbol{x}^A)$$

$$\hat{v}_{jk}(t) = E(v_{jk}(t) \mid \boldsymbol{x}^A) = P(C_{t-1} = j, C_t = k \mid \boldsymbol{x}^A)$$

These conditional expectations are computed using a small modification to the way in which the forward and backwards probabilities are calculated. As described for the direct maximization with missing data above, to calculate the forward and backward probabilities, the probability of obtaining *any* observation is set to one using $p_j(x) = 1$. The M-step proceeds as in the case of no missing data.

5.7 Hidden Markov Models used in animal movement research

Hidden Markov models are commonly used in speech recognition, psychology, protein sequencing and hydrology (Tucker, Anand 2005). In terms of animal movement studies, HMMs have been used for the analysis of caribou in Canada (Franke, Caelli & Hudson 2004), wolves in Canada (Franke et al. 2006), blue-finned tuna off the Australian coast (Patterson et al. 2009), north sea cod off the coast of the United Kingdom (Pedersen et al. 2008), bison in Canada (Langrock et al. 2012) and others.

Franke et al. (2004) use multiple-observation HMMs as an individual-based modelling technique in order to explain the use of space, movement and behaviour of caribou (*Rangifer tarandus*) in Canada. They define the "hidden" states as the animal bedding, feeding or relocating. They tracked the animals for 10 days using GPS collars and obtained location coordinates every 15 minutes. From the locations they calculated the distance travelled in 15 minutes and the turning angle between successive observations They then categorized the distances into four classes (stationary, short, medium and long) as well as the turning angle into four classes (ahead, right, back and left). They used these categorical variables as the input for the HMM. However by categorizing the distances between location and the turning angles into these 4-classes (or 6-classes in (Franke et al. 2006), it means that they lose the continuous information that has been derived from the GPS locations, and impose a sequence of 'classes' onto the data which can influence the states of the HMM. In this way, they are imposing a group structure on the observed data that may or may not be present in the data. The results from the HMMs allowed them to investigate the probabilities of remaining in or changing from a state of bedding, feeding or relocating, and the times of day associated with each of the states. They could also examine differences between the behaviour of different individual animals based on the probabilities for each animal of changing state or remaining in their current state. They found that HMMs consistently outperformed auto-regressive time series models in terms of Percent Correct and Absolute Average Distance between predicted and observed values. Their results showed that caribou tended to forage for short periods but would bed and relocate for longer periods. From the state transition probabilities they found that animals were most likely to bed after a relocating bout and forage after bedding. The models revealed a diel cycle with long periods of stationary behaviour during the hottest part of the day and relocations during cooler times. The authors did not include any ecological covariates within their models.

Following on from the caribou study which showed that HMMs can provide insight into the movement behaviour of animals, Franke et al. (2006) used HMMs to predict the locations of wolf kill sites. In addition to the GPS tracking data, aerial relocations of the wolves twice a day identified the actual kill-sites. This meant that the 'hidden' state was now known. The study tracked wolves hourly for 13 days. The aim of their study was to determine whether HMMs could be used to correctly predict the location of kill-sites while gaining insight into the underlying movement states of the wolves during this time period. Franke et al. (2006) used the GPS location coordinates to calculate the distance between locations, the turning angle between successive locations and the travel rate per hour. Once again they categorized the distance between location measures into six classes based on the distance moved. If location observations were missed, they divided the total distance moved by the number of time periods elapsed. The turning angles between successive movements were also categorized into one of four classes based on the direction of the move. The travel rate per hour was turned into a Bernoulli variable with value one if the travel rate was below a median wolf travel rate calculated in a different study, and a value of two if the rate was greater than the median. These distance, turning angle and travel rate categorical variables were then used as the input for the HMMs, similar to the Franke et al. (2004) study. The authors do not discuss whether the model estimates for either the state-dependent distribution or the transition probability matrix are stable relative to the classes selected for the input variable classification. The authors used the measures of 'Percent Correct' and 'Absolute Average Deviation' to determine the relationship between the predictions from the HMM and what was called the training data (observed data excluding the information on the location of the kill sites). They could also compare the predicted states with the aerial locations of kill-sites. The models performed well in terms of accurately predicting kill sites, with 22 of the known 23 kill sites predicted by the model. Franke et al. (2006) found that the distance between locations clearly differs between states whereas the turning angle did not. Since the turning angle does not differ between the states, this implies that it is not assisting to differentiate the observations into the underlying groupings.

Pedersen et al. (2008) studied the movement of Atlantic cod in the North Sea using an Archival tag. The tag has a logger, release section, a float and an antenna. Data such as location, depth and temperature can be recorded and it is possible for the logger to release after a set time period and float to the surface for retrieval. In this study, archival tags recorded the depth of the fish and the authors used a direct Fokker-Planck based methodology using HMMs. The aim of the analysis was to reconstruct the geographic movements of the fish. 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 depth was recorded every 10 minutes and the data spanned the time from when the fish was tagged and released until it was caught by either commercial or recreational fishermen. The tidal cycle was used in the calculations to reconstruct the migration pathway using a likelihood model.

Another application of HMMs in the fisheries field was the analysis of southern Bluefin tuna (*Thunnus maccoyii*) movements by Patterson et al. (2009). The input for their models was the displacement between successive locations, and they included temperature as a covariate in their model. They fitted a 2-state model to the data, using an exponential distribution to describe each state, and assumed that the two states corresponded to 'resident' and 'migratory' behavioural states. Daily data (24 hours between successive locations) was used as the input for their models. They also used the concept of pseudo-residuals discussed in Section 5.5 above, to investigate the fit of the model. The authors conducted simulations to determine the effectiveness of HMMs for the analysis of this type of movement data. They found that the model parameter estimates for datasets of longer than 300 observations were suitably close to the true parameters, particularly if the distributions for each state were quite different. For the application of the models to the southern Bluefin tuna data, they found that the average movement distance within each state was similar for the 3 individuals used in this study. A likelihood ratio test was used to determine that there was no significant difference in the movement between the three individuals.

Hidden Markov Models were used in the analysis of GPS tracks obtained from homing pigeons near Oxford, in the United Kingdom (Roberts et al. 2004). The authors used spectral entropy as the input for the HMMs. This spectral entropy was a measure calculated using the trajectories of the bird's path and the number of tracks which passed through a grid cell of a map. The authors used a variational Bayes learning for their models, which they believe avoided the issue of over-fitting associated with maximum likelihood methods. This is due to the likelihood always increasing as the number of parameters and complexity of the model increases (Roberts et al. 2004). The variational Bayes methodology is based on being able to integrate over computationally or analytically intractable models by minimising the distance between an approximate model which is solvable and the true intractable model distribution (Roberts et al. 2004). The authors found that a three-state model was the most suitable for the analysis of the homing pigeon navigation tracks and that differences exist between state probabilities for birds released from different sites. The three-states were assumed to correspond to high, intermediate and low-complexity behaviour on the flight paths leading to the pigeon loft. They conclude that HMMs are a powerful analysis technique for the complex data obtained from movement data and that they could be applicable to a wide variety of biological problems.

In a study using dive depth and duration data from Macaroni penguins (*Eudyptes chrysolophus*), Hart et al. (2010) used the theory and code applied in the pigeon study by Roberts et al. (2004) to determine the behavioural states of the penguins. Dive data were collected for 103 breeding penguins at Bird Island, South Georgia. They found that a two-state HMM performed the best since models with additional states had very low proportion of state allocation to the higher states. They assumed that the two states corresponded to short-shallow dives and long-deep dives, which may or may not be associated with feeding. The authors used the HMM states in association with the rate of mass gained during an expedition away from the breeding island to investigate whether the amount of time spent in the state was related to the bird's weight gain. This would have allowed for

interpretation of the states in terms of foraging. However, no evidence could be found to support this hypothesis, but this could be due to the weight gain being dependent on the reproductive stage of the bird (Hart et al. 2010). They investigated the influence of the time of day using Generalized Additive Mixed Models and found that time of day was highly significant in the probability of being in a deep dive state once the individual, sex and age of the bird had been taken into account.

Langrock et al. (2012) presented a very similar method for applying HMMs as is used for the KNP study. They fitted 2-state stationary models with a Weibull distribution for the displacement and a wrapped Cauchy distribution for the turning angle between successive locations of nine bison in Canada. They also applied an extension to the normal HMM using a hierarchical semi-Markov model applied to the movement paths of multiple bison. There was a three hour interval between successive location points, and they used data spanning one winter season which corresponded to between 1427 and 1660 observations (Langrock et al. 2012). They fitted models individually and also one model simultaneously to all nine animals using the assumption that model parameters are identical for each individual. The authors found that the Weibull distribution outperformed the Gamma distribution based on the AIC and BIC, similarly the wrapped Cauchy distribution outperformed the von Mises distribution for the turning angle. The results that they obtained were very similar to those obtained for elk movement by Morales et al. using a Bayesian framework, discussed further in Section 7.3 (Morales et al. 2004). The model fitted to all nine animals simultaneously performed better than the nine individual models based on the AIC. The hidden semi-Markov models (HSMM) relax the condition inherited from the HMMs, that the amount of time spent in a state is geometric. The amount of time spent in a state for an HSMM is modelled explicitly from a discrete distribution (Langrock et al. 2012). This means that the probability of remaining in a state, γ_{ii} is modelled separately by the distributions p_i. The HSMMs fitted to the bison movement outperformed the HMMs for seven out of the nine animals. Residual analyses

were used to show that the models provided an adequate fit to the data. The authors used a further extension to the HSMM by fitting a hierarchical model to all the movement which assumed a common shape parameter, but varying scale parameters for the Weibull distribution across all individuals. This hierarchical HSMM provided the best fit across all models fitted by Langrock et al. (2012).

From these applications in the literature, it can be seen that HMMs have consistently provided realistic results for the situation in which they were applied. They have been used to model the movement of a variety of species at a number of different time scales between successive observations. The models have been extended in some of the studies to allow for the inclusion of covariates or to build a hierarchical model structure. While most authors felt that the results they obtained provided a realistic picture of the animal's true behaviour, further validation of the predicted behaviour is need to confirm the efficacy of these models. A number of different techniques and metrics have been used to determine the goodness-of-fit of the model and inform the model selection. A consistent method to do this would be useful across the movement studies. The majority of the studies which have applied HMMs have used a short time-series of observations which negates the need to take into account the seasonality within the model. A variety of ad-hoc tests have been used to investigate the differences in movement patterns across species, and more work is needed in this area to find the most suitable methods for comparing patterns across regions and species. The HMMs can be fitted using either a Bayesian or frequentist approach and have been successfully applied to the movements of both marine and terrestrial species. They are a very promising suite of models for the application to movement tracks collected using GPS locations.

This chapter aims to investigate the effect that the time scale of the input data has on the most suitable statistical distribution to model the movement data, as well as the effect of the length of the time series on the ability of AIC and BIC to select

the correct model. The effect of including or excluding the turning angle between locations is also investigated in Chapter 6. These are areas which have not been addressed in the literature.

5.8 Analysis

Few studies have compared the results obtained from direct numerical maximization and the EM algorithm. In this study both methods are used to determine the model parameter estimates for each animal. Stationary and non-stationary models are fitted using direct numerical maximization to contrast the results based on different assumptions of stationarity of the Markov chain. The Weibull distribution has been used in a number of the animal movement studies to model the step displacement. Since the log-normal distribution was found to be the most suitable distribution for describing the movement distances using the independent mixture models, this is used as the state dependent distribution for the HMMs. However a comparison of the fit of the models using the log-normal, Gamma and the Weibull distributions was done and the log-normal distribution always outperformed the other distributions.

5.9 Data Preparation

The data preparation for the HMMs was slightly different to that of the independent mixture models. Due to the battery life of the collars, some of the collars were replaced in some of the herds. For the Pretoriuskop herds, the old collar was removed and the new collar replaced onto the same animal. For these herds, the data from before and after replacement for each herd are merged to form one dataset per herd. However for the Punda Maria Buffalo and Sable herds, multiple collars were deployed within the same herd at the same time. Unlike the mixture models where the order in which the observations were collected was

irrelevant, the HMMs are dependent on the time series nature of the data. For both the sable and the buffalo, there were times when the collared animals were moving in the same herd and times when the herds split and they were moving separately. Also for the sable herd, a 'replacement' collar was placed on an animal in the herd, in order to extend the tracking period. However, unlike the Pretoriuskop herds, this replacement collar was placed onto a different animal within the herd, and there is an overlap period when both collars in the herd collected data. To preserve the time series integrity and to avoid throwing away data, models are run separately for each animal.

In order to create a complete time series dataset, missing observations were included in the dataset and coded as such to indicate the occasions when a location was not obtained. As discussed in Section 2.3.1, the GPS units could fail to connect with sufficient satellites to obtain a location reading, very occasional overwriting of stored data if the animal remained out of cell phone range for a long time or the collars had been set to a longer time interval (variation in frequency of observations). This means that each herd has a complete time series dataset spanning the date and time of the collar deployment to the last recorded observation from the GPS unit, with missing locations coded as such. The first analysis was run using hourly data, similar to the mixture model analysis. However, a second analysis was run using daily data. In this case displacements were calculated between the locations of the animals at 24 hour intervals. This daily analysis investigates different behaviour compared with the hourly study, in particular focusing on spatial regions used by the animal. The behavioural states associated with hourly movement were assumed to correspond to the animal resting, foraging or moving. However, their displacement from one day to another will correspond to different behaviour depending on the time of day at which the daily displacements are anchored. For example, an analysis of the displacement between 8am locations is likely show how the foraging area of the herd is changing from day to day. In contrast, an analysis of the 2am locations is more likely to show how the resting sites or places of nocturnal safety are changing

over time. The analysis of the daily movement data should allow for a larger scale understanding of the inter- and intra-patch movements that the herds are making and the utilisation periods of these patches. Models are run using the data anchored at four different times of day, namely 2am, 8am, 2pm and 8pm. These times were chosen to capture information about the late night resting areas, morning foraging patches, early afternoon resting sites and the evening foraging/resting sites. However, during the late dry season, the morning feeding is disrupted by journeys to water. As mentioned in Chapter 2, during the early part of the study the collars were set to record at six hour intervals in order to prolong the battery life and to ensure that the data collection period covered the complete seasonal cycle. The times selected for daily analysis here correspond to the data collection times during these 6-hourly recording periods, in order to maximise the available data for analysis.

5.10 Model Fitting

2-, 3- and 4-state log-normal HMMs were run. The 4-state model was the most complex model used, as models with more states (and hence more parameters to estimate) become more computationally intensive to fit, are sensitive to starting parameters and difficult to interpret biologically. As mentioned above, when using maximum likelihood to determine the parameter estimates, the likelihood will always increase as the complexity of the model increases. However, when estimating a model with more than 20 parameters, the likelihood function is found to be highly multimodal with many local maxima found when starting with different initial values (Zucchini, MacDonald 2009). For this reason, it was felt that running models with more than four states would be unstable and could confuse the analysis more than it would add to it. From the biological perspective, 4-states were possible to interpret in terms of the known behaviours of these animals. The sample size also plays a role in how easily a model is able to converge, and more complex models with many states will require a larger sample before convergence can be obtained, if possible.

Initial starting values for the HMMs were chosen to be similar to the parameter estimates obtained for the log-normal mixture models for each herd since the analyses are expected to provide a reasonably similar group structure and state-dependent distributions as those obtained using the independent mixture models. The initial values for the transition probability matrix were chosen to have a high value (between 0.6 and 0.8) on the main diagonal and small values off the diagonal (Zucchini, MacDonald 2009).

5.11 Goodness-of-fit and Model Selection

The AIC and BIC were used to determine which HMM provided the best fit to the observed data. The autocorrelation was calculated for the observed sample data and for the fitted HMM to investigate whether there was a correspondence between them. qq-plots of the standard normal pseudo-residuals were plotted to assist with the model selection and to identify outliers.

5.12 Distributions for modelling step length displacement

In the analysis of the ungulate movements using independent mixture models, a mixture of log-normal models were always selected as the best performing mixture of distributions. The log-normal distribution did not require the complication of adding offsets to the model as was required for the Gamma distribution. For these reasons the log-normal distribution was chosen as the underlying distributions for the HMM analysis. Since the independent mixture models and HMMs are very similar, it is reasonable to assume that the same distribution is likely to perform well for both models given the nature of the movement data. The log-normal and Gamma distribution density function, expected value and variance were defined in Section 3.7.2.

However for the daily models, the fit of log-normal, Weibull and Gamma distributions is compared. The requirement for the offset parameters was not needed with the Gamma and Weibull distributions applied to the daily data. The Weibull distribution has been used as the state dependent distribution in a number of movement studies. The density function for the Weibull distribution with shape parameter *a* and scale parameter *b* is given by:

$f(x) = \frac{a}{b}\left(\frac{x}{b}\right)^{a-1} e^{\left(\frac{-x}{b}\right)^{a}}$ for x > 0. The expected value and variance of the distribution are given by:

$$E(X) = b\Gamma\left(1+\frac{1}{a}\right) \text{ and } Var(X) = b^2\left(\Gamma\left(1+\frac{2}{a}\right) - \left(\Gamma\left(1+\frac{1}{a}\right)\right)^2\right).$$

5.13 Stationarity

As defined above, a Markov chain has a stationary distribution **δ** if ***δ* Γ** = ***δ*** and ***δ*1'** = 1. A stationary Markov chain starts from its stationary distribution and continues with that distribution through time. This implies that the probability of being in a state does not change through time. For the studies mentioned from the literature which have used HMMs to model animal movement data, some have assumed stationarity of the Markov chain (Franke, Caelli & Hudson 2004, Franke et al. 2006, Patterson et al. 2009, Tucker, Anand 2005) while others have not, see for example (Pedersen et al. 2011a, Pedersen et al. 2011b, Zucchini, Raubenheimer & MacDonald 2008). In some applications it is not specified and it is actually not clear whether the assumption of stationarity of the Markov chain has been made.

Given that environmental conditions change throughout the year due to the changing seasons and variability of climatic factors, it seems likely that the

unconditional probability of entering a state is likely to change throughout the year which would imply not assuming stationarity of the underlying Markov chain. In order to understand the impact of assuming stationarity of the Markov chain, both stationary and non-stationary HMMs are run for each herd. The results are compared to investigate whether the parameter estimates are influenced by the decision. For homogenous HMMs, the parameter estimates should not differ much using either the assumption of stationarity or non-stationarity of the time series. The estimates should only differ if the time series is very short and contains strong serial correlation.

5.14 Results

Unlike the mixture model analysis, the datasets prepared for the HMM analyses include missing data, for observations that were missed due to changes in the observation frequency, the collar being unable to connect with sufficient satellites or other unknown reasons. An initial analysis of the amount of missing data introduced into the datasets is shown in Table 5.1 and the Appendix Section 5.16 Table 5.7 for the hourly and daily datasets respectively. A very large percentage of the data are missing for some of the hourly datasets, with some having more than 80% missed locations (Sable143, Sable149 and Zebra141). These are the older collars which were set to record at six hour intervals for much of the time they were on the animal. The maximum run length of continuous data is shown in column 4. Some of the collars only have a maximum of around three days of continuous data, while PK_Shitlave has a run of 62 days with continuous hourly observations (Collar AM286). As expected, the daily datasets, anchored to coincide with the timing of the six hourly recording periods, have fewer missed locations. The average missed locations are about 12%. The 2pm locations had the highest rate of missing data for nine of the datasets, with the other four having the highest miss rate at the 8am reading. This is consistent with the comment in Graves & Waller (2006), who mention that the fix success of collars varies by

hour of the day. They had an average success rate of 72% overall. For the daily datasets used in this study, the overall fix success is 88%.

Table 5.1 - Summary of Missing data for hourly time series

	TotalObs Incl Missing	Missing	%Missing	Max_data_run	Max_run_days
PM_Sable143_Hourly	11850	11339	95.69	89	3.71
PM_Sable146_Hourly	3894	2989	76.76	142	5.92
PM_Sable149_Hourly	8982	8417	93.71	67	2.79
PM_Sable279_Hourly	4918	1412	28.71	271	11.29
PK_Numbi_Hourly	14983	9180	61.27	237	9.88
PK_Phabeni_Hourly	24305	10220	42.05	315	13.13
PK_Shitlave_Hourly	19342	9671	50.00	1499	62.46
PM_Buffalo150_Hourly	3463	2400	69.30	164	6.83
PM_Buffalo152_Hourly	26121	11451	43.84	329	13.71
PM_Zebra141_Hourly	11317	10318	91.17	132	5.50
PM_Zebra142_Hourly	21001	10322	49.15	683	28.46
PM_Zebra277_Hourly	26008	3785	14.55	472	19.67
PM_Zebra280_Hourly	6378	1870	29.32	93	3.88

Homogenous but non-stationary and stationary models were run for every animal using direct numerical maximization, as well as non-stationary models using the EM algorithm. The models using the EM algorithm took much longer to run than those using direct numerical maximization, which was not unexpected. In many cases the EM algorithm did not converge for the more complex 4-state model, even though the estimation using numerical maximization did converge. It was possible to get the EM algorithm to converge by relaxing the convergence criterion, however this often led to the model converging to local maxima, even if the starting values used were based on the parameter estimates obtained using direct maximization. All the models were checked with a number of different starting values. If the initial values were extreme, particularly for the state

dependent distribution parameters, the model could sometimes converge to a local maximum, with many of the parameters estimates at zero or very large numbers. In this case, the model is struggling to converge correctly since the initial parameter values are very close to the edge of the boundary space. However, if reasonable parameter estimates were chosen, the models converged to the same maximum, assumed to be the global maximum. Even the 4-state models, which have the most complex likelihood surface of the models used, converged consistently to the global maximum despite the changing starting parameters. If models with more states were estimated, issues with convergence to local maxima are likely to become more problematic as the likelihood surface is even more complicated.

5.14.1 Hidden Markov Model Simulation

In order to determine the accuracy of the model selection criteria and the Viterbi algorithm for state allocation, a series of states of length 100, 1 000, 10 000 and 100 000 were simulated with the proportion of states corresponding to the stationary distribution. These states are considered the "known" sequence of states. Data observations are simulated based on the state dependent distributions corresponding to the 'known' states. For this simulation, the distribution parameters were chosen to be similar to those obtained from the previous analyses for animal movement using log-normal distributions. Once the data observations have been simulated, an HMM is fitted to the simulated data, using reasonable starting values. As with the mixture model simulation exercise, models of 2-, 3- and 4-states were fitted to each of the true 2-, 3- and 4-state mixtures. The AIC and BIC were determined for each of the fitted models and the results are shown in Table 5.2. For the true 2-state model, the BIC selected the correct model every time, while the AIC was correct in all cases except for the 1000 observation dataset where the more complex 3-state model was selected as the best. The true 3-state model was poorly predicted, with the AIC and BIC both selecting the simpler 2-state model for the 100 observation series. AIC incorrectly selected the more complex 4-state model for the 10 000 and 100 000 series, with BIC also

selecting the 4-state model for the 100 000 observations. The true 4-state model was better predicted with the AIC and BIC selecting the correct number of states for the 10 000 and 100 000 data series.

Table 5.2 - Model selection accuracy based on simulation

2-state Log-normal model				
Model	**Sample size**	***Log- like***	***AIC***	***BIC***
2-state	100	38.45	88.89	104.52
3-state	100	37.75	99.50	130.76
4-state	100	35.54	111.08	163.19
2-state	1000	449.00	910.01	939.46
3-state	1000	442.71	909.43	968.32
4-state	1000	442.16	924.31	1022.47
2-state	10000	4695.87	9403.74	9447.00
3-state	10000	4692.59	9409.19	9495.71
4-state	10000	4691.06	9422.11	9566.32
2-state	100000	47236.37	94484.75	94541.82
3-state	100000	47235.01	94494.02	94608.18
4-state	100000	47234.41	94508.81	94699.07
3-state Log-normal model				
Model	**Sample size**	***Log- like***	***AIC***	***BIC***
2-state	100	32.88	77.76	93.39
3-state	100	28.80	81.59	112.85
4-state	100	25.93	91.87	143.97
2-state	1000	677.64	1367.28	1396.73
3-state	1000	628.10	1280.21	1339.10
4-state	1000	625.04	1290.08	1388.24
2-state	10000	6417.68	12847.36	12890.62
3-state	10000	6079.67	12183.35	12269.87
4-state	10000	6060.94	12161.88	12306.09
2-state	100000	61996.20	124004.39	124061.47
3-state	100000	58507.93	117039.86	117154.02
4-state	100000	58402.89	116845.79	117036.05

4-State Log-normal model				
Model	**Sample size**	***Log- like***	***AIC***	***BIC***
2-state	100	76.38	164.76	180.39
3-state	100	70.81	165.62	196.88
4-state	100	66.18	172.35	224.46
2-state	1000	798.15	1608.29	1637.74
3-state	1000	765.61	1555.23	1614.12
4-state	1000	760.70	1561.40	1659.56
2-state	10000	7283.13	14578.25	14621.51
3-state	10000	6961.00	13945.99	14032.52
4-state	10000	6922.09	13884.18	14028.38
2-state	100000	74307.79	148627.57	148684.65
3-state	100000	71176.17	142376.34	142490.49
4-state	100000	70774.94	141589.88	141780.14

Both measures selected the incorrect simpler models for the shorter data series. In general it is expected that the accuracy of the model selection for the number of states would improve as the number of observations increases. This was true for the 2- and 4-state models; however, it was not the case for the 3-state series where the more complex 4-state models were selected as the 'best' fit to the simulated data. This result is somewhat unexpected and suggests that other measures such as the autocorrelation and the pseudo-residuals should also be used to inform the decision about how many states to use in the model. The biology of the situation being modelled is also informative and should play a role in the decision about the number of states to use.

A second analysis was done to determine the accuracy of the Viterbi algorithm for predicting the true underlying state of the observation. Once again, the true model and 'known' state sequence was based on parameters similar to those obtained during the previous analyses of the movement data. In most cases, approximately 18% of the simulated observations were allocated to the incorrect state. Increasing the number of observations used for the simulation did not improve the accuracy

of the state allocation at all. The accuracy of the local decoding was very similar to that of the global decoding (Viterbi) which indicates that the optimization over the entire sequence of states does not add much accuracy to the allocation of the state. However, the Viterbi algorithm is not more computationally intensive compared with the local decoding. If it was, it would be debatable whether the effort to obtain the most likely *sequence* of states is worthwhile in comparison to identifying the most likely state based on the observation. The accuracy of the Viterbi state allocation was checked using the code developed for this study, as well as the R package 'HiddenMarkov'. A similar accuracy level was found using both methods. While an error rate of 18% sounds reasonably high for the state allocation, it still provides a suitable method for predicting the underlying hidden state with a confidence of around 80% if the fitted model is indeed the true model. However, if the state-dependent distributions are very different from one another, then the accuracy of the state allocation process improves. This makes sense in that if the underlying groups within the data are very different from one another, it is easier to differentiate them. By simulating HMMs using different state-dependent distributions and parameters, a maximum state allocation accuracy of approximately 97.5% was found. This accuracy level was obtained for HMMs simulated using the Poisson, Normal and Log-normal distributions, with very diverse groups within the data imposed through the choice of distribution parameters. Patterson et al. (2009) found that very different groups within the data also improved the accuracy of the parameter estimates that were obtained. However, in the context of applying hidden Markov models to animal movement data similar to the sable, buffalo and zebra data, there is an expected state allocation accuracy of approximately 80%, based on a 4-state model as simulated above. True models with fewer states should have increased allocation accuracy since there is less expected overlap between the groups, whereas multi-state models are likely to have more overlap between states.

Although this simulation implies that the use of HMMs is not completely accurate for the analysis of movement data, any modelling and prediction process will have

an associated risk of incorrect prediction. However no statistical process is completely accurate! As shown above, it is possible to quantify the expected error rate for the fitted model, under the assumption that it is indeed the true model. This means that the interpretations drawn from the models can be tempered by an understanding of the expected accuracy of the state allocation. The simulation to check the accuracy of the Viterbi allocation is simple to code and quick to run, so it can easily be added to each herd's Hidden Markov movement analysis to determine the expected accuracy of the state allocation given the fitted model as the true model. Despite the potential errors, the models still provide insight into the behaviour of the animals which would not be possible with the location points alone.

5.14.2 Hourly Movement Results

The results for the EM algorithm (non-stationary) and direct numerical maximization (stationary and non-stationary) are shown in Table 5.3 for the Punda Maria Sable herd and Table 5.4 for the Punda Maria buffalo herd. Two-, three- and four-state models were run for each of the above techniques. The 'gamma1' to 'gamma4' columns show the values on the main diagonal of the transition probability matrix.

Punda Maria Sable279

For the Punda Maria herd, four datasets were created, one for each of the animals which had a collar. As discussed above, this was done to preserve the time series nature of the data, and not shift from one animal to another's movement, as was done for the mixture models. A 4-state stationary log-normal HMM is selected as the best model for the Punda Maria sable (AM279) data by AIC and BIC (highlighted in red). A 4-state model fitted using the EM algorithm failed to converge even when the convergence criterion was relaxed. This animal had the lowest percentage of missing data for the Punda Maria sable herd. A 4-state model is consistent with the results in Chapter 3, where a 4-state model was also selected

for this herd. The expected movement distance in each state is 50m, 230m, 760m and 2.10km respectively. This is in comparison to the mixture models which had expected movement distances in each state of 50m, 270m, 910m and 2.25km respectively. It is encouraging that the distributions of the movement in each of the states are similar between the two methods. The transition probability matrix has a 0.7 probability of remaining in a resting state, with 0.21 of shifting to foraging behaviour from resting. The probability of transitioning from resting to one of the more active behavioural states (moving or relocating) is very low, 0.08 and 0.01 respectively. This is a different pattern to the more active states which have higher probabilities of transitioning to other states although the diagonal probability of remaining in the same state is always the highest. From this, it can be inferred that the behavioural states usually persist for more than one hour (remain in the same state), and the bouts of resting are longer than those for the other 3 states. When the animal is foraging, it is most likely to continue foraging (0.41) or transition to resting (0.38). While making the longer relocating movements, the animal is most likely to continue in that state, or shift to the slightly shorter movement state. This confirms that these relocating treks, assumed to often be movements to water, take a few hours to complete and will place extra stress on the animal in the late dry season when they make these long journeys searching for food and water. The stationary distribution parameter for the movement state (0.18) is reasonable for the proportion of time that grazers spend moving between feeding areas (Owen-Smith 2002).

Table 5.3 - Hidden Markov Model results for Punda Maria Sable herd

	logLik	AIC	BIC	delta	mu	sigma	gamma1	gamma2	gamma3	gamma4	Mean
EM.State1	-1538.93	-3063.85	-3018.35	1.00	-3.18	1.36	0.86	0.14	NA	NA	0.10
EM.State2				0.00	-0.48	0.84	0.32	0.68	NA	NA	0.89
NLM.NonStat.State1	-1538.93	-3063.85	-3020.89	1.00	-3.18	1.36	0.86	0.14	NA	NA	0.10
NLM.NonStat.State2				0.00	-0.48	0.84	0.32	0.68	NA	NA	0.89
NLM.Stat.State1	-1538.55	-3065.10	-3028.28	0.69	-3.18	1.36	0.86	0.14	NA	NA	0.10
NLM.Stat.State2				0.31	-0.47	0.84	0.32	0.68	NA	NA	0.89
EM.State1	-1662.57	-3297.14	-3206.13	1.00	-3.52	1.20	0.76	0.22	0.02	NA	0.06
EM.State2				0.00	-1.11	0.76	0.36	0.56	0.07	NA	0.44
EM.State3				0.00	0.36	0.50	0.13	0.24	0.63	NA	1.63
NLM.NonStat.State1	-1662.57	-3297.14	-3211.22	1.00	-3.52	1.20	0.76	0.22	0.02	NA	0.06
NLM.NonStat.State2				0.00	-1.11	0.76	0.36	0.56	0.07	NA	0.44
NLM.NonStat.State3				0.00	0.36	0.50	0.13	0.24	0.63	NA	1.63
NLM.Stat.State1	-1662.02	-3300.03	-3226.39	0.57	-3.52	1.20	0.76	0.22	0.02	NA	0.06
NLM.Stat.State2				0.33	-1.11	0.76	0.36	0.56	0.07	NA	0.44
NLM.Stat.State3				0.10	0.36	0.50	0.13	0.24	0.63	NA	1.63
EM.State1	-1709.57	-3373.14	-3223.62	1.00	-3.71	1.13	0.70	0.21	0.08	0.01	0.05
EM.State2				0.00	-1.68	0.67	0.38	0.41	0.19	0.03	0.23
EM.State3				0.00	-0.40	0.50	0.25	0.25	0.41	0.08	0.76
EM.State4				0.00	0.69	0.32	0.08	0.06	0.32	0.53	2.10
NLM.NonStat.State1	-1709.57	-3373.14	-3231.99	1.00	-3.71	1.13	0.70	0.21	0.08	0.01	0.05
NLM.NonStat.State2				0.00	-1.68	0.67	0.38	0.41	0.19	0.03	0.23
NLM.NonStat.State3				0.00	-0.40	0.50	0.25	0.25	0.41	0.08	0.76

NLM.NonStat.State4				0.00	0.69	0.32	0.08	0.06	0.32	0.53	2.10
NLM.Stat.State1	-1708.93	-3377.86	-3255.12	0.50	-3.71	1.13	0.70	0.21	0.08	0.01	0.05
NLM.Stat.State2				0.26	-1.68	0.67	0.38	0.41	0.19	0.03	0.23
NLM.Stat.State3				0.18	-0.40	0.50	0.25	0.25	0.41	0.08	0.76
NLM.Stat.State4				0.06	0.69	0.32	0.08	0.06	0.32	0.53	2.10

Table 5.4 - Hidden Markov Model results for Punda Maria Buffalo herd

	logLik	AIC	BIC	delta	mu	sigma	gamma1	gamma2	gamma3	gamma4	Mean
EM.State1	-6645.42	-13276.85	-13219.65	1.00	-3.78	1.25	0.70	0.30	NA	NA	0.05
EM.State2	-6645.42	-13276.85	-13219.65	0.00	-1.00	0.79	0.25	0.75	NA	NA	0.50
NLM.NonStat.State1	-6645.42	-13276.85	-13223.90	1.00	-3.78	1.25	0.70	0.30	NA	NA	0.05
NLM.NonStat.State2	-6645.42	-13276.85	-13223.90	0.00	-1.00	0.79	0.25	0.75	NA	NA	0.50
NLM.Stat.State1	-6644.63	-13277.27	-13231.88	0.45	-3.78	1.25	0.70	0.30	NA	NA	0.05
NLM.Stat.State2	-6644.63	-13277.27	-13231.88	0.55	-1.00	0.79	0.25	0.75	NA	NA	0.50
EM.State1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EM.State2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EM.State3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NLM.NonStat.State1	-7548.01	-15068.02	-14962.12	1.00	-4.05	1.16	0.67	0.00	0.33	NA	0.03
NLM.NonStat.State2	-7548.01	-15068.02	-14962.12	0.00	-1.90	0.82	0.55	0.45	0.00	NA	0.21
NLM.NonStat.State3	-7548.01	-15068.02	-14962.12	0.00	-0.77	0.71	0.00	0.32	0.68	NA	0.59
NLM.Stat.State1	-7547.05	-15070.09	-14979.33	0.38	-4.05	1.16	0.67	0.00	0.33	NA	0.03
NLM.Stat.State2	-7547.05	-15070.09	-14979.33	0.23	-1.90	0.82	0.55	0.45	0.00	NA	0.21
NLM.Stat.State3	-7547.05	-15070.09	-14979.33	0.39	-0.77	0.71	0.00	0.32	0.68	NA	0.59

EM.State1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EM.State2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EM.State3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EM.State4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NLM.NonStat.State1	-7891.81	-15737.61	-15563.64	1.00	-4.05	1.15	0.66	0.00	0.03	0.31	0.03
NLM.NonStat.State2	-7891.81	-15737.61	-15563.64	0.00	-1.95	0.81	0.59	0.41	0.00	0.00	0.20
NLM.NonStat.State3	-7891.81	-15737.61	-15563.64	0.00	0.00	0.48	0.02	0.18	0.60	0.19	1.12
NLM.NonStat.State4	-7891.81	-15737.61	-15563.64	0.00	-1.01	0.57	0.00	0.35	0.08	0.57	0.43
NLM.Stat.State1	-7890.85	-15741.69	-15590.41	0.38	-4.05	1.15	0.66	0.00	0.03	0.31	0.03
NLM.Stat.State2	-7890.85	-15741.69	-15590.41	0.22	-1.95	0.81	0.59	0.41	0.00	0.00	0.20
NLM.Stat.State3	-7890.85	-15741.69	-15590.41	0.09	0.00	0.48	0.02	0.18	0.60	0.19	1.12
NLM.Stat.State4	-7890.85	-15741.69	-15590.41	0.31	-1.01	0.57	0.00	0.35	0.08	0.57	0.43

Punda Maria Buffalo152

The movement pattern for the buffalo herd, represented by collar AM152, is somewhat different to the sable movement. The model fitting results for this herd are Table 5.4. As with the sable herd's models, the results show almost identical distribution and transition parameters fitted using either the stationary or non-stationary assumptions. However, the EM algorithm failed to converge for the more complex 3- and 4-state models. Here the expected distances moved in each state are 30m, 200m, 430m and 1.12km. Although the 4-state model was once again selected as the best model according to AIC and BIC, the distributions associated with each state are different to the sable movement states, particularly for the active 'movement' and 'relocating' states. This clearly shows that the HMMs are sensitive to different movement strategies, and that the buffalo are not making the very long relocating movements that were seen for the sable living in the same region. Based on the stationary distribution, the buffalo spend 38% of the day resting, 22% foraging, 31% moving and 9% relocating. This is a slightly lower proportion of time allocated to resting and foraging than for the sable, but more time allocated to the two active states, although the movements within the active states are much shorter. The transition probabilities are very different though in comparison to the sable movement. Remaining in the resting state is once again the single highest transition probability, however some of the transition probabilities are actually zero within the matrix, and this seems to be as a result of the large degree of overlap between the second and third distributions. The third distribution is in fact probably better described as an active foraging state rather than a movement state. After resting, the herd is most likely to move into this active foraging phase, but never into the less active foraging state. These differences are probably due to the ruminating behaviour of buffalo where the whole herd will lie down to rest and chew the cud.

The pattern with the parameter estimates obtained using the EM algorithm and the stationary and non-stationary direct numerical maximization are very similar with almost no difference between the stationary and non-stationary models in terms of

the parameter estimates for the state-dependent distribution and transition probability matrix. This was shown for the Punda Maria sable and buffalo herds in Table 5.3 and Table 5.4 respectively. This pattern was consistent for all of the herds used in this study including the Pretoriuskop sable and Punda Maria zebra.

As discussed above, there are bouts of missing data in the hourly data series due to the changes in observation frequency. The models were re-run on portions of the data, so as to avoid including these large missing sections in the model. The parameter estimates for these shorter time series were still very consistent with the full data model, which indicates that the HMMs are robust, even when there is a lot of missing data within the dataset.

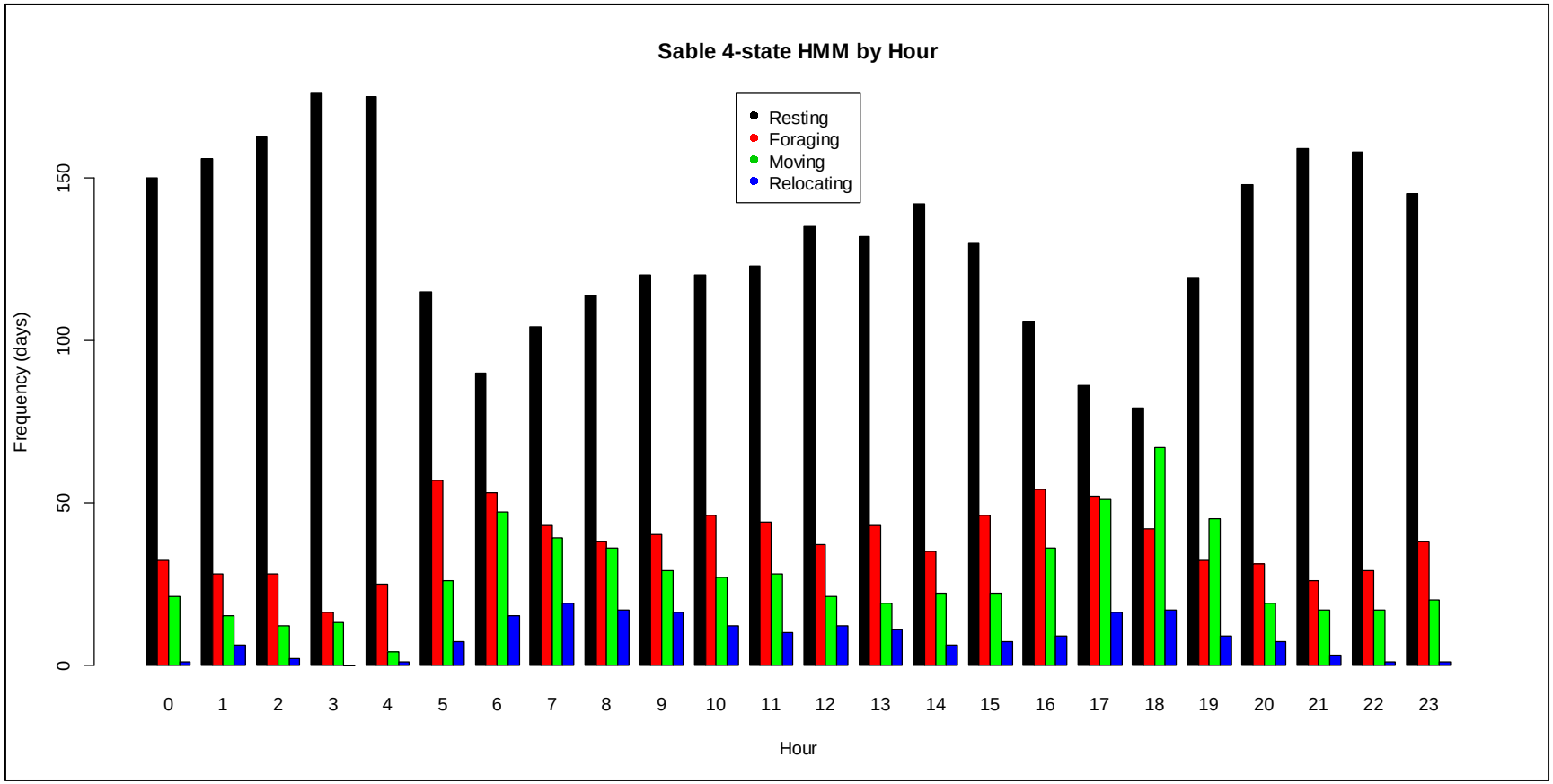

Figure 5.3 - State allocation (using Viterbi) by time of day – PM_Sable279

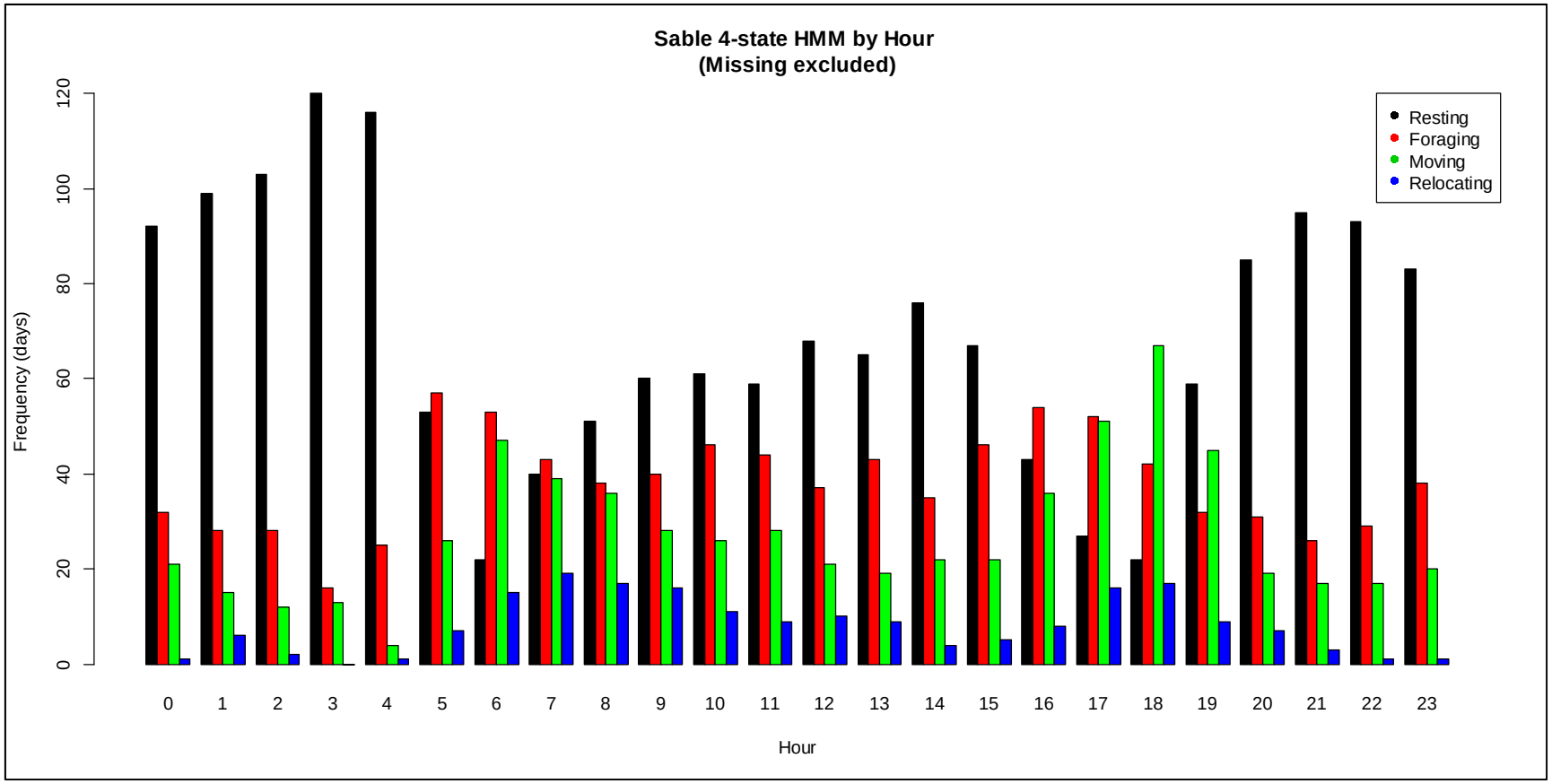

Figure 5.4 - Missing locations removed from State allocation (using Viterbi) by time of day – PM_Sable279

Local decoding and the global decoding using the Viterbi algorithm were run based on the 4-state stationary model. For the Punda Maria sable herd, in 97.7% of the cases, the Viterbi algorithm and the global decoding allocated the same state to an observation. The allocation comparison results are shown in Table 5.5.

Table 5.5 - Comparison of Local and Global decoding - PM_Sable279

	Global_decode			
Local_decode	State 1	State 2	State 3	State 4
State 1	3099	25	0	1
State 2	30	874	8	1
State 3	16	16	636	5
State 4	0	0	9	198

Since the Viterbi algorithm maximises the probability of the state sequence, rather than just the conditional probability of the observation given the state, the global decoding results will be used. The results of the global decoding are plotted against the hour of the day in Figure 5.3. The resting state is clearly dominant throughout the day, although there is a slight dip during the morning and evening times after sunset and before sunrise. This graph illustrates an over-estimation of the resting phase, which is a result of the volume of missing data within the time series. A large amount of the missing locations are allocated to the resting phase, due to the transition matrix having the highest transition probability of remaining within the resting state. This graph is re-run and shown in Figure 5.4, with the predicted states by time of day, with the frequency of states allocated to missing locations removed. This provides a far more realistic view of the animal's behaviour. In this case, the resting state is still dominant during the early hours of the morning, before the foraging spell commences around 6am. The proportion of the resting state increases somewhat during the heat of the day around midday before decreasing again around 4pm, when the evening foraging spell begins. The resting state becomes the dominant state once again around 9pm, and stays that way through to midnight. However, quite a high proportion of the darkness hours are still allocated to the foraging state. As one would expect, the movement state is uncommon during the night to avoid the risk of predation. The movement state

proportion tends to track the foraging state with an increase during the morning after sunrise, then a decrease during the heat of the day and then an increase again in the late afternoon.

These two graphs illustrate that although the volume of missing data within the time series does not seem to affect the model parameter estimates, it does alter the proportions of the allocated states in a way that does not make sense. Although this is a shortfall of the method, it is unreasonable to expect a fitted model to accurately predict the state that the animal is in, when the observation has been missed, particularly when there can be sections with a few months of observations missing continuously. This is however an extreme situation since the collars were set to switch between hourly and six hourly locations. The effect of the missing data will be investigated again for the daily data analysis, when far fewer observations are missing and there are not these large chunks of missing points. In Chapter 8, only the blocks of observed hourly data will be analysed and compared with the mixture model results for the same data series, to remove the effect of the changed location frequencies. Despite the influence of the missing data, the plot of the state allocation excluding the missing sections provides a very reasonable picture of the movement of the animal based on the fitted model. This picture is consistent with the movement expected for these species and is similar to the patterns found using the Mixture Models, which is encouraging since it confirms that the HMM fitting is robust and can cope with a large amount of missing data. This is less likely as the technology for obtaining GPS location data improves.

The ACF of the fitted HMM can be used to help with the assessment of the goodness-of-fit of the model. There is a reasonable similarity between the ACF of the model and the observed hourly data, which is shown in Figure 5.5. As shown in Chapter 4, the movement data of all of these ungulates has a strong cycle of autocorrelation within the data, which can extend to lags corresponding to more than five days' worth of movement. No cyclic or daily trend has been built into

these basic HMMs, so the fitted model does not show this repeating cycle of autocorrelation. However, for the lags up to around lag seven, the ACF of the fitted model is reasonably similar to the ACF of the observed data, shown in Figure 5.5. This means that the dependence structure built into the HMM via the Markov chain, provides a similar dependence structure up to lag five as is present in the observed hourly data.

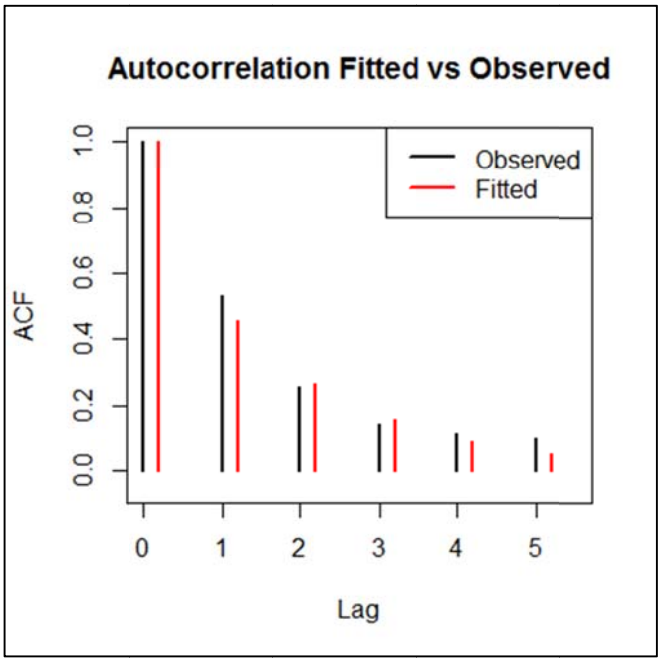

Figure 5.5 - PK_Sable279 - model and sample ACF

A daily or seasonal cycle can be included in the model by introducing time as a covariate, this will be discussed and illustrated in Chapter 6. The other measure used to assess the goodness-of-fit of a model is the pseudo-residual which was introduced in the Section 5.5 above. The pseudo-residuals for the 3-state HMM fitted using the EM algorithm are shown in Figure 5.6. The residuals fall mainly on the straight line except for the outliers at the top and bottom. The outliers at the bottom of the plot indicate that the model predicted fewer very small movements than were observed in the raw data. This is the opposite pattern to that found for the bison data studied by Langrock et al. (2012) whose models predicted more

steps of short length. However, the two clear outlier points are actually the observations where there was GPS error and identical locations (i.e. exact zero movement) were replaced by 0.0001km (1m), which is an artefact of the decision of how to handle exact zero displacements. A Shapiro-Wilks test can be done to check the normality of the pseudo-residuals. However the test is only valid for datasets with 5000 observations or less and hence it is often not applicable for the hourly movement datasets where there are sometimes many more observations. For large samples, inspecting the normal probability plot (qq-plot) for evidence of non-normality is sufficient.

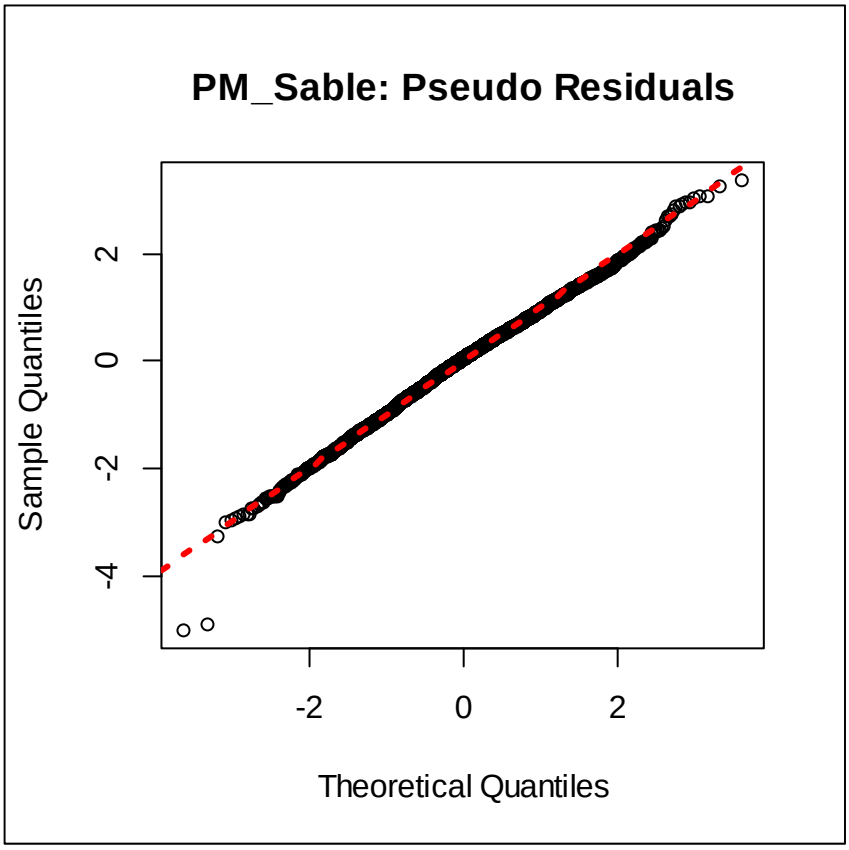

Figure 5.6 - Punda Maria Sable: Q-Q Plot of Pseudo Residuals

5.14.3 Daily Movement Results

Hidden Markov models were fitted to the daily displacements for each of the animals, with the data anchored at either 8am, 2pm, 8pm or 2am. A plot of the

locations obtained at these 6 hourly intervals is shown in Figure 5.7. Very similar areas are used at the different times of day, although the blue dots representing the 2am locations are mostly clustered into certain locations, which are assumed to be areas which provide the sable with protection at night against predators. There are also very few 2am locations in the area generally not unutilized between the northern and southern regions which are more commonly used by the Numbi sable herd. Since the models fitted using direct numerical maximization were computationally easier to run and provided almost identical model parameters, this method was used in preference to the EM algorithm for fitting HMMs to the daily data. The Markov chain was assumed to be stationary based on the fact that the hourly models had exactly the same model parameter estimates using the stationary or non-stationary assumption. Only 2- and 3-state models are fitted to the daily movement since it is biologically unreasonable to expect more complex movement processes between 24 hour displacements. Datasets are also much smaller due to the daily subsampling, so fitting a 4-state model to 200 observations could provide unstable results. In many cases, attempts to fit a 4-state model to the daily data did not converge anyway.

Effect of time of day on state-dependent distribution

The log-normal distribution provided the best fit to the hourly time series for both the mixture models and the HMMs. However, a different time scale between observations influences the distribution of the displacements. This means that different statistical distributions may provide a better fit to the data depending on the time scale of the observations. As discussed in the Hidden Markov Models used in animal movement research section (Section 5.7), the Weibull distribution has been used in many studies modelling the displacement between successive locations.

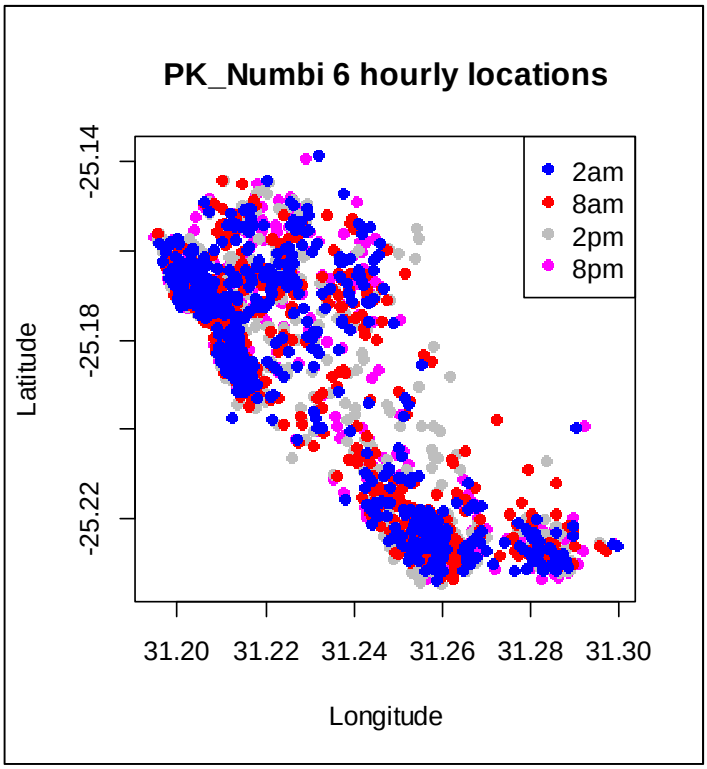

Figure 5.7 – Six-hourly locations of Numbi sable

Very little literature has been found which compares the fit of various distributions, particularly for HMMs. Langrock et al. (2012) did mention that the Weibull distribution provided a better fit than the Gamma distribution for their bison data. For this Kruger National Park study, the fit of the log-normal, Weibull and Gamma distributions are compared and shown in Table 5.8 for the daily displacement data. LN_–LL, W_-LL, G_-LL indicates the maximum log-likelihood for the Log-normal, Weibull and Gamma HMMs respectively, similarly for the AIC and BIC values. The minimum value of the AIC for an animal across the three model distributions for the 2- and 3-states models is highlighted in pink, with the minimum BIC highlighted in orange. The Gamma or Weibull HMMs are always selected as the best model according to both AIC and BIC. The log-normal HMMs are never selected as the best model fit to the data, which is completely different to the hourly data, where the models using the log-normal state-dependent distribution always outperformed the Gamma and the Weibull models. This shows that the time scale of the data certainly does affect

the probability distribution which is most suited for modelling the displacement data, and a variety of distributions should be tested during the modelling process. Fortunately HMMs are flexible enough that any distribution can be used, and packages such as 'HiddenMarkov' are already programmed for many of the common discrete and continuous distributions. Even if the HMMs are programmed individually for a specific distribution, it is reasonably easy to adjust the coding for other state-dependent distributions. An interesting pattern appears in the Appendix Section 5.16 Table 5.8 with the Gamma distribution being favoured for models run for data anchored at 8pm or 2am (the night time locations), whereas the Weibull distributions is selected more often using the 8am and 2pm datasets, particularly the 2pm data. This suggests that the different behavioural processes which occur at the different times of day influence the fit of the model and which distribution is better suited for the analysis. However, the fit of the Gamma and Weibull models is often very similar. In the cases when the Weibull distribution is selected in preference to the Gamma distribution, there is often only a marginal improvement over the fit of the Gamma HMM according to the information criteria. The Weibull distribution was somewhat harder to obtain convergence for some of the models, and these models seemed to be a lot more sensitive to the choice of starting values compared with the Gamma distribution. In some cases, a large variety of initial values for the Weibull HMMs had to be used before the model would converge to an apparent global maximum, with reasonable parameter estimates. Quite often the Weibull models would appear to converge, but some of the parameter estimates were extreme. For these reasons, it seems that the Gamma distribution is the more stable and suitable distribution for modelling the movements of these animals based on their 24 hour displacements.

Results

The model parameter estimates and expected movement distances are shown for the "best" model as selected by AIC for the Numbi herd in Table 5.6 while the fitted Gamma distribution curves are shown in Figure 5.8. These curves have not been weighted by the stationary distribution. The distribution means for state 1

and state 2 for the 2pm (daytime) movements are 0.39km and 1.01km respectively, while the night time distribution means for the 8pm and 2am locations are around 0.5 and 1.4km respectively. For the 8am movements, a 2-state model was selected by AIC and BIC as the best model, with a mean of 0.57km for the first state which is the same as the mean for the first state of the 2am model but the second state has a mean of 1.82km. These displacements within the first state represent the within-patch movements, while longer displacements will represent between-patch movements. For a 3-state model, the states are assumed to correspond to the animal being "encamped" and using a very similar late night position to the previous night, "local" for medium length displacements from one night to the next, or "relocated" when a new patch is used.

Table 5.6 – Numbi Sable: Parameter estimates for Gamma HMM based on AIC selection

Time of Day	δ	α	β	γ1	γ2	γ3	Mean
2pm	0.17	3.03	0.13	0.90	0.08	0.02	0.39
	0.34	2.05	0.49	0.05	0.90	0.06	1.01
	0.49	1.71	1.23	0.00	0.04	0.96	2.11
8pm	0.22	2.18	0.21	0.96	0.03	0.01	0.47
	0.52	1.38	0.94	0.02	0.98	0.00	1.30
	0.26	1.19	1.94	0.00	0.02	0.98	2.31
2am	0.31	1.81	0.29	0.94	0.05	0.01	0.52
	0.42	1.46	1.00	0.05	0.95	0.00	1.46
	0.27	1.12	2.05	0.00	0.02	0.98	2.30
8am	0.38	2.12	0.27	0.94	0.06		0.57
	0.62	1.62	1.13	0.04	0.96		1.82

The results presented in this section focus on the HMMs using the 2am data. The analysis of movements anchored at this time of day will provide some insights into the behaviour of the herd relative to their night time resting positions. These late night positions will be selected by the animal to provide shelter either from predators, cold or other features that the animal choses to avoid. Changes in this late night position can be influenced by shifts in the daytime feeding areas used by

the herd, the presence of predators and how secure they feel in this night time resting area. For the Numbi herd's movements anchored at 2am, the model fit is investigated using the pseudo-residuals and the ACF. These graphs are shown in Figure 5.9. The ACF of the fitted model is similar to the ACF of the observed data, with a gradual decline in correlation for increasing lag. The pseudo-residuals fit a straight line in the qq-plot well. A Shapiro-Wilks test for normality of the pseudo-residuals confirm that they are normally distributed (p = 0.1961). For a Weibull model fitted to the 2am data, the fitted ACF provided a consistent pattern with the observed ACF and the pseudo-residuals were also normally distributed according to the Shapiro-Wilks test (p = 0.1243). The Viterbi algorithm is used to determine the most likely sequence of states and this state allocation relative to the month of the year is shown in Figure 5.10. The relocation state, which corresponds to the herd spending the night on average 2.3kms from the place where they spent the previous night, is only predicted during the months of September to December. This corresponds to the times when conditions are probably the toughest for the animal, before the arrival of the first summer rains. This clearly shows that the HMMs are sensitive enough to pick up this change of behaviour that is only evident at the end of the winter season and before the summer rains. There does not appear to be a clear pattern to the timing of the 'Encamped' and 'Local' states. Since they are far more similar to one another with shorter displacements from day to day, it is not surprising that there is not a clear pattern through the course of the year.

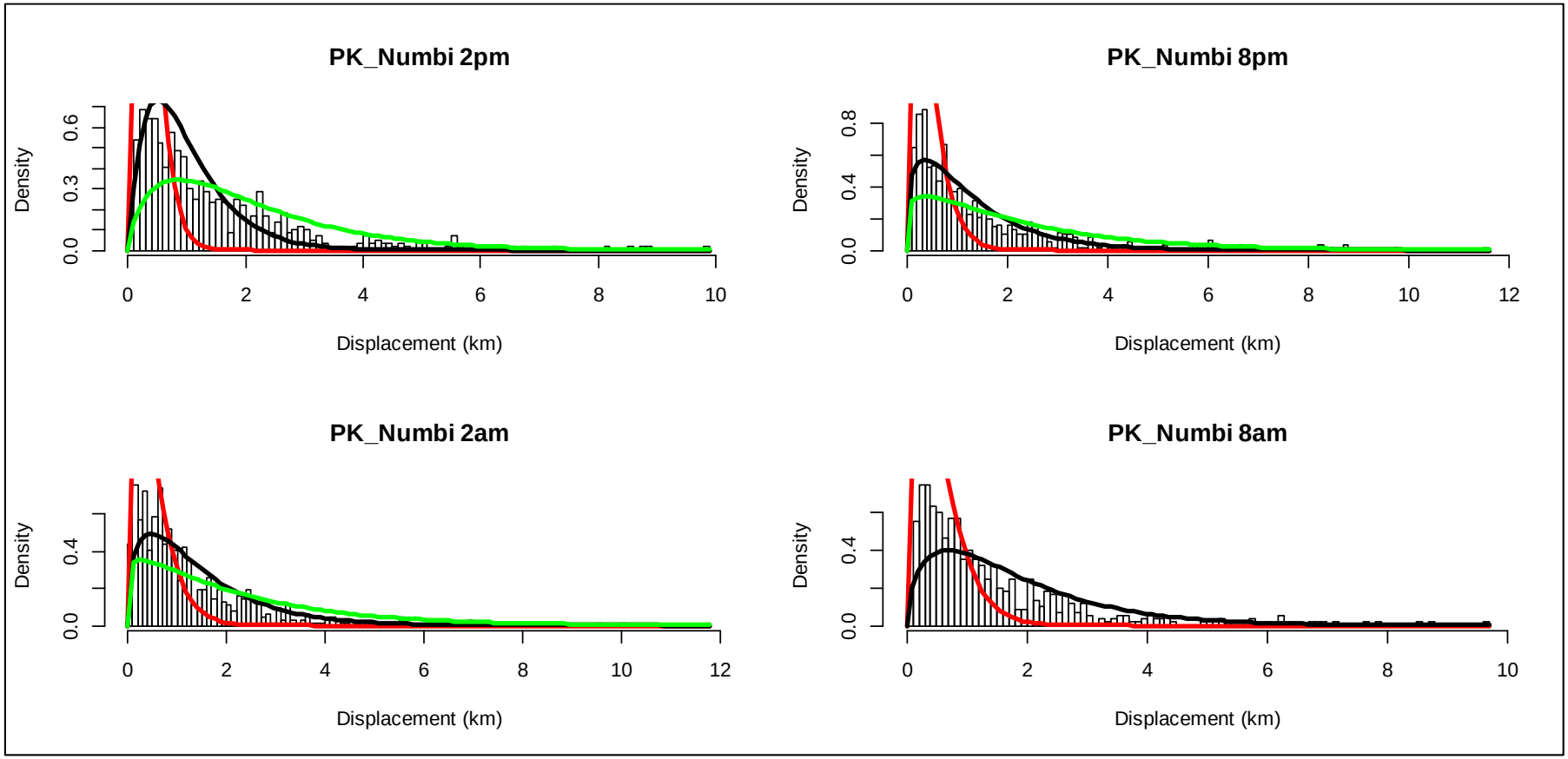

Figure 5.8 - Fitted state-dependent distribution curves - Numbi Sable - State 1 (red), State 2 (black), State 3 (green)

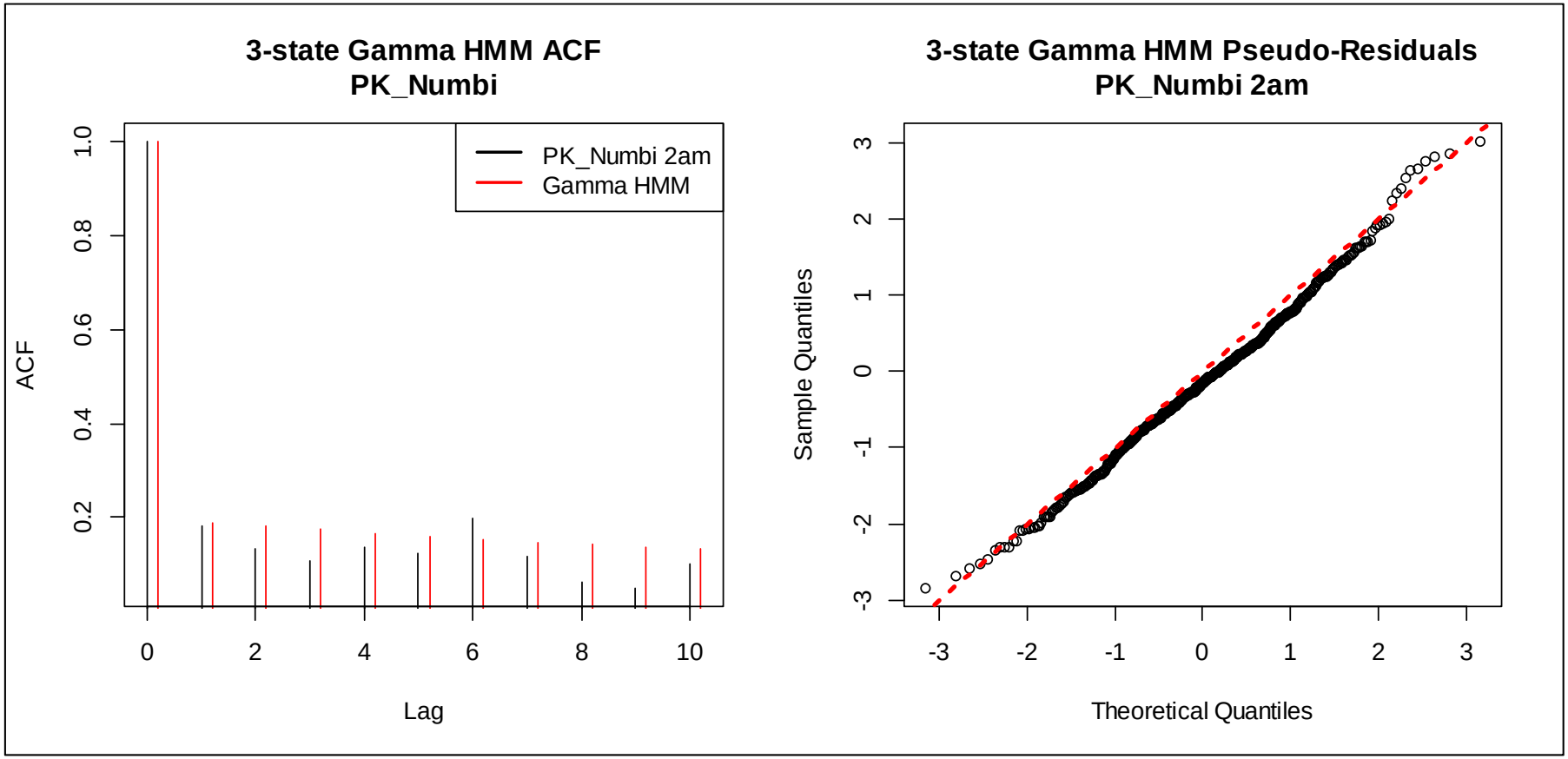

Figure 5.9 - PK_Numbi 2am ACF and Pseudo-residuals - 3-state Gamma HMM

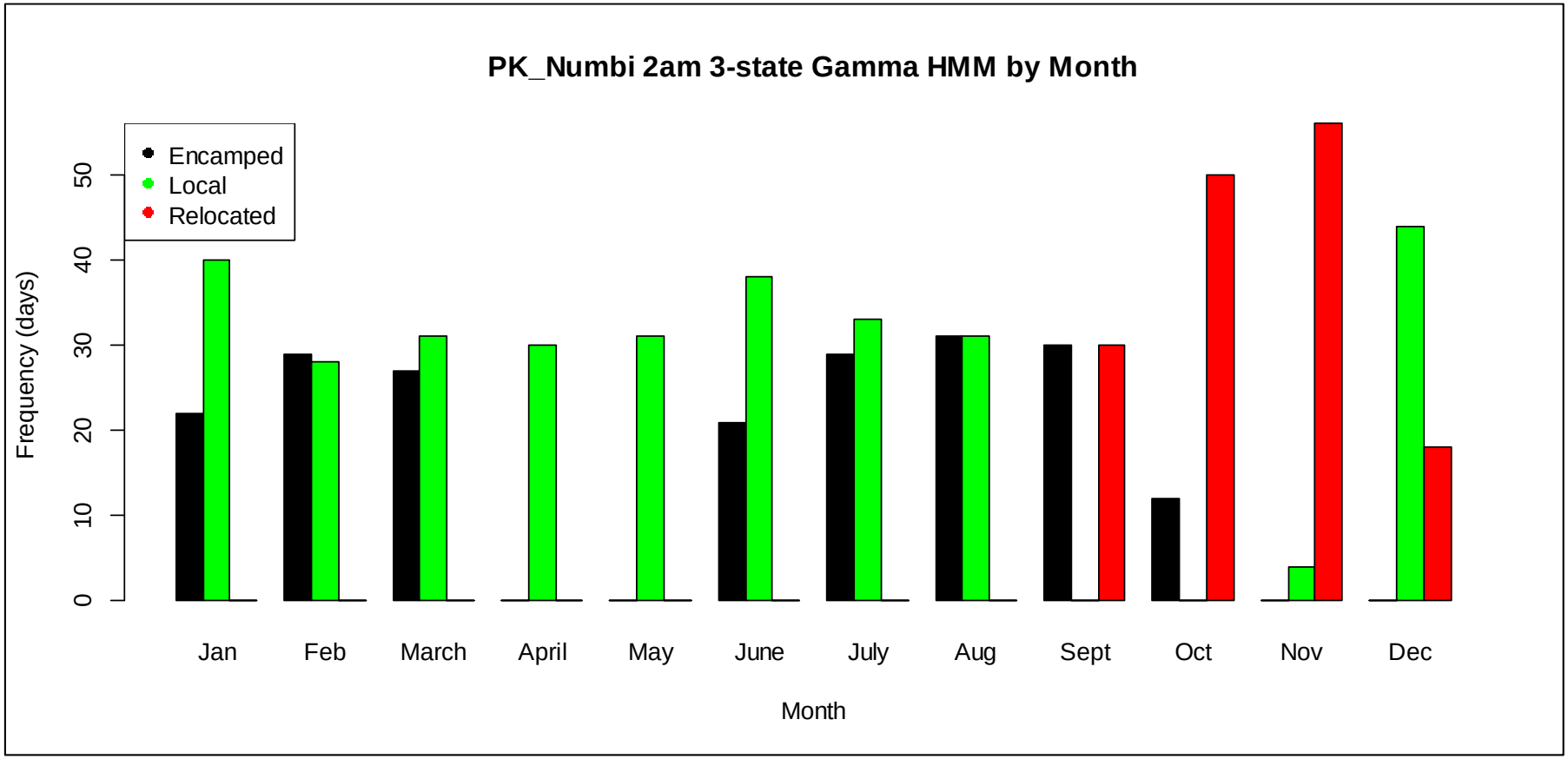

Figure 5.10 - Numbi Sable (2am) Viterbi state allocation by month of the year

The state allocation was also tested using the Weibull model. Only 27 observations out of the whole dataset (n = 665) were allocated to a different state compared to the Gamma model. The Weibull distribution models were sometimes trickier to obtain convergence and were more sensitive to the choice of starting values. For any HMM, issues with convergence are likely to be caused by the model having a flatter likelihood surface. This means that multiple local maxima can be found, with a similar fit for different parameters which would lead to different conclusions. In this case, the preferred family of distributions will be the one with the more peaked likelihood function and a clear global maximum.

5.15 Conclusion

The hidden Markov models are flexible models that can be applied to univariate or multivariate time series data (Zucchini, MacDonald 2009). The models allow for inferences to be made about the behavioural state of the animal based on its observed displacement during an interval of time, and also for investigation of the probabilities associated with shifting from one state to another. These states are assumed to correspond to the behavioural states of the animal. The models presented in this chapter are relatively simple and do not take into account the turning angle between observations, the time of year of the observation or any other ecological covariates.

In this study, it has been shown that HMMs are a technique that can be applied to animal movement data to provide meaningful results which can be biologically interpreted in terms of the movement behaviour of the animal. From this work, it seems that the time scale between observations influences which distribution provides the best fit to the data. The conclusions obtained from this study indicate that a log-normal distribution is the most suitable for these data at the hourly scale. For data at the daily scale, both the Gamma and Weibull distributions provide very similar results, with the Gamma distribution being favoured by the

model selection tests. These conclusions need further analyses from a variety of species and in a variety of locations, to confirm whether the pattern continues or if other factors are also influencing the best-fitting distribution.

It is possible to fit the models using either the EM algorithm or direct maximization, depending on preference since the two methods provided almost identical results. However, the direct numerical maximization was computationally quicker and easier to code. A further advantage of HMMs for animal movement studies is that the state allocation is a straightforward process and diagnostics such as the pseudo-residuals, Information Criteria and the autocorrelation function can all be used to investigate the fit of the model.

A useful output of the model fitting process is the predicted behavioural state. These locations and behaviour can then provide the input for further biological studies, for example to investigate the habitat used in the different states, how different species behave in the same region, a detailed space-time analysis and many others. The models provide insight into how the animals are shifting from one behavioural state to another and how the movements within a patch differ from those between patches. The HMMs were sensitive enough to pick up different movement patterns through the year, with both the hourly and daily analyses picking up longer movement states which were most common during the late dry period of the year when the animals are under the most stress. Since missing data can be included in the modelling process, it is not necessary to interpolate the observed data when missing observations occurred. This is a benefit of the HMM analysis and avoids an interpolation process that could introduce error into the study.

However, one of the disadvantages of HMMs (and independent mixture models) is that they tend to be over-interpreted (Zucchini, MacDonald 2009). As with any

clustering technique, there is uncertainty about the accuracy of the cluster allocation, as well as the interpretation of each cluster. As the number of states in the model is increased, the differentiation between the clusters is far less, and the interpretation of the states in terms of the behaviour of the animal can become uncertain. However, for the sable, buffalo and zebra data used in this study, the distinction between the states seems to be quite clear in terms of the behaviour.

Despite only applying a simple HMM structure during these analyses, the results are interpretable and make sense biologically. In many studies, the results obtained here would be sufficient for the purposes of the analysis and could feed into further ecological studies. However, an advantage of this modelling technique is that it can be extended for several individuals and include a time trend or dependence on some other covariate (MacDonald, Raubenheimer 1995). As shown by Langrock et al. (2012), HMMs can also be extended to introduce different dependence structures into the model, to include covariates or to include a hierarchical structure. The inclusion of covariates allows for the investigation of how animals are adapting their behaviour due to environmental conditions or other covariate. In Chapter 6, the analysis is extended to investigate various ways of building seasonality into the model, as well as a way to model observation data with varying time intervals between locations.

Moving forward, work is required to investigate the extension of HMMs to model group dynamics and the movement of an individual and the herd with which the animal is associated simultaneously (Langrock et al. 2013). Although very few instances of multiple collars within one herd are known, this is likely to become an area of interest for movement ecologists, particularly as the cost of collars is reducing. Work is also needed to obtain independent information on the true behavioural state of the animal which can be used to validate the state prediction from the models.

5.16 Appendix

Table 5.7 - Summary of Missing data for daily time series

	TotalObs	Missing	%Missing	Max_data_run
PM_Sable143_Daily_8pm	575	24	4.17	93
PM_Sable146_Daily_8pm	264	6	2.27	91
PM_Sable149_Daily_8pm	418	143	34.21	102
PM_Sable279_Daily_8pm	205	10	4.88	102
PK_Numbi_Daily_8pm	665	24	3.61	106
PK_Phabeni_Daily_8pm	1032	36	3.49	147
PK_Shitlave_Daily_8pm	859	21	2.44	166
PM_Buffalo150_Daily_8pm	292	45	15.41	52
PM_Buffalo152_Daily_8pm	1133	70	6.18	142
PM_Zebra141_Daily_8pm	494	14	2.83	99
PM_Zebra142_Daily_8pm	898	71	7.91	138
PM_Zebra277_Daily_8pm	1083	73	6.74	94
PM_Zebra280_Daily_8pm	448	203	45.31	69
PM_Sable143_Daily_2am	575	24	4.17	128
PM_Sable146_Daily_2am	263	5	1.90	91
PM_Sable149_Daily_2am	418	147	35.17	102
PM_Sable279_Daily_2am	205	12	5.85	86
PK_Numbi_Daily_2am	665	25	3.76	177
PK_Phabeni_Daily_2am	1032	46	4.46	211
PK_Shitlave_Daily_2am	859	22	2.56	149
PM_Buffalo150_Daily_2am	292	51	17.47	90
PM_Buffalo152_Daily_2am	1133	83	7.33	213
PM_Zebra141_Daily_2am	494	11	2.23	129
PM_Zebra142_Daily_2am	898	73	8.13	143
PM_Zebra277_Daily_2am	1084	121	11.16	79
PM_Zebra280_Daily_2am	448	207	46.21	30
PM_Sable143_Daily_8am	575	32	5.57	85
PM_Sable146_Daily_8am	263	4	1.52	134
PM_Sable149_Daily_8am	418	151	36.12	101
PM_Sable279_Daily_8am	205	14	6.83	67
PK_Numbi_Daily_8am	665	32	4.81	106
PK_Phabeni_Daily_8am	1032	48	4.65	178
PK_Shitlave_Daily_8am	859	27	3.14	198
PM_Buffalo150_Daily_8am	291	50	17.18	82

PM_Buffalo152_Daily_8am	1133	84	7.41	158
PM_Zebra141_Daily_8am	494	46	9.31	88
PM_Zebra142_Daily_8am	898	100	11.14	105
PM_Zebra277_Daily_8am	1084	83	7.66	133
PM_Zebra280_Daily_8am	447	213	47.65	33
PM_Sable143_Daily_2pm	575	62	10.78	57
PM_Sable146_Daily_2pm	263	11	4.18	56
PM_Sable149_Daily_2pm	417	146	35.01	116
PM_Sable279_Daily_2pm	205	18	8.78	50
PK_Numbi_Daily_2pm	665	38	5.71	109
PK_Phabeni_Daily_2pm	1032	63	6.10	131
PK_Shitlave_Daily_2pm	859	26	3.03	235
PM_Buffalo150_Daily_2pm	291	54	18.56	57
PM_Buffalo152_Daily_2pm	1133	104	9.18	89
PM_Zebra141_Daily_2pm	493	36	7.30	58
PM_Zebra142_Daily_2pm	898	92	10.24	160
PM_Zebra277_Daily_2pm	1084	169	15.59	109
PM_Zebra280_Daily_2pm	447	217	48.55	28

Table 5.8 - Daily HMMs - Comparison of fit: Log-normal, Weibull and Gamma state-dependent distributions

	LN_-LL	LN_AIC	LN_BIC	W_-LL	W_AIC	W_BIC	G_-LL	G_AIC	G_BIC
PM_Sable279_Daily_8pm_2-state	341.02	694.05	713.50	333.21	678.43	697.88	331.97	675.93	695.38
PM_Sable279_Daily_8pm_3-state	335.53	695.06	733.96	326.04	676.08	714.98	327.57	679.15	718.05
PK_Numbi_Daily_8pm_2-state	768.87	1549.74	1576.31	751.09	1514.17	1540.74	748.02	1508.04	1534.61
PK_Numbi_Daily_8pm_3-state	754.54	1533.08	1586.22	736.21	1496.43	1549.57	733.89	1491.78	1544.92
PK_Phabeni_Daily_8pm_2-state	664.96	1341.91	1371.12	646.75	1305.50	1334.71	645.74	1303.47	1332.68
PK_Phabeni_Daily_8pm_3-state	633.42	1290.84	1349.26	623.78	1271.56	1329.98	625.86	1275.71	1334.13
PK_Shitlave_Daily_8pm_2-state	1087.16	2186.31	2214.57	1060.74	2133.47	2161.73	1061.17	2134.35	2162.60
PK_Shitlave_Daily_8pm_3-state	1069.06	2162.12	2218.64	1049.38	2122.77	2179.28	1045.22	2114.44	2170.95
PM_Buffalo152_Daily_8pm_2-state	2028.76	4069.52	4099.10	2007.68	4027.37	4056.95	2007.16	4026.32	4055.90
PM_Buffalo152_Daily_8pm_3-state	2003.88	4031.76	4090.93	1992.86	4009.71	4068.88	2003.64	4031.28	4090.44
PM_Zebra141_Daily_8pm_2-state	587.91	1187.83	1212.69	585.62	1183.25	1208.11	584.43	1180.86	1205.73
PM_Zebra141_Daily_8pm_3-state	577.07	1178.14	1227.87	571.41	1166.83	1216.56	573.73	1171.47	1221.20
PM_Zebra142_Daily_8pm_2-state	1206.60	2425.19	2453.33	1203.21	2418.42	2446.56	1200.84	2413.69	2441.82
PM_Zebra142_Daily_8pm_3-state	1200.08	2424.15	2480.43	1192.86	2409.72	2465.99	1192.82	2409.63	2465.91
PM_Zebra277_Daily_8pm_2-state	1811.30	3634.61	3663.74	1822.20	3656.40	3685.53	1815.75	3643.50	3672.62
PM_Zebra277_Daily_8pm_3-state	1788.93	3601.87	3660.12	1777.80	3579.61	3637.86	1781.15	3586.30	3644.55
PM_Zebra280_Daily_8pm_2-state	328.10	668.20	688.67	321.54	655.07	675.54	320.57	653.14	673.61
PM_Zebra280_Daily_8pm_3-state	314.93	653.87	694.81	318.35	660.71	701.65	317.56	659.12	700.06
PM_Sable279_Daily_2am_2-state	312.23	636.46	655.75	315.15	642.30	661.59	311.78	635.57	654.86
PM_Sable279_Daily_2am_3-state	310.40	644.81	683.38	307.77	639.53	678.11	306.39	636.77	675.35
PK_Numbi_Daily_2am_2-state	781.36	1574.71	1601.26	760.65	1533.30	1559.85	759.45	1530.90	1557.45
PK_Numbi_Daily_2am_3-state	762.20	1548.40	1601.50	750.78	1525.57	1578.67	749.56	1523.13	1576.22

PK_Phabeni_Daily_2am_2-state	582.24	1176.49	1205.58	575.92	1163.84	1192.93	573.30	1158.59	1187.68
PK_Phabeni_Daily_2am_3-state	554.12	1132.25	1190.42	538.43	1100.87	1159.04	538.52	1101.04	1159.21
PK_Shitlave_Daily_2am_2-state	1110.42	2232.84	2261.06	1085.61	2183.22	2211.45	1087.55	2187.11	2215.33
PK_Shitlave_Daily_2am_3-state	1084.33	2192.66	2249.12	1081.10	2186.19	2242.65	1067.21	2158.41	2214.87
PM_Buffalo152_Daily_2am_2-state	1930.73	3873.46	3902.90	1918.93	3849.86	3879.30	1913.57	3839.14	3868.58
PM_Buffalo152_Daily_2am_3-state	1916.48	3856.95	3915.83	1905.32	3834.65	3893.53	1905.10	3834.20	3893.08
PM_Zebra141_Daily_2am_2-state	479.16	970.33	995.27	486.78	985.56	1010.50	473.13	958.26	983.21
PM_Zebra141_Daily_2am_3-state	462.88	949.77	999.65	452.61	929.21	979.10	457.56	939.13	989.01
PM_Zebra142_Daily_2am_2-state	1241.30	2494.60	2522.72	1225.40	2462.79	2490.91	1226.03	2464.07	2492.19
PM_Zebra142_Daily_2am_3-state	1228.15	2480.31	2536.55	1212.44	2448.88	2505.13	1212.80	2449.60	2505.84
PM_Zebra277_Daily_2am_2-state	1684.06	3380.13	3408.77	1670.65	3353.30	3381.94	1668.19	3348.38	3377.03
PM_Zebra277_Daily_2am_3-state	1664.77	3353.55	3410.84	1651.34	3326.69	3383.98	1654.47	3332.94	3390.23
PM_Zebra280_Daily_2am_2-state	312.89	637.78	658.06	309.84	631.68	651.96	309.64	631.27	651.55
PM_Zebra280_Daily_2am_3-state	309.42	642.85	683.41	305.08	634.16	674.71	305.60	635.20	675.76
PM_Sable279_Daily_8am_2-state	325.07	662.14	681.29	317.75	647.49	666.65	317.75	647.50	666.66
PM_Sable279_Daily_8am_3-state	319.48	662.96	701.28	312.68	649.35	687.67	314.78	653.56	691.87
PK_Numbi_Daily_8am_2-state	708.38	1428.76	1455.19	710.80	1433.60	1460.03	703.65	1419.30	1445.73
PK_Numbi_Daily_8am_3-state	702.26	1428.52	1481.38	698.46	1420.91	1473.77	699.15	1422.30	1475.16
PK_Phabeni_Daily_8am_2-state	665.30	1342.61	1371.69	639.07	1290.13	1319.21	690.93	1393.86	1422.94
PK_Phabeni_Daily_8am_3-state	640.78	1305.56	1363.73	614.13	1252.26	1310.42	619.47	1262.95	1321.11
PK_Shitlave_Daily_8am_2-state	1029.80	2071.59	2099.76	1001.78	2015.57	2043.74	999.17	2010.34	2038.51
PK_Shitlave_Daily_8am_3-state	998.89	2021.78	2078.13	988.51	2001.02	2057.37	994.62	2013.24	2069.59
PM_Buffalo152_Daily_8am_2-state	1933.26	3878.52	3907.93	1924.63	3861.26	3890.67	1919.68	3851.35	3880.76
PM_Buffalo152_Daily_8am_3-state	1918.82	3861.64	3920.46	1910.98	3845.95	3904.77	1911.65	3847.29	3906.11
PM_Zebra141_Daily_8am_2-state	644.05	1300.11	1324.29	638.27	1288.54	1312.72	637.00	1285.99	1310.18

PM_Zebra141_Daily_8am_3-state	632.76	1289.52	1337.89	627.32	1278.63	1327.00	629.18	1282.36	1330.73
PM_Zebra142_Daily_8am_2-state	1236.37	2484.74	2512.52	1224.45	2460.89	2488.67	1224.27	2460.55	2488.32
PM_Zebra142_Daily_8am_3-state	1218.36	2460.72	2516.27	1215.83	2455.67	2511.22	1216.06	2456.11	2511.67
PM_Zebra277_Daily_8am_2-state	1916.80	3845.60	3874.60	1902.82	3817.64	3846.64	1901.94	3815.88	3844.88
PM_Zebra277_Daily_8am_3-state	1893.20	3810.41	3868.42	1880.69	3785.37	3843.38	1881.07	3786.14	3844.15
PM_Zebra280_Daily_8am_2-state	347.15	706.29	726.29	342.16	696.33	716.32	343.20	698.39	718.39
PM_Zebra280_Daily_8am_3-state	343.69	711.39	751.38	342.04	708.07	748.06	337.73	699.46	739.46
PK_Numbi_Daily_2pm_2-state	738.81	1489.62	1515.94	738.75	1489.51	1515.83	730.47	1472.94	1499.26
PK_Numbi_Daily_2pm_3-state	721.16	1466.31	1518.96	719.62	1463.24	1515.88	716.20	1456.41	1509.05
PK_Phabeni_Daily_2pm_2-state	789.93	1591.86	1620.77	751.76	1515.53	1544.44	757.47	1526.95	1555.86
PK_Phabeni_Daily_2pm_3-state	756.76	1537.52	1595.34	740.54	1505.08	1562.89	746.36	1516.73	1574.54
PK_Shitlave_Daily_2pm_2-state	1092.02	2196.04	2224.21	1053.81	2119.62	2147.80	1055.89	2123.77	2151.95
PK_Shitlave_Daily_2pm_3-state	1066.50	2157.00	2213.35	1049.52	2123.04	2179.39	1049.37	2122.74	2179.09
PM_Sable279_Daily_2pm_2-state	350.17	712.35	731.27	338.45	688.90	707.82	340.40	692.80	711.72
PM_Sable279_Daily_2pm_3-state	339.61	703.22	741.06	330.65	685.30	723.14	334.58	693.16	731.00
PM_Buffalo152_Daily_2pm_2-state	1942.92	3897.84	3927.02	1894.71	3801.43	3830.61	1898.18	3808.37	3837.55
PM_Buffalo152_Daily_2pm_3-state	1895.80	3815.59	3873.96	1876.73	3777.47	3835.83	1888.02	3800.04	3858.41
PM_Zebra141_Daily_2pm_2-state	833.58	1679.17	1703.52	813.33	1638.67	1663.02	815.81	1643.61	1667.97
PM_Zebra141_Daily_2pm_3-state	812.39	1648.78	1697.49	798.34	1620.68	1669.39	811.66	1647.33	1696.04
PM_Zebra142_Daily_2pm_2-state	1334.56	2681.12	2708.95	1307.34	2626.69	2654.52	1311.99	2635.99	2663.82
PM_Zebra142_Daily_2pm_3-state	1311.65	2647.30	2702.96	1298.28	2620.56	2676.22	1303.26	2630.52	2686.18
PM_Zebra277_Daily_2pm_2-state	1713.71	3439.42	3467.53	1687.52	3387.04	3415.15	1688.00	3388.00	3416.11
PM_Zebra277_Daily_2pm_3-state	1684.86	3393.73	3449.94	1677.13	3378.26	3434.48	1678.32	3380.64	3436.86
PM_Zebra280_Daily_2pm_2-state	364.87	741.75	761.45	359.50	731.00	750.70	357.82	727.64	747.34
PM_Zebra280_Daily_2pm_3-state	350.82	725.65	765.05	349.99	723.98	763.38	355.13	734.26	773.66

6 HIDDEN MARKOV MODEL EXTENSIONS FOR ANIMAL MOVEMENT MODELLING

6.1 Introduction

GPS tracking technology is now used more commonly in ecological studies of animal movements. Tracking devices can be placed on a variety of species, even with very small species being tracked due to a reduction in size of the device. The battery life of the units has now also been extended considerably which means that data can be collected at more frequent intervals over a longer period of time. Although this is an excellent advancement and means that far more data are now available for analysis, it also adds to the statistical complications of the analysis since there is strong autocorrelation within the data as well as daily and/or seasonal cycles. More questions are now being asked of the tracking data as studies develop.

Hidden Markov models are an analysis approach which seems to be used increasingly in the studies of animal tracking behaviour, with inferences made about the behaviour of the animal given the latent state. However many of these studies only use a short period of tracking data (Franke, Caelli & Hudson 2004, Franke et al. 2006, Roberts et al. 2004) or intentionally select a portion of the data so that seasonality is not an issue within the analysis (Langrock et al. 2012). However this means that a large part of the value of long-term remotely sensed tracking data is not used. In the previous chapter, it was shown that Hidden Markov models (HMMs) applied to the sable, buffalo and zebra data spanning a number of seasons still provide realistic results and that the interpretation of the behaviour compared with known behaviour is consistent. It is encouraging that even the simple application of these models ignoring seasonality provides realistic results and was able to pick up different behavioural patterns through the seasons.

However an important feature of HMMs which make them useful for the analysis of ecological data is that covariates can be included in the models. From the literature, it seems that few movement studies have in fact built the seasonality into the HMM. In this chapter, four different techniques are used to introduce seasonality formally via the model specification. Two different methods are used to account for seasonality within the model, and these are applied separately to the state-dependent distributions and to the transition probability matrix.

Another challenge with the analysis of animal tracking data is the presence of missing data. As discussed previously, a missing data point can be obtained when the collar failed to connect to sufficient satellites in order to obtain a location reading or due to changes in the observation frequency. This results in two missed input displacement observations since the displacements cannot be calculated. In other studies, the common way to handle this is to interpolate the expected location and proceed with the analysis as normal, or to rather use the discrete displacement over variable time converted to a movement rate rather than displacement as the input measure (Morales et al. 2004). This means that the input data is an average speed over variable time periods rather than the distance moved over a regular time gap. Both these methods are approximations and can introduce additional error and uncertainty into the model. In some studies, variable time intervals are actually programmed into the collars in order to prolong the battery life. This is normally done by only obtaining regular observations during the part of the day or night that the animal is most active and fewer observations when the animal is sleeping (Tambling et al. 2010, Tambling et al. 2011). In the previous chapter, the missing data were retained as missing for model fitting and the maximum likelihood estimation was done with the missing data included in the input. It is an advantage of the HMMs that missing input data can be used in the likelihood computation using either the EM algorithm or direct numerical maximization. However, the structure of the likelihood function of the HMM makes it possible to develop a method which uses all of the locations that were obtained and the number of time steps that occurred between successive

observations that were correctly obtained. This means that when a location is missed, there are not two missed input displacements in the model, but rather the total distance moved and number of time units between obtained locations are used. This means that all of the obtained locations can actually be used in the model fitting process and studies designed with variable time steps can use HMMs as an analysis technique making use of all their data.

Many of the animal movement studies in the literature include the turning angle between successive locations as an input for their modelling. This has been done for models using both HMMs (Franke, Caelli & Hudson 2004, Franke et al. 2006, Langrock et al. 2012) and Bayesian State-space models (Morales et al. 2004, McClintock et al. 2012). Throughout this thesis, the turning angle has not been included in the analyses of the sable antelope, buffalo and zebra data. However since it is a metric commonly used in the analysis of animal movement data, its effect on the modelling needs to be investigated.

6.2 Accounting for Seasonality

Two different methods were used to formally include seasonality in the modelling process, with each method applied using two different approaches. For the first method, the seasons are divided into three blocks using the month of the year in the same way that the seasonal models were run in Section 3.8.6. This method will be called the seasonal-split model. The models are fitted in one of two ways, firstly using one transition probability matrix for all data but fitting separate state-dependent distribution parameters for each season. Alternatively, the models are fitted with a different transition probability matrix for each seasonal block and the state-dependent distribution parameters are consistent for the whole dataset. Both the transition probability matrix and the state-dependent distribution parameters could be varied simultaneously, but this would in effect be fitting separate HMMs

to the seasonal blocks, with a missing data point to indicate the split between seasons in separate years.

A different method for building seasonality into the models is to include a continuous time-dependent covariate. The covariate can be included either in the state-dependent distributions or the transition probability matrix or both. For the purposes of this study, a 'day of the year' covariate was included in one or the other parameter set but not both at the same time. Applying them to both, while being mathematically possible is likely to encounter difficulties converging and the interpretation of the states and transitions will be complex. Other covariates such as the prior rainfall or temperature could also be used for this type of application as it will have an impact on the conditions experienced by the animal.

Basic tests to formally compare the behaviour from one season to another are done, as a follow-on from the methods applied using the Independent Mixture Models.

6.2.1 Seasonal-Slit Models

All models and results discussed in the previous chapter have been calculated on the complete dataset which spans over a year in most cases meaning that a number of seasonal cycles are represented within the data. It is known that animal behaviour varies over the seasonal cycle, not only due to the changing climatic conditions and associated water availability and veld condition, but also the number of daylight hours during the day with changing sunrise and sunset times. In the independent mixture model section, separate models were run for seasonal units and compared to the overall model fitted to the full dataset. In the time dependent HMM analysis, we cannot simply split the data into seasonal datasets and re-run the models, due to the continuous time series nature of these input data. Instead modifications can be made to the HMM structure to allow for seasonal

models. A likelihood ratio test can be used to determine whether there is a significant improvement in the fit of the model with the inclusion of the covariate(s), or the model fit can be compared using AIC and BIC.

6.2.1.1 Methods

For the Kruger movement study, the year is split into three categories corresponding to the Early Dry, Late Dry and Wet season based on the months of the year, see section 2.5.1 for more details. This is the same as the seasonal split used for the mixture model analysis. This means that the full dataset is essentially split into three subsets depending on the season.

State-dependent distribution: the parameters for the state-dependent distributions are assumed to differ in each of the three seasons while the transition probability matrix remains constant across the year. In this study, the distributions for each season were from the same family (log-normal); however there is no mathematical reason why different distributions could not be used for each season if this made ecological sense. Each parameter is replaced by a vector of length three, with each element representing that model parameter during one of the seasons. For a log-normal HMM, each state-dependent distribution has six parameters, three mean and three variance parameters, with each pair representing the distribution fitted to one of the three seasonal blocks. Non-stationary models were fitted using direct numerical maximization to obtain the vector parameter estimates for the seasonal split models. The likelihood function is defined as:

$$L_T = \delta P_{j_{(1)}}(x_1)\Gamma P_{j_{(2)}}(x_2)\Gamma P_{j_{(3)}}(x_3).....\Gamma P_{j_{(T)}}(x_T)\mathbf{1}'$$

$$\text{where } j(t) = \begin{cases} 1 & \text{if } t \text{ in April, May, June, July} \\ 2 & \text{if } t \text{ in August, September, October, November} \\ 3 & \text{if } t \text{ in December, January, February, March} \end{cases}$$

The basis of this type of model specification is that the probability of a species switching behaviour remains consistent throughout the year, however at different times of the year the distribution of each behavioural state will be different. It is expected that the states associated with the Late Dry season will have longer expected values compared with the Wet season's shorter movements.

Transition probability matrix: The second approach was to have one set of state-dependent distribution parameters for the full dataset but a different probability matrix for each season. The likelihood function for this model definition is given by:

$$L_T = \delta P(x_1)\Gamma_{k(2)}P(x_2)\Gamma_{k(3)}P(x_3).....\Gamma_{k(T)}P(x_T)\mathbf{1}'$$

$$\text{where } k(t) = \begin{cases} 1 & \text{if } t \text{ in April, May, June, July} \\ 2 & \text{if } t \text{ in August, September, October, November} \\ 3 & \text{if } t \text{ in December, January, February, March} \end{cases}$$

In this case, the assumption is that the movement behaviours remain consistent throughout the year while the probabilities of moving from one state to another depend on the time of the year. It is expected that the probability of transitioning into the more active movement states would be higher during the Late Dry season when the animals are forced to search more for food and water.

6.2.1.2 Results

State-dependent distribution

The method is illustrated using the hourly data with a log-normal distribution, although it could easily be applied to the daily movement data and with the Gamma, Weibull or other distributions. The seasonal model specified in this way was more time consuming to obtain convergence than the basic model fitted to the annual data which ignores the seasonality. This is due to the increased number of parameters that have to be estimated. It is possible that the way in which the

model is coded could be improved in order to speed up the fitting process. Although it is more time consuming to run the seasonal-split models, this is certainly not a constraint to using this type of seasonal model specification and even relatively small desktop and laptop computers have sufficient computing power to fit the models. The maximization was run using the unconstrained optimizer `nlm` in R (R Development Core Team 2011), with transformations before and after optimization to ensure that valid parameter estimates were obtained. The transformations used are detailed in Section 5.4.1. The maximum likelihood and model selection criteria for some of the animals fitted using seasonal-split models with dependent state distributions are shown in Table 6.1.

Table 6.1 - Model fit of seasonal-split HMMs with state-dependent covariate (non-stationary)

		-LogLik	AIC	BIC
PK_Phabeni	2-state	-19994.26	-39956.52	-39836.22
	3-state	-20126.71	-40199.41	-39996.41
	4-state	-20189.44	-40298.89	-39998.14
	Annual	-29656.59	-59267.18	-59094.25
PM_Sable279	2-state	-1572.642	-3113.284	-3015.091
	3-state	-1672.641	-3291.283	-3125.581
	4-state	DNC		
	Annual	-1708.932	-3377.865	-3255.123
PK_Numbi	2-state	-6118.79	-12205.58	-12099.56
	3-state	-6208.784	-12363.57	-12184.67
	4-state	-6227.601	-12375.2	-12110.16
	Annual	-6186.07	-12332.14	-12199.61
PM_Buffalo152	2-state	-6733.936	-13435.87	-13314.85
	3-state	-7324.648	-14595.3	-14391.07
	4-state	-7754.053	-15428.11	-15125.55
	Annual	-7891.806	-15737.61	-15563.64

For the Phabeni herd at Pretoriuskop, a 4-state seasonal model was selected as the best model by AIC and BIC. Although the 4-state model was able to converge for the Phabeni herd, with some of the other herds (Punda Maria Sable279), the 4-state models had some difficulty converging to a global maximum. This is not

unexpected due to the increased number of parameters to be estimated. For studies with smaller amounts of data, the application of the seasonal-split models might have difficulty being able to converge due to fewer data points in the seasonal blocks. The estimated model parameters for the Numbi sable and Punda Maria sable and buffalo herds are shown in Table 6.2. Parameter estimates across the seasonal components are similar, with the expected movements for the resting and foraging states very similar across the seasons. The expected movement distance for the Late Dry season is longer for the movement and relocating states, confirming that the animals have to travel greater distances towards the end of the dry season to find food and water. This will put the individuals under a lot more pressure. The estimated model parameters and expected movement distance during the Early Dry and the Wet season are more similar to one another. For the annual HMMs with four states, the fourth relocating state has a distribution with a much longer expected value compared with the other states, and it is consistent with the seasonal-split model's movement state for the Late Dry season. For the Punda Maria sable herd, only 10 observations outside the Late Dry season were in fact allocated to the relocating state using the annual model fitted to all data. This pattern is not as evident for the state allocation of the Punda Maria Buffalo where the fourth state is regularly visited, however, the expected movement distance of the fourth state is half that of the Sable in the same region. The state allocation by season for these animals is shown in Table 6.3. Given that the fourth state is quite different for these two species, it is not surprising that the buffalo enter their 'relocating' state more often than the sable whose relocating state is more extreme and is reserved for very long movements. The buffalo behaviour inferred from the model is less seasonal and could be explained by the fact that the buffalo herd is not making the long journeys to water seen for the sable in the Late Dry season (Cain, Owen-Smith & Macandza 2012, Macandza, Owen-Smith & Cain 2012b).

Table 6.2 – Parameter estimates and distribution moments for seasonal-split models (state-dependent distribution)

	PK_Numbi				PM_Sable279				PM_Buffalo152			
	Early Dry	**Late Dry**	**Wet**	**Annual**	**Early Dry**	**Late Dry**	**Wet**	**Annual**	**Early Dry**	**Late Dry**	**Wet**	**Annual**
μ1	-4.63	-4.30	-4.02	-3.93	-3.50	-3.27	-3.61	-3.71	-4.17	-3.85	-3.96	-4.05
μ2	-3.04	-2.83	-3.62	-2.28	-1.33	-0.81	-1.32	-1.68	-1.96	-1.66	-1.83	-1.95
μ3	-1.82	-1.63	-2.13	-1.98	-0.04	0.61	-0.16	-0.40	-1.00	-0.81	-0.98	-1.01
μ4	-0.64	-0.47	-0.86	-0.61				0.69	-0.55	-0.16	-0.58	0.00
σ1	0.75	0.91	1.16	1.05	1.24	1.28	1.26	1.13	1.10	1.17	1.22	1.15
σ2	0.87	0.87	0.92	0.85	0.67	0.70	0.88	0.67	0.69	0.69	0.70	0.81
σ3	0.62	0.72	0.62	0.75	0.25	0.37	0.47	0.50	0.64	0.60	0.71	0.57
σ4	0.56	0.52	0.66	0.58				0.32	0.57	0.61	0.59	0.48
E(X1)	0.01	0.02	0.04	0.03	0.07	0.09	0.06	0.05	0.03	0.04	0.04	0.03
E(X2)	0.07	0.09	0.04	0.15	0.33	0.57	0.39	0.23	0.18	0.24	0.21	0.20
E(X3)	0.20	0.25	0.14	0.18	0.99	1.97	0.96	0.76	0.45	0.53	0.48	0.43
E(X4)	0.62	0.72	0.52	0.65				2.10	0.68	1.03	0.67	1.12
Var(X1)	0.00	0.00	0.00	0.00	0.02	0.03	0.01	0.01	0.00	0.01	0.01	0.00
Var(X2)	0.01	0.01	0.00	0.02	0.06	0.21	0.18	0.03	0.02	0.04	0.03	0.04
Var(X3)	0.02	0.04	0.01	0.02	0.06	0.57	0.23	0.16	0.10	0.12	0.15	0.07
Var(X4)	0.14	0.16	0.15	0.17				0.46	0.18	0.49	0.19	0.33

Table 6.3 - State allocation frequencies by season for Punda Maria Buffalo and Sable based on an annual 4-state HMM

		Early Dry	Late Dry	Wet	Expected Distance (km)
Punda Maria Buffalo	**State 1**	5558	5736	5608	0.03
	State 2	1296	1106	1376	0.20
	State 3	914	870	945	0.43
	State 4	833	1072	807	1.12
Punda Maria Sable	**State 1**	909	1583	653	0.05
	State 2	86	640	189	0.23
	State 3	36	510	107	0.76
	State 4	0	195	10	2.1

Transition Probability Matrix

The results for the seasonal-split transition probability models are shown in Table 6.4 and Table 6.5. In all cases the state-dependent distribution parameters and the expected values of the distributions defined by them are consistent with the models run for the annual data. The probabilities of remaining in the less active resting and foraging states are mostly consistent throughout the year. This is exactly as one would expect the animals to move based on the requirements of the conditions. 4-state models for the Punda Maria and Phabeni sable herds failed to converge. This failure to converge is likely a result of the fewer data points in the seasonal blocks available to fit the model. This appears to be the reason for the Punda Maria sable herd which did support a 4-state model for the annual data. The Phabeni herd only required a 3-state model for the annual data, so it is not surprising that a 4-state model was not possible to fit for the smaller seasonal blocks of data. For the Punda Maria Buffalo herd, the expected movement distances for the four states are almost identical to those obtained for the 4-state annual model. However the transition probabilities differ for the three seasonal blocks.

The probability of remaining within the relocating state is highest in the Late Dry season (0.50) compared to the Wet and Early Dry season (0.41 and 0.40

respectively). This pattern was repeated for all of the herds, with the probability of remaining in the most active (and hence most stressful state) being highest during the Late Dry season. Although the log-likelihood, AIC and BIC values indicated that the seasonal split model did not improve the overall fit of the model compared with the annual model, they were still able to provide insight into how the patterns of behaviour can change through the season.

Table 6.4 - Model fit of seasonal-split HMMs using transition probability matrix with 'best' seasonal-split model highlighted

		-LogLik	AIC	BIC
PK_Phabeni	2-state	-19907.51	-39791.02	-39700.8
	3-state	-20071.24	-40088.48	-39885.48
	4-state	DNC		
	Annual	-29656.59	-59267.18	-59094.25
PM_Sable279	2-state	-1548.418	-3072.836	-2999.191
	3-state	-1657.373	-3260.746	-3095.044
	4-state	DNC		
	Annual	-1708.932	-3377.865	-3255.123
PK_Numbi	2-state	-6107.227	-12190.45	-12110.94
	3-state	-6197.123	-12340.25	-12161.34
	4-state	-6205.891	-12315.78	-11997.73
	Annual	-6253.106	-12466.21	-12333.69
PM_Buffalo152	2-state	-6656.63	-13289.27	-13198.5
	3-state	-7252.544	-14451.09	-14246.86
	4-state	-7578.567	-15061.13	-14698.06
	Annual	-7891.806	-15737.61	-15563.64

6.2.1.3 Summary

It is clear that this method, applied in either way, provides very realistic results which are easily interpretable in terms of the latent states and the seasonal transition probabilities. The seasonal split method appears crude and segments the year in a way which does not take into account the true conditions using only an artificial segmentation based on the calendar months. However the results are meaningful and provide additional information about the movement patterns of the animals. HMMs split in this way are relatively easy to code, needing only

minor modifications from the basic HMM model and converge reasonably quickly.

Table 6.5 – Parameter estimates and distribution moments for Seasonal-split models (transition probabilities)

PM_Buffalo152			
E(X) - State 1	**E(X) - State 2**	**E(X) - State 3**	**E(X) - State 4**
0.03	0.18	0.43	0.99
Early Dry			
0.68	0.08	0.19	0.05
0.45	0.40	0.04	0.11
0.03	0.49	0.43	0.05
0.11	0.03	0.45	0.40
Late Dry			
0.70	0.09	0.10	0.11
0.35	0.51	0.06	0.08
0.08	0.35	0.51	0.06
0.07	0.08	0.34	0.50
Wet			
0.64	0.11	0.17	0.08
0.43	0.40	0.07	0.11
0.05	0.45	0.42	0.08
0.11	0.05	0.44	0.41

6.2.2 Continuous Time Covariate

6.2.2.1 Methods

State-dependent distribution – For a model including a continuously varying covariate, it can be assumed that the state-dependent probabilities are influenced by the covariate which is shown in graphical model form in Figure 6.1. The Markov chain is given by $\{C_t\}$, with X_t the observation and Y_t the covariate influencing the observation at time t. For example, a Gamma HMM with parameters α and θ for the shape and scale parameters respectively which have to be positive, the parameters depend on the row vector $\mathbf{y}_t$ of q covariates. For example:

$$\log {}_t\alpha_i = \beta_i \mathbf{y}'_t$$

where ${}_t\alpha_i$ indicates that the parameter α_i varies with time. The vector $\mathbf{y}_t$ can include a constant, sinusoidal components (for seasonality) and/or other relevant covariates (Zucchini, MacDonald 2009). For example, if a seasonal cycle only is included in the model, the parameters will be defined as follows, with r denoting the period of the cycle. In the case of an annual cycle the period will be 365.25 days with t starting on the day of the year of the first observation:

$$\log {}_t\alpha_i = \beta_{i1} + \beta_{i2}\cos\left(\frac{2\pi t}{r}\right) + \beta_{i3}\sin\left(\frac{2\pi t}{r}\right)$$

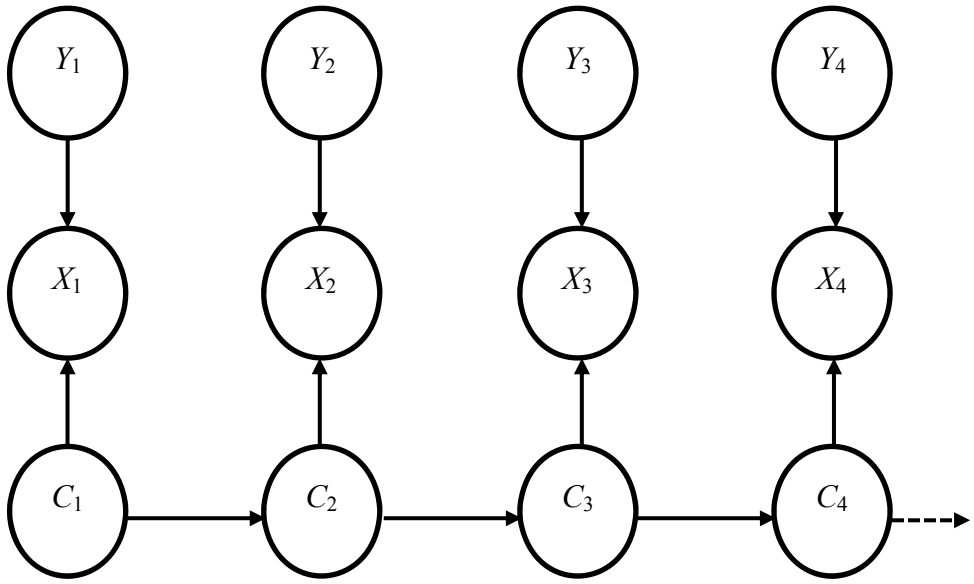

Figure 6.1 - HMM structure with covariates influencing the state-dependent probabilities (Zucchini & MacDonald, 2009)

This method will be illustrated using the daily movement data for simplicity. However if the model was fitted to hourly data, additional parameters could be included to account for a 24 hour cycle as well. The likelihood function for a model including covariates is similar to that of the standard case

$$L_T = \boldsymbol{\delta}_1 \mathbf{P}(x_1,y_1)\boldsymbol{\Gamma}_2 \mathbf{P}(x_2,y_2)\boldsymbol{\Gamma}_3 \mathbf{P}(x_3,y_3)\ldots\boldsymbol{\Gamma}_T \mathbf{P}(x_T,y_T)\mathbf{1}'$$

where ${}_t\mathbf{P}(x_t,y_t) = \text{diag}({}_tp_1(x_t,y_t), {}_tp_2(x_t,y_t),\ldots, {}_tp_m(x_t,y_t))$ which allows for the covariates in the state-dependent probability distributions that now vary with time. This is the likelihood definition for a non-stationary HMM since if X_t contains seasonal variation, the Markov chain cannot be stationary and the initial distribution cannot be derived from the transition probability matrix. With the fitting of any of these models, any state-dependent parameter constraints must be met. This is achieved in the same way as with the basic models, through the parameter definitions and fitting transformed variables which incorporate the constraints or using a constrained optimizer.

Transition probabilities – A somewhat more complicated way of including the covariates within the model is to assume that the Markov chain is not homogeneous, and instead assume that the transition probabilities are functions of time (Zucchini, MacDonald 2009). These transitions can depend on one or more covariates (Zucchini, MacDonald 2009). Once again, the Markov chain is assumed to be non-stationary. This method for including covariates in the model is more complex and computationally intensive compared with the state-dependent distributions depending on the covariates. The covariates are included in the transition matrix using a multinomial logit transformation to ensure that the parameters range between zero and one. For a 2-state model, the transition probability matrix is defined as:

$${}_t\Gamma = \begin{pmatrix} {}_t\gamma_{11} & {}_t\gamma_{12} \\ {}_t\gamma_{21} & {}_t\gamma_{22} \end{pmatrix}$$

The covariates included can include time and seasonality, or any other covariate that appears relevant. The components of the Markov chain are then given as:

$$P(C_t = 2 \mid C_{t-1} = 1) = {}_t\gamma_{12} \quad \text{and} \quad P(C_t = 1 \mid C_{t-1} = 2) = {}_t\gamma_{21}$$
$$\text{for } i = 1,2$$
$$\text{logit}\, {}_t\gamma_{i,3-i} = \beta_i y'_t$$

Once again for the *r*-period seasonality:

$$\text{logit}_t \gamma_{i,3-i} = \beta_{i1} + \beta_{i2}\cos\left(\frac{2\pi t}{r}\right) + \beta_{i3}\sin\left(\frac{2\pi t}{r}\right).$$

The transition probability matrix at time *t* is then given by:

$${}_t\Gamma = \begin{pmatrix} \dfrac{1}{1+e^{\boldsymbol{\beta}_1 \mathbf{y}_t'}} & \dfrac{e^{\boldsymbol{\beta}_1 \mathbf{y}_t'}}{1+e^{\boldsymbol{\beta}_1 \mathbf{y}_t'}} \\ \dfrac{e^{\boldsymbol{\beta}_2 \mathbf{y}_t'}}{1+e^{\boldsymbol{\beta}_2 \mathbf{y}_t'}} & \dfrac{1}{1+e^{\boldsymbol{\beta}_2 \mathbf{y}_t'}} \end{pmatrix}$$

This model can be extended for the case where $m > 2$, but this becomes much more computationally intensive (Zucchini, MacDonald 2009). The transition probability matrix at time *t* for a 3-state HMM where $\text{logit}_t \gamma_{i,j} = \tau_{ij} y'_t$; for $i \neq j$ is given by:

$${}_t\Gamma = \begin{pmatrix} \dfrac{1}{1+e^{\tau_{12}\mathbf{y}_t'}+e^{\tau_{13}\mathbf{y}_t'}} & \dfrac{e^{\tau_{12}\mathbf{y}_t'}}{1+e^{\tau_{12}\mathbf{y}_t'}+e^{\tau_{13}\mathbf{y}_t'}} & \dfrac{e^{\tau_{13}\mathbf{y}_t'}}{1+e^{\tau_{12}\mathbf{y}_t'}+e^{\tau_{13}\mathbf{y}_t'}} \\ \dfrac{e^{\tau_{21}\mathbf{y}_t'}}{1+e^{\tau_{21}\mathbf{y}_t'}+e^{\tau_{23}\mathbf{y}_t'}} & \dfrac{1}{1+e^{\tau_{21}\mathbf{y}_t'}+e^{\tau_{23}\mathbf{y}_t'}} & \dfrac{e^{\tau_{23}\mathbf{y}_t'}}{1+e^{\tau_{21}\mathbf{y}_t'}+e^{\tau_{23}\mathbf{y}_t'}} \\ \dfrac{e^{\tau_{31}\mathbf{y}_t'}}{1+e^{\tau_{31}\mathbf{y}_t'}+e^{\tau_{32}\mathbf{y}_t'}} & \dfrac{e^{\tau_{32}\mathbf{y}_t'}}{1+e^{\tau_{31}\mathbf{y}_t'}+e^{\tau_{32}\mathbf{y}_t'}} & \dfrac{1}{1+e^{\tau_{31}\mathbf{y}_t'}+e^{\tau_{32}\mathbf{y}_t'}} \end{pmatrix}$$

For each increase in the number of states, the number of parameters that need to be estimated increases rapidly which makes the likelihood surface more complicated and can lead to model fitting problems. As discussed above, it is possible to include the day of the year as a seasonal covariate in both parameter sets concurrently; however obtaining convergence and interpreting the results will be even more complicated. For these reasons it was felt that this application was beyond the scope of this study.

6.2.2.2 Results

Seasonal models using a continuously time-varying covariate were only applied to the daily data, including an annual cycle into the model. The covariate in this case is the day of the year. However, similar models could be run using the hourly data which would then need to include a daily cycle. This would mean that both the time of day and the day of the year would be included in the model as covariates. The daily cycle would be included in exactly the same way as the annual cycle, just with additional sine and cosine terms for the time of day. While this method would not introduce much more coding complexity, it would be more computationally intensive due to the large volume of data available and the additional terms would make the interpretation of the parameters more challenging. The likelihood surface will also be more complicated and convergence to the global maximum would need to be checked and issues could be found with actually getting the model to converge satisfactorily. For these reasons, it was felt that for the purposes of this study, the method could be suitably presented using the illustration of the daily data. Future work, beyond the scope of this thesis, is needed to investigate the inclusion of a daily cycle and how the model parameters can be interpreted in terms of the behaviour of the animal. Similar to the daily models fitted using the basic HMM, Gamma distributions were used to describe the latent states. Only 2-state models were fitted to these data. Once again, the additional complexity of adding additional states is mathematically possible, however the fitting process would be trickier with the amount of daily data available to find a global maximum. Interpretation of the models would also be potentially harder. For the purposes of this study, the simplest model is considered and illustrated.

State-dependent distribution

Covariate models were fitted allowing both of the parameters of the state-dependent distribution (Gamma distribution) to vary with the time of the year. The parameters themselves are difficult to interpret, however the plot of both the shape and scale parameters are shown in Figure 6.2 for the sable herd at Numbi.

(Note that in these figures the date is displayed in the format 2007.0 representing January 2007 and 2007.5 as June 2007 and so on). There is a clear seasonal pattern with the red line (State 1) being well above the black line (State 2) for most of the year except briefly during the time associated with the Wet season. The scale parameters are also quite distinct for the two states, and are furthest apart during the Late Dry season. As is, the parameters are quite difficult to interpret, but a plot of the expected movement distance for the distribution is shown in Figure 6.3. There are clear "seasons" to the data. State 1 is associated with the longer movements which will be associated with shifts between patches. The peak in the expected movement distance occurs around the end of the Late Dry season which is consistent with the time when the animals are under the most stress during the commonly driest months in this region. The expected movement for this state ranges from between 1.3km and 2.3km per day. State 2 is associated with the much shorter movements from day to day which will correspond to the animal remaining within a patch. These movements range between 0.5km and 0.75km per day. This analysis illustrates that a continuously varying covariate can be included via the state dependent probabilities mathematically and that the results are in fact consistent with the expected behaviour for the animal. Fitting a seasonal covariate to the state-dependent distribution parameters is computationally easier than via the transition probabilities, but in some cases might result in latent states which are not as easily interpretable in terms of the behaviour of the animal.

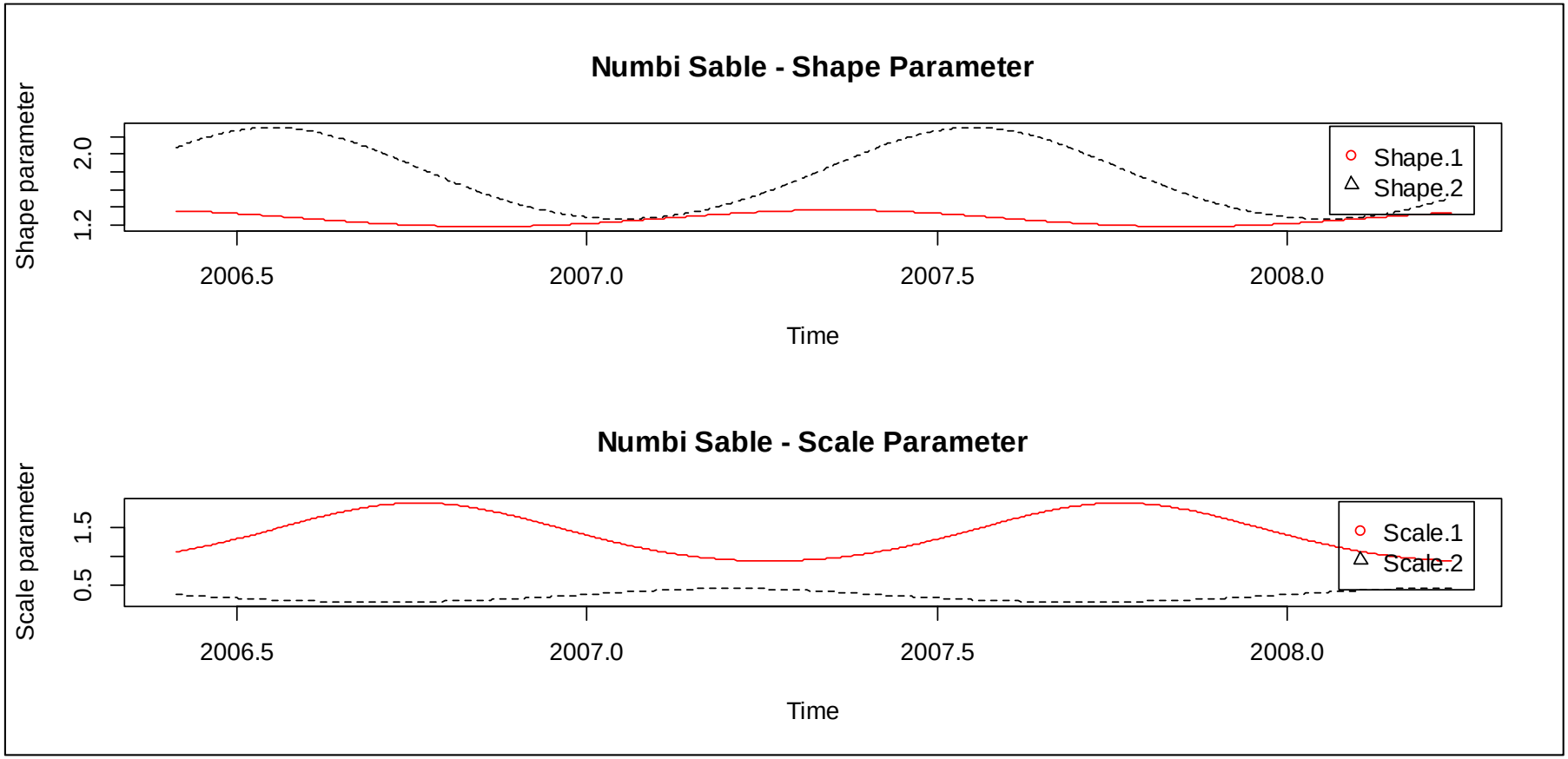

Figure 6.2 - Continuous time-varying seasonal covariate - state-dependent distribution parameters - Pretoriuskop Numbi sable herd

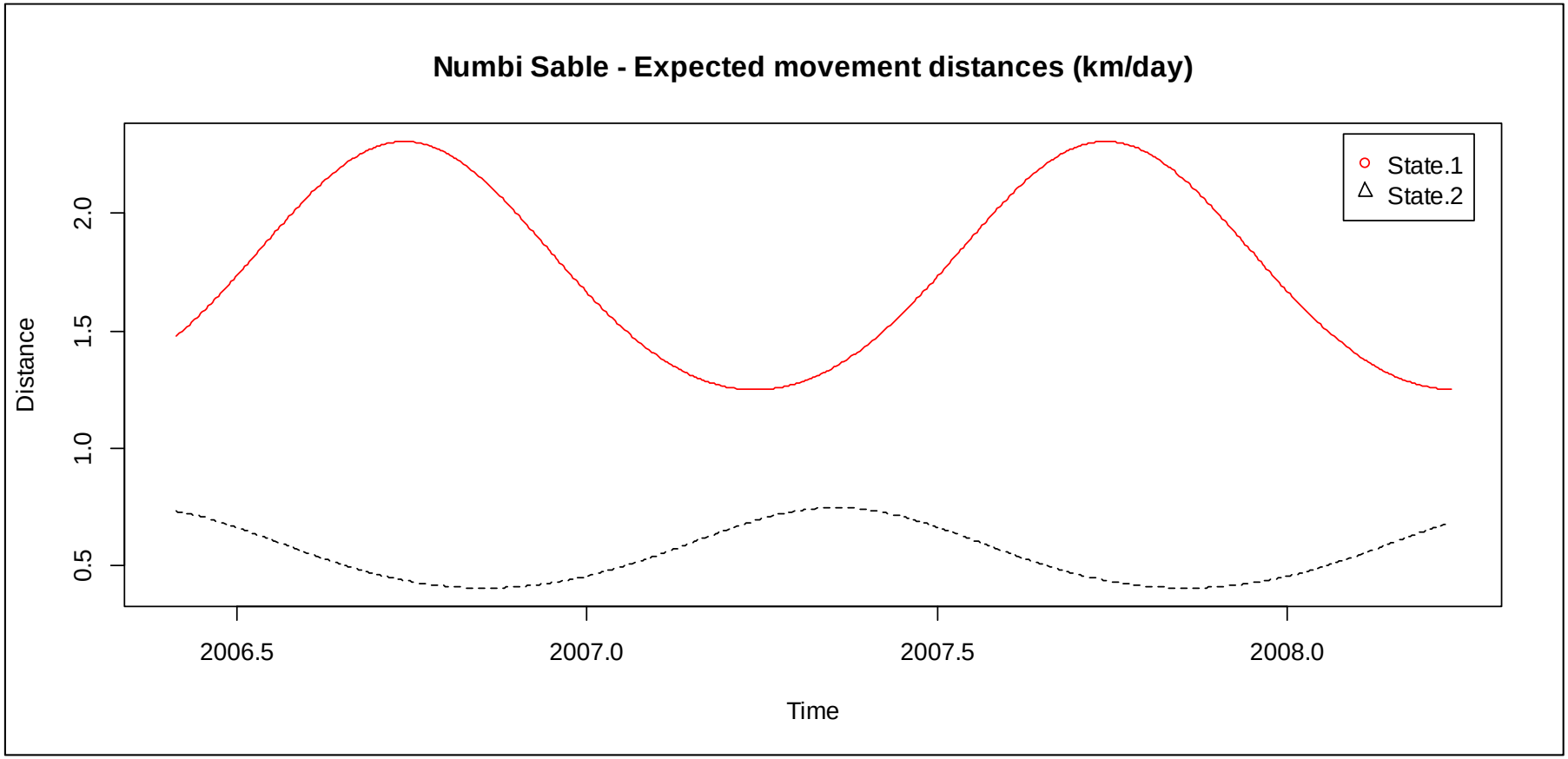

Figure 6.3 - Pretoriuskop Numbi Sable - Expected Movement Distances with continuous time varying state-dependent distributions

Table 6.6 - Fitted model parameters for time-varying seasonal covariate models (transition probability matrix)

Species	Herd	Time	$\alpha 1$	$\beta 1$	$\alpha 2$	$\beta 2$	a1	b1	c1	a2	b2	c2	E(X): State 1	E(X): State 2
Sable	Numbi	2am	1.89	0.29	1.29	1.52	-2.53	-0.34	2.06	-2.83	-0.22	2.61	0.55	1.96
Sable	Phabeni	2am	1.47	0.29	2.40	0.57	-1.97	0.07	-0.33	-0.59	-0.46	1.24	0.43	1.37
Sable	Shitlave	2am	1.50	0.62	1.31	1.53	-2.92	0.91	-0.51	-3.35	1.00	2.88	0.93	2.00
Sable	Punda Maria 143	2am	1.25	0.67	1.79	1.53	-3.08	1.27	-0.88	-2.09	0.70	2.15	0.84	2.74
Sable	Punda Maria 279	2am	Did not converge											
Buffalo	Punda Maria 152	2am	2.23	1.04	3.43	1.75	-2.95	-0.23	-1.04	-0.34	-0.19	0.51	2.32	5.99
Zebra	Punda Maria 277	2am	1.43	1.00	2.29	2.08	-2.01	0.60	-0.24	-1.35	0.14	0.17	1.00	4.35

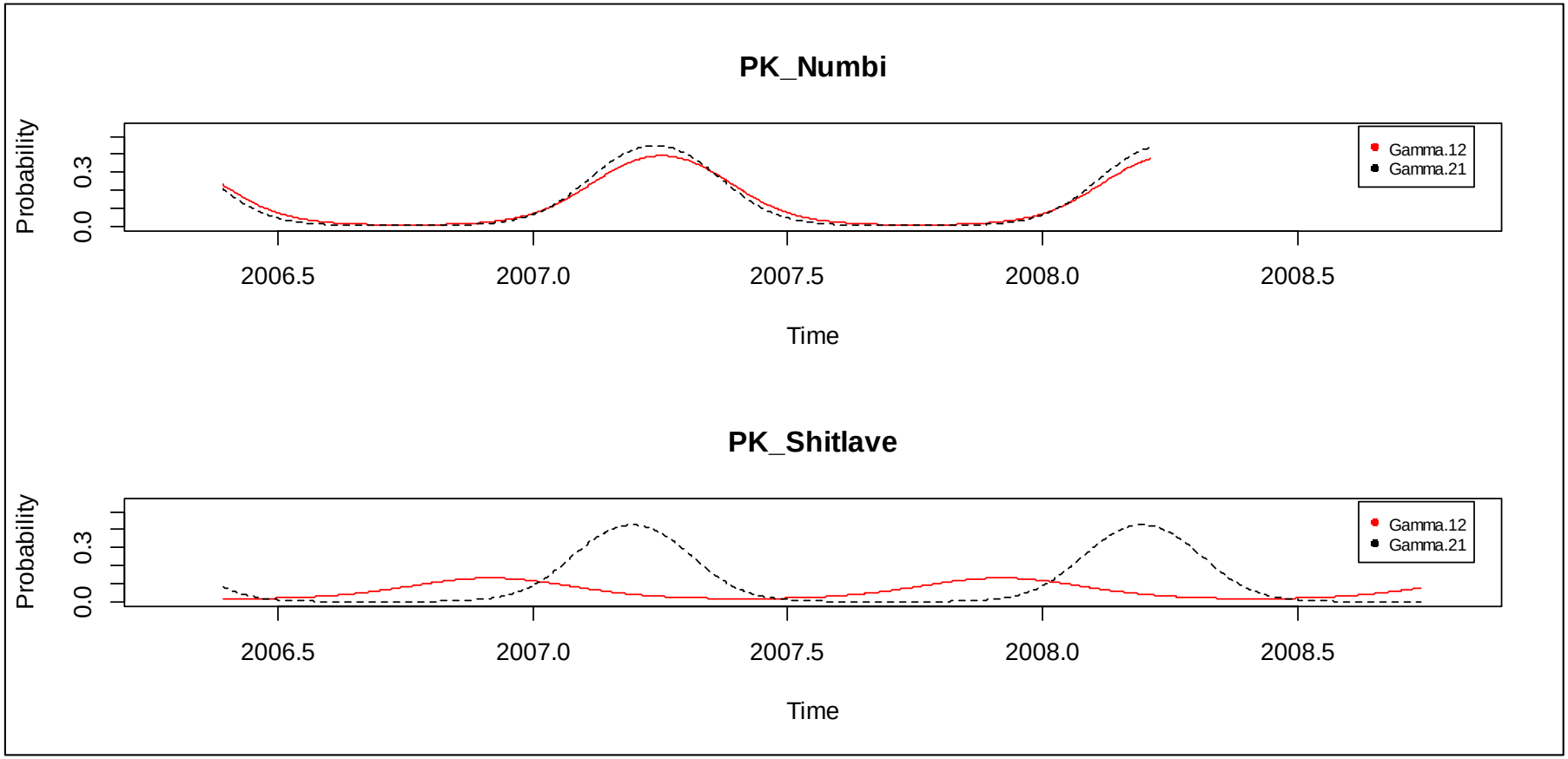

Figure 6.4 - Transition probabilities with annual trend for Pretoriuskop herds (Numbi and Shitlave)

Transition Probability Matrix

Once again, only a 2-state model was fitted with the transition probabilities dependent on a continuously time-varying covariate allowing for an annual cycle. The state-dependent distribution parameters and the constant parameters of the annual trend are shown in Table 6.6. For each herd, the expected movement distance for each state corresponds to shorter and longer displacements which are more similar to the latent states obtained for the basic Gamma models applied to the daily data. The constants for the annual trend terms are meaningless as is, however, when the transition probabilities are examined over time, they are more easily interpretable. Figure 6.4 shows the off-diagonal probabilities (i.e. the probability of switching state over time) for the Numbi and Shitlave sable herds. The patterns are quite different for the two herds, with the Numbi herd having very similar transition probabilities for the two states over time. For both latent states, the probability of switching state is highest during the late Wet and Early Dry season. While for the Shitlave herd, the probability of switching from state 2 to state 1 (Gamma.21) remains relatively constant throughout the year, however the probability of switching from state 1 to state 2 (Gamma.12) increases at almost the exact same time as the Numbi herd's, around January. The expected movement distances for these two herds are very similar for the longer state (approximately 2km), but the shorter state for the Numbi herd is 0.55km compared to 0.93km for the Shitlave herd. These plots imply that the higher probabilities of shifting states is more associated with the benign conditions at the end of the late Wet season, while the animals get stuck in one state during the tougher conditions. The actual ecological interpretation of these states would need to be investigated further but is beyond the scope of this thesis. However, the results from this analysis seem to provide more easily interpretable results from a biological point of view, and the concept of changing probabilities of shifting state throughout the year is easy to comprehend. Although the model fitting process is more computationally intensive and time consuming than the time-varying state-dependent distribution parameter method, both methods appear to be suited for this application to animal movement modelling.

6.2.3 Comparison of Seasonal Behaviour

One of the desired outcomes for a seasonal comparison is a formal method to test whether significant differences exist between the behaviours in the different seasons. In an identical test to what was done for the Mixture Models, the distribution of movement under an HMM is compared between a combined model for the full data and independent hidden Markov models fitted to the summer and winter data for the Punda Maria buffalo herd. 4-state log-normal HMMs were selected by AIC as providing the best fit for the combined and winter data, with a 3-state model for the summer movement.

$$
\begin{aligned}
G^2 &= -2\left(\text{combinedLL} - \left(\text{winterLL} + \text{summerLL}\right)\right) \\
&= -2\left(-2235.104 - \left(-1547.551 + (-694.9914)\right)\right) \\
&= 14.87633 \\
G^2 &\sim \chi^2\left(df2 - df1\right) \\
G^2 &\sim \chi^2\left(35 - 20\right) \\
G^2 &\sim \chi^2\left(15\right)
\end{aligned}
$$

At a 5% level of significance $G^2 = 14.88 < \chi^2(15) = 24.9958$, therefore the null hypothesis of no difference in movement is not rejected, and it is concluded that there is no evidence for a difference in the movement patterns of the buffalo herd between summer and winter. This is a different result to what was found in the Mixture Model chapter, where a significant difference in movement was found when model fitting was done using the Mixture Models. This suggests that the added dependency built into the model via the HMM is able to account for the movement patterns better than the IMM and hence it is not necessary to model the summer and winter movements separately. The IMM for the buffalo herd had far more active movement states compared with the HMM (will be discussed in more detail in Chapter 8) which would show more variation across the year as they are associated with much longer movements.

The test for the difference of means and the one-way analysis of variance for multiple parameters can also be applied to the HMM parameter estimates as it was

for the IMMs in Sections 3.8.7 and 3.8.8. This is an example of how formal tests can be run to investigate significant differences in the behavioural patterns across seasons. The tests could be used for all of the fitted parameters across all of the herds.

6.2.4 Summary of Seasonal Models

These results show that the models are detecting the changing movement patterns of the animals, in both the annual and the seasonal models. It is very encouraging that even the annual model is in fact able to identify behavioural states which are mainly only represented in one of the three seasons. The seasonal-split model parameters obtained using the seasons applied here, show that hidden Markov models are flexible enough to be easily extended for this type of "block" covariate, and that the modelling process is sensitive enough to pick up the different movement patterns over time. The application of the seasonal split via the state-dependent distribution parameters is relatively easy to code and is not too computationally intensive although it takes longer to converge than the basic HMM. The models fitted using a time varying covariate are more computationally intensive to fit but still provide results consistent with the expected behaviour for these species and are easy to interpret.

6.3 Turning Angle Models

Franke et al. (2006) found that the turning angle did not really differ among states which implies that the turning angle is not helping to differentiate the observations between the latent states. A similar outcome was found for the analysis of wildebeest (*Connochaetes taurinus*) GPS data, also from the Kruger National Park (Holdo, unpublished). Owen-Smith, Goodall & Fatti (2012) used only the distance between locations in an application of the Mixture Model approach. They mention that for ungulates which move in coordinated herds, the turning angles between successive linear movements become less relevant. Including the turning

angle in the model fitting process increases the complexity of the model by requiring additional parameters to be fitted and introduces potential problems for the convergence of the likelihood due to the circular nature of the data.

6.3.1 Methods

A simulation study was done in order to investigate the effect that the turning angle has on the analysis using HMMs of movement similar to those used in this study. The main focus was to investigate the impact of including the turning angle for daily movement models. Under the assumption that the turning angle and the displacements are independent, displacement and turning angle data were simulated assuming a 2-state HMM with Weibull or Gamma distributions for the step lengths and Wrapped Cauchy distributions for the turning angles. The Wrapped Cauchy distribution was used to model the turning angles as it has been used in other movement studies in the literature and was shown to outperform the von Mises distribution (Langrock et al. 2012). The states were characterized by shorter movements which tended to reverse direction and the second state with longer movements and persistent direction. These states are assumed to correspond to the animal being 'encamped' and 'relocating' and have been discussed in a number of movement studies. The simulation means that the series of "true" states is known. Weibull HMMs with two states were then fitted to the simulated data. One model included the turning angle and one other excluded it. The results of the model fitting are compared to the true known parameters and known sequence of states to identify the state allocation and parameter estimate accuracy between the two model specifications. The process was repeated using Gamma distributions for the simulated displacement distances, and then Gamma HMMs including/excluding the turning angle were fitted to the simulated data. Wrapped Cauchy distributions were still used for the turning angles.

The probability density function for the Wrapped Cauchy distribution for $0 \le \theta \le 2\pi$ is given by:

$$f(\theta;\mu,\rho)=\frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)}$$

The parameter μ is the mean direction $0\le\mu\le 2\pi$, while ρ is the mean cosine of the angular direction, with $0\le\rho\le 1$. The simulation and model fitting was done using the functions `rwrpcauchy` and `dwrpcauchy` from the CircStats package (S-plus original by Ulric Lund and R port by Claudio Agostinelli 2012) in R. The probability density functions for the Gamma and Weibull distributions were given in Section 3.7.2 and 5.12 respectively.

A 2-state Weibull HMM was simulated with the following parameters:

$$\alpha=\begin{bmatrix}0.86\\0.81\end{bmatrix}\quad \beta=\begin{bmatrix}0.04\\0.48\end{bmatrix}\quad \mu=\begin{bmatrix}3.16\\6.05\end{bmatrix}\quad \rho=\begin{bmatrix}0.21\\0.35\end{bmatrix}$$

$$\Gamma=\begin{bmatrix}0.70 & 0.30\\0.24 & 0.76\end{bmatrix}$$

These parameters were chosen to be similar to the fitted parameters obtained for a 2-state movement model fitted to bison movements (Langrock et al. 2012). Figure 6.5 shows the histograms of the displacements and turning angles for the whole dataset and split by state. State 1 is clearly associated with the short displacements and angles around π while state 2 is associated with much longer displacements and angles around 0 and 2π indicating directional persistence. A similar simulation was done assuming a true Gamma/Wrapped Cauchy model with 2-states with the histograms shown in Figure 6.6. The true parameters were defined as:

$$\alpha = \begin{bmatrix} 0.50 \\ 0.80 \end{bmatrix} \quad \beta = \begin{bmatrix} 1.5 \\ 5.00 \end{bmatrix} \quad \mu = \begin{bmatrix} 3.16 \\ 6.00 \end{bmatrix} \quad \rho = \begin{bmatrix} 0.20 \\ 0.35 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 0.70 & 0.30 \\ 0.24 & 0.76 \end{bmatrix}$$

6.3.2 Simulation Results

In order to investigate the accuracy of the state allocation, 50 simulations were run with the Viterbi algorithm used to determine the most likely sequence of states for the maximum likelihood model fitted to each of the 50 simulated datasets. The true simulated and maximum likelihood fitted parameters for one of the simulations are shown in Table 6.7. Each of the parameters has been reasonably estimated by the HMM fitting process including the turning angle. The Viterbi algorithm was used to determine the predicted state allocation for the fitted HMM model. When compared to the known states for this example, 81.7% of the observations are correctly allocated to the true state when estimated using a Weibull/Wrapped Cauchy HMM. Table 6.7 also shows the results when the model fitting process was repeated using only a Weibull HMM and excluding the turning angle. The parameter estimates for the Weibull distributions and the transition probability matrix are similar to the model with turning angles. The results of the Viterbi algorithm without the turning angle shows a 78.4% accuracy for this example, only 3.3% lower than the more complex model which included the turning angle. A direct comparison of the state allocation of the fitted models shows that 87.1% of the observations were allocated to the same state by both methods. 73.6% of the observations were allocated correctly to the true state by both methods, with 4.8% correctly allocated by the Weibull model but incorrectly by the Weibull/Wrapped Cauchy model. 8.1% of the observations were allocated correctly by the turning angle model but incorrectly by the Weibull HMM and finally 13.5% of the observations were allocated incorrectly by both methods.

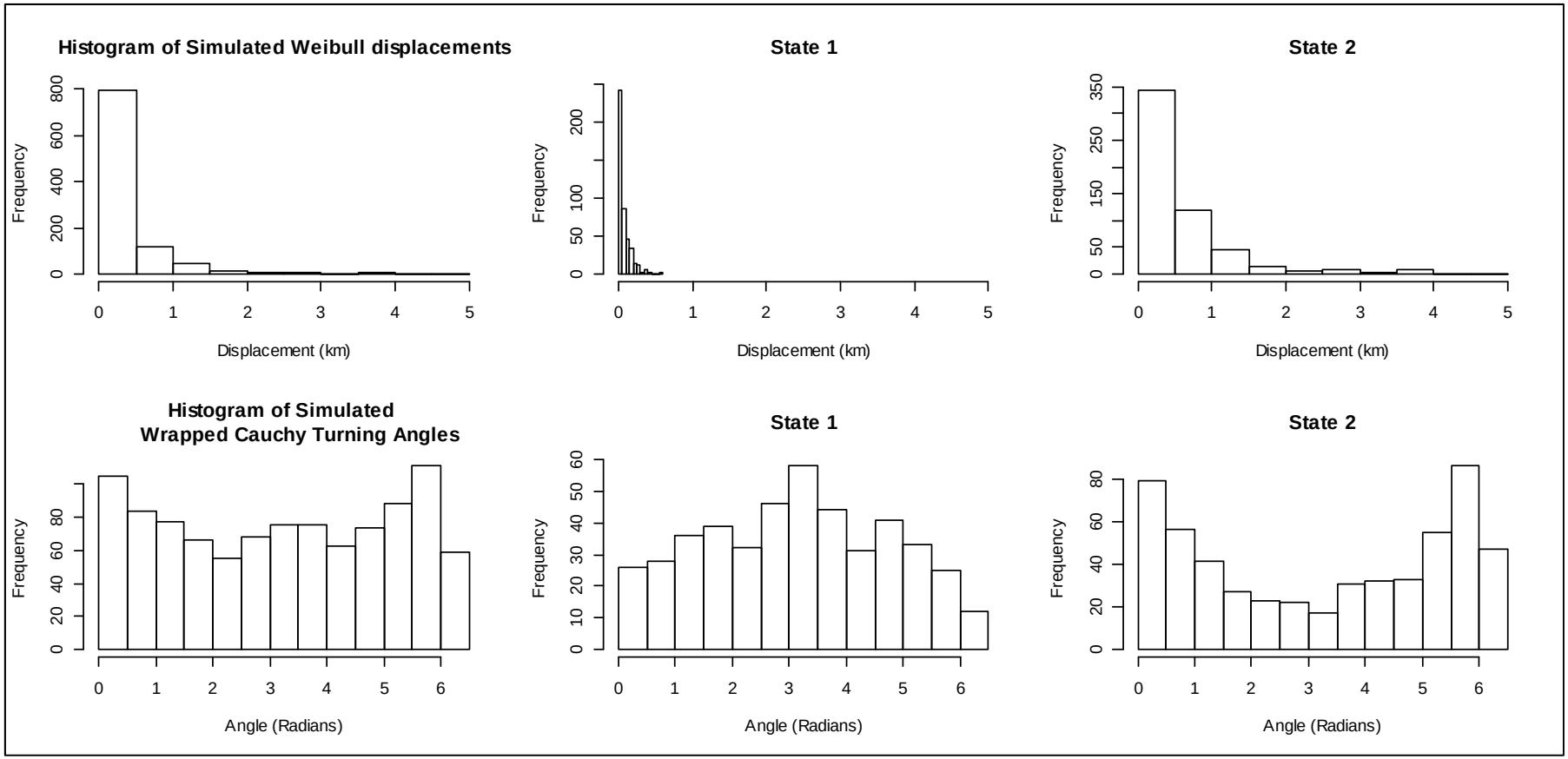

Figure 6.5 - Histograms of simulated displacements (top) and turning angles (bottom) for 2-state Weibull/Wrapped Cauchy HMM; all data (left), State 1 (centre) and State 2 (right)

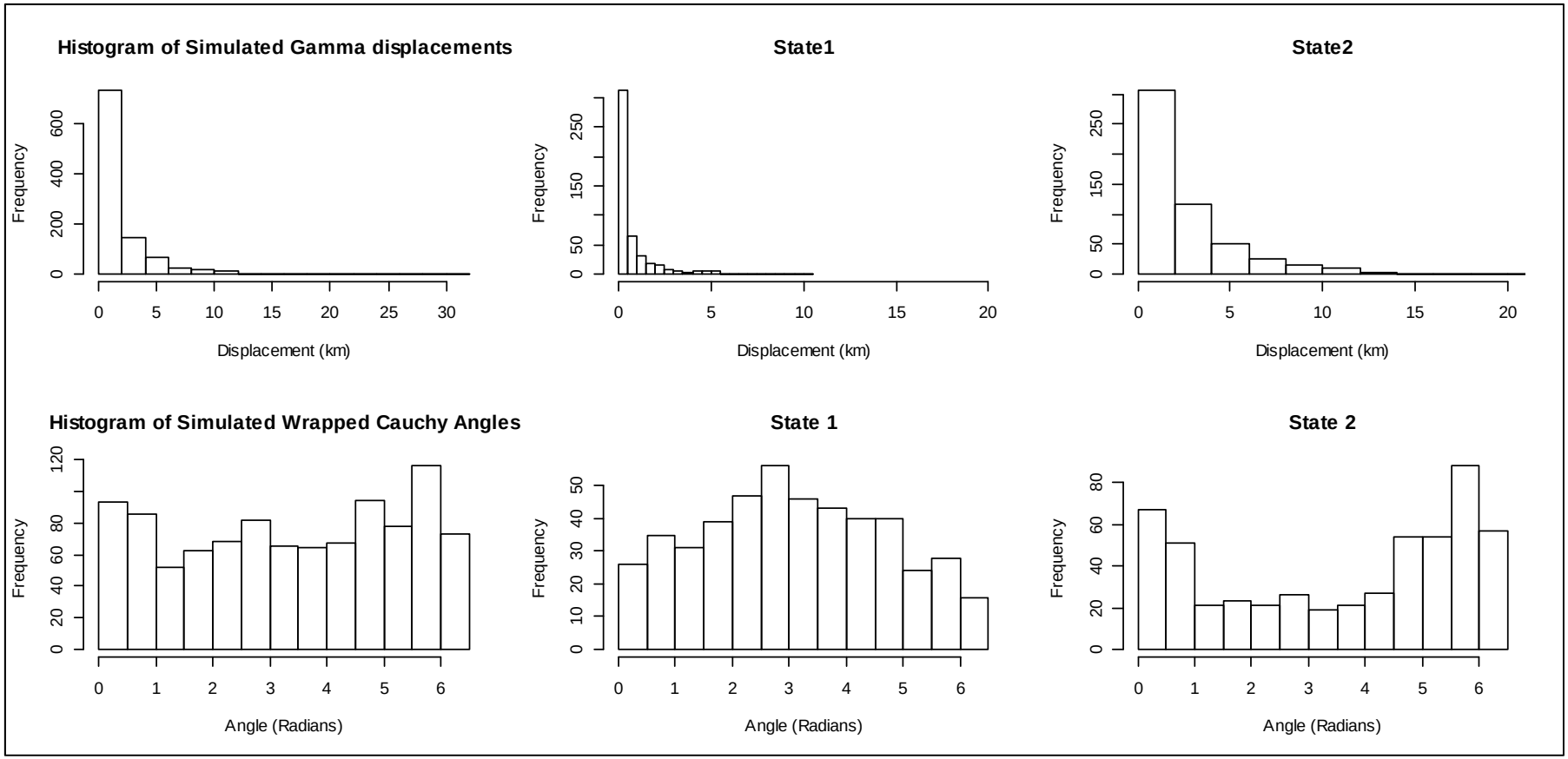

Figure 6.6 - Histograms of simulated displacements (top) and turning angles (bottom) for 2-state Gamma/Wrapped Cauchy HMM; all data (left), State 1 (centre) and State 2 (right)

Table 6.7 - True and Fitted Parameters for 2-state Weibull/Wrapped Cauchy HMM

	True Parameter	Fitted Parameter including turning angle	Fitted Parameter excluding turning angle
$\alpha 1$	0.86	0.90	0.94
$\alpha 2$	0.81	0.85	0.79
$\beta 1$	0.07	0.07	0.07
$\beta 2$	0.48	0.52	0.45
$\mu 1$	3.16	3.16	
$\mu 2$	6.05	6.12	
$\rho 1$	0.21	0.15	
$\rho 2$	0.35	0.35	
$\gamma 11$	0.70	0.65	0.62
$\gamma 12$	0.30	0.35	0.38
$\gamma 21$	0.24	0.30	0.24
$\gamma 22$	0.76	0.70	0.76

The above simulation shows the model including the turning angle does improve the accuracy of the state allocation slightly. However the Weibull/Wrapped Cauchy model is more complicated to fit. It also tends to be more sensitive to the choice of the initial values and does run into some complications when the parameters for the mean direction of the Wrapped Cauchy distribution (μ_i) are close to the boundaries due to the circular nature of the distribution.

After 50 simulations of Weibull/Wrapped Cauchy observations, the average state allocation accuracy for the 2-state fitting approach including the turning angle was 81.14%, while the method excluding the turning angle from the model fitting had an average state allocation accuracy of 75.60%. The results of the state allocation for each simulation are shown in the Appendix Section 6.7 Table 6.12. The simulation exercise was also repeated 50 times for a true 3-state Weibull/Wrapped

Cauchy model. The average state allocation accuracy including the turning angle was 81.22%, while the method excluding the turning angle had an average of 79.56%. For the 3-state models, the average accuracy was much closer between two methods which suggests that as the complexity of the true model increases, the turning angle plays less of a role in differentiating the states. This is of course dependent on the amount of overlap in the states either from the perspective of the displacement distances or the turning angles. The model complexity increases substantially with the inclusion of additional states and issues with obtaining convergence can occur for the multi-state turning angle models.

The results for one example of a simulation of the Gamma/Wrapped Cauchy HMM are shown in Table 6.8. Once again, the fitted model including the turning angle provides good estimates of the true parameters. However, the Viterbi algorithm shows that only 74.4% of the observations are correctly allocated using the fitted Gamma/Wrapped Cauchy model. This dropped to 68.2% state allocation accuracy for the fitted Gamma model excluding the turning angle. 81% of all observations were allocated to the same state by both methods. 61.8% were correctly allocated by both methods, with 19.2% incorrectly allocated by both methods. 6.4% of observations were correctly allocated by the Gamma displacement model, but incorrectly allocated by the more complex Gamma/Wrapped Cauchy model. However, this model including the turning angle allocated 12.6% of observations correctly which were incorrectly allocated by the Gamma displacement model. The state allocation accuracy results for all 50 simulations are shown in the Appendix Section 6.7 Table 6.13, where the average accuracy for the fitted models including the turning angle was 73.13% and excluding the turning angle was 70.46%. Once again, the models including the turning angle do on most occasions outperform the model excluding the turning angle, however the improvement is minimal. Although a slight forward persistence was shown in Figure 2.26 this was not included in the simulation, however it is unlikely that this slight association between the turning angle and the displacement will affect the conclusions obtained here.

Table 6.8 - True and Fitted Parameters for 2-state Gamma/Wrapped Cauchy HMM

	True	Fitted Parameter	Fitted Parameter
$\alpha 1$	0.40	0.43	0.43
$\alpha 2$	0.80	0.86	0.85
$\beta 1$	1.50	1.14	1.89
$\beta 2$	3.00	3.11	3.44
$\mu 1$	3.10	3.08	
$\mu 2$	6.00	6.01	
$\rho 1$	0.20	0.16	
$\rho 2$	0.35	0.38	
$\gamma 11$	0.70	0.68	0.86
$\gamma 12$	0.30	0.32	0.15
$\gamma 21$	0.24	0.28	0.22
$\gamma 22$	0.76	0.72	0.78

The accuracy of the state allocation was better for the Weibull models than the Gamma models, and the variation in the accuracy for the Gamma model was greater. This suggests that the Weibull model is providing more consistent results, although this will be very dependent on the true parameters and the overlap between the states.

There are a variety of methods for implementing a HMM including the turning angle in the fitting process. In their application of the HMMs for the bison data, Langrock et al. (2012) used only a 2-state model and the parameters for the Wrapped Cauchy distribution were carefully restrained to avoid the issues of convergence to the boundaries and to ensure that one state had angles which continued in direction and the other state had reversing directions. The issue with this is that it imposes a structure on the model and is impossible in its current form to extend to multiple states. If a multi-state model including the turning angle is coded, the constraints on the Wrapped Cauchy distribution parameters have to be removed. In initial tests using the two fitting methods (with and without the constraints on the Wrapped Cauchy distribution), the results and interpretations of the turning angle distributions can differ quite substantially. This suggests a tricky

problem with fitting the turning angle in the model. The circular data are challenging and the results can differ quite a lot based on the choice of which way to define the model. From the models tested, the models including the turning angle are far more sensitive to the choice of starting parameters. The mean direction parameter of the Wrapped Cauchy distribution is problematic since a value of 0 or 2π is actually the same, however with the fitting approach, the direct numerical maximization will approach it from one side only.

6.4 Irregularly Timed Observations

One of the apparent disadvantages of HMMs for the analysis of animal movement data is that the models require regularly spaced intervals between successive locations. Missing data can be introduced into the series when observations are missed due to satellite connection errors, which is becoming less common as technology improves and more satellites are available to connect to. However, as discussed earlier, changes can be made to the frequency at which the collar attempts to obtain a location fix. The fewer fixes that are attempted during a day, the longer the battery of the GPS will last which will mean that the data series will span a longer time over the year(s). For some species, it makes more sense to have irregularly spaced observations where more frequent observations would be obtained when the animal was expected to be more active and fewer during its sleeping time. For example, for a nocturnal predator, it would be beneficial to rather have more locations obtained during the night. This is when observing the animal is more difficult, and it is more likely to be active than during the day. In some cases, as was the case with the sable antelope, buffalo and zebra study done in the Kruger Park, the battery life was not certain and the collars were set to record at six hour intervals with additional intensive periods of hourly observations for a week or two at a time. The objective of the collar placements was to obtain locations throughout the annual cycle, and it was not known beforehand how long the batteries were going to last. In these cases, it would be very advantageous to extend the methods usually used for regular time interval

analysis in order to utilise these observations obtained at the irregular time intervals. Data is lost when converting irregular time locations to regular movements since two regularly spaced locations are needed to calculate one displacement. One possible way of dealing with this irregular data is to convert the input data from the displacement between successive locations to the *rate of movement* over the time period between successive obtained locations. This conversion to movement rate is useful, however it is not ideal as the behavioural processes that occurred during the time period will be blurred by the averaging of the total displacement by time elapsed. Another method used in some of the literature is to estimate the missed location through interpolation (Polansky et al. 2010). However, this has the potential to introduce more uncertainty and error into the model. A method which can combine data from various time scales would be advantageous for the analysis of animal movement data using HMM techniques, and would maximise the available data which had been collected.

6.4.1 Methods

The likelihood function of an HMM is calculated from the initial distribution, the density of the state-dependent distribution and the transition probability matrix. The transition probability matrix contains the probability of switching from one state to another and of remaining in the same state from one time period to the next. For irregularly spaced displacements, this matrix can be raised to a power corresponding to the number of time steps between locations based on the shortest time unit used in the data. For example, in a dataset of mainly hourly observations, with phases of 2-hour and 4-hour observations, the transition probability matrix is raised to the power two or four for observations obtained from the 2- and 4-hour displacements respectively. In some cases, the time differences between observations can be variable with time gaps ranging from one hour to 24 hours or more between them. In this case the transition probability matrix would be raised to the power corresponding to the time gap for that observation.

For a non-stationary HMM, the likelihood is defined as:

$L_T = \boldsymbol{\delta}\mathbf{P}(x_1)\boldsymbol{\Gamma}\mathbf{P}(x_2)\ \boldsymbol{\Gamma}\mathbf{P}(x_3)\ldots\boldsymbol{\Gamma}\mathbf{P}(x_T)\mathbf{1}'$. For the HMM with irregular time steps between successive observations, the likelihood function is defined as:

$L_{Tirreg} = \boldsymbol{\delta}\mathbf{P}(x_1)\Gamma^{h_1}\mathbf{P}(x_2)\Gamma^{h_2}\ldots\Gamma^{h_{T-1}}\mathbf{P}(x_T)1'$ where h_i is the number of hours between observations i and i+1. If the observations were recorded at a different time scale, for example minutes, the h_i is defined as a multiple of the shortest time step between the majority of observations. As the number of time steps between the observations increases, the transition probability matrix is multiplied by itself an increasing number of times. This product ends up being a matrix with identical rows equal to the stationary distribution of the Markov chain with transition probability matrix $\boldsymbol{\Gamma}$. This implies that the probability of being in a state after a number of steps remains constant for each state equal to the stationary distribution probabilities. This appears to happen after between eight to twelve time periods for the transition probabilities associated with the hourly model which implies that the probability of being in state C_i after twelve hours does not depend on the behavioural state twelve hours before. Therefore after a time gap of eight to twelve hours, the model being fitted is actually an independent mixture model. However for time gaps in the data which are smaller than this, the HMM is fitted with the transition probability matrix modified based on the number of time periods between observations and the total distance moved between obtained locations.

6.4.2 Results

While this method could be applied to the sable, buffalo and zebra data with the 6-hourly gap data included, this was an extremely artificial situation and the six time step gap is already nearing the point where the matrix ends up having identical rows of the stationary distribution. For this reason, data from a collaborative study in the Kruger National Park tracking a lion pride is used. The data were collected from a collar placed on a female lioness in the Crocodile Bridge section in the south-east of the Kruger National Park. This is part of a long-term study tracking

22 lion prides in the Park. The collar was a Hawk105 GPS/GSM device (http://www.awt.co.za) The collar recorded mostly during the night at 1- or 2-hour intervals, with mainly 4-hour intervals during the day. 5824 observations were recorded. A table of the number of observations for each time gap is shown in Table 6.9.

Table 6.9 - Summary of the frequency of time gaps between successive observations – Crocodile Bridge lioness (hours)

Hours	1	2	3	4	5	6	7	8	9	10	11	12	>12
Frequency	2519	982	398	1556	98	48	34	121	15	6	7	26	14

More than 85% of the observations had a time gap of 1-, 2- or 4-hours. Since the matrix becomes more like the stationary distribution as the time gap increases and the HMMs are able to cope with missing data, the example analysis was only run using observations with these 1-, 2- and 4-hour time gaps. All other time gaps were considered as missing data.

2-, 3- and 4-state Irregular Time HMMs were run, using the displacement and time gap between successive observations as the input to the matrix. The Viterbi algorithm was used (modified to allow the transition probability matrix to be raised to the respective power) to determine the most likely state for the fitted model. The AIC selected the 4-state model as the 'best', while the BIC was more conservative and selected the 3-state model.

In order to investigate the effect of the Irregular Time gap model specification, a comparison was done between this model and one fitted using the movement rate data. This meant that the input was calculated as the total displacement between successive obtained locations divided by the total time gap between successive

locations. For a 4-state model, a comparison of the state allocation is show in Table 6.10. This model is called the 'Movement Rate model'.

Table 6.10 - Comparison of state allocation between Movement Rate and Irregular Time Gap HMMs

	Irregular Time Gap Model				
Movement rate Model		**1**	**2**	**3**	**4**
	1	1208	1655	121	8
	2	4	387	778	437
	3	0	23	337	688
	4	0	1	29	150

Almost all of the observations which had been allocated to state 1 (resting) are allocated to the same state using the movement rate model. However, states 2, 3 and 4 have quite different allocations using the Irregular Time gap model. A comparison of the mean movement distances allocated to each state by the two models is shown in Table 6.11. The expected movement rate and displacement are the expected values for the distributions fitted to the four states using the movement rate (regular time gap) and Irregular Time gap models respectively. The average movement rate is the average observed movement rate of the observations allocated to the four states based on the Viterbi algorithm using the fitted Irregular Time gap model.

Table 6.11 - Summary of expected movement rate and displacement for regular and Irregular Time gap HMMs

	State 1	**State 2**	**State 3**	**State 4**
Expected Movement rate (km/h); regular time lag	0.013	0.046	0.267	0.830
Expected displacement (Irregular time lags)	0.016	0.162	0.619	1.453
Average Movement rate km/h (Fitted Irregular Time Lag model)	0.009	0.053	0.334	0.821

The results are similar with State 1 being a clearly inactive, 'resting' state, with the states 2, 3 and 4 becoming increasingly active. It is very encouraging that the results using the two methods are consistent, especially since the Irregular Time gap model is able to utilize more data than a regular model with missing data which can be very useful for certain studies. Plots of the state allocation by time of day for these two methods of analysis are shown in Figure 6.7 and Figure 6.8. Quite different patterns are seen, with State 2 of the Irregular Time gap model being strongly associated with daylight hours. The different patterns of state allocation and the expected values of the distributions suggest that the biological interpretation of the states from the Irregular Time gap model would be different to that of the movement rate model. This would need to be investigated further in terms of lion biology, which is beyond the scope of this study. However, this work is a demonstration of how this method can be applied to studies with irregular time gaps. Further work is needed to determine how useful it is in terms of uncovering biologically meaningful behavioural states within the data.

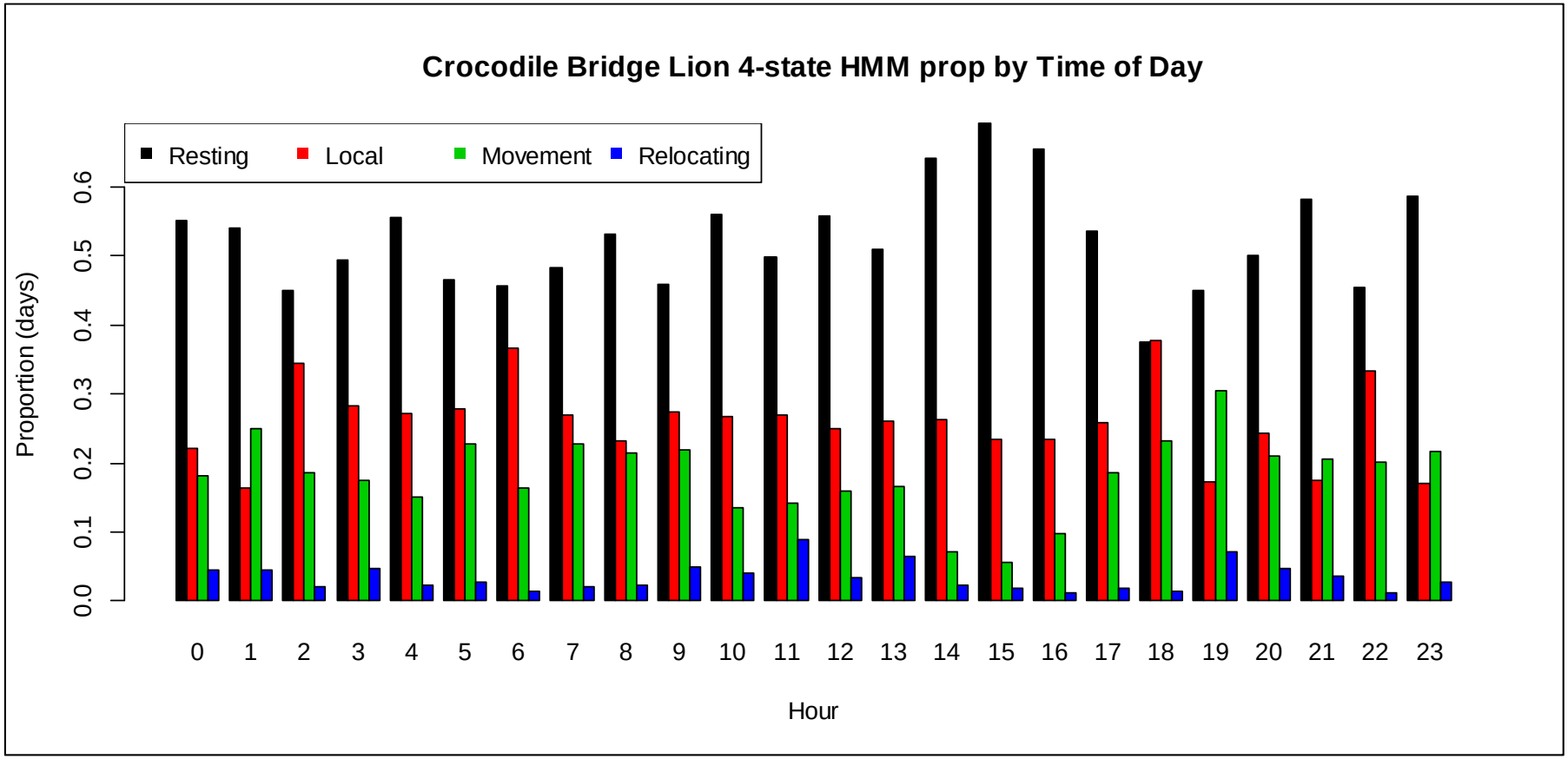

Figure 6.7 - Proportion of state allocation by time of day Movement Rate Regular time gap model - Crocodile Bridge Lion

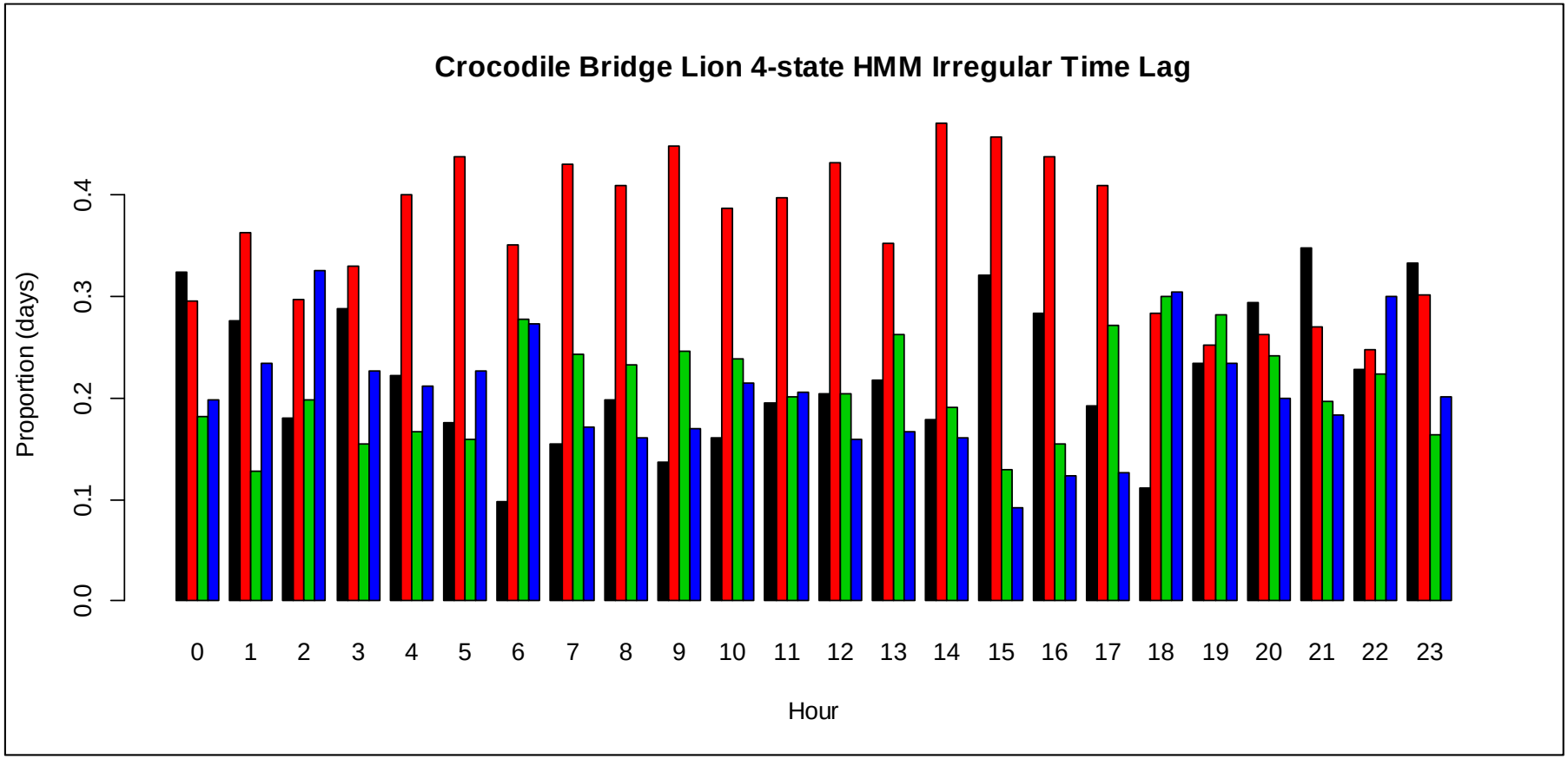

Figure 6.8 - Proportion of state allocation by time of day - Irregular Time gap model - Crocodile Bridge Lion

6.5 Conclusion

Hidden Markov models are becoming more popular in the analysis of animal movement data, which itself is a growing field often referred to as Movement Ecology (Nathan et al. 2008). More GPS collars are being deployed on a variety of species in many different parts of the world. Battery life is improving which means that more locations are obtained over longer periods of time. While this all sounds positive, it introduces challenges in terms of the data analysis such as autocorrelation, daily and seasonal cycles within the movement and missing data if locations are missed for some reason or the collar is set to record at irregular intervals. While HMMs seem to provide very interpretable and realistic results for a variety of species published in the literature discussed in Chapter 5 and for the three species studied for this thesis, extensions to the basic model will in some cases be needed.

It is well known that all of these ungulate species will follow some sort of daily cycle such as foraging after dawn and before sunset, resting during the heat of the day and being less active at night. There will also be seasonal changes to their behaviour as not only does the number of daylight hours vary during the day, but so do the climatic conditions and availability of resources. In order to be truly useful models, any model applied to animal movement needs to have the potential to be extended in order to account for these cycles.

The results presented here illustrate four different methods for including seasonality into the models. The methods vary in complexity and depending on the study one or the other might prove to be more useful. For these three ungulate species, if the simple categorization of three seasonal blocks is assumed to be sufficient to explain the seasonality, both the state-dependent distribution and the transition probability matrices based on the seasonal blocks provided realistic and interpretable results. If a continuously varying time trend is required, the models

can be extended without prohibitively increasing the computational time. In this study, the time varying transition probability matrix was found to provide more realistic results, however the state-dependent distribution method is easier to implement and in some cases may provide more insightful results. The results from the state dependent distribution time varying components appeared to find a seasonal movement latent state, rather than uncovering activity states (switching between patch utilisation) which varied through the seasonal cycle which was the intention of this study. The important feature of all this work is that HMMs are indeed very flexible and relatively simple to extend. This makes them very good candidates for the analysis of animal movement data.

A novel application of the HMMs using irregular time gaps in the input data is a very useful application of the HMM technique when it is desirable to have irregular observations to focus observations on key movement times or to be able to include periods between locations when observations have been missed. In effect, this technique fits distributions linked via a Markov process when there are a small number of missing time steps, and an independent mixing model when the time gaps are larger than around eight time gaps. In the presence of substantial amounts of missing data or an intentional variable time lag study design it is advantageous to extend the HMMs and to study the application of this irregular time lag analysis further. The unobserved states can quite easily be interpreted for the basic models, being assumed to correspond with the behavioural states. However the extension of the model for irregular time lags resulted in underlying states which require slightly different interpretations, based on the plot of the states by the time of day and the associated expected displacement distances. Further studies, beyond the scope of this thesis, are needed to understand how robust this method is and whether the latent states can provide consistently useful ecological insights into the movement patterns.

The turning angle is a metric commonly used in the analysis of animal movement studies, in conjunction with the linear displacement between successive locations. However, in the above simulation it is shown that the inferences and state allocations obtained from models, fitted with or without the turning angle included, remain very similar for some of the datasets and the inclusion does not improve the state prediction accuracy substantially.

It is expected that the importance of the turning angle will be associated with the time scale of the data. As new tracking devices are deployed with much shorter time intervals between locations (as short as a few seconds in some studies), the turning angles are likely to be far more influential in the model fitting process since the displacements will all be relatively short. For the time scales used in this thesis, the displacement distances on their own provide a suitably accurate metric for the identification of the latent states assumed to correspond to the animal's behaviour. While the turning angle does indeed improve the fit of the model, the improvement is minimal and as the complexity of the true model increases it appears that the simpler model excluding the turning angle provides a state allocation accuracy even closer to the turning angle model. The behavioural inferences drawn from the states do not seem to be changed by the inclusion of the turning angle. Since the results are very similar and the inferences are unchanged, it seems that for studies at these time scales for these three large herbivores, the inclusion of the turning angle will complicate the fitting process without adding substantial improvements to the modelling output. For these reasons, the decision to exclude the turning angle for the analyses in this thesis seems justified. The models fitted to the displacement distances only provide adequate results for the purposes of this study and the ecological questions asked of the data. The turning angle might also be more influential for marine studies where the animal can drift with the currents.

Both the seasonal and irregular time lag extensions and the ability to include the turning angle between displacements showcase the flexibility of the HMMs and their suitability for the analysis of animal movement data. While further research needs to be done, this study has shown the suitability of HMMs for extension in these ways and that the results are consistent with those found using the basic models.

6.6 Acknowledgements

South African National Parks for providing the lion data. Thanks to Dr Sam Ferreira, Dr Paul Funston and Nkabeng Maruping for their assistance and advice on the lion analysis.

6.7 Appendix

Table 6.12 - State Allocation Accuracy for Weibull/Wrapped Cauchy HMMs Fitted Including and Excluding the Turning Angle

	2-state Model		3-state Model	
	Including TA	**Excluding TA**	**Including TA**	**Excluding TA**
Simulation 1	80.6	73.5	78.4	81.4
Simulation 2	81.7	69.9	82.3	82.9
Simulation 3	81.7	75.9	80.7	63.1
Simulation 4	75.3	68.3	77.4	82.6
Simulation 5	81.8	71.9	79.1	75
Simulation 6	79.1	75.1	80.6	76.1
Simulation 7	81.4	78.2	86.5	82.2
Simulation 8	83.7	82.6	81.8	81.3
Simulation 9	82	73.1	79.1	76.8
Simulation 10	83.4	79.3	83.3	83.1
Simulation 11	79.7	78.8	85.9	83.6
Simulation 12	80.3	76.3	82	81
Simulation 13	82.8	80.3	84.5	80.8
Simulation 14	77.7	67.6	83.6	80.1

Simulation 15	82.4	71	80.6	78.7
Simulation 16	81.9	77.8	77.7	75.7
Simulation 17	83	79.3	85.6	83.7
Simulation 18	78.8	76.3	82.7	79.4
Simulation 19	78.6	71.9	71.1	77.8
Simulation 20	83.4	78.5	82.7	81.9
Simulation 21	82.3	76.6	86.6	82.6
Simulation 22	78.1	73.2	84.2	81.9
Simulation 23	85.5	81.4	82.2	81
Simulation 24	84.1	78.8	76.3	76.5
Simulation 25	79.9	72.8	82.3	81.7
Simulation 26	80.7	75.5	83.5	79.7
Simulation 27	82.3	78.9	83.5	79.7
Simulation 28	80.3	77	83.5	80.2
Simulation 29	82.4	79.3	82.3	81.9
Simulation 30	80.7	75.8	74	79.8
Simulation 31	81.5	80.3	83.2	82.3
Simulation 32	82.3	77.7	81.6	80.6
Simulation 33	80.4	71.4	72.9	73
Simulation 34	79.7	75	77.7	80.1
Simulation 35	81.8	75.9	80.9	80.1
Simulation 36	84.5	74	81.1	82.3
Simulation 37	80.6	73.5	68.8	75.9
Simulation 38	81.7	69.9	85.5	84.1
Simulation 39	81.7	75.9	86.9	85.3
Simulation 40	75.3	68.3	83.3	82.2
Simulation 42	81.8	71.9	82.2	75.7
Simulation 42	79.1	75.1	81.4	77.3
Simulation 43	81.4	78.2	82.8	80.8
Simulation 44	83.7	82.6	85.2	82.4
Simulation 45	82	73.1	83.8	78.9
Simulation 46	83.4	79.3	81.8	81.8
Simulation 47	79.7	78.8	80.6	78.1
Simulation 48	80.3	76.3	76.3	74
Simulation 49	82.8	80.3	80.4	78.3
Simulation 50	77.7	67.6	80.5	72.7
Average	**81.14**	**75.60**	**81.22**	**79.56**

Table 6.13 - State Allocation Accuracy for Gamma/Wrapped Cauchy HMMs Fitted Including and Excluding the Turning Angle

	2-state Model	
	Including TA	**Excluding TA**
Simulation 1	74.7	72.1
Simulation 2	74.3	67.7
Simulation 3	74.4	74.2
Simulation 4	69.3	61.4
Simulation 5	71.6	60.5
Simulation 6	79.6	71.7
Simulation 7	70.5	74.6
Simulation 8	75.1	74.7
Simulation 9	74.1	70.5
Simulation 10	74	73.9
Simulation 11	76.5	76.4
Simulation 12	75.8	67.3
Simulation 13	72.6	70.6
Simulation 14	72.8	73.7
Simulation 15	69.1	59.8
Simulation 16	74.9	73
Simulation 17	60.9	67.7
Simulation 18	69.4	70.7
Simulation 19	72.6	71.9
Simulation 20	74.3	70.9
Simulation 21	72.4	70.6
Simulation 22	74.6	73.8
Simulation 23	76.3	73.9
Simulation 24	72.1	70.2
Simulation 25	70.7	75.6
Simulation 26	74.5	73.3
Simulation 27	73.3	68.8
Simulation 28	71.4	70
Simulation 29	73.6	66.9
Simulation 30	70.9	69.2
Simulation 31	74.8	70
Simulation 32	73.9	70.5
Simulation 33	70.8	70.1
Simulation 34	69	67.2
Simulation 35	75.5	54.8
Simulation 36	71	77.5
Simulation 37	65.5	72.4
Simulation 38	74.6	70.2

Simulation 39	75.7	74.1
Simulation 40	78.8	73
Simulation 41	75.7	68.2
Simulation 42	74.4	60.5
Simulation 43	74.6	73.6
Simulation 44	75.6	75.5
Simulation 45	73.5	66.8
Simulation 46	74.7	70.1
Simulation 47	74.9	75.4
Simulation 48	72.4	69.6
Simulation 49	73.5	73.4
Simulation 50	71.3	74.4
Average	**73.13**	**70.46**

7 BAYESIAN STATE-SPACE MODELS

7.1 Introduction

With the advancement in computing power (Patterson et al. 2008), Bayesian methods are becoming more popular and accessible as a method for data analysis. A state-space model (SSM) is a time-series model which allows for unobserved states and biological parameters to be estimated from observed data with error (Jonsen, Myers & James 2006). The state-space model combines a statistical model of the observation method with a model of the movement dynamics, which can include effects due to the environment and other variable factors (Patterson et al. 2008). This means that there is a transition equation in which an unobservable or hidden state, representing the true behaviour of the animal, changes through time according to a Markov process and an observation equation which relates the observed data to the underlying state (Eckert et al. 2008). The Markov process implies that the conditional density of the *i*-th value in the Markov chain, conditioned on the previous sequence of observations in the chain, is only dependent on the previous value in the chain (Press 2003). The challenge is to estimate the underlying states. Essentially, the Bayesian SSM has the same framework as the hidden Markov model (HMM), it is just fitted using Bayesian techniques and can include prior information if this exists.

7.2 Bayesian Techniques

Bayesian models are usually fitted using Monte Carlo Markov Chains (Morales et al. 2004) where the Gibbs sampler is the algorithm commonly used (Cressie et al. 2009) unless the posterior distribution can be evaluated analytically. The Gibbs sampler is a special case of the Metropolis-Hastings algorithm (Chib, Greenberg

1995). A useful factorization for the Bayesian framework breaks the joint distribution into 3 phases: a data phase, a parameter phase (φ) and a process phase (x) (Clark 2005, Cressie et al. 2009, Schick et al. 2008) where $p(\ )$ denotes the probability:

$$p(x \mid \varphi, Data) \propto p(Data \mid x, \varphi) p(x \mid \varphi) p(\varphi)$$

Schick et al. (2008) note that this representation is applicable for animal movement since the data are often recorded with error and the data phase allows for a different error structure. The parameter phase allows for uncertainty in the model parameters at all levels and the process phase allows an understanding of the factors driving the movement (Schick et al. 2008). The ability to account for error in the observed locations has meant that state-space models are particularly suitable for the analysis of marine animal tracks, the most common being seal and turtle studies where ARGOS tracking technology is often used to track the movements of these animals. ARGOS tracking data are commonly used for the collection of the movements of marine animals and other oceanographic data. More information about ARGOS tracking can be found on their website (http://www.argos-system.org). A Kalman filter is used during the processing phase by the ARGOS group to improve the accuracy of the locations, however errors will still be present in the observed data which makes the application of Bayesian state-space models popular for these types of marine studies.

7.2.1 Metropolis-Hastings Algorithm

The Markov Chain Monte Carlo (MCMC) approach is based on the need to simulate samples from an unknown *target* density ($\pi(x)$), when the *invariant* density is known, but the transition kernel is not known (Chib, Greenberg 1995). The method finds a transition kernel whose n-th (for large n) iteration converges to the target density (Chib, Greenberg 1995). This approach is gaining in popularity since computers have improved in terms of computing power and it can be programmed in various languages including R (R Development Core Team

2011) or WinBUGS (Lunn et al. 2000). An aperiodic, irreducible Markov chain is established, where the stationary distribution of the chain is the posterior distribution required (Albert 2009). This means that the simulated observations are approximately the target distribution, but only after a large number of iterations when the effect of the initial starting value can be ignored (Chib, Greenberg 1995). The convergence to the target distribution can only occur under the regularity conditions that the chain is irreducible and aperiodic (Chib, Greenberg 1995). However it is unknown how large *n* needs to be to ensure convergence. Using the notation of Chib & Greenberg (1995), a candidate density or 'proposal density' (Press 2003), $q(x,y)$, is defined such that $\int q(x,y)dy = 1$. They interpret this density that when a process is at x, the density generates a value y from $q(x,y)$. If a function $p(x,y)$ satisfies the reversibility condition $\pi(x)p(x,y) = \pi(y)p(y,x)$ then $\pi(\cdot)$ is the invariant density of $P(x,\cdot)$. It is unlikely that the candidate density $q(x,y)$ would satisfy the reversibility condition which implies that further iterations are needed to determine the invariant density. Therefore for some values of x and y: $\pi(x)q(x,y) > \pi(y)q(y,x)$. Chib & Greenberg (1995) then define $\alpha(x,y)$ as the "probability of a move" from x to y. This is also known as the "acceptance probability" (Albert 2009, Press 2003). If the process does not move to y, then the process returns x as the value from the target distribution. Therefore, the authors define the transitions from x to y according to:

$$p_{MH}(x,y) \equiv q(x,y)\alpha(x,y), \qquad \text{for } x \neq y$$

where $\alpha(x,y)$ is unknown. The probability $\alpha(y,x)$ is then set at its upper limit, which for a probability is one. And since the condition of reversibility is required, the probabilities $\alpha(x,y)$ and $\alpha(y,x)$ are introduced such that:

$$\begin{aligned}\pi(x)q(x,y)\alpha(x,y) &= \pi(y)q(y,x)\alpha(y,x) \\ &= \pi(y)q(y,x) \\ \alpha(x,y) &= \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\end{aligned}$$

Therefore, in order for p_{MH} to be reversible, the probability of move is set to:

$$\alpha(x,y)=\begin{cases}\min\left[\dfrac{\pi(y)q(y,x)}{\pi(x)q(x,y)},1\right], & \text{if } \pi(x)q(x,y)>0\\ 1, & \text{otherwise}\end{cases}$$

In addition to this, there exists a nonzero probability that the process remains at x instead of moving to y. Chib & Greenberg (1995) define this probability as:

$$r(x)=1-\int_{\Re^d} q(x,y)\alpha(x,y)dy$$

From this, the Metropolis-Hastings transition kernel is defined as:

$$p_{MH}(x,dy)=q(x,y)\alpha(x,y)dy+\left[1-\int_{\Re^d} q(x,y)\alpha(x,y)dy\right]\delta_x(dy)$$

$$\text{with } \delta_x(dy)=\begin{cases}1, & \text{if } x\in dy\\ 0, & \text{otherwise}\end{cases}$$

Since, $p_{MH}(x,y)$ was constructed to be reversible, the kernel has $\pi(x)$ as its invariant density (Chib, Greenberg 1995). The simulated observations are only assumed to be a sample from the target density after sufficiently large number of iterations, so that the effect of the starting value has been negated (Chib, Greenberg 1995). The choice of a suitable candidate distribution needs to be made, with a variety of options available which have been investigated by various authors (Chib, Greenberg 1995, Tierney 1994). The choice of the candidate distribution will influence the efficiency of the Metropolis-Hastings algorithm (Chib, Greenberg 1995).

7.2.2 Gibbs Sampler

The Gibbs sampler is a special case of the Metropolis-Hastings algorithm, and is considered one of the simplest MCMC algorithms (Press 2003). Casella & George (1992) define the Gibbs sampler as "a technique for generating random variables from a (marginal) distribution indirectly, without having to calculate the density". While the Gibbs sampler is most often used in the Bayesian context, it can also be applied in classical likelihood calculations (Casella, George 1992). In the

Bayesian approach, the Gibbs sampler is used to generate posterior distributions (Casella, George 1992). The Gibbs sampler allows complex joint densities to be estimated using a sequence of easier to compute conditional densities (Albert 2009). Once the observations have been generated to approximate the joint density; the mean, variance or any other characteristic of the joint density can be calculated, including the joint density itself (Casella, George 1992). Using the notation of Casella and George (1992), if the target joint density is given by:

$$f(x) = \int ... \int f(x, y_1, ..., y_p) dy_1 ... dy_p$$

the joint density $f(x)$ is approximated by sampling from the conditional distributions $f(x|y)$ and $f(y|x)$ (Casella, George 1992). The "Gibbs sequence" for the bivariate case ($p = 1$) is then generated in the form:

$$Y_0^{'}, X_0^{'}, Y_1^{'}, X_1^{'}, Y_2^{'}, X_2^{'}, ..., Y_k^{'}, X_k^{'}$$

where the initial value $Y_0^{'} = y_0^{'}$ is specified, and the rest of the "Gibbs sequence" is obtained iteratively by alternate values from randomly generated values from the conditional denisites given by:

$$X_j^{'} \sim f(x \mid Y_j^{'} = y_j^{'})$$

$$Y_{j+1}^{'} \sim f(y \mid X_j^{'} = x_j^{'})$$

As k tends to infinity, the distribution of X'_k converges to $f(x)$. A variety of approaches exist for obtaining the approximate sample from $f(x)$. Among others, by sampling the final value from a number of Gibbs sequences, selecting every r-th value from a very long Gibbs sequence or using all of the values of the Gibbs sequence (Casella, George 1992). The Gibbs sampler is a special case of the Metropolis-Hastings algorithm, when it is applied "Block-at-a-Time", using what Chib & Greenberg (1995) call the "product of kernels principle". It requires that it is possible to generate samples from the full conditional densities (Chib, Greenberg 1995). In a frequentist approach, the Gibbs sampler can be used to calculate the likelihood function and characteristics of the likelihood estimators (Casella, George 1992).

7.2.3 Bayesian State-Space Models

State-space models can either be applied using a Bayesian or frequentist approach. Jonsen et al. (2003) note that an advantage of using the Bayesian approach is that prior information can be incorporated formally into the model through the specification of prior distributions for variables in the model (Jonsen, Myers & Flemming 2003). Vague or non-informative prior distributions are often used for the unknown parameters (Jonsen, Flemming & Myers 2005) which allows the observed data to influence the estimates (Clark 2005). In practice, it seldom happens that sufficient information exists to accurately determine the prior distribution (Robert 1994). In this case, the benefit of using Bayesian methods is reduced, and the results should be very similar to the frequentist approach. Another advantage of the Bayesian framework is that population level inference can be based on the hyper-parameters using the hyper prior distributions for a hierarchical model. However, the shape of prior distributions should be checked, since a 'flat' prior is not necessarily non-informative (Lambert et al. 2005).

In probability theory, Bayes' theorem shows that one conditional probability depends on its inverse. The equation is given by:

$$P(\theta \mid data) = \frac{P(data \mid \theta)P(\theta)}{P(data)} \propto P(data \mid \theta)P(\theta)$$

The theory is based on the concept of the probability of observing one state being updated, having already witnessed another state. For a time series, the sequence of states is modelled via a Markov process, where the observed movements are conditional on the unknown states. The conditional distribution of the state at time $t + 1$ will depend on the observed movement at time $t + 1$ and the state at time t. A Bayesian framework enables more complex problems to be analysed, through factorizing the joint probability of all unknowns into a product of their conditional probabilities. A prior probability is interpreted as a description of what is known about a variable in the absence of any other information. A posterior probability is then the conditional probability of the variable taking into account other

information. If the goal is to predict x', the predictive distribution is obtained by integrating over the posterior distribution for the estimated parameters (θ).

$$p(x' \mid x) = \int p(x' \mid \theta) p(\theta \mid x) d\theta$$

For the animal movement studies, the posterior distribution is required to describe the behavioural processes which drive the movement, based on the model parameters and the observed data. This is obtained by conditioning the observed displacement distances on a model for the movement process and parameters, which are informed by prior information and updated at each iteration of the MCMC algorithm. There is potential to model the uncertainty (measurement error) in the displacement distances, if required, as well as the uncertainty in the movement process and the model parameters. The process begins by randomly generating a starting state for the Gibbs sequence. Then using initial starting values for the parameters, their prior distributions and given the observed distance moved, simulate values for each parameter from their posterior distributions obtained via the Bayes Theorem. The next state in the sequence is also simulated given the observations and the prior distributions. The next step uses the second observed displacement and simulates values for the parameters from their updated posterior distributions and the next state in the sequence. This process is continued until the end of the time series sequence which then constitutes the first Gibbs sequence. This process is continued until a stopping criterion is reached. WinBUGS uses three different sampling algorithms, depending on the structure of the fitted model. This is either 'direct sampling' for conjugate prior-posterior models, 'adaptive rejection Gibbs sampling' for log-concave posterior probabilities or the 'Metropolis-Hastings algorithm' for non-conjugate priors and non-log-concave posterior densities (Press 2003).

A Bayesian state-space model is a general and flexible alternative when dealing with nonlinear models. Any functional form and error distribution can be used (Clark 2007) and the model accounts separately for the error in the process and the observations (Schick et al. 2008). The methods are flexible enough to deal

with complex features within the data and missing data (Jonsen, Myers & Flemming 2003, Jonsen, Myers & James 2006, Clark 2005). The structure also allows for testing of the model fit using a full likelihood based framework as well as inference to be made about the movement process (Schick et al. 2008).

There are many advantages to using Bayesian analysis. Clark (2007) summarises these advantages as follows:

- Bayes provides a structure to formally incorporate prior information.
- Bayesian inference is conditional on the observed data.
- The posterior is the goal, and it is obtained directly. Bayesian analysis generates a posterior probability by combining the observed data with a prior distribution.

An advantage of the hierarchical state-space models is that it is possible, in the estimation process, to account separately for uncertainty in the process and in the observations (Schick et al. 2008). A hierarchical framework makes it possible to tease apart population-level patterns from individual-specific patterns (Eckert et al. 2008), which is particularly useful when individuals have different time series lengths. Another important advantage of the technique, which is particularly relevant for animal movement, is the flexibility of the model in terms of allowing a complex structure in the model for both the process and the observed data. Environmental covariates can be included in the models and informative priors can be used if the covariate data is missing for a particular point (Eckert et al. 2008) which can often be the case when dealing with environmental data. However, a disadvantage for complex hierarchical models is the difficulty in specifying 'non-informative' priors and confirming convergence of the model. This makes it difficult to identify when there are problems with the complex model (Clark 2005). For both the Hidden Markov models and the Bayesian SSMs, the movement model is specified *a priori* as a correlated random walk, with a hidden underlying movement state (Barraquand, Benhamou 2008). The Bayesian state-space models are still considered powerful models in terms of animal

movement analysis (Barraquand, Benhamou 2008). In most cases non-informative priors are used, which means that the posterior distributions are mostly based on the observed data (Morales et al. 2004). For simple Bayesian models with non-informative priors, the results obtained will be similar to classical analysis in most cases (Clark 2005). Some of the issues associated with fitting Bayesian models are how to specify an appropriate prior distribution, the sensitivity of the model to that choice of prior distribution and how to test for the convergence of the MCMC chain (Jonsen, Myers & Flemming 2003). HMMs can also be fitted using Bayesian techniques. In an analysis of the movement of Southern Bluefin Tuna (*Thunnus maccoyii*), Bayesian fitting methods were used to fit classical Hidden Markov Models (Pedersen et al. 2011b).

Information-theoretic approaches to model selection are popular, using criteria such as the AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion) (Cressie et al. 2009). The Deviance Information Criterion (DIC) has been used as a measure of the goodness-of-fit of Bayesian models. It is a generalization of the AIC which was introduced due to the difficulty of counting parameters in hierarchical models (Cressie et al. 2009). However, Morales et al. (2004) did not use only the DIC as a strict measure for determining the "best model". They also used the posterior distribution of parameters generated by the Monte Carlo Markov Chain to determine whether the models were able to accurately reproduce observed properties within the observed data.

7.3 Bayesian Methods applied to animal movement studies

Many of the studies which have used Bayesian methods to analyse animal movement data have tracked the movements of marine animals. In many cases these data are obtained with the potential for larger errors than usual for land-based tracking, or the locations are actually estimated values obtained from known currents, pressure and temperature data. The ability to model the error

process within the model fitting process makes the Bayesian techniques useful for these marine movements.

A very well cited paper, with over 200 citations, used Bayesian methods to fit correlated random walk models to the GPS relocation data of Elk in Canada (Morales et al. 2004). GPS collars were fitted to four female elk in a group of 116 which were relocated from Alberta to Ontario. Locations were obtained around 2am, and the daily movement rate and turning angle between observations were used as the input for their models. Habitat type was included in one of the models as a covariate. They used Bayesian methods to fit the following models to the movement data, and used the results to infer the behavioural activity of the animal:

- *Single* - random walk. There is a single movement state which generates the whole movement path.
- *Double* – a mixture of two random walks. The movement path is generated by two movement states, and the observation is allocated to one of the two states, independently of the previous state.
- *Double with Covariates* – a mixture of two random walks with the probability of being in a movement state related to the habitat type which the animal was moving through. This is exactly the same model as the 'Double' except the probability of being in one of the two states is dependent on the habitat in which the animal is at the time of the observation.
- *Double Switch* – a mixture of two random walks where the probability of switching from one state to the other was formally incorporated into the model. This is the same as the 'Double' model except that the probability of being in a state at time t is dependent on which state was occupied at time t -1. The transition probabilities are fixed and do not vary with time.
- *Switch with covariates* – a combination of the Double Switch and Double with Covariates model. Same as the 'Double Switch' model except the transition probabilities are not fixed, but rather depend on an

environmental covariate. In this case, Morales et al. (2004) used the distance to habitat type as the covariate.

- *Switch constrained* – the Double Switch model except the mode of the more active state distribution is forced to be greater than zero through constraining the parameters.
- *Triple Switch* – a mixture of three random walks with fixed transition probabilities between each state incorporated into the model. This is the 3-state extension of the 'Double Switch' model.

The priors used were always vague, except for the case of the Switch Constrained model (Morales et al. 2004). Morales confirmed on the 6th June 2013, that the prior distributions were specified such that they seemed reasonable for the parameters, and set the parameters of the prior distributions so that they appeared to be more or less flat in the range of the parameters which would fit the data (Morales 2013). They used a Weibull distribution to model the step lengths and the Wrapped Cauchy distribution for the turning angles. Four MCMC chains were fitted using WinBUGS with 20 000 iterations. They found that the DIC values normally ranked the 3-state or switch-constrained models in the top three models for each animal. The switch constrained model also performed well when the autocorrelation structure of the model was compared to that of the observed data.

Haydon et al. (2008) used radio-collar data from the same elk herd as used in the study above (Morales et al. 2004), although they looked at the movement of all of the reintroduced elk and investigated how all the animals moved relative to one another (Haydon et al. 2008). They called these "social informed" movement models since they take into account the group structure of a population and the movement of individuals within that population. They used fine scale GPS collar data to parameterize the movement models, while the group structure was determined from VHF-collar data from all of the elk which were reintroduced. A two-state correlated random walk model with fixed switching probabilities, called the "Double Switch" above (Morales et al. 2004), was used to model the movement of a solitary elk. Once again a Weibull distribution was used for the

step lengths and a Wrapped Cauchy distribution for the turning angles. The models were fitted using MCMC methods. An animal was defined as solitary if no other collar was located within 2km of the animal within four days either side of the GPS fix. The status of every animal was recorded in terms of whether it was part of a group or not, and if so, what the size of the group was. The authors assumed that the two-state movement of grouped animals would be the same as that of solitary animals, but that the switching rates between states would differ depending on the group status of the individual. They found that elk move further and have a higher mortality rate when they are solitary than when they are part of a group. All models were fitted using WinBUGS, following the fitting process of (Morales et al. 2004).

Jonsen et al. (2003) simulated data similar to a marine turtle to illustrate the application of state-space models using Bayesian methods. They assumed that the movements in a north-south and east-west direction were independent and drawn from a normal distribution and were related to the sea surface temperature. They included measurement error in their simulation since marine tracking data from an ARGOS system (http://www.argos-system.org) does include measurement error. Non-informative priors were selected for the model parameters, while an informative prior was selected for the measurement error, where in a real situation there is likely to be information about the expected observation errors. The authors used WinBUGS software to fit the models, and defined a hierarchical structure in order to gain insights into the population movement by combining data from individuals through the use of hyper-priors.

Jonsen et al. (2005) used a similar approach as that in Jonsen et al. (2003) and applied it to the movement tracks of hooded (*Cystophora cristata*) and Grey (*Halichoerus grypus*) seals. They independently modelled the distribution of the measurement error of the longitude and latitude observations, using *t*-distributions with *a priori* parameters obtained from independent research. The authors used

non-informative (uniform, Wishart or Beta) priors for all the parameters. They found that the analysis worked well to reduce the effect of extreme, and most likely erroneous, observations. The results showed that a double switching model provided the best fit for two of the three individuals, while the simpler double model was the better model for the third individual, based on investigating the overlap of the 95% credible limits of the estimated parameters.

Jonsen et al. (2006) then used hierarchical Bayesian state-space models to test the hypothesis that leatherback turtles (*Dermochelys coriacea*) off the coast of Nova Scotia, Canada, travel faster during the day when the turtles are closer to the water surface, than at night when they dive to greater depths. A secondary analysis was used to determine whether differences could be seen in travel rates between males and females, and between those who were breeding and those who were not. The authors used a fully Bayesian state-space approach which enabled them to combine individual results in order to analyse among-individual variation in movement rates, following the methods in Jonsen et al. (2003; 2005). The hierarchical structure is specified by placing vague hyper-priors on the priors of the model parameters (Jonsen, Myers & James 2006). The joint posterior distribution of the model parameters were estimated using Monte Carlo Markov Chain methods within the WinBUGS software. 35 000 iterations were generated, for two Markov chains. The burn-in removed the first 20 000, and then every fifth sample thinned to reduce autocorrelation which resulted in two chains with a total of 6 000 samples of the posterior distribution. The authors analysed their results in different segments of the migratory cycle. They found that the diel travel rates of the turtles, both male and female or breeding, changes over different phases of the migratory cycle.

In a follow-on study, Jonsen et al. (2007) applied what they termed a Switching State-Space Model (SSSM) to the same leatherback turtle data discussed in Jonsen et al. (2006). In addition to the location data used in the above study, dive

data were included in the study. The variables were time at depth, time at temperature, maximum dive depth and dive duration. These variables did not correspond exactly to the location data, but were rather binned through various times of the day. The patterns in the diving behaviour were compared to the behavioural state estimate obtained from the SSSM, which has the same model specification as the double switching model of Jonsen et al. (2005). The authors used some vague and some informative priors to fit the model, and the analysis was run using WinBUGS. The MCMC chains contained 30 000 iterations, 10 000 removed as burn-in and thinned to 4 000 by retaining every fifth observation to remove autocorrelation. The authors compared the predicted states from the model to the dive statistics to investigate the ability of the model to predict behavioural dynamics (Jonsen, Myers & James 2007).

Blackwell (2003) used radio-tracking data from a mouse to implement a Bayesian analysis of a bivariate Ornstein-Uhlenbeck process. This is considered a 'continuous-time and continuous-space' model (McClintock et al. 2012). The input data consisted of 10-minute interval observations of the mouse's location and a classification of its behaviour at the time of the location recording. The behaviour was classified as either resting, feeding or running and the behavioural decision was based on the strength and position of the radio signal (Blackwell 2003).

Eckert et al. (2008) used a combination of two hierarchical Bayesian state-space models to analyse the location data obtained from ARGOS satellite-tags on 15 juvenile loggerhead turtles (*Caretta caretta*) in the western Mediterranean Sea. They used the analyses to estimate the movement pathways, and to assign each location to one of two behavioural states. These states were then compared with environmental features to gain insight into the behavioural patterns of the turtles. The longitude and latitude locations were estimated during the first phase of the modelling, to account for the observation error associated with ARGOS data. In

the second phase, the movement rate (km/day) and turning angle calculated from the estimated locations were used as the model input. In the second phase of the modelling process, the framework used by Morales et al. (2004) was used to model switches in the behavioural state based on turtle size and environmental covariates. The authors used a two-state behavioural model, and they inferred these states to represent an intensive search or foraging state and an extensive search or exploratory state. A Wrapped Cauchy distribution was used to model the turning angle, while a normal distribution was used for the movement rates. The decision to use the normal distribution, as opposed to the Weibull distribution used in the Morales et al. (2004) framework, was based on the fact that the normal distribution parameters are easier to constrain in order to ensure that the mean movement rate of one state is constrained to be slower than that of the other state (Eckert et al. 2008). The authors used a hierarchical Bayesian framework so that they could directly model the individual heterogeneity in terms of model parameters while population parameters for the turtles were included using the hyper-parameters. Eckert et al. (2008) used the WinBUGS software to fit the models, with phase one having two MCMC chains, each with 20 000 iterations. The first 15 000 iterations were discarded and the chains thinned by ten to reduce autocorrelation. For the second modelling phase, the chain length was 35 000, with 10 000 discarded and the chain thinned by fifty. The posterior distributions were therefore based on 1 000 independent samples. The results of their analysis showed that the turtles in the Balearic Sea in the western Mediterranean were more likely to exhibit behaviour classified as foraging, while only bigger turtles responded to variations in sea surface height (Eckert et al. 2008).

McClintock et al. (2012) used a similar framework to Morales et al. (2004), to fit generalized multi-state random walk models to the GPS tracking data for a Grey Seal in the North Sea. They also used the Weibull distribution to model the step lengths and a Wrapped Cauchy distribution for the turning angles, and they assumed that these two measures were independent of one another. They used non-informative uniform priors for the displacement and direction distribution

parameters. A reversible-jump Markov chain Monte Carlo method was used to fit the models, and the models were evaluated using the posterior model probabilities. McClintock et al. (2012) acknowledge the difficulties of assessing the absolute goodness-of-fit of mechanistic movement models. They used the posterior model probabilities to provide some assessment of the relative goodness-of-fit of the models. The authors fitted the model using one Markov chain which needed five million iterations before convergence could be obtained. The transition probability results showed that the seal had a high probability of remaining within its current movement state. And they found strong evidence to support using drift or bias towards centres of attraction, which were far better at explaining the seal movement than simple or correlated random walks.

Many of the Bayesian applications to animal movement have been done using marine tracking data which can have larger errors associated with the locations obtained from the GPS. Bayesian fitting approaches can be beneficial in these cases if informative prior distributions are used for the error distributions and this is an area for potential further study. In most cases uninformative prior distributions are used, which means that the results are likely to be similar to results that could have been obtained using frequentist methods with the data dominating. However, few studies have compared the results between Bayesian and frequentist approaches. This is investigated in Chapter 8. Bayesian state-space models have been applied in a variety of different ways, including models with covariates and a hierarchical structure. A number of the studies have been based on previous studies, using either the same movement data or the same model structure. A number of the studies make mention of the challenges associated with assessing the goodness-of-fit and the convergence of the chains. Another area for investigation in this thesis is the effect that the time scale of the input data has on which distribution is best suited to being fitted to the movement data. The desired outcomes of Hidden Markov models and Bayesian state-space models applied to animal movement are often very similar. How the results and the inferences

obtained from each of the suite of models differs from one another needs to be investigated.

7.4 Methods

The models used by Morales et al. (2004) were adapted for this study of the sable antelope, buffalo and zebra data. These models were chosen as they provide a simple structure consistent with the HMM framework but using a Bayesian fitting approach. Once again, only the distances between successive locations were used as an input to the models in order to be consistent with the work done using Independent Mixture Models (IMMs) and HMMs presented in this thesis. As discussed in Section 6.3.2, the models were all tested including the turning angle in the fitting process and this was found to not increase the state allocation accuracy much, but did increase the complexity of the model for fitting. Similar to the Mixture and Hidden Markov models, the error process associated with the GPS observation was not included in the modelling process. At the hourly and daily time scale, the error which is assumed to be random and miniscule compared to the distance moved during a day is not expected to have an effect on the modelling process. This randomness is accounted for through a Markov process as an independent identically distributed random variable error process. The GPS observation error was discussed in more depth in Chapter 2. The Morales et al. (2004) and Jonsen et al. (2005) models are examples of 'discrete-time and continuous-space' models (McClintock et al. 2012). The models from Morales et al. (2004) adapted to use only the GPS location data and no habitat covariates were used, consistent with the models run in the previous chapters. Habitat variables have been excluded from the entire thesis as other studies had been initiated to investigate the effect of the habitat on the movements of these three species (le Roux 2010). However, as discussed in previous chapters, the IMMs and HMMs as well as these Bayesian state-space models can be extended to include covariates such as the habitat or season. Models were fitted using either the Weibull or Gamma distributions to model the step length, with non-

informative uniform prior distributions specified for the distribution parameters. These two fitted distributions for the displacement distances (Gamma and Weibull) were selected as they have been used in other animal movement studies (Morales et al. 2004, Codling, Bearon & Thorn 2010, Dragon et al. 2012, Beyer et al. 2013), have a similar shape and were also shown to be the most suitable distributions for modelling the daily movement data using HMMs in the previous chapters. Only the log-normal distribution was used as an example of fitting the Bayesian SSMs to hourly data. Since the framework used here to showcase the Bayesian fitting approach is very similar to the HMM framework, the same distributions are likely to perform as they did using the frequentist fitting approach. Since non-informative prior distributions are used, the results are expected to be close to those of the frequentist approach.

In terms of the model specification, the underlying movement using displacement only is modelled as a random walk. A correlated random walk is used if there is directional persistence and the turning angle between successive locations is included in the modelling process. The 'Single' model is a one-state model, where the observed movement is assumed to have been generated from a single movement state. The 'Double' model is essentially a Mixture Model whereas the 'Double with Switch' and 'Triple with Switch' correspond to a 2- or 3-state Hidden Markov process which formally incorporates the probability of switching states into the model fitting process. At the daily scale, it was expected that the underlying states would correspond to behaviours such as "encamped" with short movements and "roaming" with long movements (Langrock et al. 2012), with a third state splitting the roaming state into shorter and longer displacements. The hourly activity states are expected to be very similar to those found using the IMM and HMMs corresponding to resting, foraging, local movement and relocation.

For the Single model using the Weibull distribution to describe the step length, with shape parameter *a* and scale parameter *b*, the likelihood function is given by:

$$L(a,b) = \prod_{i=1}^{T} \left(\frac{a}{b}\right)\left(\frac{t_i}{b}\right)^{a-1} e^{-\left(\frac{t_i}{b}\right)^{a}}$$

Similarly for the Gamma distribution, using the same notation for the shape and scale parameters, the likelihood is given by:

$$L(a,b) = \prod_{i=1}^{T} \left(\frac{1}{b^{a}}\right)\left(\frac{1}{\Gamma(a)}\right) t_i^{\,a-1} e^{-\left(\frac{t_i}{b}\right)}$$

The joint distribution, with $\varphi(a)$ and $\phi(b)$ denoting the prior distributions for the parameters *a* and *b* respectively, is given by:

$$f(a,b \mid Data) = \frac{L(a,b)\varphi(a)\phi(b)}{\int_0^{\infty}\int_0^{\infty} L(a,b)\varphi(a)\phi(b)dadb}$$

where *a* and *b* are assumed to be independent *a priori*. And the posterior distribution for a Weibull model is given by:

$$f(x, Data) = \int_0^{\infty}\int_0^{\infty} f(x,a,b) f(a,b \mid Data)dadb$$

$$\text{where } f(x, Data) = \left(\frac{a}{b}\right)\left(\frac{x}{b}\right)^{a-1} e^{-\left(\frac{x}{b}\right)^{a}}$$

For the Double and Triple model, the parameters *a* and *b* will be vectors of parameters for the different movement states, rather than single parameter values for the single model. All models were initiated and specified in R, using the packages 'R2WinBUGS' (Sturtz, Ligges & Gelman 2005) and 'coda' (Plummer et al. 2006) to run the analysis in WinBUGS from R and to analyse the output from WinBUGS in R. As with the previous analyses (Chapters 3 and 5) missing data were included in the input data when GPS locations were missed. WinBUGS

can include missing data in the model fitting process which is advantageous for animal movement studies since this is a common feature of the data. The details of the percentage of missing data for each herd were shown in the Appendix Section 5.16 Table 5.7. All models were fitted using non-informative priors and randomly varying starting values for the parameter estimates for each chain. WinBUGS uses non-standard notations for some of the built in distribution functions (Press 2003) which need to be carefully checked before using them in an analysis. For example, the Weibull distribution function in WinBUGS uses a scale parameter which R refers to as the 'rate' parameter and it is actually the inverse of the scale parameter used by R. The normal, log-normal and Gamma distributions definitions are also non-standard in WinBUGS. If used without checking, these unusual definitions of the probability density functions can cause errors to occur in the analysis and interpretations of the results.

To determine the sensitivity of the model output to the choice of the prior distributions used in the analysis, Gamma, normal and uniform prior distributions were tested as prior distributions for all of the fitted parameters. The parameters were fitted using Gamma or Weibull distributions to model the daily displacements. Lambert et al. (2005) found that in general the shape parameter was not affected much by the choice of the prior distribution. However the scale parameters were more sensitive to the choice, particularly in the case of small observed datasets. With a large amount of observed data, the choice of non-informative prior distribution was less important (Lambert et al. 2005) and the data will dominate.

For all models, three chains were fitted with 50 000 iterations. The default burn-in value of half the number of iterations was used (25 000) to remove the effect of the starting values (Press 2003), as well as the thinning rate default setting so that 1000 independent samples were collected for each chain. The Gelman-Rubin statistic was calculated and plotted for all fitted variables to investigate the

convergence of the Markov chains (Gelman, Rubin 1992). The `gelman.diag` function in the coda package in R was used, and if this scale reduction statistic was close to one it is concluded that the simulated values are close to the target distribution (Gelman, Rubin 1992) and that the Markov chain has converged. The DIC was used as an indication of the goodness-of-fit along with other posterior checks.

Daily movement data were used in the Bayesian state-space model analyses, as a natural comparison to the model results from the daily HMM analyses using the Gamma and Weibull distributions. The movements represent different behaviours depending on the time of day at which they are anchored. This was discussed further in Section 5.9. Models using the log-normal distribution were fitted to the hourly data using WinBUGS, as an example to compare the results to the hourly IMMs and HMMs. This provides a comparison of the Bayesian and frequentist approaches to fitting these types of model which will be investigated further in Chapter 8. Table 7.1 shows the prior distributions used for all parameters in the models using either the Weibull or Gamma distributions to model the step lengths. a_i and b_i represent the rate and shape parameter for both distributions fitted to the step lengths. An additional parameter (eps_i) is added to the rate parameter for the second state, in order to ensure that one of the movement states is slower and the other is faster. The transitions between states for the 'switch' models are denoted by q_i while η_i is the probability of being in state i. The prior distributions and their parameters were chosen to be similar to work done in the literature, where the Morales et al. (2004) study has been a commonly referenced study. These parameters for the prior distributions were selected by Morales et al. (2004) so that they appeared "flat" in the expected region of parameters which would fit the observed data (Morales 2013). However, due to the large dataset sizes used in this study, it is expected that the likelihood will dominate and the prior distributions will have a minimal effect on the posterior distributions (Box, Tiao 1973). Similar prior distributions have been used in other studies which adapted the Morales et al. (2004) code for their own studies.

Table 7.1 - Prior distributions used to fit Bayesian state-space models with either a Weibull or Gamma distribution for step lengths

Parameter	Prior Distribution	Description
a_i	uniform(0,10)	Rate parameter for the Weibull or Gamma distribution
eps_i	normal(0,0.01)	Difference between a_i and a_{i+1} when multiple walks fitted ($a_{i+1} = a_1 + eps_i$)
b_i	uniform(0.4,8)	Shape parameter for the Weibull or Gamma distribution
$\eta_{1,t}$	uniform(0,1)	Probability that the t-th observation is in State 1
q_{ij}	uniform(0,1)	Transition probability from the i-th to j-th movement state

7.5 Results – Daily Data

Single, double, switch and triple switch models were run for the Pretoriuskop Phabeni sable herd locations recorded at 8pm.

Table 7.2 shows the fitted parameter estimates for the three simpler models (Single, Double and Switch) using either uniform prior distributions for all parameters, or using some gamma and some normal prior distributions for the parameters. This tested the sensitivity of the models to the choice of prior distribution for the parameters of both the Weibull and the Gamma state-space models. The test was done since in some cases even though a non-informative prior is used, it can have an influence on the results, especially for smaller studies (Lambert et al. 2005). In contrast to the results presented for the Hidden Markov models, the parameters reported for the Bayesian models are the shape (b) and rate (a) parameters, where the rate parameter is the inverse of the scale parameter. For the simpler Single and Double models fitted using either the Weibull or Gamma distributions to model the step-lengths, the estimated parameters are almost identical using any of the prior distributions for the fitted parameters. This

means that the results are not influenced by the choice of the non-informative prior, and they are being driven by the data and the model only. For the more complex Switch model, there are some slight differences in the results, particularly when using some Gamma and some Normal prior distributions on the fitted parameters. However for these Gamma or Weibull movement models, it seems that the choice of the prior distribution does not influence the results too much, and they are not providing meaningful prior information to the model as required. Since the prior distributions are indeed non-informative, all the results presented hereafter are models fitted using uniform prior distributions for the fitted parameters as specified in Table 7.1.

A summary of the deviance and DIC values for each model, as calculated by WinBUGS, as well as the posterior distribution model parameters are shown in Table 7.3. The DIC values increase rapidly as the complexity of the models increase, with the Single model having the lowest DIC for both the Gamma and Weibull models. However, there is some debate about how the DIC should be calculated and its use in terms of model selection (Gelman 2011). The issues raised include the slow convergence of the DIC which requires a very large number of simulations, beyond the number required for chain convergence for the fitted parameters; various definitions for the effective number of parameters and how to select the correct loss function (Gelman 2011). This is probably an area of Bayesian models applied to animal movement which requires more research and more reliable measures of goodness-of-fit and model selection. For this reason, other posterior checks were done to investigate the states associated with the posterior distributions and to compare these to the biology of the problem in order to assist with the decision about the most suitable model.

The model parameters for the fitted Bayesian Models are shown in Table 7.3 and Table 7.4 for the Phabeni and Numbi sable herds respectively. The deviance and DIC values as calculated by WinBUGS are reported, as well as the shape and rate

parameters of the Weibull and Gamma distributions. Note that in the HMM chapter, the scale parameter was reported not the rate parameter which is used by WinBUGS for these models. The standard errors are also given. The values q1, q2 and q3 are the transition probabilities. As with the HMMs fitted to the daily data, the probability of remaining in the current state (diagonal values of the transition matrix) is high (0.69 – 0.92), and has an associated low probability of transitioning to the other state(s) across all the switching models. The expected values (or first moment) for the posterior distributions are also shown in Table 7.3, and are similar to the HMMs for the daily data. The states appear to follow the usual interpretation of "encamped" and "relocated" which have been found in other studies (Beyer et al. 2013, Morales et al. 2004, Haydon et al. 2008, Langrock et al. 2012). As was suggested in the HMM chapter, the states appear to be somewhat associated with the time of the year, with a seasonal shift from one state to another. For the Weibull Double model, State 1 has an expected value of 0.43km and State 2 an expected value of 0.99km. For the Double with Switch models the states have more extreme expected values of 0.28km and 1.07km respectively. Potentially the more separated states are due to the transitions between the states now being formally included in the model. The 3-state model with switching probabilities has states with expected values of 0.15km, 0.54km and 1.63km respectively. The standard errors for the Weibull models are mostly slightly smaller than for the Gamma models. The results for the Numbi herd are similar to the Phabeni herd with high probabilities of remaining within a state and states that corresponded to encamped and relocating behaviour and are consistent with the results obtained for the HMMs. Only the triple with Switch model converged to different states using the Gamma and Weibull distributions; however these models struggled to converge and were very sensitive to the choice of starting parameters. Once again the standard errors are very slightly smaller for the Weibull models. These patterns were consistent for all of the herds, not just the two sable herds presented in Table 7.3 and Table 7.4. From Figure 7.1 it is clear that State 1 is associated with the shorter movements and State 2 has a much fatter tail to account for the longer relocation movements. The probability of remaining within a state is very high (0.87 and 0.89 respectively), which is

consistent with the high diagonal transition probabilities obtained from the HMMs. The mean displacement calculated from the observations allocated to each state provides a similar picture, with state 1 having a mean displacement of 0.39km and state 2 a mean of 1.05km. As with the hidden Markov models, these states are inferred to correspond to the animal remaining within the same area or relocating its area of utilisation.

To investigate the convergence of the Markov chains, Figure 7.4 and Figure 7.5 show a plot of the three chains for the fitted distribution parameters and the transition probabilities for the Weibull Double with Switch model for the Phabeni herd. While there is considerable visual variation in each of the chains, the Gelman-Rubin statistic was very close to one for all of the model parameters. This was true for all of the parameters across all of the models, fitted using either the Gamma or the Weibull distributions for the step-lengths. This indicates that the chains for each of the parameters have converged with 50 000 iterations used. In some cases, using fewer iterations could result in the chains not converging. This is a trade-off though, since more iterations means more time is needed to fit the models. While this is not as problematic for the daily data, for input of the size of the hourly datasets, running a large number of iterations for multi-state models was prohibitively time consuming.

Table 7.2 – Phabeni Sable: Fitted model parameters for Gamma and Weibull state-space models using various prior distributions (a = rate parameter; b = shape parameter; DNC = did not converge)

		Fitted Weibull Distribution		Gamma Distribution	
	Parameter	Uniform Prior	Gamma/Normal Prior	Uniform Prior	Gamma/Normal Prior
Single	a	1.15	1.15	1.73	1.74
	b	1.28	1.28	1.33	1.33
Double	a1	2.11	DNC	3.22	3.14
	a2	0.97		1.29	1.29
	b1	1.47		1.83	1.77
	b2	1.13		1.24	1.25
Double with Switch	a1	3.27	3.21	3.26	3.54
	a2	0.86	0.85	2.08	1.80
	b1	1.55	1.54	2.57	1.66
	b2	1.28	1.29	2.08	2.33
	q1	0.87	0.87	0.81	0.87
	q2	0.11	0.12	0.18	0.22

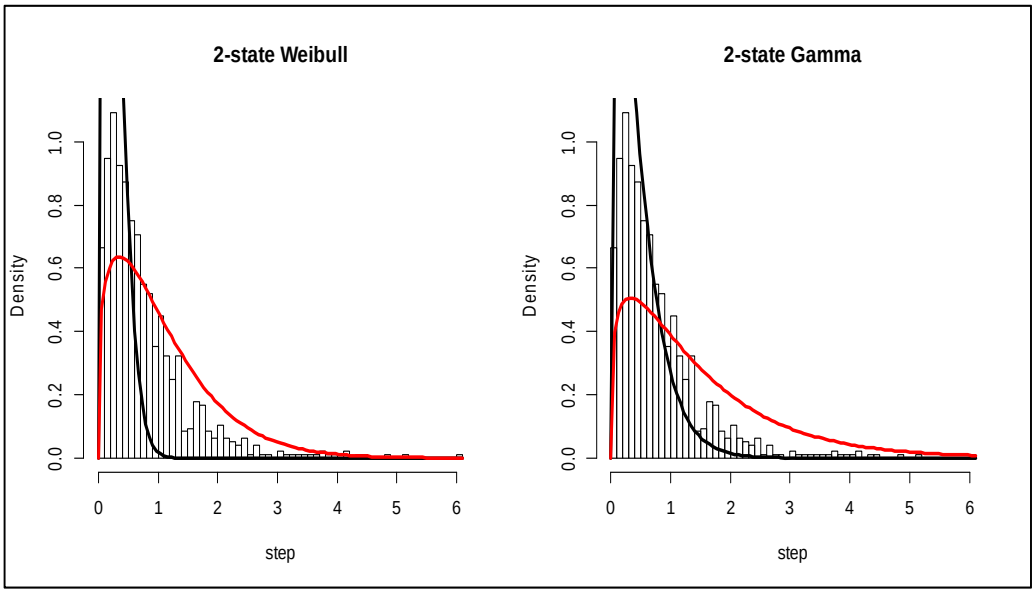

Figure 7.1 – Phabeni Pretoriuskop 8pm Double with Switch Weibull SSM and Gamma SSM – State 1 (black), State 2 (red)

The Double with Switch model was selected as the model which is consistent with the assumptions and inferences from the 2-state HMM. Both the Gamma and Weibull SSMs provided quite similar underlying states based on the inferences drawn when examining the fitted distributions in Figure 7.1, their expected values and the observations allocated to each state. If there was a large difference in results based on the two distributions, this would indicate that the choice of the underlying distribution had a dramatic effect on the interpretations which could be drawn from these types of models and reasons for this would need to be investigated. Since the probabilities of remaining within a state are very high, each state persists for a while before transitioning to the other. The state allocation by month of the year is shown in Figure 7.2. There is a clear increase in the occurrence of the relocating state during the late winter months which confirms that the animals are moving more regularly, searching for resources during this tougher time of the year. This clear behavioural change occurs around July which is at the end of the Early Dry moving into the Late Dry season. A reasonably similar pattern of state allocation was found for the same Double with Switch model fitted using Gamma distributions for the step lengths, shown in Figure 7.3.

Table 7.3 – Phabeni Pretoriuskop Sable: Weibull and Gamma SSM - fitted model results

	Deviance WinBUGS	DIC WinBUGS	Shape	Shape SE	Rate	Rate SE	q1	q1 SE	q2	q2 SE	q3	q3 SE	Expected value
	Weibull distribution for step length												
Single	1384.38	1386.29	1.15	0.03	1.28	0.04							0.75
Double - State 1	1221.34	2066.24	1.47	0.09	2.11	0.25							0.43
Double - State 2			1.13	0.07	0.97	0.08							0.99
Double Switch - State 1	1090.72	1857.28	1.55	0.10	3.27	0.56	0.87	0.04	0.13				0.28
Double Switch - State 2			1.28	0.09	0.86	0.12	0.11	0.06	0.89				1.07
Triple - State 1	914.88	2400.50	1.52	0.12	6.04	1.26	0.82	0.05	0.04	0.03	0.14	0.04	0.15
Triple - State 2			1.61	0.10	1.67	0.16	0.02	0.01	0.93	0.03	0.06	0.03	0.54
Triple - State 3			1.43	0.14	0.56	0.13	0.12	0.05	0.11	0.07	0.77	0.08	1.63
	Gamma distribution for step length												
Single	1371.37	1373.41	1.33	0.05	1.73	0.09							0.77
Double - State 1	1246.03	2316.88	1.83	0.24	3.22	0.42							0.57
Double - State 2			1.24	0.19	1.29	0.16							0.97
Double Switch - State 1	1007.85	2614.54	2.57	1.33	3.26	0.54	0.81	0.10	0.19				0.79
Double Switch - State 2			2.08	0.55	2.08	0.44	0.18	0.09	0.82				1.00
Triple - State 1	856.02	4225.06	2.40	1.39	5.64	1.55	0.78	0.12	0.08	0.09	0.14	0.08	0.43
Triple - State 2			2.00	0.45	3.64	0.64	0.04	0.05	0.88	0.08	0.08	0.06	0.55
Triple - State 3			2.60	0.72	1.73	0.31	0.15	0.11	0.14	0.09	0.71	0.13	1.50

Table 7.4 – Numbi Pretoriuskop Sable: Weibull and Gamma SSM - fitted model results

	Deviance WinBUGS	DIC WinBUGS	Shape	Shape SE	Rate	Rate SE	q1	q1 SE	q2	q2 SE	q3	q3 SE	Expected value
	Weibull distribution for step length												
Single	1633.33	1635.30	1.00	0.03	0.73	0.03							1.37
Double - State 1	1381.89	2754.93	1.41	0.09	1.58	0.24							0.58
Double - State 2			1.15	0.08	0.41	0.06							2.31
Double Switch - State 1	1401.54	1640.63	1.63	0.14	2.88	0.50	0.93	0.03					0.31
Double Switch - State 2			1.12	0.04	0.52	0.04	0.03	0.01					1.86
Triple - State 1	1327.26	1879.83	1.76	0.17	3.83	1.00	0.90	0.05	0.06	0.05	0.04	0.03	0.23
Triple - State 2			1.72	0.73	0.83	0.33	0.10	0.12	0.71	0.24	0.19	0.19	1.07
Triple - State 3			1.10	0.08	0.43	0.08	0.02	0.02	0.12	0.13	0.87	0.13	2.22
	Gamma distribution for step length												
Single	1630.77	1632.75	1.09	0.05	0.80	0.05							1.37
Double - State 1	1362.61	3170.67	1.75	0.18	2.74	0.46							0.64
Double - State 2			1.42	0.23	0.67	0.09							2.11
Double Switch - State 1	1406.22	1625.49	2.16	0.27	4.41	0.75	0.94	0.02					0.49
Double Switch - State 2			1.26	0.09	0.73	0.05	0.03	0.01					1.74
Triple - State 1	1279.09	2238.82	2.03	0.24	1.66	0.24	0.72	0.15	0.22	0.13	0.06	0.03	1.22
Triple - State 2			6.59	0.87	1.53	0.15	0.71	0.11	0.22	0.08	0.07	0.08	4.31
Triple - State 3			1.17	0.10	1.48	0.15	0.04	0.02	0.02	0.02	0.94	0.02	0.79

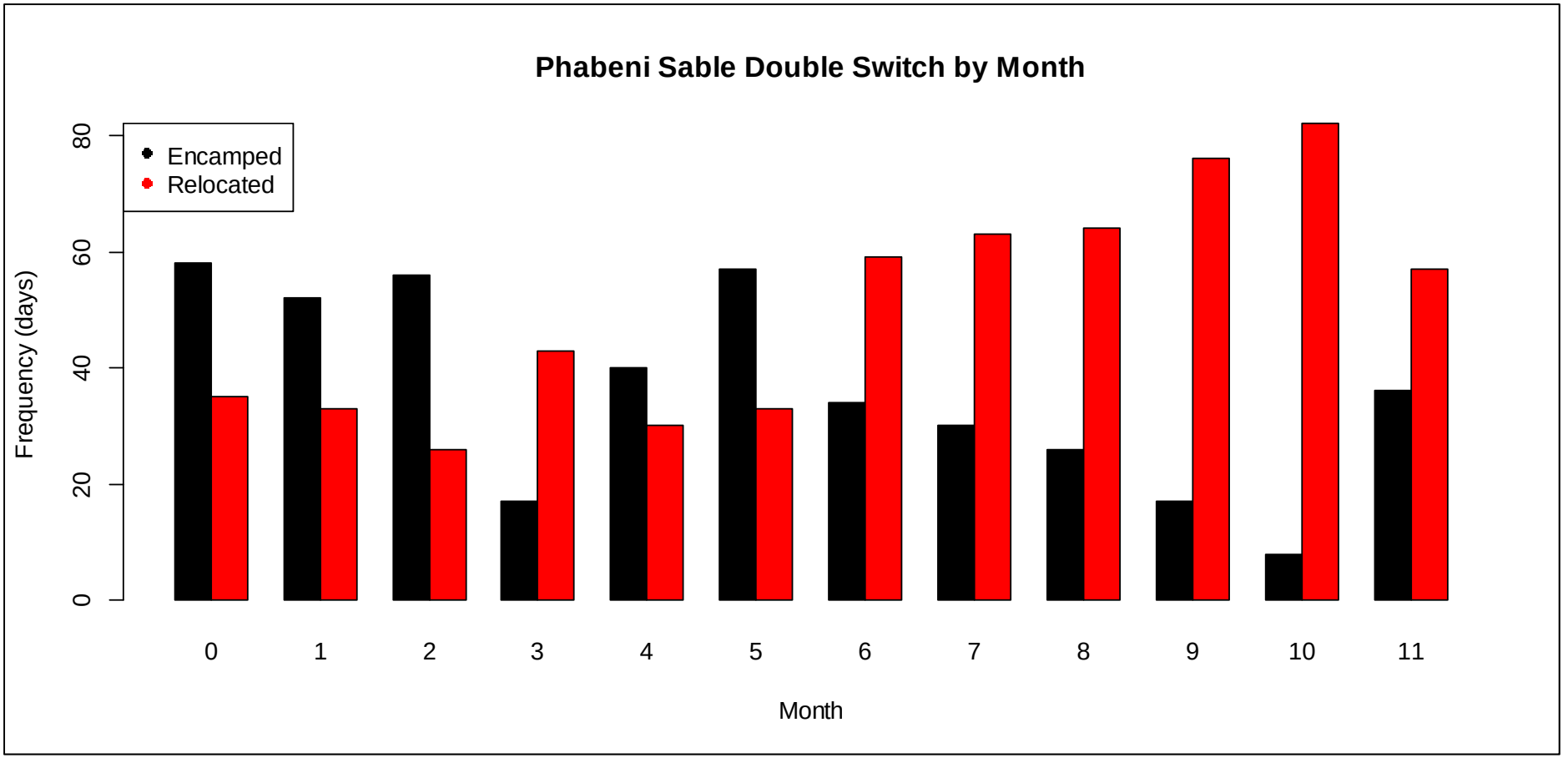

Figure 7.2 - Phabeni Sable - State allocation by month of the year - Double with Switch model (Weibull)

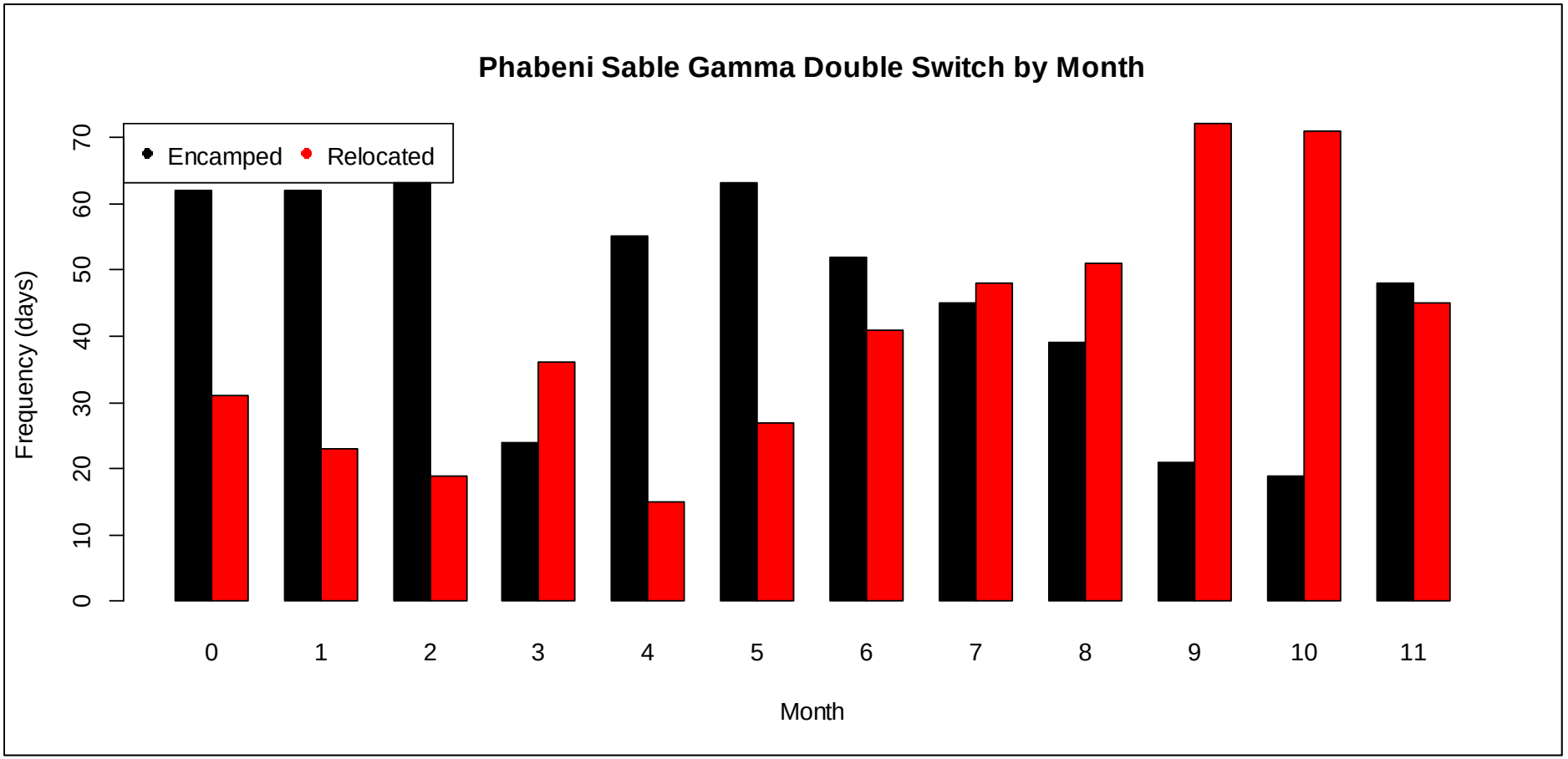

Figure 7.3 - Phabeni Sable - State allocation by month of the year - Double with Switch model (Gamma)

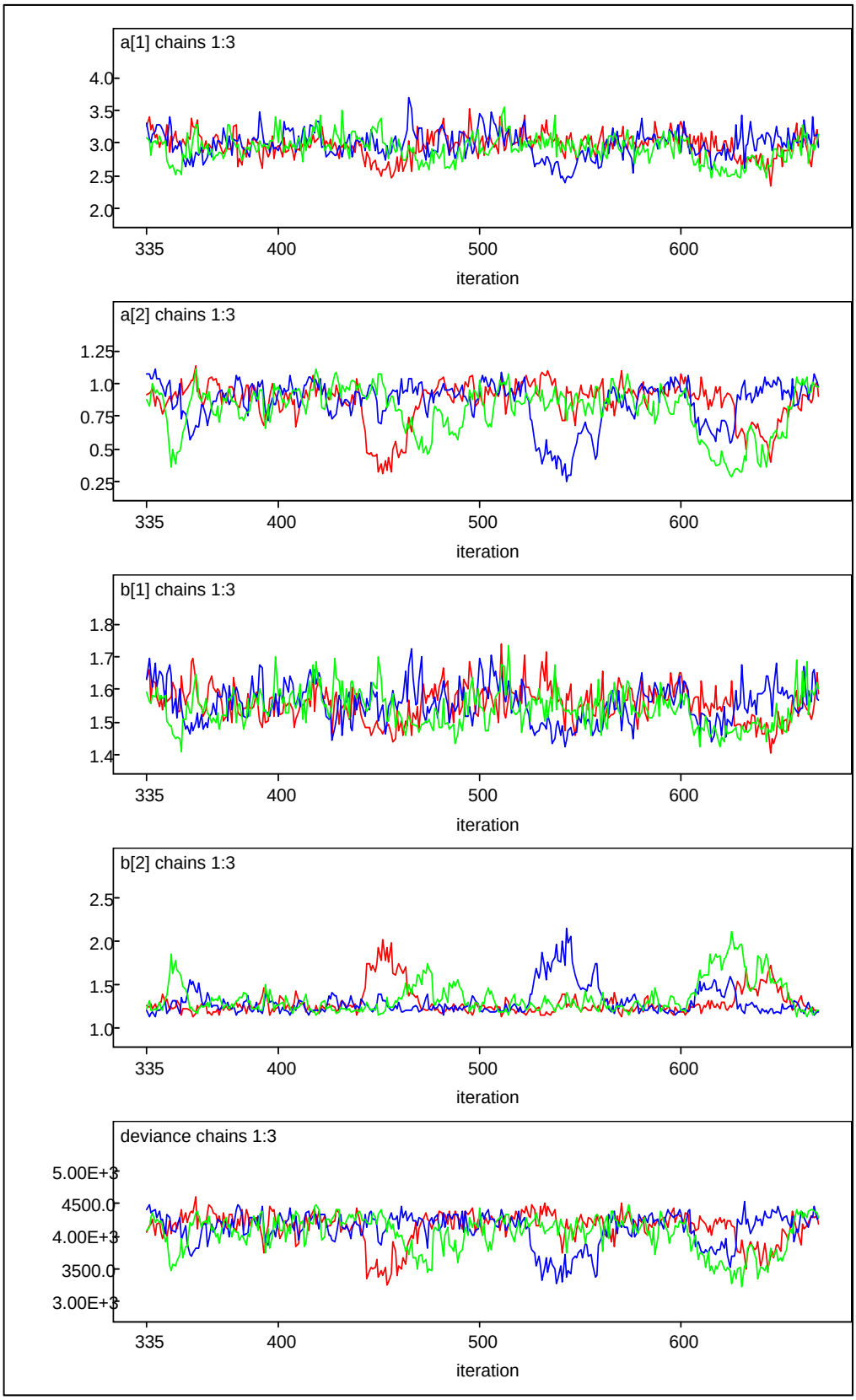

Figure 7.4 – Phabeni sable Markov chain samples for Switch model parameters with different colours indicating chains fitted using different starting values

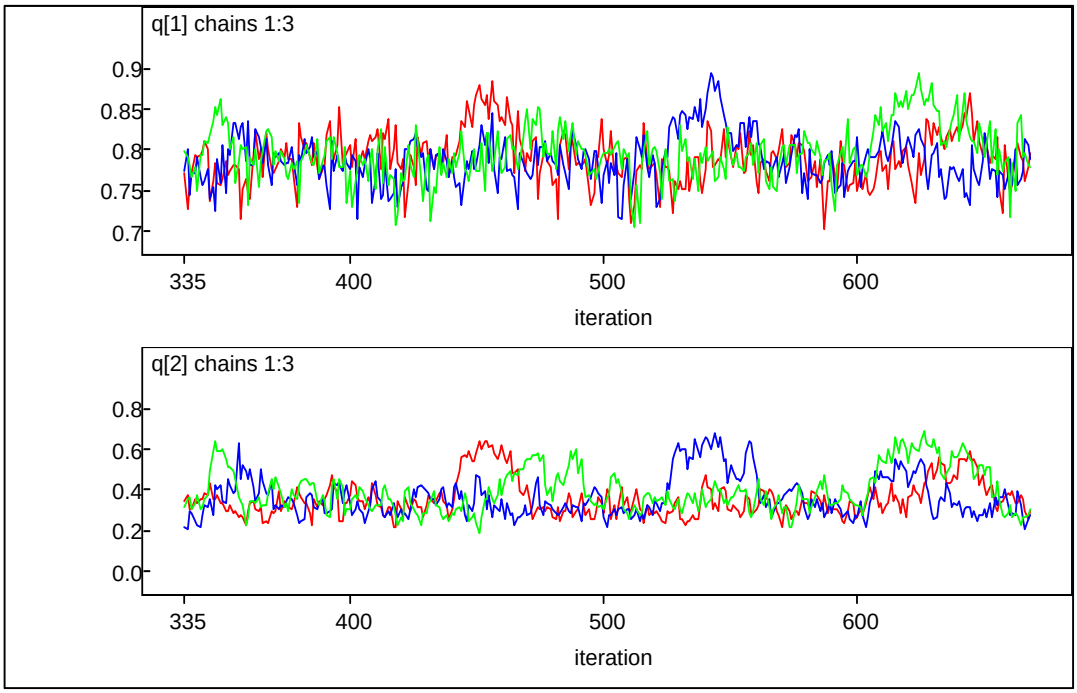

Figure 7.5 – Phabeni sable Markov chain samples for Switch model parameters - transition probabilities with different colours indicating chains fitted using different starting values

7.6 Results – Hourly Data

Log-normal models were fitted to the hourly data in order to compare the method with the IMMs and HMMs. The WinBUGS software uses an unusual version of the density function, with parameters μ and τ, where τ is the inverse of the square root of the σ parameter fitted in the IMM and HMM chapters. Fitted parameters were almost identical for models fitted using the Bayesian or Hidden Markov approaches for all the hourly models. This confirms that although the model fitting processes are different between the frequentist and the Bayesian approaches, the results when non-informative priors are used will be almost identical. However, the time taken to fit the Bayesian models was far longer. The hourly datasets are much larger than the daily datasets, and this exaggerated the time taken to fit the Bayesian models. Also, for the more complex triple model, the size of the dataset actually made it almost impossible to run a large number of iterations in order to obtain convergence of the chains. And some statistics such as

the Gelman-Rubin statistic could not be calculated in R due to the size of the dataset requiring too much computation causing the software to crash. This problem could be overcome by using a more powerful computer or running models on subsets of the data, however it is a drawback of the method. This is in contrast to the 3-state HMMs, which although they took longer to fit than the 2-state, it was not a particularly time consuming process for the hourly models and were also extended to 4-state models without computational problems. The results of the hourly log-normal models for the Punda Maria sable herd are shown in Table 7.5.

Table 7.5 - Log-normal SSM - fitted model results for PM_ Sable279 (Hourly Data)

	Deviance WinBUGS	**DIC WinBUGS**	μ	τ	$\sigma = 1/\sqrt{\tau}$	**q1**	**q2**	**Expected value**
		Log-normal distribution for step length						
Single	-2350.70	-2348.81	-2.32	0.33	1.75			0.45
Double - State 1	-5398.31	-493.22	-3.69	0.80	1.11			0.0
Double - State 2			-1.01	0.79	1.12			0.69
Double Switch - State 1	-5105.94	-857.49	-3.18	0.54	1.36	0.85	0.15	0.11
Double Switch - State 2			-0.48	1.40	0.84	0.32	0.68	0.88

7.7 Discussion

In comparison to the HMMs and the IMMs, the Bayesian SSMs are far more time consuming models to fit. The WinBUGS software is reasonably tricky to start using and error messages can be unhelpful. The R package 'R2WinBUGS' however does make it possible to initiate the model fitting from the R environment and also to do post-processing and analysis in R. The models were robust to the choice of starting values, provided that the initial values were reasonable and not extreme. The uniform distribution was found to be the most consistently applicable prior distribution for the fitted parameters although the fitted model parameter estimates were consistent despite the choice of prior distribution. This was expected since they were non-informative priors. For many analyses, the need to specify prior distributions can be an advantage to using Bayesian techniques to fit the model. However, if no suitable prior information exists, and non-informative prior distributions are selected, the observed data itself must dominate (Lambert et al. 2005, Box, Tiao 1973). This negates one of the benefits of a Bayesian approach. Many of the animal movement studies using a Bayesian approach use non-informative priors, in which case the results will be very similar to models fitted using the frequentist approach. In this case, the additional time and complexity of fitting Bayesian models makes this approach less attractive to researchers who are more comfortable with the classical methods. However, as many more animal movement studies are done, and more information is gathered about particular species, the scope for including informative priors should increase. This would make the Bayesian approach more desirable for animal movement modelling.

The complex nature of the analyses and the intricacies of the software to run these models, limits the accessibility of running them to those with statistical expertise rather than ecologists who are studying the movement patterns of an animal (Langrock et al. 2012, McClintock et al. 2012). One of the challenges for fitting these models is to confirm when convergence of the chains has been achieved or

identify when there are issues with the fitted model. There are likely to be advances in the software used to run these models which will make them more accessible to researchers, however the danger is that the software becomes a 'black box' with little understanding by the user of the processes being performed. A number of the papers that have been published which have applied Bayesian state-space models have published their code in Appendices or Supplementary Information. While this does make it a lot easier to get started with running these types of analyses, there are still challenges and complications. There is also the danger that these models are replicated without a sufficiently in depth understanding of the statistical analysis which is being performed in the background, the workings of the Gibbs sampler and the implication of the default settings.

The models were very time consuming to run, and the output from WinBUGS and R2WinBUGS is not particularly simple to work with. While the time taken to fit the models would be reduced for smaller datasets, it will still require a substantial amount of extra time in comparison to the HMM fitting approach. Error messages within WinBUGS were found to be quite cryptic and unhelpful. This makes getting used to the software more time consuming and frustrating for new users. For the hourly data, which had far larger datasets, the more complex models (Double Switch and Triple Switch) were extremely time consuming to fit. If the number of iterations were increased beyond 1000, more than three hours were needed to fit the model. Far more than 1000 iterations are needed to be more confident about the convergence of the Markov chains.

Advantages of the Bayesian approach include the flexibility of the models, the ability to cope with missing data, to include covariates and to formally include a hierarchical structure. It is these advantages and belief that they are powerful models which can reveal multiple influences which has made them popular in recent studies. However these are offset by the disadvantages of the complex

underlying fitting processes of the MCMC and the time required fitting the models. Determining the goodness-of-fit of the models is challenging, and the DIC should be used as a guide only. The pattern of autocorrelation of the observed and fitted data can also be used as a guide for model selection, in a similar way to what was done for the HMMs.

7.8 Conclusion

Bayesian analysis techniques are very useful for complicated model specifications and are becoming more popular as computing power increases and software improves. However they are complicated and are likely to be beyond the level of programming and statistical depth that most researchers are interested in. Depending on the level of complexity required from the model and the desired output, Bayesian state-space models will have their niche where they play an important role, particularly for complex hierarchical structures and multiple covariates with associations between them. In the situation where non-informative prior distributions are used, the Bayesian approach provides almost identical results to the frequentist methods, while being far more time consuming to run. This is explored further in the following chapter.

The feature that sets Bayesian methods apart from the frequentist approach is the potential for including informative prior information. In most cases so far in animal movement research this has not been the case and non-informative priors are used. However, as more movement research is done, there is the potential to collect informative data which could be included as prior information in future models. However, for models fitted using very large datasets which can be obtained using GPS collars, the volume of data are likely to overwhelm any informative prior information. The additional complexity and computing time of the Bayesian approach can be justified through the improvement in the modelling process by the inclusion of these informative prior data. The potential of building

hierarchical models which can allow for population level inferences to be drawn is a further advantage of the Bayesian approach. The results from these models provide realistic results when used to make inferences about the behaviour of the animal.

The Bayesian state-space approach provides a model framework to identify the latent states assumed to correspond to animal behaviour, based on the locations of the animal. The fitting process is challenging and time consuming, but the results provide predicted state allocations which can be used to infer the behaviour of the animal. The clear advantages of Bayesian analysis include the ability to extend the models to include covariates, a hierarchical framework and the ability to include informative prior information into the modelling process. As computing power increases, along with the ability to collect more covariate information including remotely sensed data, the popularity and scope for using Bayesian analysis will increase.

8 MODEL COMPARISON – STATE-SPACE MODELS OF ANIMAL MOVEMENT

8.1 Introduction

A vast array of different analysis types have been applied to tracking data collected using GPS tracking devices. Thus far no single method is universally accepted for these types of data analysis. However many more studies are being initiated around the world which collect GPS location data for a large variety of species, both terrestrial and marine. Numerous new techniques for the analysis of this type of data are appearing in the ecological (McClintock et al. 2012), rather than statistical literature. However, there is not an accepted suite of models which are widely accepted by the movement ecology community. While some review studies have been done (Patterson et al. 2008, Smouse et al. 2010, Schick et al. 2008) which discuss some of the various methods which have been applied to animal movement data, very few actually compare the model output and interpretations which are found when applying different methods to the same input data. The main aim of this study was to establish a statistical framework for the analysis and interpretation of GPS tracking data, using sable antelope, buffalo and zebra data from the Kruger National Park. The focus was on the methods which are most appropriate for these species, how the models compared to one another and the comparison of the movement patterns across species and regions. This chapter focuses on how the three techniques used in this thesis compare to one other and which is the most suitable for the analysis of these movement data.

In this thesis, the focus has been to investigate three different, yet similar, methods for applying what are often referred to as state-space models to the movement data in order to infer the behaviour of the animal from the underlying latent states within the model. Each latent state is assumed to correspond to a

different behavioural state during the period between successive locations. It was necessary to understand how these methods perform relative to one another and what the similarities and differences between their outputs are. This builds a baseline understanding of the applicability of these methods to model animal movement data in a behavioural context. These behavioural states can be used to investigate ecological questions about the different ways that animals are behaving, how this behaviour changes through the seasons, what similarities exist with other species or the same species in different regions, and how they are utilizing their resources, among many others.

However, while all three of the methods focused on in this thesis (Mixture, Hidden Markov and Bayesian state-space models) have been used in the literature in previous animal movement studies (Owen-Smith, Goodall & Fatti 2012, Franke, Caelli & Hudson 2004, Franke et al. 2006, Langrock et al. 2012, Morales et al. 2004) among others, very few studies compare and contrast the methods used. This would allow an investigation of how the decision about which analysis to use can affect the interpretations made about the behaviour of the animal. These inferences are based on the different latent states uncovered by the different methods. Pros and cons exist for each method including the complexity of the model, the time taken to fit the model, the availability of software and the complexity of the output. Ultimately the decision about which method to use is likely to be based on the hypotheses and desired outcome of the study (for example: simple behaviour categorization, comparisons of the transition probabilities, seasonal changes in model parameters, etc.), as well as the level of statistical expertise of the researcher and the area of statistics in which they feel most comfortable. The desired outcome should also determine the amount of complexity which is required in the model, i.e. seasonality, covariates, transition probabilities etc. The aim of this chapter is to compare and contrast the suitability of these different modelling techniques for analysing the hourly and daily movement data for the sable antelope, buffalo and zebra collected in the Kruger National Park, using the basic models for the three methods. All three methods

provided realistic estimates of the behavioural patterns for the three species presented in the previous chapters. However, how similar the fitted state-dependent distributions are and how consistent the state allocation results are at an observation level is not clear. This chapter investigates the model output and how the different methods provide similar/different interpretations and results from the same input data.

8.2 Comparative studies in the literature

Very little work has been done where a variety of different models are applied to the same data in order to investigate the output and attempt to determine which model or suite of models is most suited to the movement data which they are analysing. Some papers review a few of the methods, highlighting advantages and disadvantages to each method and then customize or develop their own model to apply to the data, rather than testing their data with all of the reviewed models (Schick et al. 2008). This applies to studies of home range and behavioural activity state analysis, where in both fields there are numerous new methods being discussed yet very little being done to formally compare the model outputs. This section is a review of some examples in the literature where comparisons are made between the model output from various home range or behavioural studies. The aim of the review is to highlight papers which have applied multiple models to a dataset in order to compare the results and inferences obtained from the models, even if the time scales are very different to the temporal resolution of the data used in this thesis. In each case, the criteria used to differentiate the models are noted.

Austin et al. (2004) used correlated random walk (CRW) and Lévy flight models to study the movement of grey seals (*Halichoerus grypus*) on Seal Island in the North Atlantic (Austin, Bowen & McMillan 2004). Their objective was to test the applicability of CRWs and Lévy flight models to predict the trajectories of

individual seals; and to investigate how covariates such as age, sex and season influenced their movement types. The authors found that a CRW model fitted the movement tracks of around half of the studied seals, while only eight of the 52 seals had step lengths which fitted the distribution of a Lévy flight. Five of these eight also fitted the CRW model. The authors concluded that while the CRWs only fitted around half the studied animals, the models did provide an opportunity to classify the general movement of the animal as either 'resident' or 'mover' and that the Lévy distribution was not appropriate for modelling their movements. Models were compared using the observed vs. predicted (determined via bootstrapped simulation) net squared displacement calculated using the mean movement displacement and the mean of the cosines of the turning angles.

Franke et al. (2004) used Hidden Markov and autoregressive time series models to investigate the movement of caribou (*Rangifer tarandus*) at 15 minute intervals (Franke, Caelli & Hudson 2004). They categorized the distance and turning angle between locations into four categories each and fitted a 3-state hidden Markov model to the data. They assumed these states corresponded to the animal bedding, feeding and relocating. They also fitted autoregressive time series models to each individual caribou, which ranged between second and tenth order models. However, the first term of the model always had the greatest predictive power which supports a first-order Markov process. The authors trained, estimated and evaluated their models using two methods, the full dataset and half the data (training on the first half and testing on the remaining data). They used measures of percent correct and absolute average difference to test the fit of the models. They found that the HMMs outperformed the autoregressive time series on both measures. The true behavioural state of the animal was never known though, so the "observed" states were presumably estimated using the raw GPS locations and then compared to the "predicted" states obtained from the Viterbi algorithm for the fitted model. The model input was categories of the distance between locations and turning angle determined using the GPS locations. They used Monte Carlo sampling to estimate the average percent correct classification.

In a study not related to inferring behavioural states but which formally compared different methods to determine the 'best' home range model, Horne & Garton (2006) discussed the application of information theoretic approaches to model selection (Horne, Garton 2006). They used a modification of the AIC, especially for the case where the ratio of sample size to the number of fitted parameters is small, called the AICc and compared this to the 'likelihood cross validation' called CVC. To use either of these criteria, the fitted model has to be defined in terms of a likelihood function. They tested the model selection using six different home range models including the bivariate normal, nonparametric fixed and adaptive kernel methods, two models based on the "multinuclear" home range and a third based on the uniform distribution of space use. The best model was determined using the lowest AICc (parametric methods only) or CVC. They also simulated data to actually compare the model selection by the two criteria, for various sample sizes. They found that the two criteria performed similarly in correctly selecting the true data-generating model, with less uncertainty as the sample size increased. The authors did not find any evidence to favour using one criterion for model selection over the other, although they noted that the AIC is not as computationally intensive as the CVC. However, they noted that the AIC is only applicable when the number of parameters is specified, which means that it cannot be calculated for kernel methods.

In an application of a home range analysis using data from a black bear (*Ursus americanus*), Horne et al. (2007) estimated the home range of the bear using a Brownian Bridge Movement Model and a fixed kernel density with smoothing parameter (Horne et al. 2007). They used graphical output to compare the two home range estimates, and found that there was a 77% overlap in the areas represented by the two estimates. The authors explained the differences in home range estimates being due to the differences in assumptions made by both methods. They felt that the Brownian Bridge Movement Model provided the more realistic estimate of the space use by the bear.

Gutenkunst et al. (2007) studied the movement tracks of Atlantic Bluefin tuna (*Thunnus thynnus* L.) within the Gulf of Maine (Gutenkunst et al. 2007). The logger records the position of an individual within its school tracked using acoustic transmitters from a boat. Data were recorded at one minute intervals for up to two days. The authors used two different movement models in order to identify resource patches used by the tuna from these fine scale data. They segmented the movement tracks using the ratio of net displacement to length of track over a time window *t*. The best segmentation was found by optimizing the fit of a random-walk movement model with two states for localized and long-range movement. Two random-walk movement models were used; one was a biased turning model and the other a biased speed model. For the biased turning model, the movement of the tuna was biased in a particular direction by changes in the turning angle from a random component and a bias component. In the biased speed model, the movement was biased by changes in speed. In a comparison of the results from the two models, the authors found that both methods had similar distributions of long-range movement durations, but the biased turning model provided far better resolution of localized movement patches. They used graphical methods to illustrate how the biased speed model identified large utilisation patches, which the biased turning model was able to segment into smaller patches of utilisation. They concluded that the biased turning model provided the better model and was able to identify resource patches with higher resolution. Gutenkunst et al. (2007) also applied some of the other movement models from the literature to their movement data to investigate how these other methods performed. They found that the first passage time method (Fauchald, Tveraa 2003) found patches with far larger radius than the patches using the biased turning model. They expected that some of the patches they identified may have been included in the first passage time patches. The authors also fitted the 'double switch' model of Morales et al. (2004) using WinBUGS (Lunn et al. 2000). They used graphical methods to compare the results and found that the Bayesian method provided relatively similar patch segmentations. However some of the long-range movements of the Gutenkunst et al. (2007)

method were identified as local movement by the Morales et al. (2004) model and the patch sizes were larger using the Morales et al. model.

In a comparative study, Dragon et al. (2012) compared the performance of the Morales et al. (2004) Bayesian fitted two-state models which they referred to as Hidden Markov models (HMM) fitted using a Bayesian framework (with and without covariates included in the model), first passage time (Fauchald, Tveraa 2003), first bottom time (a modification of the first passage time analysis) and empirical descriptors of the foraging effort (step length, number of dives, bottom time residuals and maximum diving depth) (Dragon et al. 2012). They applied these models to Argos and GPS movement tracks of six Elephant Seals (*Mirounga leonine*) which were captured on Kerguelen Island, recorded at a minimum of 20 minute intervals spanning one to two years. They used time depth recorder data to identify drift dives made by the seal and then a drift rate of the animal was determined using the slope coefficient of a linear regression between the depth and time. The drift rate data were used to identify periods when the seals were in an intensive forage phase due to an improvement in body condition and buoyancy. These periods were used as the "known" states to compare with and validate the behavioural state prediction from the fitted Bayesian SSMs, the first passage time analyses and the empirical descriptors. The authors used two measures of similarity, a Jaccard index and an index derived from the Levenshtein distance, to compare the similarity of the time series of intensive foraging periods from the "known" behaviour and the predicted behavioural state. They found that their 2-state Bayesian 'HMM' with or without an environmental covariate (Sea Level Anomalies data) provided the most accurate behavioural predictions of all the methods tested. The simpler 2-state model with no covariate performed the best, with the next simplest SSM including a covariate for the Sea Level Anomalies data as the next best predictor of the foraging behaviour of the seal. The authors used the DIC criterion to investigate the fit of the models. The first passage time and first bottom time methods did not perform as well as the simple Bayesian 'HMMs', but did perform better than the empirical descriptors. The authors also

investigated the ability to predict behaviour using all these models between the GPS and Argos data. Once again the Bayesian 'HMM' analyses provided the most similar results from the GPS and Argos data of the locations of intensive foraging.

Langrock et al. (2012) applied two analysis techniques to data from a study of bison movement in Canada (Langrock et al. 2012). They applied basic HMMs fitted using a frequentist approach and a modification of the HMM where the probabilities of remaining in the same state are modelled separately by the probability mass functions of the dwell-time distributions (time spent in a state before transitioning to another state). This modification is called the Hidden Semi-Markov Model (HSMM). The authors fitted the 2-state HMMs and HSMMs to individual data for nine bison, and then fitted three different models simultaneously to all nine movement paths using complete pooling and random effects. Their aim was to compare the fit of these various hidden Markov methods to see which one provided the 'best' fit to the observed data using the AIC to determine the fit of the model. They found that the combination of Weibull distribution for the step lengths and wrapped Cauchy for the turning angles outperformed the Gamma and von Mises distributions respectively in terms of the AIC. The authors used AIC to investigate the fit of the different models, and found that the HSMM outperformed the HMMs in most cases. Q-Q plots were also used to assess the model fit.

Although not a study of animal movement, Pedersen et al. used simulated data from a logistic population growth model to investigate the performance of three methods which have also been used in the analysis of animal movement (Pedersen et al. 2011a). Although not a study of animal movement, this study involved a formal comparison of three methods applied to the same input data. The models they compared were a hidden Markov model, a mixed effects model and a Bayesian state space model. The mixed effects model was fitted using AD Model

Builder software, with Matlab used for the HMM and WinBUGS for the Bayesian model. The authors compared the models in terms of the complexity of the implementation, the time taken to compute the output, the accuracy of the parameter estimation, limiting assumptions of the fitted model and the amount of subjective manipulation required before the model can be fitted. Two tests were carried out, firstly the ability to estimate the correct state when the model parameters were known, and secondly, the ability of the model to estimate the model parameters when these were unknown. The root mean square error was used to compare the models for the first test. The authors found that the state estimation accuracy for all three methods was almost identical, with the mixed effects model being the quickest to fit, followed by the HMM and then the Bayesian model. For the parameter estimation, the mixed effects model and the HMM provided almost identical results, while the Bayesian methods were found to be quite sensitive to the choice of the prior distribution and provided much wider confidence intervals than the other two methods. They also found an unexpected result that the time taken to fit the models using WinBUGS was very dependent on the choice of prior distribution, with an inverse Gamma prior on the variance being six times faster to fit than a uniform prior on the log-standard deviations. They concluded that each method has its pros and cons, and in certain situations one might be preferable over another, but the results obtained from all three methods were very similar.

As tracking technology improves and the number of collars placed on a variety of species grows, the volume of data recorded increases. It is expected that more methods will be published which deal with this type of tracking data, as researchers strive to find the most appropriate statistical models and ask different questions of the movement data. One of the key challenges will be to compare these methods, to identify the most suitable ones based on the objectives of the study and to understand the differences in output that they might provide. Very little work has been done on direct comparisons of models applied to a single tracking dataset. These movement data have been collected on very different time

scales, with some as short as a minute to up to two days between locations. In this chapter, the effect of the time scale on the most suitable distribution to fit the data is investigated, using Information Criteria to assess the fit of the distributions. This different scale of input data is likely to influence the ability of the models to fit the observed data. Simulations are used to assess the accuracy of the state allocation and the parameter estimates under the assumption that the true model is known. How the output differs from one type of model to another is not known. It is important to have an understanding of how the decision about which analysis to use can influence the results and inferences obtained.

8.3 Model Comparison

8.3.1 Methods

In order to compare the performance of these three methods which have been applied in this thesis, a similar comparison to the Pedersen et al. study has been done (Pedersen et al. 2011a). The methods are compared using two criteria, one is the classification accuracy of the model and the other is the accuracy of the parameter estimates.

The classification accuracy compares the true known states of the model, which can either be known if the data are simulated or if additional observed data are available which allows the true state to be observed. In some cases, the true state can be obtained by observing the animal during its natural behaviour or calculated from metrics other than the GPS locations which provides another estimate of the true behaviour. This was done in the Dragon et al. (2012) study where periods of foraging were identified from drift rate data which corresponded to an improvement in the body condition of the seal. The classification accuracy is determined by comparing the true states and the predicted states given the fitted model. For this chapter, the classification accuracy is calculated for models fitted to simulated data where the true state is known. The accuracy of the classification

is very dependent on the overlap between the distributions which describe the latent states. If the states are highly separated, the accuracy will be much higher, since there is very little overlap. However, if the states are more similar, there will be an overlap in their distributions and hence more than one state with a high probability of state membership. At the time scales used in this thesis, the expected state allocation is slightly lower since the behavioural states underlying the movement distances are closely related not highly separated. This is evident from the histogram of the displacements which for the movement studies show a sweeping curve compared to a clearly bimodal distribution with clearly distinct states, see Figure 8.1. The further apart the states are, the higher the state allocation accuracy will be.

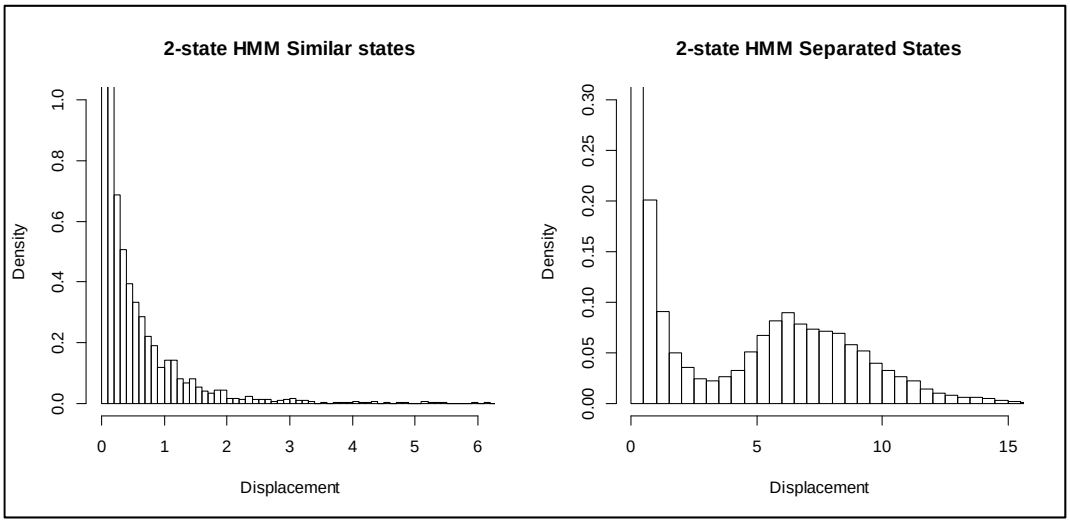

Figure 8.1 - Histograms showing observations from 2-state HMMs with similar (left) and separated states (right)

The second way to compare the accuracy of the model fitting process is via the accuracy of the estimated parameters. For a simulated dataset, the true parameters are known, and the fitted parameters with their 90% confidence intervals and standard errors can be compared to the known parameters. This can be done for the distribution parameters as well as the transition matrix probabilities. Standard errors and confidence intervals can be calculated for all of these parameter

estimates using a parametric bootstrap process. For real world studies where the true parameters are not known, the magnitude of the standard error and the width of the confidence interval can be used as an indicator of the accuracy of the estimate. If the confidence intervals are very wide, there is greater uncertainty about the accuracy of the parameters. This comparison is potentially of more interest to the ecologists since it will provide an idea of the uncertainty around the parameters and hence the confidence with which inferences can be made about the movement states of the animal, and the probabilities of shifting from one state to another. For the simulated data, the parameter estimates should be very close to the true parameter for the fitting process to be considered accurate.

8.3.2 Simulated Data

For the simulation section, only HMMs were used as the 'true' models to simulate data from. Firstly, to investigate the performance of fitted models for data similar to the observed hourly movement data, observations from a 4-state log-normal HMM were simulated. The true parameters were selected to be similar to those obtained from the fitted model for the Punda Maria sable herd using hourly data. The likelihood of the stationary true model is given by:

$$L_T = \delta\Gamma P(x_1)\Gamma P(x_2)\Gamma P(x_3)...\Gamma P(x_T)\mathbf{1}'$$

where $\delta = \begin{bmatrix} 0.50 \\ 0.26 \\ 0.18 \\ 0.06 \end{bmatrix}$, $\Gamma = \begin{bmatrix} 0.70 & 0.21 & 0.08 & 0.01 \\ 0.37 & 0.41 & 0.19 & 0.03 \\ 0.25 & 0.25 & 0.41 & 0.09 \\ 0.08 & 0.06 & 0.32 & 0.54 \end{bmatrix}$

and $P(x) = \text{diag}(p_i(\mathrm{x}_t))$ is the diagonal matrix of state-dependent log-normal probabiltity density functions if the Markov chain is in state i at time t for $i = 1, 2, ...m.$

The density function for the log-normal distribution was given in Section 3.7.2.

The simulation was run with μ_i and σ_i the log-normal distribution parameters

for the *i*-th state defined as:

$$\mu = \begin{bmatrix} -3.71 \\ -1.68 \\ -0.40 \\ 0.69 \end{bmatrix} \quad \text{and} \quad \sigma = \begin{bmatrix} 1.13 \\ 0.67 \\ 0.50 \\ 0.32 \end{bmatrix}$$

This meant that there was a "known" simulated sequence of states with associated simulated observed displacements. In the Correlation Chapter, it was seen that successive displacements are correlated for the movements of the sable, buffalo and zebra. For the mixture model results, the simulation comparison will indicate how well the simple model is able to describe the underlying states within the data, even though the model does not take into account the serial correlation between the observations. This will indicate whether this model is a useful approximation to the real model, under the assumption that in practice the animals behave according to an HMM. The more complex HMM and Bayesian SSM analyses take into account the correlation and should provide more accurate results. The density functions for the IMM and HMM respectively are given by:

$$f(x) = \sum_{i=1}^{m} \delta_i P(x_i) \quad \text{and} \quad g(x) = \sum_{i=1}^{m} \delta_i \Gamma P(x_i)$$

with Γ representing the state transition matrix. The Bayesian state-space model is given by:

$$f(\mu, \sigma \mid Data) = \frac{L(\mu, \sigma)\varphi(\mu)\phi(\sigma)}{\int_0^\infty \int_0^\infty L(\mu, \sigma)\varphi(\mu)\phi(\sigma) d\mu d\sigma}$$

2-, 3- and 4-state mixture, hidden Markov and Bayesian state-space models were fitted to these simulated data, and the "best" model for each method identified using the AIC and BIC criteria. Local decoding and the Viterbi algorithm were

applied to the fitted IMMs and HMMs respectively and along with the predicted sequence from the Bayesian model were compared to the "known" sequence of states. The accuracy of the fitted parameters was compared using their bootstrap standard errors and the ability of the models to accurately predict the true underlying state was investigated. Gamma and Weibull HMMs were fitted to the true-log-normal model to see how well and incorrect model could actually fit the data.

A similar approach was used to investigate the performance of the HMMs and Bayesian state-space models (SSMs) for daily movement data. Raw data (states and observations) were simulated using a 2-state Gamma HMM with parameters similar to those obtained for the 2-state model fitted Punda Maria daily sable movement.

The probability of the Gamma distribution was given in Section 3.7.2.

For the simulated data $\Gamma = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$

with α_i and β_i the Gamma distribution parameters for the *i*-th state

$\alpha = \begin{bmatrix} 1.55 \\ 1.25 \end{bmatrix}$ and $\beta = \begin{bmatrix} 0.50 \\ 1.10 \end{bmatrix}$

2- and 3-state Gamma and Weibull HMMs were fitted to the simulated 2-state data using R (R Development Core Team 2011), and the same simulated data were used as the input for 2-state and 3-state Double with Switch models of Morales et al. (2004). These models are essentially 2- and 3-state HMMs fitted using a Bayesian approach. Once again the model selection, estimated parameters, confidence intervals, standard errors and accuracy of state allocation were compared for the two methods. The PDF of the Weibull distribution was given in Section 5.12. The shape and scale parameters for the Gamma and Weibull

distributions are both denoted with α and β, so as to make the tabulation of the results easier to compare, see Table 8.5.

8.3.3 Simulation Results

Hourly Data Simulation

The maximum likelihood (or deviance), AIC and BIC (or DIC) are shown in Table 8.1, with the lowest values of the AIC, BIC and DIC highlighted. For both the IMM and all the HMMs, both the AIC and the BIC criteria correctly identify the 4-state model as the 'best' model. The Bayesian models were very slow to fit, and in order to fit the models in a reasonable time only 2000 of the 5000 simulated observations were used for the Bayesian model fitting. Also, due to the very long time taken to fit the models in WinBUGS, only 20 000 model iterations were used. The DIC values indicate that the 2-state model would be selected as the best model based on the DIC alone. Although as discussed in Chapter 7, the calculation of the DIC as a measure for model selection is still being debated. The 4-state IMM and HMM were selected as the 'best' models by AIC and BIC, while the 4-state Bayesian SSM did not converge. The Log-normal distribution models always outperformed the incorrect models, and both the AIC and BIC indicated that the Gamma models were preferred to the Weibull models.

Table 8.1 - Maximum likelihood and model selection criteria for models fitted to the true 4-state log-normal HMM

True 4-state Log-normal HMM					
	Number of states	Likelihood/ Deviance	AIC	BIC	DIC
Independent Mixture Model Log-normal	2-state	-1881.89	-3753.78	-3721.19	
	3-state	-1917.47	-3818.94	-3766.81	
	4-state	-1933.50	-3844.99	-3773.30	
Hidden Markov Model Log-normal	2-state	-2253.29	-4494.57	-4455.47	
	3-state	-2415.58	-4807.15	-4728.94	
	4-state	-2476.01	-4912.03	-4781.68	
Bayesian SSM Log-normal	2-state	-3335.00			240.00
	3-state	-3985.56			704.20
Hidden Markov Model Gamma	2-state	-2106.66	-4201.33	-4162.22	
	3-state	-2351.59	-4679.17	-4600.97	
	4-state	-2446.33	-4852.66	-4722.31	
Hidden Markov Model Weibull	2-state	-2131.75	-4251.51	-4212.40	
	3-state	-2343.04	-4662.07	-4583.87	
	4-state	-2432.12	-4824.23	-4693.89	

Table 8.2 shows the true HMM parameters with the fitted parameters using a 4-state log-normal IMM and 4-state HMM. The fit of the model can be judged using the AIC and BIC criteria and also examining the fitted parameter confidence intervals and standard errors. The fitted parameters are very close to the true parameters using both methods, and the 90% confidence limits for both models always includes the true parameter value. These confidence limits are always narrower for the HMM. The standard errors for every parameter are small for both methods, but once again the standard errors for the HMM parameters are always smaller than those for the IMM parameters. These model fitting results suggest that the IMM is a very adequate method for approximating the distributions which describe the underlying states, however, as expected the HMM is always the better method and provides results with greater certainty. The Bayesian SSMs were not included in this comparison due to the inability to fit a 4-state model to the simulated data known to include four distinct latent states. The state allocations determined using the fitted model from the log-normal IMM and HMM methods are shown in Table 8.3.

Table 8.2 - True 4-state HMM parameters and fitted values using log-normal IMM and HMM

True Model		Fitted Models					
Hidden Markov Model		Independent Mixture Model	Independent Mixture Model 90% Confidence Limits	Std Error - IMM	Hidden Markov Model	Hidden Markov Model 90% Confidence Limits	Std Error - HMM
δ_1	0.50	0.50	(0.41, 0.56)	0.05	0.51	(0.47, 0.55)	0.02
δ_2	0.26	0.24	(0.10, 0.41)	0.09	0.23	(0.19, 0.29)	0.03
δ_3	0.18	0.19	(0.09, 0.31)	0.07	0.19	(0.16, 0.23)	0.02
δ_4	0.06	0.07	(0.05, 0.09)	0.01	0.06	(0.05, 0.08)	0.01
μ_1	-3.71	-3.66	(-3.92, -3.47)	0.14	-3.63	(-3.75, -3.53)	0.07
μ_2	-1.68	-1.79	(-2.07, -1.54)	0.17	-1.75	(-1.86, -1.61)	0.08
μ_3	-0.4	-0.55	(-0.77, -0.42)	0.11	-0.52	(-0.60, -0.44)	0.05
μ_4	0.69	0.62	(0.51, 0.71)	0.06	0.64	(0.59, 0.69)	0.03
σ_1	1.13	1.16	(1.05, 1.25)	0.06	1.17	(1.10, 1.23)	0.04
σ_2	0.67	0.64	(0.40, 0.92)	0.16	0.65	(0.56, 0.75)	0.06
σ_3	0.5	0.48	(0.34, 0.71)	0.11	0.52	(0.46, 0.58)	0.04
σ_4	0.32	0.33	(0.27, 0.38)	0.03	0.32	(0.29, 0.36)	0.02
γ_{11}	0.70				0.71	(0.67, 0.75)	0.03
γ_{12}	0.21				0.19	(0.15, 0.25)	0.03
γ_{13}	0.08				0.08	(0.05, 0.10)	0.02
γ_{14}	0.01				0.01	(0.01, 0.02)	0.00
γ_{21}	0.37				0.34	(0.29, 0.39)	0.03
γ_{22}	0.41				0.39	(0.33, 0.47)	0.04
γ_{23}	0.19				0.23	(0.17, 0.29)	0.04
γ_{24}	0.03				0.03	(0.02, 0.05)	0.01
γ_{31}	0.25				0.32	(0.27, 0.37)	0.03
γ_{32}	0.25				0.19	(0.14, 0.25)	0.03
γ_{33}	0.41				0.40	(0.35, 0.45)	0.03
γ_{34}	0.09				0.09	(0.07, 0.12)	0.01
γ_{41}	0.08				0.08	(0.04, 0.12)	0.03
γ_{42}	0.06				0.06	(0.00, 0.12)	0.03
γ_{43}	0.32				0.34	(0.28, 0.41)	0.04
γ_{44}	0.54				0.51	(0.46, 0.57)	0.03

The diagonal values in Table 8.3 are by far the highest for each state, which represents observations correctly allocated to their true state. The incorrect state allocation is nearly always to the state before or after the correct state. There does not appear to be any bias towards either the lower or upper state for either model, with similar numbers of incorrectly allocated observations going to the states above and below.

Table 8.3 - True Hidden Markov model state allocation for fitted IMM (left) and HMM (right)

		State allocation - IMM			
		State 1	**State 2**	**State 3**	**State 4**
True State	**State 1**	2158	294	26	0
	State 2	204	876	257	5
	State 3	0	78	725	86
	State 4	0	0	23	268

		State allocation - HMM			
		State 1	**State 2**	**State 3**	**State 4**
True State	**State 1**	2233	222	23	0
	State 2	223	901	216	2
	State 3	0	78	750	61
	State 4	0	0	27	264

The fitted IMM provides 80.54% state allocation accuracy with the HMM slightly better at 82.96%. On the basis of this, it seems reasonable to conclude that inferences drawn from the underlying states are suitable for use in further analyses. Beyer et al. (2013) found that the accuracy for state allocation of state-space models could range from close to 100% when the states were highly separated to close to 0% when the states were very similar or if one state was very rare. Table 8.4 shows the percentage of observations from each state which were accurately allocated to the correct state using the fitted model. The HMM performs better than the IMM for all states except the most active fourth state. The allocation accuracy is low for the second state, normally assumed to correspond to the foraging state. It is shown in Table 8.3 that the errors in state allocation for the second state are evenly spread between the first and third states for both the IMMs and the HMMs. This lower state allocation accuracy for this state is not necessarily as a result of a poor model, but rather the overlap in

displacement distributions for the foraging state with the states on either side of it. This overlap will result in a higher probability of an incorrect state allocation. When compared to the state allocation accuracy of the incorrect HMMs, a 4-state Gamma model had 67.54% of observations correctly allocated while the 4-state Weibull model's accuracy dropped to 61.34%. This shows as expected that in terms of the state allocation accuracy, the log-normal models fitted to a known log-normal HMM clearly outperform Gamma or Weibull HMMs. Although the AIC and BIC showed that the Gamma or Weibull HMMs would be selected in preference to the log-normal Mixture Model, the IMM outperforms these HMMs in terms of state allocation accuracy. The incorrect models perform better for the observations allocated to the more active states rather than the foraging and resting states.

Table 8.4 - Percentage correct allocation of observations to the true state

Fitted model	True State 1	True State 2	True State 3	True State 4
Log-normal IMM	87.1%	65.3%	81.6%	92.1%
Log-normal HMM	90.1%	67.1%	84.4%	90.7%
Gamma HMM	61.99%	57.60%	89.31%	94.16%
Weibull HMM	57.10%	45.23%	85.94%	96.56%

The accuracy for the simulated movement data at 80% is high although not perfect, but it provides reasonable levels from which to draw inferences about the movement of the animal. It is very interesting to note that the improvement in accuracy from the independent to dependent model is only 2.42%, which indicates that the IMM is performing very well when it comes to predicting the true state of the model which does contain serial correlation between successive observations. From Table 8.3, it seems that the improvement in accuracy comes from observations incorrectly allocated to the higher state (longer displacements) by the IMM, which are then accurately classified by the Viterbi algorithm applied to the

fitted HMM. These results suggest that the additional fitting complexity for the HMM could in some cases be unnecessary, if the only desired outcome from the model is the behavioural prediction, rather than investigating the probability of moving from one state to another. Even though the mixture model violates the assumption of dependence within the input data, the results provide a very accurate representation of the underlying groupings within the model. It also suggests that the magnitude of the observed displacement is influencing the state allocation probabilities more than the probability of transitioning from one state to another, which allows the IMM to provide very adequate results using the displacement at each time step only and not the previous state.

Out of interest, it was investigated whether it was possible to fit an HMM to data simulated from an IMM. It was not surprising that it was not possible to obtain convergence of the HMM since the model is trying to fit a dependency between observations which does not exist in reality. The results from the Bayesian analysis were not particularly useful due to the challenges of fitting multiple-state models using WinBUGS to a dataset of 5000 observations. The models were sensitive to the starting parameters, time consuming to run and often required numerous attempts to fit the model using slightly different initial vales. If the DIC is used as the model selection criterion for these large datasets fitted using WinBUGS, it appears to favour the simplest models. In this case the DIC favoured the simplest 2-state model, although it is known due to the simulation that there are four latent states within the data. For these very large datasets, the HMM model fitting process is the more accurate and is much more time efficient than the Bayesian approach.

True Daily Movement – Gamma Hidden Markov Model

2- and 3-state Gamma and Weibull HMMs were fitted to the data simulated from a 2-state Gamma HMM, with state 1 being the shorter ‘encamped’ state and state 2 the ‘exploratory’ state with longer movements. Compared with the simulated

hourly movement states, the daily states are more separated from one another (expected values for the two states 0.49km and 1.32km respectively), so a slightly higher accuracy of state allocation is expected. 2-state 'Double with Switch' (Morales et al. 2004) Bayesian state-space models were also fitted to the data, using either the Gamma or Weibull distributions to model the step length and with uniform prior distributions for the model parameters as was done in Chapter 7. The Υ_{ij}'s in Table 8.5 represent the transition probabilities from state i to state j, while the α_i and β_i represent the shape and scale parameters for both the Gamma and Weibull distributions. For the Bayesian state-space models, the standard errors for the β_i's (italicised) are the standard errors for the rate parameter as calculated by WinBUGS, not the scale parameter.

Table 8.5 – Fitted parameters and standard errors for Gamma and Weibull HHM and Bayesian SSMs fitted to data simulated from a Gamma HMM

True Model		**Fitted Models**			
Hidden Markov Model		**Hidden Markov Model**		**Bayesian State-space model**	
Gamma		**Gamma**	**Weibull**	**Gamma**	**Weibull**
δ_1	0.65	0.64 (0.03)	0.60		
δ_2	0.35	0.36 (0.03)	0.40		
α_1	1.62	1.66 (0.05)	1.44	1.67 (0.05)	1.44 (0.03)
α_2	2.31	2.22 (0.16)	1.47	2.25 (0.16)	1.47 (0.05)
β_1	0.30	0.29 (0.01)	0.51	0.29 (*0.15*)	0.38 (*0.11*)
β_2	0.57	0.60 (0.03)	1.42	0.60 (*0.09*)	1.68 (*0.04*)
γ_{11}	0.88	0.87 (0.01)	0.86	0.87 (0.01)	0.86 (0.01)
γ_{12}	0.12	0.13 (0.01)	0.14	0.13(0.01)	0.14 (0.01)
γ_{21}	0.22	0.23 (0.02)	0.21	0.23 (0.03)	0.22 (0.02)
γ_{22}	0.78	0.77 (0.02)	0.79	0.77 (0.03)	0.78 (0.02)
AIC		7041.97	7047.72		
BIC		7081.08	7086.83		
DIC				14250.68	12592.2
State Allocation Accuracy		84.14%	83.92%	85.20%	84.66%
Expected Value – State 1	0.49	0.49	0.47	0.49	0.35
Expected Value – State 1	1.32	1.34	1.29	1.35	1.52

Bootstrap simulations to compute the standard errors for the Weibull estimates broke down and resulted in unrealistically large standard errors and hence are not reported in Table 8.5.

In reality, the true distribution is not known, so once again the two different distributions were tested to see how well an 'incorrect' distribution could actually fit the data. The Gamma and Weibull distributions are both commonly used for animal movement analyses and their fit to the data can be similar for various combinations of the distribution parameters. The model parameter estimates are shown in Table 8.5 while the pseudo-residuals used to check the goodness-of-fit of models are shown in Figure 8.2. Both the AIC and BIC correctly identify the Gamma HMM as the preferred model to the Weibull HMM, while the DIC seems to indicate that the Weibull Bayesian SSM is preferred to the Gamma Bayesian SSM.

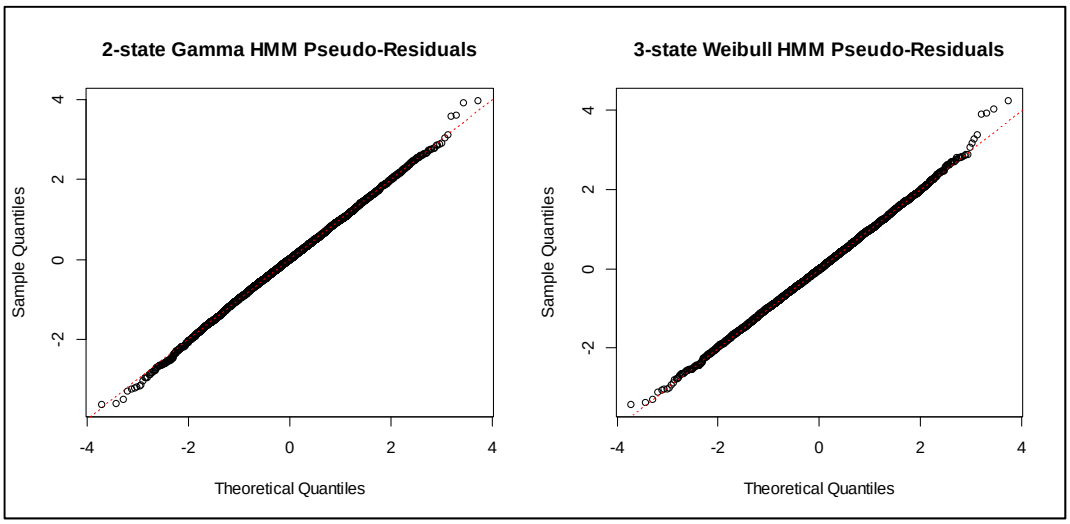

Figure 8.2 - Pseudo-residuals for 2-state Gamma and Weibull HMMs fitted to true 2-state Gamma HMM

The pseudo-residuals for both the Gamma and Weibull models are normally distributed and both provide similar and good fits to the data, as shown in Figure

8.2. The HMM and Bayesian SSM fitted using the Gamma distribution provided good estimates of the true parameters, and the standard errors for both fitting approaches are very similar. As expected, the parameters fitted by the Weibull state-space models are very different. However the expected values for the two Weibull distributions are very similar to the expected values for the true model which suggests that the Weibull distributions also provide a good fit to the simulated data (0.47km and 1.29km respectively). It is encouraging that although two different distributions could be used, the underlying states and inferences drawn from the models would be very similar. Figure 8.3 shows a plot of the fitted Gamma and Weibull distributions for the fitted HMMs. It shows how similar the fitted distributions are, even though they are from a different family of distributions. An examination of the accuracy of the state allocation though for both the Gamma and the Weibull distributions fitted using either the Bayesian SSMs or HMMs is approximately 84% - 85% for all of the models. The standard errors are slightly larger for some of the daily model parameters compared to the hourly parameters (see Table 8.5 and Table 8.2 respectively), which might be due to the much larger dataset size used for the hourly movement's simulation.

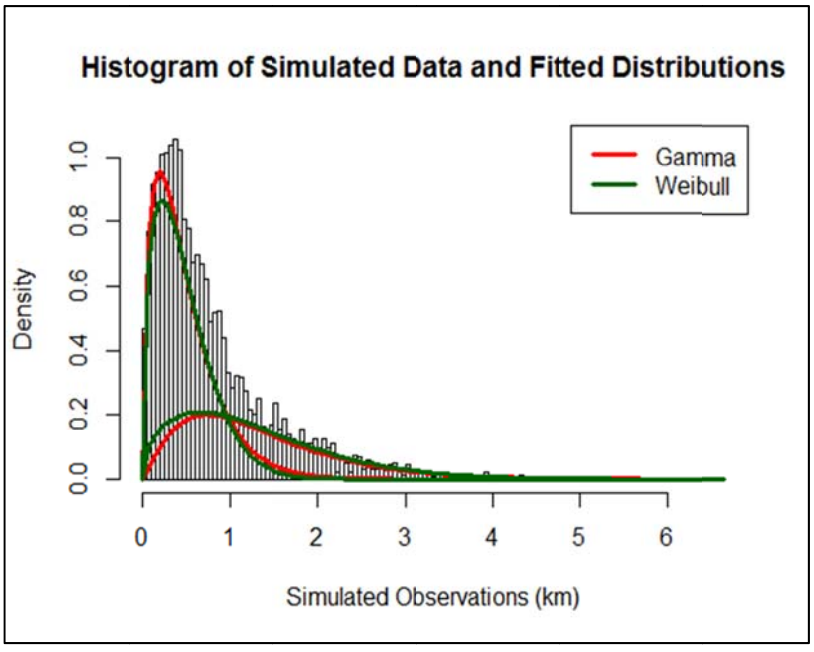

Figure 8.3 - Simulated observations and fitted distributions for the Gamma and Weibull HMMs

A caution to the accuracy rates stated here for the simulations is that these are based on the true distributions being known, and the estimated model being fitted with initial starting values quite similar to the known values. As shown above, it is also possible when fitting Gamma or Weibull distributions to observed data that different combinations of the parameter values could result in very similar distributions (Beyer et al. 2013). When conducting a movement study, checks need to be done to ensure that even though the parameters for competing models might look very different, whether in fact the distributions defined by them are similar and would result in similar interpretations being drawn.

8.4 Animal Movement Application - Hourly Movement Data

8.4.1 Methods

A subset of the data was chosen to compare the fit of the IMMs and HMMs. This was done to ensure the results were directly comparable and that any effect due to the chunks of missing data (changes in the frequency of observation) was avoided. The data for the comparative analysis was selected at an ad hoc time when the collars consistently recorded hourly. Details of the datasets are shown in Table 8.6. This meant that the percentage of missing observations in these comparative analyses was much lower than for the analysis presented in the Basic HMM Chapter. Stationary HMMs were used for the comparative analysis so that the mixing parameters from the IMM could be compared to the stationary distribution from the HMM. In both cases, the inference from the stationary distribution and the mixing parameters is in terms of the amount of time that the animal will remain in a particular state. Log-normal distributions were used as the state-dependent distributions for this comparison. The state allocation used the Viterbi algorithm (Zucchini, MacDonald 2009) for the HMMs and the 'Mixture Likelihood' clustering method (McLachlan, Peel 2000) for the Mixture Models as discussed in the respective chapters. Models were compared using the likelihood, AIC, BIC and the time taken to fit the models. The HMMs are more complex models, formally including the probability of transitioning from one state to

another and require these additional parameters to be fitted. The time taken to fit the models will be a good indicator of the additional computational requirements of each model. The model outputs are compared using the fitted distribution parameters, the confidence intervals for each parameter obtained using the parametric bootstrap method, the expected movement distance in each state and the observed mean movement distance of the observations allocated to each state.

8.4.2 Results

2-, 3- and 4-state IMMs and HMMs were run for a Punda Maria sable, buffalo and zebra herd, as well as the Phabeni sable herd at Pretoriuskop. The time series length and percentage of missing observations for each herd are shown in Table 8.6. The percentage of missing data for the HMM analysis ranged between 10.77% and 16.16%.

Table 8.6 - Summary of subset data used for Mixture Model and Hidden Markov Model comparison

	PM_Sable279	PM_Buffalo152	PM_Zebra277	PK_Phabeni
Date range	17 August 2007 – 9 January 2008	21 January 2008 – 29 June 2009	15 July 2007 – 6 June 2010	5 August 2007 – 20 March 2009
Time series length	3 500	12 608	25 382	14 196
% Missing data	12.46%	10.77%	16.16%	11.31%
Best IMM based on AIC	4-state	4-state	4-state	3-state
Best HMM based on AIC	4-state	4-state	4-state	4-state

The AIC for both IMM and HMM selected a model with four states in all cases except for the Phabeni sable herd where the AIC for the IMM results selected a 3-

state model. The AIC for the Phabeni HMM preferred a 4-state model, however based on the ease of interpretation of the states, comparison to the IMM and the fact that two of the states for the 4-state HMM were very similar, it was decided to use the 3-state model specifications for the comparison for this herd.

The log-likelihood, AIC, BIC and fitted model parameters for the 2-, 3- and 4-state IMM and HMMs for the Punda Maria Sable herd (PM_Sable 279) are shown in the Appendix Section 8.8 Table 8.12. δ_i denotes the stationary distribution or mixing parameters for the HMM or IMM respectively, and μ_i, and σ_i, denote the mean and variance parameters respectively of the log-normal distributions for the i-th state of either model. λ_{ij} is the transition probability from the i-th to j-th state for the HMM. The 'proc.time' is the time taken to fit the 2-, 3- or 4-state models in R using a Toshiba NB520 netbook. The model fitting times would vary from computer to computer, however the ratio between the two is likely to remain quite consistent due to the differences in model complexity. The HMM takes between three and ten times longer to fit compared with the IMM for all of the herds. This is a significant increase in time, although for datasets of these sizes and with a simple model specification, it is not prohibitive to fitting the model. It could however become more problematic for very large datasets or models with a more complex structure and more states. This is in contrast to the Bayesian state-space models which were already taking a substantial amount of time to fit for multi-state models with large datasets and was considered prohibitive for using them as an analytical technique for the hourly models. Pedersen et al. also found that their Bayesian models took much longer to fit than the HMMs (Pedersen et al. 2011a).

For the Punda Maria Sable herd, the fitted model parameters for the distributions and the mixing parameters/stationary distribution are very similar between the two fitting approaches (Section 8.8 Appendix Table 8.12). For the 4-state model, both models predict above 50% of the time resting, over 20% foraging, under 20% moving and only around 5% of the time taken up with long relocating moves. As

with the models fitted to the simulated data, the 90% confidence intervals and the standard errors of the parameter estimates are always narrower and smaller for the HMMs. This confirms that there is more certainty around the parameter estimates for the HMM. Although, once again the IMM provides very adequate parameter estimates of the distributions which describe the underlying states.

Table 8.7 - Punda Maria sable - expected movement distances and mean displacement by state

	Hidden Markov Model		**Mixture Model**	
	Expected movement (km)	Average displacement using state allocation (km)	Expected movement (km)	Average displacement using state allocation (km)
State1	0.05	0.03	0.07	0.05
State2	0.24	0.22	0.31	0.27
State3	0.75	0.75	0.83	0.81
State4	2.10	2.15	2.17	2.26

Table 8.7 shows the expected movement distance in an hour for each state using the expected value for the state based on the fitted log-normal distribution parameters and the average displacement of the observations allocated to each state using the Viterbi and Maximum Likelihood clustering methods. The results are very similar for the two analysis methods, and the interpretations of the latent states from these two methods would be the same. In both cases, only the fourth state has a higher average displacement based on the state allocation than the expected value of the distribution describing that state. This could be as a result of all the extreme observed values from the right tail of the distribution being allocated to the fourth state, while some of the observed values to the left of the distribution's expected value would be allocated to the third state instead. This would mean that the observations were allocated as large third state values rather than small fourth state values.

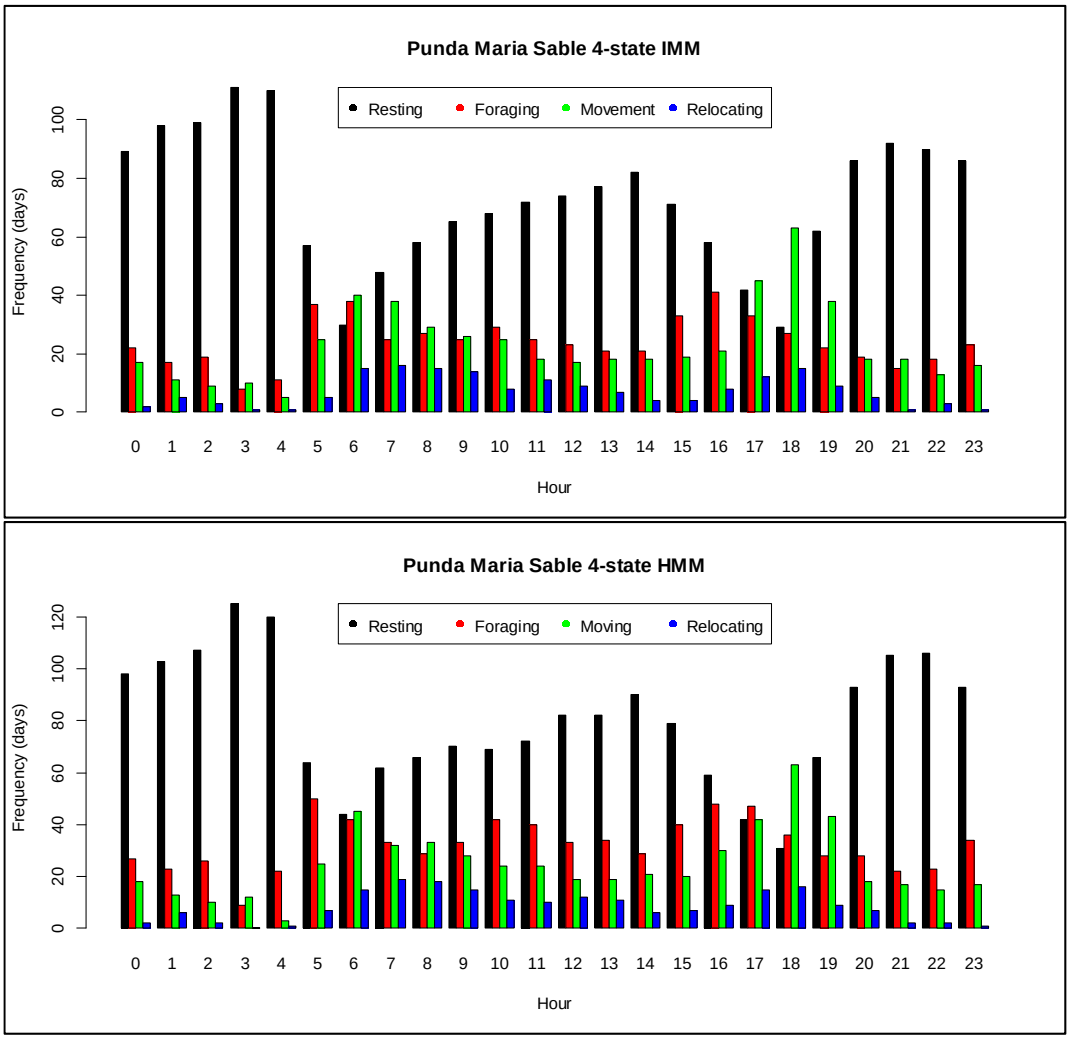

Figure 8.4 - Punda Maria Sable - State allocation by time of day using 4-state Mixture Model and Hidden Markov model

Figure 8.4 shows the comparative state allocation by time of day for the observations using these two methods, with each bar in the plot representing an hour time interval. The pattern is almost identical, with resting dominating until 5am, followed by a bout of foraging or movement. Resting then increases during the morning and early afternoon until the evening foraging or movement session becomes the dominant activity from 5pm. Resting returns as the dominant state from 7pm. Relocating nearly always occurs during the day, peaking around 8am and again at 6pm, with a dip during the middle of the day. At a global level, the interpretations obtained from these two methods appear to be exactly the same.

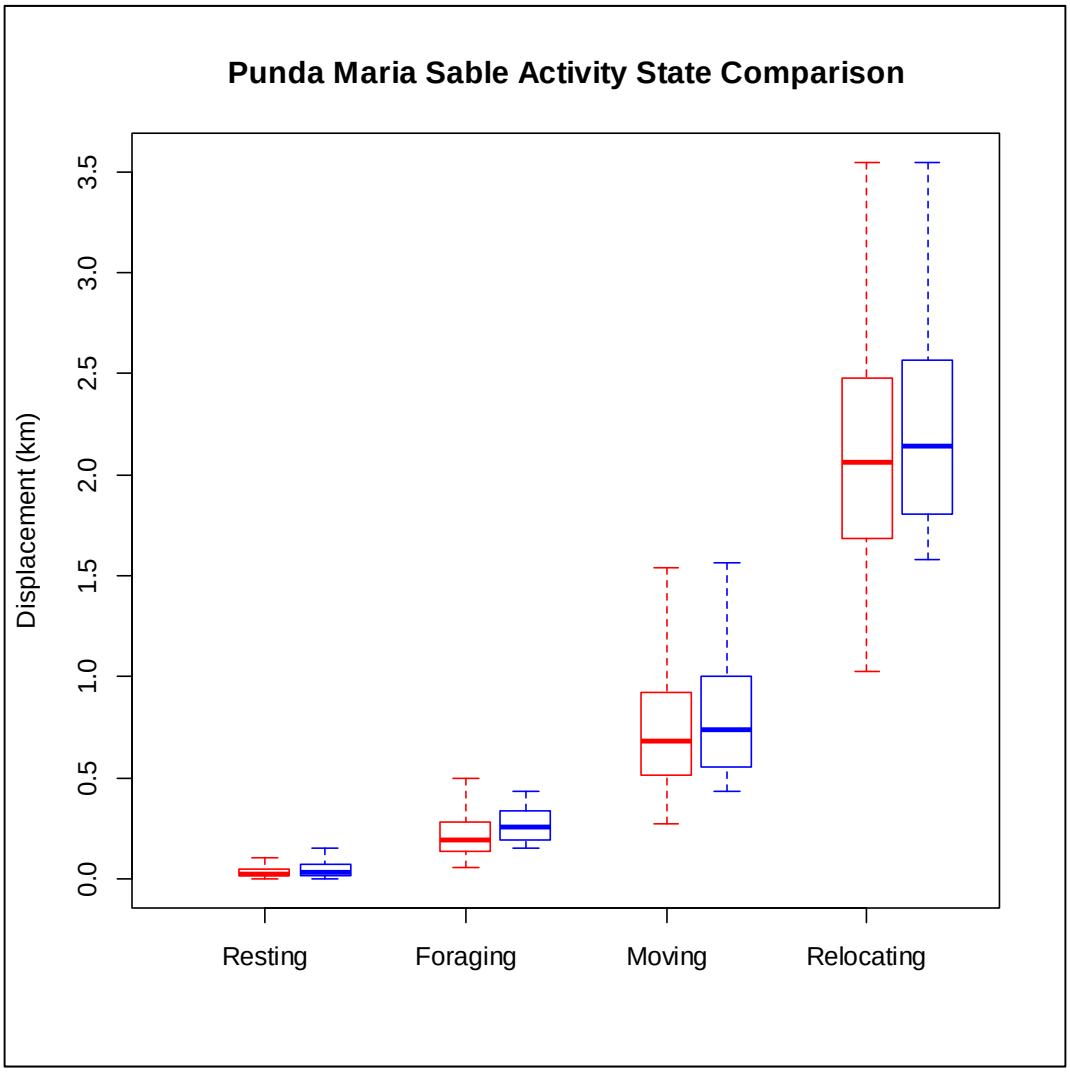

Figure 8.5 - Punda Maria Sable - Comparison of observations allocated to each state. Red (HMM) and blue (IMM)

Figure 8.5 shows a box plot of the observations allocated to each of the states using the two different methods. One of the most noticeable differences is the lack of overlap across the different states of the whiskers for the IMMs. In effect, this method has thresholds where any value above a specific level will be allocated to a higher state. This is in contrast to the HMM models where an observation can be allocated to a higher or lower state with the transition probabilities playing a role in the shifts from one state to another. Although the whiskers of the HMM plots do overlap, the overlap is sensible and small. For example, a very short displacement is never allocated to the relocating or movement state. For all states,

the spread of observations allocated to a state is wider for the HMM than that of the IMM. Table 8.8 shows a cross tabulation of the state allocation for the two different models. The values on the diagonal of the table are by far the highest indicating that the vast majority of the observations are allocated to the same state, no matter which method of fitting is used. The observations not allocated to the same state are most likely as a result of the transition probabilities influencing the state allocation. These observations tend to be more likely to be allocated to the less active state for the Mixture Models and the more active state for the HMMs.

Table 8.8 - Punda Maria Sable - comparison of state allocation using Maximum Likelihood clustering and the Viterbi algorithm

		Mixture Models			
		State 1	**State 2**	**State 3**	**State 4**
Hidden Markov Models	**State 1**	1495	10	0	0
	State 2	259	487	32	0
	State 3	0	82	494	13
	State 4	0	0	31	161

Although the results for the Punda Maria sable herd are very similar using the two methods, this is not always the case for the other herds. The Phabeni sable herd had very similar results using both techniques however the buffalo and one zebra herd did not. Fitting the Bayesian model to the Phabeni sable herd was also attempted, however, for this very large dataset (14196 observations), the model took a very long time to run, struggled to converge and it was decided that this was not a viable modelling option for the multi-state models required for the hourly analyses using these large datasets. In contrast to the results for the Punda Maria sable herd, the results between the IMMs and HMMs differ quite dramatically for the Punda Maria buffalo (PM_Buffalo152) and zebra (PM_Zebra277). This zebra herd did tend to move more than the other zebra herds in the same area, but it is clear that the results of the two methods are not always going to be as similar, and can fit distributions with very different shapes and

have different mixing parameters/stationary distribution. These differences are highlighted in Figure 8.6 and Table 8.9.

Table 8.9 - Punda Maria buffalo - expected movement distances and mean displacement by state

	Hidden Markov Model		**Mixture Model**	
	Expected movement (km)	Average displacement using state allocation (km)	Expected movement (km)	Average displacement using state allocation (km)
State1	0.04	0.03	0.04	0.03
State2	0.20	0.18	0.29	0.22
State3	0.46	0.40	0.64	0.72
State4	0.81	0.86	1.93	2.13

Figure 8.6 shows the state allocation for the buffalo herd by time of day using the two different methods. The resting state (black) has a very similar pattern in both plots with clear peaks in resting before sunrise and during the heat of the day. The pattern of the foraging is also similar with peaks mid-morning and late afternoon and continues into the evening. However, the state allocation pattern of the more active states is very different. Far more observations are allocated to the fourth state for the HMM. This is a result of the expected displacement distance for this state's distribution being far shorter than for the fourth state in the fitted IMM (0.81km and 1.93km respectively, see Table 8.9). The interpretation of the third and fourth states is clearly very different for these two models and so is the state allocation and proportion of time spent in these states. Both models were re-run using starting values similar to the fitted results for the alternative method, but in all cases the models converged to the results presented here. While it might seem alarming that the results presented here are so different for the same herd, the allocation and interpretation of the states would be different for the two models. While the Mixture Model could be interpreted as resting, foraging, local movement and relocating, the HMM could represent the herd resting, foraging, mixture of feeding and movement, with the fourth state representing local

movement, respectively. It is interesting to note that the fitted models for the simpler 3-state HMM and IMM also describe different states which would be associated with different behavioural inferences for the two approaches. The question about which is the most appropriate model for the analysis of these buffalo data is a tricky one to answer. It is likely that both methods provide sensible results in terms of the behaviour of the animal since the interpretation of the states will differ for the two models' output. The choice of which model to use would most likely be based on what state interpretations were preferred and made more biological sense. If any actual behavioural field observations existed, then it would depend on which model interpretation best suited the observations recorded about the true behaviour. However, as discussed for the simulation, due to the fact that it is known that there is strong serial correlation in the movement data, the HMM has to be the better and more accurate model as it accounts for this dependency within the data. It is expected that the HMM will always provide the more sensible model for the analysis of animal movement than the IMM, except in the situation of little serial correlation in the movement data which is unlikely. Both the buffalo and this zebra herd had very strong patterns of autocorrelation shown in the Chapter 4 which might explain why these differences in model output were found. However, an alternative view is that the HMM will mostly persist in a state for more than one time step. In reality this might not be the case for the more active movement states. For buffalo, a travelling state rarely persists for more than one hour, and by anticipating the autocorrelation the HMM might not perform as well for this state. Obtaining field observations to validate these state predictions might provide some valuable insight into this.

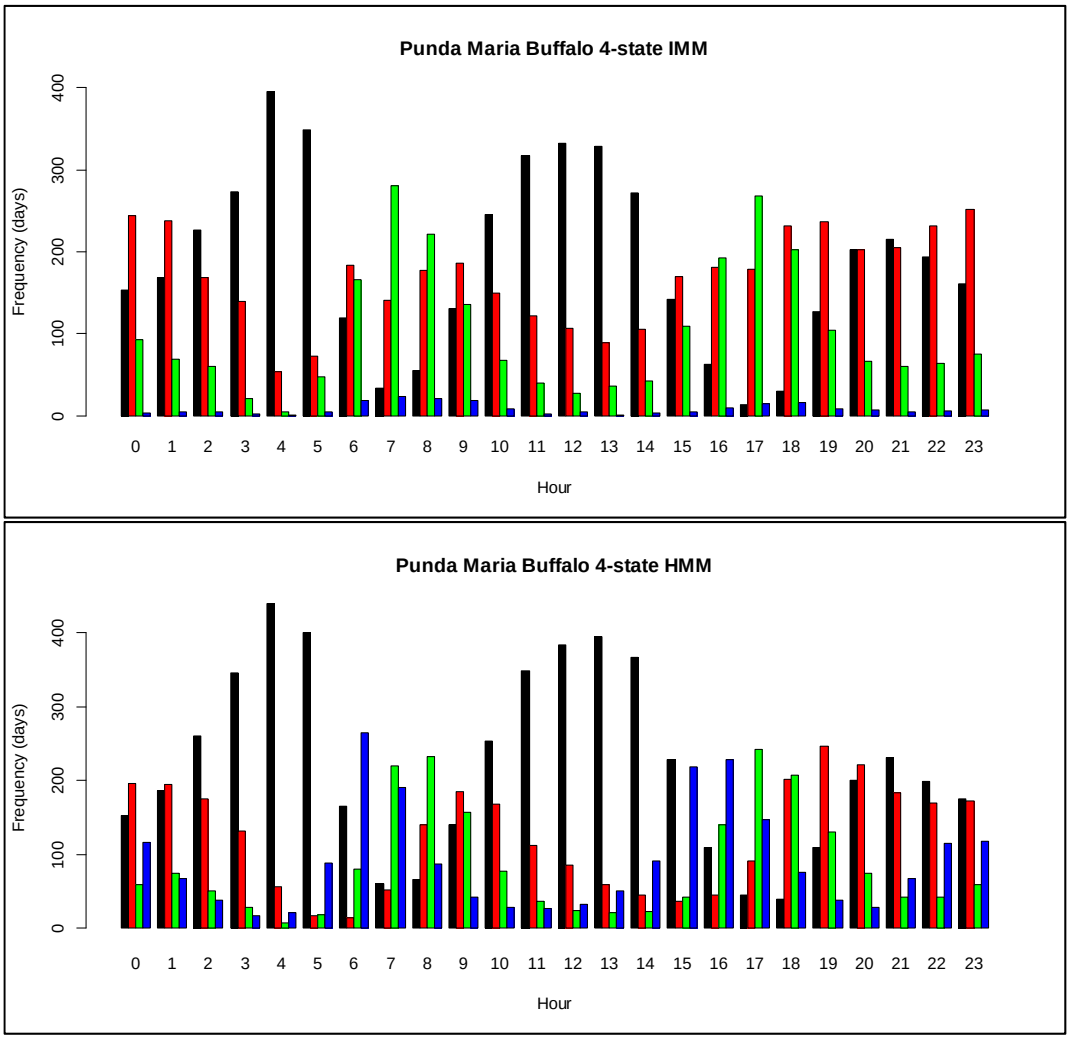

Figure 8.6 - Punda Maria buffalo - State Allocation by time of day for IMM and HMM. Black (state1); red (state2); green (state3); blue (state 4)

8.5 Animal Movement Application - Daily Movement Data

8.5.1 Methods

The comparative work on the daily movement models compared the output of the HMMs and the Bayesian State-space models (SSMs). The results used were presented in the Basic HMM and Bayesian SSM chapters since changes in the frequency of observation did not affect the daily datasets. Models were compared which fitted Gamma distributions to the daily displacement distances. A 2-state HMM is compared with the Bayesian Double with Switch model using a Gamma

distribution for the step lengths and uniform prior distributions applied to each parameter. The Double Switch model was defined in Page 270. For the rest of this chapter, it will be referred to as the 2-state Bayesian model. Stationary HMMs were used for the comparison due to the very similar parameter values which were found between the stationary and non-stationary models in Section 5.13. Due to the different model specifications, the likelihood functions are not compared nor are values such as the AIC, BIC or DIC. The distribution parameters, standard errors, the expected movement distance in each state and the state allocation are compared for both models.

8.5.2 Results

The 8pm locations for the Phabeni sable herd from Pretoriuskop was used as the case study herd to compare and contrast the results obtained by these two models. The same process could be used to compare the fit of these approaches using the data anchored at the other time points during the day, but it is expected they would perform similarly although the distributions and parameter estimates might differ for the different input data collected at each time. Table 8.10 shows the fitted parameters and their standard errors for the two models. An initial inspection of the parameters suggests that the results are slightly different when looking at the distributional parameters and the expected values from the distributions. However the average displacements of the observations allocated to each state are more similar. State 1, the encamped state, has observations with an average daily displacement of 0.48km and 0.39km allocated to it for the HMM and the Bayesian SSM respectively. The relocating state is associated with daily movements with an average of 1.42km and 1.2km respectively. The state transition probabilities are quite similar for the two methods, although the results for the HMM are higher for the diagonal probabilities which represent remaining within the same state. The standard errors for the HMM parameters are smaller than for the Bayesian model. The comparison of state allocation in Table 8.11 shows that the models are quite consistent, with 83.6% of observations being allocated to the same state using the two different methods. These differences are

likely to have been caused by the more extreme transition probabilities for the HMM which will result in longer runs of observations in one state before a switch to the other state. Interestingly, 88% of the observations were allocated to the same state using the Weibull distribution instead of the Gamma distribution for the step-lengths for the Bayesian and HM models. A comparison of the state allocation for both methods over time using the Gamma distribution is shown in Figure 8.7, showing a similar pattern over time. The encamped state is the dominant state during the summer months and through the early dry season, with the relocating state being more dominant from August/September through until December. The state allocation at the end of the dry season is slightly more extreme for the HMM which is probably caused by the higher state transition probabilities for remaining within a state resulting in fewer transitions from one state to another.

Table 8.10 - Phabeni Sable - Daily 8pm - Fitted parameters using a Gamma Hidden Markov model and Bayesian state-space model

PK_Phabeni_8pm							
		Parameter	**Std Error**			**Parameter**	**Std Error**
2- state HMM	$\delta 1$	0.65	0.06	**Double with switch - Bayesian SSM**			
	$\delta 2$	0.35	0.06				
	$\alpha 1$	1.62	0.10		$\alpha 1$	2.57	1.33
	$\alpha 2$	2.31	0.47		$\alpha 2$	2.08	0.55
	$\theta 1$	0.30	0.03		$\theta 1$	0.31	0.05
	$\theta 2$	0.57	0.07		$\theta 2$	0.48	0.10
	$\Upsilon 11$	0.88	0.03		$\Upsilon 11$	0.81	0.10
	$\Upsilon 12$	0.12	0.03		$\Upsilon 12$	0.19	0.10
	$\Upsilon 21$	0.22	0.06		$\Upsilon 21$	0.18	0.09
	$\Upsilon 22$	0.78	0.06		$\Upsilon 22$	0.82	0.09
	Expected Value - State 1	0.48			Expected Value - State 1	0.79	
	Expected Value - State 2	1.30			Expected Value - State 2	1.00	
	Average displacement 1	0.48			Average displacement 1	0.39	
	Average displacement 2	1.42			Average displacement 2	1.20	

It is interesting to note that the expected value for State 1's distribution is shorter using the fitted HMM than when using the Bayesian SSM, however the average displacement is longer. For State 2 the average displacement from the HMM is also longer, however in this case the expected value for the distribution is also longer than for the Bayesian SSM. The pattern of average displacement can probably be explained by the state allocation shown in Table 8.11. 169 observations allocated to State 1 of the HMM were allocated to State 2 of the Bayesian SSM. These are likely to be the longer State 1 displacements from the HMM allocation which when moved to State 2 in the Bayesian SSM allocation will decrease the average displacement of both states. No observations allocated to State 1 for the Bayesian SSM were allocated to State 2 of the HMM.

These comparisons of the modelling approaches are consistent with the results obtained from the simulation study. Comparisons run using the other herd's data also showed very similar results. Although the distributions parameters are slightly different, the shape of the distribution, expected values and state allocation are consistent. Both methods seem suitable for modelling the movements of the animals using the daily movement data, and both provide state allocations which are very similar and appear accurate according to the simulation.

Table 8.11 - Phabeni Sable – daily movement comparison of state allocation using Hidden Markov and Bayesian state-space models, with Gamma distribution for step length

		Hidden Markov Model	
		State 1	**State 2**
Bayesian SSM	**State 1**	553	0
	State 2	169	310

The underlying behavioural state inferences which can be drawn from the two methods are very similar, although the Bayesian model took a lot longer time to

fit. The computational time, the complexity of working with WinBUGS, the sensitivity to the choice of starting values and prior distributions, the subjective parameters which need to be defined (iterations, burn-in etc.) and the difficulties of assessing the convergence of the Markov chains make the HMM a far simpler model to fit for these types of data in comparison to the Bayesian state-space model. For modellers who prefer the Bayesian paradigm, the challenges might be worth the effort in order to use models they are more familiar with. However, in terms of the inferences which can be obtained from both analyses, the results are similar and imply that the choice of the model fitting process will not dramatically influence the interpretation of the analysis. More work has been published which extends the basic Bayesian state-space models through the addition of covariates or specifying a hierarchical structure with hyper-priors (Clark 2005, Cressie et al. 2009, Eckert et al. 2008, Haydon et al. 2008). In a situation where an extension of the models is required, the Bayesian framework could be easier. However, as shown in Chapter 6, the HMMs can also be customized and extended to include covariates. From the analyses conducted in this thesis, there does not seem to be any evidence of the Bayesian model providing better results than the HMM.

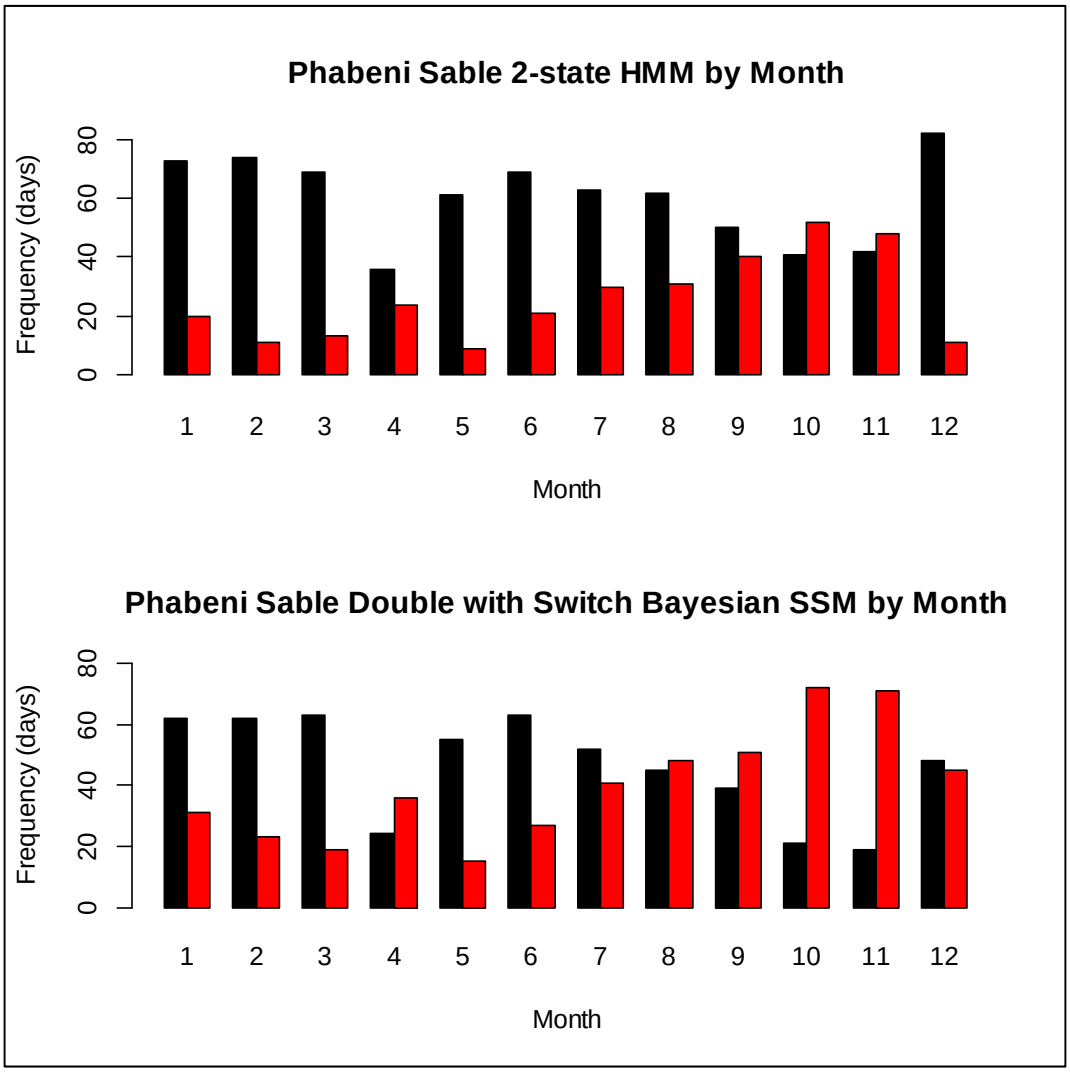

Figure 8.7 - Phabeni sable - Daily movement state allocation by month. Black (encamped), red (exploratory)

8.6 Discussion

The results for the comparison of the independent and the dependent Mixture Models show mostly similar findings, which is expected given the similarities of the model specification. However, this is not always the case and careful consideration needs to be given to the interpretation of the model's latent states. The HMMs provide information about the probability of shifting from one state to another, and the state allocation is not only based on the probability of observing a

displacement between locations of a particular distance. The transition matrix also determines how long an animal is likely to remain within a state. This persistence in one state will in reality be governed by how quickly the resources in an area are depleted. In effect the Mixture Model is a threshold method where observations are allocated to a state with maximum and minimum displacement values associated with each state. This is a result of using the Maximum Likelihood clustering method for the state allocations, since there is a point where the probability that an observation belongs to a particular state switches from being highest for one state to the next state. This was illustrated in Figure 3.8 and Figure 3.9. In contrast, the HMMs use the previous state to inform the current state allocation which makes more sense biologically for animals which will persist in a behavioural state which spans more than one time lag. As the time lags between successive locations decrease, the strength of the autocorrelation within the time series will increase and this should lead to HMMs being the more effective modelling technique. Depending on the desired outcome from the analysis, the time dependent model might not be necessary if a simple categorization only of the observations is required. As the simulation analysis showed, the state allocation accuracy of the IMM is only marginally worse than that of the HMM. The results of the Bayesian SSMs and the HMMs applied to the daily movement data provided similar interpretations of the states. Hidden Markov Models are relatively simple to implement and are computationally not too intensive which makes them a very strong analytical technique for animal movement data. The Bayesian state-space models are far more complicated to apply in comparison (Langrock et al. 2012) and they do not seem to provide any additional insight into the behaviour of the animal. This could be different if informative prior information exists which can be included in the Bayesian model.

Independent Mixture Models are by far the quickest models to fit, and in turn the HMMs are far quicker than the Bayesian models. The Bayesian framework requires specification of the prior distributions and subjective input from the modeller in terms of the number of iterations, burn-in period, thinning etc.

(Pedersen et al. 2011a). Given the simplicity of the HMMs to run, this technique seems preferable and the accuracy of the parameter estimates based on the bootstrap standard errors was always better than the IMMs. Simple exploratory and graphical techniques have been used in this comparison to understand the differences between the methods and the inferences which can be drawn from them. The strength of the model fit was assessed using two techniques, the accuracy of the state allocations provided that the true state was simulated or known from other information, and the confidence intervals and standard errors of the parameter estimates.

An aim of these comparative studies is to identify which model or suite of models appears to be the most suitable for the analysis of these types of data. Models with similar model specifications should provide similar results and interpretations of the states in terms of the behavioural patterns of the animals but this is not always the case. Further work is likely to lead to more techniques or adaptations of current methods being published with animal movement applications, given that the field of movement ecology is relatively new and advancing rapidly. If the results of different analyses consistently provide very different results and contradictory interpretations, there is a fundamental issue with at least one of the models. Ways to determine which the better model is are needed. While some of the work in the literature has compared the performance of more than one model to their data, no consistent way to measure the fit of the various models is used. In many cases the underlying assumptions of the different methods are very different. In some cases, basic graphical comparisons of the output are the only methods presented to compare the model output (Gutenkunst et al. 2007). These are then compared to the known biology for the species. Most often, the model comparison that is done in a study uses nested models, mostly increasing in complexity with the addition of covariates or additional states. These models can be more easily compared using criteria such as the AIC, BIC or DIC (Forester et al. 2007, Morales et al. 2004). However, without knowing the true behaviour, there will always be model selection uncertainty (Horne, Garton 2006). As

discussed in all of the previous chapters, in most cases the true behaviour of the animal is not known, so absolute validation of the models is impossible. However, with advances in technology and tracking equipment, it would not be unexpected for more behavioural data to be remotely sensed. This could be done using cameras or activity collars to gain more "true behavioural" data with which to validate the model output, particularly for studies with short time intervals between observations. While behavioural data collected by an observer is in some cases possible, this is unlikely to provide long-term data and the behaviour of the animal has the potential to be influenced by the observer.

8.7 Conclusion

Hidden Markov models appear to provide a very good balance between computational requirements and model simplicity which allows for reasonably simple interpretation of the latent states in terms of the behaviour of an animal. A number of authors have identified them as promising for modelling animal movements in a heterogeneous environment (Dragon et al. 2012, Barraquand, Benhamou 2008, Patterson et al. 2008, Langrock et al. 2012).

In the end, it is expected that the decision about which method of analysis to use will be based on the questions being addressed about the ecology of the animal including the scale and regularity of the movement and the distances travelled and the desired outcomes of the model (Dragon et al. 2012). If a simple state allocation is required from which the behaviour of the animal will be inferred, then the Independent Mixture Model may be sufficient. If a complex hierarchical model structure is required, the Bayesian methods can be the most suitable. If informative prior information does not exist, the results of the Bayes models with non-informative prior distributions are almost identical to the frequentist approach. In this situation, the added computation time and model complexity for the Bayesian models is not warranted. However, if informative prior information

does exist, this approach could be advantageous. For most animal movement studies, it is expected the Hidden Markov models will provide a satisfactory balance between model complexity and extensibility, while being simple enough to fit and efficient in terms of the need for processing power requirements and model fitting time. Gutenkunst et al. (2007) suggested that the major uncertainty in the analysis of animal movement data comes from an ignorance of the "prefect movement model". This is a fundamental aspect of statistics, that the underlying true model is seldom known. However, with improvements in remote sensing technology, it is likely that the available data will become richer and have the potential to uncover the true behavioural states, which can then be used to validate and refine the movement models.

8.8 Appendix

Table 8.12 – PM_Sable 279 - Fitted model parameters for 2-, 3- and 4-state IMMs and HMMs

	Parameter		90% CI	Std error		Parameter		90% CI	Std error
2-state HMM	-LL	-1319.98			**2-state IMM**	-LL	-1077.06		
	AIC	-2627.95				AIC	-2144.11		
	BIC	-2591.79				BIC	-2113.98		
	$\delta 1$	0.70	(0.67,0.73)	0.02		$\delta 1$	0.66	(0.57,0.69)	0.04
	$\delta 2$	0.30	(0.27,0.33)	0.02		$\delta 2$	0.34	(0.31,0.43)	0.04
	$\mu 1$	-3.14	(-3.21,-3.06)	0.05		$\mu 1$	-3.20	(-3.52,-3.19)	0.10
	$\mu 2$	-0.40	(-0.47,-0.35)	0.04		$\mu 2$	-0.53	(-1.03,-0.73)	0.09
	$\sigma 1$	1.36	(1.31,1.42)	0.03		$\sigma 1$	1.36	(1.22,1.38)	0.05
	$\sigma 2$	0.82	(0.78,0.87)	0.03		$\sigma 2$	0.88	(0.91,1.05)	0.04
	$\Upsilon 11$	0.86	(0.84,0.88)	0.01					
	$\Upsilon 12$	0.14	(0.12,0.16)	0.01					
	$\Upsilon 21$	0.32	(0.29,0.35)	0.02					
	$\Upsilon 22$	0.68	(0.65,0.71)	0.02					
	Proc. time	25.34				Proc. time	1.56		
		6.16							
3-state HMM	-LL	-1419.10			**3-state IMM**	-LL	-1098.38		
	AIC	-2814.19				AIC	-2180.77		
	BIC	-2741.86				BIC	-2132.55		
	$\delta 1$	0.64	(0.61,0.67)	0.02		$\delta 1$	0.65	(0.6,0.72)	0.04
	$\delta 2$	0.28	(0.25,0.31)	0.02		$\delta 2$	0.30	(0.22,0.36)	0.04
	$\delta 3$	0.08	(0.06,0.09)	0.01		$\delta 3$	0.04	(0.03,0.07)	0.01
	$\mu 1$	-3.29	(-3.38,-3.2)	0.05		$\mu 1$	-3.24	(-3.39,-3.05)	0.10
	$\mu 2$	-0.80	(-0.87,-0.74)	0.04		$\mu 2$	-0.71	(-0.86,-0.61)	0.08
	$\mu 3$	0.61	(0.56,0.66)	0.03		$\mu 3$	0.72	(0.6,0.8)	0.06
	$\sigma 1$	1.29	(1.24,1.35)	0.03		$\sigma 1$	1.34	(1.25,1.44)	0.06
	$\sigma 2$	0.71	(0.64,0.75)	0.03		$\sigma 2$	0.78	(0.61,0.87)	0.08
	$\sigma 3$	0.37	(0.33,0.41)	0.02		$\sigma 3$	0.27	(0.19,0.37)	0.06
	$\Upsilon 11$	0.82	(0.8,0.85)	0.01					
	$\Upsilon 12$	0.17	(0.14,0.19)	0.02					
	$\Upsilon 13$	0.01	(0.01,0.02)	0.00					
	$\Upsilon 21$	0.41	(0.37,0.44)	0.02					
	$\Upsilon 22$	0.49	(0.45,0.53)	0.03					

	ϒ23	0.10	(0.08,0.12)	0.01					
	ϒ31	0.01	(0,0.04)	0.02					
	ϒ32	0.45	(0.39,0.51)	0.04					
	ϒ33	0.54	(0.48,0.6)	0.04					
	Proc. time	117.16				Proc. time	4.61		
		3.93							
4-state HMM	-LL	-1477.34			**4-state IMM**	-LL	-1102.32		
	AIC	-2914.69				AIC	-2182.64		
	BIC	-2794.14				BIC	-2116.33		
	$\delta 1$	0.51	(0.46,0.55)	0.03		$\delta 1$	0.59	(0.45,0.66)	0.08
	$\delta 2$	0.25	(0.2,0.31)	0.03		$\delta 2$	0.20	(0.03,0.34)	0.10
	$\delta 3$	0.18	(0.14,0.22)	0.02		$\delta 3$	0.17	(0.05,0.31)	0.08
	$\delta 4$	0.06	(0.05,0.08)	0.01		$\delta 4$	0.05	(0.03,0.08)	0.02
	$\mu 1$	-3.68	(-3.81,-3.56)	0.08		$\mu 1$	-3.42	(-3.76,-3.2)	0.24
	$\mu 2$	-1.66	(-1.79,-1.51)	0.09		$\mu 2$	-1.42	(-2.15,-0.97)	0.38
	$\mu 3$	-0.41	(-0.5,-0.33)	0.05		$\mu 3$	-0.34	(-0.71,-0.04)	0.25
	$\mu 4$	0.69	(0.63,0.75)	0.04		$\mu 4$	0.73	(0.59,0.84)	0.08
	$\sigma 1$	1.13	(1.06,1.19)	0.04		$\sigma 1$	1.26	(1.12,1.36)	0.09
	$\sigma 2$	0.68	(0.57,0.81)	0.07		$\sigma 2$	0.71	(0.23,0.93)	0.22
	$\sigma 3$	0.50	(0.43,0.56)	0.04		$\sigma 3$	0.55	(0.28,0.73)	0.16
	$\sigma 4$	0.31	(0.27,0.35)	0.02		$\sigma 4$	0.29	(0.2,0.36)	0.05
	ϒ11	0.71	(0.66,0.75)	0.03					
	ϒ12	0.20	(0.15,0.26)	0.03					
	ϒ13	0.08	(0.05,0.11)	0.02					
	ϒ14	0.01	(0,0.02)	0.00					
	ϒ21	0.36	(0.32,0.41)	0.03					
	ϒ22	0.42	(0.34,0.5)	0.05					
	ϒ23	0.18	(0.12,0.24)	0.04					
	ϒ24	0.03	(0.01,0.05)	0.01					
	ϒ31	0.27	(0.22,0.33)	0.04					
	ϒ32	0.22	(0.16,0.29)	0.04					
	ϒ33	0.41	(0.36,0.48)	0.04					
	ϒ34	0.09	(0.06,0.12)	0.02					
	ϒ41	0.09	(0.04,0.13)	0.03					
	ϒ42	0.06	(0,0.13)	0.04					
	ϒ43	0.32	(0.25,0.4)	0.05					
	ϒ44	0.53	(0.46,0.59)	0.04					
	Proc. time	302.55				Proc. time	13.66		

9 CONCLUSION

9.1 Summary of the study

This thesis aimed to explore the applications of various statistical methods to movement data obtained using GPS coordinates for the movement of sable antelope, buffalo and zebra in the Kruger National Park, South Africa. It has been acknowledged that the ability to collect GPS tracking data has outpaced the statistical ability to analyse these data. This study has focused on the application of Hidden Markov and other state-space models to these data. These models have been used in previous studies using animal tracking input data and show potential to provide easily interpretable results in terms of the behaviour of the animal. Barraquand & Benhamou (2008) commented that "SSMs and HMMs are thus certainly the most powerful models for segmenting a long movement through a heterogeneous landscape into homogeneous movement bouts when the possible types of behavioural changes occurring along the path in relation to the environment can be a priori anticipated and modelled in terms of Correlated Random Walks or Biased Correlated Random Walks" (Barraquand, Benhamou 2008). A number of different methods were considered at the outset of this study, however the HMMs stood out as a candidate model for these types of analysis. HMMs make some simple assumptions about the data, can be extended to include covariates and provide a simple classification of the observations into latent states. These states can then be interpreted in terms of the behaviour of the animal. To complete the picture, the simpler Independent Mixture Models which assume independence between observations were used as a basic modelling approach. The more complex Bayesian approach was also used to model the same input data. A focus of this study was to compare these methods, which show a natural increase in complexity, and to investigate how the results differ when applied to the same input data. All three methods were shown to provide realistic results and all of

them have their place in the toolbox of methods which can be applied to animal movement data.

In Chapter 2, the study was introduced and background information provided about the sable antelope, buffalo and zebra species. The need for the broad ecological study as a result of the decline in the sable antelope population was discussed. Various aspects relating to animal movement tracking using GPS were investigated, including the expected error associated with each GPS location point using stationary GPS tracking collars.

Chapter 3 provided a basic analysis of the movement tracks using Independent Mixture models. The models were used as a clustering technique to allocate observations to one of a number of latent states found by the model. The models were applied to the hourly movement data for all of the species and mostly identified a 4-state model as the best model fit to the data. These states were assumed to correspond to the animal resting, foraging, mixed movement and travelling. When investigated over the course of a day, the states provided a realistic pattern of movement that one would expect for these types of ungulates. A log-normal distribution was found to provide the best fit. It was clear that the sable herd in the far north of the Park was moving far more than the other species in the area or the sable in the south-west. A novel modification of the state allocation was applied so that each hourly observation could be proportioned into the different behavioural states which had contributed to the observed displacement. For this modification, an offset Gamma mixture model was required. The modification allows for a different interpretation of how the behaviours of the animal are mixed during the time period of each observation.

One of the acknowledged features of animal movement data is that the locations are correlated in space and time. Chapter 4 investigates the strength and pattern of

the temporal autocorrelation of the linear displacements for all of the herds and species. At the hourly time scale, there is a clear pattern of autocorrelation with multiple cycles during the 24 hour period. For the large buffalo herd, this pattern is noticeably consistent throughout the seasons and years. It was as strong at a lag of five days as it was on day one. It was speculated that the patterns of autocorrelation will be stronger for large ruminant herds like the buffalo, while the smaller herds will have slightly weaker correlation. However a pattern is always likely to exist. Further research is needed to show whether these patterns are consistent for other herds and species. This chapter highlighted the need for the analysis to take into account the time series nature of the data and the autocorrelation of movements for these animals. This leads into Chapter 5 which extends the concept of the Mixture Model to the dependent Hidden Markov model.

In Chapter 5 the theory of the HMMs was reviewed as well as some its applications to animal movement in the literature. The log-normal distribution was used to model the hourly movement data, consistent with the IMMs. The Viterbi algorithm was used to determine the most likely sequence of states associated with the observed movement distances. Two methods for fitting HMMs, direct numerical maximization and the EM algorithm, were discussed. Adjustments needed to these methods in order to allow for missing observation data were discussed since missing data are a common feature of remotely sensed movement tracks. A simulation exercise was conducted which showed around 80% of observations were allocated to the correct state, for a true model with parameters similar to those obtained for the sable movement data. The HMMs were also applied to daily movement data which used the linear displacement between locations 24 hours apart, anchored at various times of day. For these daily data, Weibull or Gamma distributions were found to provide the best fit to the data. The HMMs provided results that appear plausible in terms of the expected movement of the animals and since this method formally incorporates

the serial correlation of the input data, it should be an improvement over the simpler IMM.

One of the advantages of using HMMs for the analysis of animal movement data is that the models are flexible and can be customized to include covariates. In Chapter 6, two different methods were used to include a seasonal covariate in the models. Both methods were applied separately to the state-dependent distributions and to the transition probability matrix. The two approaches used either a discrete segmentation of the year into seasonal blocks or a continuously time-varying covariate for the day of the year. The four variations provided reasonable results and are an example of a straightforward method of including covariates into the HMM framework. A reasonably common feature of remote sensing data is missed locations or tracking devices set to record at irregular time intervals to maximise the obtained locations during focus periods for that animal. In a novel modification of the HMM for animal movement, the likelihood function was adapted to allow for input data of irregularly spaced observations. The results were encouraging and this simple modification allows for all of the data to be used in the modelling process when irregular locations have been obtained. Further work is needed to investigate the biological interpretations of the states for the irregular spaced HMMs, but the models provide a positive advance for movement research which does not require interpolation or averaging when locations are missed due to satellite error or changes in frequency of observation.

Chapter 7 introduced Bayesian state-space models which have been used in a number of animal movement studies, often building on the work of Morales et al. (2004) who fitted Bayesian models to the movement of elk in Canada. These models have a very similar framework to the IMMs and HMMs, with the models fitted using MCMC methods. This technique is a natural follow-on from the HMMs and the Bayesian methods have the potential to be extended to hierarchical

models which can include population level parameters, as well as the inclusion of informative prior information in the fitting process.

All three of these methods are compared in Chapter 8. The models were run for exactly the same subsets of data and the model parameters, their standard errors and the state allocation obtained from the models were compared. While these methods have all been used in other studies of animal movement using GPS tracking data, very few studies have actually investigated how multiple methods perform when applied to the same dataset, and how the inferences about the results would change based on the decision of which method to use. A simulation exercise was done in order to compare the results of the methods when the true model was actually known. The HMM outperformed the IMM in terms of the accuracy of the state allocation, and the standard errors for all of the parameter estimates were narrower for the HMM. While this confirms, as expected, that the HMM is the better model when serial correlation exists within the data, the improvement in parameter estimates and state allocation is quite minimal. Similar underlying states were found for the sable herds as were found using the Mixture Models, although some differences were found for the models of the buffalo and zebra data. It was suggested that these differences could be due to the very strong patterns of autocorrelation that were seen in Chapter 4 for the buffalo herd. IMMs are by far the quickest models to fit. The time taken to fit the Bayesian state-space models for large datasets, which are common for remotely sensed tracking data, was a serious challenge towards using this method of analysis. The HMMs appear to provide a good balance between the two, accounting for the correlation but without taking a prohibitive time to fit the model. The results from all three methods are consistent and provide results which are easy to interpret in terms of the behaviour of the animal. Depending on the desired depth of the analysis all three techniques are useful for the analysis of animal movement data. Independent Mixture Models provide a simple classification of the displacement distances into the latent states, which can be improved upon by accounting for the correlation using the Markov property of HMMs. The Bayesian methods add a lot of

computational complexity and would be more valuable when informative prior information exists. Without any informative prior information, the HMMs are the better model choice for these data studied in this thesis. The modifications made to the HMMs for the seasonal data analysis and the irregular time gap are novel extensions which are particularly useful for the application of HMMs to animal movement data. These results concur with the conclusions made by Dragon et al. (2012) who found that HMMs were the most efficient method for inferring foraging behaviour from either ARGOS or GPS tracking data.

9.2 Outcomes

At the outset, this study aimed to achieve six broad objectives. The first objective was to compare and contrast the suitability of different modelling approaches for the analysis of animal movement data. After a review of the variety of models that have been used in the literature, Hidden Markov models and related models were identified as a promising suite of models for these animal movement studies. The IMMs were simple to fit and provide a basic output and categorization of the observations into latent states. The models can be extended to include offsets which allow for alternate interpretations of the state allocation results. These models provided realistic results at both the annual and seasonal scale. The HMMs are the natural extension to the IMM which include the transition probabilities and account for the autocorrelation within the input data. The results were mostly consistent with the IMMs except for a few herds which tended to have the stronger patterns of correlation. These models account for the time series nature of the data and can be extended to include covariates and seem to be the most suitable for the application to movements at this scale for these species. The structure of the Bayesian state-space model is a Markov process, which means that it is essentially an HMM fitted using a Bayesian approach. No informative prior information was used, and with the size of the datasets obtained using the remote sensing collars, the data would dominate the model anyway. These models

appeared to be a computational waste when almost identical results were obtained using frequentist methods.

The second objective was to differentiate the movement patterns among species. The simple IMMs were able to identify different movements between the Punda Maria and Pretoriuskop sable herds, with 4-state models selected for the Punda Maria herd but only 3-state models for the Pretoriuskop herds. Very similar resting and foraging states were identified for all of the herds and species, although the more active states differed across the species. The sable and buffalo moved longer distances in the travelling state compared to the zebra in the same region; while the Phabeni sable herd moved far less than the other Pretoriuskop sable herds. The HMMs and Bayesian SSMs were also able to identify these different active movement states across the herds. A visual inspection of the autocorrelation patterns seems to differentiate the species with a very distinctive double hump associated with the buffalo herd. All the species showed a daily cycle, with rounded peaks which suggested shifting behaviours through the days.

The third objective was to identify movement within foraging patches from those pathways between patches. The daily movement models (both HMM and Bayesian) identified states which persisted longer than one would expect for movements within and between foraging patches, but certainly could be used to identify areas of utilisation. Further spatial interrogation of the movements and fitted states is needed to understand their role in patch identification. These persistent states are as a result of the high probabilities of remaining in the current state.

The fourth objective was to establish how movements varied over the seasonal cycle. Models were fitted in seasonal blocks for the IMMs and identified distributions with different parameters to describe the movements in the different

seasonal blocks. Foraging and resting stayed consistent throughout the year but much longer movements were associated with the active states in the Late Dry season. HMMs were fitted using the day of the year as a covariate or a seasonal categorisation. These were novel applications of the HMM in the animal movement context.

The fifth objective of the study was to identify unusual movement patterns to reveal stressful times for the animal. Other than the longer movements associated with the Late Dry season, this objective was not pursued further. However the potential for further work exists with the opportunity to study the run lengths of each state to identify unusual changes in behaviour. New technology included in current studies also records the temperature of the animal at the same time as the locations are obtained. These data will provide an opportunity to study how the animal may or may not change its behavioural patterns to cope with extreme conditions.

The final objective of the study was to customize and improve the models for their application in animal movement work. Time-varying covariates are not uncommon in the HMM literature, however they have not been seen to have been applied to animal movements in this way, nor has the seasonal block analysis. Both methods provide a practical solution to account for the seasonality in the Hidden Markov movement model. The modification of the likelihood function is a novel way of dealing with the missed locations which are a feature of remotely sensed data, especially since one missed location results in two missing displacements. This application requires further work to improve the interpretation of the underlying states. The IMMs using offsets was used to facilitate a different interpretation of the latent states using proportionate state allocation.

9.3 Recommendations and Future Work

An area of work that would be beneficial to the statistical analyses of animal movement data is around model selection. In this study, it was found that different statistical distributions provided the 'best' fit to the data at different time scales. This needs to be investigated in other studies to determine whether the pattern of log-normal or offset Gamma distributions for the hourly data and Weibull or Gamma distributions for the daily data is consistent for other species and which distributions are the most appropriate at other time intervals. From a biological point of view, an independent way of determining the true behaviour of the animal either through observation or activity collars would provide valuable information to validate the inferences made about the underlying movement states. It would also be possible to use the true behaviour to provide direct estimates of the state dependent distribution parameters.

One of the issues of using these state-space models to uncover the behavioural state of the animal is that the models are based on the assumption that the animal remains within a single state during the time period between successive locations. The posterior probabilities of the Independent Mixture Models were used as a proxy for the proportion of time during the period of observation that the animal spent in each state. However, this is not an ideal solution and violates the assumptions of the model. This issue has not been satisfactorily solved in this thesis. Further work would be valuable in this area, to find statistical answers to the ecological question of how the different states contribute to the observed displacements. A statistical solution to this problem would be very useful to the ecologists trying to understand the smaller scale movements of the animals and how the various states contribute to the observed displacements. As the technology and battery life improve, it will be possible to collect location estimates at a finer resolution which will mean the intervals could be shorter than the time the behavioural state persists for. However shorter time intervals used for the input data will identify different underlying groupings, and the same issue of a

combination of states contributing to the observation may still occur. Statistical methods to tease apart these combinations are important and are also a high priority for future research.

The outputs from this thesis, while providing some basic insight into the behavioural patterns of these three species, can now be included in more elaborate ecological studies with the aim of answering new and different questions about the ecology and behaviour of the animals. These movement models could be used in association with environmental features and other movement data to see how animals behave relative to mountains, rivers or waterholes, availability of food, territorial boundaries, other species, the presence of predators and human disturbance. Of particular concern at the moment is the poaching of black (*Diceros bicornis*) and white rhinoceros (*Ceratotherium simum*). Movement models could potentially be used to identify areas to focus anti-poaching efforts, the patterns of "normal" rhino behaviour and an early warning system triggered by unusual behaviour if near real-time monitoring of the locations was possible.

The issue of irregularly spaced observations is also an area for future work. Although this was touched upon in this thesis, more work is necessary to improve the modified models and provide more clarity on the interpretations of the underlying states. This will enable these types of analyses to be used more usefully for data recorded with finer resolution of observations for certain periods during the day or season. These periods of more frequent recordings will mostly be associated with important bouts of behaviour which the ecologist is trying to investigate.

As mentioned a number of times during this thesis, there is a clear daily cycle to the movements of these animals. All of these models can be modified to include this daily cycle as a covariate. A high priority for future work is to focus on the

implementation of these models including the daily cycle, investigation of the resulting parameter estimates and the influence on the interpretations of the underlying states. This could also tie in with the investigation of allocating more than one state to a single observation, since the combination of likely states contributing to a single observed movement may change through the course of a day.

Statistical modifications to the models will continue to be necessary as different ways of tracking animal movements are employed, for example the altitude and acceleration of birds along with the GPS locations of their flights.

9.4 Conclusion

As data collection capabilities increase both of the tracking data and supplementary environmental data, the challenges and opportunities for statistical modelling work will increase as well. This study contributes to a baseline understanding of how these models perform relative to one another. It has highlighted how modifications can make the most of the available data and include covariates in the models. The paradigm of 'movement ecology' defined by Nathan (2008) will extend and be redefined as more research is done in this exciting and rapidly advancing area of ecological analysis.

REFERENCES

Agostinelli, C. & Lund, U. 2011, *R package 'circular': Circular Statistics (version 0.4-3)*.

Albert, J. 2009, *Bayesian Computation with R,* 2nd edn, Springer, New York.

Apps, P. 1992, *Field Guide to the Behaviour of Southern African Mammals Wild Ways,* Southern Book Publishers, Halfway House, South Africa.

Atkinson, R.P.D., Rhodes, C.J., Macdonald, D.W. & Anderson, R.M. 2002, "Scale-free dynamics in the movement patterns of jackals", *Oikos,* vol. 92, pp. 134-140.

Austin, D., Bowen, W.D. & McMillan, J.I. 2004, "Intraspecific variation in movement patterns: modeling individual behaviour in a large marine predator", *Oikos,* vol. 105, pp. 15-30.

Barraquand, F. & Benhamou, S. 2008, "Animal movements in heterogeneous landscapes: Identifying profitable places and homogenous movement bouts", *Ecology,* vol. 89, no. 12, pp. 3336-3348.

Beichelt, F.E. & Fatti, L.P. 2002, *Stochastic Processes and Their Applications,* 1st edn, CRC Press, Boca Raton.

Bergman, C.M., Schaefer, J.A. & Luttich, S.N. 2000, "Caribou movement as a correlated random walk", *Oecologia,* vol. 123, pp. 364-374.

Beyer, H.L., Morales, J.M., Murray, D. & Fortin, M.-. 2013, "The effectiveness of Bayesian state-space models for estimating behavioural states frommovement paths", *Methods in Ecology and Evolution,* vol. 4, no. 5, pp. 433-441.

Blackwell, P.G. 2003, "Bayesian inference for Markov processes with diffusion and discrete components", *Biometrika,* vol. 90, no. 3, pp. 613-627.

Bolker, B.M. 2008, *Ecological Models and Data in R,* Princeton University Press, Princeton.

Box, G.E.P. & Tiao, G.C. 1973, *Bayesian Inference in Statistical Analysis,* Addison-Wesley, Reading.

Boyce, M.S., Pitt, J., Northrup, J.M., Morehouse, A.T., Knopff, K.H., Cristescu, B. & Stenhouse, G.B. 2010, "Temporal autocorrelation functions for movement rates from global positioning system radiotelemetry data", *Philosophical Transactions of the Royal Society B,* vol. 365, pp. 2213-2219.

Brillinger, D.R., Preisler, H.K., Ager, A.A. & Kie, J.G. 2004, "An exploratory data analysis (EDA) of the paths of moving animals", *Journal of Statistical Planning and Inference,* vol. 122, pp. 43-63.

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,* vol. 365, pp. 2157-2162.

Cain, J.W., III, Owen-Smith, N. & Macandza, V.A. 2012, "The costs of drinking: comparative water dependency of sable antelope and zebra", *Journal of Zoology,* vol. 286, pp. 58-67.

Calenge, C. 2006, "The package “adehabitat” for the R software: A tool for the analysis of space and habitat use by animals", *Ecological Modelling,* vol. 197, pp. 516-519.

Cappé, O., Moulines, E. & Rydén, T. 2005, *Inference in Hidden Markov Models,* Springer, New York.

Casella, G. & George, E.I. 1992, "Explaining the Gibbs Sampler", *The American Statistician,* vol. 46, no. 3, pp. 167-174.

Chib, S. & Greenberg, E. 1995, "Understanding the Metropolis-Hastings Algorithm", *The American Statistician,* vol. 49, no. 4, pp. 327-335.

Chirima, J.G. 2009, *Habitat Suitability Assessments for Sable Antelope*, PhD edn, University of the Witwatersrand, Johannesburg.

Clark, J.S. 2007, *Models for Ecological Data,* Princeton University Press, New Jersey.

Clark, J.S. 2005, "Why environmental scientists are becoming Bayesians", *Ecology Letters,* vol. 8, pp. 2-14.

Clarke, J.S. 2007, *Models for Ecological Data – An Introduction,* Princeton University Press, Princeton.

Codling, E.A., Bearon, R.N. & Thorn, G.J. 2010, "Diffusion about the mean drift location in a biased random walk", *Ecology,* vol. 91, no. 10, pp. 3106-3113.

Crawley, M.J. 2009, *The R Book,* John Wiley & Sons, Chichester.

Cressie, N., Calder, C.A., Clark, J.S., Clark, J.S., Ver Hoef, J.M. & Wikle, C.K. 2009, "Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling", *Ecological Applications,* vol. 19, no. 3, pp. 553-570.

Cushman, S.A., Chase, M. & Griffin, C. 2005, "Elephants in space and time", *Oikos,* vol. 109, pp. 331-341.

Dragon, A.-., Bar-Hen, A., Monestiez, P. & Guinet, C. 2012, "Comparative analysis of methods for inferring successful foraging areas from Argos and GPS tracking data", *Marine Ecology Progress Series,* vol. 452, pp. 253-267.

Eckert, S.A., Moore, J.E., Dunn, D.C., Sagarminaga van Buiten, R., Eckert, K.L. & Halpin, P.N. 2008, "Modeling loggerhead turtle movement in the Mediterranean: Importance of body size and oceanography", *Ecological Applications,* vol. 18, no. 2, pp. 290-308.

Efron, B. & Tibshirani, R.J. 1993, *An Introduction to the Bootstrap,* Chapman & Hall, New York.

Ephraim, Y. & Merhav, N. 2002, "Hidden Markov Processes", *IEEE Transactions on Information Theory,* vol. 48, no. 6, pp. 1518-1569.

Everitt, B.S. & Hand, D.J. 1981, *Finite Mixture Distributions,* Chapman & Hall, London.

Fauchald, P. & Tveraa, T. 2006, "Hierarchical patch dynamics and animal movement pattern", *Oecologia,* vol. 149, pp. 383-395.

Fauchald, P. & Tveraa, T. 2003, "Using first-passage time in the analysis of area-restricted search and habitat selection", *Ecology,* vol. 84, no. 2, pp. 282-288.

Forester, J.D., Ives, A.R., Turner, M.G., Anderson, D.P., Fortin, D., Beyer, H.L., Smith, D.W. & Boyce, M.S. 2007, "State-space models link elk movement patterns to landscape characteristics in Yellowstone National Park", *Ecological Monogrpahs,* vol. 77, no. 2, pp. 285-289.

Franke, A., Caelli, T. & Hudson, R.J. 2004, "Analysis of movements and behaviour of caribou (*Rangifer tarandus*) using hidden Markov models", *Ecological Modelling,* vol. 173, pp. 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,* vol. 197, pp. 237-246.

Friar, J.L., Merrill, E.H., Visscher, D.R., Fortin, D., Beyer, H.L. & Morales, J.M. 2005, "Scales of movement by elk (*Cervus elaphus*) in response to

heterogeneity in forage resources and predation risk", *Landscape Ecology,* vol. 20, pp. 273-287.

Funston, P.J., Mills, M.G.L. & Biggs, H.C. 2001, "Factors affecting the hunting success of male and female lions in the Kruger National Park", *Journal of Zoology (London),* vol. 253, pp. 419-431.

Gelman, A. 2011, 22 June-last update*, Deviance, DIC, AIC, cross-validation, etc*. Available: http://andrewgelman.com/2011/06/22/deviance_dic_ai/ [2013, April/02].

Gelman, A. & Rubin, D.B. 1992, "Inference from Iterative Simulation Using Multiple Sequences", *Statistical Science,* vol. 7, no. 4, pp. 457-472.

Gertenbach, W.P.D. 1983, "Landscapes of the Kruger National Park", *Koedoe - African Protected Area Conservation and Science,* vol. 26, no. 1, pp. 9-121.

Getz, W.M. & Saltz, D. 2008, "A framework for generating and analyzing movement paths on ecological landscapes", *Proceedings of the National Academy of Sciences USA,* vol. 105, no. 49, pp. 19066-19071.

Grant, C.C. & van der Walt, J.L. 2000, "Towards an adaptive management approach for the conservation of rare antelope in the Kruger National Park - outcome of a workshop held in May 2000", *Koedoe - African Protected Area Conservation and Science,* vol. 43, no. 2, pp. 103-112.

Graves, T.A. & Waller, J.S. 2006, "Understanding the Causes of Missed Global Positioning System Telemetry Fixes", *The Journal of Wildlife Management,* vol. 70, no. 3, pp. 844-851.

Gutenkunst, R., Newlands, N., Lutcavage, M. & Edelstein-Keshet, L. 2007, "Inferring resource distributions from Atlantic bluefin tuna movements: An analysis based on net displacement and length of track", *Journal of Theoretical Biology,* vol. 245, pp. 243-257.

Hart, T., Mann, R., Coulson, T., Pettorelli, N. & Trathan, P. 2010, "Behavioural switching in a central place forager: patterns of diving behaviour in the macaroni penguin (*Eudyptes chrysolophus*)", *Marine Biology,* vol. 157, pp. 1543-1553.

Harte, D. 2011, *HiddenMarkov: Hidden Markov Models*, Statistics Research Associates, Wellington.

Haydon, D.T., Morales, J.M., Yott, A., Jenkins, D.A., Rosarre, R. & Fryxell, J.M. 2008, "Socially informed random walks: incorporating group dynamics into models of population spread and growth", *Proceedings of the Royal Society B,* vol. 275, pp. 1101-1109.

Horne, J.S. & Garton, E.O. 2006, "Selecting the Best Home Range Model: An Information-Theoretic Approach", *Ecology,* vol. 87, no. 5, pp. 1146-1152.

Horne, J.S., Garton, E.O., Krone, S.M. & Lewis, J.S. 2007, "Analyzing animal movements using Brownian Bridges", *Ecology,* vol. 88, no. 9, pp. 2354-2363.

Jackson, C.H. 2011, "Multi-State Models for Panel Data: The msm Package for R", *Journal of Statistical Software,* vol. 38, pp. 1-29.

Janis, C. 1976, "The Evolutionary Strategy of the Equidae and the Origins of Rumen and Cecal Digestion", *Evolution,* vol. 30, no. 4, pp. 757-774.

Johnson, C.J., Parker, K.L., Heard, D.C. & Gillingham, M.P. 2002, "Movement parameters of ungulates and scale-specific responses to the environment", *Journal of Animal Ecology,* vol. 71, pp. 225-235.

Jonsen, I.D., Flemming, J.M. & Myers, R.A. 2005, "Robust state-space modelling of animal movement data", *Ecology,* vol. 86, no. 11, pp. 2874-2880.

Jonsen, I.D., Myers, R.A. & Flemming, J.M. 2003, "Meta-analysis of animal movement using state-space models", *Ecology,* vol. 84, no. 11, pp. 3055-3063.

Jonsen, I.D., Myers, R.A. & James, M.C. 2007, "Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model", *Marine Ecology Progress Series,* vol. 337, pp. 255-264.

Jonsen, I.D., Myers, R.A. & James, M.C. 2006, "Robust Hierarchical state-space models reveal diel variation in travel rates of migrating leatherback turtles", *Journal of Animal Ecology,* vol. 75, pp. 1046-1057.

Joubert, S. 2007a, *The Kruger National Park - A History - Volume II,* First edn, High Branching, Johannesburg.

Joubert, S. 2007b, *The Kruger National Park - A History - Volume III,* First edn, High Branching, Johannesburg.

Lambert, P.C., Sutton, A.J., Burton, P.R., Abrams, K.R. & Jones, D.R. 2005, "How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS", *Statistics in Medicine,* vol. 24, pp. 2401-2428.

Langrock, R., Hopcraft, J.G.C., Blackwell, P.G., Goodall, V.L., King, R., Niu, M., Patterson, T.A., Pedersen, M.W., Skarin, A. & Schick, R.S. 2013, "Modelling group dynamic animal movement", *Methods in Ecology and Evolution*, doi: 10.1111/2041-210X.12155.

Langrock, R., King, R., Matthiolpoulos, J., Thomas, L., Fortin, D. & Morales, J.M. 2012, "Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions", *Ecology,* vol. 93, no. 11, pp. 2336-2342.

le Roux, E. 2010, *Habitiat and forage dependency of Sable antelope (Hippotragus Niger) in the Pretorius Kop region of the Kruger National Park*, MSc edn, University of the Witwatersrand, Johannesburg.

Leroux, B.G. & Puterman, M.L. 1992, "Maximum-Penalized-Likelihood Estimation for Independent and Markov- Dependent Mixture Models", *Biometrics,* vol. 48, no. 2, pp. 545-558.

Little, R.J.A. & Rubin, D.B. 2002, *Statistical Analysis with Missing Data,* 2nd edn, Wiley, New York.

Lunn, D.J., Thomas, A., Best, N. & Spiegelhalter, D. 2000, "WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility", *Statistics and Computing,* vol. 10, pp. 325-337.

Macandza, V.A., Owen-Smith, N. & Cain, J.W., III 2012a, "Dynamic spatial partitioning and coexistence among tall grass grazers in an African savanna", *Oikos,* vol. 121, pp. 891-898.

Macandza, V.A., Owen-Smith, N. & Cain, J.W., III 2012b, "Habitat and resource partitioning between abundant and relatively rare grazing ungulates", *Journal of Zoology,* vol. 287, pp. 175-185.

MacDonald, I.L. & Raubenheimer, D. 1995, "Hidden Markov Models and Animal Behaviour", *Biometrical Journal,* vol. 6, pp. 701-712.

MacDonald, I.L. & Zucchini, W. 1997, *Hidden Markov Models and Other Models for Discrete-valued Time Series,* Chapman & Hall, London.

Mårell, A., Ball, J.P. & Hofgaard, A. 2002, "Foraging and movement paths of female reindeer: insights from fractal analysis, correlated random walks, and Lévy flights", *Canadian Journal of Zoology,* vol. 80, no. 5, pp. 854-865.

McClintock, B.T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B.J. & Morales, J.M. 2012, "A general discrete-time modeling framework for animal movement using multistate random walks", *Ecological Monogrpahs,* vol. 82, no. 3, pp. 335-349.

McLachlan, G. & Peel, D. 2000, *Finite Mixture Models,* John Wiley & Sons, New York.

McLachlan, G.J. 1987, "On boostrapping the likelihood-ratio test statistic for the number of components in a normal mixture", *Applied Statistics,* vol. 36, pp. 318-324.

McLachlan, G.J. & Basford, K.E. 1988, *Mixture Models Inference and Applications to Clustering,* Marcel Dekker, New York.

McLachlan, G.J. & Jones, P.N. 1988, "Fitting Mixture Models to Grouped and Truncated Data via the EM Algorithm", *Biometrics,* vol. 44, pp. 571-578.

Morales, J.M. 2013, Personal Communication.

Morales, J.M., Fortin, D., Friar, J. & Merrill, E.H. 2005, "Adaptive models for large herbivore movements in heterogeneous landscapes", *Landscape Ecology,* vol. 20, pp. 301-316.

Morales, J.M., Haydon, D.T., Friar, J., Holsinger, K.E. & Fryxell, J.M. 2004, "Extracting more out of relocation data: building movement models as mixtures of random walks", *Ecology,* vol. 85, pp. 2436-2445.

Muggeo, V.M.R. 2008, *segmented: an R Package to Fit Regression Models with Broken-Line Relationships*, R News, 8/1, 20-25 edn.

Muggeo, V.M.R. 2003, "Estimating regression models with unknown break-points", *Statistics in Medicine,* vol. 22, pp. 3055-3071.

Nathan, R. 2008, "An emerging ecology paradigm", *Proceedings of the National Academy of Sciences USA,* vol. 105, pp. 19050-19051.

Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. & Smouse, P.E. 2008, "A movement ecology paradigm for unifying organismal movement research", *Proceedings of the National Academy of Sciences USA,* vol. 105, pp. 19052-19059.

Ogutu, J. & Owen-Smith, N. 2003, "ENSO, rainfall and temperature influences on extreme population declines among African savanna ungulates", *Ecology Letters,* vol. 6, pp. 412-419.

Ott, T. & van Aarde, R.J. 2010, "Inferred spatial use by elephants is robust to landscape effects on GPS telemetry", *South African Journal of Wildlife Research,* vol. 40, no. 2, pp. 130-138.

Owen-Smith, N. 2013, "Daily movement responses by African savanna ungulates as an indicator of seasonal and annual food stress", *Wildlife Research,* vol. 40, no. 3, pp. 232-240.

Owen-Smith, N. 2002, *Adaptive Herbivore Ecology. From Resources to Populations in Variable Environments,* Cambridge University Press, Cambridge.

Owen-Smith, N., Goodall, V.L. & Fatti, L.P. 2012, "Applying mixture models to derive activity states of large herbivores from movement rates obtained using GPS telemetry", *Wildlife Research,* vol. 39, pp. 452-462.

Patterson, T.A., Basson, M., Bravington, M.V. & Gunn, J.S. 2009, "Classifying movement behaviour in relation to environmental conditions using hidden Markov models", *Journal of Animal Ecology,* vol. 78, pp. 1113-1123.

Patterson, T.A., Thomas, L., Wilcox, C., Ovaskainen, O. & Matthiopoulos, J. 2008, "State-space models of individual animal movement", *Trends in Ecology and Evolution,* vol. 23, no. 2, pp. 87-94.

Pedersen, M.W., Berg, C.W., Thygesen, U.H., Nielsen, A. & Madsen, H. 2011a, "Estimation methods for nonlinear state-space models in ecology", *Ecological Modelling,* vol. 222, pp. 1394-1400.

Pedersen, M.W., Patterson, T.A., Thygesen, U.H. & Madsen, H. 2011b, "Estimating animal behaviour and residency from movement data", *Oikos,* vol. 120, no. 9, pp. 1281-1290.

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,* vol. 65, pp. 2367-2377.

Plummer, M., Best, N., Cowles, K. & Vines, K. 2006, "CODA: Convergence Diagnosis and Output Analysis for MCMC", *R News,* vol. 6, pp. 7-11.

Polansky, L., Wittemyer, G., Cross, P.C., Tambling, C.J. & Getz, W.M. 2010, "From moonlight to movement and synchronized randomness: Fourier and wavelet analyses of animal location time series data", *Ecology,* vol. 91, no. 5, pp. 1506-1518.

Preisler, H.K., Ager, A.A., Johnson, B.K. & Kie, J.G. 2004, "Modeling animal movements using stochastic differential equations", *Environmetrics,* vol. 15, pp. 643-657.

Press, S.J. 2003, *Subjective and Objective Bayesian Statistics,* 2nd edn, John Wiley & Sons, Hoboken, New Jersey.

Quinn, G.P. & Keough, M.J. 2002, *Experimental Design and Data Analysis for Biologists,* Cambridge University Press, Cambridge.

R Development Core Team 2011, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing edn, Vienna, Austria.

R Development Core Team 2010, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rahimi, S. & Owen-Smith, N. 2007, "Movement patterns of sable antelope in the Kruger National Park from GPS/GSM collars: a preliminary assessment", *South African Journal of Wildlife,* vol. 37, no. 2, pp. 143-151.

Redfern, J.V., Grant, R., Biggs, H.C. & Getz, W.M. 2003, "Surface-Water Constraints on Herbivore Foraging in the Kruger National Park, South Africa", *Ecology,* vol. 84, no. 8, pp. 2092-2107.

Robert, C.P. 1994, *The Bayesian Choice,* Springer-Verlag, New York.

Roberts, S., Guilford, T., Rezek, I. & Biro, D. 2004, "Positional entropy during pigeon homing I: application of Bayesian latent state modelling", *Journal of Theoretical Biology,* vol. 227, pp. 39-50.

Schick, R.S., Loarie, S.R., Colchero, F., Best, B.D., Boustany, A., Conde, D.A., Halpin, P.N., Joppa, L.N., McClellen, C.M. & Clark, J.S. 2008, "Understanding movement data and movement processes: current and emerging directions", *Ecology Letters,* vol. 11, pp. 1338-1350.

Scientific Software Development - Dr. Lin Himmelmann and www.linhi.com 2010, *HMM: HMM - Hidden Markov Models*.

Sibert, J.R., Lutcavage, M.E., Nielsen, A., Brill, R.W. & Wilson, S.G. 2006, "Interannual variation in large-scale movement of Atlantic bluefin tuna (*Thunnus thynnus*) determined from pop-up satellite archival tags", *Canadian Journal of Fisheries and Aquatic Sciences,* vol. 63, pp. 2154-2166.

Sibly, R.M., Nott, H.M.R. & Fletcher, D.J. 1990, "Splitting behaviour into bouts", *Animal Behaviour,* vol. 39, pp. 63-69.

Slater, P.J.B. & Lester, N.P. 1982, "Minimising errors in splitting behaviour into bouts", *Behaviour,* vol. 79, no. 2/4, pp. 153-161.

Smouse, P.E., Focardi, S., Moorcroft, P.R., Kie, J.G., Forester, J.D. & Morales, J.M. 2010, "Stochastic modelling of animal movement", *Philosophical Transactions of the Royal Society B,* vol. 365, pp. 2201-2211.

S-plus original by Ulric Lund and R port by Claudio Agostinelli 2012, *CircStats: Circular Statistics, from "Topics in circular Statistics" (2001)*, R package version 0.2-4 edn, http://CRAN.R-project.org/package=CircStats.

Spreij, P. 2001, "On the Markov property of a "nite hidden Markov chain", *Statistics & Probability Letters,* vol. 52, pp. 279-288.

Sturtz, S., Ligges, U. & Gelman, A. 2005, "R2WinBUGS: A package for running WinBUGS from R", *Journal of Statistical Software,* vol. 12, no. 3, pp. 1-16.

Tambling, C.J., Cameron, E.Z., du Toit, J.T. & Getz, W.M. 2010, "Methods for Locating African Lion Kills Using Global Positioning System Movement Data", *Journal of Wildlife Management,* vol. 74, no. 3, pp. 549-556.

Tambling, C.J., Laurence, S.D., Bellan, S.E., Cameron, E.Z., du Toit, J.T. & Getz, W.M. 2011, "Estimating carnivoran diets using a combination of carcass observations and scats from GPS clusters", *Journal of Zoology,* vol. 286, no. 2, pp. 102-109.

Tierney, L. 1994, "Markov Chains for Exploring Posterior Distributions", *Annals of Statistics,* vol. 22, pp. 1701-1762.

Titterington, D.M., Smith, A.F. & Makov, U.E. 1985, *Statistical Analysis of Finite Mixture Distributions,* John Wiley & Sons, Chichester.

Tolkamp, B.J. & Kyriazakis, I. 1999, "To split behaviour into bouts, log-transform the intervals", *Animal Behaviour,* vol. 57, pp. 807-817.

Tomkiewicz, S.M., Fuller, M.R., Kie, J.G. & Bates, K.K. 2010, "Global positioning system and associated technologies in animal behaviour and ecological research", *Philosophical Transactions of the Royal Society B,* vol. 365, pp. 2163-2176.

Tucker, B.C. & Anand, M. 2005, "On the use of stationary versus hidden Markov models to detect simple versus complex ecological dynamics", *Ecological Modelling,* vol. 185, pp. 177-193.

Viswanathan, G.M., Afanasyev, V., Buldyrev, S.V., Murphy, E.J., Prince, P.A. & Stanley, H.E. 1996, "Levy flight search patterns of wandering albatrosses", *Nature,* vol. 381, pp. 413-415.

Viswanathan, G.M., Buldyrev, S.V., Havlin, S., Da Luz, M.G.E., Raposo, E.P. & Stanley, H.E. 1999, "Optimizing the success of random searches", *Nature,* vol. 401, pp. 911-914.

Vuong, W.H. 1989, "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses", *Econometrica,* vol. 57, no. 2, pp. 307-333.

Wackerly, D.D., Mendenhall, W. & Scheaffer, R.L. 1996, *Mathematical Statistics with Applications,* Duxbury Press, Belmont.

Wittemyer, G., Polansky, L., Douglas-Hamilton, I. & Getz, W.M. 2008, "Disentangling the effects of forage, social rank, and risk on movement autocorrelation of elephants using Fourier and wavelet analyses", *Proceedings of the National Academy of Sciences USA,* vol. 105, no. 49, pp. 19108-19113.

Zucchini, W. 2000, "An introduction to model selection", *Journal of Mathematical Psychology,* vol. 44, pp. 41-61.

Zucchini, W. & MacDonald, I.L. 2009, *Hidden Markov Models for Time Series: An Introduction Using R,* Chapman & Hall/CRC, Boca Raton.

Zucchini, W., Raubenheimer, D. & MacDonald, I.L. 2008, "Modeling Time Series of Animal Behavior by Means of a Latent-State Model with Feedback", *Biometrics,* vol. 64, pp. 807-815.