Bayesian Inference for Animal Movement in Continuous Time

The University Of Sheffield.

Hajar Alkhezi

Supervisor: Paul Blackwell

The University of Sheffield
Faculty of Science
School of Mathematics and Statistics

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

September 2019

I would like to dedicate this thesis to my loving parents ...

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This thesis contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Hajar Alkhezi
Supervisor: Paul Blackwell
September 2019

Acknowledgements

I am grateful to my great supervisor Prof. Paul Blackwell for all the help, support, knowledge and for his patience. It was fantastic to have the opportunity to work with you. Also, I thank my colleagues, my friends, my family especially my lovely niece.

Abstract

Movement is an essential process for almost all species in the animal kingdom. For example, survival is the result of successfully finding food and avoiding predators. Also, reproductive success depends on encounters with mates. Over the last century, scientists have developed statistical methods to understand the nature and behaviour of species and have thus shed light on animal movement. The more we study the details of animal movement, the more we understand animal behaviour. Naturally and realistically, animal movement happens in continuous time, but movement data on tagged animals are observed in discrete time. Many statistical methods ignore this fact or consider animal movement to be time-continuous but use a discrete time approximation, which involves errors that are difficult to calculate. Blackwell et al. (2016; Methods in Ecology and Evolution, 7, 184-195) introduced a new statistical method to analyse complex animal movement with environmental information in continuous time, specifically without the need for approximation. In this thesis, continuous time Ornstein-Uhlenbeck (OU) diffusion processes have been used to model animal movement data, and Markov Chain Monte Carlo (MCMC) Bayesian methods have been applied to make inferences about the model. The purpose of this thesis is to extend and refine recently developed methods for statistically analysing animal movement data in continuous time. In practical terms, it aims to improve the efficiency of current algorithms and to allow more general models to be applied. The goals of this thesis are to extend the current continuous time models to allow for the estimation of unknown boundaries for the animal's home range or between different habitats, to extend the possible range of prior distributions that can be used for the behavioural process, and to generalise the model to allow for semi-Markov modelling. The information gained using these methods will help ecologists to learn more about animal behaviour and the environment. Ecologists study animal movement to explain the relationship between animal movement and major habitat features, interactions between species and how the animals use the habitat. We apply our methods to real data and explore their performance using simulated data.

Keyword: Animal Movement, Ornstein-Uhlenbeck process, Exact simulation, Markov Chain Monte Carlo, Switch diffusion models, semi-Markov model, boundaries.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Thesis aim

Animal movement includes any movement or method that an animal uses to move or transport itself from one place to another, such as running, jumping, swimming, flying and hopping. Animals move for several reasons: to find food, to find mates, to explore, to escape from a predator or for migration. As a result, the ability to move is essential to survival. Animal movements have been extensively investigated because movement is one of the most important parameters of ecological processes. Furthermore, movement studies assist ecologists in improving their understanding about animal behaviour and how an individual's movement is affected by the surrounding environment and interaction with other animals.

The overall goal of this thesis is to develop, improve and generalise continuous time movement models. The specific aims of this thesis are:

1. To develop a better tool for analysing animal movement data, in order to answer ecologists' questions.
2. To improve the efficiency of continuous time modelling and continuous time estimation.
3. To extend the range of the continuous time models.

We achieve these aims by extending the range of the continuous time models to include estimation an unknown boundary and semi-Markov modelling, by proposing an approximation to the existing exact method of inference (converting it from a variable-dimension problem to a fixed-dimension one), and by reformulation of the exact method to allow it to be more adaptive.

1.2 Thesis outline

This thesis is structured as follows. Chapter 1 presents the general ideas used in animal movement studies by describing the home range, territory, utilisation distribution and centre of attraction, providing details about how to collect and present animal movement data, and the history of the beginning of animal movement methods, which focused mainly on home range estimation. Chapter 2 discusses definitions of animal behaviour, presents a literature review of the more widespread recent methods used to fit movement data that incorporate animal behaviour, such as hidden Markov models (HMMs) and state space models (SSMs), and compares discrete and continuous time models. Chapter 3 presents a detailed discussion of the Ornstein-Uhlenbeck (OU) diffusion model, describes an existing recent algorithm and discusses some of the implementation issues. Chapter 4 presents a new approximation idea in which there is one potential change in behaviour between each pair of observations to speed the algorithm. In Chapter 5, the exact method is improved by allowing kappa, a key quantity in the exact estimation of continuous-time models, generally taken to be constant, to vary over the iterations of the algorithm. The aim is to save computational cost and increase the accuracy of the parameter estimation. In Chapter 6, the exact method is adjusted to estimate an unknown boundary between regions in space in order to learn more about the environment from movement data. In Chapter 7, the variable kappa idea is generalised to deal with semi-Markov models, where the duration time for behaviours can follow any continuous distribution instead of being limited to the exponential distribution as in the Markov case. Chapter 8 presents a summary and future extensions.

1.3 Home range and territory

In the first animal movement studies, the primary task was to estimate the animals home ranges (Jennrich and Turner 1969). Burt (1943) stated that the first and essential point is to understand the difference between the home range and the territory. Burt (1943) distinguished between home range and territory, defining the home range as the region that a single animal or group of animals typically uses around the den or nest within which the animal accomplishes daily activities and the territory as the guarded, defended and protected area. Burt (1943) defined "the home range as the area that the animal uses for its normal activities, such as collecting food, mating and raising young". Areas of nature that are simply explored by an animal should not be considered a part of the home range. Seton (1909) said no animal moves completely at random. Rather, each animal covers a consistent general area, which is called its home range. Jennrich and Turner (1969) defined "the home range as the smallest

region that accounts for 95% of an animal's habitat utilisation". Burt (1943) demonstrated that most animals have a home range but that not all animals have a territory; also, animals do not all engage in territorial behaviour to the same degree. Burt (1943) states there are two main types of territoriality: areas reserved for storing food and shelter and those intended for breeding and the rearing of young animals. These two categories may be further divided in special cases.

Burt (1943) explained that individuals of most species spend their lives in a specific area, which is called the home range. However, some animals migrate from one area to another during certain seasons. Burt (1943) clarified that animals do not necessarily have the same home range throughout their lives; this is especially true for migrant animals, which have different home ranges in winter and summer. Every species chooses and deals with its home range and territory area in different ways. A territory can be the same area as a home range, or it can serve only as a nest site. Burt (1943) has found that territories are often smaller than home ranges because they require a great deal of energy and time to maintain. The sizes of a home range and territory depend on an animal's size, age and sex; the season; the type of habitat; food availability and interactions with other species. Cowlishaw (2014) stated that identifying the size of a home range and territory is complicated. Cowlishaw (2014) explained that the two important ecological factors affecting territory area are population size and food availability. For example, Cowlishaw's (2014) studies of damselfish show that when the population is high and there is a large amount of food available, the territory area tends to be small. However, when there is a high population size but a low amount of food available, the animals will leave and relinquish their territory area because protecting the territory area hinders the animals from exploring other areas in search of food.

Burt (1943) wrote that home ranges may sometimes overlap within species. This overlap is not part of the territory area. Some species will tend to avoid one another rather than seeking to expel one another. MacDonald et al. (1980) stated that interactions between animals can be split into two categories based on their home ranges and movement. The first is static interaction, which pertains to the proportion of home range overlap between two or more animals of the same species. The second is dynamic interaction, which refers to the relation between the movements of two animals, or the way in which the movements of two animals are connected in terms of attraction or avoidance. Dunn and Gipson (1977) examined the movements of twin deer and found that most of the time the twins stayed very close to one another. In contrast, MacDonald et al. (1980) investigated the dynamic interaction between male and female foxes, observing that these foxes typically avoided each other.

1.3.1 Utilisation distribution and the centre

Don and Rennolls (1983) stated that a home range can be delineated as one region or divided into more than one area, each with well-defined boundaries. They referred to these regions as animal domains. Don and Rennolls (1983) explained that species do not utilise all of their home ranges with equivalent density. Specific areas are used more than others. Jennrich and Turner (1969) explained that another term related to the home range is utilisation distribution (occupation density distribution), which is used to identify where the animal may be at a specific time. In other words, it indicates the probability of discovering an animal in specific parts of its home range. Van Winkle (1975) referred to utilisation distribution as the "proportional to the bivariate probability distribution of an animal's position over time". Considering this, even a home range that is formed of two or more disjointed area can be estimated. Anderson (1982) stated that utilisation distribution describes space use but not necessarily resource use.

Dixon and Chapman (1980) referred to "the activity centre as the zone inside the home range with the highest level of activity". Every home range has a statistical centre and typically one or more biological centres. Hayne (1949) identified the statistical centre as the mean coordinates of the observed animal location points or the mean of the utilisation distribution. This statistical centre of activity usually has little or no biological meaning. Don and Rennolls (1983) explained that the true biological centres of activity (centre of attraction) are biological attraction points, such as nest sites, food sources, small pools, resting sites and feeding places. Don and Rennolls (1983) called these centres nuclei, and an animal typically has several nuclei within a home range, as reflected in the animal's choice of multiple resting sites. Don and Rennolls (1983) assumed that the locations of these biological centres are often known by a field worker. None of these nuclei are necessarily located at the statistical activity centre. These nuclei have an important effect on the habitat utilisation intensity of a specific animal because the animal frequently returns to these attraction sites.

1.4 Movement data

Animal tracking data help ecologists to understand the movement of animals within their habitats or as they migrate. This information helps ecologists to understand the effect of environmental challenges such as the passing of infection to other animals and humans, the effects of enviromental change on animal behaviours and species loss. Scientists study animal movement to gain more information about animal populations, major habitat features, interactions between species and how the animals use the habitat. Understanding an animal's

life requires studying its behaviour. Ecologists often want to study the causes and reasons for animal movement in a habitat as well as to determine what affects the animals' movement. In addition, they seek to clarify how the animals interact with other species and the habitat (Hooten et al. 2017). In addition, recently, ecologists have been wanting to extract more detail from the movement data, particularly about where animals go, their behaviour and how they interact with each other and the environment (Hooten et al. 2017). The statistical methods used to understand the nature and behaviour of species require data on animal locations across time (Worton 1987).

Movement data can be collected using various techniques depending on the study's budget and the animal in question. Spatial data may be collected via radio tracking, live trapping or other methods. At the beginning of the twentieth century, the live trapping method was typically used to collect data on animal location by attaching a band, collar or ring with a code or number to the animal, then releasing the animal back into the habitat. To record the data, the animal had to be seen or captured again. The tag did not contain any electronics and was typically used with birds. The tag was lightweight and cheap. However, a large number of animals had to be tagged. because the tags yield only very limited data on location as animals are only observed very intermittently. The tag itself is inexpensive, but collecting the data is very labour-intensive. This form of data is called capture-recapture data and will not be discussed any further in this thesis.

The other method involves attaching an electronic tag which will provide telemetry tracking such as via radio or satellite, such as a global positioning system (GPS) and the Advanced Research and Global Observation Satellite (Argos), an acoustic monitor, an accelerometer, geolocation/TDR tags or other technologies. The electronic tracking approach also became popular because it can provide extensive data, and the time interval between the observations is often minimal. Electronic tracking data are collected by immobilising the animal, attaching a small electronic tag and then releasing the animal. GPS data can supply high-resolution, longitudinal data over massive distances.

Hooten et al. (2017) explained that animal movement data are the result of recording series of spatial points. Worton (1987) states that animal movement data are collections of animal locations or positions over time (discrete points), representing a type of time series data. Most animal data are on a horizontal plane, which is two dimensional. The exceptions to this are aerial and marine wildlife, where the data can be three dimensional or vertical (height or depth), representing a movement in one dimension. For example, Natvig and

Subbey (2011) analysed vertical movement data collected by electronic bands attached to fish. The tags recorded the depth, then the researchers used a mixed OU process with a transition matrix for states in which to analyse the data. Marine data will not be discussed any further in this thesis. Movement data are generated by frequently recording the location of an animal. The time interval between the observed locations can be regular or irregular because of missing data, the design of the model or limitations on the sensor. The time interval can be seconds, minutes, hours or days. If studying behaviour is the aim of the research, the time interval between observations should be short. Thus, there should be a balance between the time interval and aim of the research. The accuracy of estimating animal behaviour is affected by the sample interval between observations; if there is a large interval, it may be difficult to understand some of the behaviour. Thus, ecologists should be careful when collecting the data to choose a time interval suitable for the aim of the study.

Movement data can be represented in different forms. Most tags produced raw (unprocessed) data in the form of animals' positions/locations over time, but most studies use processed data, such as measuring the distance between two locations, which is called the step length, determining the direction of the movement or any change in the direction between locations, which is called the turning angle, by calculating the angles between three locations. The step length may be divided by the length of the time interval to give a measure of speed. The step length and turning angle depend on the animal's initial location and the direction the animal is heading (bearing).

Tags that record only accelerometry data do not directly give locations, and are outside the scope of this thesis. The electronic tracking method provides data on an animal's location and, in some monitors, on what it is doing and with which other animal it is interacting. Various monitors are used to supply knowledge and guidance about animal behaviours that cannot be identified from location data alone. GPS location data do not provide direct information on animal behaviour. Data on animal movement and behaviour complement one other, thereby enabling scholars to explain the social system that underlies species interactions. For example, Nams (2014) combined animal movements and behavioural data. Pedersen and Weng (2013) used SSMs and OU processes to estimate animal movements from observation network data; they used fixed monitors to collect animal presence and absence data in specific areas. Their observation network data were similar to the capture-recapture idea but without the need for great effort from ecologists; moreover, their approach provides continuous information. Description about animal behaviours can be found in Section 2.1.1.

Sometimes, an ecologist probes into the movement of a group of animals of the same species and combines data regarding many individual animals. The animal movement data considered in this thesis are for a single animal and provide information only about animal location and the time of observation.

1.5 History of home range estimation methods

Macdonald et al. (1980) stated that the starting point for understanding animal movement is studying the size and shape of the home range and how the animal utilises their home range. The methods discussed in this chapter have been used for analysing animal movement data to provide the ecologist with a range of essential information, such as the time an animal spends in a specific location and the size and shape of the home range.

In the early movement methods literature, the animal movement data were usually analysed using three main methods: polygon methods, nonparametric methods for utilisation distribution and parametric methods, such as a bivariate normal distribution. Worton (1987) explained that the polygon method is a nonparametric method that provides information only about the size and shape of the home range, whereas the other approaches provide information about the animal's home range use intensity in addition to the size and shape of the home range.

1.5.1 Polygon method

In the early stages of home range scholarship, researchers employed polygon methods. At that time, polygon methods were the most popular and convenient approach to estimate home range. The polygon method assumes that the region is in the form of a convex polygon and derives this representation by drawing the boundary around the outermost capture observation points. Worton (1987) said the polygon method defines "the home range as the smallest polygon comprising all the position points", the minimum convex polygon (MCP) or convex hull. The attractiveness of this method lies in its simplicity; it is easy to calculate by hand, quick and reasonably statistically stable. However, the polygon method has two disadvantages: first, this method is characterised by biases in terms of sample size; an increase in sample size expands the area a home range occupies. Second, it make an unreasonable assumption that the home range shape should be a convex polygon, which is often implausible, especially when the surrounding environment is nonhomogeneous. This method produces overestimates of home range size when the home range is not convex.

This discrepancy in size is attributed to the inclusion of areas that an animal never visits or can never visit. The polygon methods were once widely used but are now regarded as inappropriate because these methods provide only the home range size and shape but do not provide any information about the utilisation distribution and which parts of the home range are most utilised.

1.5.2 Nonparametric utilisation distribution methods

The nonparametric methods described in this section produce more information about how an animal utilises the environment than the polygon method. Worton (1987) explained that nonparametric home range methods have no assumption about the format of the utilisation distribution and make it easier to investigate the effect of different factors on the animal's movement. Voigt and Tinline (1980) presented the grid method as a simple nonparametric method that draws the animal's movement according to the frequency with which it is located in each grid cell or square. Worton (1986) did not consider the grid method a statistical method or a viable model, adding that in the grid method animal movements are restricted solely to the types of moves that are allowed in a game of chess. However, if the main interest is in exploring the overall spatial distribution of directions of movement within the data, this method is very helpful.

To deal with the issue of small sample size, Ford and Krumme (1979) established an approach known as population utilisation distribution (PUD), which calculates the average utilisation for all the inhabitants by merging data for individual animals from the population. They estimated the population utilisation distribution by assuming that the habitat is divided into regions that are equal in size, and assuming that the total of the probabilities of all regions was equal to one, which means it is impossible for the animal to depart from the study area. They then calculated the population utilisation distribution by solving equations that equate the empirical Relocation distance function (RDF) to the RDF implied by the population utilisation distribution (PUD). Ford and Krumme (1979) calculated the Relocation distance function which is the equation for the frequency of finding an animal at each possible distance from its previous location. The series of equations can be solved by minimising the squared difference between the empirical Relocation distance function and the Relocation distance function implied by the population utilisation distribution.

$$\text{Empirical RDF} - \text{RDF implied by the PUD} = 0.$$

The advantage of this method is that it can be used to calculate the utilisation distribution for an individual animal and a group of animals. The disadvantage is that it requires high computation cost, and the results are complicated to explain because the method incorporates data for many animals that could be from different areas.

Anderson (1982) developed a nonparametric method to estimate the utilisation distribution for a single animal using Fourier transform methods for smoothing two-dimensional locations. Worton (1987) suggested the kernel method as another nonparametric method that can be applied to estimate the home range by smoothing the animal's location. Worton (1987) compared the Fourier transform with the kernel method and found that both have disadvantages and give similar estimation results; nevertheless, biologists prefer the kernel method because it is easily interpretable and is more reliable than other nonparametric methods. Worton (1987) concluded that to estimate utilisation distribution, it is best to use the method that makes no assumptions about the underlying distribution. Additionally, Worton (1987) suggested that when using nonparametric methods to analyse telemetry data, a test for independence should be conducted because these methods assume independence between observations. There are other nonparametric methods that are not mentioned in this section due to their limited relevance, such as the harmonic mean (Dixon and Chapman,1980).

1.5.3 Bivariate normal methods

Because the probabilities of an individual appearing in specific areas of a home range are unequal, scientists have for a long time been striving to develop a model to calculate such probabilities. Calhoun and Casby (1958) assumed the animal locations are independently sampled from a bivariate normal distribution, which is the utilisation distribution. Calhoun and Casby (1958) developed three assumptions regarding the home range: first, the home range is stationary that is, it is fixed and exhibits no shift in its centre of activity; second, a true centre of activity (biological attraction) exists, but the statistical centre may differ from it and, third, the probability of an animal appearing in a specific location of the home range decreases with increasing distance from the true centre of activity (Calhoun and Casby 1958). These assumptions suggest a bivariate normal distribution. Calhoun and Casby (1958) defined the capture radius as the distance between the biological centre and the observation point. This distance can be measured to understand how the animals interact with the environment and to estimate the home range boundaries. Calhoun and Casby (1958) excluded any sample that was characterised by a short time interval between two observations because such durations cause dependence between observations, and their model assumes independence between observations. This model can estimate the home range size and boundaries, the

probability of an animal appearing in a specific area and the time an animal spends in a certain location. Van Winkle (1975) said this method measures how far the animal moves from the centre, not the area covered during daily movement.

Calhoun and Casby (1958) restricted their models by assuming that the home range is circular. However, the credibility of such a finding is doubtful because the method affects the determination of the shape of the home range. If circular symmetry fails, the home range will be overestimated. The assumption by Calhoun and Casby of a circular home range thereby makes it unsuitable as a general home range model (Don and Rennolls 1983). Jennrich and Turner (1969) created a more general model by removing the circularity assumption from Calhoun and Casby's method. The unrestricted bivariate normal model became popular because it was the only parametric method that was free of the assumption of circularity and has no sample size bias. However, Macdonald et al. (1980) explain that Jennrich and Turner's method has little biological explanatory power because it does not include any information about different behaviours (modes of movement) at different locations. Given the lack of independence between radio tracking observations, this method cannot be generalised to all home range data. If the utilisation distribution is unknown, then Jennrich and Turner's method may not fit all data well.

1.5.4 Conclusion

Although the conception of a home range has changed over the last century, researchers have not reached a consensus regarding the approach that best represents movement data. Various methods are applied to analyse movement data, and each approach can yield different results. All the previous methods assumed a fixed home range, meaning the animal did not travel outside the study area. Each technique is characterised by limitations and biases in the estimation of home ranges. Van Winkle (1975) states that none of the parametric methods can be used with a non-homogeneous habitat. Anderson (1982) compared his nonparametric method with polygon convex and bivariate normal distribution, concluding that none of these models are free from problems. In his comparison of the nonparametric and bivariate home range models, Van Winkle (1975) argues for the bivariate normal method's appropriateness as an approximation of the utilisation distribution. The bivariate normal method is more informative than the polygon method because the polygon method provides information regarding only the home range's size, shape and position. The polygon method provides information about only the home range boundary, but it does not provide information about what happens inside that region. In the polygon method, the time of the observation is usually completely ignored (i.e. not included in the model), which can potentially be very

misleading. The bivariate normal method is free from sample size bias, but it is based on a very strong assumption about the shape of the utilisation distribution. A home range's size, shape, position and orientation are insufficient for an ecologist who is particularly interested in the estimation of the utilisation distribution because it has additional information that can explain habitat use. Macdonald et al. (1980) state that a complicated method does not necessarily yield a better estimate. As none of the previous methods is without faults, none of them is appropriate for all cases.

Trapping data with long periods between capture observations provide independent location records, but the use of electronic tracking data and decreased time between captures may led to dependence between the location observations, which means the current positions of an animal will be correlated with the previous locations. Dunn and Gipson (1977) suggest the use of a diffusion model to solve the autocorrelation problem, which will be discussed in detail in Chapter 3. Otherwise, when using electronic tracking data, the time interval between observations must be large to ensure independence. The nature of animal movement may be insufficiently explained in the previous methods because of a lack of any information about animal behaviour. Frequently, the boundaries of a home range calculated as described above cannot be interpreted biologically (MacDonald et al. 1980). A statistician endeavours to develop a home range model by delving into how changes in physical and biological variables affect animal movement. An ecologist may know the location of the biological centre of activity for some animals, but he or she is unable to incorporate this information into a home range model. Macdonald et al. (1980) stated that the main drawback of all the methods described above is that they do not necessarily distinguish between different behaviours at different locations. Van Winkle (1975) recommends generating a model to study how animal movement is affected by changes in the environment.

Don and Rennolls (1983) stated that there is still no agreement regarding which method gives the most valuable results. The choice of a method for analysing spatial data depends on the purpose of the study and the type of data. Anderson (1982) stated that various methods can be used to estimate home range but that these estimates cannot be compared, because there is no agreement about the definition of a home range. All of the methods described here assumed independence between observations and do not include any behavioural information. Because of this, nowadays, the researcher is seeking more suitable models to analyse animal movement data. Models that incorporate dependence between observations and different behaviours can give a more realistic estimation of the home range. Statisticians have developed more advanced approaches to tackle these problems, which will be described

in Chapter 2. The best movement models include the effects of the environment on animal behaviour in order to understand how the animal utilises the space.

Chapter 2

Recent Animal Movement Methods

2.1 Introduction

Different data require different analytical methods, and data collection methods have developed significantly in recent years. However, a gap between the technology used in collecting data and the statistical methods used remains. Don and Rennolls (1983) stated that the choice of model is determined by the aims of the research, the type of data and the ecological question being pursued in the research.

The previous methods described in Section 1.5 were aimed at estimating the home range. If scientists want to learn only about the home range, they use these methods. However, most of the models described in Chapter 1 are no longer in use because they are too simple and do not provide enough information. For this reason, statisticians have incorporated behaviour into the movement models, which results in a more complicated but more realistic model in regard to utilisation distribution. The methods described in the current chapter are used to model a more complicated and realisitc model for utilisation distribution. Moreover, when electronic tags are used, the time between observations in the movement data is greatly reduced, which results in dependency between observations. However, all the methods described in Chapter 1 assume that the observations are independent.

The models described in this chapter are more widespread than those developed in the past. They can deal with large datasets and allow for behavioural movement processes. The most widespread groups of statistical methods for analysing animal movement data are as follows; all are stochastic processes:

- Hidden Markov models (HMMs): These models involve discrete time, continuous space and no measurement errors.
- State-space models (SSMs): These models represent discrete time, continuous space and measurement errors.
- Diffusion process: These models comprise continuous time and continuous space. Diffusion processes can incorporate measurement error but in practice the measurement error is often ignored or neglected.

These classes of models are described in the following Sections 2.2 to 2.4, and Section 2.5 outlines the limitations and advantages of both the continuous and discrete time models. But first a description of animal behaviours will be presented in the next subsection.

2.1.1 Animal behaviours

In examinations of animal movement, certain animal behaviours or habitat features can be incorporated into models because interactions between animals and the environment affect animal movement behaviour. Studying animal behaviour helps in understanding how the organism utilises different habitats and resources. Ecologists study animal movement to understand animal behaviour. The more extensively we examine an animal's behaviour, the more insights we obtain into its inherent characteristics.

In this thesis, we think of behaviour as being expressed by different modes of movement, for example moving at different speeds, moving towards particular features, etc. In movement modelling, this is often represented by an animal switching between discrete behavioural states, with different movement characteristics in each state. Blackwell (1997) explained that there are many factors that affect an animal's movement, such as the type of activity, e.g. foraging, resting, exploring and socialising; various physiological states, such as experiencing hunger or thirst, or the part of the home range being used. For convenience, we refer to these as 'behavioural states'. Moreover, an animal might be foraging in an area and then return to its nest or home range to rest or sleep; thus, the animal will not sleep outside its home range or nest. Animal behaviour changes over time. An animal can have certain behavioural states that depend on it being in a particular habitat, so it moves in a different way in a different habitat. In most of the models considered in this thesis, the behavioural states are set up in a particular way. For example, In the fisher data (described in Chapter 5), each state is associated with a particular habitat. In the ibex data (described in Chapter 6), one state represents locations close to the centre of attraction and the other state represents

locations farther away from the centre of attraction. Also, there may be some states without any biological meaning. For example, in the homogenous model, the behavioural state does not depend on the animal's location (details of the homogenous model are given in Chapter 3). Hence, there is sometimes no immediate link between the states and anything that is external or biological. Therefore, no biological interpretation of the state is necessary.

Animal behavioural states can be observed, partially observed or non-observed. These terms refer to whether or not an ecologist can observe that an animal is in a particular behavioural state. For example, an ecologist can observe an animal's behaviour by following it around, as was done in early experiments before researchers used electronic monitors. However, in a large habitat area, it is difficult and time consuming to follow the animal around to record the different behaviours. With recent technology, some sensors can record animal activity, heart rate and body temperature, which helps to estimate different behavioural state. A significant amount of GPS and radio-tracking data do not record any behaviour directly, but an ecologist can use modern sensors to record extra data to determine how an animal is moving, as described in Section 1.4. Of course, if the behavioural state is internal, such as the animal being hungry or thirsty, the state cannot be observed. In this thesis, animal behaviours are not observed; rather we try to learn about the behaviours from the location data.

2.2 Hidden Markov models (HMMs)

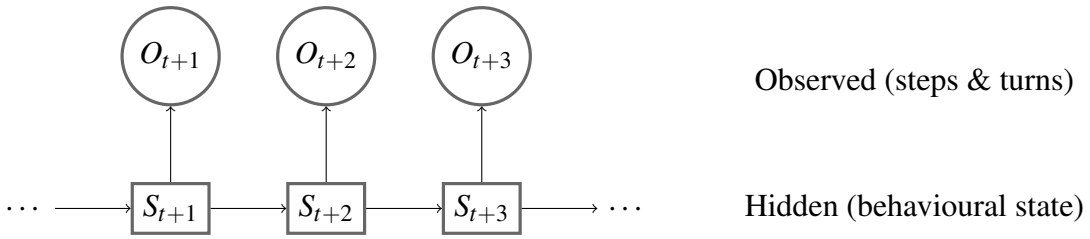

Fig. 2.1 Basic structure of Hidden Markov Model.

HMMs are widely used for analysing animal movement data with underlying behaviour. An HMM is an SSM with a discrete finite number of hidden behavioural states and the assumption of no errors in the location data. Figure 2.1 provides basic structure for the HMM. The HMM structure has two mains processes. First, a state dependent process O_t using time series data is usually employed in the form of step lengths l_t and turning angles a_t, which are assumed to come from distributions that depend on the underlying state with regular time intervals between observations and the assumption that the data have no measurement errors

or only small, insignificant errors.

$$O_t = (l_t, a_t), \tag{2.1}$$

Second, the hidden behaviour process $S_t\{1,\ldots,N\}$ in the HMM includes only one part, namely the hidden behavioural state. In the HMM, the behavioural states are assumed to be hidden (unobserved), and they can be represented using a Markov chain with fixed number of states. The Markov chain behaviour matrix contains the probability of switching from one state to another (Zucchini and MacDonald 2009).

$$\lambda_{ij} = P(S_t = j \mid S_{t-1} = i), \quad i, j = 1, ..., N. \tag{2.2}$$

Also, the observations are assumed to be conditionally independent, given the states (Zucchini and MacDonald 2009).

$$f_i(O_t) = f(O_t \mid S_t = i) = f((l_t, a_t) \mid S_t = i) = f(l_t \mid S_t = i) f(a_t \mid S_t = i). \tag{2.3}$$

The duration time for each behavioural state must follow a geometric distribution. Each behaviour is connected with a different random walk. Often, HMMs uses a homogenous two-state Markov chain and two separate correlated random walks (CRWs). For example, Patterson et.al (2017) explained that the two states could be represented as the first is "encamped" with small step lengths and a lot of turnings and the second is "exploring" with extended step lengths and few smaller turning angles.

The likelihood for the HMM is:

$$\begin{aligned} L &= f(O_1, \ldots, O_t) \\ &= \sum_{S_1=1}^{N} \ldots \sum_{S_T=1}^{N} f(O_1, \ldots, O_t \mid S_1, \ldots, S_t) f(S1, \ldots, S_t). \\ &= \sum_{S_1=1}^{N} \ldots \sum_{S_T=1}^{N} P(S_1) \prod_{t=1}^{T} f(l_t \mid S_t) f(a_t \mid S_t) \prod_{t=2}^{T} P(S_t \mid S_{t-1}). \end{aligned} \tag{2.4}$$

There are many recent extensions and developments for the HMM. For example, Morales et al. (2004) extended the HMM to allow for a mixture of random walks, each representing different animal behaviours, using step length and turning angle data. For the state dependent processes, which generate steps and turns, there are certain assumptions and restrictions for the types of distribution that can be used. However, Langrock et al. (2015) solved this problem by providing a nonparametric method for estimating the state dependent distributions. The HMM has been extended to include spatial heterogeneity by adding environmental

covariates that affect animal behaviour to understand the interaction between animals and their environment, using non-homogenous Markov chains (Langrock et al. 2012). Langrock et al. (2011) developed a model to remove the limitation that the duration time for each behavioural state must follow a geometric distribution by assuming that the time spent can follow any distribution, which gives a hidden semi-Markov model (HSMM). Bulla et al. (2010) produced an R package called HSMM to provide inference for this model. Langrock et al. (2014) extended the hidden Markov model with multiple states to estimate group movement and how the animals interact within the group. Most HMM methods are easily accessible to ecologists because of the R package called moveHMM (Michelot et al. 2015).

The HMM has the following limitations:

- It can fit only with regular time interval data because discrete time Markov chains are uninterpretable without time units. A continuous time HMM can be used to solve this limitation. However, the problem with the continuous time HMM is that it can not easily use the current method of inference (discussion the next subsection).
- It is useful only when the location data include no measurement errors or very small errors. The HMM models are usually applied to GPS data, thereby assuming that there are no measurement errors in the data is realistic in this case. In general, only GPS monitors give data with small measurement errors; some monitors usually contain large measurement errors in the data. The solution to this problem is to use the SSMs, although the SSMs are also only suitable with regular time data.

2.2.1 Continuous time-hidden Markov model

In the discrete-time HMM, it is assumed that the animal can switch or change its behaviour only at the time of the observation, not during the interval. The animal's next location depends only on the states at these times (i.e. the times of the observations) and these states are the only hidden quantities. In a continuous-time HMM, the animal can change its behaviour at any time. Also because it uses continuous time, then the time between observations does not have to be regular (i.e. it can be regular or irregular). Hence, the animal's new location does not depend only on what the animal is doing at this instant but rather on everything that happens in the interval between the two observations. There are additional sets of hidden quantities - the states, locations and times of switches. It is more realistic to construct the model in this way than to use the discrete-time HMM. The problem with the continuous-time HMM is the difficulty in using the same tools and algorithm that are used in the discrete-time HMM. Blackwell (2018) developed a method to deal with this problem, which allows for

a calculation tool that is similar to the discrete-time HMM, which is called the integrated continuous-time HMM. It involves sampling the switching times and locations by MCMC and then using the same algorithm as in discrete-time HMMs to calculate the likelihood.

2.3 State space models (SSMs)

SSMs are hierarchical discrete time and continuous space models. SSMs represent an extension of the HMM that can deal with measurement errors in the data. Both the HMM and SSM contain two stochastic processes. The first is the state dependent process O_t which has a very similar structure in both the HMM and SSM. The second is the hidden state process z_t, which in the HMM represents a fixed number of states, while in the SSM it represents an unlimited number of states with continuous value (Patterson et al. 2017). The hidden state process z_t in the SSM includes the real continuous movement path x_t, and can also contain a second part relating to discrete behavioual states S_t (such that $z_t = \{S_t, x_t\}$). This means the SSM is modelling the true locations as hidden states. In the SSM, there is always a distinction between the true location and the observed location. The Markovian assumption is that the true future location depends on only the true current location. The observed location is connected to the true location through the dependent process. Figure 2.2 shows simple structure for SSM.

$$X_t = f(x_{t-1}, \varepsilon_t), \tag{2.5}$$

$$O_t = g(x_t, \theta_t). \tag{2.6}$$

Where ε and θ are the hidden state process and the observation (i.e. state dependent) process error, respectively.

In both the SSM and HMM, the behavioural state does not exactly represent one behaviour; instead, it represents a group of behaviours that are similar in movement. For example, if there is an SSM with two states, encampment and exploring, the "encampment" state will represent a group of behaviours related to small movements, such as foraging and resting. Meanwhile, the "exploring" state will represent behaviours related to large movements, such as running and hunting.

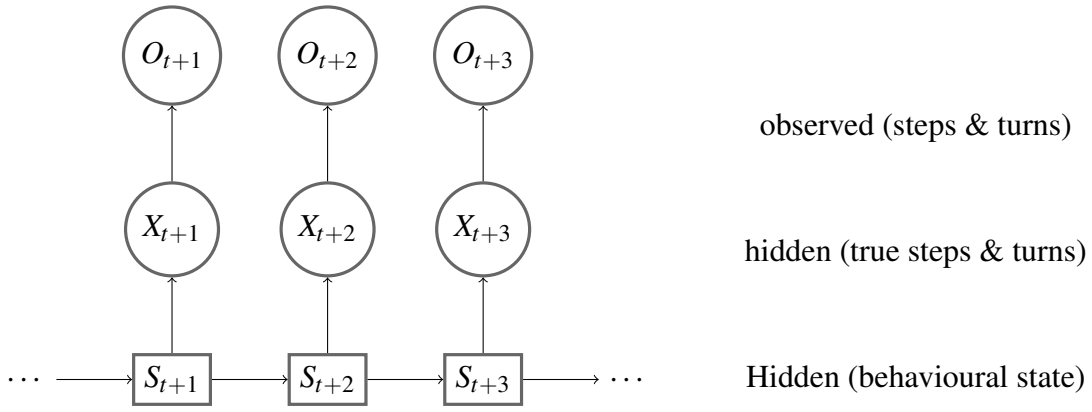

Fig. 2.2 Basic structure of State Space Model.

The likelihood for the SSM is:

$$\begin{aligned} L &= f(O_1,\ldots,O_T) \\ &= \sum_{S_1=1}^{N}\int_{z_1}\ldots\sum_{S_T=1}^{N}\int_{z_T} f(O_1,\ldots,O_T \mid x_1,\ldots,x_T) \\ &\quad f(x_1,\ldots,x_T \mid S_1,\ldots,S_T) f(S1,\ldots,S_T) dz_T \ldots dz_1. \end{aligned} \tag{2.7}$$

The SSM is more challenging to apply than the HMM because the HMM likelihood can be evaluated easily, while the SSM requires complicated inference to approximate the likelihood. It is important to note that the SSM includes location errors, but the limitations for the SSM include that it is more complicated to use than the HMM. For more details about the SSM, see Patterson et al. (2007).

Beyer et al. (2013) used a Bayesian SSM (CRW) to estimate behavioural states from movement data. These researchers improved the effectiveness of Morales et al.'s (2004) idea (introduced on page 16) by using stochastic simulations to estimate behavioural states from movement trajectories. Avgar et al. (2013) added heterogeneous habitats to the SSM and replaced the discrete hidden behavioural state with a continuous hidden behavioural state. They also incorporated environmental data with movement data to understand how the animals interact with the environment.

The limitation of both the HMM and SSM is that they are applied only to regular time interval data. In contrast, the diffusion model can be applied to regular and irregular time data.

2.4 Diffusion processes

"A diffusion process is a continuous-time Markov process with continuous sample paths" (Cox and Miller 1965). Diffusion modelling uses either standard models (for example, Brownian motion) or formulation in terms of stochastic differential equations (SDEs). Here, we consider some standard models first (Sections 2.4.1 to 2.4.3), then briefly discuss SDEs (Section 2.4.4).

2.4.1 The Wiener process: Brownian motion

The basic diffusion model is Brownian motion, which in one dimension has:

$$X(s+t)|X(s) \sim N(x(s), tv), \tag{2.8}$$

where v is the diffusion rate, or variance per unit time. It is a Gaussian stochastic process in continuous time, and can be obtained as the limit of a random walk in discrete time (Iacus, 2008). In d dimensions, we have:

$$X(s+t)|X(s) \sim N(x(s), t\Sigma), \tag{2.9}$$

where Σ is the variance-covariance matrix. Usually, we take $\Sigma = \sigma^2 I_d$, where I_d is the identity matrix. As a movement model, Brownian motion is very simple, with only a single parameter that controls how fast the animal moves.

A Brownian bridge is a Brownian motion over a time interval, conditioned on its position at both endpoints. It arises naturally concerning the location of an animal between two observations. The probability of the animal's presence at a particular location depends on the locations at the beginning and end of the interval, the time between the observations and the movement parameter (Horne et al. 2007) . In an analysis concentrating on this approach, the Brownian bridge is sometimes discussed in the literature as if it were a different model in itself, the Brownian bridge movement model (BBMM). Horne et al. (2007) introduced the BBMM terminology and extended the model to estimate the movement trajectory for an animal using movement data with measurement errors.

A key disadvantage of movement modelling based on Brownian motion or the BBMM is that assumes that the animal moves randomly, unaffected by any points of attraction or other features of its environment. The process is not stationary, so the animal cannot have a consistent home range.

2.4.2 Other models of location

Some of the limitations of the Brownian motion model can be addressed by allowing directed movement, in which the animal is always pulled toward (or pushed away from) a specific location. The Brownian motion is a limiting case of the Ornstein–Uhlenbeck OU process (Uhlenbeck and Ornstein, 1930). Dunn and Gipson (1977) introduced the use of the Ornstein–Uhlenbeck OU process as a movement model. The OU process has a centre point μ to which the animal is attracted. For example, μ can signify a food source, a nest or the centre of the home range, as described in Chapter 1. Throughout this thesis, the OU process is used for the analysis of animal location. The detailed description for this model, and the methods used to fit it, are given in Chapter 3.

The Dunn and Gipson model was extended by Blackwell (1997, 2003) to allow switching between different behavioural states, giving a switching diffusion (or mixed diffusion) model. This is analogous to having different behaviours in an HMM. A formal definition follows in Chapter 3. Harris and Blackwell (2013) extended the switching OU model further by adding spatial heterogeneity. Kranstauber et al. (2012) added behavioural switching in a similar way in the BBMM to give what they called a dynamic BBMM.

2.4.3 Modelling velocity

All the diffusion models thus far are formulated directly in terms of position. However, in this section we look at modelling the velocity and then integrating with respect to time to get the position. Johnson et al. (2008) used the OU process; however, instead of modelling locations directly, they modelled velocity. They used a continuous time OU velocity process to predict the location using maximum likelihood estimation while incorporating measurement errors. The limitation of this model is that it does not include different behavioural states. Michelot and Blackwell (2018) extended Johnson et al.'s (2008) continuous time model for velocity to include multiple behavioural states.

Parton et al. (2016) fit a stochastic continuous time model for both speed and bearing. Here, the bearing is represented by Brownian motion, and the speed is a one-dimensional OU process. This is similar to Johnson et al.'s (2008) approach, but Johnson et al. modelled velocity as a two-dimensional OU process. Note that the Johnson model is solved exactly, like the OU model for position; the Parton et al. model solved using a numerical approximation. Parton et al. (2016) assumed that location error has a multivariate normal distribution and is independent; however, the independent assumption is unrealistic because

location observations are usually auto-correlated. Breed and Severns (2015) explained that if the sampling frequency is high in GPS movement data, the observation error might be very correlated. Another limitation of this study is that it assumes animals have a single behavioural state. However, in their recent paper, Parton and Blackwell (2017) extended this model to include multiple behaviours but assumed there were no errors in the location data. The limitation of this model is the high computational cost.

2.4.4 Stochastic differential equations (SDEs)

The OU process and Brownian motion are the solutions to linear SDEs, which means that the equation can be solved exactly. Other approaches used general non-linear SDEs that cannot be solved exactly but can only be approximated numerically, which inhibits the description of complicated animal movements (Brillinger et al. 2002; Brillinger and Stewart, 1998; Preisler et al. 2004). Parton et al.'s (2016) continuous time-step and turn-angles model (as mentioned above in Section 2.4.3) uses an SDE for the speed and the bearing over time, which cannot be solved exactly, but only numerically. In general, the published non-linear SDE models do not include behavioural switching, though in principle switching could be added. Some studies have used nonparametric estimation for SDEs to analyse animal movement data and study how movement is affected by the environment. For more detail, see Brillinger et al. (2002); Brillinger and Stewart (1998) and Preisler et al. (2004).

2.5 Discrete or continuous time

McClintock et al. (2014) discussed the difference between continuous and discrete time models. Animal movement happens in continuous time, but animal locations are collected at discrete, fixed intervals. In discrete time, the time interval at which the data are recorded, for example every 10 minutes or every 1 hour, does not mean anything in regard to the animal. Thus, it is natural and more realistic to think about animal movement in continuous time. However, the discrete time model is more widely used than the continuous time model because it is usually less complicated to implement than the continuous time model. Moreover, a continuous time model is more complicated. Not all ecologists have the technical background to use a continuous time model on their own. Another problem currently with the continuous time models is the intense computational cost when used to analyse large datasets. Whilst, the discrete time modelling benefits from readily available implementations of fast algorithms for model fitting and the straightforward interpretation of the resulting parameters.

Discrete time models have three limitations. Firstly, the time intervals between observations should be regular; the models cannot be fitted in a meaningful way when there are different time intervals between observations. This can be a problem with real data in various ways. Real data can be irregular because observations are delayed (e.g. when a GPS fix fails) or missing. They can also have a time step that is not constant by design because more information is recorded at a certain time of the day. For example, for the wild boar data of Quy et al. (2014), re-analysed by Blackwell et al. (2016), some of the time measurements were collected every 15 minutes, while others were collected every 30 minutes. It may be possible to fit a discrete time model by treating some observations as missing, but this is potentially highly inefficient. A continuous-time model, in contrast, can fit irregular data without being slower than regular data.

Secondly, the interpretation of the parameters in the discrete time model depends on the time interval, so it is impossible to compare different data collected at different time scales, even if they are for the same species. Thus it is often impossible to compare different studies. For a continuous time model, the parameters are interpreted independently of the time scale of observations, so comparing or combining studies is possible.

A third difficulty in using the discrete time model is that combining the movement data with other types of data using other sensors is impossible unless the time scales in the different test data match. In contrast, in the continuous time model, this combination can be done easily. In the continuous time model, the animal movement data can be combined with other data that provide information about behaviour, such as those derived from the accelerometer, which has a completely different time scale. Sometimes, the observations are of the location and at other times they are of the behaviour. A similar issue arises in thinking about the behavioural state that underlies the observed movement. In a discrete time model, behaviour can change only at the time of an observation, whereas in reality behaviours occur at random or irregular time intervals. This can make the estimation of behaviour parameters using a discrete time model difficult.

There are two circumstances in which the discrete time method can be used to approximate the continuous time method: 1) the observations are equally spaced (which means the time between observations is the same) so that discrete time can fit these models and 2) the switching rate is very slow in comparison to the number of observations such that there are many observations for each time visit behaviour and states. In this case, a discrete time approximation can be used in these models; however, if the switch is very quick, as it is in

some of the examples in the present work, the discrete time approach will not work well. For example, if the switching in behaviour is very slow in comparison to the observations, each visit to a behavioural state will contain a substantial number of observations. In such a scenario, adding a discrete time approximation to the continuous time model would probably be reasonable.

Chapter 3

Ornstein-Uhlenbeck (OU) Diffusion Process

3.1 Introduction

We use the OU process to model animal location data throughout this thesis. The OU process is a continuous time process used to analyse animal movement data. We prefer a continuous time model because it is more realistic than a discrete time model (for more details see Section 2.5). In this chapter, we describe an existing class of models based on OU processes and give additional explanation and discussion in addition to that in Blackwell et al. (2016). This chapter is not only a literature review; it also provides more explanation and detail about how the model works.

Chapter 2 gave an overview of models that allow for dependence; here we focus only on a specific set of continuous time models. To deal with the problem of dependence between observations, Dunn and Gipson (1977) generalised and developed the bivariate movement method of Jennrich and Turner (1969), which we described in Section 1.5. They assumed the animal movement "data are generated from a continuous, stationary, Gaussian stochastic process" that has Markovian properties, which is called the multivariate OU diffusion process. The continuous Gaussian assumption means that the animal's utilisation distribution still follows a bivariate normal distribution, but attention is paid to the correlation between observations over the animal's path, which is assumed to be a continuous path. The stationary home range assumption means that the home range is fixed (stable) and that there is an assumption the animal is not in a migrant state. When an animal migrates during a study, the model fails to estimate the position because the process becomes non-stationary. The

Markovian assumption means that the animal's future or next location depends only on the current location, not on the past.

The observation (animal location) for a single animal at time t is:

$$X(t) = [x_1(t), x_2(t)]'. \tag{3.1}$$

The animal location is a sample from a bivariate normal distribution called the equilibrium (unconditional distribution):

$$X(t) \sim N(\mu, \Lambda), \tag{3.2}$$

where μ is the mean vector, and Λ is the variance covariance matrix (Dunn and Gipson 1977). This corresponds to the normal utilisation distribution discussed by Jennrich and Turner (1969). The conditional distribution of an animal's position at time $(s+t)$, given its position at time (s), has a bivariate normal distribution (Dunn and Gipson, 1977; Blackwell, 1997):

$$X(s+t) \mid X(s) = x(s) \sim N(\mu + e^{Bt}(x(s) - \mu),\ \ \Lambda - e^{Bt}\Lambda e^{B't}), \tag{3.3}$$

where μ is 2-vector, and Λ, B are 2×2 matrices, and B is a stable matrix; $e^{Bt} \to 0$ as $t \to \infty$, which controls the central tendency, and B' is the transpose of B (Dunn and Gipson 1977; Blackwell 1997). The movement process has three parameters: (μ, Λ, B), where μ is the centre of attraction (for more detail see Section 1.3.1), Λ is the covariance matrix of the animal location, i.e. the variation in location, and B is a centralisation tendency matrix or rate of attraction toward μ. In the OU model, when the animal is farther away from the centre of attraction, it moves faster toward the centre of attraction. In some cases B and Λ have an isotropic assumption, which assumes $B = bI$ for $b < 0$, and $\Lambda = vI$, where I is the identity matrix. Adding this constraint minimises the number of parameters and simplifies the inference.

The Dunn and Gipson model can be applied to $m > 1$ animals simultaneously, defining:

$$X(t) = [X_1(t), X_2(t), ..., X_m(t)], \tag{3.4}$$

and

$$X_i(t) = [x_{i1}(t), x_{i2}(t)]', \qquad i = 1, 2, ..., m. \tag{3.5}$$

where X_i is the location for the i-th animal at time (t). In this case μ is a vector of length $2m$, and Λ and B are $2m \times 2m$ matrices. Animal movement can be modelled using an OU process for a number of animals collected together as a multivariate OU process (in this case, the process will also describe the interaction between these animals). This enables the simultaneous analysis of groups of species and allows the study of interactions between two or more animals.

Dunn and Gipson (1977) and Worton (1995) find that when using the diffusion model it is necessary to decide how to specify the initial location. The likelihood for the diffusion model is:

$$p\{x(0),x(1),x(2),.....|\theta\} = p\{x(0)|\theta\}\prod_{i=1} p\{x(i)|\theta,x(i-1)\}. \tag{3.6}$$

where $\theta = (\mu,\Lambda,B)$ represents all movement parameters and the term $p\{x(0)|\theta\}$ refers to the initial location. Dunn and Gipson (1977) explained that when the dataset is large, the initial term $p\{x(0)|\theta\}$ can be ignored or deleted, which is equivalent to treating the initial location as fixed and known. However, Dunn and Gipson (1977) argued that the initial information is highly significant and that the initial term in the likelihood contains more information than the conditional term for a later observation. To solve this problem, Dunn and Gipson (1977) assumed that the initial term can be generated from the equilibrium distribution. The assumption is that the initial term is a sample from a multivariate normal distribution called the equilibrium (unconditional distribution).

There are two main disadvantages of Dunn and Gipson's (1977) OU process. First, the OU process assumes that the home range's shape is always elliptical, but some home ranges are not elliptical. Second, the OU process always assumes a unimodal utilisation distribution, which is unrealistic because some data are multimodal. Another limitation for Dunn and Gipson's method is that it does not include behavioural information (Blackwell 1997). Moreover, the model does not account for any habitat features, such as rivers or highways. The OU process can incorporate measurement error but in practice the measurement error is often ignored or neglected.

3.2 Mixed diffusion process

The aim of the Dunn and Gipson (1977) model was to study the animal's home range. However, more realistic modelling of animal movement includes information about different behaviours that animals use in daily activity. Blackwell (1997) developed Dunn and Gipson's

model to incorporate animal behaviour. Blackwell (1997) said that rather than an animal's movement being represented by a single diffusion process, an animal switches between several diffusion processes, with each one representing a different behaviour (more detail about animal behaviour is given in Section 2.1.1). This process is called a random or mixed diffusion process.

Blackwell's (1997) model has two processes:

- Behaviour or switching process: The discrete finite set of behavioural states follows a continuous time Markov process, and is independent of position. The behaviour process is defined by a generator matrix $G = (\lambda_{ij})$, indicating the transition rate between states:

$$P(J(t+\delta t) = j \mid J(t) = i, x(t) = x) \approx \lambda_{ij}\delta t, \quad i \neq j. \tag{3.7}$$

  If n states exist, the matrix dimension will be $n \times n$, where

$$\begin{aligned} &\lambda_{ij} \geq 0 \quad \text{when} \quad i \neq j, \\ \text{and} \quad &\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}. \end{aligned} \tag{3.8}$$

  The sum of the generator row should equal zero. The transition rate λ_{ij} can take any positive value. If the λ_{ij} is high, there are many switches, which means that the animal changes its behaviour quickly. If the λ_{ij} is low, it takes a long time to switch. The animal remains in one state for some period of time and then switches to different state and remains there for some period of time, etc. The probability of switching to a new state depends only on the current state. Blackwell (1997) stated that when number of states equals n, the behaviour process can represent the average time spent in every state and the probabilities of switching among several behavioural states. The duration of time spent in a state follows an exponential distribution, with the parameters (transition rates) do not depend on (location or time) only the state.

- Movement process: The animal location follows an OU diffusion process, as in the Dunn and Gipson model, but the movement parameters depend on the states, which means having more than one OU process, each one connected to a different state. The movement process is stationary. This point is different than Dunn and Gipson's model (1977), in which there is only one diffusion process, and it should be stationary. In Blackwell's (1997) method, the mixture of diffusion processes should be stationary, and this can be satisfied when all the processes are stationary or when only some of them are stationary OU processes, and the others involve Brownian motion (a diffusion

process without any drift), which is non-stationary. For more detail about Brownian motion see Section 2.4.1. For each state, there are different movement parameters for Equation (3.3):

$$\mu_i, \ \Lambda_i, \ B_i, \quad \text{for} \quad i = 1, ..., n. \tag{3.9}$$

Blackwell (1997) fitted the model for a single animal data. Blackwell (2003) described a full Bayesian approach to analysing a mixed diffusion process via the hybrid MCMC method. Blackwell (2003) assumed that the states are observed at the same times as the locations. However, he estimated transition rates, and also estimated behaviour between the observations. Dunn and Gipson's (1977) model is a special case of Blackwell's model, specifically when the number of behavioural states (n) is equal to one.

3.3 Spatially heterogeneous diffusion method

Animals move differently in different habitats, and the previous approaches do not take account of this fact. Harris and Blackwell (2013) extended Blackwell's (1997, 2003) models by adding spatial heterogeneity. Harris and Blackwell (2013) introduced new continuous time models that allow animal movement to depend on animal location over several regions and underlying behaviour states. The new spatially heterogeneous model divides the environment into a fixed group of regions and uses different continuous time diffusion processes $X(t)$ in each region. They extend Blackwell's (1997, 2003) model by allowing the movement process to depend not only on the animal's behavioural state but also on the animal's location. Therefore, three elements are used to describe the new model: the switching process for animal behaviour, the diffusion movement process for the animal's position and the boundaries of each region, which divide the habitat. If there are n states and the habitat is divided into l regions, the behaviour process depends on the animal's location, with a transition rate dependent on region; thus, it will need l generator matrices G_r, $r = 1, ..., l$ each of dimension $n \times n$. Blackwell et al. (2016) used the same idea as Harris and Blackwell (2013) by let $X(t)$ be the animal location at time t and $J(t)$ be the behavioural state at time t. $\lambda_{ij}(t,x)$ is the transition rate from behaviour state i to behaviour state j at time t and location x, so the probability of an animal switching from behaviour state i to behaviour state j in a short, finite interval is as follows:

$$P(J(t+\delta t) = j \mid J(t) = i, x(t) = x) \approx \lambda_{ij}(t,x)\delta t, \quad i \neq j. \tag{3.10}$$

The movement process will depend on both location and behaviour, which will give $n \times l$ OU diffusion movement process for the combined state and region:

$$\mu_{ij}, \quad \Lambda_{ij}, \quad B_{ij}, \quad i = 1, ..., n, \quad j = 1, ..., l. \tag{3.11}$$

For example, two regions and three states will have two generator matrices; thus, each model comprises two 3×3 matrices and six OU diffusion processes. Dunn and Gipson's (1977) model is a special case of Harris and Blackwell's model, specifically when $l = 1$ one habitat region, and $n = 1$ one behavioural state.

3.3.1 The separable model

To simplify the model and its inference, Harris and Blackwell (2013) suggested that the movement process should depend only on the state, and that the transition rate (behaviour process) should depend only on the region, meaning that the movement indirectly depends on the region. This newly defined model is called the separable model. When there are n behavioural states and the environment is divided into l regions, the separable model will have n OU movement processes and l generator matrices for the behaviour process. The separable model has been used throughout this thesis.

3.4 Behavioural states

Unknown behavioural states are used throughout this thesis. If behavioural states were known, e.g. from accelerometer data, it would be straightforward to allow for that by conditioning the state updates on the observations. Typically, for the data collected using GPS, the behavioural states will be unobserved. Usually, for GPS or radio-tracking data, it is not possible to know the animal's behavioural states. Therefore, how can the number of states be determined in the proposed model? Usually, a number of behaviours are known. The number of states is affected by type of data. There could be an explicit number when the state is related to the feature in the environment. In the adaptive model (detailed section 3.6.2), the behavioural state depends on the animal's habitat; for example, in the fisher data (detailed in Chapter 5), there were only three different habitats, so there will only be three behavioural states.

For some other cases, it might not be possible to know how many behavioural states there are. One commonly used way to determine the number is by fitting two or more separate models with different numbers of states and comparing how well they fit using deviance information criterion (DIC). However, the problem with the DIC method is the difficulty

of calculating the likelihood, which can be done approximately within the MCMC. This is straightforward using the animal movement model in discrete time but not in continuous time. The calculation of DIC for a continuous time model is described briefly by Blackwell et al. (2016). For example in the fisher experiment in Blackwell et al. (2016), the researchers fit a model with one state which is not dependent on habitat and another model with three states, and compared the two models using DIC; then they concluded that the one state model was not fitting very well. Another option is to use Reversible Jump (RJ) MCMC to obtain posterior probabilities for each model instead of fitting different models separately. The DIC compares models by looking at how well they explain the data and how complex they are. RJ treats the choice of model as an uncertain quantity that we can learn about directly from the data; if we assume that one of the models is correct, then we can calculate posterior probabilities for which one it is. However, updating the trajectory is difficult using RJ MCMC.In this thesis, we only use the DIC approach.

3.5 Introduction to MCMC

3.5.1 MCMC

Throughout this thesis, we used an MCMC algorithm to estimate the movement and behaviour parameters via mixture of Metropolis-Hastings or Gibbs sampling. The Gibbs sampling method can be used when the full conditional distribution for each parameters is known. The Gibbs sampling method is faster and easier than the Metropolis-Hastings algorithm. However, the Metropolis-Hastings algorithm covers more cases and is more general than Gibbs sampling. The Metropolis-Hastings algorithm is used when the posterior density does not follow any standard distribution as well as when the full conditional distribution is unknown. The movement parameters can be updated using Gibbs sampling or a standard random walk Metropolis-Hastings. Throughout this thesis, we used the Metropolis-Hastings algorithm to estimate the movement parameters and Gibbs sampling or a standard random walk Metropolis-Hastings to estimate the behaviour parameters. The details are given in each chapter and also in next Section 3.6 and Section 3.7.

3.5.2 Convergence

The MCMC sample has converged when the posterior result does not depend on the initial value. We diagnosed convergence in different ways: First, we ran multiple chains, each with different starting values and checked whether we got the same posterior result. Second, we used the R package called “coda”, which has many different tests for convergence. Also,

we checked the effective sample size, and autocorrelation plots and the trace plots for the posterior parameters. Third, we ran the model for a long interval to make sure it converged. We used a long run with 10 million iterations. The same number of iterations was used in all experiments in this thesis. We used a long run because the model is complicated, and the algorithm takes a long time to converge. Blackwell et al. (2016) likewise used the same number of iterations. Complicated models like this usually do not mix well. For the same reason, we used burn-in in all the experiments in this thesis. Burn-in is a method used to remove the initial part of the sample posterior (MCMC) result to remove the effect of the initial value on the posterior inference. Also, to check the stability of our results, we did not change anything; we just ran the same model with new seeds. In R we used set.seed() with different parameters to generate new random numbers. Then we run the code again to check that we get the similar result, even with different sequence of random numbers.

3.5.3 Thinning

Thinning is a technique used to handle the autocorrelation in the data by saving only every kth sample from the posterior. In this thesis, we thin for several reasons mainly to save space, storage and time in doing the actual calculation, which is the estimation of the posterior distribution. We thin to reduce the size of the file and make the results easy to work with. The only other reason for thinning in such a way is for autocorrelation. However, thinning is not needed for making theoretical points about allowing for autocorrelation. Developing a histogram or calculating the posterior density based on un-thinned data should yield the same results that are obtained using thinned data. Link and Eaton (2012) suggested that un-thinned MCMC result is often more precise than thinned result. Based on the un-thinned data, the posterior density is slightly more accurate because thinning always causes the loss of some information. However, because of the very strong autocorrelation that is typically found in MCMC samples, the posterior does not lose very much information. Hence thinning leads to minimal information loss. Therefore, fitting a model without thinning will yield a slightly higher effective sample size than with thinning. However, the difference is small, so it is only a practical consideration.

3.6 Bayesian Inference

Blackwell (2003) introduced exact Bayesian inference for mixed diffusion processes, assuming the states are observed. Blackwell et al. (2016) developed the idea of Blackwell (2003) by using Harris and Blackwell's (2013) separable model. Blackwell et al. (2016) aim to fully

implement the Bayesian inference for the separable model without any approximation due to discrete time errors using the MCMC method. Each iteration updated three main sets:

- The trajectory; switching times, locations and states.
- The movement (OU) diffusion parameters; $\mu_i,\ v_i,\ b_i$.
- The behaviour parameters; $\lambda_{ij}(x,t)$.

Every set is updated individually by the Metropolis-Hasting algorithm or Gibbs sampling. The first step in this MCMC approach is to simulate the potential path of switches, locations and states along some interval of time conditioned on the path outside the interval and on movement and behaviour parameters. Fundamentally, this involves the simulation of the augmented trajectories. This simulation of the states and the locations at the times of potential switches can be used as a form of data augmentation in an MCMC algorithm (Blackwell et al. 2016).

Blackwell (2003) explained that the exact inference for the mixed diffusion model is difficult because the switching time and some duration times of the behaviour process are unobserved, and the conditional distribution depends on all the parameters. Blackwell (2003) used data augmentation to solve this problem and treated the trajectory as missing data. Simulating a trajectory is straightforward when the switching is completely random and the behaviour is independent of location and time such that $\lambda_{ij}(t,x) = \lambda_{ij}$ (Blackwell 2003). However, in Blackwell et al.'s (2016) model when the transition rate $\lambda_{ij}(t,x)$ depends on the location $x(t)$, the time of the switch is unknown unless the animal location $x(t)$ is known. The easiest way to resolve this is to use approximation via discrete time, but Blackwell et al. (2016) wanted to avoid discretisation errors through the exact simulation of the trajectory. The important assumption Blackwell et al. (2016) made was that the transition rates were bounded. This means transitions in behaviour never happen immediately when an animal changes its habitat. They first calculate the global upper bound:

$$\kappa = \max_{j,t,x} \lambda_j(t,x), \tag{3.12}$$

where κ is constant and fixed. They define:

$$\lambda_j(t,x) = \sum_{i \neq j} \lambda_{ji}(t,x), \tag{3.13}$$

which describes the transition rate out from behaviour j at time t and position x. This allowed them to define a random variable that is a lower bound for the waiting time until the next

switch in behaviour:

$$T \sim \exp(\kappa). \tag{3.14}$$

This represents the time at which the first switch in behaviour can happen. When we consider all the times at which changes in behaviour may occur, depending on the animal's movement, this can be termed a Poisson process of rate κ. Then in an interval (t_a, t_b), $M \sim \text{Poisson}(\kappa(t_b - t_a))$ is the number of potential switches within the interval $t_a < T_k < t_b$, where $k = 1, ..., M$. Knowing the potential switching times, a trajectory for both behaviour and location can be simulated as in Blackwell (2003) and Blackwell et al. (2016) as follows: sample times, locations and states for all potential changes during an interval (t_a, t_b) are dependent on the path out the interval based on states at the beginning $J(t_a)$ and at the end $J(t_b)$. We propose times $T_1, ..., T_M$ as above, which represent the proposed times of potential switches. Starting with the location of the observation $x(t_a)$ and the behavioural state $J(t_a) = j$, we simulate the location of the first potential switch $x(T_1)$ using the movement model related to state $J(t_a)$, then check whether an actual switch occurred. If it has, the new state $J(T_1)$ is sampled. The potential switch at T_1 is an actual switch with the following probability:

$$\frac{\lambda_j(T_1, x(T_1))}{\kappa}. \tag{3.15}$$

If so, the new state i is sampled with probability

$$\frac{\lambda_{ji}(T_1, x(T_1))}{\lambda_j(T_1, x(T_1))}. \tag{3.16}$$

Otherwise, the state continues to be j. We then iterate this process forward, through each potential switch in turn. To propose a new trajectory, Blackwell et al. (2016) used an independence sampler with an idea similar to the one just described to create an exact simulation. Then, they used a Metropolis-Hasting update to accept or reject the simulated path. It was required that the last simulation state $J(T_M)$ match the existing state $J(t_b)$ at the end of the interval. In practice, we need to be able to update in this way over an interval that contains other observations of location. In that case, we can assume that the state at the time of one of these observations equals the state at the previous potential switch, and the movement process is restarted at the known location.

To simulate a single transition rate out from state j, we need the global bounded κ because the transition rate away from other states affects the other choices and because we

are considering the entire interval between observations. For example, if we have a high transition rate for one of the states, we know the animal will not remain in this state and will switch very quickly to a new state, which affects the choice of the state for the new potential switch. Additionally, we know the current state, but the next state is unknown. When simulating a process for which the current state is known, we need to determine when the first change occurs by using the rate λ_j to generate a potential switch and determine the location of that switch in space using the actual transition probability. If a switch to a new state occurs, then by using a global upper bound $\kappa = \max_{i,x}(\lambda_j(x))$, we can find the time of all potential switches and simulate forward until the next switch.

After updating the trajectory, at each observation, we know which movement processes to follow. We then infer the movement parameter in two ways: using a conjugate prior with Gibbs sampling or using a standard random walk Metropolis-Hastings update for the parameters, as according to Blackwell (2003). Blackwell (2003) explained that the movement parameters for each OU process are updated together using the Metropolis-Hastings algorithm. Each movement parameter has an independent proposal, for the centre of attraction μ_i the proposal is bivariate normal distribution, while for b_i and v_i, because of the isotropic assumption, the proposal is univariate truncated normal distribution centred on the current value.

Similarly, the switching parameters can be updated in a separate MCMC step. The details depend on the structure of the model.

3.6.1 Homogeneous model

The homogenous model means that the habitat has one region and the transition rates from one state to another $\lambda_{ij}(x,t) = \lambda_{ij}$ do not depend on the animal's location. This is the same model described by Blackwell (1997, 2003). The model will have one generator:

$$G = \begin{pmatrix} -\sum_{i=2}^{d} \lambda_{1i} & \lambda_{12} & \lambda_{13} & \dots & \lambda_{1d} \\ \lambda_{21} & -\sum_{i\neq 2}^{d} \lambda_{2i} & \lambda_{23} & \dots & \lambda_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{d1} & \lambda_{d2} & \lambda_{d3} & \dots & -\sum_{i=1}^{d-1} \lambda_{di} \end{pmatrix}. \tag{3.17}$$

At a potential switch, the probability for the switch from state i to state j is:

$$p_{ij} = \frac{\lambda_{ij}}{\kappa}, \tag{3.18}$$

and the probability for no switch in the behaviour is:

$$p_{ii} = 1 - \frac{\sum_{i \neq j} \lambda_{ij}}{\kappa}, \tag{3.19}$$

where

$$\kappa = \max_i \sum_{i \neq j} \lambda_{ij}. \tag{3.20}$$

The likelihood for the transition rate is a multinomial product of the elements in each row of the probability matrix. There are likelihoods for each state; for example, the likelihood for state d is :

$$p(\text{data}|\lambda_{ij}) = \left(1 - \frac{\lambda_{dd}}{\kappa}\right)^{n_{dd}} \times \prod_{j=1}^{d-1} \left(\frac{\lambda_{dj}}{\kappa}\right)^{n_{dj}}, \tag{3.21}$$

where n_{ij} is the number of switches from state i to j, and n_{ii} is the number of switches that stay at state i. After simulating the trajectory, the inference for the behaviour parameters is straightforward. Blackwell (2003) used a Dirichlet distribution as the joint prior to update the transition rate in the Gibbs sampling algorithm.

3.6.2 Adaptive model

The adaptive model is an example of a separable model, which is similar to Harris and Blackwell (2013) and Blackwell et al. (2016), where the habitat is divided into multiple regions, and the transition rate between behaviours depends on the region $\lambda_{ij}(x,t)$. The adaptive model means the animal can only switch to the behaviour that matches its region. In any given region, there is only one possible behaviour to switch to. This means the number of behavioural states is equal to the number of regions. The way that the adaptive model works is that if the animal enters into region one, the animal will change to behaviour one quite quickly, but not immediately. When the animal crosses the boundary into a different habitat, it switches to the behaviour that matches the new region but not immediately. However, we want the animal to switch quickly so that we do not spend a long time in the "wrong" state. The model has multiple generator matrices, each one associated to a different region G_l where l represents the region:

$$G_1 = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ \lambda_{21} & -\lambda_{21} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{d1} & 0 & 0 & \dots & -\lambda_{d1} \end{pmatrix}. \tag{3.22}$$

For example, when the animal is in region one and state one, then nothing happens; the animal should stay in state one. When the animal is in region one and state two, the animal should switch to state one with rate $\lambda_{21}(l=1)$. However, there is always a chance that the animal will stay in state two for a while and take some time before switching to state one or just leave the region. When the animal is in region one and state d, then it will either switch to state one or stay in state d a bit longer.

$$G_2 = \begin{pmatrix} -\lambda_{11} & \lambda_{12} & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \lambda_{d2} & 0 & \dots & -\lambda_{d2} \end{pmatrix}. \tag{3.23}$$

$$G_d = \begin{pmatrix} -\lambda_{1d} & 0 & 0 & \dots & \lambda_{1d} \\ 0 & -\lambda_{2d} & 0 & \dots & \lambda_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{3.24}$$

This means the diagonal element in the generator matrix always equals the transition rate:

$$\lambda_i(x) = \lambda_{ii}(x) = -\lambda_{ij}(x), \tag{3.25}$$

where

$$\kappa = \max_{i,x} \sum_{i \neq j} \lambda_{ij}(x). \tag{3.26}$$

$$\kappa = \max_{i,x} \lambda_i(x). \tag{3.27}$$

The probability can be calculated in the same way as in the homogenous model; however, the probability in the adaptive model depends on the region as well. The likelihood for the transition rate is a binomial product of the elements in each row of the probability matrix. Each row has only two possibilities: to stay in the same state for a little while or switch to the behavioural state that matches the region. There are likelihoods for each region l:

$$p(\text{data}|\lambda_{ij}(x)) = \prod_{x=1}^{l} \prod_{i=1}^{d} \left(\frac{\lambda_{ix}(x)}{\kappa}\right)^{n_{ix,x}} \times \left(1 - \frac{\lambda_{ii}(x)}{\kappa}\right)^{n_{ii,x}}, \tag{3.28}$$

where $n_{ix,x}$ is the number of switches from state i to x that matches the region, and $n_{ii,x}$ is the number of switches staying at state i. After simulating the trajectory, the inference for the

behaviour parameters is straightforward. Blackwell et al. (2016) used the beta distribution as a marginal prior to update the transition rate in the form of (λ/κ) in the Gibbs sampling algorithm.

3.6.3 Initial value

Initial values for the movement parameters μ_i and v_i can be learned from the data, while other parameters can be learned from running the model more than once. The details are given in each chapter and the code.

We sampled the initial states using a hidden Markov model. We used the R package called "moveHMM" to help define initial values for the states. The starting configuration of the states was taken into consideration. In theory, any possible sequence of initial states is acceptable. In practice, the initial set of states can significantly influence the estimated posterior results in any finite run. One of the limitations of the Blackwell et al. (2016) approach (exact method) is that it has difficulty constructing a reasonable trajectory - or at least takes a very large number of iterations – to reach a state sequence that is a good fit to the data if we start from all states being the same. Alternatively, if there is a set of possible state sequences that are the best fitting ones, it takes a long time to reach any of them from that starting point. we are updating the transition rates separately from the states, then the values of the transition rates in the early iterations will be misleadingly low. The algorithm will have trouble picking the correct value for the transition rate λ because it does not properly change the initial sequence of states. Thus, the initial set of states should not start with all observations assigned to one state, as this will slow the mixing. Also, it would be difficult to determine a set of parameters that easily enable transition between states.

Starting the initial sequence of states with a mix of states, rather than having all the observations in one state, will improve the mixing and help the model converge more quickly than before. To do This, we used a hidden Markov model to start the configuration for the behaviour states, which gave us a better estimate for the transition rate and more consistent results. We chose the hidden Markov model because it is a simple and fast method. The HMM move package (moveHMM) can be used to estimate the initial states of the model by fitting a discrete-time HMM and using the fitted states as the initial states of the continuous-time model. Although this method does not specifically fit the correct model, it does allow for a sensible sequence of initial states. Of course, we did not have any actual information about the states, but using a hidden Markov model as part of the set up could provide a better initial sequence of behavioural states to classify each observation into one of different states, and

this is something that we could do with real data. The package ignore the actual times, even if the data are not regular in time, the HMM package will treat the data as if it is regularly space in time.

3.7 The prior distribution

Some of the information used in setting the prior distributions were obtained from discussions with ecologists. In some situations, if the species is well studied, the ecologist may know a great deal about its behaviour and its movements. If the species has not been studied extensively, the ecologist probably has only a limited idea of how the animal moves. For this reason, the statistician might want to use an uninformative prior distribution. An uninformative prior helps us to know what we can learn just from the data, without being strongly influenced by the prior of one particular ecologist. Statisticians might want to try to convince or demonstrate to a wide range of scientists that something interesting has emerged from the dataset, which is the reason they use uninformative prior distributions. Using an uninformative prior distribution that leads to a clear conclusion can convince others that using a different prior distribution will lead to a similar conclusion. It is generally better to use the informative prior distribution of the ecologist. However, if it is not possible or the ecologist does not have strong information about the species, then the statistician can use a less informative prior distribution. The general idea is that when the data sample size is large, then the posterior distribution is dominated by the data, so using an informative prior or an uninformative prior distribution should yield similar results.

Part of the prior information can also be obtained from thinking about the interpretation and meaning of the model and thinking about what it represents: i.e. what the model is meant to represent. For example, in the fisher data (Section 5.1.1), the adaptive model (for more detail see section 3.6.2) only makes sense if the animal changes its behaviour quite quickly to match the habitat that it is in. Because of this, Blackwell et al. (2016) used prior information that made it very unlikely that the transition rate would be very low. They tried to ensure that the transition rate was not too low because when the animal enters a new habitat, its behaviour should switch to match the habitat it is in. Thus, if the transition rate is too low, then the animal can spend a long time in a habitat without changing its behaviour to match the habitat. This is not what the model is meant to represent.

Prior distributions may be based upon previously published choices, as is the case for most of the examples in this thesis. All data used in this thesis have already been published

and analysed. Therefore, we will attempt to obtain a prior distribution similar to those in published experiments so that we can compare the results. Blackwell (2003) and Blackwell et al. (2016) used the Metropolis-Hastings algorithm to estimate the movement parameters. The prior distribution for μ was bivariate normal. The prior distributions for each v_i and b_i were truncated univariate normal distributions. These priors are used throughout this thesis.

Blackwell (2003) and Blackwell et al. (2016) used Gibbs sampling to estimate the behaviour parameters. In the homogenous model, Blackwell (2003) used the Beta distribution as a marginal prior distribution to estimate the transition rate in the form of (λ/κ). In the adaptive model, Blackwell et al. (2016) used a gamma distribution as the marginal prior distribution for each transition rate. In this thesis we used the Metropolis-Hastings algorithm to update the transition rate but with similar priors to the published papers. The details are given in each Chapter.

3.8 Model extensions

Single animal

This thesis aims to study the movement of a single animal. Other studies that have simultaneously modelled a group of tracked of animals have previously been published. For example, Mu Niu et al. (2016) presented a new approach for modelling group animal movements in continuous time using a multivariate OU process to implement inferences exactly, without any time discretisation or approximation. They developed a method using a multivariate OU process in high-dimensional space by assuming there is a hidden leading point that all the animals in the group are pulled toward. The leading point follows an OU process pulled to an unknown centre of attraction.

The migration path in continuous time

Throughout this thesis, we assume the animal does not migrate during the study. The animal needs to be motivated to migrate. There are two approaches to modelling animal migration. The first is the Brownian motion with drift. For example, if the animal is migrating to the north but in a steady way, then the Brownian motion with drift is used with no particular centre, only the drift to travel in a particular direction, to model the migration. The Brownian motion with drift is a limiting, special case of the OU process. The Brownian motion with drift has variance and linear drift, which is always in the same direction. The advantage of the Brownian motion model is that it can be applied using only data with a small sample size.

The second approach to modelling the migration could be an OU process with a very large v and a very small b. Then the attraction toward the centre would be less and slower so that over the timescale of the data, the animal would always be attracted in the same direction as the location to which it would eventually migrate. Therefore, the centre of the attraction could be the new location, which would yield results similar to those obtained by the Brownian motion.

Chapter 4

Approximation Method

4.1 Introduction

Throughout this thesis we tried to extend Blackwell et al.'s (2016) method (for more detail see Section 3.4), which we called the exact or full method. The main point of this chapter is increasing the speed of the inference and thus its efficiency. As described in the previous chapter (Section 3.4), to update the trajectory in the exact case, the number of potential switches is sampled from the Poisson distribution, whose mean number of potential switches can take any value. In this chapter, however, we assume that between every two observations there is only one potential switch, and we call this the approximation method. In the implementation of the exact case, the estimated parameters are exact for that model. However, in the approximation case, the estimation of the parameters is approximate. We also approximate the number of switches as a step toward approximately estimating the parameters.

In the approximation case, we have fewer potential switches; therefore, we know that the analysis will not be exactly right because it excludes some possible occurrences in the real unobserved path. For example, in using this approximation, it is no longer possible to have two actual switches in the interval, which rules out some possible sequences of behaviours. We expected the model to be faster because there will be fewer potential switches in most cases, so less computation, and the number of potential switches will be fixed, which also gives an opportunity to simplify/speed up the code.

4.2 The model

We update the trajectory of the sort described in Section 3.4, and we choose the interval to include at least three observations, and assume the states at the beginning and end of the interval are known. Now we assume that there is only one potential switch between two observations. Using the approximate method, we take a random interval with three to six observations and assume that the states at the beginning and end of the interval are known and that the states of the observations between them are unknown. The interval should contain at least 3 observations, because it needs to contain at least one observation in addition to the observations that lie at the start and end of the interval, at which the behavioural state is assumed to be known. The reason for not including too many observations is that the acceptance probability is going to decrease as we cover more observations, and for long intervals we will nearly always reject the proposed set of switches and states, which is inefficient. The "known" states represent conditioning within the MCMC and are not actual data, for which states are unobserved. For example, if we take an interval of five observations from t_1 to t_5, we know the states at $J(t_1)$ and $J(t_5)$ only. We can then assume that there is only one potential switch between each observation, which gives four potential switches in this example $x(T_1),...,x(T_4)$. We sample the time for the first potential switch to determine when the switch occurs, at exactly $T_1 \sim exp(\kappa)$, where $t_1 < T_1 < t_2$, and sample the location of the potential switch $x(T_1)$. We subsequently check whether or not this potential switch is an actual switch. If it is, the state of the switch will be $J(T_1)$. For this purpose, we assume the state of the second observation $x(t_2)$ will equal the new state of the previous potential switch $J(t_2) = J(T_1)$, and we repeat the iteration for each switch. The final state must equal the state of the end observation, and the state in the fourth potential switch should equal the state in the fifth observation $J(T_4) = J(t_5)$.

4.3 Implementation

We consider a single interval from t_a to t_{a+k}, and write $t = t_{a+k} - t_a$ for its length, where k is some interval of time , $\Omega = \kappa t$ for the expected number of potential switches in the exact method it contains and

$$M \sim \text{Poisson}(\Omega). \tag{4.1}$$

for the random number of potential switches, in the exact method. We then have

$$M = \begin{cases} 0 & \text{with probability } \exp(-\Omega) \\ 1 & \text{with probability } \Omega\exp(-\Omega) \\ 2 & \text{with probability } \Omega^2\exp(-\Omega)/2! \\ & \text{etc.} \end{cases} \tag{4.2}$$

These probabilities are derived from the probability mass function for the Poisson distribution. The idea of the approximation is to condition on $M \leq 1$, so that

$$M|M \leq 1 = \begin{cases} 0 & \text{with probability } 1/(1+\Omega) \\ 1 & \text{with probability } \Omega/(1+\Omega). \end{cases} \tag{4.3}$$

This probability is related to whether there is one potential switch. The probability that there is no actual switch given there is at least one potential switch can be shown to be:

$$\begin{aligned} p(M=0|M\leq 1) &= \frac{p(M=0 \ \& \ M\leq 1)}{p(M\leq 1)} \\ &= \frac{p(M=0)}{p(M\leq 1)} \\ &= \frac{p(M=0)}{p(M=0)+p(M=1)} \\ &= \frac{\exp(-\Omega)}{\exp(-\Omega)+\Omega\exp(-\Omega)} \\ &= \frac{1}{1+\Omega}. \end{aligned} \tag{4.4}$$

The same holds for the probability that there is one actual switch given there is at least one potential switch:

$$\begin{aligned} p(M=1|M\leq 1) &= \frac{p(M=1 \ \& \ M\leq 1)}{p(M\leq 1)} \\ &= \frac{p(M=1)}{p(M\leq 1)} \\ &= \frac{p(M=1)}{p(M=0)+p(M=1)} \\ &= \frac{\Omega\exp(-\Omega)}{\exp(-\Omega)+\Omega\exp(-\Omega)} \\ &= \frac{\Omega}{1+\Omega}. \end{aligned} \tag{4.5}$$

Obviously this will only be a useful approximation if κ and the intervals between the observations are not too large. To allow for a simpler implementation of this approximate algorithm, we want to arrange for the dimension of the reconstructed trajectory to be fixed. We can do this by representing the case $M=0$ as one potential switch with zero probability of being an actual switch.

In the exact method, a potential switch is an actual switch from state i to state j with probability p_{ij} ; $i \neq j$, where

$$p_{ij} = \frac{\lambda_{ij}}{\kappa}. \tag{4.6}$$

Combining these expressions, in the approximation case, the probability of an actual switch from state i to j within the interval is as follows:

$$p_{ij} = \frac{\Omega\lambda_{ij}}{\kappa(1+\Omega)}. \tag{4.7}$$

So, the likelihood for the behaviour process for each behaviour i will be:

$$\begin{aligned} p(\mathtt{data}|\lambda_{\mathtt{ij}}) &= (p_{ii})^{n_{ii}} \times \prod_{j\neq i}(p_{ij})^{n_{ij}} \\ &= \Big(1-\frac{\lambda_i}{\kappa(1+\Omega)}\Big)^{n_{ii}} \times \prod_{j\neq i}\Big(\frac{\Omega\lambda_{ij}}{\kappa(1+\Omega)}\Big)^{n_{ij}}, \end{aligned} \tag{4.8}$$

where n_{ii} is the number of switches out of state i, and n_{ij} is the number of switches from state i to state j. We sample the potential switch time T_a, location $x(T_a)$ and state $J(T_a)$

conditional on the trajectory outside the interval where potential state $J(T_a)$ should equal the state at the end of the interval $J(t_{a+k})$, where is k is the number of observations in the interval." ; otherwise the proposal is rejected. Then we can perform an update to that part of the trajectory by proposing new values T_a' and then performing a Metropolis-Hastings accept/reject step. Usually we would do this for a longer interval, eg. from t_a to t_{a+k}. In the approximation approach, we proposed the number of switches in a way different from that proposed in the exact approach, which affects the way the trajectory is updated. Also, updating the behaviour parameters is different in the two approaches because the probability has changed. However, the updating for the movement parameters is the same in the two approaches.

4.3.1 Other approaches of approximation

Here, we present two different approach approximations to the Poisson distribution. Conditioning is not a very accurate way of approximating the Poisson distribution for the number of the switches M with a Bernoulli distribution. An alternative method involves having a potential switch whenever $M \geq 1$ and having a non-switch when $M = 0$, with probabilities of $1 - \exp(-\Omega)$ and $\exp(-\Omega)$, respectively. In this case, the probability of having an actual switch from i to j is:

$$(1 - \exp(-\Omega))\frac{\lambda_{ij}}{\kappa}. \tag{4.9}$$

Another possibility is to try to match the mean number of potential switches. For example, we want a potential switch with a probability equal to Ω if possible; in practice, the probability would have to be $\min\{\Omega, 1\}$. The probability of an actual switch from i to j is:

$$\min\{\Omega, 1\}\frac{\lambda_{ij}}{\kappa}. \tag{4.10}$$

4.4 Simulated experiments

4.4.1 Data

We used simulated data that were similar to the data used by Patterson et al. (2017) in which the movement model was assumed to follow an OU process with three behavioural states. We used these data because Patterson et al. (2017) analyse them using the exact method of Blackwell et al. (2016). We used these simulated data to test and compare the exact method with the new approximate approach. In this experiment, the model was

homogeneous (Section 3.4.1), so there is only one region in the habitat. There are three behavioural states, so we will have three OU processes, each associated with one state. The movement parameters for each OU process are: the variance for each state $v_i = (0.05, 0.5, 5)$, and all the three states share the same rate of attraction $b_i = 0.1$ where $i = 1,2,3$ and also share the same centre of attraction $\mu_i = (0,0)$. The behaviour parameters λ_{ij} are determined by the generator matrix:

$$G = \begin{pmatrix} -0.10 & 0.04 & 0.06 \\ 0.025 & -0.05 & 0.025 \\ 0.2 & 0 & -0.20 \end{pmatrix}. \tag{4.11}$$

The sample size for the simulated data was 100, and the time interval between observations was one unit of time.

4.4.2 The model

For each of the following experiments, we ran the data twice: once with the exact approach and then using the same data with the approximation approach to compare the two approaches. We ran each experiment for 10 million iterations using a thinning ratio of 1000 and a burn-in period of 5000 iterations after thinning. The value of κ is fixed at 0.4. We are taking kappa to be 0.4 because, in reality, we would not know the true parameter values, and we need to give a reasonable range for their possible values.

In the exact method Gibbs sampling is used to update the behaviour parameter, while in the approximation method we used a Metropolis-Hastings algorithm to update the behaviour process for the transition rates. In the exact method the probability of an actual switch was fixed and does not depend on the interval length between observations. However, in the approximation method the probability of an actual switch is more complicated and depends on the interval length between observations.

In both approaches, the transition rates, expressed relative to κ (i.e. λ/κ) are assumed to have a joint prior distribution that is Dirichlet, so that the marginal prior distribution for each λ/κ is a Beta distribution with parameters $(\alpha = 1, \beta = 3)$. Because we need the prior to be bounded. The transition rates λ can take any value in the range $(0,\infty)$. Instead of assigning priors to the transition rates themselves, however, we assign them to the transition rates divided by kappa (when kappa is fixed) - i.e. to λ/κ. Because kappa is the upper bound for the transition rates the value of λ/κ must be in the range $(0,1)$. We assume that the joint prior distribution for λ/κ is a Dirichlet distribution, because this ensures the prior for each

marginal distribution is bounded to $(0,1)$ and has a Beta distribution.In the approximation approach, we used a truncated normal distribution centred on the current value to propose a new value for the transition rate. Updating the movement parameters is the same in both methods, as described in Section 3.6. For the movement parameters: The prior distribution for μ is bivariate normal with a mean vector $(0,0)$ and SD 2. The prior distribution for each v_i is a univariate normal distribution with mean$=0.1$ and SD =20 truncated to $(0,\text{inf})$. The prior distribution for each b_i is a univariate normal distribution with mean$=0.1$ and SD =2 truncated to $(0,1)$. The initial states are unknown and sampled from a HMM package, as in Section 3.6.3. The initial value for simulation and the MCMC can be accessed from the code.

4.4.3 Example (1): Compare exact method with approximation method

We compared the results from both, the exact and the approximate methods. Figure 4.1 shows the MCMC sample of the posterior distribution for the two main parameters that control the movement b_i and v_i in the model for each behavioural state. We can note from Figure 4.1 and Table 4.1 that the result for the approximate method was very similar to the result for the exact method. The estimation for the state one parameters is more precise than for the other states. This is a property of the data itself, not the method, because both approaches get a similar result. Figure 4.2 shows a map of the trajectory data, where the animal spends little time in state 3, which makes estimating the parameters for state 3 imprecise. The standard deviation for v_3 is very large and that because the animal spend few time in state 3. Both approaches get a more precise estimate for the movement parameters for state one than for the other states. Also, both approaches overestimate the transition rate; again, this issue is related to the small sample size, not the method. Figure 4.2 shows there are 9 switches in total between the different states. To improve the mixing and to get a more precise estimate, we need to increase the sample size because the 100 observations contain very little information about the states and switches as there are not many transitions between different behaviours, and that makes trying to estimate the transition rate difficult. To solve this, we increase the sample size and the transition rate. Increasing the transition rate means having more switches.

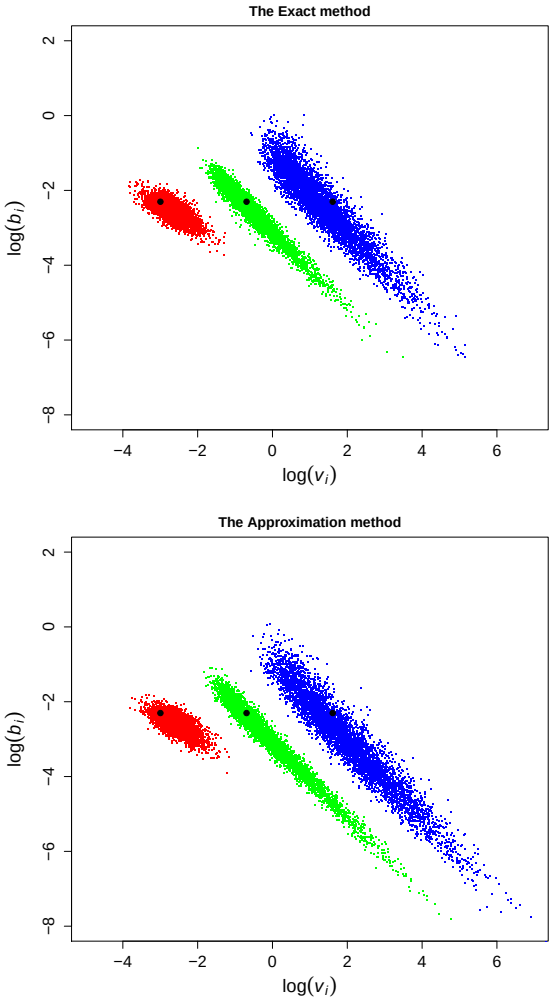

Fig. 4.1 Posterior distributions for the log of the movement parameters of all the behavioural states where the sample size is 101, Example (1). The movement parameters for states 1, 2 and 3 can be seen in red, green and blue, respectively. The real values are displayed as black dots.

	real value	Exact	Approximation
μ_x	0	-0.29(0.24)	-0.32(0.26)
μ_y	0	-0.12(0.29)	-0.17(0.3)
b_1	0.1	0.079(0.018)	0.077(0.017)
b_2	0.1	0.091(0.05)	0.071(0.05)
b_3	0.1	0.136(0.12)	0.092(0.11)
v_1	0.05	0.07(0.025)	0.078(0.028)
v_2	0.5	0.79(1.08)	1.77(4.25)
v_3	5	6.4(10.8)	17.12(44.61)
λ_{12}	0.04	0.087(0.05)	0.091(0.05)
λ_{13}	0.06	0.095 (0.53)	0.099(0.52)
λ_{21}	0.025	0.036(0.035)	0.046(0.037)
λ_{23}	0.025	0.05(0.035)	0.044(0.035)
λ_{31}	0.2	0.18(0.066)	0.18(0.066)
λ_{32}	0	0.05(0.043)	0.049(0.045)
running time	-	1h 38min	1h 53min

Table 4.1 Posterior means and standard deviations (SDs) for the parameters in both approaches, where the sample size is 101, Example (1).

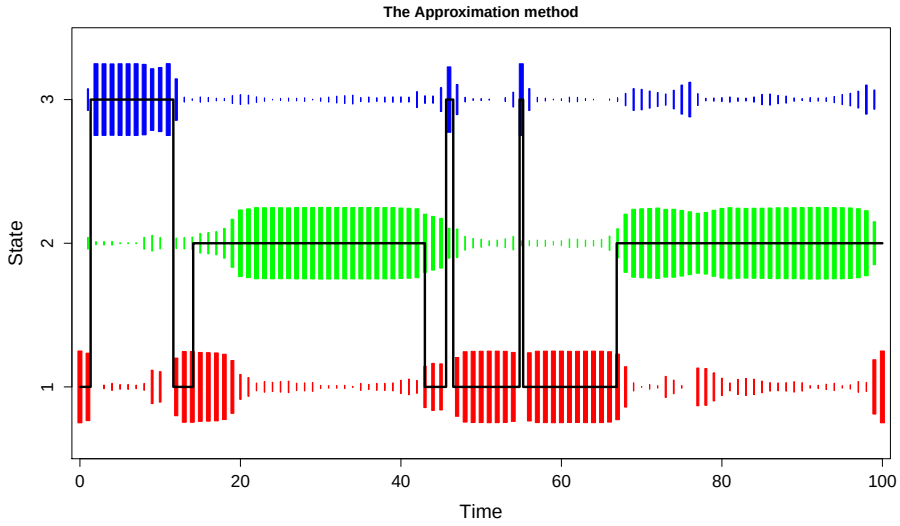

Fig. 4.2 The trajectory for each observation to follow state 1, 2 or 3, For example (1).

4.4.4 Example (2): Approximation method with higher transition rate and larger sample size

Example 2.a: Larger sample size.

In this example, we used data and methods similar to those in Example 1, but first the data sample size was increased from 100 to 300. After we increase the sample size to 300 observations, the number of switches between different states increase to 31 switches. Figure 4.4 shows there are 31 switches in total between the different states. The results in Figure 4.3 and Table 4.2 show that the posterior results are more precise than the results in Example 1 for both the exact and approximation methods for all the parameters. However, the approximation approach overestimates v_2, which can be seen in Figure 4.3 and Table 4.2. Figure 4.5 compare the posterior results for the transition rate values from the approximation approach with the marginal prior distribution.

	real value	Exact	Approximation
μ_x	0	0.008(0.084)	0.01 (0.088)
μ_y	0	0.044 (0.082)	0.0375 (0.083)
b_1	0.1	0.1(0.006)	0.1(0.006)
b_2	0.1	0.13 (0.05)	0.1 (0.06)
b_3	0.1	0.099 (0.04)	0.09 (0.04)
v_1	0.05	0.046 (0.005)	0.048 (0.005)
v_2	0.5	0.59 (0.78)	2.34 (10.21)
v_3	5	7.407 (5.747)	10.58 (14.23)
λ_{12}	0.04	0.049 (0.025)	0.05(0.024)
λ_{13}	0.06	0.064 (0.026)	0.065(0.025)
λ_{21}	0.025	0.06 (0.036)	0.092(0.043)
λ_{23}	0.025	0.06 (0.034)	0.053(0.035)
λ_{31}	0.2	0.1 (0.034)	0.09 (0.035)
λ_{32}	0	0.024 (0.02)	0.052 (0.033)
running time	-	1h 50 min	2h 8min

Table 4.2 Posterior means and SDs for the parameters in both approaches for Example 2.a, where the sample size has been increases from 100 to 300.

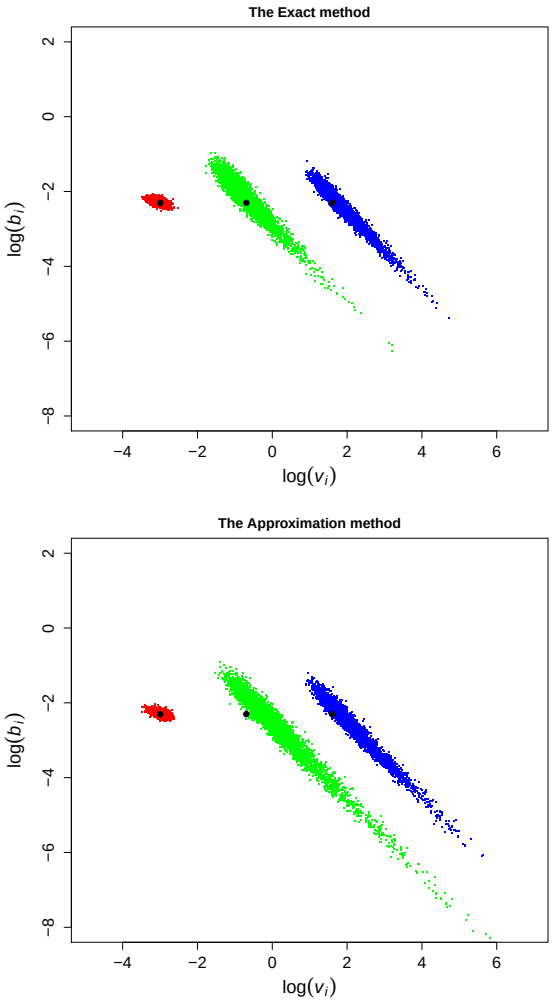

Fig. 4.3 Posterior distributions for the log movement parameters of all the behavioural states, where the sample sizes is 301, Example 2.a. The parameters for states 1, 2 and 3 can be seen in red, green and blue, respectively. The real values are displayed as black dots.

Example 2.b: Larger sample size and doubled the transition rate.

Then we did another run with 300 observations, and the transition rates in the generator matrix were doubled except λ_{31} to improve the estimation; the generator matrix is as follow:

$$G = \begin{pmatrix} -0.20 & 0.08 & 0.12 \\ 0.05 & -0.1 & 0.05 \\ 0.2 & 0 & -0.2 \end{pmatrix}. \tag{4.12}$$

The results shown in Figure 4.6 and Table 4.3 are similar to those in Example 2.a; however, after doubling the value for the transition rates, the CV for the posterior distributions for

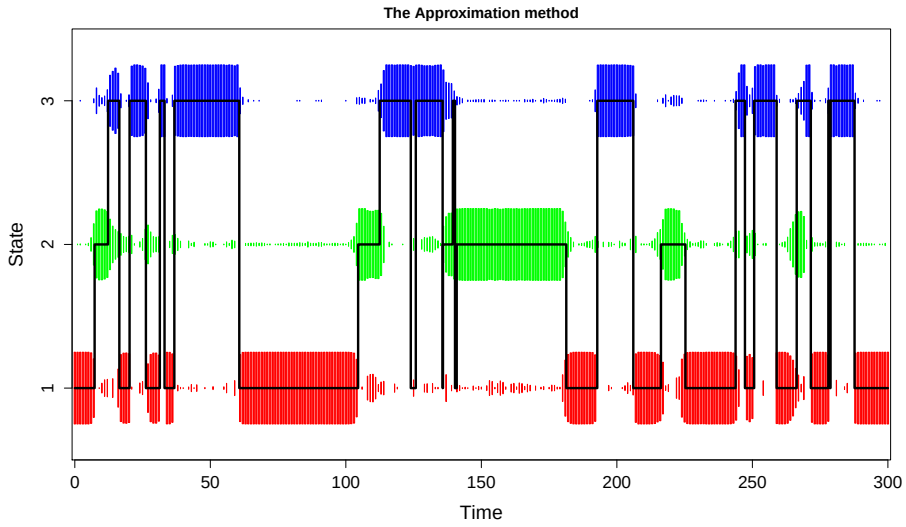

Fig. 4.4 The trajectory for each observation to follow state 1, 2 or 3, For example (2).

the transition rates were slightly smaller than the results in Example 2.a for both the exact and approximation methods. The MCMC algorithm mixined better when we increased the sample size from 100 to 300. Also, by increasing the transition rates, the number of switches increases, which results in more precise estimates for the transition rates. However, because we have not double all the transition rates the number of switches between different states in this example stay the same as Example 2.a.

4.4.5 Example (3): Approximation method with a similar variance for all the states

In this example, we used data and methods similar to those in Example 2, but the v_i values were closer to each other $v_i = (0.25, 0.5, 1)$, which meant that there were no large differences between the three states. In Figure 4.7 we can see that in the exact approach there were only small differences between the movement posterior for the three states. In the approximation approach the movement posterior for state one and state two overlap. This example shows that the model cannot recognise the difference between the three states when the v_i have very small differences in the value between all the states. Table 4.4 shows the posterior results for all the parameters in both approaches.

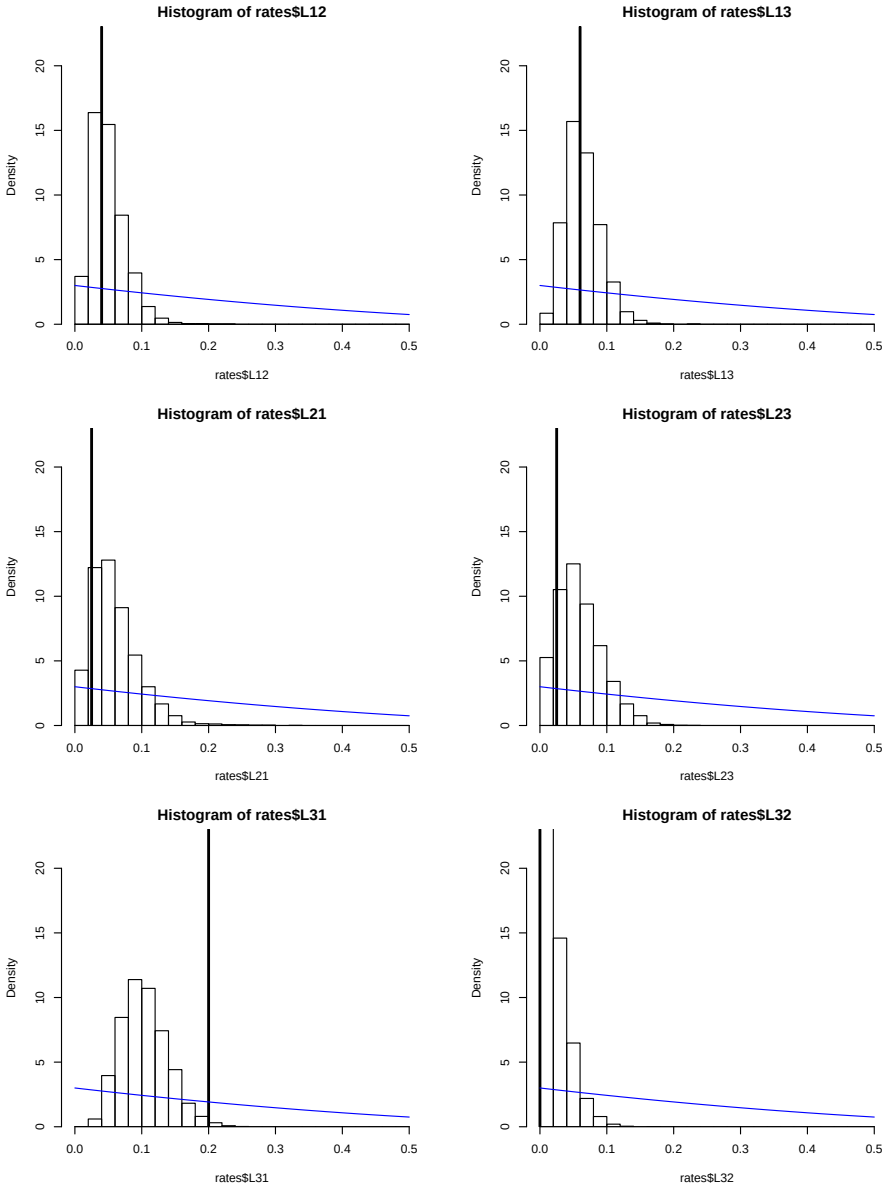

Fig. 4.5 For example 2, the posterior results for the transition rate of the three behavioural states. The true values are displayed as vertical black lines. The blue curve is the marginal prior distribution.

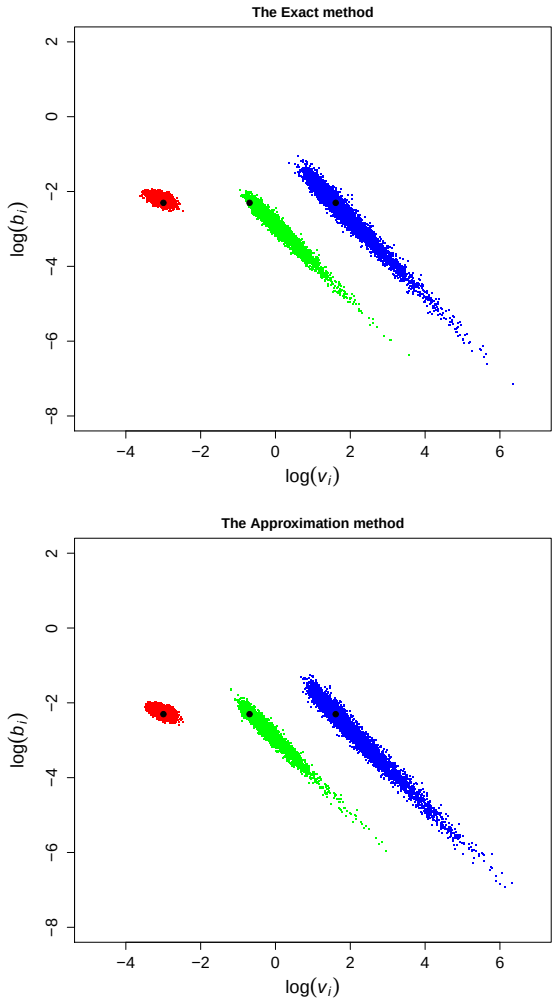

Fig. 4.6 Posterior distributions for the log movement parameters of the all behavioural states, for Example 2.b. The parameters for states 1, 2 and 3 can be seen in red, green and blue, respectively. The real values are displayed as black dots.

4.4.6 Example (4): Different way of approximation

Here, we used the same data and methods as in Example 2, but we used the two different approaches of approximations described in Section 4.3.1. From Figure 4.8 we can see that the minimum approximation and the exponential approximation produced a result similar to that of the exact and approximate methods, as shown in Figure 4.6. Also, Table 4.4 shows that the posterior results for all the parameters in both approaches are very similar. We concluded that the results from the approximate model were very close to the results from the exact method. The results are robust in regard to the details of how the approximation is done.

	real value	Exact	Approximation
μ_x	0	-0.12(0.12)	-0.13 (0.12)
μ_y	0	0.19(0.13)	0.156(0.12)
b_1	0.1	0.11(0.009)	0.11(0.0009)
b_2	0.1	0.058(0.023)	0.069(0.025)
b_3	0.1	0.086(0.051)	0.074(0.05)
v_1	0.05	0.046(0.007)	0.049(0.007)
v_2	0.5	1.243(1.145)	0.94(0.82)
v_3	5	10.22(18.21)	14.79(27.37)
λ_{12}	0.08	0.1 (0.045)	0.091(0.039)
λ_{13}	0.12	0.1(0.043)	0.095(0.037)
λ_{21}	0.05	0.07(0.036)	0.041(0.026)
λ_{23}	0.05	0.06(0.033)	0.082(0.031)
λ_{31}	0.2	0.14(0.045)	0.16(0.05)
λ_{32}	0	0.09(0.044)	0.1(0.051)
running time	-	1h 58min	2h 14 min

Table 4.3 Posterior means and SDs for the parameters in both approaches for Example 2.b, where the transition rate values have been doubled.

	real value	Exact	Approximation
μ_x	0	-0.05(0.19)	-0.061 (0.19)
μ_y	0	-0.2(0.18)	-0.22(0.19)
b_1	0.1	0.093(0.05)	0.134(0.069)
b_2	0.1	0.17(2.4)	0.04(0.046)
b_3	0.1	0.14(0.07)	0.15(0.069)
v_1	0.25	0.4(0.4)	0.31(0.4)
v_2	0.5	1(1.56)	3.43(7.22)
v_3	1	2.02(7.1)	1.52(7.1)
λ_{12}	0.08	0.081(0.064)	0.09(0.067)
λ_{13}	0.12	0.08(0.055)	0.081(0.056)
λ_{21}	0.05	0.087(0.068)	0.072(0.057)
λ_{23}	0.05	0.1(0.064)	0.13(0.065)
λ_{31}	0.2	0.175(0.064)	0.15(0.069)
λ_{32}	0	0.058(0.051)	0.08(0.055)
running time	-	1h 51 min	2h 19 min

Table 4.4 Posterior means and SDs for the parameters in both approaches for Example 3.

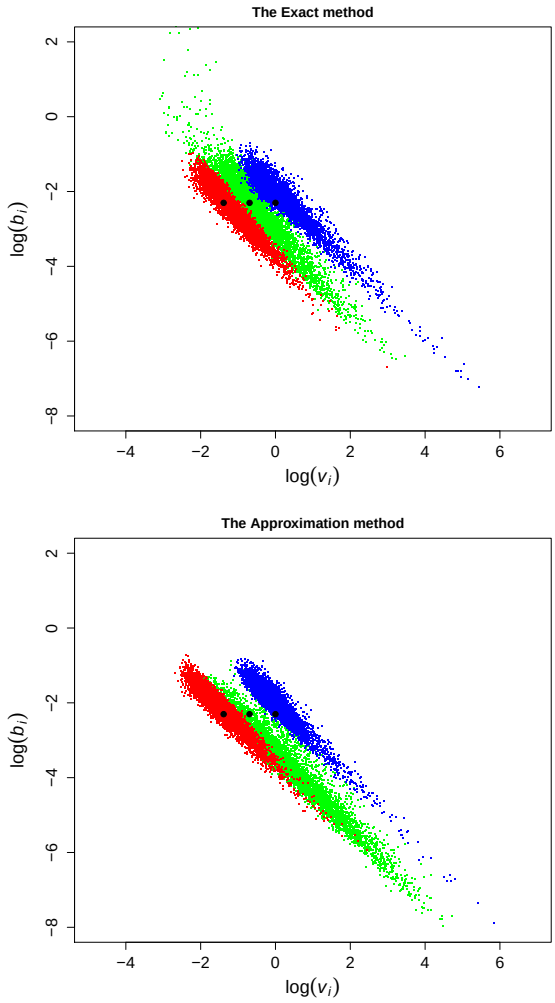

Fig. 4.7 Posterior distributions for the log movement parameters of all the behavioural states, where the v_i values are close to each other, Example 3. The parameters for states 1, 2 and 3 can be seen in red, green and blue, respectively. The real values are displayed as black dots.

4.5 Discussion

In this chapter we tried to improve the efficiency for Blackwell et al.'s (2016) exact method by approximating the posterior parameters. We made the approximation by assuming there is only one potential switch between each pair of observations rather than generating the number of potential switches from the Poisson distribution. We presented three different approximation approaches, and all gave results very close to each other and to the exact method. However, none of them were faster than the exact approach. Maybe this occurred because we had to use the Metropolis-Hastings method to update the transition rates.

	real value	Exponential	minimum
μ_x	0	-0.13(0.125)	-0.12(0.13)
μ_y	0	0.14(0.127)	0.16(0.13)
b_1	0.1	0.11(0.009)	0.11(0.0095)
b_2	0.1	0.066(0.026)	0.062(0.027)
b_3	0.1	0.072(0.048)	0.079(0.053)
v_1	0.05	0.049(0.008)	0.05(0.009)
v_2	0.5	1.127(1.69)	1.43(1.75)
v_3	5	13.82 (21.45)	17.56(43.39)
λ_{12}	0.08	0.11(0.04)	0.11(0.04)
λ_{13}	0.12	0.091(0.036)	0.09(0.035)
λ_{21}	0.05	0.047(0.025)	0.087(0.04)
λ_{23}	0.05	0.082(0.029)	0.053(0.026)
λ_{31}	0.2	0.17(0.047)	0.13(0.043)
λ_{32}	0	0.077(0.045)	0.1(0.04)
running time	-	2h 14 min	2h 16 min

Table 4.5 Posterior means and SDs for the parameters in both approaches for Example 4.

The strength of the approximation method is that the model performs very well and gives us very close results to those of the exact case, but the weakness was that it did not speed the inference up as much as we had hoped. In these scenarios we attained results similar to those of the exact method when we used the approximate method, it took the same amount of time. In other words, we could not generate a new method that was faster than the exact method. one main limitation for the experiments presented in this chapter is with kappa 0.4, most of the time the approximation method will have the same number of potential switches as the exact method, so of course it is a good approximation, and of course it is not much faster. The probability of being two or more switches/transition between the observations is only about 0.06. Part of the reason we expected to save some time even in this case, not because of those fewer potential switches but because the number of switches/transitions is fixed so the dimension of the problem is fixed. In practice, as we have shown, the results of both experiments are the same. The positive point is that the approximation idea works well for estimation, and different versions of the approximation approach give the same answer. Unless the speed is increased, there is not much to gain. Regarding the present algorithm, it is not worth pursuing because of the computational cost of the present models. However, in future research on different continuous time models, the idea presented here might be useful. This idea may work with other classes of models or in other extensions or variations. Alternatively, there may be cases where the approximation is more useful and beneficial.

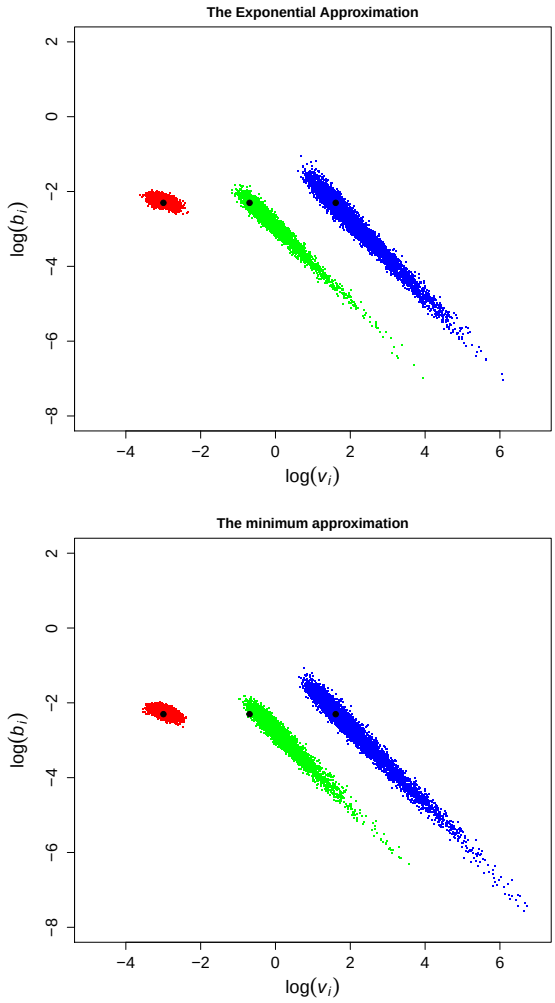

Fig. 4.8 Posterior distributions for log movement parameters of all the behavioural states in both exponential and minimum approaches, Example 4. Parameters for states 1,2 and 3 can be seen in red, green and blue respectively. The real values are displayed as black dot.

Accordingly, in the next chapter, we use the exact method with a variable kappa rather than a fixed global upper bound (kappa) for the transition rates. This is to increase efficiency as the kappa can be closer to the actual switching rate in the model. This is also to reduce the computational cost. This approach is also more general than the existing exact method because we do not need a fixed upper bound on the prior support for the transition rates.

Chapter 5

Variable Kappa

5.1 Introduction

The aim of this chapter is to improve the Blackwell et al.'s (2016) algorithm by making it more efficient. As mentioned in Section 3.4, the use of kappa is essential to fitting with the exact method, in which kappa is an upper bound for the transition rates. Blackwell et al. (2016) used a fixed value of kappa, so the transition rates are globally bounded above. However, in this chapter, we allow kappa to vary so that we do not need to fix an upper bound on the prior distribution for the transition rates. Allowing kappa to vary will potentially make the model more efficient, universal and flexible by saving time. Kappa will potentially vary from each iteration to the next, in a way that depends on the data through the lambda parameters. Kappa is treated either as a function of the parameters or as a parameter in its own right, that will be estimated from the data and part of inference

Blackwell et al. (2016) stated that the value of the fixed kappa should be sufficiently large, but in a real dataset, we can never be sure if the fixed kappa value is large enough. Additionally, if the fixed kappa value is too large, it will give the correct answer, but it will not be efficient because the run will take too long. Using the fixed kappa approach with a value that is too high results in too many potential switches between observations, slowing down the algorithm and increasing the computational cost. Allowing kappa to vary is potentially faster than using a fixed kappa because the kappa value does not need to be made too large; the new kappa value will adapt to the values of the transition rates λ_{ij}, as they are learnt during the inference. Another benefit of the variable kappa is that when starting with a very small kappa value, the kappa will increase or decrease according to the values of the transition rates. In the case of fixed kappa, starting with a very small value will causes flawed results.

One of the limitations of the exact method with fixed kappa is that if the transition rate prior is unbounded, approximation is needed to fit the model. With variable kappa, the exact method can be used with any form of prior for the transition rates, i.e. bounded or unbounded.

In Sections 5.2 and 5.3, we describe how to the model can be implemented using variable kappa. In Section 5.3, we use simulated data to test the idea. In Section 5.5, we use real data to compare our results with those of published work. Section 5.6, provides the conclusions.

5.2 The model

There are two ways to implement the variable kappa idea, described in the following two subsections.

5.2.1 Kappa as a function of the transition rate

In this case, kappa is a deterministic function of the transition rates λ_{ij}. As the transition rates λ vary in the MCMC updating, so does kappa because at any given iteration, we need:

$$\kappa \geq \max_{i,x} \lambda_i(x), \tag{5.1}$$

where

$$\lambda_i(x) = \sum_{i \neq j} \lambda_{ij}(x). \tag{5.2}$$

is the transition rate out of state i. The simplest idea is to always take:

$$\kappa = \max_{i,x} \lambda_i(x). \tag{5.3}$$

i.e. kappa is given by the maximum row sum of the (off-diagonal) lambdas in the generator matrix across all locations.

In many cases, individual λ_{ij} values will be strictly less than kappa, as will all but one the λ_is. However, there are some particular models in which $\lambda_i(x)$ is equal to one of the $\lambda_{ij}(x)$, either for particular values of i and x or more generally. For example, in a two-state model, $\lambda_1(x) = \lambda_{12}(x)$ and $\lambda_2(x) = \lambda_{21}(x)$. Similarly, in the adaptive model (described in

Section 3.4.2),

$$\lambda_i(x) = \lambda_{ij}(x), \tag{5.4}$$

in region j, so kappa will be equal to one of the lambdas. This can lead to situations in which the probability of a specific transition is one at a potential switch, leading to difficulties in updating state sequences and, therefore, to poor mixing. The probability for a switch happening from state$i$ to state j is λ_{ij}/κ. If kappa is equal to one of the lambdas, then the probability to switch to this state will be one and the probability to switch to other states will be zero. In this case, it is not possible to generate a feasible trajectory for behaviour which then makes it difficult to update the lambdas, since the probability is equal to zero for a wide range of values of the lambdas. In particular, for example, runs with the three-state adaptive model (details not included) show that this 'obvious' way of defining kappa does not work in practice. This problem can be avoided by defining kappa as a different function of the transition rate λs. To improve the mixing, we make kappa slightly higher than required:

$$\kappa = (1+\varepsilon)\max_{i,x}\lambda_i(x). \tag{5.5}$$

where ε is some small value. This small addition makes kappa a little larger than the lambda and slows down the run slightly, but it seems to improve the mixing. An alternative would be:

$$\kappa = \max_{i,x}\lambda_i(x) + \varepsilon, \tag{5.6}$$

but multiplying by $(1+\varepsilon)$ instead of adding ε is better, as this rate will be different depending on the example. In the multiplicative case, if the lambdas are completely different in each example, using the same value for epsilon still makes sense. We will fix this epsilon value for all of the examples. Currently, we have lambda values that are around 3.0 and an epsilon of 0.1, which is related to the definition of an epsilon in Equation 5.5; this is fine for the current experiment. In experiments with a much larger or smaller lambda, an additive epsilon value would need to be changed, whereas a multiplicative term is more likely to be transferable between examples.

We add Equation 5.5 to the MCMC algorithm for updating the transition rates λ so that every time the lambda updates, the kappa will also change. Adding a small value to the maximum works well with homogenous and adaptive models (details in Section 3.4), and it is better than just using the maximum. The results are omitted because we are going to focus on the second approach. To improve the algorithm, we want a more general idea that can work with any model.

5.2.2 Kappa as a parameter

The second approach is to think of kappa as a parameter and allow it to update separately from the lambdas. Obviously, kappa and lambda are connected because kappa needs to satisfy Equation 5.1 at each iteration. Kappa is difficult to interpret through this approach, as it is not really a parameter; changing kappa does not change the true posterior for the transition rate parameters, provided Equation 5.1 is satisfied. However, treating kappa as if it was a parameter makes it much easier to update the lambda and kappa separately. We add an extra MCMC step to the algorithm to update kappa. We will have two likelihood calculations and two Hastings ratios for the separate updates. By having separate proposals, kappa will typically change quite slowly, whereas the lambdas change relatively quickly. Within this approach, we need to specify a joint prior for the lambdas and for kappa. The prior for the λs potentially represents 'real' prior information, whereas that for kappa is an artefact of this version of the algorithm. This suggests an independent form for the joint prior, which would look as follows:

$$p(\underset{\sim}{\lambda}, \kappa) \propto f(\underset{\sim}{\lambda})g(\kappa), \tag{5.7}$$

where $f(\underset{\sim}{\lambda})$ is a function only of the transition rates $\underset{\sim}{\lambda}$, and $g(\kappa)$ is a function only of kappa κ; and $\underset{\sim}{\lambda}$ represents the vector of λ_{ij}s. However, the joint prior distributions cannot be independent, as $\underset{\sim}{\lambda}$ and κ must satisfy Equation 5.1. Instead, to keep the prior as simple as possible whilst satisfying the inequality, we let:

$$p(\underset{\sim}{\lambda}, \kappa) \propto f(\underset{\sim}{\lambda})g(\kappa)I_{\{\kappa \geq \max_{i,x} \lambda_i(x)\}}, \tag{5.8}$$

where I_A is the indicator function of the event A, taking a value of 1 if A occurs and zero otherwise. In this case,

$$I_{\{\kappa \geq \max_{i,x} \lambda_i(x)\}} = \begin{cases} 1 & \text{if} \quad \kappa \geq \max_{i,x} \lambda_i(x) \\ 0 & \text{otherwise.} \end{cases} \tag{5.9}$$

To do separate updates for lambda and kappa, we need both the conditional prior for the lambdas given a particular kappa,

$$p(\underset{\sim}{\lambda} \mid \kappa) \propto f(\underset{\sim}{\lambda})I_{\{\kappa \geq \max_{i,x} \lambda_i(x)\}}. \tag{5.10}$$

and the conditional prior for kappa given the lambdas,

$$p(\kappa \mid \underset{\sim}{\lambda}) \propto g(\kappa)I_{\{\kappa \geq \max_{i,x} \lambda_i(x)\}}. \tag{5.11}$$

We also need the marginal prior for lambda, integrating over all values of kappa, to match the required 'true' prior for $\underset{\sim}{\lambda}$.

$$\begin{aligned} p(\underset{\sim}{\lambda}) &= \int_{\kappa} p(\underset{\sim}{\lambda}, \kappa)\, d\kappa, \\ &= \int_{\kappa=0}^{\infty} f(\underset{\sim}{\lambda})\, g(\kappa)\, I_{\{\kappa \geq \max_{i,x}\, \lambda_i(x)\}}\, d\kappa, \\ &= \int_{\kappa=\max_{i,x} \lambda_i(x)}^{\infty} f(\underset{\sim}{\lambda})\, g(\kappa)\, d\kappa, \\ &= f(\underset{\sim}{\lambda}) \int_{\kappa=\max_{i,x} \lambda_i(x)}^{\infty} g(\kappa)\, d\kappa. \end{aligned} \tag{5.12}$$

We do not care what the prior distribution is for kappa, except that it needs to satisfy the condition that kappa is always greater than the appropriate sum of the lambdas, as shown in Equation 5.1. However, the kappa value should not make any difference as long this condition is always satisfied. In principle, if the kappa value increases, there will be no difference in the posterior result for the transition rate, only in the efficiency of the algorithm. Furthermore, we want the prior to give a slightly greater probability to the small values because we want to try keeping the kappa value small for efficiency. The marginal prior for kappa is :

$$p(\kappa) = \int_{\lambda} p(\underset{\sim}{\lambda}, \kappa) d\lambda. \tag{5.13}$$

For the MCMC algorithm, we need to consider how the other quantities depend on λ and κ. Only the behavioural states depend directly on $\underset{\sim}{\lambda}$, and only the potential switching times depend on κ. The conditional posterior to update lambda, which is the density of lambda given the states and kappa, is therefore:

$$\begin{aligned} p(\underset{\sim}{\lambda} \mid \kappa, S) &\propto p(\underset{\sim}{\lambda} \mid \kappa) \;\times\; p(S \mid \underset{\sim}{\lambda}), \\ &\propto f(\underset{\sim}{\lambda})\, I_{\{\kappa \geq \max_{i,x}\, \lambda_i(x)\}} \;\times\; p(S \mid \underset{\sim}{\lambda}). \end{aligned} \tag{5.14}$$

The conditional posterior to update kappa, which is the density for lambda given kappa and the switching time, is:

$$\begin{aligned} p(\kappa \mid \underset{\sim}{\lambda}, T) &\propto p(\kappa \mid \underset{\sim}{\lambda}) \;\times\; p(T \mid \kappa), \\ &\propto g(\kappa)\, I_{\{\kappa \geq \max_{i,x}\, \lambda_i(x)\}} \;\times\; p(T \mid \kappa). \end{aligned} \tag{5.15}$$

Here, $p(S \mid \underset{\sim}{\lambda})$ is the probability of the current state sequence occurring S, given the transition rates $\underset{\sim}{\lambda}$, and $p(T \mid \kappa)$ is the probability of the current set of potential switching times T as a realisation of a Poisson(κ) process.

These conditional posteriors are known only up to proportionality, but we are interested only in the Hastings ratio calculation. Therefore, all terms not shown in Equation 5.18 and 5.19 will cancel. The Hastings ratio for a proposed value $\underset{\sim}{\lambda}'$ given a current value $\underset{\sim}{\lambda}$ is:

$$\begin{aligned} &= \frac{q(\lambda \mid \lambda')}{q(\lambda' \mid \lambda)} . \frac{p(\underset{\sim}{\lambda}' \mid \kappa, S)}{p(\underset{\sim}{\lambda} \mid \kappa, S)} \\ &= \frac{q(\lambda \mid \lambda')}{q(\lambda' \mid \lambda)} . \frac{f(\underset{\sim}{\lambda}')\, I_{\{\kappa \geq \max_{i,x}\, \lambda_i(x)\}}\, p(S \mid \underset{\sim}{\lambda}')}{f(\underset{\sim}{\lambda})\, p(S \mid \underset{\sim}{\lambda})}, \end{aligned} \tag{5.16}$$

assuming λ ,κ satisfy $\kappa \geq \max_{i,x} \lambda_i(x)$, and for κ' given κ, it is:

$$\frac{q(\kappa \mid \kappa')}{q(\kappa' \mid \kappa)} . \frac{g(\kappa')\, I_{\{\kappa \geq \max_{i,x}\, \lambda_i(x)\}}\, p(T \mid \kappa')}{g(\kappa)\, p(T \mid \kappa)}, \tag{5.17}$$

where $q(.|.)$ is the proposal forward and backward functions, and κ' and λ' are the proposed values for kappa and lambda, respectively. In the code, we incorporate the indicator function by truncating the proposal distribution for both lambda and kappa. We need to have a proposal for kappa that is bounded below by the maximum lambda value. We let kappa have a normal proposal centred on the current value but truncated below at the maximum lambda. In other words, the kappa will be able to change slightly and will be more likely to move toward the transition rates value, although it cannot go below the maximum. We will have both forward and backward proposal terms in the Hastings ratio because we have this truncation. We have:

$$q(\kappa' \mid \kappa) = \frac{\phi(\kappa' \mid \kappa, \sigma_\kappa^2)}{1 - \Phi(\max(\lambda_i) \mid \kappa, \sigma_\kappa^2)}, \tag{5.18}$$

where $\phi(\kappa' \mid \kappa, \sigma_\kappa^2)$ and $\Phi(\max(\lambda_i) \mid \kappa, \sigma_\kappa^2)$ are the p.d.f and c.d.f, respectively, of the untruncated normal distribution with a specified mean and variance, and σ_κ^2 is the proposal variance. Thus, the Hastings ratio includes a term:

$$\frac{q(\kappa \mid \kappa')}{q(\kappa' \mid \kappa)} = \frac{1 - \Phi(\max(\lambda_i) \mid \kappa, \sigma_\kappa^2)}{1 - \Phi(\max(\lambda_i) \mid \kappa', \sigma_\kappa^2)}, \tag{5.19}$$

as the $\phi(\kappa' \mid \kappa, \sigma_\kappa^2)$ terms cancel by symmetry.

This approach, in which kappa is a parameter, also enables us to update each lambda separately, which is closer to the fixed case and is definitely better than updating all the lambda togather, but there is no considerable difference in performance. Updating each transition rate separately may be useful in some cases.

5.3 Implementation

We use the second approach with kappa as a parameter, to fit the data. All the implementations should be the same as in Section 3.4, except for updating the behaviour parameter. With a fixed value of kappa, as described in Section 3.4, we can use the Gibbs sampling algorithm to update the behaviour parameters λ_{ij}. In the variable kappa approach, the Metropolis-Hastings algorithm has been used to update the behaviour parameters lambda and kappa. We want to compare the exact method with a fixed value of kappa and our new idea with kappa as a parameter; all parameters are the same for the two methods, except for the lambda prior. We want to obtain similar results between the fixed and variable kappa approaches because statistically, the only difference between the two is the prior for the transition rates and the way of updating the behaviour parameters. However, the fixed and variable kappa approach have different algorithms to update the behaviour parameters—Gibbs sampling and Metropolis–Hastings, respectively—which might mean that one does a better job exploring the behaviour parameter space than the other.

The important point when comparing the two methods is that the prior used for the transition rate should be similar in each case. Currently, the fixed kappa (the exact method) uses a Beta distribution as the prior for the transition rate in the adaptive model; in the homogenous model, a Dirichlet distribution is used as the prior for the transition rate (details in Section 3.4). However, in the current approach, when kappa is a parameter, using gamma as the marginal prior for the transition rates is natural; therefore, the gamma prior distribution should match the beta distribution in the adaptive model (Dirichlet distribution in the homogenous model). The transition rate prior in the current approach with kappa as a parameter is designed to match the prior in the fixed kappa case (the exact method), with the gamma distribution mean being equal to the beta distribution (Dirichlet distribution) mean. They are not exactly the same, but they look close enough.

For example, the transition rate prior for the current approach follows:

$$\begin{aligned} X &\sim \text{Gamma}(\text{Shape} = \alpha, \text{rate} = \beta). \\ E[X] &= \frac{\alpha}{\beta}, \qquad V[X] = \frac{\alpha}{\beta^2}. \end{aligned} \tag{5.20}$$

while, the transition rate prior for the fixed case (adaptive model) follow:

$$\begin{aligned} &Y \sim \text{Beta Distribution}(\eta, \zeta) \\ &E[Y] = \frac{\eta}{\eta+\zeta}, \qquad V[Y] = \frac{\eta\zeta}{(\eta+\zeta)^2()\eta+\zeta+1)}. \end{aligned} \tag{5.21}$$

The mean for the current prior should be equal to the mean of the prior for the fixed case:

$$\begin{aligned} E[X] &= E[Y] \\ \frac{\alpha}{\beta} &= E[Y] \\ \alpha &= E[Y]\beta. \end{aligned} \tag{5.22}$$

Furthermore, the variance:

$$\begin{aligned} V[X] &= V[Y] \\ \frac{\alpha}{\beta^2} &= V[Y] \\ \alpha &= V[Y]\beta^2. \end{aligned} \tag{5.23}$$

Then, the Equations 5.22 and 5.23 can be solved numerically. We can follow the same process when we have a homogenous model.

Choosing the kappa prior is different; therefore, the kappa prior can be chosen in any way that makes sense. For simplicity and for ease in combining the two densities, we choose $g(\kappa)$ as an exponential density:

$$g(\kappa) \propto exp(-\kappa\theta). \tag{5.24}$$

where θ is the parameter of the exponential. If the λ_{ij}s have independent Gamma priors, then,

$$f(\underset{\sim}{\lambda}) = \frac{p(\underset{\sim}{\lambda})}{H(\underset{\sim}{\lambda})}. \tag{5.25}$$

where,

$$H(\underset{\sim}{\lambda}) = \int_{\kappa=\max_{i,x}\lambda_i(x)}^{\infty} g(\kappa)d\kappa. \tag{5.26}$$

and

$$p(\underset{\sim}{\lambda}) \propto \Pi_{i\neq j} \quad \lambda_{ij}^{\alpha-1} exp(-\beta\lambda_{ij}). \tag{5.27}$$

where α and β are the parameters for gamma distribution. We divided $p(\underset{\sim}{\lambda})$ by $H(\underset{\sim}{\lambda})$ because we want the marginal prior gamma for lambda and the marginal prior exponential for kappa, as specifying the joint prior is difficult. When $f(\underset{\sim}{\lambda})$ and $g(\kappa)$ are combined, we still get the required form for this lambda's marginal prior. In practice, the ecologist will give a prior for lambda $p(\underset{\sim}{\lambda})$. From there, we can work out some combination for $f(\underset{\sim}{\lambda})$ and $g(\kappa)$ that is consistent with the prior. There are several ways to choose $f(\underset{\sim}{\lambda})$ and $g(\kappa)$, but this was an attempt to make it simple and give us something tractable for this calculation.

5.4 Simulation experiments

We started with the simulated data to check how well our idea of estimating the parameters performed and whether the model converged. We wanted to keep the algorithm as efficient as possible. We did other runs with different starting values, and we also did some runs with different seeds and checked the convergence, as described in Section 3.5.

5.4.1 Simulated data

The simulated data (similar to those decribed in Section 4.4.1) are from a homogeneous model which has one region and three behavioural states, with the following movement parameters: the three states share the same centre of attraction $\mu_i = (0,0)$. Furthermore, the three states have the same rate of attraction $b_i = 0.1$, where $i = 1,2,3$ and $v_1 = 0.05$, $v_2 = 0.5$, $v_3 = 5$. The sample size was 300 observations with one time unit interval between observations. The generator matrix for the transition rates is as follows:

$$G = \begin{pmatrix} -0.1 & 0.04 & 0.06 \\ 0.025 & -0.05 & 0.025 \\ 0.2 & 0 & -0.2 \end{pmatrix}. \tag{5.28}$$

5.4.2 The model

We fitted the data twice: one with a fixed value of kappa and again with the variable kappa approach. In both versions, we did the same things: all initial values of the transition rates were equal to one another ($\lambda_{ij} = 0.15$), initial states were set using as an HMM run, as described in Section 3.5, priors and the proposal distribution for all the movement parameters are normally distrbuted centered on the current value with a standard deviation of 2, as described in Section 3.4, and both approaches run for 10 million iterations using a thinning ratio of 1000 and a burn-in of 5000 iterations.

In the fixed (exact method) approach, the fixed value of κ is 0.4, and the prior for the transition rate is Dirichlet (α = 1); in the variable kappa approach, the marginal prior for κ has an exponential distribution with ($\theta = 3$) , the marginal distribution for each transition rate has a Gamma distribution with ($\alpha = 2$, $\beta = 0.1$), the proposal distribution for κ is truncated normal centred on the current value and the proposal SD is 0.4, and the proposal distribution for λ is truncated normal centered on the current value and the proposal SD is 0.1. The values were chosen to obtain approximately optimal acceptance rates. The acceptance rate for the transition rate can be found in the result section.

The generator matrix for the homogeneous model was the same as that in Section 3.4:

$$G = \begin{pmatrix} -(\lambda_{12}+\lambda_{13}) & \lambda_{12} & \lambda_{13} \\ \lambda_{13} & -(\lambda_{21}+\lambda_{23}) & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & -(\lambda_{31}+\lambda_{32}) \end{pmatrix} \tag{5.29}$$

5.4.3 Results

	real value	Fixed	Variable
μ_x	0	0.009(0.089)	0.014(0.086)
μ_y	0	0.045(0.082)	0.042(0.084)
b_1	0.1	0.1(0.006)	0.1(0.006)
b_2	0.1	0.12(0.05)	0.14(0.05)
b_3	0.1	0.1(0.04)	0.097(0.043)
v_1	0.05	0.046(0.005)	0.046(0.005)
v_2	0.5	0.55(0.33)	0.49(0.24)
v_3	5	6.7(3.66)	8.21(6.92)
λ_{12}	0.04	0.059(0.024)	0.057(0.045)
λ_{13}	0.06	0.06(0.026)	0.08(0.06)
λ_{21}	0.025	0.058(0.03)	0.034(0.019)
λ_{23}	0.025	0.07(0.037)	0.038(0.02)
λ_{31}	0.2	0.11(0.04)	0.1(0.06)
λ_{32}	0	0.03(0.03)	0.075(0.057)
κ	-	-	0.26(0.11)
running time	-	1h 55 min	1h 35 min

Table 5.1 Posterior means and SDs for the parameters obtained using both approaches for the simulated data.

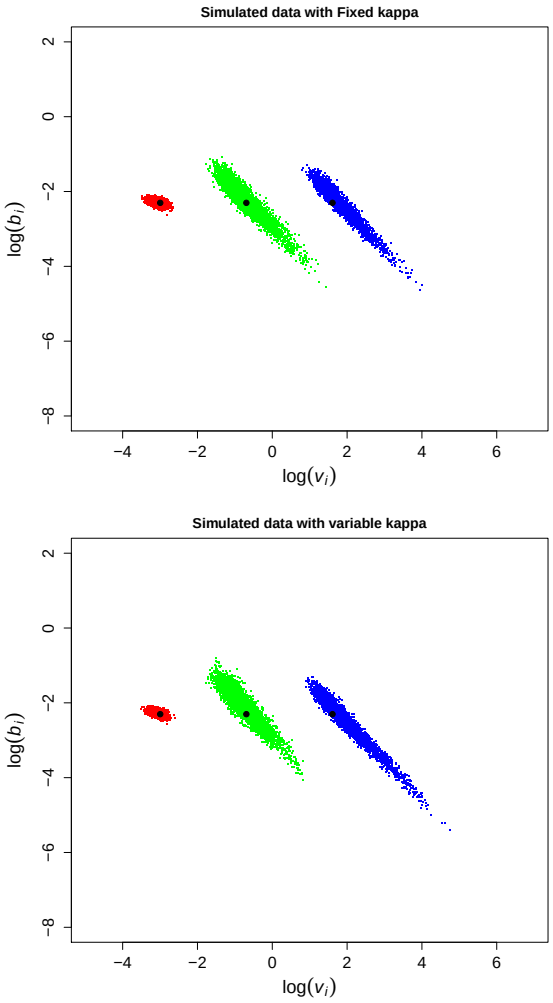

Fig. 5.1 Posterior distributions for the log movement parameters of all the behavioural states with the simulated data. Parameters for states 1, 2 and 3 can be seen in red, green and blue, respectively. The real values are displayed as black dots.

Figure 5.1 shows that both the fixed and variable kappa approaches recognised three different states and made precise estimates for the movement parameters. Furthermore, Table 5.1 shows that both approaches reconstructed the state and estimated the movement rate accurately. From Figure 5.1 and Table 5.1, we can see how well we estimated the movement parameters. The model precisely estimated the parameters for state 1, whereas for state 2 and 3, the posterior mean for the movement parameters is very close to the real value but has a high SD, especially for state 3 .

Figure 5.3 and Figure 5.4 compare the posterior results for the transition rate values from the variable kappa approach with those from the fixed kappa approach. On the his-

togram plot for the posterior results, the vertical line is the true value. The fixed kappa approach estimates the posterior of the transition rate for state one more precisely and accurately than the variable kappa approach does; by contrast, the variable kappa approach gives a more precise estimate for the transition rate for state two than the fixed kappa approach does. However, both the fixed and variable kappa approaches have a similar estimation for some of the transition rate parameters.

For the variable kappa approach, the acceptance rate for the behaviour parameters; κ is 0.18, that for λ_{1j} is 0.28, that for λ_{2j} is 0.17 and for that λ_{3j} is 0.29. The results here are for a run carried out after some pilot runs to set the proposal standard deviations in order to give reasonable acceptance rates.

Figure 5.2 shows the true trajectory of behaviour and the reconstruction in the vari-

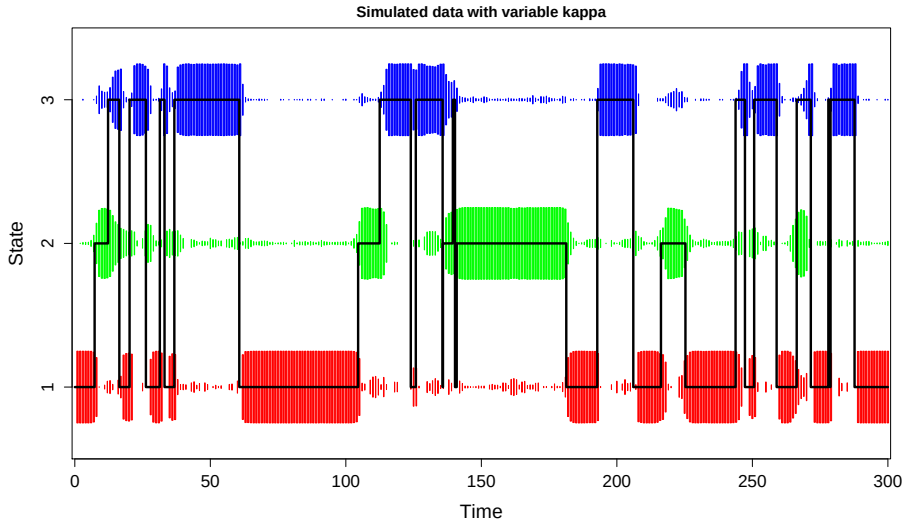

Fig. 5.2 For the variable kappa approach, the trajectory for each observation to follow state 1, 2 or 3.

able kappa case, which is very close to the true one. The black curve represents the true trajectory states, and the coloured line represents our reconstruction of the trajectory with the area representing the posterior probability of a state at each time. I do not show the plot for the fixed kappa approach because it is very similar to the variable kappa approach.

The variable kappa approach took only 1 hour and 35 minutes; whereas the fixed kappa approach took 1 hour and 55 minutes. The variable kappa approach was faster than the fixed kappa approach by just 20 minutes or about 13%.

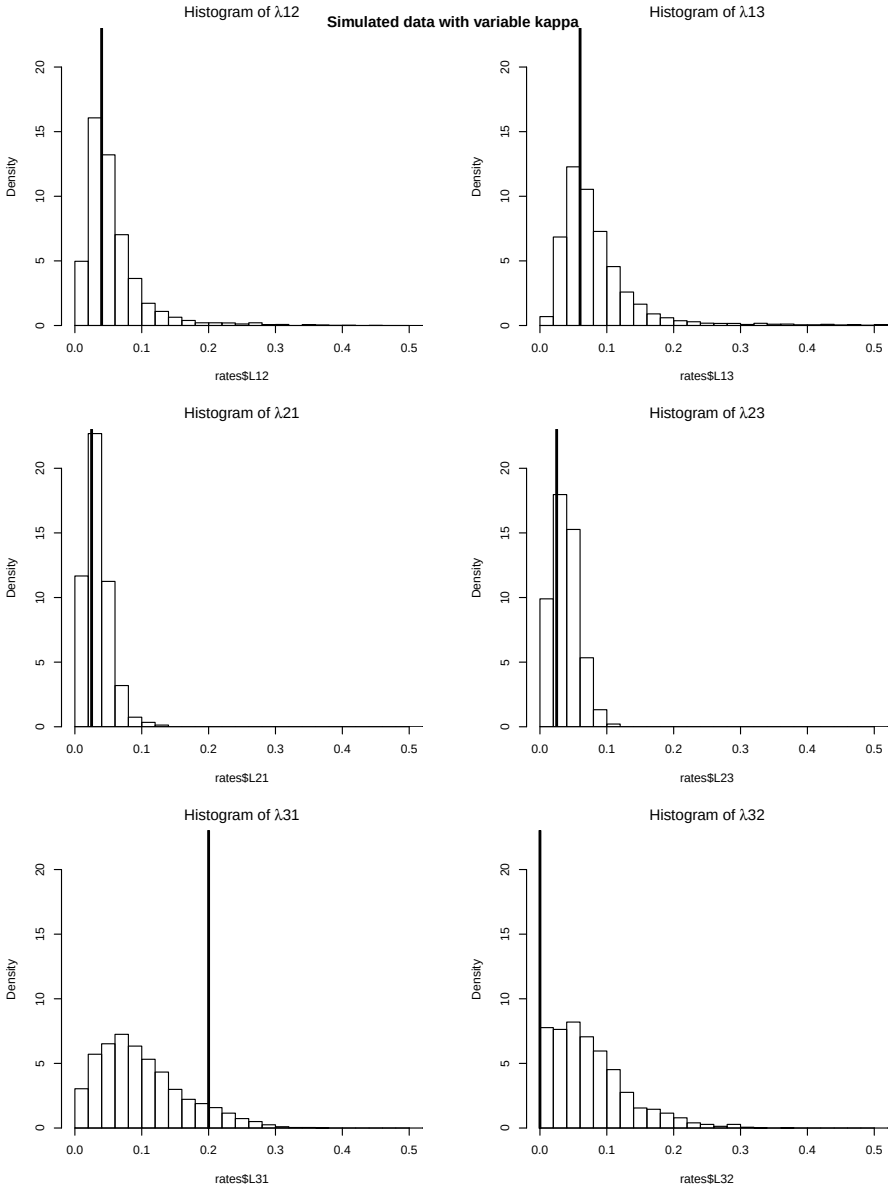

Fig. 5.3 For the variable kappa approach, the posterior results for the transition rate of the three behavioural states. The true values are displayed as vertical black lines.

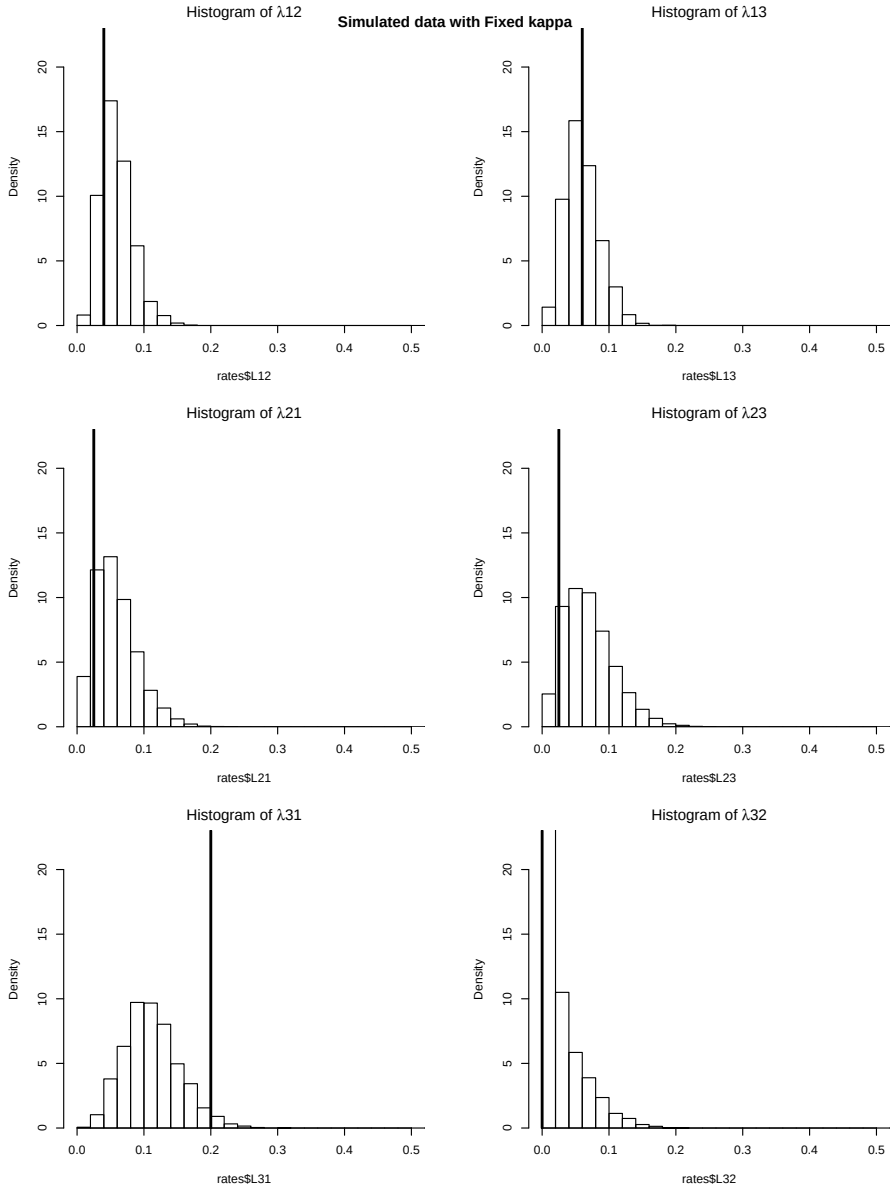

Fig. 5.4 For the fixed kappa approach, the posterior results for the transition rate of the three behavioural states. The true values are displayed as vertical black lines.

5.5 Application

In this section, I apply the the two approaches to a real data example with an adaptive model.

5.5.1 Fisher data

The data were collected using a GPS collar attached to a male fisher. A fisher (scientific name Pekania pennanti) is a medium-size terrestrial mammal living in North America. Fisher locations were recorded every 10 minutes; however, there are some missing data, which result in a longer interval between observations. These are the same data as those analysed by Blackwell et al. (2016). The data are massive, but Blackwell et al. (2016) choose one subset of the original dataset; the subset size is 128 observations for a 24 hour period. The reason for choosing one subset is to obtain a subset in which the animal is attracted to the same centre of attraction for the entire subset length of 24 hours (LaPoint et al. 2013a, b). The map for the study area is available with the known boundary for each habitat region; there are three different regions. The study area is divided into three different habitats: developed open space, forest and wetland. The map information comes from the US National Land Cover Database (NLCD 2006 Land Cover layer; Fry et al. 2011). We used the same subset as Blackwell et al. (2016) because we want to compare our results from the variable kappa approach with those from the fixed kappa approach.

5.5.2 The model

The analysis of the fisher data provides an example of an adaptive model; in this model, the transition rate between behaviours depends on the location. In this particular case, the way the behaviours work is that the animal can only switch to the behaviour that matches its habitat. The way that the adaptive model works is if the animal enters a region, the animal will change to a behaviour that matches the region quite quickly, but not immediately (for details, see Section 3.4.2)

The study area has 3 different habitat regions, which means that the model will have three different behavioural states, and each one of these is connected to one region. We therefore model the data by using three different OU processes that each has an associated behavioural state. The three OU processes share the same centre of attraction μ, with a different value for each variation v_i and b_i rate of attraction. The behaviour process depends on the animal location (habitat region), so we had three G_i generator matrices; each one is associated with one region. Therefore, the model has only six non-zero transition rates. The generator for

region one is as follows:

$$G_1 = \begin{pmatrix} 0 & 0 & 0 \\ \lambda_{21} & -\lambda_{21} & 0 \\ \lambda_{31} & 0 & -\lambda_{31} \end{pmatrix}. \tag{5.30}$$

The generator for region two is as follows:

$$G_2 = \begin{pmatrix} -\lambda_{12} & \lambda_{12} & 0 \\ 0 & 0 & 0 \\ 0 & \lambda_{32} & -\lambda_{32} \end{pmatrix}. \tag{5.31}$$

The generator for region three is as follows:

$$G_3 = \begin{pmatrix} -\lambda_{13} & 0 & \lambda_{13} \\ 0 & -\lambda_{23} & \lambda_{23} \\ 0 & 0 & 0 \end{pmatrix}. \tag{5.32}$$

I fitted both approaches to the data, first with fixed kappa and then with variable kappa, as described in Section 5.2.2. The two approaches share the same updating for the movement parameter μ, b_i, v_i, the prior for μ is bivariate normal distribution and the prior for b_i and v_i is truncated normal; all the three states have the same prior. The proposal distribution for all the movement parameters is a normal distribution centred on the current value with $SD = 2$. The details for the movement parameter prior and proposal are as described in Sections 3.4 and 3.5. The movement parameters are updated using the Metropolis-Hastings algorthim. Furthermore, the two approaches update the trajectory in the same way (more details in Sections 3.4 and 3.5). Both approaches run for 10 million iterations with a 5000 burn-in. Both approaches have the same intial value for the movement parameters and behaviour parameters.

The only difference between the two approahes is updating the behaviour parameters. In the fixed kappa approach, kappa is fixed at a value of 4, and the marginal prior for the transition rate has a truncated beta (12, 4). The behaviour parameters are updated using Gibbs sampling. By contrast, in the variable kappa approach, the kappa marginal prior is exponential with a rate of 4, and the proposal distribution for kappa is truncated normal centered on the current value with SD $= 0.2$. The marginal prior for each of the transition rates has a gamma distribution with parameters 51.14 and 17.04, the proposal distribution

for the transition rate is truncated normal centered on the current value with sd 0.3. The behaviour parameters are updated using the Metropolis-Hastings algorithm.

5.5.3 Results

Figure 5.5 shows that both algorithms gave results for the movement parameters that were very close to each other. For estimating the movement parameters, both the fixed and variable kappa experiments gave results that were close to each other as expected. Both approaches have nearly identical estimations for the centre of attraction μ; both also have a very similar estimation for the movement parameters of state one. However, there are some slight differences in estimating the movement parameters for states 2 and 3 between the two approaches; the details are shown in Table 5.2. The problem is that these particular habitats (regions two and three) do not have enough visits, so we cannot learn much about the estimates of their movement parameters.

Figure 5.6 shows that the fisher spends the most time in state one, with few visits to state three and limited visits to state two. There were not many switches or clear changes in the states, so the data do not give us much information about the states 2 and 3. One of the limitations of the fisher dataset was that the results for states two and three varied considerably because there was not much information about these two regions. This made it difficult to estimate state three parameters and also affected state two. Blackwell et al. (2016) mentioned in their results that the parameters for state two are not well estimated. Blackwell et al. (2016) explained that this issue happened because much of the data from the subset are in region one, meaning that the animal spends the most time in behaviour one and state one, which results in an uncertain estimate for the parameters of states two and three. When comparing the fixed and variable kappa for the states over time, Figure 5.6 shows that the animal spends most of the time in state one, with few visits to state three, but the probability of being in state two remains very low. We can conclude that we obtained similar results as Blackwell et al (2016) did and that the poor estimation for some of the parameters was related to the dataset and not to the approach used. Furthermore, we have a more efficient approach because of the saving in the computational cost.

Table 5.2 shows the posterior mean and standard deviation for the parameters in both the fixed kappa and variable kappa cases. The values for the posterior of the transition rate are not identical in different experiments, but they are quite close. Blackwell et al. (2016) used fixed kappa with a particular prior for their fisher results and did not allow lambda to go below two or above 4 in the beta distribution. Blackwell et al. (2016) had a restriction

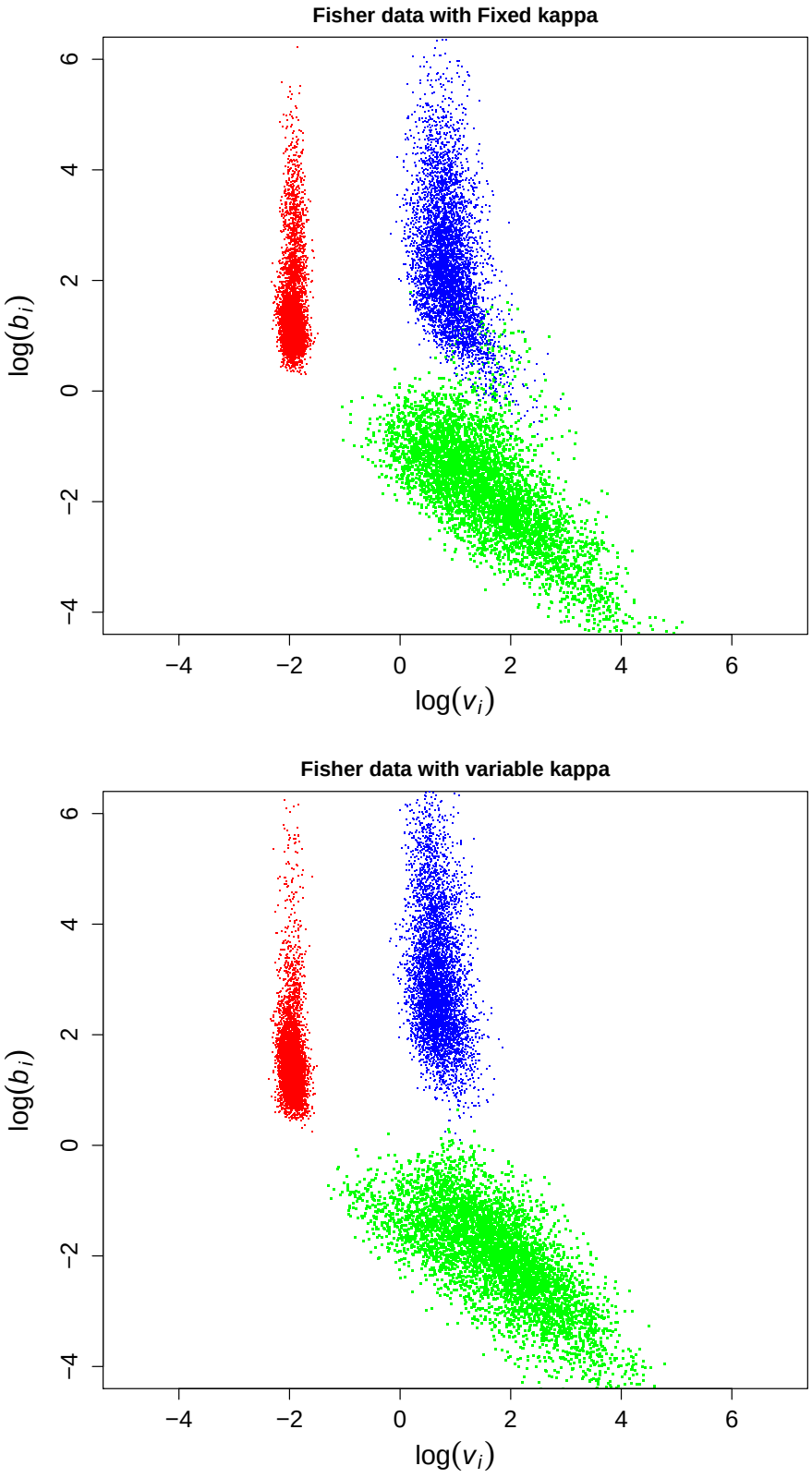

Fig. 5.5 Posterior distributions for the log movement parameters of all the behavioural states for the fisher data . Parameters for states 1, 2 and 3 can be seen in red, green and blue, respectively.

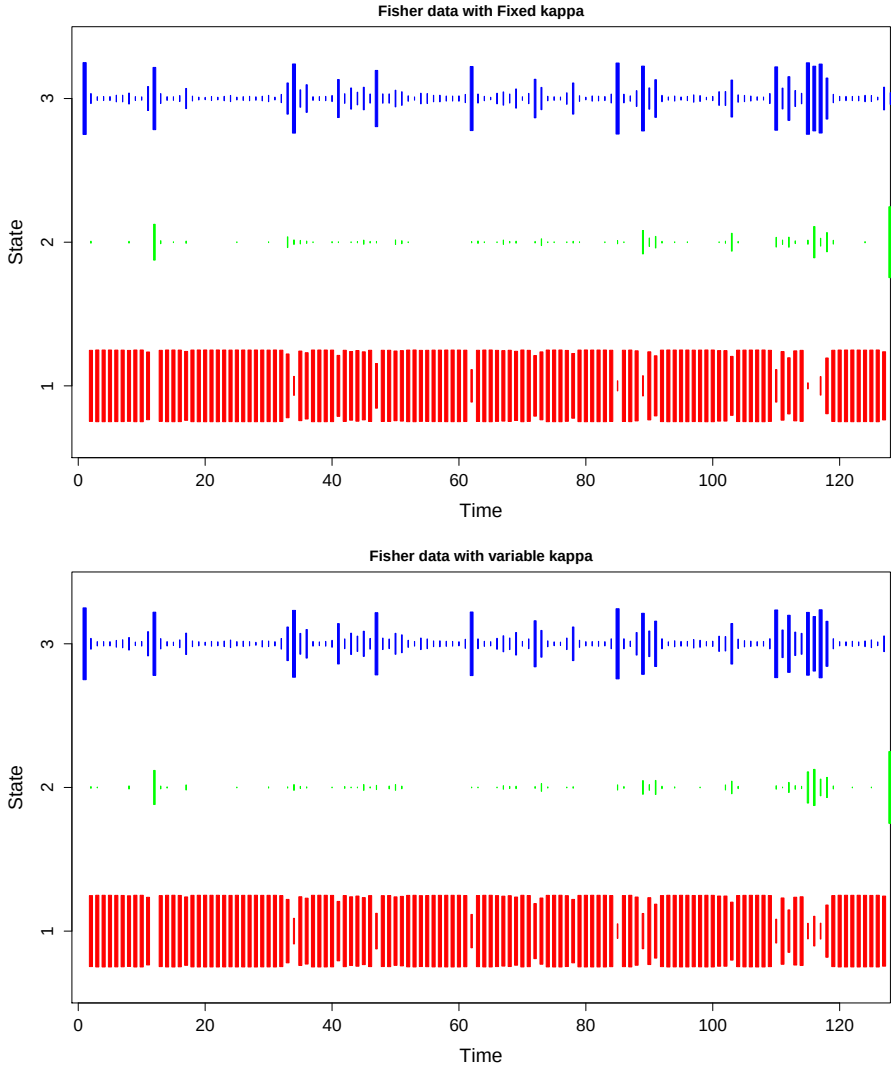

Fig. 5.6 Trajectory for each observation to follow states one, two or three, for the fisher data.

on the lambda value because they want that when the animal enters a habitat, it quickly switches its behaviour to the state that matches the habitat. However, we allowed lambda to go below two and above 4; we still obtain similar results in terms of the posterior distribution for the transition rate, with not much posterior probability outside the interval (2.0,4.0).

First, we applied the fixed kappa model to the fisher data; it took six hours with a fixed

	Fixed	Variable
μ_x	9.02(0.04)	9.02(0.04)
μ_y	3.45(0.043)	3.46(0.04)
b_1	7.94(17.52)	9.1(20.79)
b_2	0.2975(0.51)	0.188(0.16)
b_3	23.56(50.83)	51.122(130.146)
v_1	0.14(0.016)	0.14(0.017)
v_2	8.33(13.01)	9.53(12.11)
v_3	2.58(1.12)	2.04(0.59)
λ_{12}	3.02 (0.39)	2.82 (0.39)
λ_{13}	3.21 (0.31)	2.91 (0.45)
λ_{21}	2.97 (0.39)	2.94 (0.4)
λ_{23}	3.03 (0.39)	2.94 (0.41)
λ_{31}	3.1 (0.34)	3.23 (0.45)
λ_{32}	2.98 (0.38)	2.96 (0.41)
κ	-	3.79 (0.42)
running time	6h	5h 35 min

Table 5.2 Posterior means and SDs for the parameters of the fisher data in both approaches.

$\kappa = 4$. Then, I ran it with variable kappa; it took 5 hours and 35 minutes. The fixed kappa run was very slow because the kappa value affects the time it takes to run. If the kappa goes up, it generates more switches, so it runs more slowly. Details of the reduction in κ are given below in Table 5.2. The variable kappa approach reduced the computation time by 11%.

5.6 Discussion

The idea developed in this chapter is to use a variable kappa approach, which allows kappa to simultaneously update with the behavioural transition rate rather than having a fixed value upper bound on the transition rates. Doing this helped us make more efficient estimates and saves time. Kappa can be nearly equivalent to the current transition rates in the model to provide more efficiency, thereby decreasing the number of potential switches and the

computational costs (Blackwell et al. 2016). A variable kappa idea also provides an efficient algorithm and methodology to investigate how well the method can mix. We used the idea with two datasets; the fisher data (adaptive model) and homogenous simulated data, to test how well the estimate performs. We also developed and tested other methods to calculate the kappa.

In this chapter, we tried to improve the existing algorithm. The numerical results not only show us that variable kappa is better in terms of estimation and efficiency but also demonstrate that the benefit is in terms of using more realistic priors for lambda, as it is difficult to give a fixed value of kappa and ensure that this fixed value is reasonable. We wanted to put on prior with a very high lambda and see what the data say. Scientists do not need to worry about the fixed value for kappa; they just need to think about the prior for the lambdas. The other benefit is in terms of computational cost; in the real data examples from Blackwell et al. (2016), the variable kappa approach reduced the computation time by 11%.

The use of a variable kappa does not involve much extra computation and gives us the flexibility to use a general form of prior. In the next chapters, we will try to extend the model's range; we want to have an efficient algorithm for kappa in order to enable us to draw inferences about the boundary and in the semi-Markov case.

Chapter 6

Estimating an unknown boundary

6.1 Introduction

In this chapter, we try to extend the Blackwell et al. (2016) movement model to estimate the unknown boundaries of the animal habitat regions. In some movement studies, the boundary between different habitats is known, as we see in the fisher data in Section 5.5.1. The boundary between habitats is usually represented by a map for a geographical habitat with different zones, such as: forest, wetland and grassland etc. However, in this chapter, the boundary may represent an unknown boundary between different habitat types or may be more of a mental boundary for the animal, in which the animal thinks it is within its own territory or home range.

In ecological landscape studies, understanding current and future habitat fragmentation plays a significant role in managing and maintaining the ecosystem and the species in the specific area. Boutin and Hebert (2002) defined 'habitat fragmentation' as the process of dividing the landscape into small separate patches; this breakup could be result from human action, such as building urban areas, transferring forest to agriculture or constructing reservoirs, highways and so on. Alternatively, habitat fragmentation could be result from any geological action that may affect the geographic environment. Habitat fragmentation affects the species of the original landscape, which could result in their migration and extinction, or increases in some kinds of species compared to others who lived in this area. Ecologists study the new small patches to understand and predict new mixtures of species. Different sizes, shapes and boundaries of the patch affect the species' populations and their communities as well as the ecosystem in general (see Boutin and Hebert 2002).

Methods for estimating the home range that are described in Section 1.5 are similar to

the idea of estimating the boundary in this chapter. Previously published papers studied how the animal utilises the habitats and estimates the home range from the movement data. However, the authors of these studies were interested in the utilisation distribution and which area is most used, not the animal behaviour inside and outside the home range. For more details about the previous home range method, see Section 1.5. The estimate of the mental boundary of the animal is similar to the estimate of its home range; however, in an early study by Dunn and Gipson (1977), they did not examine different behaviours. They only fit the utilisation distribution as a bivariate normal, the contour of which was circular. One way of thinking about estimating the boundary as a home range model is that, if the animal changes its behaviour when it is more than a given distance outside the circle, the shape of the utilisation distribution could be changed. In Dunn and Gipson's model, there is a normal distribution with one behavioural state. When the animal is allowed to behave differently in the area outside the boundary, then the new movement behaviour pushes the animal back toward the centre. In other words, when the animal is outside the boundary, its behaviour changes, and it returns quickly to the centre. This boundary represents the home range, and the behaviour differs inside and outside that region, which helps us to generate a general behaviour model. By placing these two behaviours in two different regions, we have a more flexible idea of the appearance of a home range, how the animal behaves regarding its habitat and how the utilisation decreases as the animal moves away. Dunn and Gipson (1977) identified the problem early, and this chapter is our attempt to solve it.

In this approach, the number of regions in the environment can be known or unknown. For example, in cases where the number of regions is known, the boundaries between these regions need to be estimated, including how the animal utilises each region. However, in other cases, the number of regions inside the habitat is unknown because the terrain is hard to measure or observe or the animal has a mental map of its home range or territory. The reason for it being a distinct region might be purely mental or might be resources of food or water, which the scientists are unaware. Therefore, the model described in this chapter could be used to estimate some features of the environment regarding the animal's movement.

Different habitats have distinct effects on animals' behaviour and movements. Tishkovskaya and Blackwell (In prep.) used telemetry location data to improve a continuous time model for analysing animal movements in a heterogeneous environment. The aim of this chapter is to use the same idea as Tishkovskaya and Blackwell (In prep.), but with the Harris and Blackwell (2013) separable model (as discussed in Section 3.3.1) where the Blackwell et al. (2016) exact method can be used. The main target is to use location data to learn more about

the environment and how different habitats affect the animal movement. The variable kappa approach has been used within this chapter to estimate the boundary (described in Chapter 5). In this chapter, we try to estimate two types of boundaries: the circle/disc boundary where one behaviour is inside a circle and the other behaviour is outside it, and linear boundary. The new method has been tried with two different datasets: simulated data and ibex data. These can be seen in the next sections.

6.2 The model

Tishkovskaya and Blackwell (In prep.) introduced habitat heterogeneity by assuming that the habitat is divided into a fixed number of non-intersecting regions and that animals move differently in each region. Tishkovskaya and Blackwell (In prep.) assumed that animals have one behavioural state for each region. Therefore, each region has a different OU process. Tishkovskaya and Blackwell (In prep.) explained that, if two observed locations x_t and x_{t-1} are in separate regions, then on crossing the boundary the conditional distribution of $P(x_t|x_{t-1})$ is a weighted mixture of two OU processes. The weights could be a function of the location, but in practice they use equal weights. The authors introduced the new idea by considering a simple case in which the studied landscape consists of two regions: region one, where the animal spends most of its time, with a circular boundary, and region two, which is outside the circle and has less use. Region one, which is inside the circle, can be explained as the animal's territory, the home range; it could have a particular biological purpose. The two regions share the same centre of attraction, which means that, they share the common centre: $\mu_1 = \mu_2$. Tishkovskaya and Blackwell (In prep.) assumed that b_1 is larger than b_2. The b value means attraction towards the centre, as in Section 3.1. Note that values of b are negative, so the smaller (more negative) value represents stronger attraction. Accordingly, when the animal is farther away from the centre of attraction, it moves faster toward the centre of attraction. Thus, when the animal crosses the boundary from region one into region two, it will have a stronger attraction to return. Consequently, even though the animal can travel outside the circle boundary, it has pressure to return back to region one.

We will use a concept similar to Tishkovskaya and Blackwell's (In prep.) to extend the Blackwell et.al's (2016) model, to estimate the unknown boundaries between different regions. The unknown boundaries are estimated by dividing the environment into regions, and each region has only one behavioural state. Therefore, the number of regions is equals to the number of behavioural states. We assume the habitat is divided into two regions and each region is associated with one behavioural state, which is an adaptive model (for more details

see Section 3.6.2). Furthermore, a change in behaviour means that the animal has crossed the boundary and entered the next region. However, as discussed in Section 3.6, in the exact method, the transition rate has an upper bound κ, which means the animal does not change its behaviour immediately after crossing the boundary.

In this chapter, we estimate two different types of boundaries:

6.2.1 Circular boundary

The first idea is a circular boundary. In a situation where the animal spends most of its time inside the circle and only occasionally leaves it, this model will be more suitable. Several ecological home range studies have concluded that the shape of an animal's home range is a circle. Gurnell (1984) explained that the red squirrel's home range is circular; moreover, Russell et al. (2012) showed that the lion's home range is circular. Harris and Blackwell (2013) used two simulated datasets with a circular home range.

6.2.2 Linear boundary

We tried to generalise the boundary estimation idea. Because having a circular region boundary was rather restrictive, a more general idea for the boundary was to use a straight line. The straight line boundary is the simplest boundary, but it is also quite realistic, because it can be seen as a local approximation to any smooth boundary. We wanted to keep the structure simple. One of the possibilities was merely to create a straight line boundary so the space was cut into two regions, as part of the realistic example. In that case, we did not know where the boundary was, and we wanted to learn the boundary's location from the movement. Two regions were separated by a linear boundary. The boundary line could be oriented in any direction because it represented the edge of a forest, field, territory or home range etc. The linear boundary can be represented as a linear equation as:

$$y = ax + b. \tag{6.1}$$

where a and b are some numerical values for the intercept and slope. The linear boundary can be generalized (given suitable data) to a smooth nonlinear boundary.

6.3 Implementation

We used Bayesian inference to estimate the parameters in the model, similar to Chapter 5, except with the addition of an extra MCMC step to estimate the boundary parameters. Updating the movement parameters μ, v, b and the trajectory uses a similar approach to that used in Section 3.6. In addition, we updated the behaviour parameters $\lambda_{ij}(x)$ similar to the adaptive model in Section 3.6.2 and treating kappa as a parameter in Section 5.2.2.

To update the boundary parameters, we use a Metropolis Hastings step that proposes new parameters for the boundary. The idea of estimating an unknown boundary, is implemented by adding an extra MCMC step to update the parameters to describe the current boundary. We propose the new boundary, and after each potential switch, we calculate the new probability of making that transition and calculate the Hastings ratio from that. Subsequently, we accept or reject the proposed boundary based on the Hastings ratio.

6.3.1 Circular boundary

To estimate the boundary, we add two more parameters to the current model. The boundary parameters are the centre point (x_0, y_0) and radius R. Tishkovskaya and Blackwell (In prep.) clarified the circle border radius R as the maximum area of attraction, around a food source, nest site, etc. Using the centre point, radius and an observed location, (x, y), we can test whether the particular (x, y) location point is outside or inside the new circle boundary. The next step is to update the boundary by proposing a new value for the centre point (x_0, y_0) and the radius R, and then deciding if those values are either accepted or rejected by using the Metropolis Hastings algorithm. Hence, when the boundary is updated, any of the points may change from being in one region to being in the other; sometimes, that means the proposed boundary is not possible because the sequence of behaviours does not make sense, given the region. We have an adaptive model, which means the animal's behavioural states should only change to be the same as the region where the animal is. For example, if the animal is inside the circle's boundary, it is in behavioural state one, which is settled. The animal cannot switch to behavioural state two, which is exploring, until it leaves the circle region. The model allows the animal to switch to behavioural state two after crossing the boundary. The transition from behaviourally settled to behaviourally exploring while the animal is still in region one inside the boundary is impossible, which would lead to a rejection of the newly proposed boundary.

6.3.2 Linear boundary

The line boundary idea will have the same implementation as the circular boundary, except it replaces the radius and centre in the circle boundary by the intercept and slope a and b, as the linear boundary parameters for Equation 6.1.

6.4 Model checking

To check our results, we used the deviance information criterion (DIC) to compare a two-state model with a one-state model. The DIC, defined below, is a very useful tool for model selection in Bayesian inference where the posterior distribution for the parameters is estimated using MCMC. The DIC is defined as:

$$DIC = p_v + \overline{D}. \tag{6.2}$$

The deviance is based on the log-likelihood for a particular parameter value of θ:

$$D(\theta) = -2\log[P(y|\theta)]. \tag{6.3}$$

where θ is a vector of the unknown parameters that would be estimated. The deviance mean is given by:

$$\overline{D} = E[D(\theta)], \tag{6.4}$$

averaging over the posterior distribution for θ. The mean measures how well the model fits the data: the larger the value, the worse the fit. Of course, when the number of the parameters in the model increases, the deviance mean will decrease. Therefore, the DIC included a term related to the number of parameters. The effective number of the parameters for the model can be estimated as:

$$p_v = \frac{1}{2}Var[D(\theta)]. \tag{6.5}$$

The model with the smallest value for the DIC is the best, considering the trade-off between the fit and the complexity of the model. For more details see, for example: Spiegelhalter et al. (2002) and Spiegelhalter et al. (2014).

6.4.1 Approximating the DIC

In this chapter, the DIC is approximated similarly to both Blackwell et al. (2016) and Tishkovskaya and Blackwell (In prep.). Calculating the DIC for the one state model is straightforward however the calculation for the two-state model is not completely straightforward. The problem with the two-state model is that there are switches in behaviours and the path between observations is unknown. Even the numbers of potential switches in the two-state model is unknown, so calculating the probability of going from one observed point to another is quite complicated. Blackwell et al. (2016) and Tishkovskaya and Blackwell (In prep.) solve this problem by making the simple assumption that the behaviour and trajectory are known, which means that the DIC approximation calculation concerns only the movement parameters. We assume that, in each observation, the behaviour is known and that it stays the same until the next observation. If we do not make this assumption, we allow the behaviour to change at different times during the interval. Then the likelihoods for the DIC calculation are not comparable between the two-state model and one-state model, because the distinctions in the fit will be due to variations in the reconstruction of the behaviour much more than to the differences in the number of states.

In addition, we considered which state we wanted to use, from the MCMC inference, we had a posterior distribution over possible states. We examined the output to decide the most likely state at that time. We merely recorded the state at the time of the observation, not the complete trajectory. The state is fixed based on the result of the MCMC, by taking the state with the highest posterior probability at the actual observation. When, within the DIC calculation, we took the average over the samples from the posterior distribution, we used the same most likely state in each case, combined with parameters from the posterior distribution to calculate the probability of going from this observation to the next one.

By approximating the DIC in this way, we are using only likelihood terms that directly relate to the observations. The parameters controlling the behaviour, the transition rate λ and upper bounded κ, do not have any effect. This is consistent with the effective number of parameters p_v, which approximately matches the number of movement parameters. Similarly, the likelihood terms regarding the variable boundary do not enter into the DIC. Therefore, this approach in itself does not allow us to compare the fixed and variable boundary models. However, we have no biological justification for fixing the boundary in advance.

Therefore, the approximate DIC enables us to make sensible comparisons between the movement parameters in models with varying numbers of states. We used this method

because the difference in DICs between the two models was very large, so the approximation did not change the model order. Although this method of calculation did have some effect on the values obtained for the DIC, it did not affect the conclusion when there was a distinction between DIC values for the two models at large. If the difference in the DIC between the two models had been small, however, we should be careful, because the approximate DIC may not correctly indicate the conclusion that exact DIC values (the more sophisticated calculation) would have reached about whether one model is better than the other model.

6.5 Application

6.5.1 Ibex data

The ibex data are telemetry data consisting of ibex (wild goat) locations for 15 days in the Belledonne Mountain (French Alps), as recorded using a GPS tracker. There were four ibexes, but we studied the movement of only one of them. A total of 71 observations were recorded, and ibex locations were updated every four hours. The data are available in an R package called adehabitat (Calenge et al. 2009). This is the same dataset used by Tishkovskaya and Blackwell (In prep.).

The actual numerical values for the (x,y) locations in the data are very large, which could potentially cause numerical problems in the analysis, particularly as the variation within the data is relatively small. To make the numbers easier to work with, we rescaled and shifted the origin to define a new set of co-ordinates, as in the following:

$$x_{\text{new}} = \frac{x_{\text{old}} - 9 \times 10^5}{1000} \tag{6.6}$$

$$y_{\text{new}} = \frac{y_{\text{old}} - 2 \times 10^6}{1000} \tag{6.7}$$

6.5.2 The model

In this example, we assumed a circular boundary between two regions, as in Tishkovskaya and Blackwell (In prep.) and as described in detail above in Section 6.2.1. Figure 6.1 shows the movement path for an ibex in the original coordinates, in which the blue triangle is the starting location. From this figure, we can see that the ibex has a strong attraction toward the centre area at the upper part of the figure. This core area can be represented as a foraging region or home range, with the ibex making shorter movements away from this area. A centre of attraction for the foraging region could be a food source or (in the context of the other

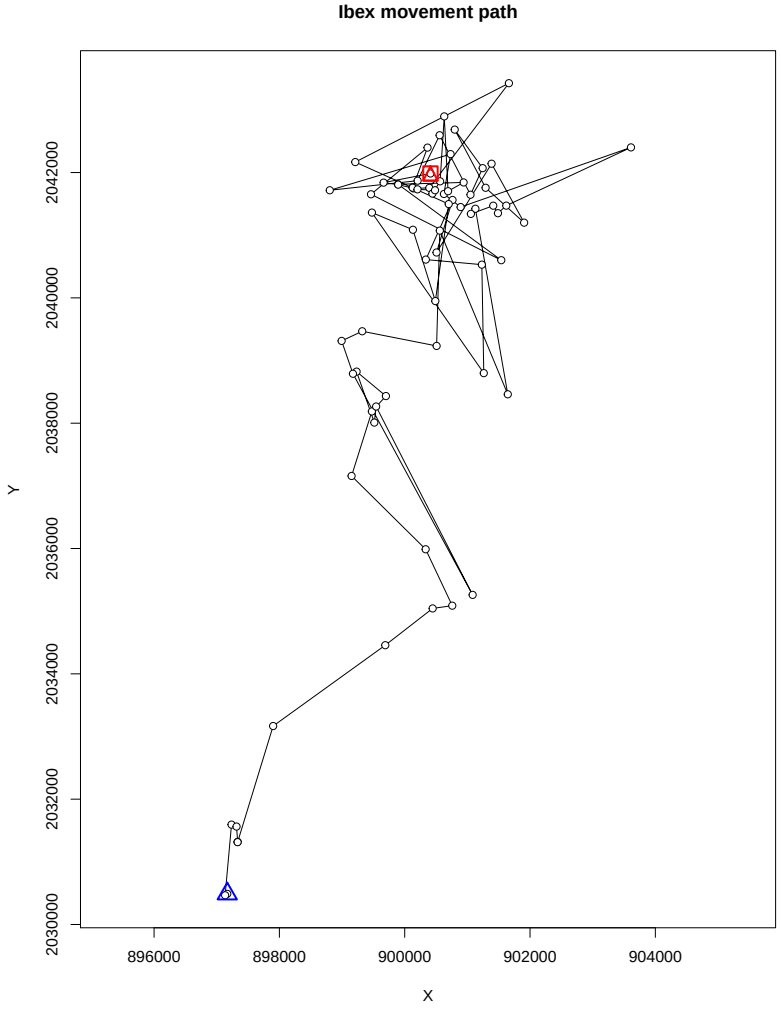

Fig. 6.1 The observed Ibex movement path. The blue triangle is the starting location, while the red is the end point.

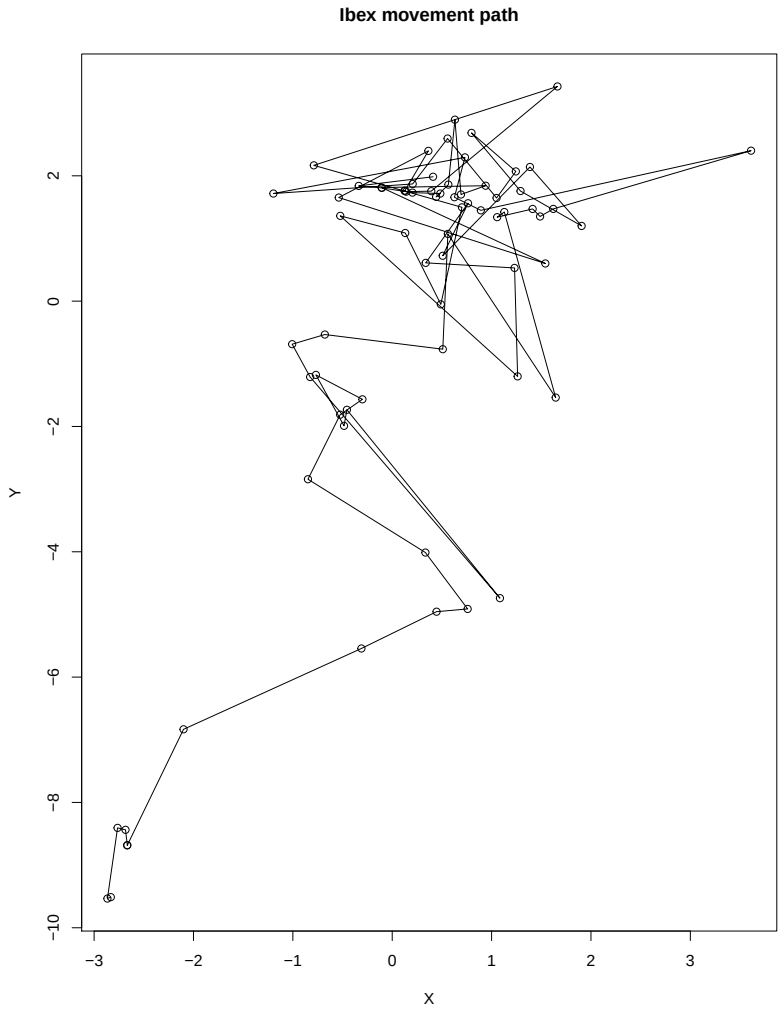

Fig. 6.2 The observed Ibex movement path after transformation of the coordinates.

species) a den or nest site. For more details about centre of attraction, see Section 1.3.1. The ibex's movements outside the boundary can be understood as exploration or incursions into other home ranges. Figure 6.2 shows the movement path for an ibex after transformation of the coordinates.

We fit the data with two OU processes, each one connected to a specific region; both OU processes share the same centre of attraction. The first OU process represents the exploratory movement (outside the boundary) from the initial location to the core area, whereas the second OU process describes the shorter animal movement inside the circular foraging area, which could be called resident or foraging movement.

6.5.3 Implementation

We updated four separate sets of parameters: the trajectory, the movement parameters, the behaviour parameters and the boundary parameters. The updating of the trajectory and movement parameters was carried out as described in Section 3.6. The movement parameters: The prior for μ_x was a normal distribution with mean 0 and sd 0.9. The prior for μ_y was a normal distribution with mean 40 and sd 0.9. The prior for b_1 was a normal distribution with mean 3 and sd 0.9. The prior for b_2 was a normal distribution with mean 0.4 and sd 0.9. The prior for v_1 was a normal distribution with mean 0.9 and sd 0.9. The prior for v_2 was a normal distribution with mean 5 and sd 0.9.

For updating the behaviour parameters, we used the variable kappa approach, where kappa was treated as a parameter, as described in Sections 5.2.2 and 5.3. The forms of the priors for kappa and the transition rates were the same as in the previous chapter (Chapter 5), the marginal prior for κ was an exponential distribution with rate 2 and the marginal prior for λ was a Gamma distribution with shape 2 and rate 6. The proposals for kappa and the transition rate were similar to those in the previous chapter. We ran the model with an initial value for all λ of around 0.5, while the initial value for κ was 1

To update the boundary parameters (radius and centre point), a truncated normal distribution was used as the prior for the radius with mean$=1.2$ and SD$=0.9$ truncated on $(0,\infty)$. The proposal for the radius was a truncated normal distribution centred on the current value, with SD 0.9. We choose a truncated normal distribution because the radius should always be positive. The truncated distribution resulted in a proposal that was no longer symmetrical, so we had to add forward and backward terms for the proposal density of the radius to the MCMC calculation for the Metropolis Hastings ratio, just as in the truncated proposals in

Section 5.2.2. The prior distribution for the centre point was normal distribution. The prior for y_0 was a normal distribution with mean 41 and sd 0.9. The prior for x_0 was a normal distribution with mean 0 and sd 0.9. The proposal distribution for the centre point was normal distribution centred on the current value with SD 0.5.

For the actual data observations, we initially labelled them according to the region that they were in. For example, if the observation was in region one, the animal was in behavioural state one. We know that the behaviour does not change exactly at the time of the observation. This trajectory (the data alone) did not make sense; thus, before we could start the algorithm, we had to add in at least some potential switches similar to Section 3.6.

The posterior estimations were a result of running ten million iterations after a five million iterations burn-in. To check our posterior estimation for the boundary regard whether or not it converged, we used the same model with a different initial value for the boundary parameters. We also made longer updates. Before, we made updates over an interval containing three to six observations, as described in Section 3.6, but we changed it to five to fifteen observations. From the results of both updates, there was no difference between longer updates or shorter updates.

We considered deleting one point from the data because we thought it may be an outlier. We ran a similar analysis excluding that point to check if it made a difference to the estimation of the parameters. After deleting point number 18, there was no difference in the estimates for the movement parameters. As such, the outlier did not have a major effect.

6.5.4 Results

Figure 6.5 shows the estimates of movement parameters b and v for the two states, which together explain that the model can distinguish between the two regions parameters precisely. The posterior mean and SD for the movement parameters were the followings: $b_1 = -5.027(0.497)$, $b_2 = -0.748(0.057)$, $v_1 = 0.58(0.053)$, $v_2 = 7.5(0.6)$. The posterior mean for the movement parameters, v_2, was more than twice the size of v_1, suggesting that the movement outside the circle was more spread out. Also b_2 was larger than b_1, which means the movement inside had a weaker attraction towards the centre point than the movement outside the circle. This agrees with Tishkovskaya and Blackwell's results.

For the behaviour parameters estimate, the posterior mean and SD were as follows: $\lambda_{12} = 0.135(0.088)$, $\lambda_{21} = 0.83(0.074)$, $\kappa = 0.27(0.16)$. The acceptance rate for λ_{ij} was 0.3 and

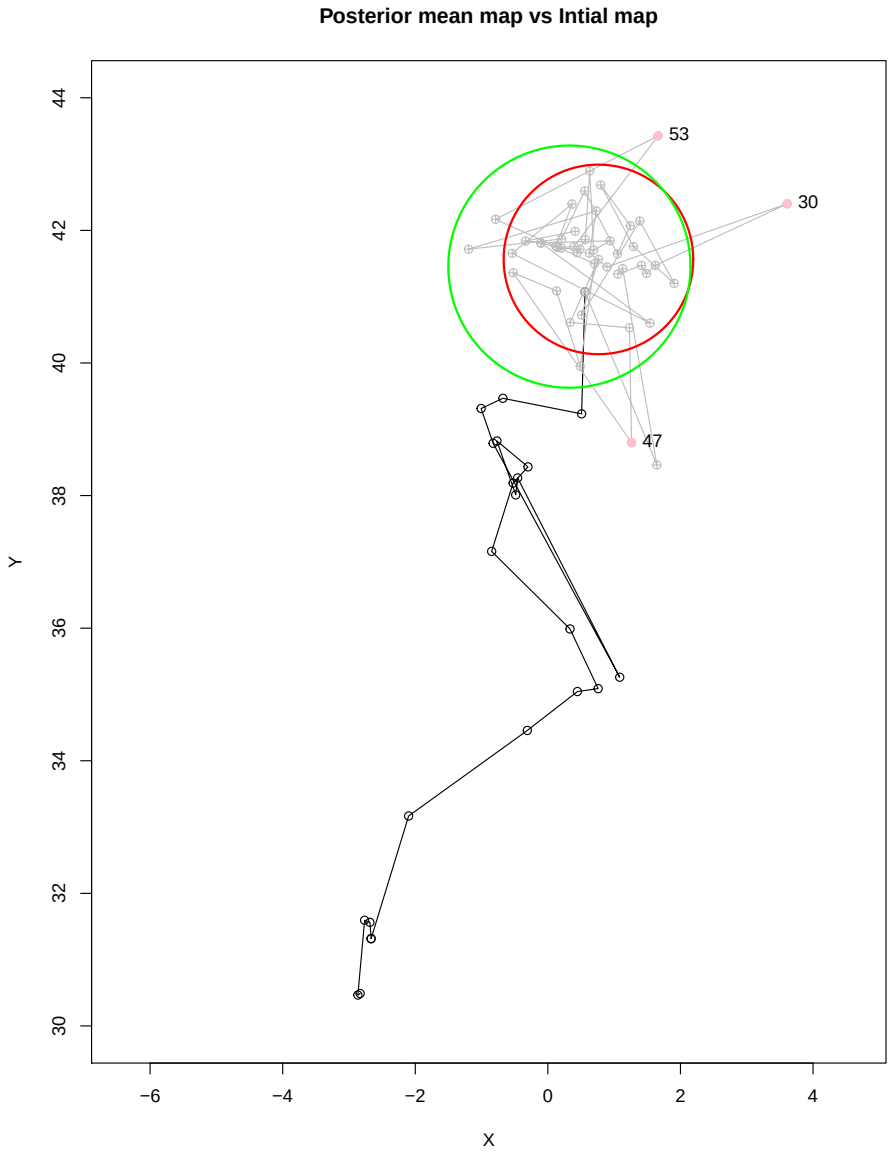

Fig. 6.3 Ibex movement data with the circular boundary between two regions. Red for the initial boundary and green for the posterior mean boundary.

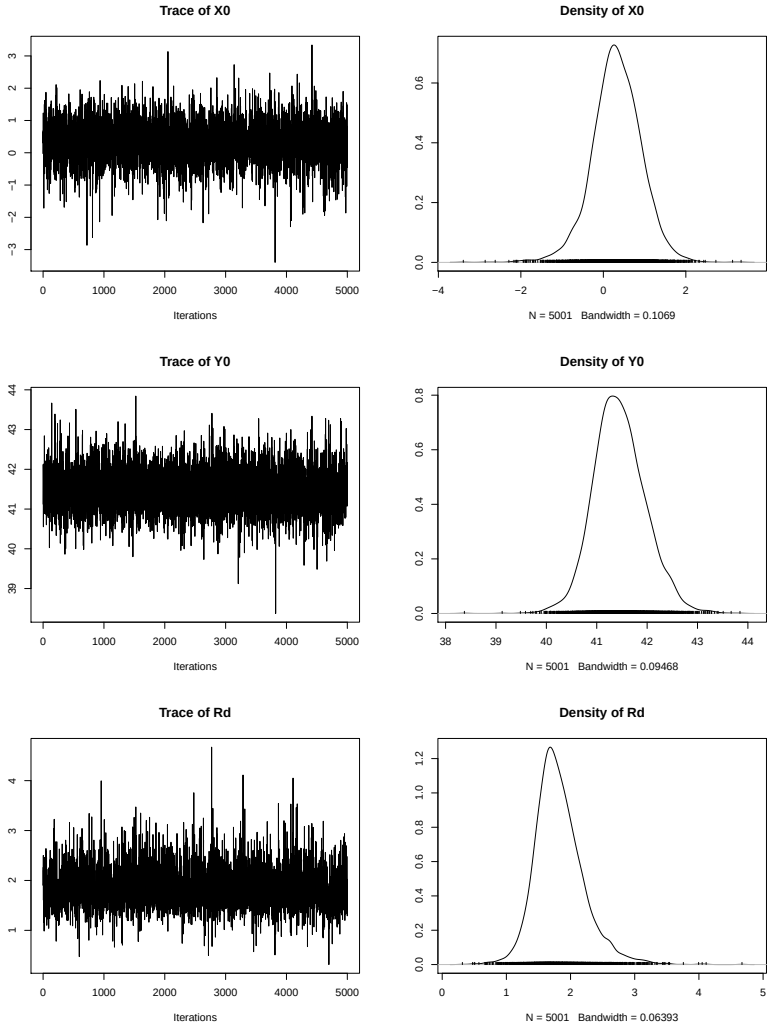

Fig. 6.4 Trace plots for the boundary parameters for the ibex data.

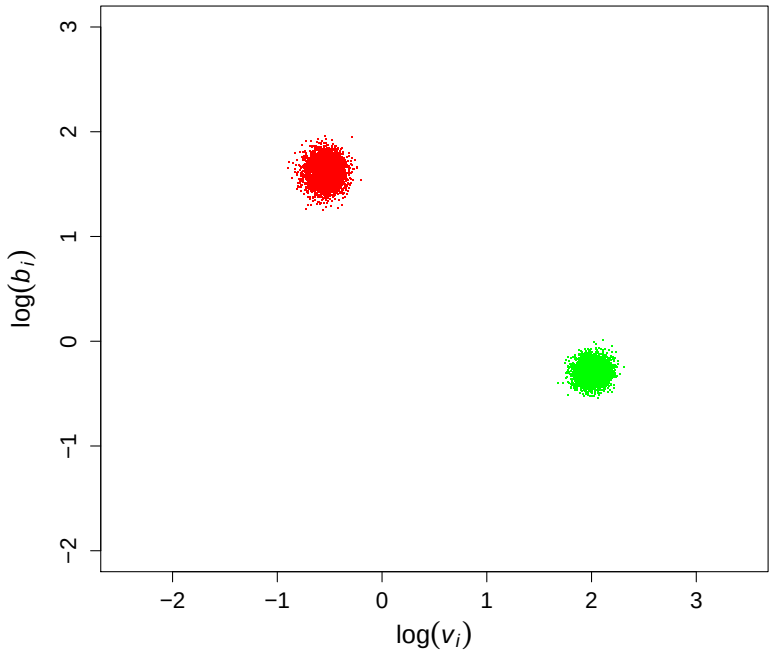

Fig. 6.5 Posterior results for the movement parameters for the ibex data. The two clusters correspond to states 1 and 2 (red and green, respectively).

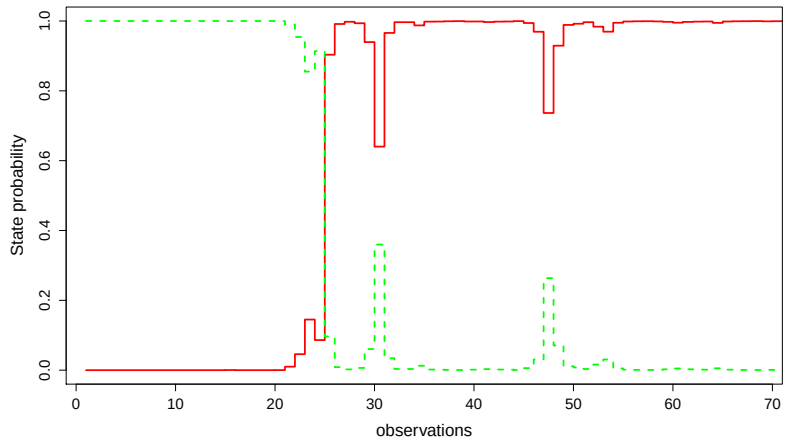

Fig. 6.6 The probability for each observation following state one (red) or two (green dashed) for the ibex data.

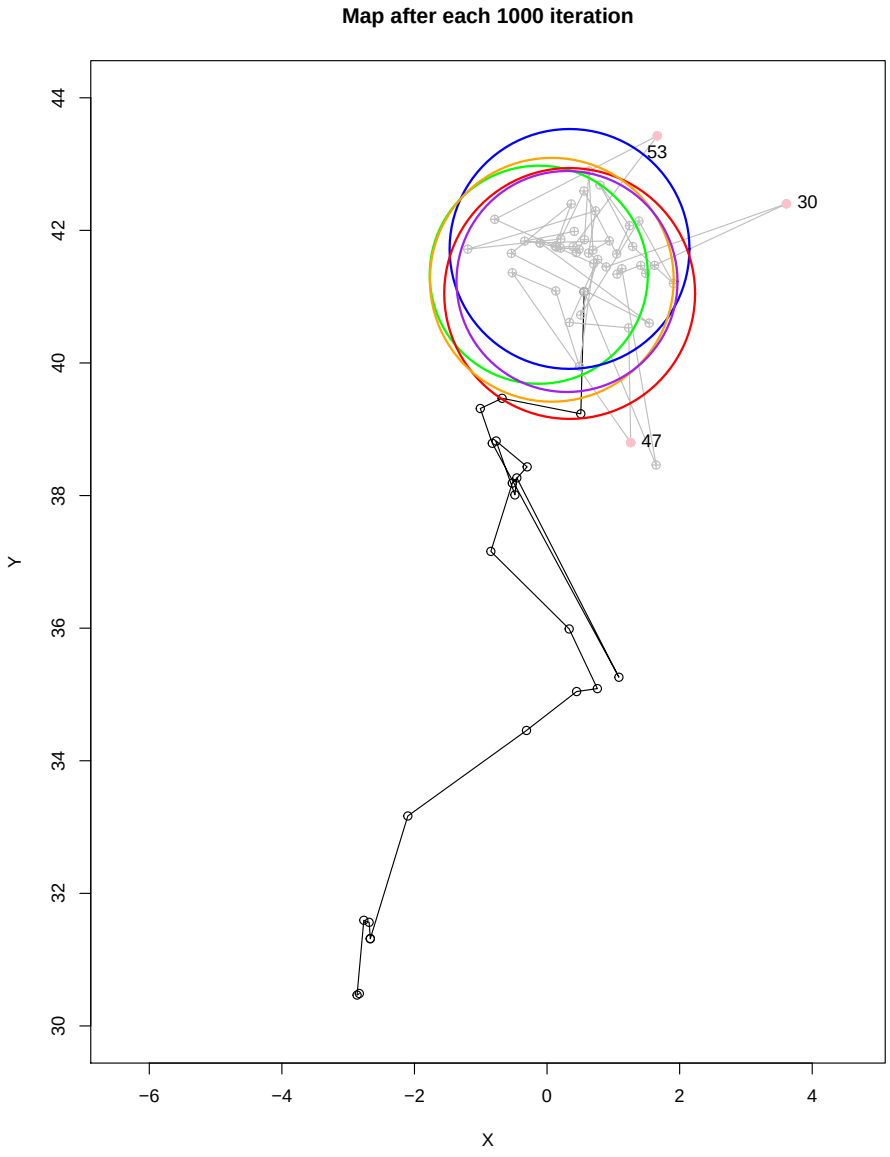

Fig. 6.7 The ibex movement data with a few samples from the posterior distribution of circle boundaries.

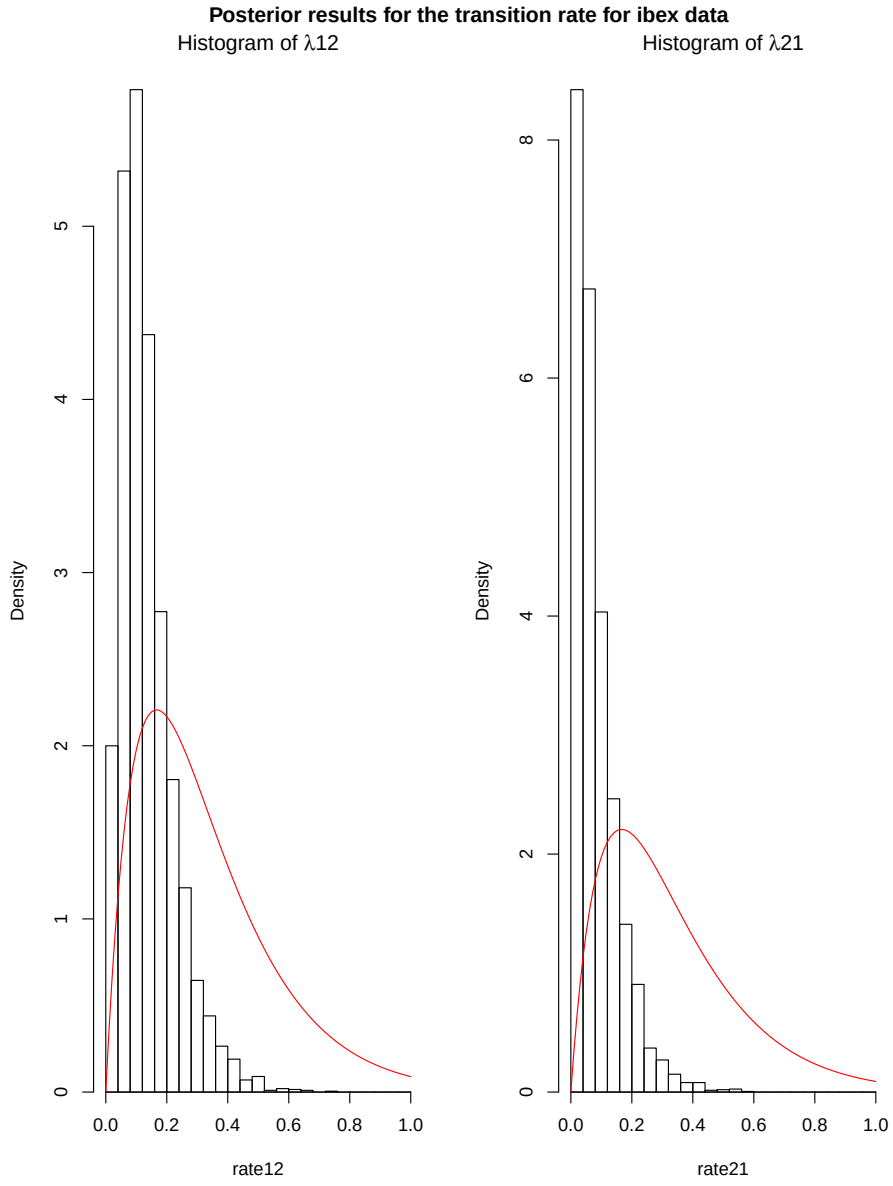

Fig. 6.8 Ibex data: The marginal posterior distribution for the transition rates between the two behavioural states. The marginal priors are displayed as red curve.

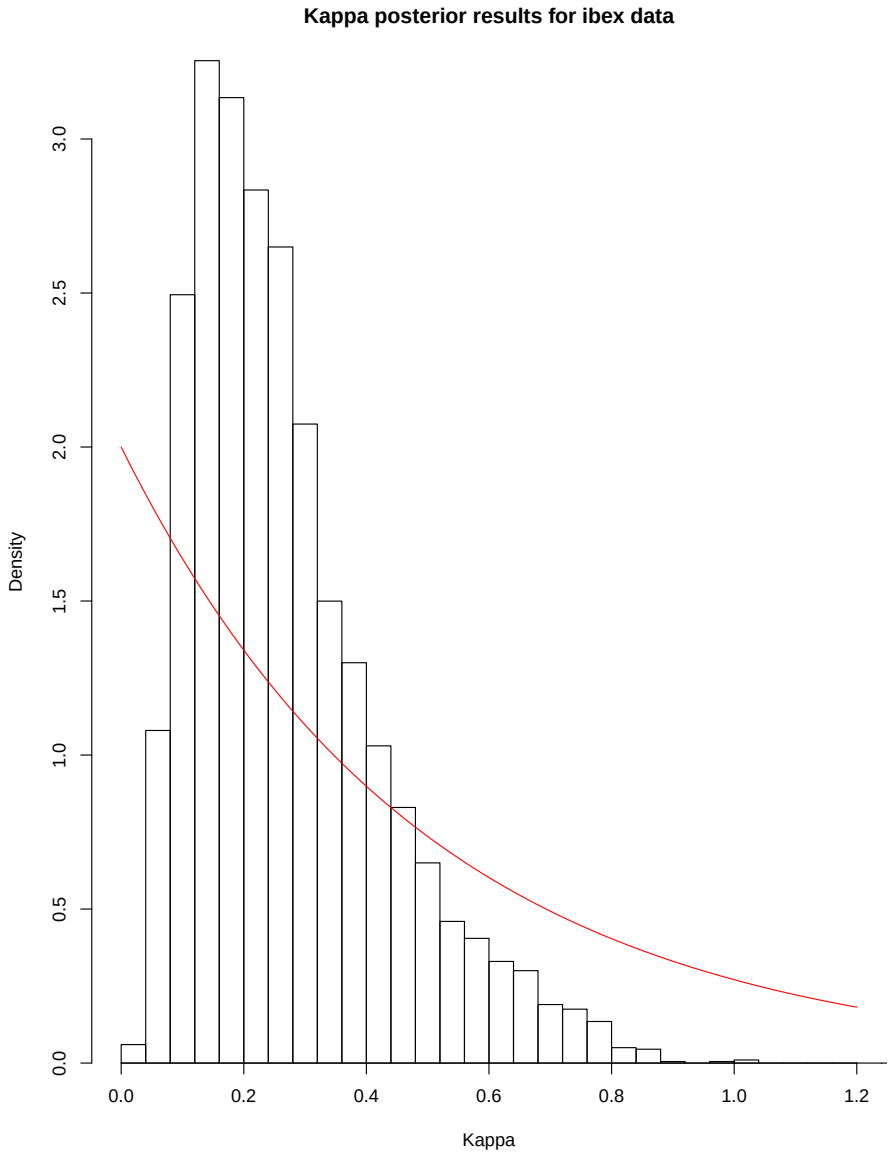

Fig. 6.9 Ibex data: The posterior distribution for the kappa. The prior is displayed as red curve.

for κ was 0.35. Figure 6.8 shows the marginal posterior distribution of the transition rates between the two states and the marginal priors displayed as a red curve. Figure 6.9 shows the posterior distribution of kappa and the prior displayed as a red curve. These results suggest that the data do not give us much informative about these parameters. We expected this because we had little information about the switches and the transitions. We had few points to tell us about the transition from behaviour state two to one, which means that the choice of prior becomes very important. As such, we are uncertain about the value of lambda. We can only learn about the value of lambda when the animals are in a region and have not yet changed their behaviour to match that region. Unfortunately, this state does not happen very much in this dataset. The ibex only crossed the boundary three or four times, so it is very hard to learn about what the values of the lambdas are, which is expected from this dataset. The location points tell us more about the boundary. We cannot omit the transition rates, because they are part of the model, and the estimation of the movement parameters is precise, so we should not worry about the large uncertainty on the estimate value of the lambda transition rate.

Figure 6.3 shows the estimated posterior boundary. The green circle represents the estimated posterior mean boundary between the two regions, while the red one is the initial boundary. The posterior mean estimates of the boundary parameters with the standard deviations (SD) are as follows: $X_0 = 0.325(0.6)$, $Y_0 = 41.453(0.52)$, $R = 1.827(0.38)$. Figure 6.10 shows the posterior distribution for the centre point and the marginal prior displayed as a red curve. Figure 6.11 shows the posterior distribution of the radius and the prior displayed as a red curve. In Figure 6.4, we see the trace plot for the boundary parameters, which show good mixing, with an acceptance rate of 0.34.

In Figure 6.7, the plot shows not only the posterior mean boundary but also a few boundary samples from the posterior distribution after the burn-in. Every few thousand iterations, we plotted the posterior boundary. Presenting the posterior uncertainty about the regions in the form of a plot gives us a geometric understanding rather than a purely numerical one in terms of the parameters. Figure 6.7 shows that all the posterior samples for the boundaries are very close to each other and that they give us a reasonable result. We get a slightly larger map boundary than the map in Tishkovskaya and Blackwell (In prep.).

Figure 6.6 shows the probability of each observation following each behavioural state. We can see that, from the first observation to the 20^{th} observation, the ibex was in state two, exploring, with probability one, because the ibex spent that time outside the boundary. After

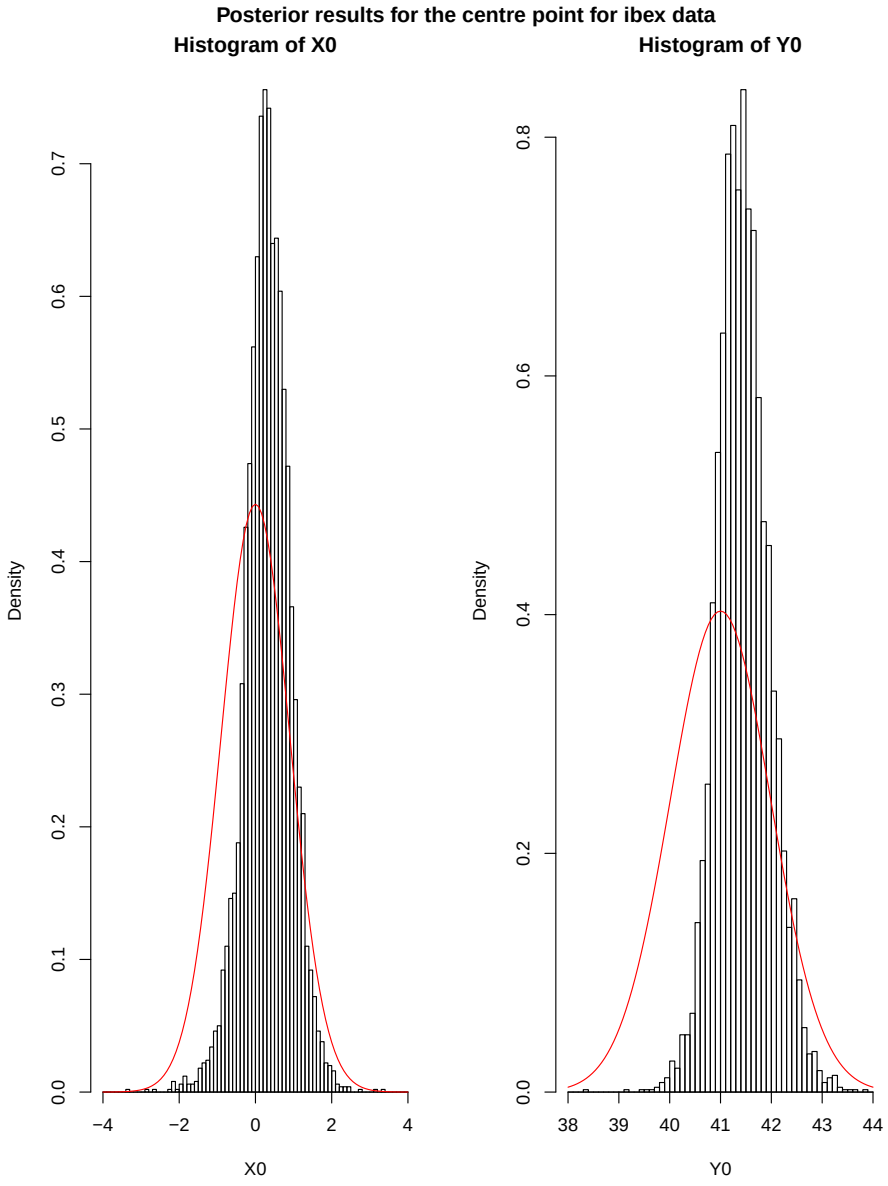

Fig. 6.10 Ibex data: The marginal posterior distribution for the centre point between for the circular boundary between two regions. The marginal priors are displayed as red curve.

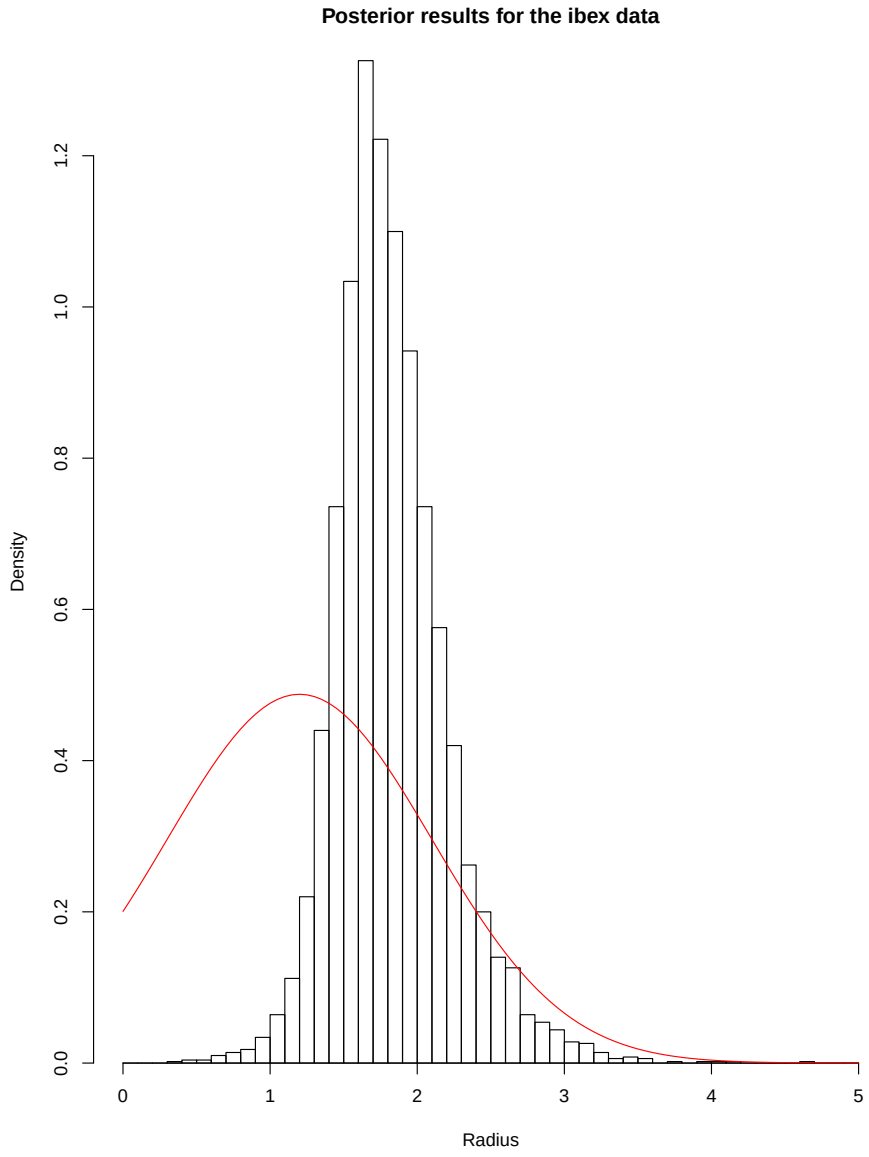

Fig. 6.11 Ibex data: The posterior distribution for the radius. The prior is displayed as red curve.

that, the ibex began switching to behaviour state one, foraging, with very high probability, around 0.9, by observation 25. This implies that the ibex had crossed the boundary. There is some uncertainty with observation numbers 30 and 50, which correspond with Figure 6.3, because these two points are outside the boundary, but still have some probability of following behaviour one. The results show that some of the points close to the boundary but outside the circle (region one), have the same behaviour as in region one because an animal does not switch its behaviour immediately after crossing the boundary.

In summary, we gained precise results for the movement parameters but uncertain results for the transition rates. This is acceptable, as we are using a small dataset and we have few switches between the two regions. We checked our parameter estimates by running the same model with a fixed boundary to estimate the transition rate and movement parameters, and it gave us a similar result. It took around five hours to run the model.

The DIC for ibex data

A much simpler model was fitted with only one behavioural state, one region, no boundary, no kappa, no switches, no transition rate and therefore no need for MCMC to update the behaviour or the boundary. Then, we compared it with the two behavioural state model and checked if the two state model fits better using DIC.

Before formal assessment using the DIC, we wanted to compare the estimated values for the movement parameters between the two-state model and the one-state model. Figure 6.12 shows the joint posterior for b and v. When we compared it with the two-state case in Figure 6.5, the parameters for the one-state case were "in between" those for the two-states in the switching model, which is as expected. As such, those estimates do seem reasonable when compared with the two-state case.

For the two-state model, we had eight movement parameters, and for the one-state model, we had four movement parameters. For the one-state model, the effective number of parameters was $p_v = 3.09$, the mean deviance $\overline{D} = 648.583$ and the DIC= 651.673. The effective number of parameters for the two-state model should be greater than for the one-state model. For the two-state model, the effective number of parameters was $p_v = 8.5$, the mean deviance $\overline{D} = 509.11$ and the DIC= 517.535.

In summary, we have an approximate form of the DIC that is easily calculated, is directly comparable with the simpler one-state model and which seems to behave sensibly based on

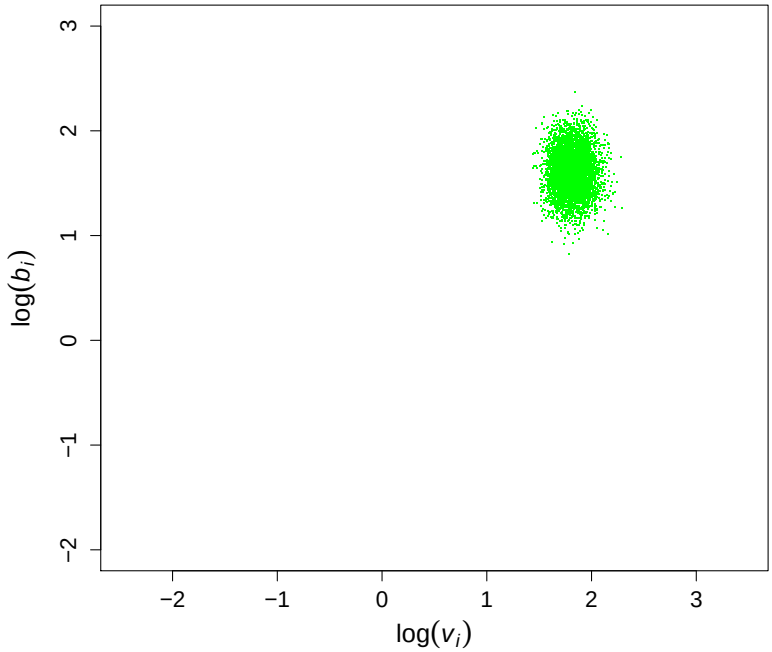

Fig. 6.12 Posterior movement parameters for the ibex data for the one state case.

the effective number of parameters. It shows that the two-state model is clearly better than the one-state model, even allowing for its extra complexity. In this example, because the difference between the DIC values for both models is large, we expect that the conclusions of both will be the same regardless of how we calculated the DIC.

6.6 Simulated experiments: linear boundary

We used the idea described in 6.2.2, where we had a linear boundary between two regions. The size of the simulated data was 100 observations, and the time interval between each pair of observations was one time unit. We assumed the habitat was divdied into two regions and that each region each region was associated with one behavioural state, so that we had two behavioural states with two transition rates $\lambda_{12} = \lambda_{21} = 0.3$. In all the experiments, the area above the line was region one, while the area under the line was region two. The two regions shared the same centre of attraction $\mu = (10,5)$, which was above the line in region one, while the start point (initial location) $(x_0, y_0) = (-20, -10)$ was under the line in region two. The linear boundary had parameters are $a = 0$ and $b = 0$.

We did three different experiments with the linear boundary. All the three simulated data

have the same behaviour parameters (transition rates λ) and the same boundary parameters (intercept a and slope b). The only different between the three data is the movement parameters for state 2 b_2 and v_2. First, one where b_1 was larger than b_2 and v_2 was larger than v_1; second, one where b_1 was larger than b_2 but v_2 was equal to v_1; and third, one where b_1 was equal to b_2 and v_2 was larger than v_1. We used these experiments to check how the linear boundary estimation was affected by the strength of attraction and also demonstrate the effect of the location variation.

6.6.1 Implementation

We updated four separate sets of parameters: the trajectory, the movement parameters, the behaviour parameters and the boundary parameters. All the three experiments used the same code with the same priors for all the parameters except the movement parameters for region two. The updating of the trajectory and movement parameters was carried out as described in Section 3.6. For updating the movement parameters, The prior for b_1 was a normal distribution with mean 0.1 and sd 0.1. The prior for v_1 was a normal distribution with mean 0.2 and sd 0.1.

For updating the behaviour parameters, we used the variable kappa approach, where kappa was treated as a parameter, as described in Sections 5.2.2 and 5.3. The forms of the priors for kappa and the transition rates were the same as in the previous chapter (Chapter 5), the marginal prior for κ was an exponential distribution with rate 3 and the marginal prior for λ was a Gamma distribution with shape 3 and rate 10. The proposals for kappa and the transition rate were similar to those in the previous chapter. We ran the model with an initial value for all λ of around 0.5, while the initial value for κ was 1

To update the linear boundary parameters (intercept a and slope b), a normal distribution was used as the prior for the intercept and the slope with mean$= 0$ and SD$= 0.25$. The proposal for the intercept and the slope were a normal distribution centred on the current value, with SD 0.9. The posterior estimations were a result of running ten million iterations after a five million iterations burn-in.

6.6.2 Experiment 1: Simulated data with b_2 larger than b_1 and v_2 is larger than v_1

In the first experiment, the real movement parameters values were $b_1 = -0.1 > b_2 = -0.5$ and $v_1 = 0.2 < v_2 = 1$, so there was a very large difference between the v_i values for states

one and two. The posterior behaviour parameters acceptance rate is 0.35. Figure 6.13

	Real value	Simulated
μ_x	10	10.1 (0.11)
μ_y	2	1.9 (0.11)
b_1	-0.1	0.1 (0.008)
b_2	-0.5	0.53 (0.03)
v_1	0.2	0.21 (0.02)
v_2	1	0.98 (0.1)
λ_{12}	0.3	0.3 (0.14)
λ_{21}	0.3	0.31 (0.16)
κ	0.8	0.7 (0.33)
a	0	-0.01(0.19)
b	0	-0.11 (0.14)

Table 6.1 Posterior means and SDs for the parameters of the simulated data with a linear boundary in Experiment 1.

shows the movement trajectory in black, with the real boundary in blue and the posterior mean boundary in red. There was some uncertainty in the boundary. Even if we assume the boundary to be a straight line, we have some uncertainty. For the posterior movement parameters, Table 6.1 shows that the results are accurate and precise.

For the posterior behaviour parameters, Table 6.1 shows the estimates are accurate but have high standard deviation. For the posterior boundary parameters, Table 6.1 shows the results are uncertain about the boundary, with high but still reasonable standard deviation.

Figure 6.14 shows the marginal posterior distribution of the transition rates between the two states and the marginal priors displayed as a red curve. Figure 6.15 shows the posterior distribution of kappa and the prior displayed as a red curve. Figure 6.16 shows the posterior distribution of the boundary parameters and the marginal priors displayed as a red curve. Figure **??** shows the posterior distribution of the movement parameters and the prior displayed as a red curve. These results suggest that the data do not give us much informative about these parameters. We expected this because we had little information about the switches and the transitions. We had few points to tell us about the transition from behaviour state two to one, which means that the choice of prior becomes very important.

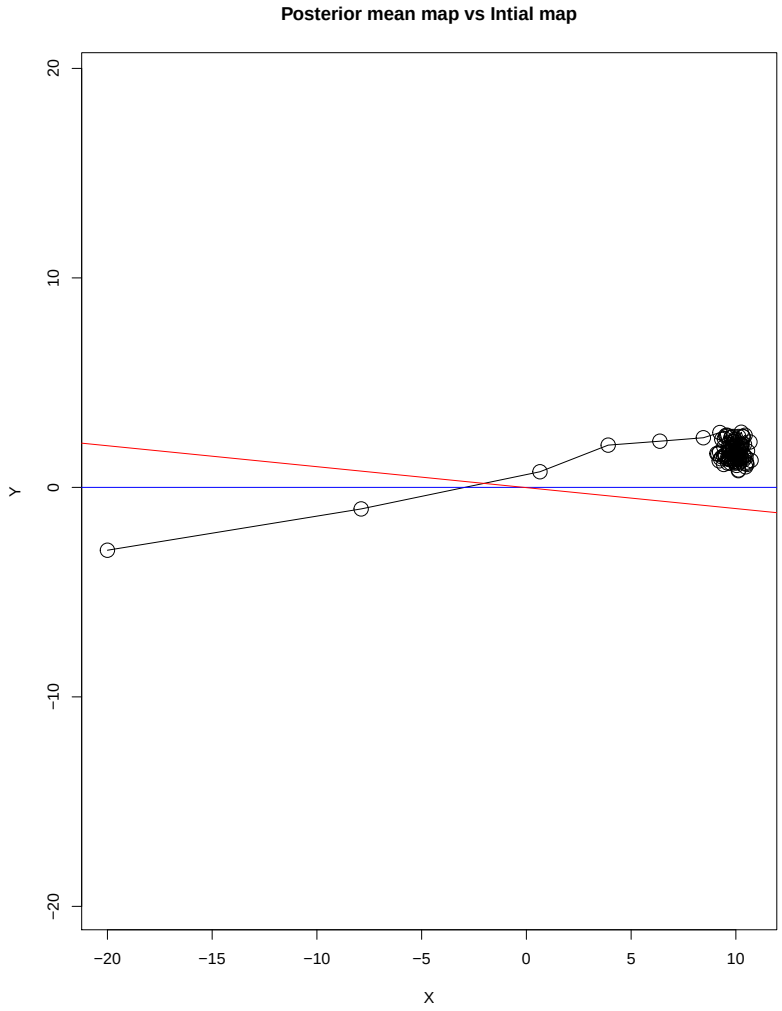

Fig. 6.13 The movement path for the simulated data for Experiment 1. Blue line for the real boundary and red for the posterior mean

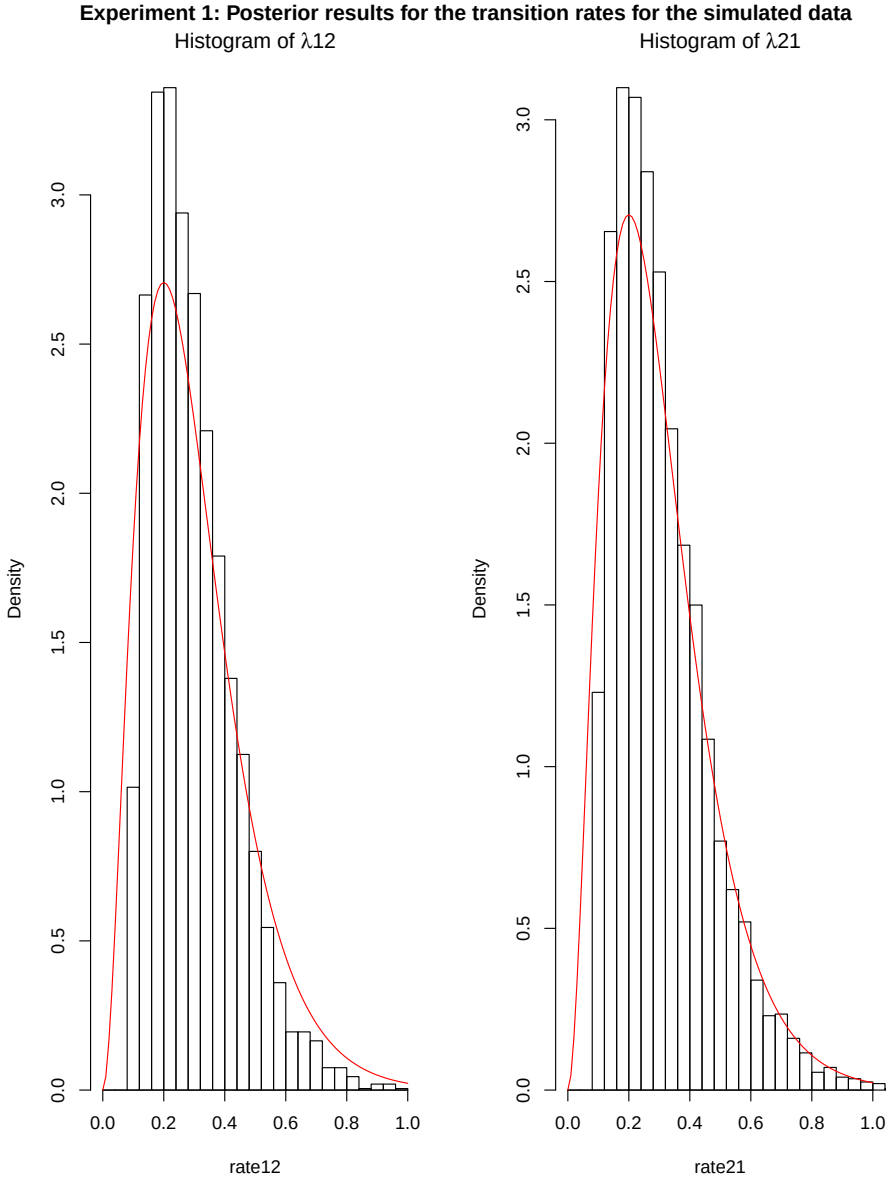

Fig. 6.14 Experiment 1: The marginal posterior distribution for the transition rates between the two behavioural states. The marginal priors are displayed as red curve.

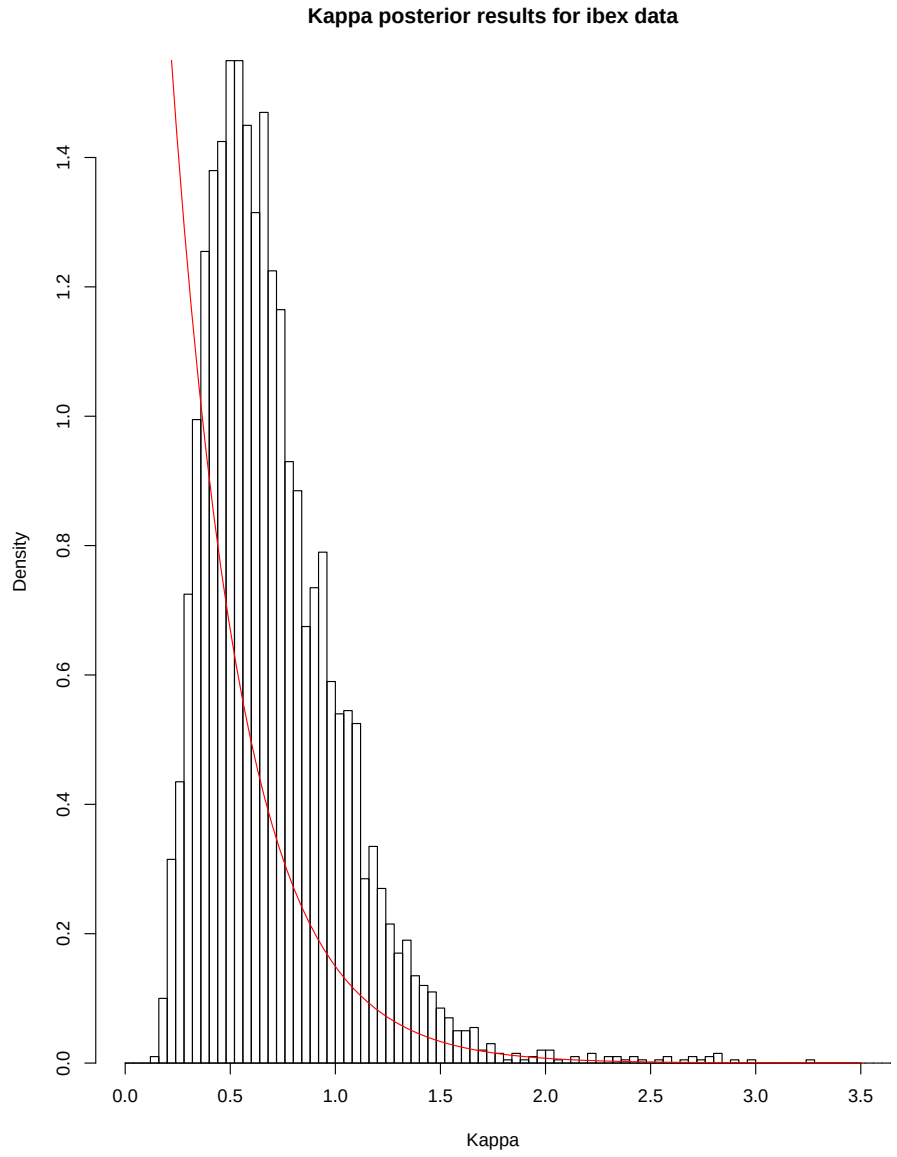

Fig. 6.15 Experiment 1: The posterior distribution for the kappa. The prior is displayed as red curve.

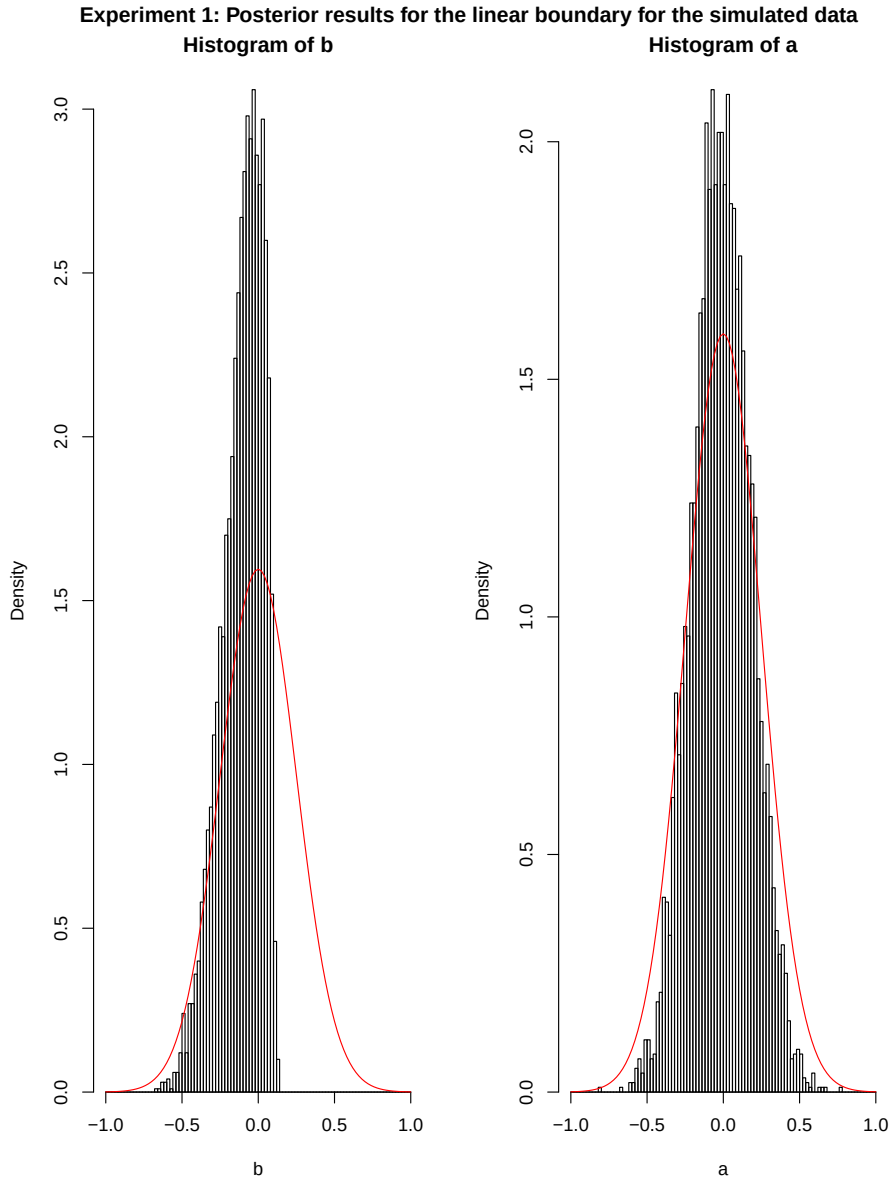

Fig. 6.16 Experiment 1: The posterior distribution for the boundary parameters between the two regions. The priors is displayed as red curve.

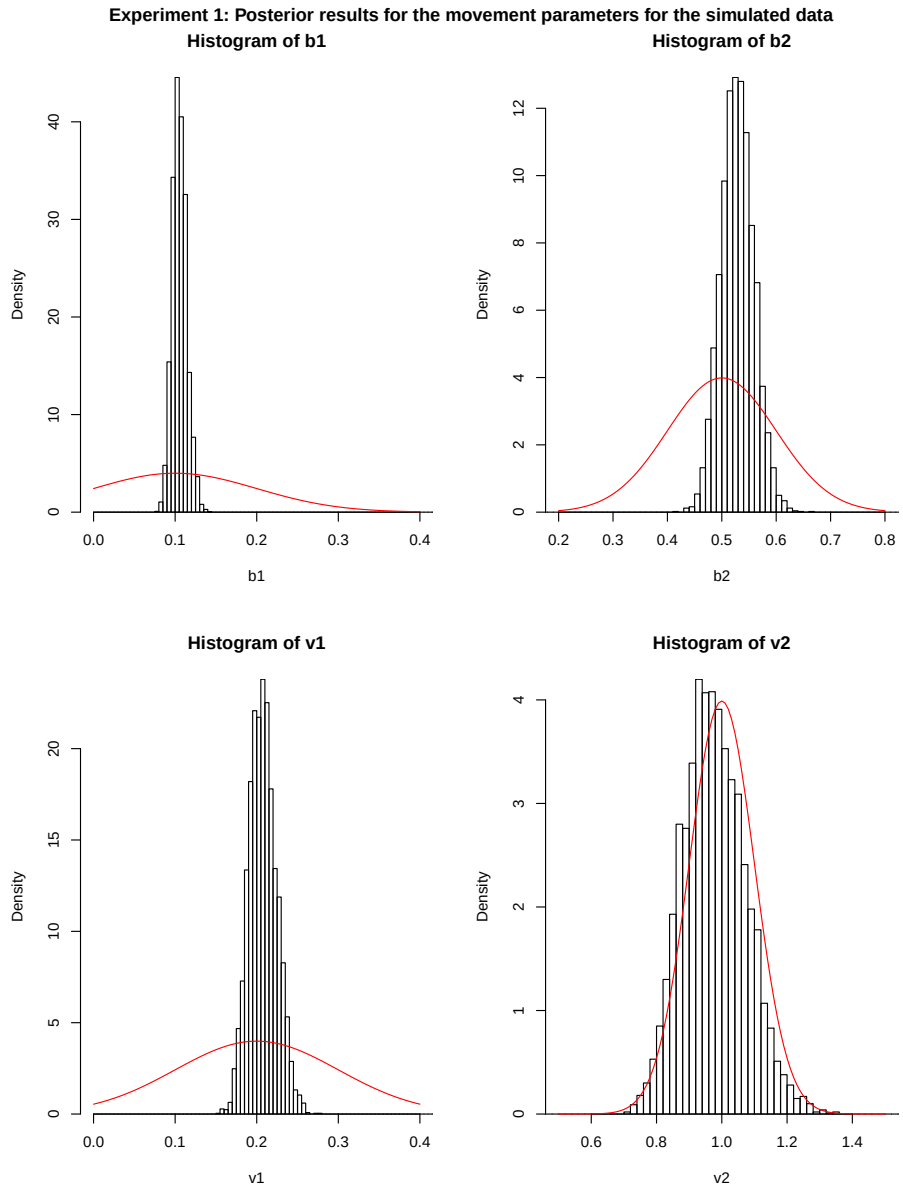

Fig. 6.17 Experiment 1: The posterior distribution for the movement parameters. The prior is displayed as red curve.

6.6.3 Experiment 2: Simulated data with b_2 is larger than b_1 but v_2 is equal to v_1.

In this experiment the real movement parameter values were $b_1 = -0.1 > b_2 = -0.5$ and $v_1 = v_2 = 0.2$. From Table 6.2 and Figure 6.23, CV for some of the parameters is much

	Real value	Simulated
μ_x	10	9.9 (0.12)
μ_y	5	4.9 (0.11)
b_1	-0.1	-0.1 (0.008)
b_2	-0.5	-0.5(0.014)
v_1	0.2	0.2(0.016)
v_2	0.2	0.2(0.02)
λ_{12}	0.3	0.31(0.16)
λ_{21}	0.3	0.3(0.17)
κ	0.8	0.8 (0.4)
a	0	0.001 (0.2)
b	0	0.009 (0.2)

Table 6.2 Posterior means and SDs for the parameters of the simulated data with a linear boundary in Experiment 2.

higher than the previous example. Figure 6.19 shows the marginal posterior distribution of the transition rates between the two states and the marginal priors displayed as a red curve. Figure 6.20 shows the posterior distribution of kappa and the prior displayed as a red curve. Figure 6.21 shows the posterior distribution of the boundary parameters and the marginal priors displayed as a red curve. Figure **??** shows the posterior distribution of the movement parameters and the prior displayed as a red curve. These results suggest that the data do not give us much informative about these parameters. We expected this because we had little information about the switches and the transitions. We had few points to tell us about the transition from behaviour state two to one, which means that the choice of prior becomes very important.

6.6.4 Experiment 3: Simulated data with b_1 equal to b_2 and v_2 is larger than v_1

In this experiment, the real movement parameter values were $b_1 = b_2 = -0.1$ and $v_1 = 0.2 < v_2 = 1$. Thus, there was a very large difference between the v_i values for state one and two, but the two regions had the same attraction toward the centre.

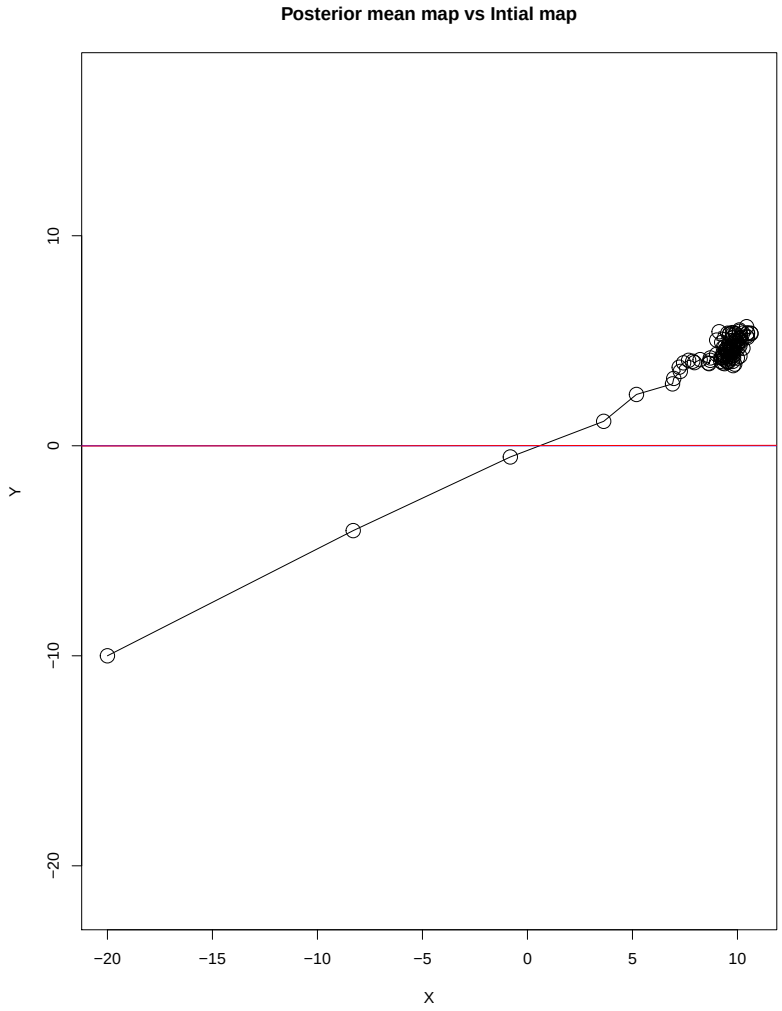

Fig. 6.18 The movement path for the simulated data for Experiment 2. Blue line for the real boundary and red for the posterior mean.

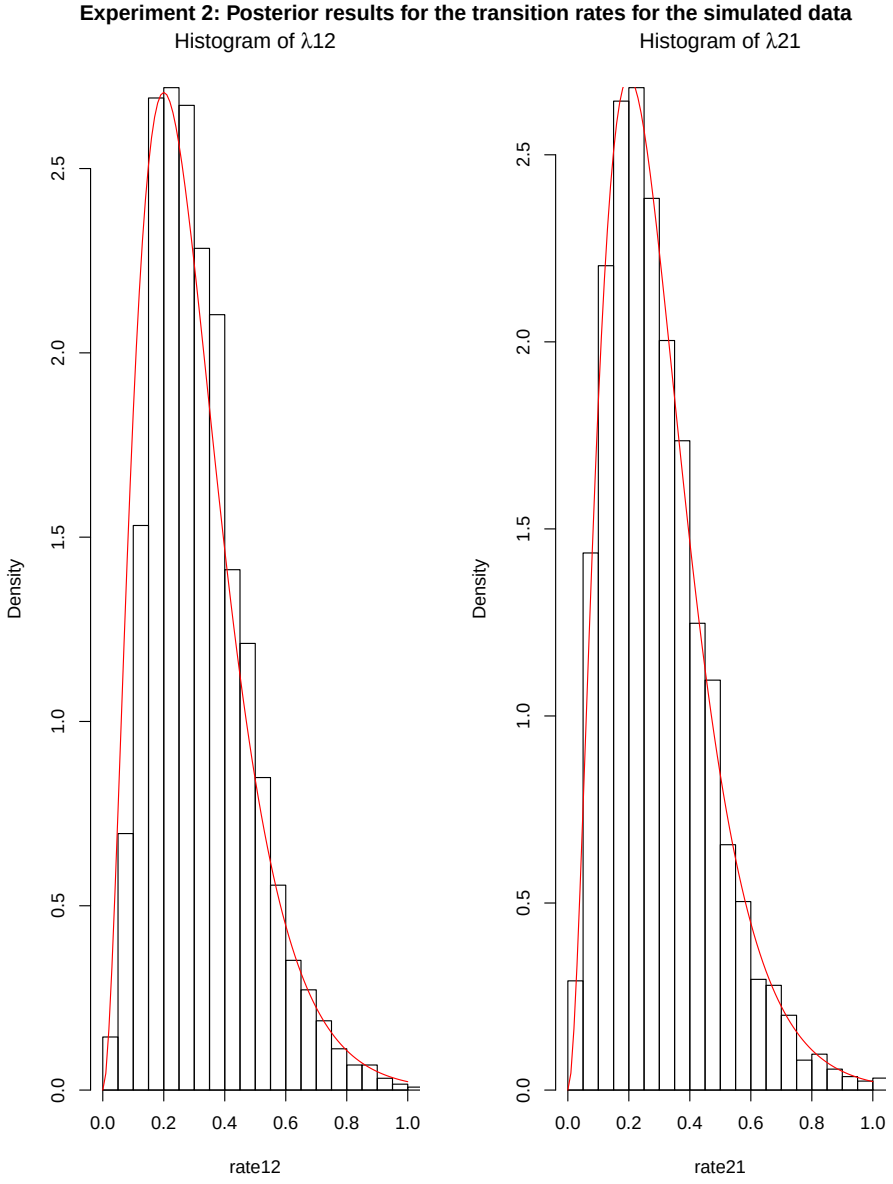

Fig. 6.19 Experiment 2: The marginal posterior distribution for the transition rates between the two behavioural states. The marginal priors are displayed as red curve.

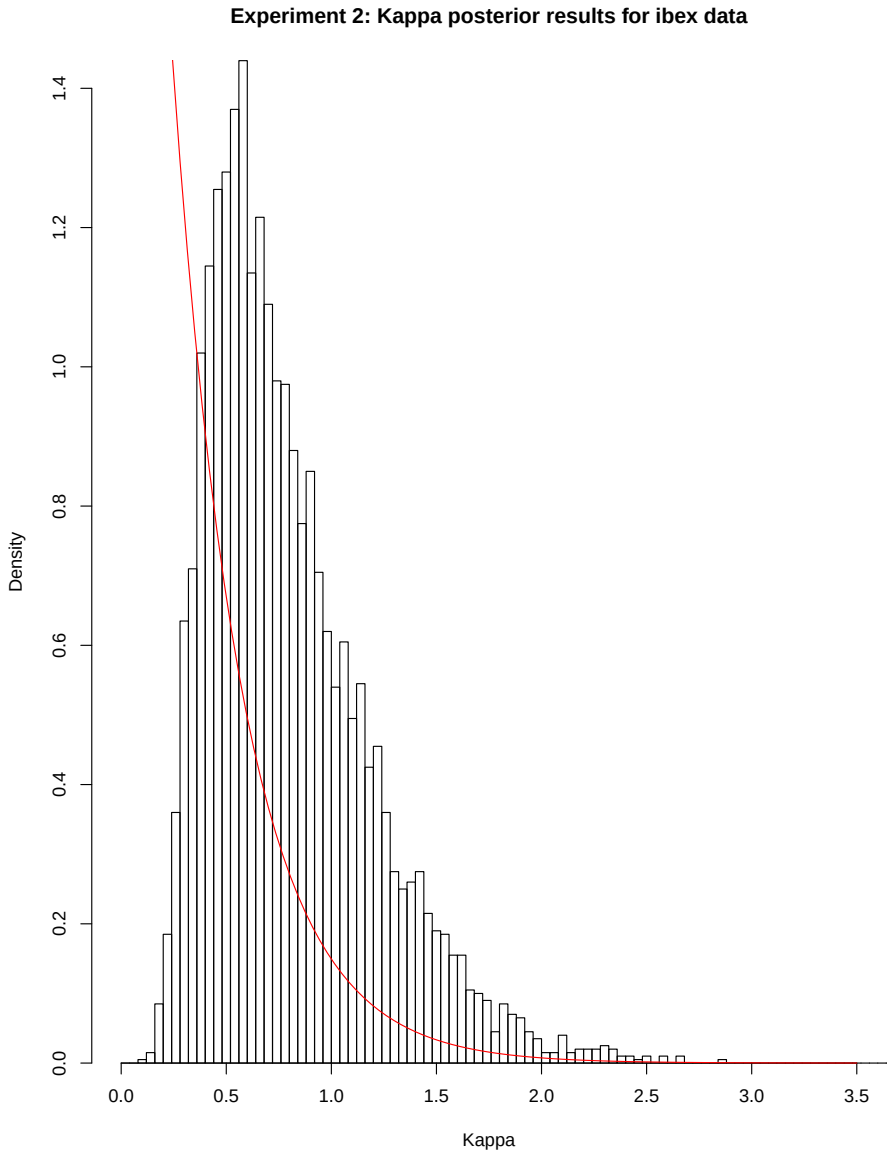

Fig. 6.20 Experiment 2: The posterior distribution for the kappa. The prior is displayed as red curve.

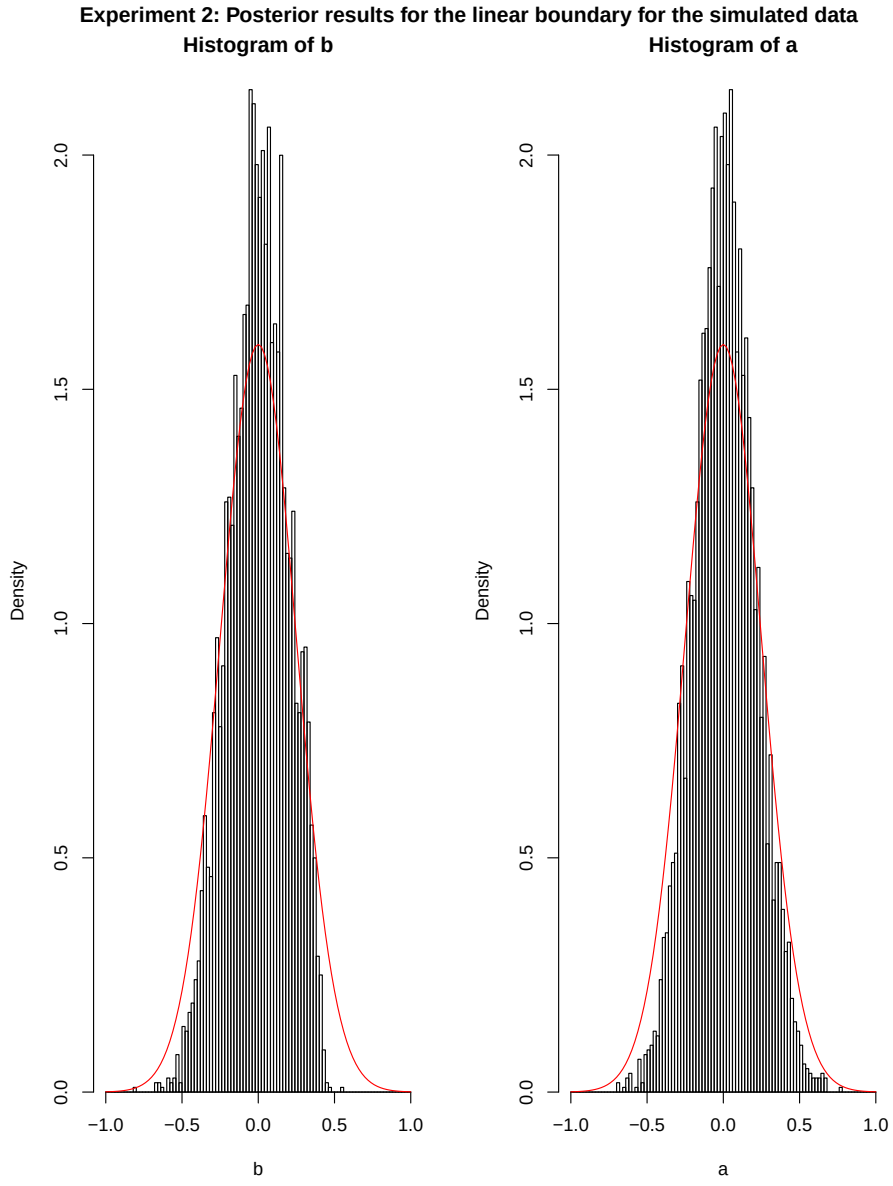

Fig. 6.21 Experiment 2: The posterior distribution for the boundary parameters between the two regions. The priors is displayed as red curve.

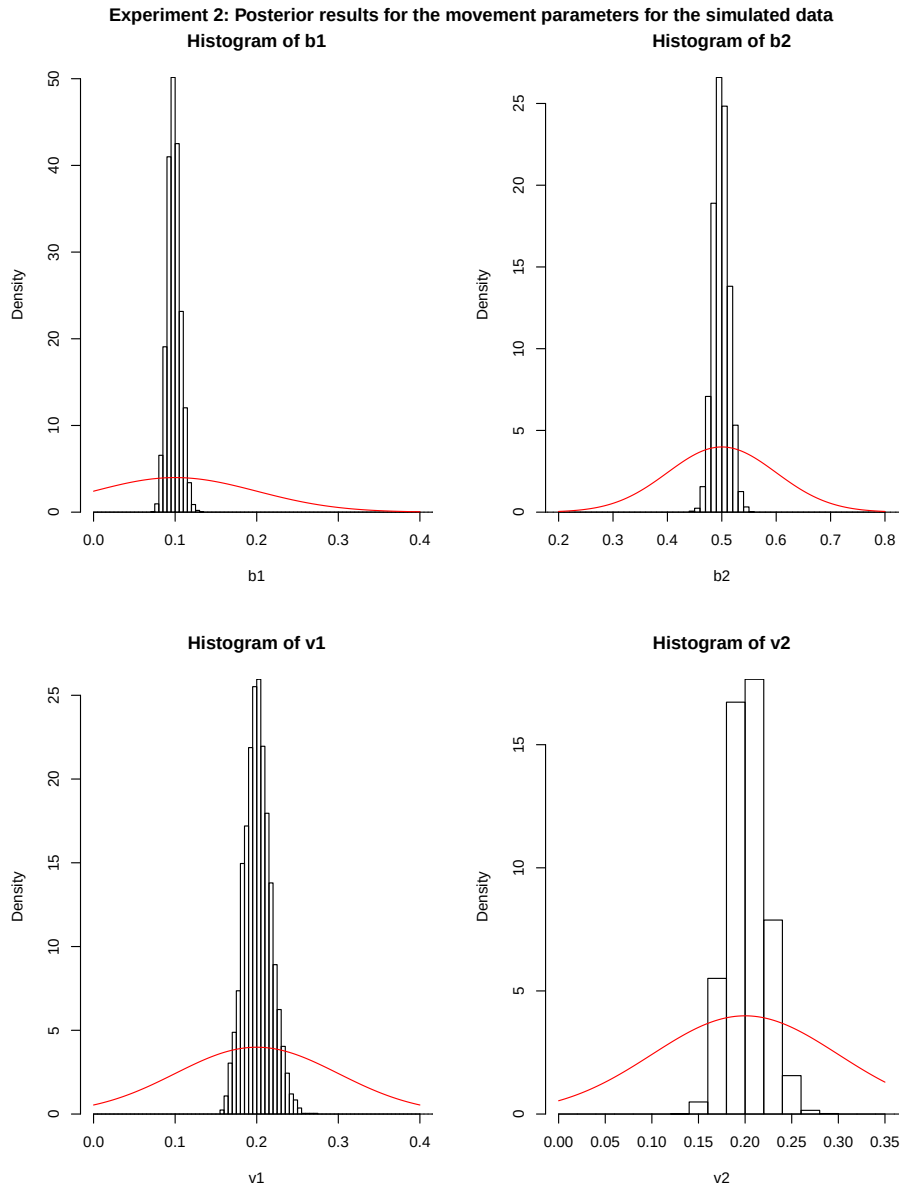

Fig. 6.22 Experiment 2: The posterior distribution for the movement parameters. The prior is displayed as red curve.

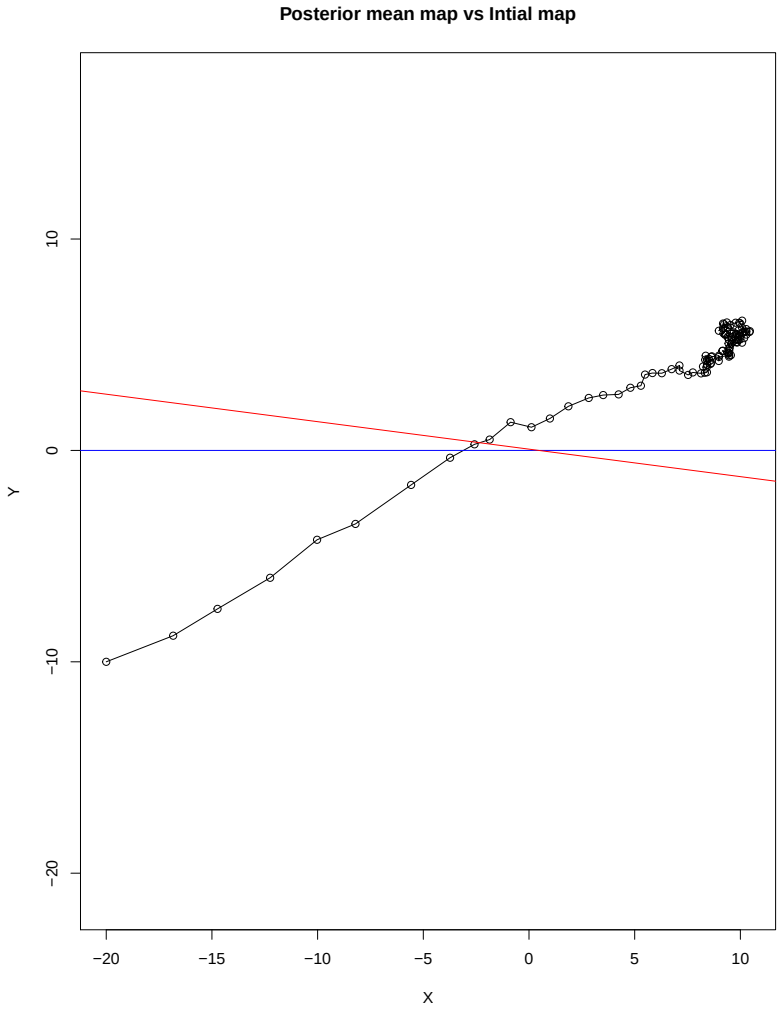

Fig. 6.23 The movement path for the simulated data for Experiment 3. Blue line for the real boundary and red for the posterior mean

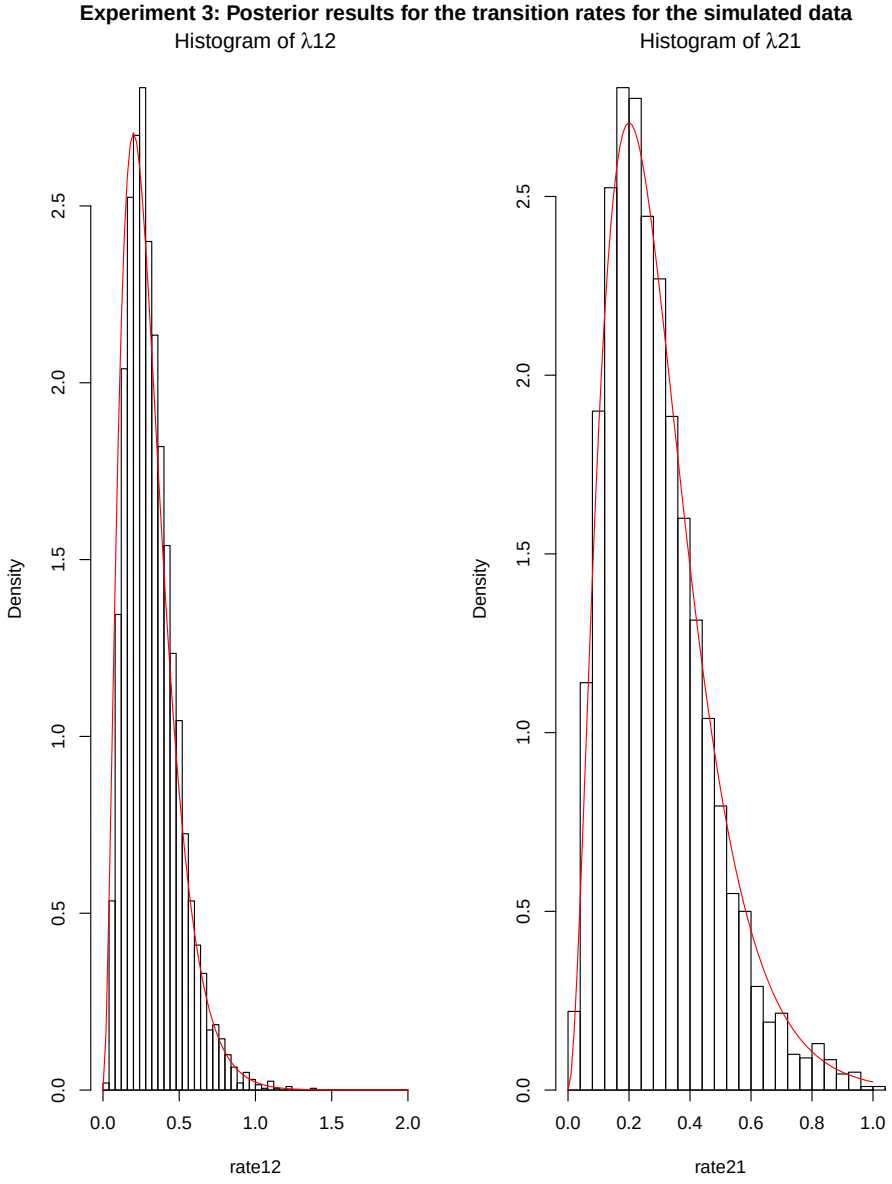

Fig. 6.24 Experiment 3: The marginal posterior distribution for the transition rates between the two behavioural states. The marginal priors are displayed as red curve.

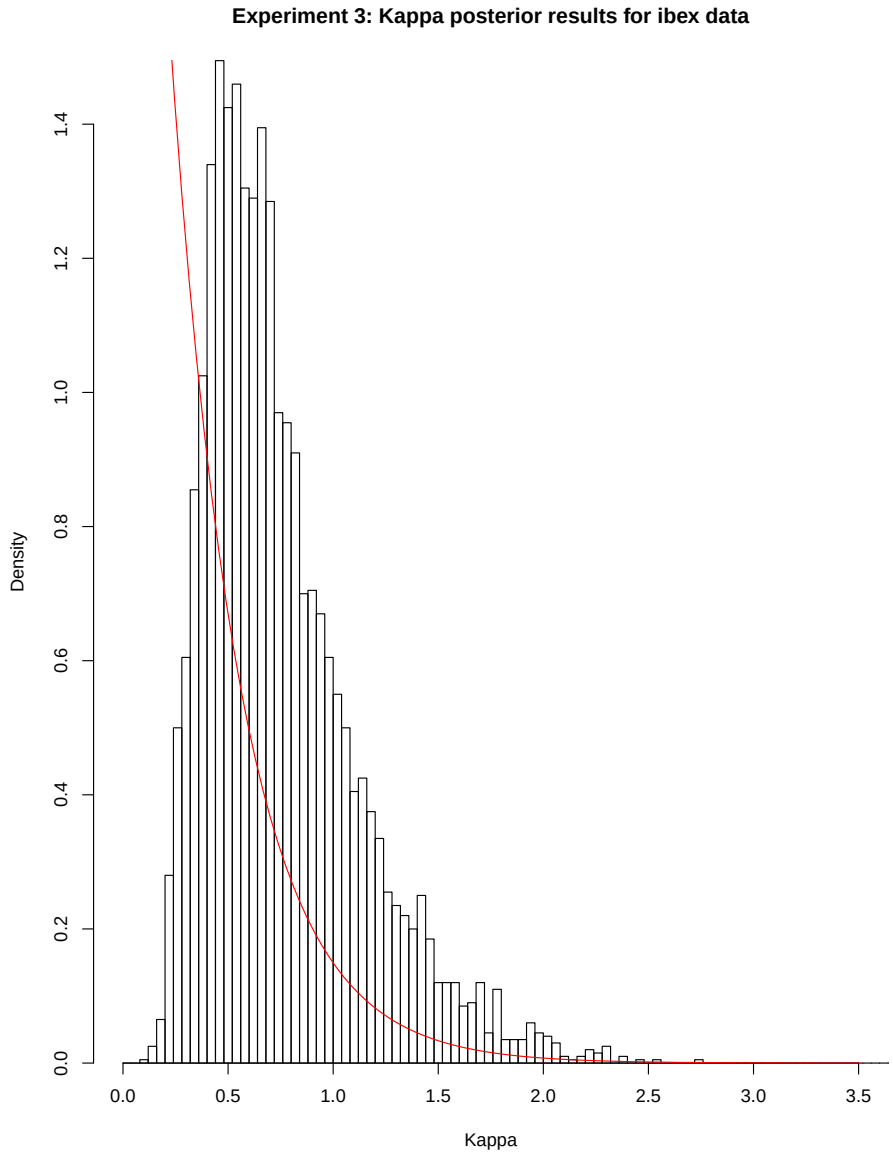

Fig. 6.25 Experiment 3: The posterior distribution for the kappa. The prior is displayed as red curve.

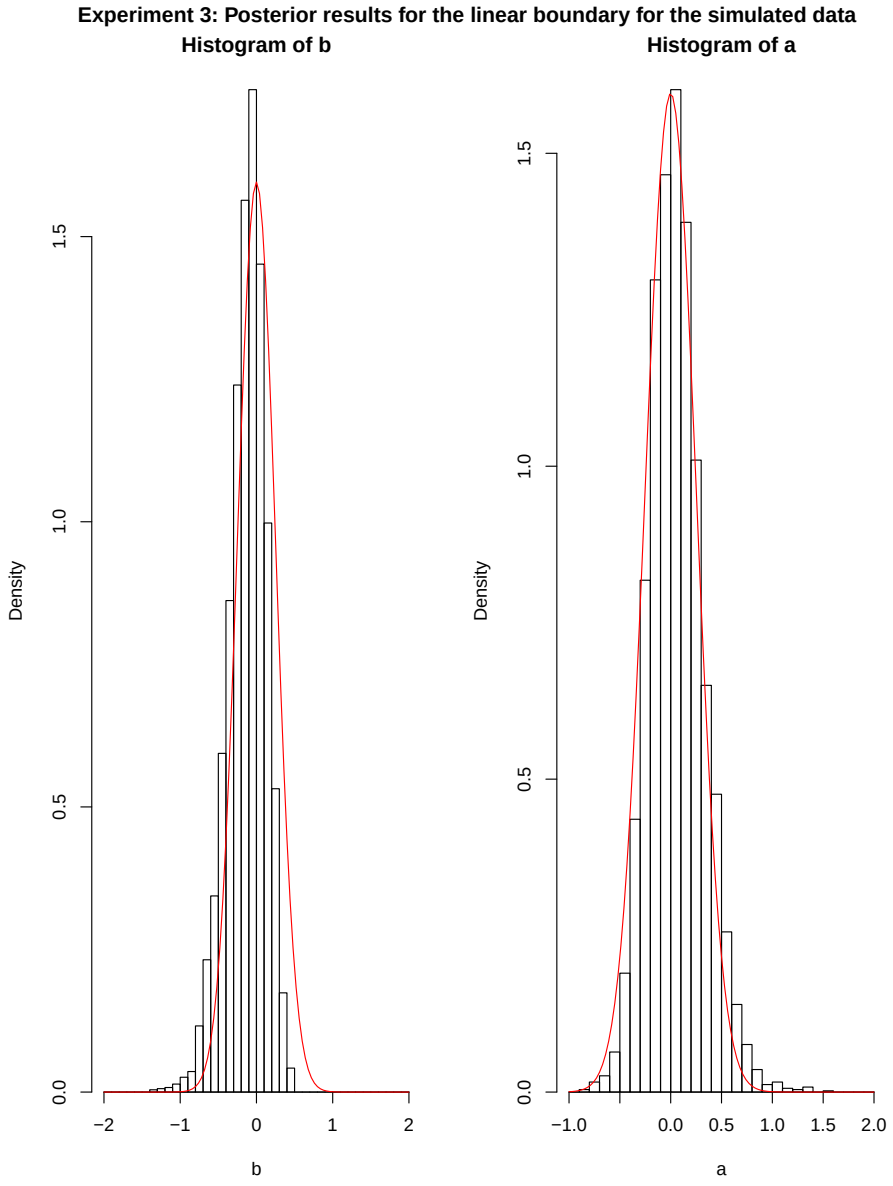

Fig. 6.26 Experiment 3: The posterior distribution for the boundary parameters between the two regions. The priors is displayed as red curve.

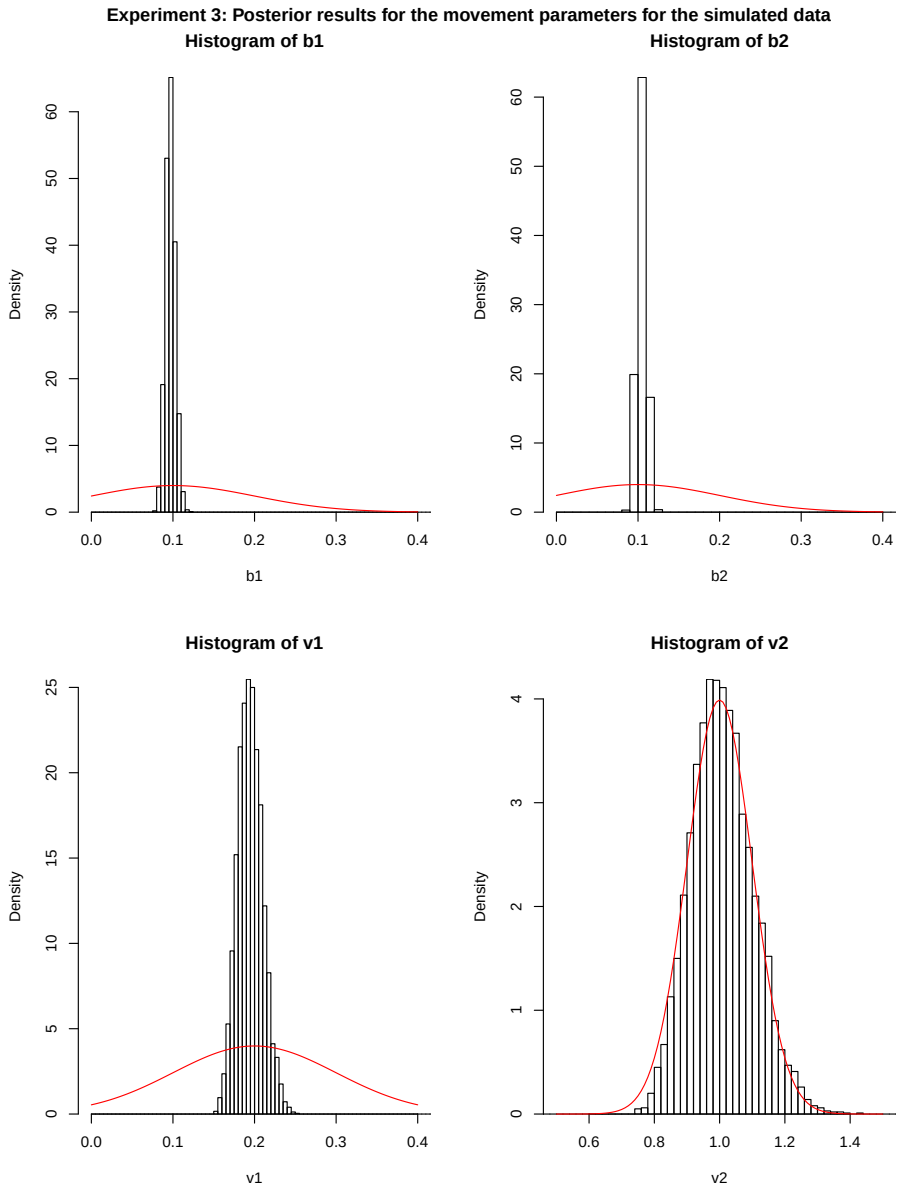

Fig. 6.27 Experiment 3: The posterior distribution for the movement parameters. The prior is displayed as red curve.

	Real value	Simulated
μ_x	10	9.9 (0.12)
μ_y	5	5.1 (0.12)
b_1	-0.1	-0.09 (0.006)
b_2	-0.1	-0.11 (0.005)
v_1	0.2	0.19 (0.015)
v_2	1	1.01 (0.1)
λ_{12}	0.3	0.32(0.17)
λ_{21}	0.3	0.29 (0.17)
κ	0.8	0.75 (0.37)
a	0	0.06 (0.3)
b	0	-0.13 (0.25)

Table 6.3 Posterior means and SDs for the parameters of the simulated data with a linear boundary in Experiment 3.

The model results can recognise that there were two states. Figure 6.23 shows the real boundary in blue and the estimated boundary in red. The posterior results for the boundary behaved in an expected way. For the posterior movement parameters, Table 6.3 shows that the results are accurate. For the posterior behaviour parameters, Table 6.3 shows that the results are accurate with a high standard deviation. That may be because there is only one switch between the two regions. Further, we had few data points under the line, which was also because of the small dataset, and we had only two switches. This makes estimating the transition rate very difficult. For the posterior boundary parameters, Table 6.3 shows that the results are uncertain about the boundary, with a high standard deviation, but they are still accurate. Figure 6.24 shows the marginal posterior distribution of the transition rates between the two states and the marginal priors displayed as a red curve. Figure 6.25 shows the posterior distribution of kappa and the prior displayed as a red curve. Figure 6.26 shows the posterior distribution of the boundary parameters and the marginal priors displayed as a red curve. Figure **??** shows the posterior distribution of the movement parameters and the prior displayed as a red curve. These results suggest that the data do not give us much informative about these parameters. We expected this because we had little information about the switches and the transitions. We had few points to tell us about the transition from behaviour state two to one, which means that the choice of prior becomes very important.

6.6.5 Conclusion for simulated experiment

For all the experiments above, the movement parameter estimation was accurate and precise, and it did a reasonable job of separating the states. The behaviour parameters estimation was accurate but not certain. The boundary estimate contained some uncertainty, but it was at a reasonable level. We tried other experiments with the real value for the movement parameters b_i and v_i for both states being relatively similar, but it was not possible to estimate the boundary, and the model cannot separate the two states. It is clear in this case that it is difficult to detect this difference.

To check the results, we used the same data but with a fixed boundary. In the fixed boundary, we got the same results, so there was no problem regarding the estimate boundary idea. As such, even in the two-state model, the movement behaviours were quite similar in their being attracted back toward the same centre of attraction. The results separated out the two states and the parameters of the two states.

The idea of estimating the boundary worked very well when there was a large difference between the states. The simulated experiments told us which cases were possible and which cases were difficult.

6.7 Discussion

We extended the Blackwell et al.'s (2016) concept to include the estimation of an unknown boundary between two regions. The boundary idea was included in the model by adding an extra MCMC step to estimate the boundary parameters. We fit two types of boundary: one with a circular habitat boundary and one with a linear boundary dividing the environment into two different habitats. Our extension of this research will help the ecologists learn more about using movement data to study the environment and how different habitats affect the animal movement. This idea can be extended to a more complicated model, where the habitat consists of many different patches rather than just two regions. In this case, the habitat could be split into many regions, but the same idea would apply. The linear boundary can be generalized (given suitable data) to a smooth nonlinear boundary.

Chapter 7

Generalization of models for duration of behaviour

7.1 Introduction

As discussed in Chapter 3, the Markov assumption is important for the Ornstein-Uhlenbeck (OU) process with behavioural switching. For the Markov assumption to be correct, the behaviour process constrains the time spent in a given behavioural state (the duration-time) to follow an exponential distribution in the continuous time models, or a geometric distribution in the discrete time models. This is not reasonable or appropriate in some situations. One solution to this problem is to implement a semi-Markov process, wherein duration can follow any distribution with positive support. The downside to this solution is that the model loses the Markov property, which results in more complicated likelihood calculations. This means that the probability of transitioning from one state to another in the semi-Markov model depends on the duration-time of that state $\lambda_{ij}(t)$, while in the Markov model, the probability of transition is constant in any given state of $\lambda_{ij}(t) = \lambda_{ij}$, which does not depend on time. We do not consider models in which the movement within a state is non-Markov.

Many studies have attempted to deal with the duration-time distribution limitation without losing the Markovian property. Cox and Miller (1965) introduced the method of stages as a solution for non-Markovian cases. The method of stages assumes that the process will go through sub-states in a series. Each state is divided into k sub-states. The time spent in each sub-state follows an exponential distribution so that the Markovian property is maintained, but the total of the sub-states in the series is not exponential. Cox and Miller (1965) stated that the simplest case where all sub-states share the same $\lambda_{ij} = \lambda$, or transition rate. The total

time spent in each state is proportional to the Gamma/Erlang distribution with parameters (k, λ). This process is Markovian. The partition of states into sub-states is a mathematical strategy and does not have any biological interpretation. Many recent studies have used the distribution function of the Erlang exponential staging method to model the duration-time of states as the sum of independent exponential sub-states.

Cox and Miller (1965) concluded that the incorporation of sub-states in series or parallel can approximate any distribution, but an accurate approximation requires a larger number of sub-states. Using this method, Cox and Miller (1965) noted that when the number of sub-states k is very large, an accurate approximation of any distribution can be ensured. However, in practice the number of sub-states is usually small to simplify the calculation. Cox and Miller (1965) concluded that the important point is to increase the flexibility in the presentation of duration-time distribution which can be obtained with a small number of sub-states

This idea developed into the concept of the phase-type distribution, which has made the Markov process an effective and widely used stochastic model. The distribution of time in the joining of an absorbing state from a group of transition states in a continuous time Markov process is called the phase type distribution. Each transition sub-state is called a phase. Thus, if there are n transition sub-states, the process has the phase type distribution with order n. The time spent in each phase is exponentially distributed, and the process begins with phase one. Each transient phase is visited at least once before absorption. Any positive distribution in $(0, \infty)$ can be approximated by the phase type distribution. Discrete distributions can be approximated, but it probably needs a lot of phases to get something close to a discrete distribution.

The phase type distribution is a generalization of the exponential and Erlang distributions. The phase type distribution with only one transition phase is equivalent to an exponential distribution with a rate of λ. More generally, the Erlang distribution $E(k, \lambda)$ is the sum of k, exponential phases having the same or equal rate of λ. The Erlang distribution is further generalized by allowing each exponential phase to have different transition rates of λ_{ij}. This is called the hypo-exponential distribution. Allowing the process to start in any phase and permitting each phase to have different transition rates is called the hyper-exponential distribution. A combination of the hypo and hyper-exponential distribution is called a Coxian distribution. A Coxian distribution begins in phase one, moving through the various phases with different rates (λ_i). In each phase, there are two possibilities: moving into to the next

phase, or going into the absorbing state. A Coxian distribution can be extended by allowing the process to start in any state. This process is called a generalized Coxian distribution (Buchholz et al., 2014).

Rubino and Sericola (1989) discussed the duration-time time of a homogeneous finite Markov process in both discrete and continuous time models without an absorbing state. Rubino and Sericola (1989) used a Markov chain, grouping the sub-states into state aggregates. Each aggregate corresponded to one behavioural state. In general, the time spent in one aggregate by the Markov chain is not geometric or exponential, and the sequence of transitions between the aggregates produces a Markov chain. This is done to protect the Markovian property, even if the group of the sub-states is not Markovian.

These approaches to more flexible duration times have been applied in a discrete-time setting to hidden Markov models by several authors. Russell and Cook (1987) compared two solutions to the limitation imposed by the geometric assumption. First, they considered a hidden semi-Markov model (HSMM), defined to be a generalisation of an HMM in which the duration of a visit to a state has a general discrete distribution. Bulla et al. (2009) introduced an R package for analysing the hidden semi-Markov model in discrete time. Second, they considered an expanded state hidden Markov model (ESHMM). Russell and Cook (1987) defined the ESHMM as replacing each state of an HMM with a group of Markovian sub-states. Thus, the distribution function for the time in an existing state is equivalent to the distribution of the sum of the times in the sub-states. Johnson (2005) also experimented with HSMM, ESHMM; he also considered a further extension of HSMM, called Variable Transition HMMs, which we do not consider any further here. Both Russell and cook (1987) and Johnson (2005) showed that when the number of sub-states is large enough, the performance of the HSMM and the ESHMM are equivalent. For the ESHMM, the Markovian property is maintained so that the standard Markovian algorithm can be used. As the number of sub-states increases, computational complexity also increases. This is a limitation of the ESHMM. However, when the total number of sub-states increases, so does the ability of improving target distribution estimation. For this reason, the ESHMM is seen to be an efficient option in comparison to developing a complex and expensive computational model.

Aparna et al. (2014) used the ESHMM to predict the farrowing of sows. This was done by assuming that sows go through behavioural states, such as: a exploring state, nesting state, resting and farrowing. Each state was divided into sub-states, and each sub-state followed an

exponential distribution. As such, the total duration-time in the state was non-exponentially distributed, and farrowing was the absorbing state. These behavioural states were not directly observed.

The main problem with the HSMM is the loss of Markovian property, which leads to complex computation, difficulty including covariates, and issues with making predictions. In the wild, there is a lack of information regarding the animal being studied. The Markovian property helps to deal with this lack of information, as future states depend only on the current ones. Langrock and Zucchini (2011) structured a new application of the ESHMM that fits and approximates any duration-time distribution. This is called the state-aggregate process. This concept is similar to those discussed by Russell and Cook (1987) and Johnson (2005). Langrock and Zucchini (2011) argued that an HMM with aggregate sub-states is appropriate, as it will not change with differing state duration-time distributions. It can also use the same statistical tools and methods for making predictions or also for including covariates that already exist.

The aim of this chapter is to address this limitation by allowing the duration-time of each behavioural state to follow any continuous distribution, by allowing distribution to be approximated. This is done by expanding the behavioural states into multiple sub-states. Langrock and Zucchini (2011) used this idea for animal movement in a discrete-time model. While this has been suggested before, it has not been implemented for the animal movement model under a continuous-time OU process. This chapter proposes expanding upon the state-aggregate approach introduced by Langrock and Zucchini (2011) by extending it to continuous time. We want to take the same developments that have been made in discrete time models and applying them to the OU model.

Blackwell et al. (2016) discussed the use of different distributions for the duration-time in a continuous-time model under a semi-Markov model. Although this study mentioned the use of the hazard function as the transition rate, they did not attempt to implement it. In the hazard idea, inference is possibly to be hard. This is because a change in the update of an interval can affect transition rates once an interval has ended. Thus, to facilitate the inference, a better approach and a more elaborate Markov method must be established.

7.2 The model

The aim of this chapter is to present a behavioural process that represents any continuous duration-time distributions. The proposed model expands a non-exponential behavioural state to multiple sub-states, forming a state-aggregate. Within each aggregate, each sub-state follows an exponential distribution.

For example, if the time spent in the feeding state does not follow an exponential distribution, the feeding behaviour can be divided into two sub-states: feeding A and feeding B. Each of these sub-states would have an exponential distribution, and the distribution of the total time spent in the two sub-states (the general feeding state) would be a non-exponential distribution. This allows for more flexible modelling of the time spent feeding. The differences between the feeding A and feeding B states would not be observed in the field. The expansion of behaviour does not have any biological meaning. Rather, this is a mathematical process used to generalize the model and simplify the inference.

To obtain a precise approximation of the real distribution for the duration time, there must be a large number of sub-states. In other words, if one of the states does not follow an exponential distribution, it can be divided into multiple sub-states. By creating extra sub-states, the distribution of time spent in a behavioural state can be controlled. When the behaviour time does not follow an exponential distribution, the true distribution can be accurately approximated through the use of several sub-states. However, even if the behavioural state is expanded to just two sub-states,this model will still be more flexible than the model without any sub-states.

7.2.1 The method

Langrock et al. (2012) added several sub-states for each behaviour. They divided the non-geometric behavioural states into 30 sub-states, and the process progressed from each sub-state into the next until the process left one aggregate to transfer to the other. Because Langrock et al. (2012) used a discrete-time model (HMM), its formulation is slightly different than the OU process. The inference computation for 30 sub-states presents a problem for the OU process, as the large number of sub-states requires a lot of parameters, and this makes the inference very slow. In our new approach, used throughout this chapter, we assume that each behaviour is represented by only two sub-states. The use of two sub-states provided flexibility, and sufficient for most purposes. Further, even dividing the behaviour state into two sub-states can offer a worthwhile improvement. Expanding the behaviour state

into more than two sub-states is straightforward, but we use only two sub-states to save on computational costs.

In the simplest case, where the process enters an aggregate sub-states, the process should go through all the sub-states in series with a fixed transition rate for all the sub-states $\lambda_{ij} = \lambda$. The animal must cycle through all the sub-states before switching to a different behaviour state. For example, when the animal enters the substate feeding A, the animal can't switch to different behavioural state until it visits sub-state feeding B. Then, from the substate feeding B, it can switch to a different behaviour state. In this example, the generator would appear as follows:

$$G = \begin{pmatrix} -\lambda_{fa,fb} & \lambda_{fa,fb} & 0 \\ 0 & -\lambda_{fb,E} & \lambda_{fb,E} \\ \lambda_{E,fa} & 0 & -\lambda_{E,fa} \end{pmatrix} \tag{7.1}$$

If $\lambda_{fa,fb} = \lambda_{fb,E}$, the duration time for the feeding state is a gamma distribution. In this example, the animal would begin their duration-time in the exploratory state. They would then transition to the feeding A sub-state, followed by the feeding B sub-state, before returning to the exploratory state. This one route case can be generalised to the case where each sub-states has a different transition rate $\lambda_{fa,fb} \neq \lambda_{fb,E}$ In such case, the duration time in the behavioural state follows hypo-exponential distribution.

The alternative to the one route case is a more flexible case in which, when the process enters a sub-state, it can switch to the next sub-state or switch out of this behaviour state. The duration distribution for the behaviour state in this case would follow a phase type distribution. For example, from the exploratory state, the animal can transition to feeding A. Moreover, if the behaviour transitions to feeding A, it could either move forward to feeding B, or leave the feeding behaviour state by switching back over to the exploratory state. The generator matrix for this case would be flexible.

$$G = \begin{pmatrix} -\sum_i \lambda_{fa,i} & \lambda_{fa,fb} & \lambda_{fa,E} \\ 0 & -\lambda_{fb,E} & \lambda_{fb,E} \\ \lambda_{E,fa} & 0 & -\lambda_{E,fa} \end{pmatrix} \tag{7.2}$$

In this case, time spent in the feeding behaviour state follows a phase-type distribution composed of a combination of exponential distributions. Any continuous distribution can be approximated in this way if there are enough sub-states. If the number of sub-states is large enough, duration-time distribution can be estimated regardless of distribution shape. ˜

To fit a continous time animal movement model using this idea, we assume that we have a homogenous model (for more details, see Section 3.4.1). The movement process has an OU process for each sub-state. The two sub-states representing the same behaviour share the same movement parameters, but for the behaviour process each substate had different transition rates. For example, there are two behavioural states (feeding and exploring), and feeding has two sub-states (feeding A and feeding B). The movement process has three OU processes: for the exploring state, feeding A and feeding B. However, the two OU processes for the sub-states feeding A and feeding B will share the same movement parameters $\mu_{f_a} = \mu_{f_b}$, $b_{f_a} = b_{f_b}$ and $v_{f_a} = v_{f_b}$. To examine the new idea, even when the data is Markovian and the time spent in a specific behaviour follows an exponential distribution, we can fit the data using the group of sub-states as a special case. If the state is Markovian, then the transition rate for the second sub-state will be close to zero or, at the least, small.

7.3 Implementation

In this chapter, we use the variable kappa approach by treating kappa as a parameter. For more details, see Section 5.2.2. The implementation in this chapter is similar to that of previous chapters; updating the movement parameters and the trajectory is similar to the description in Section 3.4. The Metropolis Hastings algorithm was used for updating the behaviour parameters in a similar way to the variable kappa approach seen in Section 5.3. The algorithm was run for 10 million iterations. We ran each experiment for 10 million iterations using a thinning ratio of 1000 and a burn-in period of 5000 iterations after thinning.

7.4 Simulation experiments

In this section, there are three homogenous simulated datasets: a two state simulated dataset, a three state simulated dataset and a four-state simulated dataset. The details are given in Sections 7.4.3 to 7.4.5 below. We fitted the same model for all of the three datasets. The model has two behavioural states, (a and b), and each behavioural state is divided into two substates, (a_1, a_2 and b_1, b_2), respectively. This was done by implementing a state-aggregate method with two aggregates, wherein each set of two sub-states represents one behaviour. The model allows for the selection of the duration-time spent in each state. The two sub-states representing the same behaviour share the same movement parameter, but have different

transition rates. Thus, each state has two sub-states, Figure 7.1, with the generator:

$$G = \begin{pmatrix} -\sum_i \lambda_{a1,i} & \lambda_{a_1,a2} & \lambda_{a_1,b_1} & 0 \\ 0 & -\lambda_{a_2,b_1} & \lambda_{a_2,b_1} & 0 \\ \lambda_{b_1,a_1} & 0 & -\sum_i \lambda_{b_1,i} & \lambda_{b_1,b_2} \\ \lambda_{b_2,a_1} & 0 & 0 & -\lambda_{b_2,a_1} \end{pmatrix} \tag{7.3}$$

The data was fitted with a two aggregate substates OU process, with all the four substates sharing the same centre of attraction. The prior mean and SD for movement parameters for substate 1 equals the movement parameters for substate 2, sharing the same prior.

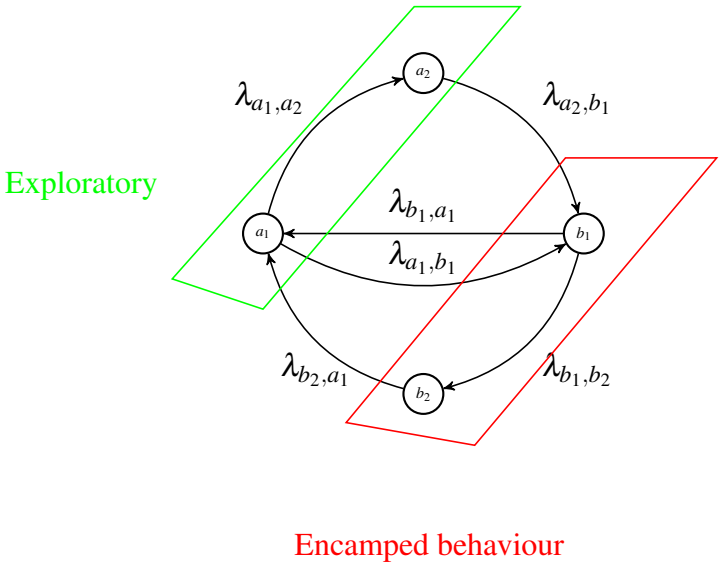

Fig. 7.1 Two behavioural state Markov chains with two sub-states for the simulated data. Left, aggregate states a_1 and $2a$. Right, aggregate states b_1 and b_2

The likelihood associated with λ_{a_2,b_1} is:

$$p(data|\lambda) = \left(\frac{\lambda_{a_2,b_1}}{\kappa}\right)^{n_{a_2,b_1}} \times \left(1 - \frac{\lambda_{a_2,b_1}}{\kappa}\right)^{n_{a_2,a_2}} \tag{7.4}$$

where n_{ij} is the number of switches from sub-state i to sub-state j. This is similarly the case for λ_{b_2,a_1}, because the likelihood for these states is binomial, while for λ_{a_1,a_2} and λ_{a_1,b_1}, the

likelihood is multinomial :

$$p(data|\lambda) = \left(\frac{\lambda_{a_1,a_2}}{\kappa}\right)^{n_{a_1,a_2}} \times \left(\frac{\lambda_{a_1,b_1}}{\kappa}\right)^{n_{a_1,b_1}} \times \left(1 - \frac{\lambda_{a_1,a_2} + \lambda_{a_1,b_1}}{\kappa}\right)^{n_{a_1,a_1}} \tag{7.5}$$

similarly, this is the case for λ_{b_1,a_1} and λ_{b_1,b_2}.

After reaching the posterior result, the duration-time distribution was reviewed within each behaviour by plotting the distribution of each posterior. The Bernoulli distribution was used so that substate a_1 could transition into substate a_2 or b_1. As such, the Bernoulli model yielded 1 if substate a_1 transitioned to sub-state a_2 and 0 if it transitioned to sub-state b_1. Simulation of duration time in a phase type- process.

- Simulate $X_i \sim \text{Exp}(\lambda_{a_1,a_2} + \lambda_{a_1,b_1})$.
- Simulate $B_i \sim \text{Bernoulli}\left(\frac{\lambda_{a_1,a_2}}{\lambda_{a_1,a_2}+\lambda_{a_1,b_1}}\right)$.
- Simulate $Y_i \sim \text{Exp}(\lambda_{a_2,b_1})$.
- $Z_i = \begin{cases} X_i + Y_i & B_i = 1 \\ X_i & B_i = 0 \end{cases}$

This prompted an idea regarding the duration-time distribution shape of the posterior distribution. The duration-time distribution was compared against an exponential distribution with the same mean, and the mean was calculated from the simulation sample. The density was then plotted for the exponential with the same mean. This made it easier to see the differences between the estimated distribution and exponential distribution of the posterior mean. Also, a QQ-plot provides another method of comparing the estimated result with the real exponential distribution.

7.4.1 Two-state Markov model

Simulated data

First, the OU model was used to simulate data from the homogenous model. The behavioural process was demonstrated through the use of a two-state Markov process, wherein the time spent in each state followed an exponential distribution with the generator:

$$G = \begin{pmatrix} -0.3 & 0.3 \\ 0.5 & -0.5 \end{pmatrix} \tag{7.6}$$

The movement parameters shared the centre of attraction for both states $\mu_x = 95$, $\mu_y = 55$, $b_a = 0.00023$, $b_b = 0.0033$, $v_a = 3$, $v_b = 75$.
The simulation was completed with observations every 1 time units, providing 901 observations in total.

Results

After running the MCMC algorithm described above; the mean posterior transition rates as follows:

$$\lambda = \begin{pmatrix} -0.24 & 0.09 & 0.15 & 0 \\ 0 & -0.01 & 0.01 & 0 \\ 0.17 & 0 & -0.2 & 0.03 \\ 0.025 & 0 & 0 & -0.025 \end{pmatrix} \tag{7.7}$$

The model results indicate that the λ_{a_1,a_2} and λ_{b_1,b_2} values are very small, close to zero. The results suggest that this is a two-state model without the need to expand into substates, wherein the duration-time for each state represents the exponential. When the animal is in substate a_1, it does not have to switch to substate a_2, but it can leave this behaviour and switch directly to the next behaviour b_1. This allows for an exponential distribution for the total time in a behaviour.

The posterior means for the movement parameters vary between the two behaviours. b_i and v_i are shown in Figure 7.2. There are clear differences in the movement parameters between the two behaviours. The black dot is the real value, and the posterior results show estimates for the movement parameters as coloured points. The estimation for the movement parameters was accurate but not precise.

Figure 7.3 shows the QQ-plot to examine if the time spent in each behaviour follows the exponential distribution. Figure 7.4 shows a histogram defining the time spent in each behaviour. The red line is the exponential distribution with a rate=$\frac{1}{\text{mean}}$. Figure 7.4 and Figure 7.3 shows that states a and b are Markovian and follow an exponential distribution. The message from Figure 7.4 and Figure 7.3 is that even though the parameter estimation is not very good, the distribution of durations using those parameters is quite close to the truth.

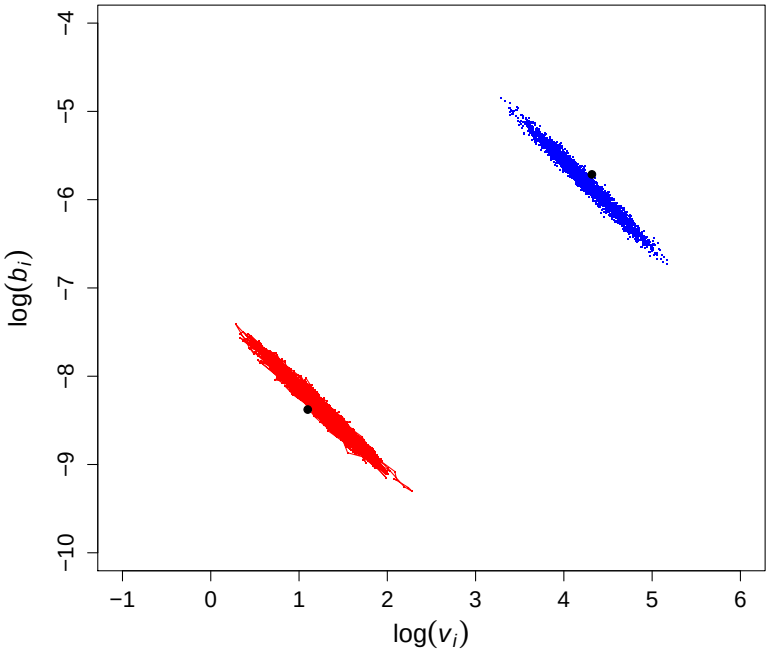

Fig. 7.2 Posterior movement parameters for the simulated dataset with two states. The two clusters correspond to behaviour states a and b (red and blue, respectively).

7.4.2 Three-states Markov model

Simulated data

The second simulated data-set was conducted through the simulation of a three-states model, wherein two of the states represent the same behaviour. This is known as a combined state. The behavioural process has given by a three-state Markov process in which the duration-time followed an exponential distribution, with the generator:

$$G = \begin{pmatrix} -0.3 & 0.3 & 0 \\ 0.4 & -0.6 & 0.2 \\ 0.3 & 0 & -0.3 \end{pmatrix} \tag{7.8}$$

The movement parameters were sharing the centre of attraction for three states $\mu_x = 95$, $\mu_y = 55$, movement parameters for state 2 equal the movement parameters for state 3 because they represent the same behaviour. $b_1 = 0.00023$, $b_2 = b_3 = 0.0033$, $v_1 = 3$, $v_2 = v_3 = 75$. The simulation was run with observations every 1 time units, giving 901 observations in total. The 4-state model defined in Equation 7.3 was fitted, as before. The MCMC algorithm was run for 10 million iterations.

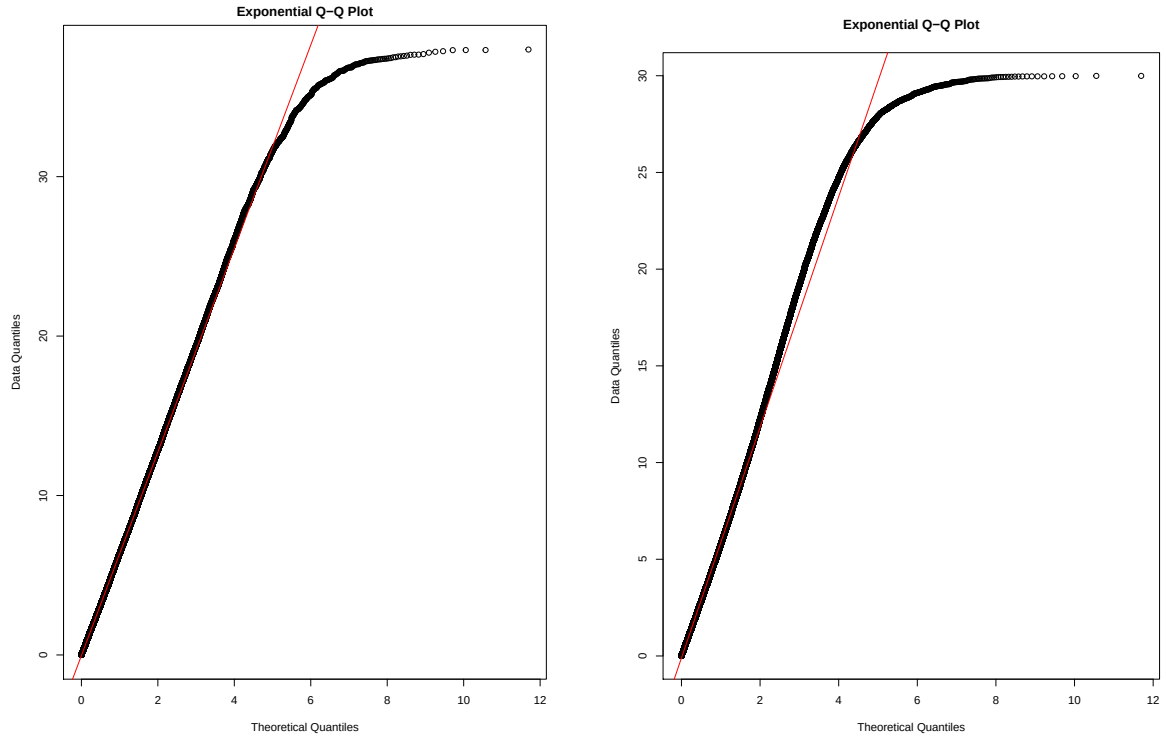

Fig. 7.3 Exponential QQ-plot for the posterior result for both behaviour for the simulated dataset with two states. The left hand graph shows the aggregate state associated with a_1 and a_2, the right hand graph the aggregate state associated with b_1 and b_2.

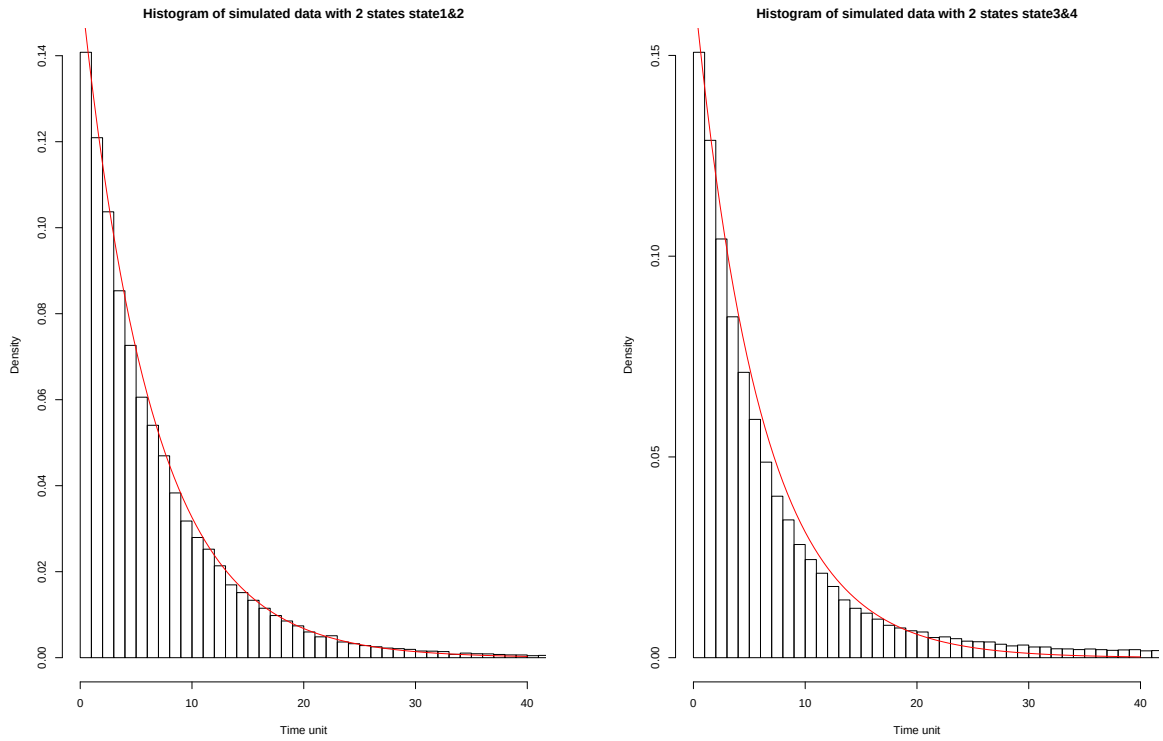

Fig. 7.4 Histogram for the posterior time spent in each behaviour for the simulated dataset with two states. The left hand graph shows the aggregate state associated with a_1 and a_2, and the right hand graph the aggregate state associated with b_1 and b_2. The red line is the exponential distribution.

Results

Thus, each state has two sub-states with the mean posterior transition rate generator:

$$\lambda = \begin{pmatrix} -0.25 & 0.02 & 0.23 & 0 \\ 0 & -0.13 & 0.13 & 0 \\ 0.19 & 0 & -0.56 & 0.37 \\ 0.22 & 0 & 0 & -0.22 \end{pmatrix} \tag{7.9}$$

The model results show that $\lambda_{a_1,a_2} = 0.02$ is small. This means that there are very few switches to substate a_2. Meanwhile, $\lambda_{b_1,b_2} = 0.37$ is large. This means that there are many switches to substate b_2. This suggests that the process is a two-states model, wherein substate a_1 is a regular Markov state without the need to extra substate while substates b_1 and b_2 use an sub-states aggregate model, and the duration time in total b state do not follow exponential distribution. When the animal is in substate a_1, it does not have to transition to substate a_2, but can go directly to substate b_1. This allows for an exponential distribution for the duration time in behaviour a. So the structure in the estimated model matches the structure used to simulate the data. The posterior mean for the movement parameters varies between the two behaviours. b_i and v_i are shown in Figure 7.5. There are clear differences in the movement parameters between the two behaviours. The black dot is the real value, and the posterior results are shown as colored dots. Figure 7.6 shows the QQ-plot examining whether or not the dwell-time of each behaviour follows an exponential distribution. Figure 7.7 and Figure 7.6 shows that duration time for state a follow exponential distribution while duration time for state b is not exponential. Figure 7.7 shows a histogram of the time spent in each behaviour. The red line is the exponential distribution with rate=$\frac{1}{\text{mean}}$.

7.4.3 Four-states Markov model

Simulated data

This simulated datasets is similar to the bison example in the next section. A four-states model was simulated, wherein each pair of two states represented one behaviour. The OU model with a four-state Markov process was implemented, and the duration-time of each

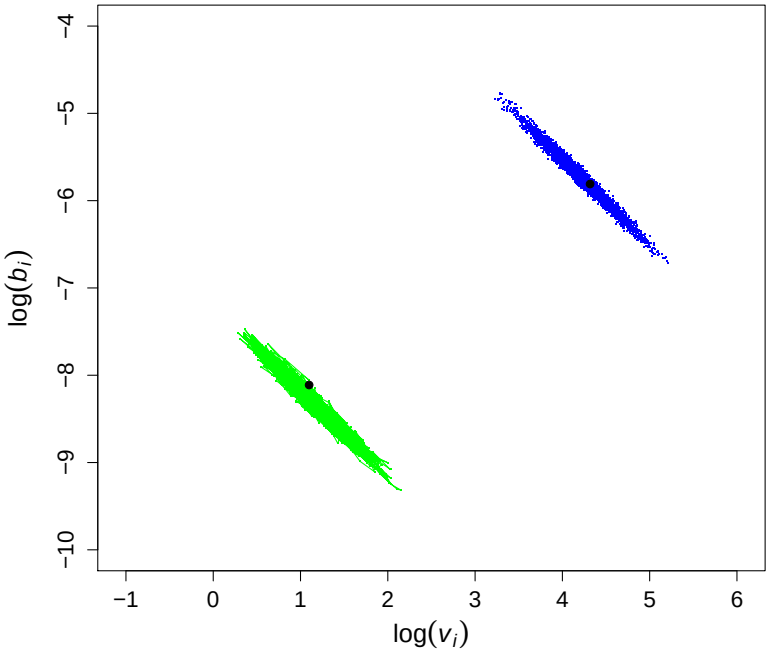

Fig. 7.5 Posterior Movement parameters for the simulated dataset with three states. The two clusters correspond to states a and b (green, blue respectively).

state followed an exponential distribution with the generator:

$$G = \begin{pmatrix} -0.5 & 0.3 & 0.2 & 0 \\ 0 & -0.2 & 0.2 & 0 \\ 0.3 & 0 & -0.7 & 0.4 \\ 0.5 & 0 & 0 & -0.5 \end{pmatrix} \tag{7.10}$$

The OU parameters shared the centre of attraction in the four-states model $\mu_x = 95$, $\mu_y = 55$. Movement parameters for state 3 were equal to the movement parameters of state 4 because they represent the same behaviour. The movement parameters of state 1 are also equal to the movement parameters of state 2, as they represent the same behaviour $b_1 = b_2 = 0.00023$, $b_3 = b_4 = 0.0033$, $v_1 = v_2 = 3$, $v_3 = v_4 = 75$. The simulation was completed with observations every 1 time units, providing 901 observations in total. The estimation algorithm used was run for 10 million iterations.

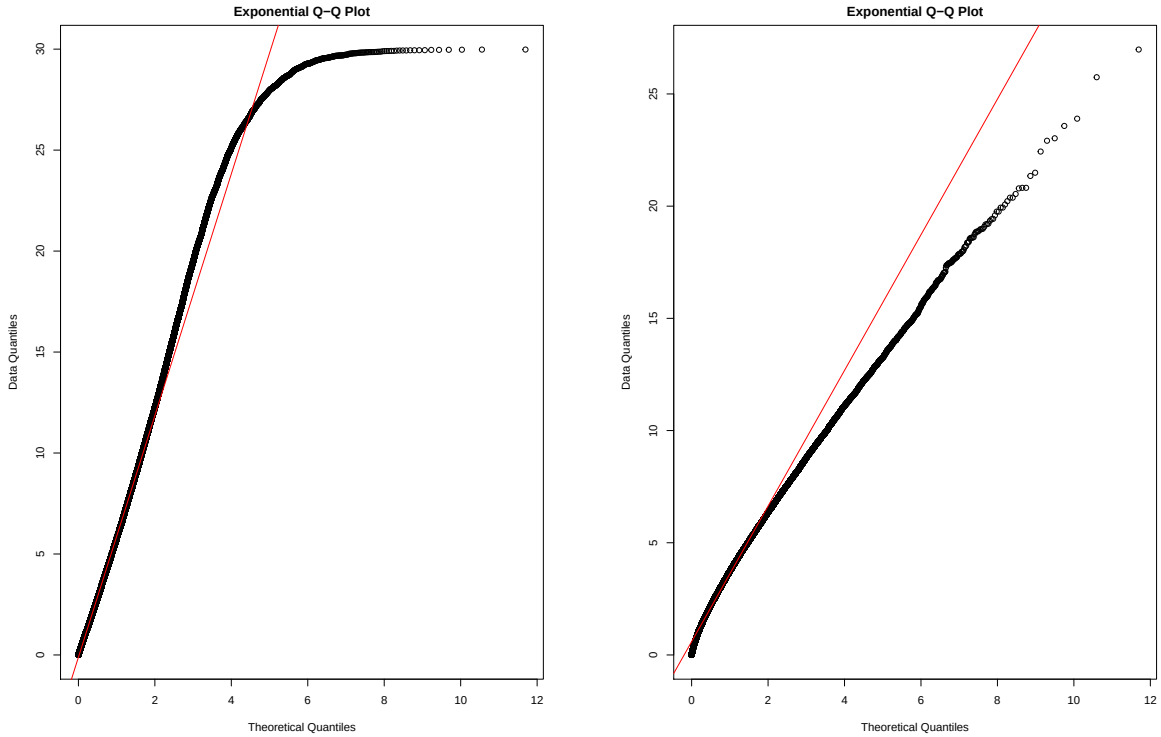

Fig. 7.6 Exponential QQ-plot for the posterior result for both behaviour for the simulated dataset with three states. The left hand graph shows the aggregate state associated with a_1 and a_2, the right hand graph the aggregate state associated with b_1 and b_2.

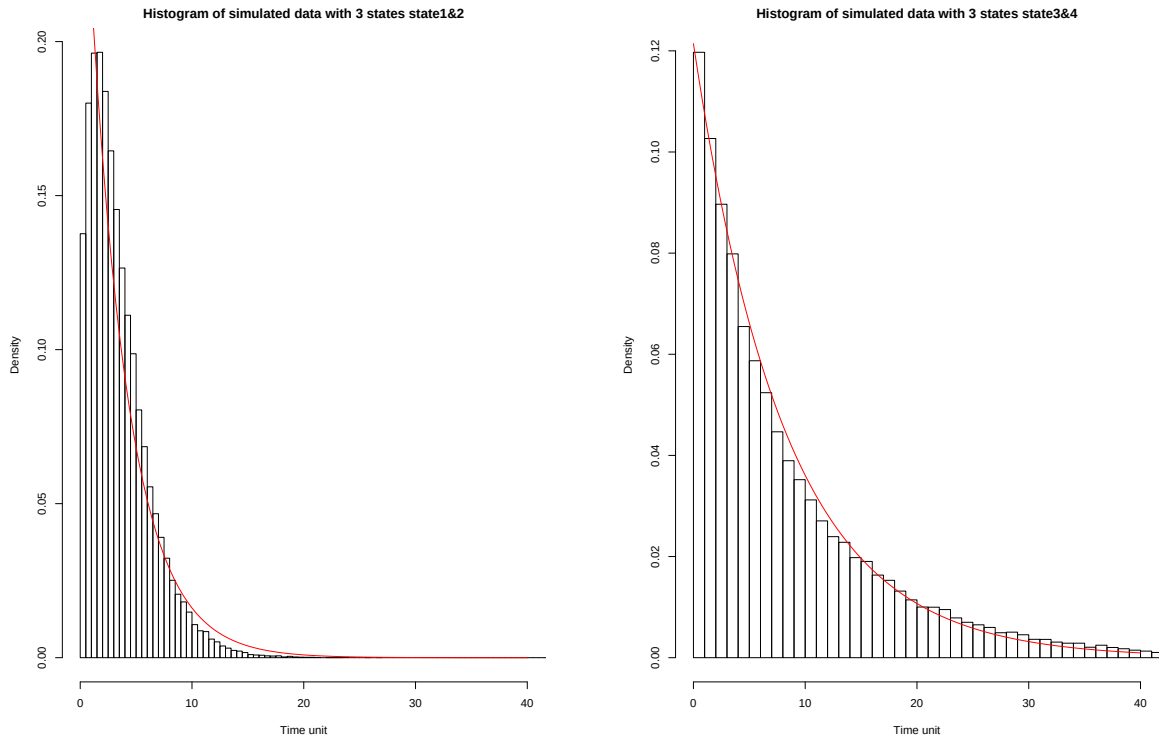

Fig. 7.7 Histogram for the time spent in each behaviour for the simulated dataset with three states. The left hand graph shows the aggregate state associated with a_1 and a_2, the right hand graph the aggregate state associated with b_1 and b_2, the red line is the exponential distribution

Results

We fit the 2 state model in which each state has two sub-states, with the mean posterior transition rate generator:

$$\lambda = \begin{pmatrix} -0.36 & 0.27 & 0.09 & 0 \\ 0 & -0.21 & 0.21 & 0 \\ 0.08 & 0 & -0.25 & 0.17 \\ 0.35 & 0 & 0 & -0.35 \end{pmatrix} \tag{7.11}$$

The posterior distributions of the movement parameter varied between the two behaviours.

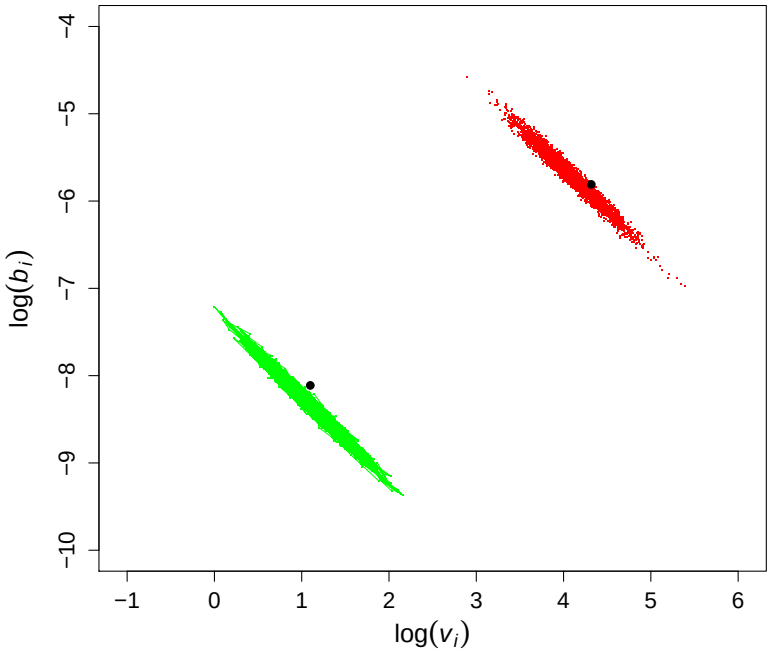

Fig. 7.8 Posterior movement parameters for the simulated dataset with four states. The two clusters correspond to states a and b (red, green respectively).

b_i and v_i are shown in Figure 7.8. There are clear differences in the movement parameters between the two behaviours. The black dot is the real value, and the posterior results shows an colored dots.

Figure 7.9 shows the QQ-plot examining whether or not the duration-time of each behaviour follows an exponential distribution. Figure 7.10 shows a histogram of the time spent in each behaviour for the four states simulated data. The red line is the exponential

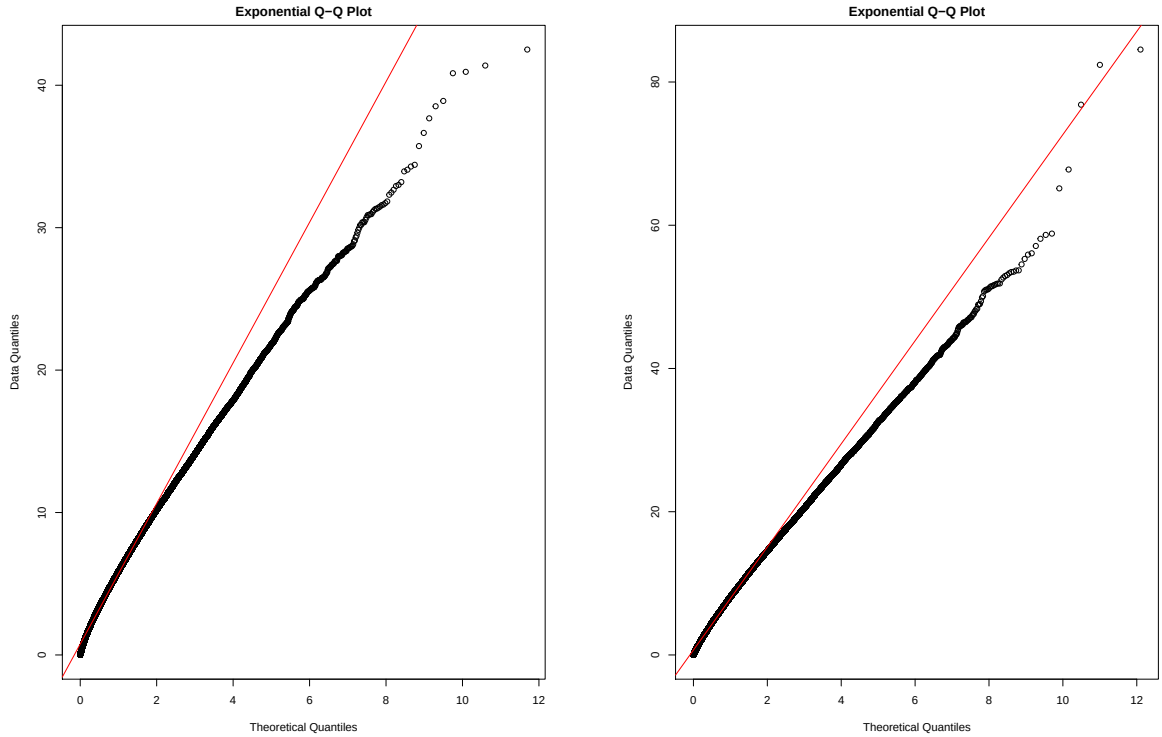

Fig. 7.9 Exponential QQ-plot for the posterior results for both behaviour for the simulated dataset with four states. The left hand graph shows the aggregate state associated with a_1 and a_2, the right hand graph the aggregate state associated with b_1 and b_2.

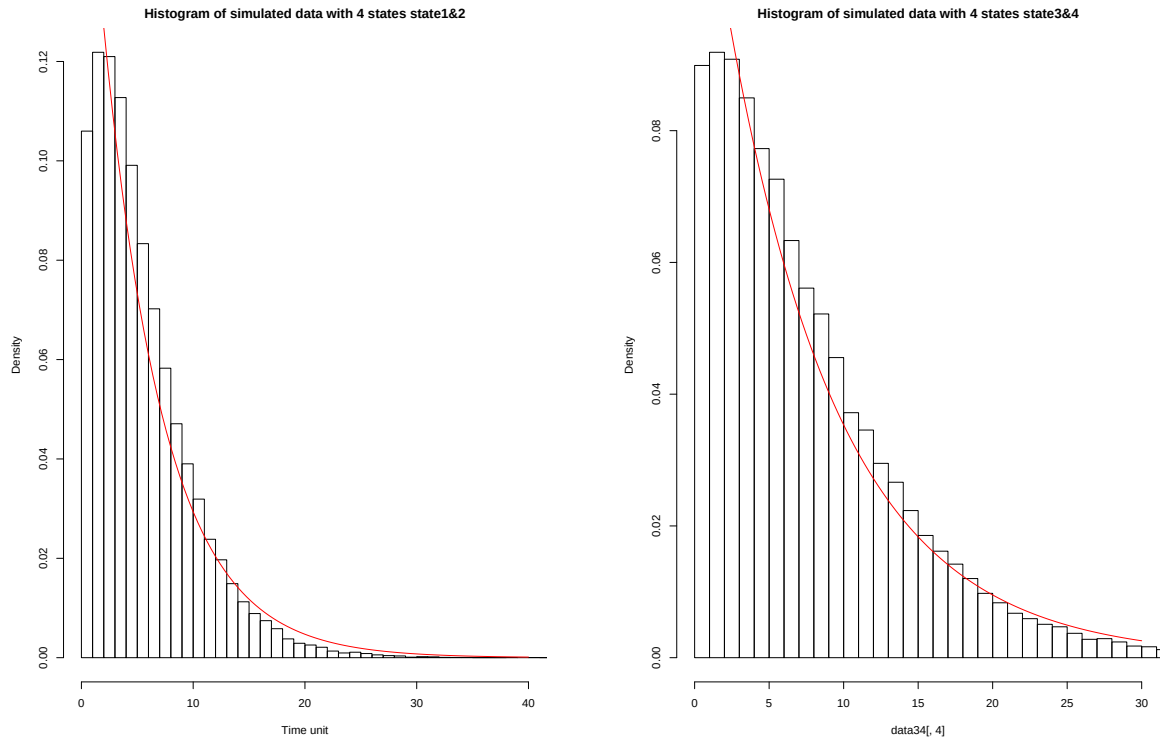

Fig. 7.10 Histogram for the posterior time spent in each behaviour for the simulated dataset with four states. The left hand graph shows the aggregate state associated with a_1 and a_2, the right hand graph the aggregate state associated with b_1 and b_2. The red line is the exponential distribution.

distribution. Figure 7.10 and Figure 7.9 shows that state *a* and *b* are not Markovian and the lengths of time spent there do not follow an exponential distribution.

7.5 Application

7.5.1 Bison data

American bison were observed using GPS radio collars within the Prince Albert National Park in Saskatchewan, Canada. This study was conducted between October 2005 and April 2006. The locations of the bison were collected every hour. For more details about the data see; Babin et.al. (2011). Langrock et al. (2012) used this datasets but with three-hour intervals to ensure the data followed regular time-intervals. There are some missing data so they thinned the data to a three-hour interval to solve this problem. This is because the discrete time approach cannot address irregular data, for more details see Section 2.5. While Langrock et al. (2012) analysed the movement paths of nine bison but each animal separately, this chapter focuses on the movement of only one bison. There were 1654 observations. A plot of the bison movement is provided in Figure 7.11. Langrock et al. (2012) suggest that the bison displayed two behaviours: encamped with slower movement, and exploratory with faster movement.

Langrock et al. (2012) choose to use part of the dataset that ware collected over one winter. This is because though the behaviour of bison varies within the season, it is reasonable to assume that it is homogeneous throughout the winter. In other words, the animal will not change its behaviour over time within this season. For example, the transition rate will change with the season over a full year, but the movement parameters and transition rate during the winter will remain constant. The same data as in Langrock et al. (2012) are used here to compare the results of our new approach with their results.

7.5.2 Approaches to analysis

The proposed model will be spatially homogeneous. There will be a single transition generator matrix, and the behavioural states will be exploratory and encamped. Langrock et al. (2012) suggested the bison in these data have two behaviours and divided each behaviour into 30 sub-states. Then they used a hidden Markov model to fit the data. Here, we assume the same thing, that the bison has only two behaviours, but we expanded each behaviour to only two sub-states. This is because using the OU model, the 30 sub-states will complicate and slow the algorithm. Also, even dividing the behaviour state to only two sub-states can

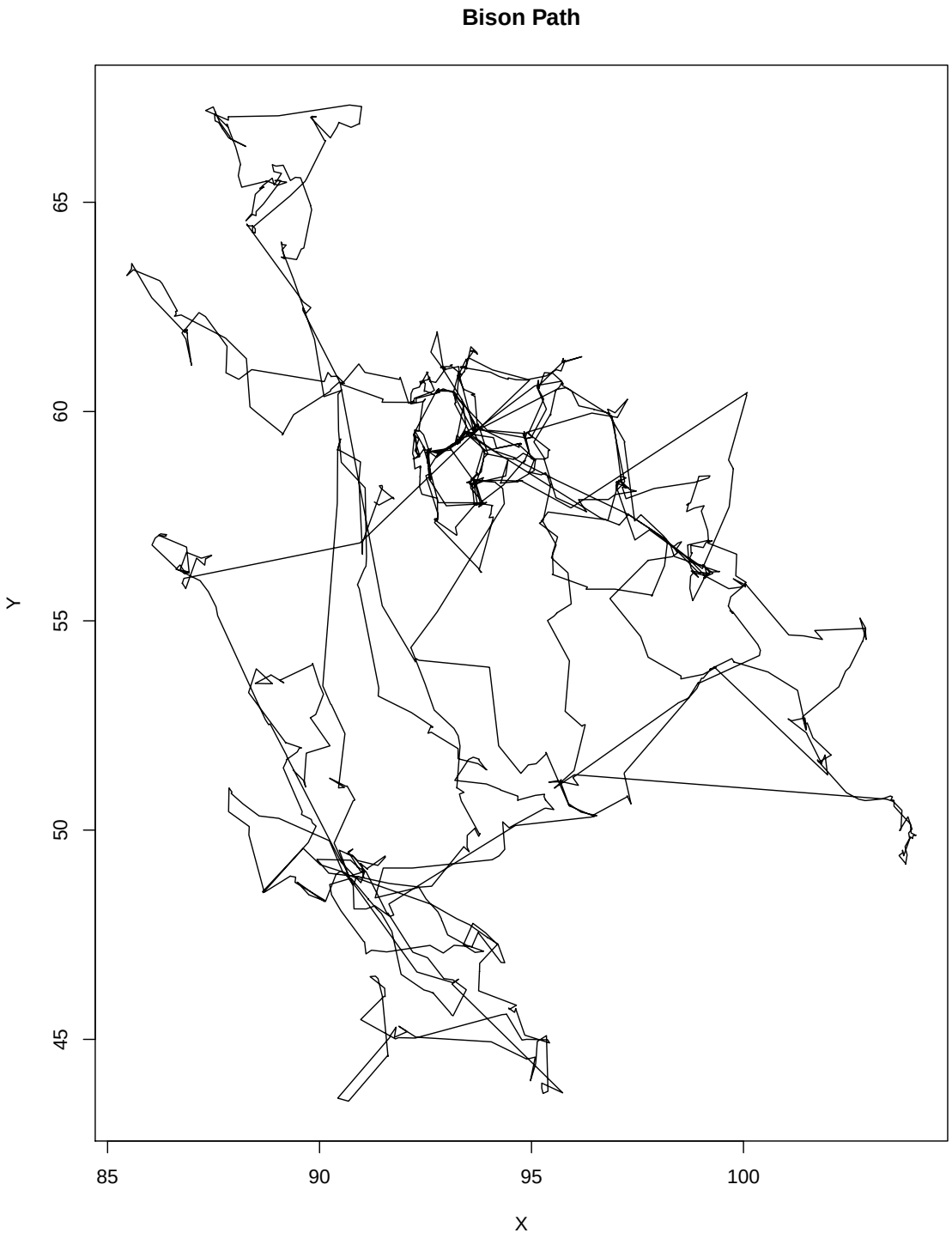

Fig. 7.11 Observed movement path for the bison.

give a worthwhile improvement. We will use the same model describe in the simulated experiment with 2 states. Each state has two substates, and the substate share the same OU process. For more details see Section 7.4.

7.5.3 Results

The movement parameters of state a_1 were equal to the movement parameters of state a_2, as they shared the same behaviour. The movement parameters of state b_1 were equal to the movement parameters of state b_2.
The posterior mean for the transition rates generator matrix is:

$$\lambda = \begin{pmatrix} -0.49 & 0.3 & 0.19 & 0 \\ 0 & -0.26 & 0.26 & 0 \\ 0.27 & 0 & -0.3 & 0.03 \\ 0.19 & 0 & 0 & -0.19 \end{pmatrix} \tag{7.12}$$

The kappa posterior mean is $\kappa = 0.61$.
From the above matrix it can be noted that, λ_{a_1,a_2}, λ_{a_1,b_1} are large which means there are many switches to state a_2 and b_1 from state a_1, while the transition rate $\lambda_{b_1,b_2} = 0.03$ is small, which means very few switches happen from state b_1 to state b_2. That means most of the time the process travels through only 3 states which are a_1, a_2 and b_1 with very few visits to state b_4. That means the first a behaviour does not follow an exponential distribution and needs an aggregate state to fit it. And the second b behaviour is following an exponential distribution and does not need any sub-states.

The result of Langrock et al. (2012) suggest that the duration time for the encamped state follows a geometric distrbution, while the duration time for the exploratory state follows a negative binomial distribution. The first behaviour was Markov, while the second behaviour was not, and that agrees with our results. As this study did not implement the same model used by Langrock et al., there was no guarantee that the same distribution would be obtained. As with the results in the previous chapters, there was a strong correlation between the b and v movement parameters. The posterior distributions of the parameters vary between b_i and v_i. They are shown in Figure 7.12. There are clear differences in the movements parameters of the two behaviours. The parameters of states b_1 and b_2 are less accurately estimated due to the small amount of time spent in those states.

Figure 7.13 shows the QQ-plot that examines whether or not the time spent in each

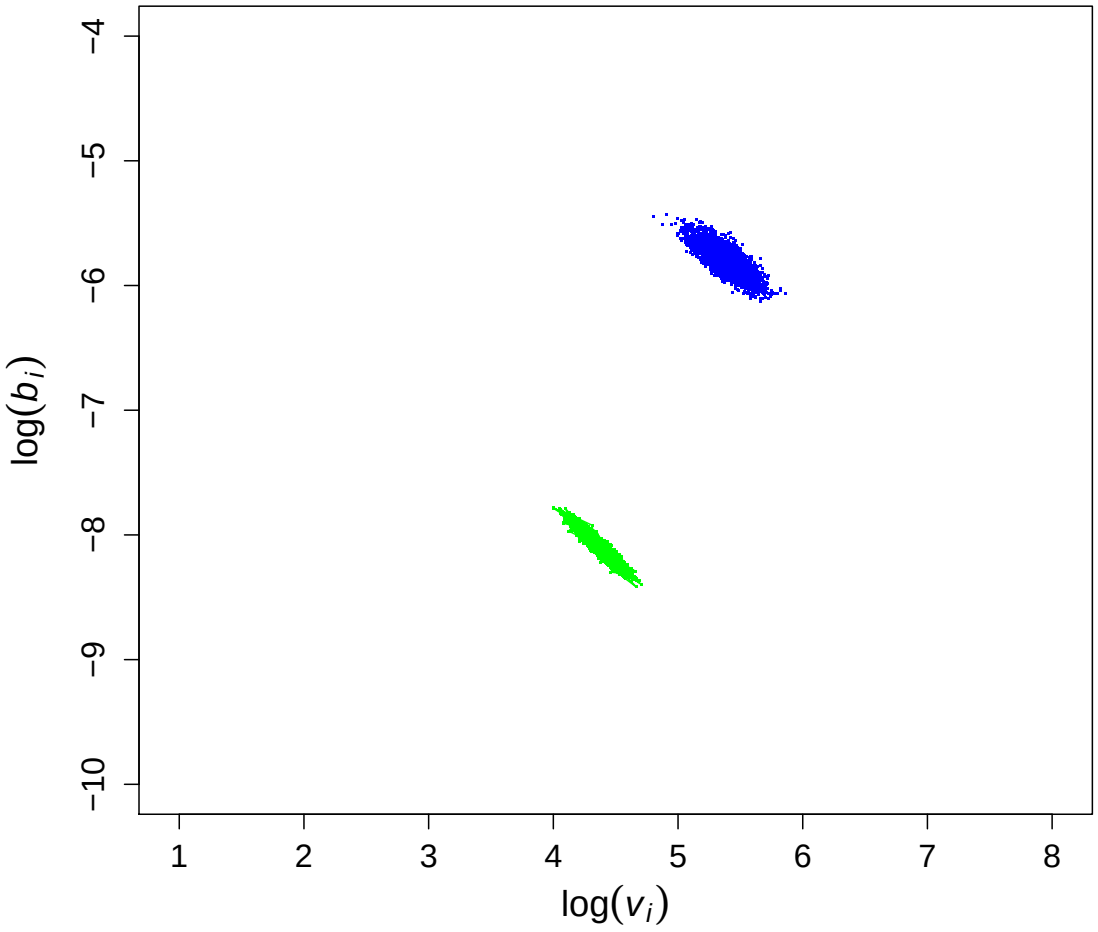

Fig. 7.12 Posterior movement parameters for the bison data. The two clusters correspond to state a and b (green, blue, respectively).

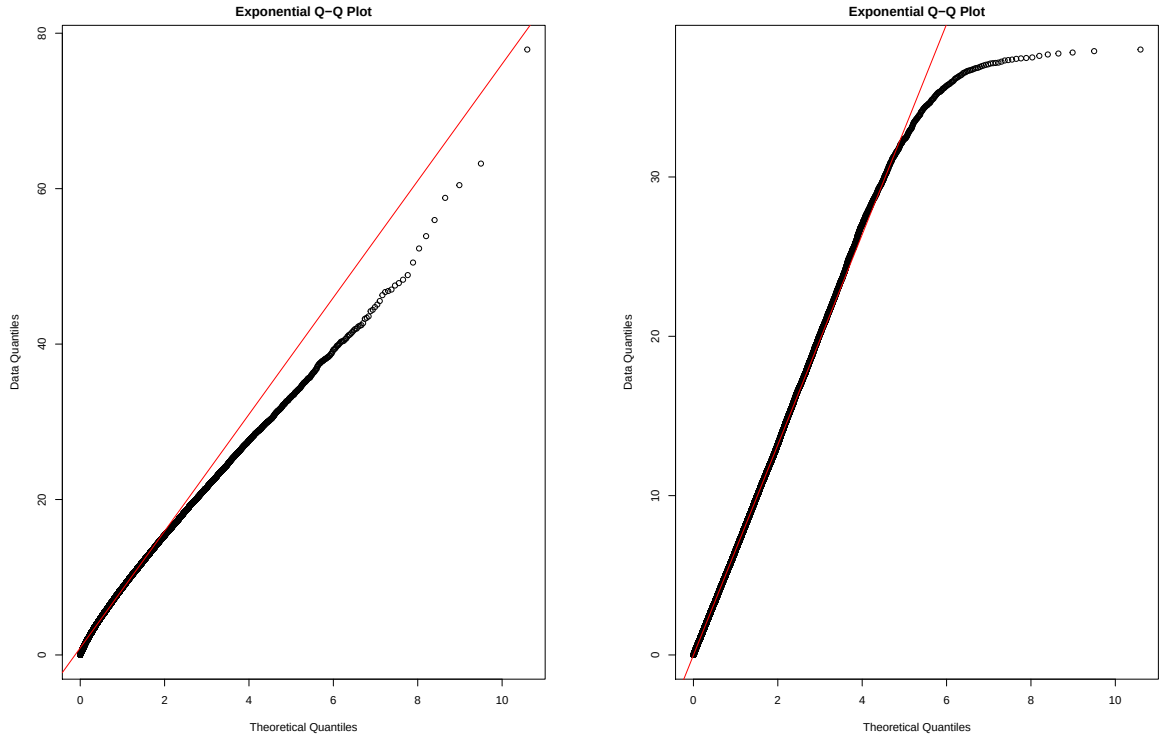

Fig. 7.13 Exponential QQ-plot for the posterior results for both behaviours for the bison data. The left hand graph shows the state associated with a_1 and a_2, and the right hand graph shows the state associated with b_1 and b_2.

behaviour follows the exponential distribution for the bison data. It is clear that the combination of states a_1 and a_2 do not follow an exponential distribution (left), while the combination of states b_1 and b_2 follow an exponential distribution with a few outliers (right). This plot allows for the comparison of estimated distributions for the time spent within a state against an exponential distribution. Figure 7.14 shows a histogram of the duration-time for each behaviour. The red line is the exponential distribution with a rate=$\frac{1}{\text{mean}}$. The histogram represents the time spent in each behaviour, as well as its density and probability. It is clear that the combination of sub-states a_1 and a_2 do not follow an exponential distribution (left), while the combination of sub-states b_1 and b_2 follows an exponential distribution (right).

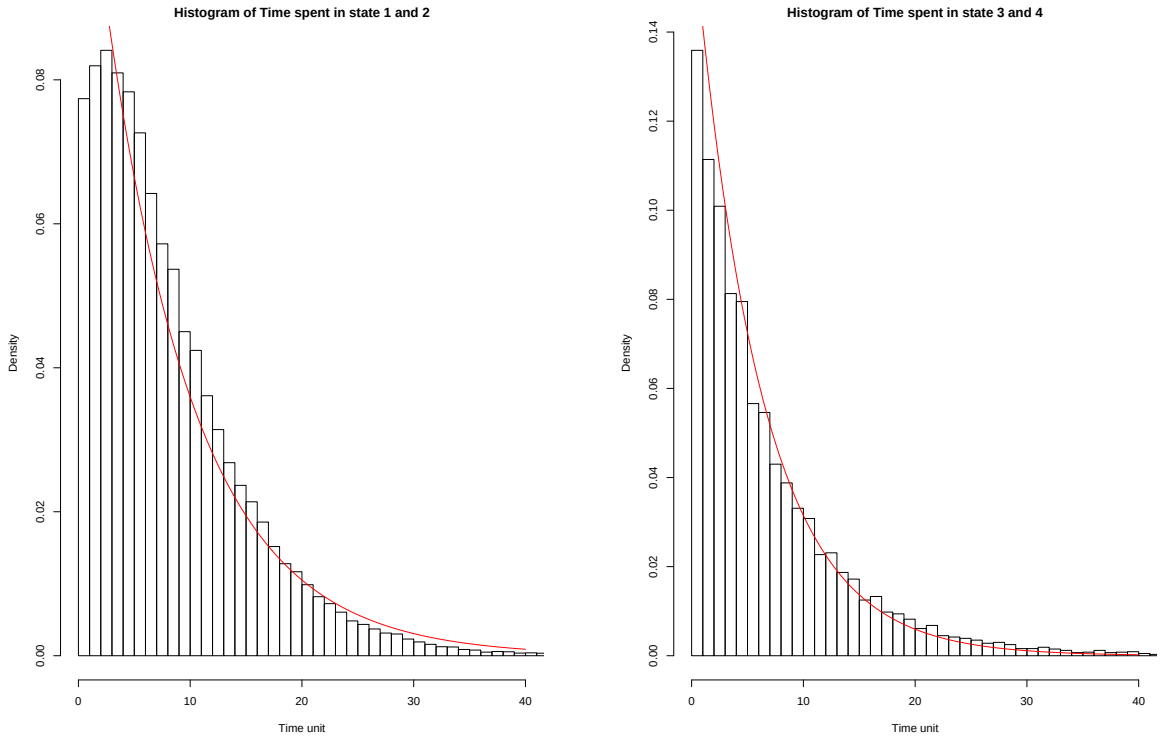

Fig. 7.14 Histogram for the posterior time spent in each behaviour for the bison data. The left hand graph shows the state associated with a_1 and a_2, and the right hand graph shows the state associated with b_1 and b_2. The red line is the exponential distribution.

7.6 Discussion

A limited property of the Markov process is that the duration-time distributions for behavioural states must follow an exponential distribution in the continuous time. Relaxing this assumption leads to a loss of the Markovian property, making the analysis more difficult.

This chapter has considered how to address this limitation by allowing the duration-time to follow any continuous distribution, while simultaneously maintaining the Markov property. This was achieved by expanding the existing behaviour state to multiple sub-states.

In the simulated experiments, three different simulated datasets were used to examine the new idea and all of them gave reasonable results. Then real data were used to compare the results we get with other studies that used the same data and the same technique but in a discrete time model, and the result was satisfactory.

The advantage of the proposed model is that it provides more flexibility and generality, while also saving on computational costs when compared against semi-Markov models. Furthermore, this method maintains the Markov property, which is lost when the semi-Markov model is used. Losing the Markov property leads to difficulties in fitting the model and the inference, despite the Markov model being straightforward in principle.

This extension could be an important innovation within the field of ecology, as it not only improves the model in a realistic way, but also has the potential to bring forth important behavioural insights that basic Markov models cannot provide, as it allows the aggregate behaviours in the model more flexibility in matching actual behaviours based on observed movement.

Chapter 8

Conclusion

8.1 Summary

Animal movement data represent animal locations and their observation times. Global Positioning System (GPS) and radio tracking technology used in animal tracking have produced complex datasets of animal movement behaviour. The availability of high-resolution datasets is the result of these developed telemetry technologies, which are used to describe specific characteristics of wildlife movement and are linked to behaviour and environment. Researchers want to obtain more information from these data to help them understand not only about the animal location (home range, territory, etc.) but also about the animal dynamic (movement) itself. When an ecologist records animal locations using the capture-recapture idea or recent tracking technologies but the time interval between these observations is large, then these locations can typically be assumed to be independent. That is fine, if the ecologist is only interested in learning about where an animal wants to spend its time (home range, territory) or to simply study utilisation distribution. However, if the ecologist is interested in studying an animal's behaviour and the effect of the environment on its behaviour, then location observations should be close together in time, making the data autocorrelated. If the movement data are correlated, then a statistical approach is needed that allows for the autocorrelation between observations.

Most existing literature uses discrete time models to analyse the animal movement data. Ecologists find it easier to use discrete time models due to the availability of statistical software. There are some limitations of the discrete time models. The interpretation of the parameters depends on the time scale, which makes it difficult to compare different studies or to combine data that have different time intervals. Also, discrete time models cannot be used with irregular time data. In contrast, continuous time models can fit regular

and irregular time data, and the interpretation of the parameters does not depend on time. However, the continuous time models have two main limitations. First, the analysis of data using continuous time models is still quite complicated at the moment and most ecologists need assistance from statisticians. Second, there is a high computational cost when fitting continuous time models to large datasets. Ideally, one of the things researchers would like to achieve is to provide software that makes it easier for the ecologists to fit continuous time models. Unfortunately, the method is still in development currently, thus this is hard to accomplish.

Throughout this thesis, the OU diffusion process was used to model the animal movement data in continuous time. Dunn and Gipson (1977) were the first to present the OU process for the purpose of animal movement analysis and in order to deal with dependence between observations. Blackwell (1997, 2003) further developed the Dunn and Gipson idea to include switching between discrete modes of movement. Harris and Blackwell (2013) additionally extended the model by adding spatial heterogeneity and they also introduced the separable model to simplify inference and algorithm. Blackwell et al. (2016) introduced the exact Bayesian inference method to analyse the separable model and called it the exact method. This thesis attempted to extend and generalise the method presented by Blackwell et al. (2016).

For some cases, the exact method can be inefficient. Thus, this thesis attempted to improve the algorithm and extend the range of the models that can be fitted using this method. In practice, the goal of this thesis is to improve the efficiency of current algorithms and to allow more general models to be applied, and this thesis aimed to ensure that the exact method is practically feasible. This thesis shown some ways in which computational efficiency have been improved and some ways in which the models can be made much more general. It attempted to improve the efficiency of the computational costs of animal movement in a continuous time model as well as to increase the speed and the range of the model. The following is a summary of each experimental chapter (i.e. of each chapter that discussed how the aims of the thesis can be achieved).

In Chapter 4, the approximation method, the goal was to increase the efficiency of the current (exact) method. In the current method, the number of potential switches between observations are generated from the Poisson distribution. However, this chapter tried to approximate the number of switches in order to approximate the estimation of the posterior parameters by assuming that there is only one potential switch between each pair of

observations. This method was tested using simulated data. The results suggested that both methods (exact and approximation) give very similar results in terms of estimating the posterior parameters. However, the approximation method did not save time because other parts of the algorithm become more complicated. It can be concluded from the results of the approximation method chapter that some approximations can be safely conducted and results similar to those of the exact method can be obtained, but the method does not save as much time as hoped.

In Chapter 5, variable kappa, the aim was also to increase the efficiency of the current algorithm (exact method) as well as to broaden the range of possible priors for the transition rates. Blackwell et al. (2016) assumed that the transition rate has a global upper bound, which they defined as κ. Furthermore, they assumed that the kappa value is known and fixed even though kappa does not have any biological meaning. It was attempted to improve upon this method here by allowing kappa to vary rather than have a fixed value. This was achieved by treating kappa as a parameter and making it adjust to the value of the transition rates so that, when the transition rate is updated at every iteration, the kappa is updated as well. The reason for that is that there are a lot of potential switches between observations when kappa has a large value, which slows the algorithm. The variable kappa approach was used with various data, such as the fisher dataset (used for fitting the adaptive model) and some simulated data. The results showed that the variable kappa improves the current method by saving time and gives similar posterior results to the current approach while allowing unbounded prior distributions for all parameters.

In Chapter 6, estimating unknown boundaries, the purpose was to extend the current model to estimate the unknown boundaries between different regions. This boundary can be psychological, such as an animal's home range boundary or territory boundary, or a physical boundary of unknown habitat spatial structures. The boundary estimation idea increased the flexibility in modelling the ways in which an animal uses the space. The idea worked with two different data sets: first, the simulated dataset in which there was an unknown straight-line boundary between two regions; second, the real dataset for Ibex (a type of wild goat) in which a circular boundary between two regions was estimated. In estimating the boundaries between regions, more can be learned about the environment and how it affects the movement of the animal using it.

The exact method assumed that the duration time for each behaviour follows an exponential distribution. In Chapter 7, the current model was further extended to include application

in cases where behaviours have general continuous distributions in time. In practice, there are biological reasons for the duration of the time distribution to differ from the exponential distribution, such as an animal doing some task it needs to complete. When there is reason to think that an animal is exhibiting a particular behaviour that involves completing a task, for example feeding until it is no longer hungry, then semi-Markovian behaviour is expected. By applying the semi-Markovian model, a new model was developed here that has a more realistic behaviour pattern. This more flexible behavioural model lets researchers learn about how an animal behaves and the effect of this behaviour on its movement. These extensions were illustrated by presenting a re-analyses of two movement datasets, one simulated and one from a radio-tracking experiment on bison.

Continuous time modelling was successfully expanded in different directions by allowing the kappa to vary, allowing for semi-Markovian states and estimating unknown boundaries. It is hoped that these ideas will make modelling animal movement in continuous time more widespread. Statistical methods were developed to enable us to learn more from animal movement data. It is hoped that these new methods will help ecologists use movement data easily to learn more about animal behaviour and the environment.

8.2 Future work

In the approximation chapter, the hoped for improvements in speed were not obtained as the speed of the analysis did not increase by very much. However, the results did show that the underlying idea worked and that a very good approximation was obtained. Therefore, it may be possible to combine this approach in the future with other ideas about increasing the speed of the code. However, in isolation, it was shown that the approximation idea is not useful. Blackwell (2018) tried to use hidden Markov models to increase the speed of the exact analysis, which can be combined effectively with the approximation method. It is possible to combine the approximation method with other techniques to speed up the inference. Hence, these two separate improvements for the method could be combined in the future.

In the chapter that estimates the boundary, two examples were provided: the line boundary and the circle boundary. However, it would be possible to extend the model to include estimating the boundaries of multiple patches. It could be assumed that that the number of patches was known but that the location of these patches was unknown; thus, in the author's future work, an estimation of the location of the patches is going to be attempted based on

the movement data.

In the semi-Markov model chapter, it was assumed that the model is homogenous so that the transition rate does not depend on location. The model can be extended further by allowing the semi-Markov model to be non-homogenous and to have different transition rates for different regions. Extending the semi-Markov idea by allowing for spatial variation would increase the flexibility of the model. However, the difficultly lies in ensuring that the parameters are still interpretable. The semi-Markovian model could be extended through a combination with the spatial heterogonous case, which would result in an extremely flexible model with the potential for many parameters. Therefore, it would be necessary to restrict the parameters to maintain the interpretation and meaning of the model. If the model had a separate λ_{ij} for each region, then there would be too many parameters to estimate and the shape of the duration time distribution would vary, depending on the region in which the animal was located, thus complicating both the algorithm and the model. Therefore, this would be done only if there was a biological reason. The model should not have too many parameters to estimate, because the aim is to think about what an ecologist wants to represent as well as about the states and transitions that need to be considered. Reducing the number of parameters would be a natural way to simplify the model; in other words, some transitions would not be allowed.

Langrock et al. (2016) added a large number of sub-states to the semi-Markovian model. They allowed each state to have several sub-states in order to obtain a good approximation for the duration time distribution. For example, in the simplest model with two states, each one has a series of 30 sub-states inside a particular behavioural state. Their aim was to achieve a close mathematical approximation of a particular distribution. However, in most cases, it is not known in advance what this distribution should be. Hence, in practice, flexibility is needed to approximate a range of distributions. An expanded states idea is needed to check that the number of the sub-states required to estimate the duration distribution is enough. The model always yields better estimates when it includes more sub-states. However, the real question here is when the inclusion of additional states makes a difference? One possibility is to investigate how well the model is fitted. In the expanded states chapter, a simple model with two sub-states was applied to determine the distribution of the duration time. In future work, models could be compared to see how well they approximate the duration distribution for different numbers of sub-states. By fitting the same model to three sub-states and then to two sub-states, a comparison of the two duration distributions could help determine whether they have the same flexibility or not. The two cases would not have exactly the

same distributions, but they may be very close, in which case the simpler model would be preferred.

Resource selection

Animal movement is affected by habitat features such as food and water sources. In addition, barriers in the environment influence an animal's movement, including highways, fences, railroads and other barriers that limit animal movement. Scientists have extended the animal movement models to include habitat resource selection, barriers and measurement errors. The aim of the work by Brost et al. (2015) was to estimate the probability of resource use that was conditional on resource availability. In this type of research on resource selection, scientists often study complex maps on which there is a grid and separate resource measurements in each pixel. These maps are used to estimate how attractive each resource is to an animal.

In some studies, there may not be any information about the environmental resources or about which resources are important for an animal. Examples of such information include details about when an animal's movements show exploratory behaviour through the localisation of it slower movements, which may indicate that the animal wants to explore food or water in a patch of resources found in the habitat. Then, the analysis in the chapter estimating unknown boundaries can be extended to identify and estimate these resources.

In the discrete time model, using the hidden Markov model often allows the transition probabilities to depend on covariates. In future work, the continuous time model could be extended to allow the transition, depending on the spatial covariate, where λ_{ij} is a function of x. For example, $\lambda_{ij}(x,t)$ might be $\exp(\alpha_{ij} + \beta_{ij}\ x)$, where x is the distance from the current location to some particular point, such as a river, a food patch resource or water. The distance of an animal from this point would determine how likely the animal would be to switch to a particular behaviour. This example is similar to the idea proposed by Wang et al. (in preparation): a continuous time model was fit with covariates, but the switching model was restricted by assuming that an animal would always switch to the best centre of attraction, depending on the current location and the patch quality. In their research, Wang et al.'s study area had several patches and each patch had a centre of attraction different from the other patches. Moreover, the quality of the food density depended on the number of revisits for each patch. If an animal was in a specific location and there was a potential switch, the switch depended on the distance between the current location, the patches and the quality of the patches. In some cases, the quality of these patches was varied, depending on the time of the year. At certain times of the year, the quality of one patch decreased

while the quality of another patch increased. Hence, at some point, the animal stops heading towards a low-quality patch and starts moving to a high-quality patch. The idea of Wang et al. represents a special case of spatial covariates, because the animal could switch to only one patch every time and there is no randomness in behaviour. In the idea of multiple patches, a behaviour exists for each patch, as well as for moving between the patches, which might result in having twice as many behaviours for a given number of patches. Some behaviours represent an animal's movement inside each patch, while other behaviours represent its movement towards each patch.

Measurement error for observations

As mentioned in Chapter 3, this thesis assumed that there is no measurement error in locations. An animal does not move very much in a small interval, which indicates that a risk exists of an error being larger relative to its actual movement. If an error is substantial, then an animal might be moving in a completely different direction. When the data have a fine-scale observations, it is more important to think about the movement error. The model presented here can be extended to include the measurement error.

In the current model, some potential switches exist and the locations at the time of the switches need to be reconstructed. When including a measurement error, the true locations at the observation times also need to be reconstructed by allowing for measurement uncertainty, instead of the current practice of treating the observations as true locations. In this model, the true location $x(t)$ would follow the OU process and, at a given time. The observed location y_t could follow a normal distribution or t-distribution that is centred on the true location and with some variance. A new set of variables would need to be added to represent the true locations and an extra step would be needed in the MCMC to update those true locations.

If a measurement error is small, then both models with or without this measurement error should yield very similar results. In reality there is always some measurement error, but the question this notion raises concerns whether this error is small enough to ignore or whether it should be incorporated in the model. However, including a measurement error further complicates the model and it becomes harder to make inferences; thus, an error should be omitted if it is small enough. This approach is often applied to GPS data, because the GPS measurement error is relatively small.

Even if a measurement error is not small, scientists can often determine its size—e.g. the variance in the error term. In that case, the known variance or an informative prior

distribution of its value can directly be included in the model. On the other hand, if there is a measurement error but its size is unknown, then an uninformative prior must be used for the actual error variance. This case is unavoidably difficult—if an error is large, then the observed path might not resemble the actual path and learning about the details of an animal's movement and behaviour would be impossible.

References

[1] Anderson, D. J. (1982). The home range: A new nonparametric estimation technique. *Ecology*, 63(1):103–112.

[2] Aparna, U., Pedersen, L. J., and Jørgensen, E. (2014). Hidden phase- type Markov model for the prediction of onset of farrowing for loose-housed sows. *Computers and Electronics in Agriculture*, 108:135–147.

[3] Avgar, T., Deardon, R., and Fryxell, J. (2013). An empirically parameterized individual based model of animal movement, perception, and memory. *Ecological Modelling*, 251:158–172.

[4] Babin, J., Fortin, D., Wilmshurst, J. F., and Fortin., M.-E. (2011). Energy gains predict the distribution of plains bison across populations and ecosystems. *Ecology*, 92:240–252.

[5] Beyer, H., Morales, J., Murray, D., and Fortin, M. (2013). The effectiveness of Bayesian state-space models for estimating behavioural states from movement paths. *Methods in Ecology and Evolution*, 4(5):433–441.

[6] Blackwell, P. (2018). Integrated continuous-time hidden markov models. *https://arxiv.org/abs/1807.11907*.

[7] Blackwell, P. G. (1997). Random diffusion models for animal movement. *Ecological Modelling*, 100(1-3):87–102.

[8] Blackwell, P. G. (2003). Bayesian inference for Markov processes with diffusion and discrete components. *Biometrika*, 90(3):613–627.

[9] Blackwell, P. G., Niu, M., Lambert, C., and LaPoint, S. D. (2016). Exact Bayesian inference for animal movement in continuous time. *Methods in Ecology and Evolution*, 7:184–195.

[10] Boutin, S. and Hebert, D. (2002). Landscape ecology and forest management: Developing an effective partnership. *Ecological Applications*, 12(2):390–397.

[11] Breed, G. and Severns, P. (2015). Low relative error in consumer-grade gps units make them ideal for measuring small-scale animal movement patterns. *PeerJ*, (3):e1205.

[12] Brillinger, D. R., Preisler, H. K., Ager, A. A., Kie, J. G., and Stewart, B. S. (2002). Employing stochastic differential equations to model wildlife motion. *Bulletin of the Brazilian Mathematical Society*, 33(3):385–408.

[13] Brillinger, D. R. and Stewart, B. S. (1998). Elephant-seal movements: modelling migration. *The Canadian Journal of Statistics*, 26(3):431–443.

[14] Brost, B., Hooten, M., Hanks, E., and Small, R. (2015). Animal movement constraints improve resource selection inference in the presence of telemetry error. *Ecology*, 96(10):2590–2597.

[15] Buchholz, P., Kriege, J., and Felko, I. (2014). *Input Modeling with Phase-Type Distributions and Markov Models: Theory and Applications*. Springer, Berlin, Germany.

[16] Bulla, J., Bulla, I., and Nenadic, O. (2009). An R package for analyzing hidden semi-Markov models. *Computational Statistics and Data Analysis*, 54(3):611–619.

[17] Burt, W. H. (1943). Territoriality and home range concepts as applied to mammals. *Journal of Mammalogy*, 24(3):346–352.

[18] Calenge, C., Dray, S., and Royer-Carenzi, M. (2009). The concept of animals' trajectories from a data analysis perspective. *Ecological Informatics*, 4(1):34–41.

[19] Calhoun, J. B. and Casby, J. U. (1958). Calculation of home range and density of small mammals. Technical Report 55, U.S. Public Health Service Publication.

[20] Choquet, R., Bechet, A., and Guedon, Y. (2014). Applications of hidden hybrid Markov/semi-Markov models: From stopover duration to breeding success dynamics. *Ecology and Evolution*, 4:817–826.

[21] Cowlishaw, M. (2014). *Determinants of home range and territory size in coral reef fishes*. PhD thesis, James Cook University, Australia.

[22] Cox, D. (1955). A use of complex probabilities in the theory of stochastic process. *In Proc. Cambridge philosophical Soc.*, 51(2):313–319.

[23] Cox, D. and Miller, H. (1965). *The theory of stochastic process*. Chapman and Hall, London.

[24] Dixon, K. R. and Chapman, J. A. (1980). Harmonic mean measure of animal activity areas. *Ecology*, 61:1040–1044.

[25] Don, B. A. C. and Rennolls, K. (1983). A home range model incorporating biological attraction points. *Journal of Animal Ecology*, 52(1):69–81.

[26] Dunn, J. E. and Gipson, P. S. (1977). Analysis of radio telemetry data in studies of home range. *Biometrics*, 33(1):85–101.

[27] Ford, R. G. and Krumme, D. W. (1979). The analysis of space use patterns. *Journal of Theoretical Biology*, 76:125–155.

[28] Frair, J. L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N. J., and Pedrotti, L. (2010). Resolving issues of imprecise and habitat-biased locations in ecological analyses using gps telemetry data. *Philosophical Transactions of the Royal Society B*, 365:2187–2200.

[29] Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., and Wickham, J. (2011). Completion of the 2006 national land cover database for the conterminous united states. *Photogrammetric Engineering and Remote Sensing*, 77:858–864.

[30] Guédon, Y. (2003). Estimating hidden semi-Markov chains from discrete sequences. *Journal of Computational and Graphical Statistics*, 12(3):604–639.

[31] Guédon, Y. (2005). Hidden hybrid Markov/semi-Markov chains. *Computational Statistics and data analysis*, 49(3):663–688.

[32] Gurnell, J. (1984). Home range, territoriality, caching behaviour and food supply of the red squirrel (tamiasciurus hudsonicus fremonti) in a subalpine lodgepole pine forest. *Animal Behaviour*, 32(4):1119–1131.

[33] Harris, K. J. and Blackwell, P. G. (2013). Flexible continuous-time modelling for heterogeneous animal movement. *Ecological Modelling*, 255:29–37.

[34] Hayne, D. W. (1949). Calculation of size of home range. *Journal of mammalogy*, 30:1–18.

[35] Hooten, M. B., Johnson, D. S., McClintock, B. T., and Morales, J. M. (2017). *Animal Movement: Statistical Models for Telemetry Data*. CRC Press, Boca Raton, Florida.

[36] Horne, J. S., Garton, E. O., Krone, S. M., and Lewis, J. S. (2007). Analyzing animal movements using brownian bridges. *Ecology*, 88(9):2354–2363.

[37] Iacus, S. M. (2008). *Simulation and inference for stochastic differential equations*. Springer Series in Statistics, New York.

[38] Jennrich, R. I. and Turner, F. B. (1969). Measurement of non-circular home range. *Journal of Theoretical Biology*, 22(2):227–237.

[39] Johnson, D. S., London, J. M., Lea, M. A., and Durban, J. W. (2008). Continuous time correlated random walk model for animal telemetry data. *Ecology*, 89(5):1208–1215.

[40] Johnson, M. T. (2005). Capacity and complexity of HMM duration modelling techniques. *IEEE Signal Processing Letters*, 12:407– 410.

[41] Kranstauber, B., Kays, R., LaPoint, S., Wikelski, M., and Safi, K. (2012). A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *Journal of Animal Ecology*, 81(4):738–746.

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

[43] Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., and Morales, J. (2012). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*, 93(11):2336–2342.

[44] Langrock, R., Kneib, T., Sohn, A., and DeRuiter, S. (2015). Nonparametric inference in hidden markov models using p-splines. *Biometrics*, 71(2):520–528.

[45] Langrock, R. and Zucchini, W. (2011). Hidden Markov models with arbitrary dwell-time distributions. *Computational Statistics and Data Analysis*, 55:715–724.

[46] LaPoint, S., Gallery, P., Wikelski, M., and Kays, R. (2013a). Animal behavior, cost-based corridor models, and real corridors. *Landscape Ecology*, 28:1615–1630.

[47] LaPoint, S., Gallery, P., Wikelski, M., and Kays, R. (2013b). Data from: Animal behavior, cost-based corridor models, and real corridors. *Movebank Data Repository*.

[48] Liechty, J. C. and Robert, G. O. (2001). Markov chain Monte Carlo methods for switching diffusion models. *Biometrika*, 88(2):299–315.

[49] Link, W. A. and Eaton, M. J. (2011). On thinning of chains in mcmc. *Methods in Ecology and Evolution*, 3(1):112–115.

[50] MacDonald, D. W., Ball, F. G., and Hough, N. G. (1980). The evaluation of home range size and configuration using radio tracking data. In Amalner, C. J. and MacDonald, D. W., editors, *A handbook on biotelemetry and radio tracking*, pages 405–424. Pergamon, Oxford.

[51] McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M., and Morales, J. M. (2014). When to be discrete: the importance of time formulation in understanding animal movement. *Movement ecology*, 2(1):21.

[52] Michelot, T., Langrock, R., and Patterson, T. A. (2016). moveHMM: an R package for the statistical modelling of animal movement data using hidden Markov models. *Methods in Ecology and Evolution*, 7(11):1308–1315.

[53] Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. (2004). Extracting more out of relocation data: Building movement models as mixtures of random walks. *Ecology*, 85(9):2436–2445.

[54] Nams, V. (2014). Combining animal movements and behavioural data to detect behavioural states. *Ecology Letters*, 17(10):1228–1237.

[55] Natvig, E. and Subbey (2011). Modelling vertical fish migration using mixed ornstein-uhlenbeck processes. In Proceedings Norsk. Informatikkonferanse.

[56] Niu, M., Blackwell, P. G., and Skarin, A. (2016). Modeling interdependent animal movement in continuous time. *Biometrics*, 72:315–324.

[57] Parton, A., Blackwell, P., and Skarin, A. (2016). Bayesian inference for continuous time animal movement based on steps and turns. *arXiv preprint, arXiv:1608.05583*.

[58] Parton, A. and Blackwell, P. G. (2017). Bayesian inference for multistate 'step and turn' animal movement in continuous time. *Journal of Agricultural, Biological and Environmental Statistics*, 22:373–392.

[59] Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., and King, R. (2017). Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *Advances in Statistical Analysis*, 101:399–438.

[60] Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. (2008). State–space models of individual animal movement. *Trends in Ecology & Evolution*, 23(2):87–94.

[61] Pedersen, M. W. and Weng, K. C. (2013). Estimating individual animal movement from observation networks. *Methods Ecol. Evol*, 4:920–929.

[62] Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1):7–11.

[63] Preisler, H., Ager, A., Johnson, B., and Kie, J. (2004). Modeling animal movements using stochastic differential equations. *Environmetrics*, 15(7):643–657.

[64] Quy, R., Massei, G., Lambert, M., Coats, J., Miller, L., and Cowan, D. (2014). Effects of a gnrh vaccine on the movement and activity of free-living wild boar (sus scrofa). *Wildlife Research*, 41:185–193.

[65] Rubino, G. and Sericola, B. (1989c). Sojourn times in finite Markov processes. *Journal of Applied Probability*, 26:744–756.

[66] Russell, M. J. and Cook, A. E. (1987). Experimental evaluation of duration modelling techniques for automatic speech recognition. *In Proc. Int. Conf. Acoust., Speech, Signal Process*, pages 2376–2379.

[67] Russell, M. J. and Moore, R. K. (1985). Explicit modelling of state occupancy in hidden Markov models for automatic speech recognition. *In Proc. Int. Conf. on Acoust. Speech and Signal Processing*, Tampa, FL:5–8.

[68] Russell, R. E., Royle, J. A., Desimone, R., Schwartz, M. K., Edwards, V. L., Pilgrim, K. P., and McKelvey, K. S. (2012). Estimating abundance of mountain lions from unstructured spatial sampling. *Journal of Wildlife Management*, 76(8):1551–1561.

[69] Seton, E. T. (1909). *Life histories of northern animals. An account of the mammals of manitoba.* Charles Scribners sons, New York city.

[70] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

[71] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):485–493.

[72] Tishkovskaya, S. and Blackwell, P. G. (in prep). Bayesian estimation of heterogeneous environments from animal movement data.

[73] Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of brownian motion. *Physical Review Letters*, 36:823–841.

[74] Van Winkle, W. (1975). Comparison of several probabilistic home-range models. *Journal of Wildlife Management*, 39(1):118–123.

[75] Voigt, D. R. and Tinline, R. R. (1980). Strategies for analysis radio tracking movement data. In Amalner, C. J. and MacDonald, D. W., editors, *A Handbook on Biotelemetry and Radio Tracking*, pages 387–404. Pergamon, Oxford.

[76] Wang, Y., Blackwell, P., Merkle, J., and Potts, J. (In Prep). Continuous time resource selection analysis for moving animals.

[77] Worton, B. J. (1987). A review of models of home range for animal movement. *Ecological Modelling*, 38:277–298.

[78] Worton, B. J. (1995). Modelling radio-tracking data. *Environmental and Ecological Statistics*, 2:15–23.

[79] Zucchini, W. and MacDonald, I. (2009). Hidden Markov Models for Time Series: An Introduction Using R. volume 110 of *Monographs on Statistics and Applied Probability*.