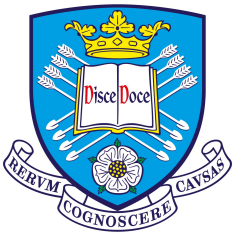
Modelling and Inference for the Movement of Interacting Animals



The University Of Sheffield.

Jordan Elliott Milner

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

This dissertation is submitted for the degree of
Doctor of Philosophy

November 2021

Declaration

I, the author, confirm that this Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not previously been presented for an award at this, or any other, university.

This thesis 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 below. Parts of Chapters 3 and 4 (specified in the introductions of those chapters) have been published in Milner *et al.* (2021), which was co-authored with Paul Blackwell and Mu Niu. Additionally, some of the work on non-stationary movement in Chapter 3 was developed in conjunction with the model detailed in Niu *et al.* (2020), which is co-authored with Mu Niu, Fay Frost, Anna Skarin and Paul Blackwell.

Acknowledgements

I wish to extend my special thanks to my supervisor, Paul Blackwell. Paul took the time to answer my speculative email several years ago and then encouraged and aided me through the application process — despite my lack of statistical background! I am grateful for his continued support and guidance throughout my PhD. I would also like to thank the Engineering and Physical Sciences Research Council (EPSRC) for funding this work.

A proportion of Chapters 3 and 4 are published in Milner *et al.* (2021), which is co-authored with Paul and Mu Niu. Additionally, some of the work on non-stationary movement in Chapter 3 was developed in conjunction with the model detailed in Niu *et al.* (2020), which is co-authored with Mu, Fay Frost, Anna Skarin and Paul. I would like to thank: my co-authors, who were a pleasure to work with; the various referees for their constructive comments; Eloise Bray for useful discussions on some aspects of Milner *et al.* (2021).

With regards to the wider academic community, I am thankful to the team behind Movebank and everyone that contributes to the open data repository — it has proved to be a most valuable resource. I am also appreciative of the opportunities and training provided by my department and the various development teams at the university.

I am grateful for the start in life that my parents gave me which has provided me with, as well as many other things, opportunities that were beyond their own. Thanks to Jake and Paddy for the (remote) weekly stress release, particularly over the past year.

Finally, I would like to thank Marianne for always turning the latter part of 'for better, for worse' into the former.

Abstract

As animal tracking data is becoming more readily available, statistically modelling the movement of animals is an increasingly utilised approach with which we can analyse their behaviours. Typically though, these models have been developed for the analysis of individual animals and so they fail to account for the social drivers of movement. In this thesis, we aim to build those social drivers into our movement models. Not only will doing so provide a more complete explanation for their behaviours, but it will also provide useful insight into their social structure in general.

Our solution is inspired by orderly social hierarchies — a simple, widely used construct that is easy to interpret whilst providing an in-depth view of social behaviours. The flexibility and level of insight gained from this approach is increased as we capture the dynamism of social behaviours through a continuous-time behavioural switching process. Alongside this framework, we define a multivariate diffusion process that can model the collective movement that results from such social interaction. We first develop our model in the simpler context of spatial homogeneity and explore Markov chain Monte Carlo inference methods with which we can estimate the model parameters in a Bayesian setting.

We then extend the above model to the spatially heterogeneous case, increasing the scope of its applications. We develop a novel model-fitting algorithm which allows us to circumvent a sizeable portion of the increased complexity and computational cost resulting from this extension. We are then able to explore additional customisation of our approach, such as building in a radius of interaction — a feature which can provide some further biological realism to the model.

The above models have been developed in continuous time to gain the resulting flexibility with regards to the temporal resolution and completeness of the data.

Contents

List of Figures

List of Tables

Nomenclature

List of Abbreviations

BM	Brownian motion
CDF	Cumulative distribution function
CTMC	Continuous-time Markov chain
DAG	Directed acyclic graph
DTMC	Discrete-time Markov chain
ESS	Effective sample size
FFD	Fission-fusion dynamics
GPS	Global positioning system
HMM	Hidden Markov model
LDB	Latent diffusion bridge
LDP	Latent diffusion process
MCMC	Markov chain Monte Carlo
MH	Metropolis-Hastings
MPSRF	Multivariate potential scale reduction factor
OU	Ornstein-Uhlenbeck
PSRF	Potential scale reduction factor
SPP	Self-propelled particles
SNA	Social network analysis
SDE	Stochastic differential equation

Chapter 1

Ecological Motivation

In this chapter, we provide a broad overview of the statistical movement models with which we can analyse animal movement behaviours. We do so mostly with respect to modelling the collective behaviours of multiple animals, but we also cover some key developments that were made with individual-based models. Additionally, we cast an eye on other methods which are used to study the social behaviours of animals in order to obtain an overview of the techniques used and what information will be valuable output from a movement model. We summarise the state of play in Section 1.4, along with outlining the aims of this thesis.

1.1 Tracking Animal Movement

1.1.1 Movement Data

Animal movement is heavily interlinked with other natural processes in our world. Understanding the hows and the whys of that movement can provide essential insight, from species-specific behaviours to ecosystem function (Kays *et al.*, 2015). Typically, the animal movement trajectories analysed are expressed as a series of geographical coordinates collected by tracking devices (Patterson *et al.*, 2017), which are attached to the animals themselves. Advances in these tracking technologies are providing more opportunities to study moving animals (Kays *et al.*, 2015).

In particular, advances in GPS technology are facilitating these opportunities. Studies on animal movement have traditionally been undertaken on data collected with very high frequency (VHF) devices. However, data collected with these devices are subject to a large measurement error (200m – 600m) and range limitations have generally restricted them to investigations of home ranges — ones that are accessible at that (Frair *et al.*, 2010; Kays *et al.*, 2015). GPS devices, on the other hand, have a smaller error (typically less than 10m), can provide a continuous record of an animal's location over long distances and are increasingly accessible (i.e. cheaper and smaller) (Kays *et al.*, 2015). As a result, animal movement can be studied at a global scale using these tags (Tucker

et al., 2018).

This increase in accessibility also provides new opportunities to tag multiple individuals within a group of animals simultaneously. Moreover, the high resolution of the data has the ability to capture fine-scale interactions between animals — thereby allowing us to study the social drivers of movement behaviours in the wild (Hughey *et al.*, 2018). Researching collective movements had, until recently, been largely restricted to laboratory or simulation approaches (Westley *et al.*, 2018). However, the scenario may still arise where the duration of an interaction is smaller than the temporal resolution of the data. Care is therefore needed when undertaking such studies and ideally the rate of data collection should be matched with the behaviours being investigated (Kays *et al.*, 2015).

Whilst data collection is typically undertaken at regular intervals, irregularities can be introduced both accidentally and by design. For instance, GPS devices can fail to record locations when an animal moves under a dense canopy cover. When tracking multiple animals, the individual tags may not be in sync with each other. Due to battery and/or storage constraints in some devices, the tracking device might be programmed to collect less data during periods of low activity in order to monitor a longer period of time (Williams *et al.*, 2020).

Aside from tracking tags, other methods of data collection to study collective behaviours include using software to track the movement of animals throughout a video (e.g. Katz *et al.* (2011)) or a series of photographs (e.g Ballerini *et al.* (2008)). Whilst these methods were previously limited to captive studies (as in Katz *et al.* (2011)) or capturing particular behaviours (e.g. location specific starling murmurations as in Ballerini *et al.* (2008)), technological developments have also allowed for these approaches of data collection to be expanded upon. For instance, Torney *et al.* (2018) use a commercial unmanned aerial system (UAS) to film multiple caribou in the wild. Such a method has advantages over GPS tags in that the animals don't need to be tagged and so it is perhaps better suited to capturing complete information on larger, more fluid social groups. That is, tagging all of the animals in a social group can create a sizeable logistical and ethical challenge. Aerial footage can also provide additional information, such as the headings of the animals. However, the battery life of UASs are currently a limiting factor as the longest individual track length in Torney *et al.* (2018) was 9 minutes, whilst tracking devices have been used to monitor animals over their entire life span (Kays *et al.*, 2015).

Whilst we will discuss some studies that utilise the above alternative methods of data collection, we will focus on modelling movement trajectories obtained through tracking tags due to their prominent use in monitoring animals and the scale of data they can provide. That is, they are becoming ever more accessible, they can provide high-frequency data for long periods of time and they are not bound by range or habitat accessibility. We will only discuss two-dimensional data (i.e. longitudinal and latitudinal), though one-dimensional (e.g. diving depth) or even three-dimensional trajectories are sometimes collected. When analysing two-dimensional longitudinal and latitudinal data, we

will convert it into the Universal Transverse Mercator (UTM) coordinate system to navigate the spherical nature of geographic coordinates.

It is also worth prefacing further discussions with a note on a key debate within animal movement modelling which concerns time formulation. That is, should we treat time as discrete or continuous (McClintock *et al.*, 2014)? Although animal movement is a continuous-time process, it is perhaps more intuitive to think of movement as a discrete-time one. For instance, visualising movement as a series of steps and turns, rather than, say, a diffusion process (McClintock *et al.*, 2014). Though, work has been undertaken to bring the intuition of a discrete-time formulation into a continuous-time framework (Parton & Blackwell, 2017). A disadvantage of discrete-time models is that they require temporally-regular data, which often does not occur as mentioned above. Additionally, discrete-time analyses are not time invariant, whereas their continuous-time counterparts are (within limits). Thus, matching the temporal scale of the model with the temporal scale of the movement behaviours to be studied (along with obtaining data with a suitable resolution) is key (McClintock *et al.*, 2014). Currently, discrete-time models are the more prominent of the two approaches. Software making the models available, and the efficiency at which they can be fitted to data, has largely driven that trend.

1.1.2 Individual-Based Statistical Movement Models

As previously stated, we are focusing on modelling collective movement. However, key developments in movement modelling that have informed our work have been made in the context of individual-based models, and so we will discuss them here. We will be focusing on the behavioural state switching class of models, which go some way to accounting for the complexity and dynamism of animal movement by allowing it to be expressed by a number of different processes.

Various metrics can be used when modelling animal movement. For instance, the step lengths and turning angles as mentioned above, velocity or the positions themselves (Patterson *et al.*, 2017). Regardless of the approach taken, movement models were initially developed with a single process or behaviour in mind. In discrete-time models, random walks (RW) are a popular mechanism to model movement. Whilst basic random walks, where each movement step is independent, may be too simplistic, variations such as correlated (CRW) and biased (BRW) random walks can represent more complex behaviours (Morales & Ellner, 2002). That is, CRWs can capture persistence in movement whilst BRWs can capture some preference towards a certain direction. Often, these more complex RWs are characterised by a distribution for step lengths and a distribution for either turning angles (CRW) or bearings (BRW).

In continuous time, diffusion processes are commonly used to model movement. Dunn & Gipson (1977) model the locations of an animal with an Ornstein-Uhlenbeck (OU) process in order to estimate their home range. An OU process was deemed a natural fit for this purpose as it is a mean-reverting process with some drift towards a central location, continuous (movement is

continuous by nature), stationary (matching the physical requirement of a home range), Gaussian (matching their assumption on the nature of a home range) and Markovian (accounting for the lack of independence between successive locations). Alternatively, the velocity of an animal can be formulated as an OU process to give rise to a continuous-time analogue of a correlated random walk (Johnson *et al.*, 2008). This approach captures the inertia that the animals are assumed to have. That is, they will move at a similar rate over successive intervals. This velocity process can be integrated to produce the positional process.

However, animals are unlikely to exhibit constant movement behaviour — they may alter their behaviour according to their physiological state or due to encountering a different habitat for example (Morales & Ellner, 2002). As such, developing methods that can accommodate those changes in behaviour will be better suited to modelling the complexity of animal movement. Morales *et al.* (2004) do just that by modelling the movement of elk as a mixture of random walks, where each observation is generated by one of the RWs. Each random walk is associated with an unobserved behavioural state ('encamped' or 'exploratory' in the two state case) and each have their own distributions for the step lengths and turning angles. 'Encamped' corresponds to short step lengths and frequent changes in direction, whilst 'exploratory' corresponds to longer, more persistent movement — differences that would be lost in a single state model. Morales *et al.* (2004) explore different models for switching between the behaviours, with the more structured approaches (e.g. Markovian) outperforming less structured approaches where the state could change freely. Note, whilst behavioural states are usually given biological labels, they should only be interpreted as statistical descriptions of the movement (e.g. short step lengths in the 'encamped' case) as opposed to biological descriptions of an animal's behaviour.

Some of the models examined by Morales *et al.* were early applications of a hidden Markov model (HMM) to animal movement data. In a HMM, the unobserved behavioural process is modelled with a discrete-time Markov chain, with the state of an animal at any one time only conditional on the previous state. The models in Morales *et al.* (2004) were fitted to data with Markov chain Monte Carlo (MCMC) methods. However, HMMs have assumed a dominant place in the animal movement modelling world in part due to an efficient likelihood maximisation algorithm with which we can estimate the model parameters and decode the behavioural states (Patterson *et al.*, 2017). HMMs can also be extended to include covariates (Patterson *et al.*, 2009) and a semi-Markovian state process when Markovian behaviour is not biologically reasonable (Langrock *et al.*, 2012). That is, the dwell time in a state is not geometrically distributed. The two main disadvantages of HMMs (as well as the aforementioned limitations of discrete-time models in general) are that they require regularly-spaced data and that they assume negligible measurement error. When the latter cannot be assumed, the more flexible, but complex, state-space models can provide an alternative route (Patterson *et al.*, 2017).

In a continuous-time framework, Blackwell (1997) combines behavioural state switching with diffusion processes — OU positional processes in particular. They also include Brownian motion (BM),

which can be considered as a limiting case of an OU process — one that doesn't contain any drift towards a central location. Whilst a perhaps overly simplistic model on its own, the inclusion of BM can provide some flexibility to a diffusion-switching model — capturing moments of non-stationary movement for instance. This diffusion-switching approach can navigate the overly simplistic unimodal home range estimation, which results from a single state model (Dunn & Gipson, 1977), as the different processes may correspond to different patches of the home range. Alternatively, the different processes may represent different behaviours altogether (e.g. foraging and travelling). Again, the behavioural process is assumed to be Markovian as state switches are modelled with a continuous-time Markov chain. Blackwell (2003) introduces an inference algorithm to fit this class of models to data, though only the case where both the locations and behaviours are observed is discussed in detail.

In the method outlined by Blackwell (2003), the behavioural process is considered missing (between the observations) and it is reconstructed in continuous time in order to determine exactly when state switches occurred. The location of the animal at these switching times is not considered though, and so this method is generally suited to the spatially homogeneous case. However, Harris & Blackwell (2013) demonstrate with numerous examples the benefits of extending these diffusion-switching models to spatial heterogeneity. Blackwell *et al.* (2016) present a much more flexible method that accounts for a completely unobserved behavioural process and both the state switches and locations of an animal are reconstructed in continuous time. Subsequently, the transitions between states can now depend on the location of the animal. This approach, fitted with MCMC methods, allows for the exact inference of the model parameters in a Bayesian setting.

1.2 Studying Social Animals

Whilst a broad range of statistical movement models have been developed to investigate individual behaviour and how that behaviour is dependent on factors such as the surrounding environment, they inherently fail to account for the social factors that drive social groups. Conspecifics of various taxa form social groups. The motivation for this may most commonly be due to kinship ('inclusive fitness' (Hamilton, 1963, 1964a,b; Clutton-Brock, 2009)), but there are other mechanisms which bring animals together, related or not. These include: reciprocity (e.g. repeated grooming assistance), mutualism (e.g. cooperative foraging or defence) and manipulation (e.g. subordinates grooming dominants to earn favour/protection) (Clutton-Brock, 2009). This sociality has a profound impact upon the fitness of the animals. Social behaviours can: reduce the cost of reproduction and rearing (Clutton-Brock, 2009); reduce navigational error (Codling *et al.*, 2007); have implications for disease transmission (Hamede *et al.*, 2009) and how information propagates through a population (Voelkl & Noë, 2008); decrease the risk of predation (Sueur *et al.*, 2011); dictate spatial use (Wittemyer *et al.*, 2007); impact obstacle avoidance (Croft *et al.*, 2015); enhance their environmental gradient-tracking ability (Berdahl *et al.*, 2013) and many more.

It is clearly prudent then to account for the sociality of animals when modelling their movement (Bode *et al.*, 2010). Not only will doing so help better explain that movement, but we will also be able to investigate the nature of the social interactions that produced it. For instance, identifying animals which have high levels of influence on a group's movement will provide us with important information on their social structure, which then has useful applications in conservation efforts (King *et al.*, 2018; Westley *et al.*, 2018). Studies of social structure can also help highlight the pressures placed on a group of animals from external factors or change, which can then impact their fitness. For example, giraffe communities closer to humans have been shown to contain weaker social relationships (Bond *et al.*, 2020). Additionally, accounting for social behaviours in animal movement can improve our analysis of population dynamics (Haydon *et al.*, 2008; Morales *et al.*, 2010).

Whilst previous developments in animal movement modelling were limited by the data available, with an increasing ability to obtain simultaneous tracking data from multiple animals within a group (Westley *et al.*, 2018), we now have the opportunity to build these social interactions into our models. In order to do so though, we must develop some social framework — one that is ideally flexible enough to be able to model a broad array of taxa.

1.2.1 Social Network Analysis

Social network analysis (SNA) is a commonly used method with which we can form a picture of the social structure of a group of animals. SNA provides a framework "with which we can study the social organisation of animals at all levels (individual, dyadic, group, population) and for all types of interaction (aggressive, cooperative, sexual etc)" (Krause *et al.*, 2009). It therefore can be deployed to better understand: the social structures in general; the causes and consequences of individual variation on network position; the implications of network structure of transmission of information and disease; and the relationship between the environment and social structure (Farine & Whitehead, 2015).

The social networks consist of nodes (individual animals), which are connected by edges. The edges can be directed or undirected and they can quantify the strength of a relationship between two animals (weighted) or merely acknowledge whether there is a relationship or not (unweighted). These networks are typically constructed from data that has been pooled over some time period and they are not limited to the acyclic case. For example, there can be directed edges going each way between two animals.

Various types of data (or some combination of) are used to construct these networks which broadly fall into two categories: interaction data and association/proximity data (Castles *et al.*, 2014). Interaction data is obtained from directly observing the animals and thus different types of interaction, such as cooperative, agonistic and affiliative, can be differentiated (Farine & Whitehead, 2015). These observations of dyadic interactions, such as grooming or pecking, can then be used to

construct interaction-based networks (Farine & Whitehead, 2015) or be distilled to create a dominance hierarchy (Chase & Lindquist, 2016). Proximity networks are constructed in much the same way — using dyadic information to construct a wider group structure — but here the strength of a dyadic relationship is dictated by spatial proximity or shared resource use (Castles *et al.*, 2014).

With SNA, we are able to investigate the properties of a social network, such as whether the network is more structured than expected from random (Farine & Whitehead, 2015). Group and individual-level metrics are derived from the network to describe the social structure and the position of individuals within it respectively. For example, there are measures to determine the density of a network and how central each animal is to it (Farine & Whitehead, 2015). The individual-level metrics provide a tool to link social position to age, sex, reproductive success and disease status amongst others (Silk *et al.*, 2015). For instance, Hamede *et al.* (2009) investigate if there are highly connected individuals or demographics in a population of Tasmanian devils, with the view that targeted culling or treatment could control the transmission of an infectious disease. Ellis *et al.* (2017) found that the survival probability of male killer whales is related to measures of social 'centrality'. However, care is needed when interpreting metrics as they depend on how the population is structured and how the data is collected (Farine & Whitehead, 2015).

Much of the SNA literature presents a static overview of the social structure, obtained by collating interactions over some time period, but there is a branch of it that aims to capture the dynamics of social behaviours (Blonder *et al.*, 2012; Psorakis *et al.*, 2012; Hobson *et al.*, 2013; Farine, 2018). After all, all animal social networks are dynamic to a certain extent (Farine & Whitehead, 2015). At the very least social groups will change with deaths and births within it. Additionally, there is much interest in the interplay between social behaviours and dynamic phenomena such as the environment and transmission of diseases in order to aid conservation efforts (Snijders *et al.*, 2017). We will be able to gain a more complete picture of those links by accounting for the dynamism of social structures. For example, analysing disease transmission with a dynamic approach will provide more information as to the exact nature of the transmission (Farine, 2018).

Blonder *et al.* (2012) discuss two common approaches to studying dynamic social networks: time-ordered networks and time-aggregated networks. The former is the more dynamic of the two as time-ordered networks represent the interactions (and therefore the network) that occurred at any given particular time — they retain all information regarding the timing, duration and ordering of the interactions. Thus, for instance, they are a useful method to investigate how information or a disease propagates through a group and highlight important individuals for doing so. Alternatively, time-ordered data can be pooled into discrete time windows to provide time-aggregated networks — essentially a series of static structures. As well as being used to assess changes in network structure in response to external factors or identify persistent communities within a group (Blonder *et al.* (2012), Hobson *et al.* (2013) and Mbizah *et al.* (2020) for example), time-aggregated networks have been used to develop stochastic actor-based models (Snijders *et al.*, 2010). In the approach of

Snijders *et al.* (2010), the network evolves stochastically according to network (e.g. a tendency to reciprocate an interaction), individual (e.g. a preference to interact with others of the same sex) and dyadic (e.g. kinship) characteristics.

McDonald & Shizuka (2013) and Shizuka & McDonald (2015) look to assess the orderliness of interaction networks across a range of taxa with regards to dominance (the interaction observed being dominance based). The orderliness of a group depends on the transitivity of the dominance relationships, as well as the stability and 'steepness' of the dominance rankings — that is, the level of disparity in the dominance scores for the animals (McDonald & Shizuka, 2013). They investigate this by breaking down the network into $\binom{n}{3}$ triads (for n animals) and classify each one as one of seven directed network motifs. Five of these motifs are of particular interest here. The double-dominant (one animal is dominant to the other two), double-subordinate (one animal is subordinate to the other two) and transitive triads are orderly. The cyclic triad is disorderly whilst the pass-along triad (a cyclic triad with a missing edge) has the potential to go either way depending on how the missing edge is filled in. Both studies find that there is a significant excess of the orderly double-dominant and transitive triads when compared to the expected number in a random network. Similarly, there is a significant deficit of the pass-along and cyclic triads. This finding was consistent across the range of taxa analysed in the aforementioned studies (birds, primates, nonprimate mammals, fish and invertebrates) and across group size (Shizuka & McDonald, 2015). The analysis of McDonald & Shizuka (2013) also concludes that groups are typically orderly from the point of view of rank stability and steepness (i.e. there is a large disparity in the rank scores). Thus, whilst cyclic structures do sometimes occur, they find social networks are highly orderly. This finding coincides with those of Chase & Lindquist (2016): that disorderly structures quickly transition to an orderly one.

Assuming transitivity across all observed relationships, interaction networks can be reduced to a dominance hierarchy in the form of a directed acyclic graph (DAG). When there is complete data on all possible pairwise interactions, these structures can be reduced further to a linear hierarchy — where the animals are effectively ordered by dominance (Chase & Lindquist, 2016). Wittemyer *et al.* (2007) take this approach after observing various agonistic interactions, such as tusk pokes and trunk slaps, between African elephants. They then use this data to produce a dominance ranking of the animals and use that ranking to investigate how social dominance interplays with spatial use. This approach again only utilises a static view of the network.

Both types of data (interaction and proximity) used to produce a social network tend to require labour intensive collection as they involve direct observation of the animals, though the collection of proximity data can now be more readily automated through the use of tracking tags (Farine & Whitehead, 2015). Not only is direct observation labour intensive, it is open to human error. For example, misidentification of animals or biases in the data collectors, such as identifying brightly coloured animals more often (Davis *et al.*, 2018). Furthermore, when using interaction data, more robust results may be obtained when combining different types of interaction (Hobson *et al.*, 2013)

— further increasing the cost of data collection. On the other hand, tagging can collect data for longer periods of time for animals which may be hard to reach or monitor or whose behaviour may be affected by observation (Scharf *et al.*, 2016). Farine *et al.* (2016) suggests that spatial data alone may be a good proxy for encompassing different types of interactive behaviours, even though it loses sight of the type of interaction (e.g. agonistic or aggressive). However, assuming interaction based on spatial proximity alone, whilst based on a fair assumption that interaction in nonhuman animals requires some sort of proximity, does not account for nonsocial motivation for that proximity. For example, converging onto a resting site does not necessarily require interaction (Pinter-Wollman *et al.*, 2013).

1.2.2 Fission-Fusion Dynamics

Fission-fusion dynamics (FFD) refers to the temporal variation in cohesion, subgroup size and composition of animal groups and it is a phenomenon widespread across taxa (Ramos-Fernández *et al.*, 2018) — thus highlighting the importance of modelling dynamic social behaviour. Such variability is considered advantageous when animals are seeking resources in a spatially and temporally heterogeneous landscape (Ramos-Fernández & Morales, 2014) and when individuals have a conflict of interest (Sueur *et al.*, 2011). Though, Sueur *et al.* (2011) point out that studies of collective behaviours have largely been concentrated on relatively stable groups.

In theory, groups exhibiting high levels of fission-fusion dynamics should not contain consistent leadership. The individual animals in these groups are free to choose their temporary peers and subgroups that will maximise their fitness at the time. However, there is evidence that in certain situations (e.g. resources are sparse or a smaller population size facilities frequent interaction), consistent leadership is found in groups that also demonstrate considerable variation in their group compositions. Lewis *et al.* (2011) obtained these findings through observing how long individual dolphins spent in leading positions in a group.

Ramos-Fernández & Morales (2014) investigated which factors determined individual fissions or fusions in spider monkeys. The most important factors were whether specific other individuals are doing the same and sex. The sex of an individual was an important factor in whether it joined or left a group (males were more likely to fission, females were more likely to fuse), as was the overall gender composition of the group (individuals were more likely to join subgroups that were predominately their own sex). Neither group size or dispersal had a significant effect on individual fissions or fusions. Thus, Ramos-Fernández & Morales (2014) argue that their results show that subgroup patterns in groups with high rates of FFD are born out of interactions between specific individuals, rather than the animals paying attention to subgroup properties.

Other studies of FFD include Ramos-Fernández *et al.* (2018), who quantified temporal variation in social behaviours through measuring the entropy of subgroup composition. Similarly to much of the SNA literature, all the studies mentioned in this section were carried out on directly observed

data. As such, there is a similar discussion to be had on the positives and negatives of the different data collection (and subsequent analysis) methods. For example, direct observation can provide information on the nature of an interaction, but tagging can monitor animals for longer time periods. See the above discussion in Section 1.2.1 for more details.

1.2.3 Local Interaction Rules

Collective behaviours and movement models are brought together in a commonly used class of models based on the 'boids' approach by Reynolds (1987). Here, animals are modelled as self-propelled particles (SPP) which interact according to local, fine-scale, interaction rules. Typically, those rules are based on repulsion, alignment and attraction (or some subset of the three) depending on the distance between the animals. That is, an animal seeks to maintain some distance to all others to avoid collisions with them. If no animals are within their 'zone of repulsion', they interact with those in their 'zone of alignment' and 'zone of attraction' as detailed by Couzin *et al.* (2002) — the radii of the zones being parameters of the model. Typically, an animal interacts with all of those in the appropriate zone(s), except for those not in their field of perception, and the effects are averaged out across all interactions. Usually these models are utilised in the analysis of large groups of animals, such as flocking birds or schooling fish, and the types of collective behaviours these models can produce are backed up by experimental evidence (Tunstrøm *et al.*, 2013).

This approach has led to numerous insights, such as investigating how information is shared within a group on the move and how conflicting information is resolved (Couzin *et al.*, 2005). This work was taken further by Couzin *et al.* (2011) to understand the role of uninformed individuals in the decision making process of a group and del Mar Delgado *et al.* (2018) investigated how individual heterogeneity impacts collective movement behaviours. Berdahl *et al.* (2013) utilises this approach to ascertain that collective dynamics are the driving forces that allow schooling fish to respond to environmental gradients whilst Croft *et al.* (2015) examined how the social structures within a flock of birds impacts its ability to avoid obstacles. Codling *et al.* (2007) show that the pooling of information between animals helps to reduce navigational error.

A variation of this modelling approach is to use topological distance, as opposed to metric distance, to neighbouring animals to dictate the rules of interaction (Ballerini *et al.*, 2008; Camperi *et al.*, 2012). That is, rather than interacting with those within x metres, an animal interacts with its nearest y neighbours. They argue that topological distance is more robust than metric distance as it is not at the mercy of the density of the group. Additionally, movements that happen to result in two animals being further apart than a defined metric distance of interaction will not immediately terminate that interaction.

Interacting with all animals (aside from those in blind spots) within some proximity ranges, either metric or topological, won't provide sufficient biological realism in many cases though (e.g. Katz *et al.* (2011), Strandburg-Peshkin *et al.* (2013) and Rosenthal *et al.* (2015)). For example, Rosenthal

et al. (2015) find that the two most important factors in predicting a response of a focal individual to a startled fish are: how much of the startled fish they can see relative to other fish and the metric distance to it (provided they can see it). Katz *et al.* (2011) find that the interactive effects of neighbouring animals on a focal animal combine in a nontrivial manner — they don't simply average out. Herbert-Read *et al.* (2011) supports this by finding that the social response of their study species is dominated by the single nearest neighbour. In a similar vein, Nagy *et al.* (2010) look at the time-delayed correlation in velocity between each pair of homing pigeons in a small flock to find a hierarchical structure within their interactions. In baboons, Farine *et al.* (2016) find that an individual's longer-term movement is best predicted by their top 4 to 6 'spatial affiliates' — those with whom it is most often the nearest neighbour with. Additionally, Tang *et al.* (2018) find that adding some network structure into these local-interaction-based models can lead to quicker swarming behaviour.

1.2.4 Statistical Collective Movement Models

Much of the above analysis using local-interaction models is not undertaken by fitting a model to data. For instance, Couzin *et al.* (2002) evaluate global properties of simulations in order to investigate the effects of different parameter values on the resulting collective behaviours. Katz *et al.* (2011) map out features of the movement data to investigate the social forces at play between fish. For example, they plot the acceleration of a fish against the distance to its neighbour to find that speed modulation is a key component of interaction that is absent in many local-interaction models. However, there have been some developments in this area. Lukeman *et al.* (2010), for example, weigh up various candidate models (based on metric distance interaction zones) before fitting the most appropriate one to data of a flock of birds in order to ascertain the weighting of each interactive force (repulsion, alignment, attraction or some interaction with a frontal neighbour). Though, the radii for the zones were still inputted as opposed to inferred.

Mann (2011) and Torney *et al.* (2018) bring the local-interaction modelling framework into the world of Bayesian statistics in order to obtain the posterior distributions of the model parameters. Torney *et al.* (2018) use data collected from unmanned aerial systems (UAS) to investigate the social cues from neighbours driving migrating caribou. Models using metric distance, topological distance and exponential decay (influence decreases with distance) are compared to find social cues based on decay best explain the data. Mann (2011) fit various models (with or without alignment; metric of topological distance-based interaction) to simulations (derived from the relevant interaction rules) of swarming animals. The posteriors of the model parameters, such as the attraction and alignment parameters, are consistent with the true values used for the simulations. Though, when the data contains a limited snapshot of interaction (e.g. the animals are always in a particular collective state), they struggle to estimate the interaction radius.

Outside of the local-interaction-rules framework, dyadic interaction metrics are used to statistically measure how interlinked the movement trajectories of two animals are (Long *et al.*, 2014; Joo

et al., 2018). For instance: how often two animals were within a certain proximity of each other; how simultaneous is the use of some area; how correlated certain features of their movement (e.g. velocity) are. Whilst these are quick to calculate and easy to use, they are limited to the analysis of two animals, each metric provides an incomplete view of interaction and some are heavily dependent on assumptions/prior knowledge (e.g. what constitutes proximity) (Joo *et al.*, 2018).

Potts *et al.* (2014) present a flexible framework in which the movement of an individual is modelled through a step selection function which contains terms for their movement process, environmental weighting and social behaviours. All three terms can be configured for the analysis at hand. For example, the social behaviours could represent local interactions, as covered in Section 1.2.3, or memories of past territorial conflicts in order to facilitate avoidance. Russell *et al.* (2016) devise a similarly flexible approach where an animal's movement is modelled with a weighted distribution incorporating: the movement process (a continuous-time correlated random walk), observation error and social interactions. The social element contributes to the weighted distribution through a pairwise interaction function, which depends only on the distances between the animals. This function can be defined to capture the assumed nature of the interactive behaviour being studied. For instance, the function Russell *et al.* (2016) showcase is based on repulsion and attraction — similar to the models in Section 1.2.3. Subsequently, a small distance between two animals will produce a small interaction function value, which will then discourage those two animals from moving towards each other at that time.

Schlägel *et al.* (2019) are able to infer asymmetric dyadic interactions through including occurrence distributions into their step selection function model. An occurrence distribution being the spatial distribution of where an animal has recently been. The model is fitted to each animal individually to ascertain whether it is attracted to, repulsed by or neutral to the occurrence distributions of other animals. Through this approach, they can see if an interaction between two animals is driven by them both or just one of them and whether it is mutual or not (e.g. whether both are attracted to each other or one is neutral). Additionally, in part due to the inclusion of repulsion, their model is not limited to same-species analysis. Similarly to the above methods of Potts *et al.* (2014) and Russell *et al.* (2016) though, the model detailed by Schlägel *et al.* (2019) does not allow for animals to exhibit different social behaviours throughout the observation period.

In a discrete-time movement model fitted to ant data, Russell *et al.* (2017) incorporate social interactions into individual behavioural state transitions. They do so by making the transition probabilities functions of various social covariates: whether a moving ant is nearby; whether a non-moving ant is nearby and whether the queen is nearby. An example of their findings is that a stationary ant was more likely to transition to a moving state when there are moving ants nearby. Whilst this framework requires an interaction radius, Russell *et al.* (2017) only explore different values for it as opposed to inferring it. McKellar *et al.* (2015) develop a similar concept where the transition probabilities of a HMM are functions of a social covariate (group size), as well as environmental covariates. Through this, they find woodpeckers are more likely to be in a resting

state when part of a larger group.

A discrete-time movement model (a HMM) developed by Langrock *et al.* (2014) allows animals to switch between behavioural states that correspond to membership of a group or not. When a member of the group, animals move with some bias (specifically, a biased random walk) towards the group centroid — either a mathematical centroid or the group leader. The movement of animals not in the group is modelled as a correlated random walk, representing the desire to forage independently. Niu *et al.* (2016) developed a continuous-time analogue of the above model. The movement of a group of animals is jointly modelled as a multivariate OU process, the drift of which is directed towards a leading point — the movement of which is also modelled with an OU process. Again, this leading point can be some central abstraction or an actual leading animal. Niu *et al.* (2020) builds upon the framework of Niu *et al.* (2016) by allowing the animals to switch between being a member of the group or being independent — modelled as BM. Additionally, Niu *et al.* (2020) explore the scenario when the leading point does not have a point of attraction. The movement of the leader is modelled with BM and, thus, is no longer stationary. This extension to the non-stationary case was developed in conjunction with the work we will present in Chapter 3.

Scharf *et al.* (2016) jointly model the movement of multiple animals in discrete time using a Gaussian Markov random field (GMRF), where the movement is conditional on the underlying social network. Their GMRF contains two social mechanisms: attraction towards the mean position of those an animal is connected to and connected animals moving in the same general direction. The social network is dynamic and Markovian as the probability of two animals being connected at time T is modelled as a Bernoulli random variable, with dependency on the their connection status at time $T-1$. Though, whilst this formulation offers temporal stability in social connections (as opposed to a network based solely on proximity), it gives rise to a large number of model parameters.

Back in the realm of continuous time, Scharf *et al.* (2018) jointly model the movement of multiple animals through a process convolution, which gives rise to a Gaussian process. One of the kernels of the process convolution relates to social behaviours, using weighted pairwise connections. That social kernel for an individual at a particular time is a weighted average of the convoluted process (so far) of those it is connected to at that moment (including itself). The network weights are dynamic and are obtained through a latent space model. Hooten *et al.* (2018) detail a similar approach. As well as accounting for the dependency amongst individuals, this framework can reduce uncertainty in individual trajectories. That is, when an individual lacks data, we can be more certain of their trajectories if they are well connected to animals who don't lack data.

Scharf & Buderman (2020) note that, despite the development of various modelling approaches as discussed above, much of the work on collective movement models has so far concentrated on 'positive' interactions (e.g. attraction or coordination), with more 'negative' interactions (e.g. territorial battles) receiving relatively little attention. Though, there are exceptions — Schlägel

et al. (2019) for instance.

1.2.5 Ad Hoc Analysis

Another route to investigate collective movement behaviours is through a more ad hoc approach. For instance, Strandburg-Peshkin *et al.* (2015) investigated how a troop of baboons collectively made their movement decisions through identifying 'movement initiations' in the spatial data. 'Movement initiations' were extracted through a method based on minima and maxima distances between a pair of animals. Through this approach, they quantified the probability of a baboon following a movement initiation in the context of the number of initiators and their consensus in direction. Though, whilst drawing out particular features in the data can allow flexible and easy to interpret analysis, it does not necessarily allow for causal relationships to be inferred. That is, in this particular example, a movement sequence of an initiator being followed doesn't necessarily imply a causal relationship between those two animals.

1.3 Heterogeneity

As mentioned throughout Section 1.2, individual heterogeneity plays a key role in social groups and thus accounting for such differences can provide a more complete picture of collective movement. Features we may want to consider from the point of view of individual heterogeneity include: physical differences (e.g. size); established social affiliations (e.g. kinship); physiological state (e.g. nutritional deprivation); resource acquisition vs predation risk (e.g. vulnerable animals will take a central, protected position in a group); aggression (e.g. lower ranking animals arriving at feeding sites late to avoid confrontation); information (e.g. experienced/informed group members position themselves at the front of the group) (del Mar Delgado *et al.*, 2018).

To briefly recap some examples of when individual variation has been taken into account in collective movement models: Delgado *et al.* (2014) define a sociability measure as the difference in the observed proximity of an individual to conspecifics and the expected proximity from a non-social ('null') model. Through this they are able to estimate individual heterogeneity in sociability. Using a model where the expected heading of movement is a weighted average of directional persistence, environmental features and social cues, Torney *et al.* (2018) investigate variation in social behaviour according to life stage. After classifying each caribou as either a calf, an adult or a large bull, they find that calves are more reliant on social cues than the other two categories and that their interaction is based more on maintaining proximity than alignment. Couzin *et al.* (2002) show that variation in movement parameters, such as speed and turning rate, relate to the position of an animal within a group. Schlägel *et al.* (2019) fit their movement model to each individual separately so they can obtain the different strengths at which they interact with others. The methods of Scharf *et al.* (2018), Hooten *et al.* (2018) and Schlägel *et al.* (2019) are able to ascertain the different strength/weighting of each dyadic interaction

Aside from providing more insight into social behaviours, utilising prior information on individual or demographic preferences can also help reduce the number of social parameters in the model (see Scharf *et al.* (2016) for a similar discussion). For example, if it is known which animals are still dependants, it may be assumed they mostly interact with their parents and so we can disregard other potential interactions.

There is also an interest in studying how environmental variation interplays with collective movement. Not only can doing so provide important insights into how social groups respond to that variation or change (Snijders *et al.*, 2017; King *et al.*, 2018), but it can also help disentangle social and nonsocial drivers of collective movement (Bode *et al.*, 2012). As we have seen above, some approaches to collective movement modelling have been designed to account for spatial heterogeneity. For example, the step selection approaches of Potts *et al.* (2014) and Strandburg-Peshkin *et al.* (2017) can account for the importance or preference of spatial features such as habitat types, landscape features (e.g. slope or forest canopy height), barriers to movement etc. Similarly, as mentioned above, the weighted heading of an animal can be influenced by features of the environment in the approach of Torney *et al.* (2018). Russell *et al.* (2017) model movement with a spatially varying stochastic differential equation to account for spatially dependent movement. However, in the models developed so far, the environmental and social drivers of movement are generally treated independently. We are then unable to examine how social behaviours transition in response to environmental change. Though, Strandburg-Peshkin *et al.* (2017) do measure group-level properties, such as the speed and alignment of a group, as a function of the environmental context.

Environmental variation in collective movement models has generally been limited to spatial heterogeneity. However, temporal heterogeneity can also be an important factor in social behaviours (Sueur *et al.*, 2011). For example, seasonal changes in prey abundance impact the social behaviours of lions (Mbizah *et al.*, 2020). Whilst some models account for the dynamism of social behaviours (Langrock *et al.*, 2014; Scharf *et al.*, 2018; Hooten *et al.*, 2018; Niu *et al.*, 2020), there has been little progress in making them dependent on temporal information.

1.4 Aim of this Thesis

We aim to develop movement models in the same vein as those discussed in Section 1.1.2, but with a flexible social framework so that we can capture the collective movement behaviours. We will do so in continuous time, in part because of the continuous nature of movement, but also because many of the issues with discrete-time modelling can be amplified when tracking multiple animals. That is, missing data is inherently more likely to occur, tracking data may not be synchronised between the animals and the temporal scale of the data may not be appropriate for the social behaviours under examination.

The behavioural switching models discussed in Section 1.1.2 provide an ideal platform to build upon. Behavioural switching in a collective movement model can account for the dynamism of social

behaviour, something which is largely ignored in the work discussed in this chapter despite frequent calls for its inclusion and its prevalence in nature (e.g. fission-fusion dynamics). Additionally, frameworks already exist that can be extended (or have already been extended) to include some heterogeneity (e.g. Blackwell *et al.* (2016)). In the context of analysing individual movement, this is typically done with spatial or temporal heterogeneity in mind but, in multivariate versions of these frameworks, extensions to individual heterogeneity should also be possible.

Sociality covers a vast scale, from dyadic relationships to swarms of locusts, and so some compromise will be needed as one modelling framework cannot possibly be flexible enough to suit all needs. So far, much of the collective movement modelling literature is focused on the extremes of this scale. For instance, the methods of Long *et al.* (2014), Joo *et al.* (2018) and Schlägel *et al.* (2019) are limited to studying dyadic relationships in isolation of other group members. The local-interaction-based models are typically developed with larger groups in mind, such as schooling fish or flocks of birds. Whilst those models work well in those contexts, they will have limited use for more complex or sparser social structures such as dominance hierarchies and fission-fusion dynamics. Additionally, they are not well suited to the coarser, longer-term data that is common in wildlife tracking studies (Calabrese *et al.*, 2018). Whilst the models developed by Langrock *et al.* (2014), Niu *et al.* (2016) and Niu *et al.* (2020) could be deployed across a range of social group sizes, their strengths lie in modelling larger groups too — such as reindeer/caribou herds as in their case studies. Moreover, the social structure in those models is always represented as a 'star' network, which may not be suitable for other types of social behaviours.

As both King *et al.* (2018) and Westley *et al.* (2018) discuss, there is much to be gained from obtaining more granular information about the social drivers of collective movement. For example, consideration of social hierarchies and 'keystone' animals can help guide conservation or population management efforts. Understanding if or how the fission-fusion dynamics of a group changes in response to environmental change can indicate how resilient that group is to that change. We are therefore focusing on capturing these complex behaviours in our work.

As discussed in Section 1.2.1, there is a large body of work that drills down to the pairwise relationships in order to piece together a view of the wider group structure. SNA and movement ecology is currently quite disconnected, but there is a wealth of insight to be gained by bridging that gap. Though, incorporating tournament-style networks into movement models, and the corresponding dependencies in movement, may be an overly complex endeavour (both in terms of model formulation and computational effort). However, the more tournament-style networks typically presented and analysed in SNA encompass all of the social interactions over some time period. In time-ordered SNA, the network at any given time is the result of the interactions occurring just at that moment. Such a framework in a behavioural switching movement model has the potential to provide detailed insight into complex, dynamic social behaviours. The methods presented by Scharf *et al.* (2018) and Hooten *et al.* (2018) capture a dynamic network, but these methods are better suited to scenarios when the social behaviours vary slowly relative to the movement process

and the inferred network is undirected. Capturing the direction of relationships is important for establishing asymmetric interaction, which is in turn important to obtaining insight such as whether there are key animals and if there is some hierarchical structure in the group.

1.4.1 Outline of this Thesis

In the following chapter, we provide a detailed overview of the modelling methods we will utilise to undertake the above aim. In particular, we will discuss: diffusion processes, both univariate and multivariate cases; behavioural state switching with a continuous-time Markov chain; algorithms to reconstruct behaviours and locations in continuous time; the Markov chain Monte Carlo methods used to fit the models to data.

In Chapter 3, we introduce the social framework and movement model that we use throughout this thesis. We present the model in detail, along with the corresponding algorithm we use in that chapter to fit it to data. This algorithm is largely based on work by Blackwell *et al.* (2016), which will be discussed in more detail in Chapter 2. We also discuss various model extensions that could better represent the analysis at hand. We trial our model with both simulated and zebra data.

The above approach is refined in Chapter 4. In particular, we develop an improved algorithm to reconstruct the animals' behaviour in continuous time. With this improvement, we investigate the robustness of the model to different social behaviours. Baboon data is analysed in this chapter to showcase the insights that we can gain when fitting our model to data from animals that exhibit complex social behaviours. Additionally, we investigate how the temporal scale of the data impacts the results and we compare the relationships our model inferred with those when applying dyadic metrics to the same data set.

In Chapter 5, we extend the model to the spatially heterogeneous case. We develop and present a new algorithm with which we can reconstruct both the behaviours and locations of the animals in continuous time. Analysis is undertaken to compare this new algorithm against other methods. With this new functionality in mind, and inspired by the models in Section 1.2.3, we also extend the model to include an inferred radius of interaction — a proximity that animals must be within in order to interact. Restricting overly distant interaction both adds some biological realism into the model and reduces the number of social interactions that need to be evaluated. We fit this new model to the same baboon data set as in Chapter 4 in order to examine this new functionality.

In Chapter 6, we discuss to what extent our models meet the above aims and where future developments should lie.

Chapter 2

Statistical Background

Methods, processes and results that are frequently referenced in this thesis are discussed in detail in this chapter. Namely, diffusion processes in Section 2.1, Markov chain Monte Carlo methods in Section 2.2 and methods with which we can reconstruct an animal's behaviour and location in continuous time in Section 2.3.

2.1 Diffusion Processes

The general form of a stochastic differential equation (SDE) comprises of a drift term and a noise term:

$$\mathrm{d}X_t = \mu(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}W_t \tag{2.1}$$

where X_t is the value of some process X at time t with $t \geq 0$. The function μ is the drift coefficient and, as the name suggests, its form dictates the nature of the drift (if any) exhibited by the process X. The function σ is the diffusion or 'noise' coefficient. It is considered the noise coefficient because W_t denotes a Wiener process, which is used to represent 'white noise' in X (Øksendal, 2003). The Wiener process is continuous, with independent, Gaussian increments such that:

$$W_t - W_0 \sim \mathrm{N}(0, t) \tag{2.2}$$

for all $t \geq 0$ where W_0 is the value of the process at time 0.

A diffusion process is a solution to an SDE (Øksendal, 2003). Diffusion processes are continuous-time, stochastic processes which also satisfy the Markov property: $P(X_{t_{n+1}}|X_{t_n}, X_{t_{n-1}}, ..., X_{t_1}) = P(X_{t_{n+1}}|X_{t_n})$ when $t_{n+1} \geq t_n \geq t_{n-1} \geq ... \geq t_1 \geq 0$ (Øksendal, 2003). The nature of the diffusion process depends on the form of the μ and σ functions in equation 2.1. We are focusing on two cases in particular, Brownian motion (BM) and Ornstein-Uhlenbeck (OU) processes, which are

both linear and Gaussian (Patterson *et al.*, 2017).

As an aside, SDEs that don't have tractable solutions can still be used to describe movement. Though they provide a flexible modelling option, it comes at the expense of computation and statistical tractability and thus they are not commonly used for movement models (Patterson *et al.*, 2017). As such, we will not discuss that approach.

2.1.1 Ornstein-Uhlenbeck

As mentioned in the previous chapter, OU processes are commonly used to model movement in continuous time — either as a positional or velocity process. We are focusing on the positional process, which was first used to model animal movement by Dunn & Gipson (1977) in order to estimate home ranges (see Chapter 1). In a single dimension, an OU process is described by the following SDE, which contains a linear drift or 'attraction' term:

$$\mathrm{d}X_t = -\beta(X_t - \theta)\mathrm{d}t + \rho\mathrm{d}W_t \tag{2.3}$$

An OU process is a stationary process that is mean-reverting to some central location. θ represents that location whilst β represents the rate of attraction to it. The diffusion coefficient is simply a constant value, ρ. There is a closed-form solution to equation 2.3, which is normally distributed:

$$X_t|X_0 \sim \mathrm{N}(e^{-\beta t}(X_0 - \theta) + \theta, \frac{\rho^2}{2\beta}(1 - e^{-2\beta t})) \tag{2.4}$$

with the distribution of X_t as $t \longrightarrow \infty$ being $\mathrm{N}(\theta, \frac{\rho^2}{2\beta})$ (Blackwell, 1997) — a unimodal distribution around the central location with $\frac{\rho^2}{2\beta}$ as the stationary variance. This follows on from the restriction that $\beta > 0$ so that $e^{-\beta t} \longrightarrow 0$ as $t \longrightarrow \infty$ (Dunn & Gipson, 1977). The expected value of the conditional distribution in equation 2.4 can be rearranged to the following: $e^{-\beta t}X_0 + (1 - e^{-\beta t})\theta$. Through this form, we can see that the expected location at X_t is a weighted mean of the location at X_0 and θ, one which is dictated by the time interval t and the strength of the attraction to θ (β).

Equations 2.3 and 2.4 both relate to the univariate case. We will consider two motivations for extending them to the multivariate case: to model the movement of multiple animals, which we will discuss in Section 2.1.3, and to model the movement of an individual in multiple dimensions, which we will discuss now.

The movement we are focusing on modelling is two-dimensional — longitudinal and latitudinal. Due to the linearity of the SDE in equation 2.3, two of them can be combined to create a two-dimensional OU process (or, in general, d of them combined to create a d-dimensional OU process).

The solution to this is:

$$\boldsymbol{X}_t|\boldsymbol{X}_0 \sim \text{MVN}(e^{\boldsymbol{B}t}(\boldsymbol{X}_0 - \boldsymbol{\Theta}) + \boldsymbol{\Theta}, \boldsymbol{\Sigma} - e^{\boldsymbol{B}t}\boldsymbol{\Sigma}e^{\boldsymbol{B}^T t}) \tag{2.5}$$

but now $\boldsymbol{\Theta}$ is a vector containing the centre of the process in each dimension whilst $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ are (2×2)-matrices — with $\boldsymbol{\Sigma}$ being the stationary covariance matrix. For biological plausibility, the $\boldsymbol{B}$ matrix is taken to be isotropic — symmetric under rotation and reflection (Blackwell, 1997). This is so that the drift is determined only by the distance to $\boldsymbol{\Theta}$, regardless of the direction to it, meaning that we avoid placing any significance on the coordinate system used for the measurements or by the animals themselves (Dunn & Gipson, 1977; Blackwell, 1997). Thus, $\boldsymbol{B} = \beta\boldsymbol{I}$ where $\boldsymbol{I}$ is the identity matrix and $\beta > 0$. Additionally, the noise coefficients of each dimension are typically treated independently ($\boldsymbol{\Sigma} = \frac{\rho^2}{2\beta}\boldsymbol{I}$). Due to the isotropic $\boldsymbol{B}$ and independent noise, each dimension is independent and can be modelled separately.

2.1.2 Brownian Motion

BM is the simplest diffusion process (Patterson *et al.*, 2017) and it is a continuous-time analogue of a random walk model. It is described by equation 2.1 when the drift coefficient is set to 0 and the diffusion coefficient is set to a constant, ρ say. As such, BM only has a single parameter, ρ, which scales the variance of a BM step:

$$X_t|X_0 \sim \text{N}(X_0, \rho^2 t) \tag{2.6}$$

for all $t \geq 0$. In the context of movement modelling and when X is a positional process, the ρ parameter can be considered to represent the speed of the movement.

BM is the limiting case of an OU process as the attraction parameter tends to 0 (Blackwell, 1997). To show this, let's take the conditional distribution of an OU process as in equation 2.4:

$$X_t|X_0 \sim \text{N}(e^{-\beta t}(X_0 - \theta) + \theta, \frac{\rho^2}{2\beta}(1 - e^{-2\beta t})) \tag{2.7}$$

As $\beta \longrightarrow 0$, $e^{-\beta t}(X_0 - \theta) + \theta \longrightarrow X_0$ and $\frac{\rho^2}{2\beta}(1 - e^{-2\beta t}) \longrightarrow \rho^2 t$ after expanding the power series of $e^{-2\beta t}$ and cancelling out the βs. Thus, the resulting distribution is the same as in equation 2.6.

Due to its simplicity, BM has limitations for modelling long-term movement. However, it does have useful applications in diffusion-switching models in order to provide some flexibility (Blackwell, 1997). That is, it may not be realistic to assume an animal is always moving with some drift to a central location and in a stationary manner — when an animal is exploring or foraging for example. As BM doesn't have drift, nor is it stationary, it may be better suited to modelling that type of movement.

2.1.3 Multivariate Diffusion

As mentioned above, the univariate SDEs can be combined to jointly model the movement of multiple animals. Note, different spatial dimensions, or axes, are treated independently and so the discussion below just relates to a single axis, with corresponding equations for further axes as required.

To model the social behaviours in collective movement, Niu *et al.* (2016) utilise a multivariate OU process to capture interaction explicitly. With the 'leading' animal (which can be an actual animal or some abstract leading point) moving according to equation 2.3, with θ being some fixed location such as a nesting site, they introduce a second animal — the 'follower.' The movement of the follower, F, is also modelled with an SDE of the form seen in equation 2.3, the centre of which is the location of the leader (L):

$$\mathrm{d}F_t = -\alpha(F_t - L_t)\mathrm{d}t + \sigma \mathrm{d}V_t \tag{2.8}$$

where α is the rate of attraction to the leader and σ is the diffusion coefficient for the follower. V_t is a distinct Wiener process from that of the leader. The bivariate OU process of these two animals for a single axis is then:

$$\mathrm{d}\boldsymbol{X}_t = \boldsymbol{A}(\boldsymbol{X}_t - \boldsymbol{\Theta})\mathrm{d}t + \boldsymbol{\Sigma}\mathrm{d}\boldsymbol{B}_t \tag{2.9}$$

where, using mostly the same notation as Niu *et al.* (2016),

$$\boldsymbol{X}_t = \begin{pmatrix} L_t \\ F_t \end{pmatrix}, \boldsymbol{A} = \begin{pmatrix} -\beta & 0 \\ \alpha & -\alpha \end{pmatrix}, \boldsymbol{\Theta}_t = \begin{pmatrix} \theta \\ \theta \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \rho & 0 \\ 0 & \sigma \end{pmatrix}, \boldsymbol{B}_t = \begin{pmatrix} W_t \\ V_t \end{pmatrix}.$$

L_t is the positional process of the leader, as described by equation 2.3, and so the $\boldsymbol{X}_t$ vector represents the locations of both animals. $\boldsymbol{A}$ is the attraction matrix, which encodes the interactions within a group. In general for larger groups, which Niu *et al.* (2016) take to consist of a single leader with all the remaining animals as its followers, the first row of $\boldsymbol{A}$ relates to the 'leader' and it consists of zeros except for $-\beta$ on the diagonal. All other rows relate to followers, with $-\alpha$ on the diagonal, α in the first column and all other columns set to zero. $\boldsymbol{\Theta}$ is a vector with all entries being θ, indicating that followers are indirectly attracted to θ too. $\boldsymbol{\Sigma}$ is a diagonal matrix as we take the noise for each animal to be independent.

This too has a closed-form solution (Niu *et al.*, 2016), which is a multivariate normal distribution:

$$\boldsymbol{X}_t | \boldsymbol{X}_0 \sim \mathrm{MVN}(e^{\boldsymbol{A}t}(\boldsymbol{X}_0 - \boldsymbol{\Theta}) + \boldsymbol{\Theta}, \boldsymbol{\Delta} - e^{\boldsymbol{A}t}\boldsymbol{\Delta}e^{\boldsymbol{A}^T t}) \tag{2.10}$$

where $\boldsymbol{\Delta}$ is the multivariate stationary covariance. In order to obtain the conditional covariance

then, we first need to derive $\boldsymbol{\Delta}$. This can be done using the following result (Niu *et al.*, 2016):

$$\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} = -(\boldsymbol{A}\boldsymbol{\Delta} + \boldsymbol{\Delta}\boldsymbol{A}^T) \tag{2.11}$$

We will not go into the derivation here, but we take a similar approach to derive the conditional covariance for the modelling framework we introduce in Chapter 3 (the full details of which are in Appendix A).

Niu *et al.* (2016) go on to expand this social framework for more animals where each subsequent animal is a follower, resulting in a larger, single group where all followers are attracted to the same animal (or leading point). However, there is scope for this approach to be more flexible in the social behaviours it can model. For instance, there could be multiple leaders in order to capture subgroups. Then, not all followers have to follow the same leader, with the $\boldsymbol{A}$ matrix detailing the specific interactions. Equations 2.10 and 2.11 can then be utilised to derive the corresponding conditional covariance matrix.

Additionally, as it is the limiting case of an OU process, there is also scope to include BM in this framework. For instance, when combining this framework with state switching, switching to BM would allow followers some time independent of the group (as in Niu *et al.* (2020)). BM could also be used to model the movement of leading animals when collective movement without drift is a suitable model. However, in both of these cases, some care is needed to avoid division-by-zero errors in both the stationary and conditional covariance matrices as some rates of attraction will be zero. See Section 2.1.2 for an example. We explore all of the above extensions in Chapter 3.

2.2 Markov Chain Monte Carlo

In this thesis, we are focusing on Bayesian inference, which involves estimating the posterior distribution of the model parameters, conditional on the data:

$$P(\Omega|x) = \frac{P(x|\Omega)P(\Omega)}{\int P(x|\Omega)P(\Omega)d\Omega} \tag{2.12}$$

where Ω represents the model parameters; x is the observed data; $P(x|\Omega)$ is the likelihood; $P(\Omega)$ is the prior (a distribution encompassing prior knowledge of Ω); the denominator is the marginal of the data and a proportionality constant. Due to the integral in the proportionality constant, the posterior distribution is often not analytically obtainable. Additionally, the data itself might not be fully observed. For example, with a state-switching model, the posterior of the movement parameters depends on the sequence of a typically unobserved behavioural process. It also depends on the timings of the state switches which are, again, typically unobserved in continuous-time models. Thus, in order to analytically derive the posterior distributions of the movement parameters, we would also need to integrate out the possible state sequences and timings of the switches.

Monte Carlo methods can be used to estimate the properties of a distribution that cannot be obtained analytically, thus providing an approximation of that distribution (Gilks *et al.*, 1996). This is achieved by drawing a large enough number of samples from the 'target' distribution, π, in some way. $\pi = P(\Omega|x)$ in the above Bayesian context. When the samples are generated through a (discrete-time) Markov chain, with π as its stationary distribution, this gives rise to a Markov chain Monte Carlo (MCMC) algorithm (Gilks *et al.*, 1996). This can be done with the Metropolis-Hastings (MH) algorithm. Let Ω_t denote the value of the Markov chain at time t, for $t \geq 1$. A candidate for the value at $t+1$, say S, is sampled from a proposal distribution, Q. S is accepted as the value of Ω_{t+1} with probability $\alpha(\Omega_t, S)$ where

$$\alpha(\Omega_t, S) = \min\left(1, \frac{\pi(S)Q(\Omega_t|S)}{\pi(\Omega_t)Q(S|\Omega_t)}\right) \tag{2.13}$$

Otherwise, S is rejected and $\Omega_{t+1} = \Omega_t$. The second term passed in the minimum function is known as the 'MH ratio'. The resulting transition kernel of this Markov chain is:

$$P(\Omega_{t+1}|\Omega_t) = Q(\Omega_{t+1}|\Omega_t)\alpha(\Omega_t, \Omega_{t+1}) + \mathbf{1}_{\Omega_{t+1}=\Omega_t}\left(1 - \int Q(S|\Omega_t)\alpha(\Omega_t, S)dS\right) \tag{2.14}$$

with the first term of the kernel relating to the acceptance of the proposal and the indicator function term relating to the rejection of all possible candidates (Gilks *et al.*, 1996). The proposal distribution Q can take any form and the stationary distribution of this Markov chain will still be π (Gilks *et al.*, 1996). Note, though, that the MH ratio will simplify when Q is symmetric (as $Q(\Omega_t|S) = Q(S|\Omega_t)$). Additionally, when Q is a Gaussian distribution centred on Ω_t, the resulting process is known as a random-walk MH.

As mentioned above, Monte Carlo methods require some means of generating samples from the target distribution — a distribution which most likely won't be analytically tractable due to the integral in the proportionality constant. However, the MH algorithm circumvents this issue as, when calculating the ratio of the posterior distribution evaluations, the proportionality constants cancel out.

The above MH algorithm works regardless of whether Ω is a single parameter or a vector of parameters. If it is a vector, the algorithm can either be implemented to update all of the parameters in a single step, or the vector can be split up into disjoint sets — with each set being updated in turn (Gilks *et al.*, 1996). Whilst we are updating a set of parameters, we condition on the latest values of the other parameters. Say the set of parameters we are updating in the current step is Ω_u, with Ω_{-u} denoting the remaining parameters. Let $\Omega_{t,u}$ represent the values of Ω_u at time t and $\Omega_{t,-u}$ represent the latest values of Ω_{-u}:

$$\Omega_{t,-u} = (\Omega_{t+1,1}, ..., \Omega_{t+1,u-1}, \Omega_{t,u+1}, ..., \Omega_{t,n}) \tag{2.15}$$

where n is the number of sets and sets $\Omega_1, ..., \Omega_{u-1}$ have already been updated in this iteration. New

candidates, S_u, are proposed for $\Omega_{t+1,u}$ and they are accepted with the following probability:

$$\alpha(\Omega_{t,u}, \Omega_{t,-u}, S_u) = \min\left(1, \frac{\pi(S_u|\Omega_{t,-u})Q(\Omega_{t,u}|S_u, \Omega_{t,-u})}{\pi(\Omega_{t,u}|\Omega_{t,-u})Q(S_u|\Omega_{t,u}, \Omega_{t,-u})}\right) \tag{2.16}$$

Otherwise, $\Omega_{t+1,u} = \Omega_{t,u}$. $\pi(S_u|\Omega_{t,-u})$ is called the 'full conditional' distribution for S_u. If it is possible to propose new candidates from the full conditional distribution (i.e. $Q(S_u|\Omega_{t,u}, \Omega_{t,-u}) = \pi(S_u|\Omega_{t,-u})$), then the acceptance probability is always 1 — a method known as Gibbs sampling. A common scenario in which Gibbs sampling is used is when there is a suitable conjugate prior for the likelihood of Ω_u.

Whilst the MH algorithm is a simple means to construct the required Markov chain, the implementation of it and the subsequent analysis of the output requires careful attention (van Ravenzwaaij *et al.*, 2018). Consideration must be made to the starting value of the Markov chain, Ω_1. It is considered good practice to initialise the Markov chain with 'over-dispersed' values from the target distribution, in part so that a wider range of the parameter space can be more readily explored. As such, the early part of the Markov chain will not have been sampled from the stationary distribution. This part of the chain is considered the 'burn-in' period and is discarded before any analysis of the chain is undertaken. Inspection of the chain's trace plots can help determine this period.

The choice of proposal distribution effects how well the algorithm 'mixes' (van Ravenzwaaij *et al.*, 2018). For example, a proposal with an overly small variance will produce a high acceptance rate, but it will take a long time to explore the parameter space and the samples may get stuck at local maxima. A proposal distribution with an overly large variance will result in a low acceptance rate as it will frequently propose values that are beyond the tails of the target distribution. Both of these extremes decrease the independence between successive values of the Markov chain. Markov chains with a high level of autocorrelation will need to run for longer than a counterpart with a low level of autocorrelation in order to achieve the same effective sample size (ESS) with which we can approximate the target distribution. The parameters of the proposal distributions are therefore called 'tuning parameters.'

We also need to evaluate whether the chain (beyond the burn-in period) has converged to the target distribution. Informal techniques to do this include inspecting the trace plots and running the MCMC algorithm multiple times from different starting values to confirm (or not) if all runs produce the same posterior properties. A common tool used to assess convergence is the Gelman-Rubin diagnostic (Gelman & Rubin, 1992; Brooks & Gelman, 1998). The Gelman-Rubin diagnostic evaluates the convergence of multiple chains (relating to the same parameter(s)), with each one being initialised with over-dispersed values. The diagnostic is called the 'potential scale reduction factor' (PSRF) and it is a ratio of a weighted average of the within-chain and between-chain variances compared to the within-chain variance. The PSRF therefore tends to 1 (from above) as the multiple chains converge to the same distribution. There is no set figure to indicate convergence,

but PSRF values less than 1.1 or 1.2 (depending on the level of rigour required) are sometimes used (Brooks & Gelman, 1998). The univariate PSRF regards the convergence of a single parameter in Ω. There is also a multivariate PSRF (MPSRF), which evaluates within-chain and between-chain covariance, for when there is expected to be some correlation between parameters. The MPSRF is the maximum scale reduction factor of any linear combination of Ω and bounds above the largest univariate PSRF (Brooks & Gelman, 1998).

2.3 Modelling Behavioural Switching in Continuous Time

2.3.1 Continuous-Time Markov Chain

As mentioned in Chapter 1, we are focusing on movement models that allow the animals to switch between a finite number of 'behaviour' states. Regardless of whether time is being treated as discrete or continuous, a common approach is to assume the behavioural process is Markovian. Whilst the Markov assumption is simplistic, it still accounts for temporal stability in an animal's behaviour (Scharf *et al.*, 2016) and so it represents a good compromise between model complexity and tractability (Patterson *et al.*, 2017). A continuous-time Markov chain (CTMC) is therefore a natural choice with which to model switching behaviours in continuous time.

A CTMC, with a finite state space, is described by its generator matrix, Λ. The off-diagonal elements of Λ (λ_{uv} for $u, v = 1, ..., n$ where n is the number of states and $u \neq v$) represent the infinitesimal transition rate from state u to state v. The diagonal elements of Λ are determined by the off-diagonal ones: $\lambda_{uu} = -\lambda_u$ where λ_u is the rate of leaving state u ($\lambda_u = \sum_{u \neq v} \lambda_{uv}$).

Whilst a generator matrix and its transition rates are perhaps not as immediately intuitive as the transition probability matrix of a discrete-time Markov chain, there is a simple interpretation of a CTMC. That is, the Markovian process will stay in state u for a holding time, which is exponentially distributed with rate λ_u. After this holding time, the animal switches to state v with probability $p_{uv} = \lambda_{uv}/\lambda_u$. State switches are thus drawn from a single-trial multinomial distribution. When discussing CTMCs, we will always do so with regards to this parameterisation of it.

Let m be a completely observed CTMC at times $0 < t_1 < ... < t_n$, with $t_0 = 0$; m_i representing the state of the process at time t_i; h_i representing the holding time from time t_i such that $h_i = t_{i+1} - t_i$. The likelihood of a generator matrix, Λ, given the observed CTMC, is a product of the likelihoods of each holding time and the subsequent state switch:

$$L(\Lambda|m) = \prod_{i=0}^{n-1} \lambda_{m_i} e^{-\lambda_{m_i} h_i} p_{m_i m_{i+1}} \tag{2.17}$$

2.3.2 Reconstructing Switching Diffusion Processes

Movement data is usually collected at discrete intervals. Even if it's not, it is unlikely to be collected at times of behaviour changes. Therefore, to facilitate modelling behaviour in continuous time, we need to reconstruct the behavioural process between the observations. Blackwell (2003) detail a method to do this in the spatially homogeneous case and when the behavioural process is known at the observations. However, the behavioural process will typically be completely unobserved and thus needs to be reconstructed in its entirety. Additionally, we may want the switches in behaviour to be dependent on location, as in the spatially heterogeneous case. We will then also need to reconstruct the movement of an animal at the times of those switches. A flexible method to do all of that was introduced by Blackwell *et al.* (2016) — a method which we will refer to as the 'kappa' method. In both Blackwell (2003) and Blackwell *et al.* (2016), inference is undertaken with MCMC techniques, with the posteriors of the movement parameters, transition parameters and trajectories being sampled from separately. For example, new behavioural/movement trajectories are proposed conditional on the current movement and transition parameter estimations. Both of the above methods are presented in the context of modelling the movement of an individual animal.

The kappa method was developed with the aim of enabling exact Bayesian inference (that is, avoiding approximations due to discrete time) for a flexible class of models (Blackwell *et al.*, 2016). The key to this method is the reconstruction/simulation of where an animal is and what it is doing when it is unobserved. This is done by augmenting the observed data with sampled state switching times and locations. These movement and behavioural trajectories are repeatedly sampled and subsequently evaluated with a MH ratio, a process which accounts for the uncertainty regarding the unobserved parts of an animal's trajectory.

A key component to this method is the parameter κ. κ has no biological interpretation, but it serves as an upper bound of the transition rates: $\kappa \geq \max(\lambda_u)$ for all states u. The holding times of all of the states are therefore probabilistically bounded below by the exponential distribution with rate κ (Blackwell *et al.*, 2016). This means, in the spatially heterogeneous case, an animal's behaviour is not forced to change instantaneously when it changes habitat — a concept that aligns with the notion of 'separable' models as discussed by Harris & Blackwell (2013). That is, whilst transition rates can depend on location, the movement parameters and trajectories do not (though they are indirectly dependent on location via the behaviour states). Say, then, the behavioural process starts at $t_0 = 0$ in state m_0. Rather than sample a holding time with rate λ_{m_0}, we do so with rate κ ($t_1 \sim \text{Exponential}(\kappa)$). A state switch does not necessarily occur at t_1 as $\kappa \geq \lambda_{m_0}$, merely t_1 is a potential switching time. A state switch occurs at t_1 with probability λ_{m_0}/κ. If that's the case, the new state is state v with probability $\lambda_{m_0 v}/\lambda_{m_0}$. In general, $t_i - t_{i-1} \sim \text{Exponential}(\kappa)$ and t_i is only ever a potential switching time, regardless of the state at t_{i-1}. The state at t_{i+1} is

drawn from the single-trial multinomial distribution:

$$m_i = \begin{cases} v & v \text{ with probability } \lambda_{m_{i-1}v}/\kappa \text{ when } v \neq m_{i-1} \\ m_{i-1} & \text{otherwise.} \end{cases} \tag{2.18}$$

As the time difference between each sequential pair of potential switching times is exponentially distributed with rate κ, the set of potential switching times over some time interval can be represented as a Poisson point process, also with rate κ (Blackwell *et al.*, 2016). Note, in the above discussion of state switches in the kappa method, we have used a simplified notation. The transition rates can be dependent on location and time (see Blackwell *et al.* (2016)). For example, the rate $\lambda_{m_{i-1}v}$ could read $\lambda_{m_{i-1}v}(t_i, x_i)$ — the rate of switching to state v at time t_i when at location x_i.

With this method, trajectories are typically simulated through small segments of data, rather than the entire data set, as simulations tend to deviate further from a reasonable realisation the longer it is. So, say we want to simulate both the movement and behavioural processes through the interval $[t_a, t_b]$, where t_a and t_b are observed times. We first propose a new set of potential switching times, $\hat{\tau}_s$, for this interval from a Poisson(κ) process. Then, let τ_o represent the observed data times in the interval $[t_a, t_b]$ and $\hat{\tau} = \tau_o \cup \hat{\tau}_s$ so that $\hat{\tau} = \{\hat{\tau}_1 < ... < \hat{\tau}_{\hat{p}}\}$, where $\hat{p}$ is the size of $\hat{\tau}$. Additionally, let $\hat{x}_i$ and $\hat{m}_i$ represent the location and state respectively at $\hat{\tau}_i$. Starting with x_1 and $\hat{m}_1 = m_1$, we simulate the trajectory forwards through each time in $\hat{\tau}$:

- If $\hat{\tau}_i \in \tau_o$ for $i = 2, ..., \hat{p}$, the behavioural state at $\hat{\tau}_i$ is carried forward from $\hat{\tau}_{i-1}$ ($\hat{m}_{i-1}$) as observed times are not part of the behavioural process.
- If $\hat{\tau}_i \in \hat{\tau}_s$ for $i = 2, ..., \hat{p}-1$, we simulate $\hat{x}_i$ from the movement process corresponding to state $\hat{m}_{i-1}$. Then, the state $\hat{m}_i$ is sampled using equation 2.18

As this trajectory proposal only concerns the interval $[t_a, t_b]$, it must remain consistent with the augmented data outside of that interval — hence why we initialised the simulation with state m_1. Additionally, the state at the final switching time, $\hat{\tau}_{\hat{p}-1}$, must be consistent with the currently estimated state at t_b. If not, the simulation is automatically rejected.

If that condition is met, the trajectory proposal is accepted or rejected with a MH ratio. For each interval in the segment of data being updated, the sampled trajectory (both in terms of the movement and behaviours) is proposed from the model itself, and so the distribution of that trajectory up to (and including) the final switching time cancels out with the proposal distribution. The MH ratio then only concerns the likelihood of the observed data conditional on the locations sampled at those final switching times (or the previous observed location if an interval doesn't contain a switching time):

$$\frac{\prod_{i \in [2:\hat{p}] \wedge \hat{\tau}_i \in \tau_o} f(\hat{x}_i | \hat{x}_{i-1}, \hat{m}_{i-1}, \hat{\tau}_i, \hat{\tau}_{i-1}, \Omega)}{\prod_{i \in [2:p] \wedge \tau_i \in \tau_o} f(x_i | x_{i-1}, m_{i-1}, \tau_i, \tau_{i-1}, \Omega)} \tag{2.19}$$

where Ω represents all movement parameters. p, x_{i-1}, m_{i-1} and τ_{i-1} relate to the size, locations, states and times of the previously accepted trajectory.

Whilst the kappa method was designed for spatial heterogeneity, the same framework can also be utilised in the spatially homogeneous case. In the spatially homogeneous case, when proposing a new trajectory, only the behavioural process is required to be reconstructed. The MH ratio thus consists of terms relating to the likelihood of the observed movement through the proposed state switches. Blackwell (2003) details how to calculate those likelihood terms when the movement is modelled with the diffusion processes outlined in Section 2.1 — an approach we utilise and discuss further in Chapter 3 and Appendix C.

κ can be defined to be fixed or it can be allowed to fluctuate — either within the MCMC process (during the burn-in period) or dependent on time. Typically, it is taken to be fixed. As such, during transition rate updates, the likelihood of the rates consists only of a product of single-trial multinomial likelihoods, similar to the probabilities given in equation 2.18, with a term for each potential switching time. In the case where the rates do not depend on location (or anything else), each potential switching time has an identical likelihood. Therefore, we can use a conjugate Dirichlet prior to obtain the full conditional distribution of the rates and resample them using Gibbs sampling (Blackwell *et al.*, 2016).

Chapter 3

Modelling Social Interaction

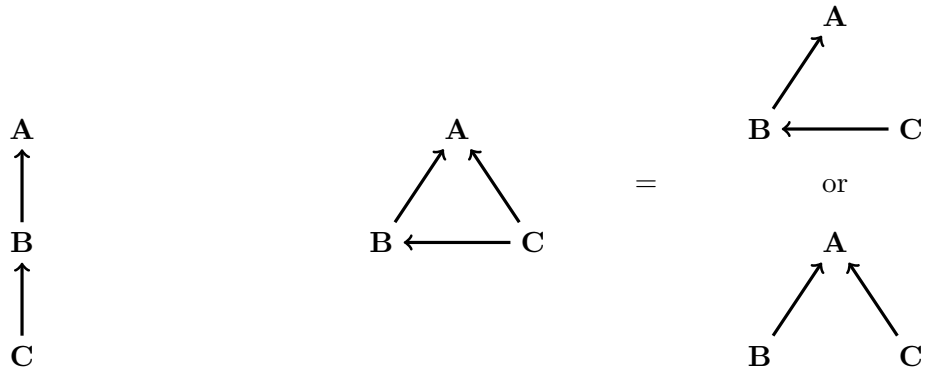
In this chapter, we introduce a continuous-time movement model containing a social framework that allows us to capture the social drivers of animal movement. We describe the social framework and assumptions in Section 3.1 before detailing the movement model within which that framework resides (Section 3.2). The intuition of our approach is presented in Section 3.3. The Markov chain that facilitates dynamic social behaviours is detailed in Section 3.4, followed by a discussion of various extensions that can be made to the model in Section 3.5. The inference algorithm with which we fit the model to data is presented in Section 3.6 before finally doing so to simulated data (Section 3.7) and zebra tracking data (Section 3.8). We then make some concluding remarks on this chapter in Section 3.9. The sections that introduce the concept of the model (Sections 3.1, 3.2 and 3.4) and some of Section 3.9 have been presented in Milner *et al.* (2021).

3.1 Influence Hierarchies

To capture the social drivers of a group's movement, our assumption is: a period of direct interaction between two animals can be characterised by the movement of one of those animals being attracted to the other. This is an assumption shared with other areas of animal movement literature (Long *et al.*, 2014) and one that considers there to be some social ordering in an interaction.

For ease of reference, we will refer to the roles in this dyadic relationship as 'dominant' and 'subordinate' (see the Social Definitions box on page 31) but, as with all behaviour modelling, we need to be careful not to over-interpret the behaviour labelling. That is, 'subordinate' or 'attracted to' have certain connotations but the movement behaviour just broadly translates to the movement of an animal being influenced by the other in some sense. Similar considerations also need to be made for our subgroup-level labels of 'leader' and 'follower'.

This dyadic concept can be naturally extended for larger social groups to give rise to social hierarchies (see Figure 3.1a). In order to keep this framework tractable and easy to interpret, we are



(a) (b)

Figure 3.1: In (a), whilst C is subordinate to B, we see B is in turn subordinate to A which enables us to learn how the influence in movement is cascaded through the group. In (b), we have a transitive triad where A dominates B, B dominates C and A dominates C. We only capture the influence that best describes C's movement and consequently the structure we estimate will either be A dominates B, B dominates C or A dominates B, A dominates C. In all of the above hierarchies, A is the leader whilst B and C are followers.

restricting these hierarchies to essentially a thinned network that contains the edges representing the most causal interaction. That is, an animal can have at most one dominant but it can have multiple subordinates. This is as opposed to a tournament-style network where there is some degree of relationship between every pair of nodes, which is a rare occurrence in nature (McDonald & Shizuka, 2013). Thus, our resulting social structure will represent the most causal influence to explain a group's movement — hence 'influence hierarchies' (see Figure 3.1b). We restrict the possible hierarchies to avoid any cycles, to ensure that the pattern of relationships is meaningful and that the movement models will be well-defined. The hierarchical structure is therefore what is often known in statistical contexts as a directed acyclic graph (DAG).

Social Definitions

Dominant: with respect to an edge in the hierarchy between animals i and j where j is attracted to i, i is dominant towards j.
Follower: an animal in a subgroup that is not the leader.
Group: all animals in the data set.
Independent: an animal is independent if it is neither a dominant or a subordinate.
Leader: the focal point of a subgroup. That is, this animal is a dominant to at least one animal, but a subordinate to none. Animal i's leader is the leading animal of the subgroup i is in.
Subgroup: all animals in a single hierarchy structure, i.e. a component of the graph formed as in Figure 3.1. Independent animals are their own subgroup.
Subordinate: with respect to an edge in the hierarchy between animals i and j where j is attracted to i, j is subordinate to i.

3.2 Movement Process

To model the movement of the animals, we use a diffusion process including a linear attraction term to represent the attraction-based interaction we have assumed. Then the movement of an individual animal i that is attracted to animal j can be described by the following stochastic differential equation (SDE):

$$\mathrm{d}A^y_{i_t} = -\alpha(A^y_{i_t} - A^y_{j_t})\mathrm{d}t + \sigma \mathrm{d}W^y_{i_t} \tag{3.1}$$

where $A^y_{i_t}$ is the location of animal i at time t in the y coordinate; α is the rate of attraction towards $A^y_{j_t}$ where $i \neq j$; σ is the coefficient of 'noise', the component of movement modelled not in terms of social interaction but as Brownian motion (BM, with $W^y_{i_t}$ representing the Wiener process). Equation 3.1 has two components: the noise term, which is a continuous-time analogue of a random walk, and the attraction term, which captures any persistence in the movement towards another animal. For leading and independent (that is, non-subordinate) animals, this reduces to BM as they have no attraction term:

$$\mathrm{d}A^y_{j_t} = \rho \mathrm{d}W^y_{j_t} \tag{3.2}$$

where ρ is a distinct noise parameter. The movement of non-subordinate animals is therefore non-stationary. This approach is similar to that of Niu *et al.* (2020), and covers the cases when self-driven animals either have no attraction to some particular point/area or the attraction is not relevant to the time scale of the data. Note, though, compared to Niu *et al.* (2020) (and Niu *et al.* (2016)), our definition of a 'follower' is more relaxed. We do not restrict followers to only be attracted to a leading animal, they can be attracted to any animal. As such, we can capture social

structures other than 'star' shaped networks.

The x and y coordinates are treated as independent (see Blackwell (1997) and Chapter 2 for justification) and so there are corresponding equations for the x axis with common parameters α, σ and ρ but with independent Wiener processes. That is, $W_{i_t}^y$ is independent of $W_{i_t}^x$ and $W_{j_t}^y$ is independent of $W_{j_t}^x$.

Because of their linearity, these univariate SDEs can be combined into a multivariate Ornstein-Uhlenbeck process to model the group's movement jointly as detailed by Niu *et al.* (2016). For n animals, say, in the y axis:

$$\mathrm{d}\boldsymbol{G}_t = \boldsymbol{F}_t(\boldsymbol{G}_t - \boldsymbol{\Theta}_t)\mathrm{d}t + \boldsymbol{\Sigma}\mathrm{d}\boldsymbol{B}_t \tag{3.3}$$

where

$$\boldsymbol{G}_t = \begin{pmatrix} A_{1_t}^y \\ \vdots \\ A_{n_t}^y \end{pmatrix}, \boldsymbol{\Theta}_t = \begin{pmatrix} L_{1_t}^y \\ \vdots \\ L_{n_t}^y \end{pmatrix}, \boldsymbol{B}_t = \begin{pmatrix} W_{1_t}^y \\ \vdots \\ W_{n_t}^y \end{pmatrix}$$

and $L_{i_t}^y$ is the location of animal i's leader at time t. $\boldsymbol{F}_t$ and $\boldsymbol{\Sigma}$ are $(n \times n)-$matrices where

$$\boldsymbol{F}_{t_{i,j}} = \begin{cases} -\alpha, & i = j \text{ and } i \text{ is a subordinate} \\ \alpha, & i \neq j \text{ and } i \text{ is subordinate to } j \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\boldsymbol{\Sigma}_{i,j} = \begin{cases} \sigma, & i = j \text{ and } i \text{ is a subordinate} \\ \rho, & i = j \text{ and } i \text{ is a leader or independent} \\ 0, & \text{otherwise.} \end{cases}$$

$\boldsymbol{G}_t$ is a vector of locations for all animals at time t and matrix $\boldsymbol{F}_t$ is the attraction matrix for the group which details the interactions within the hierarchy at time t, that is, who is subordinate to whom. All of the social structures possible in Section 3.1 are able to be represented in $\boldsymbol{F}_t$. $\boldsymbol{\Theta}_t$ is a vector which contains the location of each animal's leader at time t and the diagonal matrix $\boldsymbol{\Sigma}$ contains the coefficient of noise for each animal.

The solution to the multivariate SDE is given by

$$\boldsymbol{G}_t = e^{\boldsymbol{F}_0 t}(\boldsymbol{G}_0 - \boldsymbol{\Theta}_0) + \boldsymbol{\Theta}_0 + \int_0^t \boldsymbol{\Sigma} e^{\boldsymbol{F}_0(t-s)}\mathrm{d}\boldsymbol{B}_s \tag{3.4}$$

with the following closed form solution — a multivariate normal distribution (Niu *et al.*, 2016):

$$\boldsymbol{G}_t|\boldsymbol{G}_0 \sim \mathrm{MVN}(\boldsymbol{\mu}(\boldsymbol{G}_0, \boldsymbol{F}_0, t), \boldsymbol{\Xi}(\boldsymbol{F}_0, t)) \tag{3.5}$$

where $\boldsymbol{G}_0$, $\boldsymbol{F}_0$ and $\boldsymbol{\Theta}_0$ correspond to the animals' locations, the attraction matrix and the leaders' locations at time 0 respectively. The OU process is Markovian with the animals' locations at time t conditional on their previous locations at time 0.

The expected value of this distribution is given by:

$$E[\boldsymbol{G}_t|\boldsymbol{G}_0] = \boldsymbol{\mu}(\boldsymbol{G}_0, \boldsymbol{F}_0, t) = e^{\boldsymbol{F}_0 t}(\boldsymbol{G}_0 - \boldsymbol{\Theta}_0) + \boldsymbol{\Theta}_0 \tag{3.6}$$

and $\text{Var}[\boldsymbol{G}_t|\boldsymbol{G}_0]$ is given by $\boldsymbol{\Xi}(\boldsymbol{F}_0, t)$ which consists of the following five expressions (to ease notation, we simplify $\boldsymbol{\Xi}(\boldsymbol{F}_0, t)$ to $\boldsymbol{\Xi}$ and $\boldsymbol{F}_0$ to $\boldsymbol{F}$):

$$\boldsymbol{\Xi}_{i,j} = \begin{cases} \rho^2 t, & \text{(a)} \\ \boldsymbol{\Xi}_{\text{dom}(i),j} - \frac{\rho^2 e^{\boldsymbol{F}t}_{i,j}}{\alpha}, & \text{(b)} \\ \frac{\boldsymbol{\Xi}_{\text{dom}(i),j} + \boldsymbol{\Xi}_{i,\text{dom}(j)}}{2} - \frac{\rho^2 e^{\boldsymbol{F}t}_{i,l} e^{\boldsymbol{F}t}_{j,l}}{2\alpha} - \frac{\sigma^2 (e^{\boldsymbol{F}t}_{i,-l} \cdot e^{\boldsymbol{F}t}_{j,-l})}{2\alpha}, & \text{(c)} \\ \boldsymbol{\Xi}_{i,\text{dom}(i)} - \frac{\rho^2 e^{\boldsymbol{F}t}_{i,l} e^{\boldsymbol{F}t}_{i,l}}{2\alpha} + \frac{\sigma^2 (1 - e^{\boldsymbol{F}t}_{i,-l} \cdot e^{\boldsymbol{F}t}_{i,-l})}{2\alpha}, & \text{(d)} \\ 0, & \text{(e)} \end{cases} \tag{3.7}$$

where $\text{dom}(i)$ is the dominant of animal i; l is the leader of the subgroup both i and j are in; $-l$ indicates all animals except l. The scenarios of (a) to (e) are as follows:

(a) $i = j$ and i is a leading or independent animal.

(b) $i \neq j$ and j is the leader of i's subgroup.

(c) $i \neq j$, i and j are in the same subgroup and neither are the leader.

(d) $i = j$ and i is a subordinate.

(e) $i \neq j$ and i and j are in different subgroups.

The derivation of these terms is provided in Appendix A. In practice, the terms of $\boldsymbol{\Xi}(\boldsymbol{F}_0, t)$ need to be calculated in a specific order as some expressions rely on other values within $\boldsymbol{\Xi}(\boldsymbol{F}_0, t)$. See Appendix B for the algorithm.

3.3 Model Interpretation

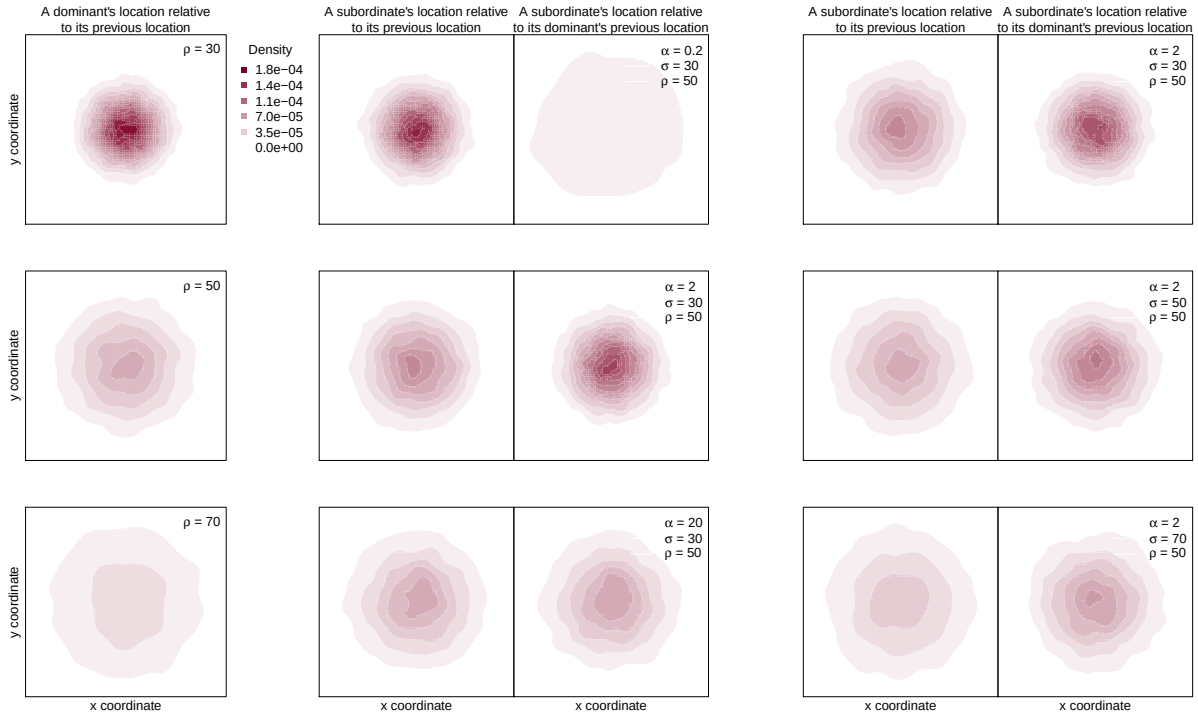
Statistical models can often be abstract and difficult to interpret. Therefore, in this section we will discuss the intuition and interpretation of the above framework. We will do so in the context of just two interacting animals: a dominant and its subordinate.

To begin with, we will discuss the restrictions we place on the parameters. Firstly, we restrict α to be positive as discussed in Chapter 2. Therefore, as $t \longrightarrow \infty$, the expected value of a subordinate's

movement at time t is the location of its dominant at time 0. Both σ and ρ are also positive as they are coefficients of noise.

To help explain the movement and interaction that this model produces, and the role of each of the parameters, we simulated the movement of a dominant and its subordinate for 10,000 steps (each step being a unit of time) and we repeated that simulation for a number of different parameter values. In order to capture the scenario that interaction may be initiated when the two animals are some distance apart, we began the simulations with the subordinate 100 units of distance away from the dominant in both axes.

The dominant only has one associated movement parameter: ρ. Figure 3.2a displays the effect the value of ρ has on the movement of the dominant. The densities relate to the difference between its location at time i and at time $i-1$ for $i = 2, ..., 10000$, with a difference of $(0, 0)$ (the centre of each plot) indicating that it didn't move. Simply, the greater the value of ρ, the greater the variability in the movement. Thus, ρ can be thought of as a representation of the speed the animal is travelling at: larger values of ρ increase the probability of larger movement steps.



(a) The effect of ρ (b) The effect of α (c) The effect of σ

Figure 3.2: (a) displays the effect of different ρ values on the movement step of the dominant animal relative to its own previous location. (b) displays the effect different α values have on the movement step of the subordinate relative to its own previous location (left column) and that of its dominant (right column). Similarly, (c) displays the effect different σ values have on the movement step of the subordinate relative to its own previous location (left column) and that of its dominant (right column). Note, the axes are consistent across all plots but the labels are omitted for ease of presentation and the specific values are not important here.

Subordinate animals have two associated movement parameters: α and σ. The expected value of the subordinate's movement at time t is $S_t = e^{-\alpha t}S_0 + (1 - e^{-\alpha t})D_0$ where S_0 and D_0 are the subordinate's and dominant's previous location respectively at time 0. α thus represents the strength of the influence of its dominant: the greater the value of α, the more the expected value of the subordinate's movement weighs towards D_0 as opposed to S_0. This effect is shown in Figure 3.2b, where the left column displays densities relating to the difference between the subordinate's location at time i and at time $i-1$, whilst the right column displays densities relating to the difference between the subordinate's location at time i and the dominant's location at time $i-1$ for $i = 2, ..., 10000$. As we can see when $\alpha = 0.2$, a relatively weak attraction in the context of a unit time interval, the subordinate's movement is still heavily concentrated around its own previous location. Though, it is clearly influenced by the dominant to some degree. When $\alpha = 2$, the movement becomes more concentrated towards the dominant's previous location — the subordinate is moving with a greater degree of persistence towards it.

However, when $\alpha = 20$, the densities in both columns look similar. The model loses some interpretability for larger values of α. In fact, as α tends to ∞, the marginal distribution of the dominant's and subordinate's movement converge:

$$\mathrm{E}[D_t|D_0, S_0] = D_0 \tag{3.8}$$

$$\lim_{\alpha\to\infty} \mathrm{E}[S_t|D_0, S_0] = \lim_{\alpha\to\infty} e^{-\alpha t}(S_0 - D_0) + D_0 \longrightarrow D_0 \tag{3.9}$$

$$\mathrm{Var}[D_t|D_0, S_0] = \rho^2 t \tag{3.10}$$

$$\lim_{\alpha\to\infty} \mathrm{Cov}[S_t, D_t|D_0, S_0] = \lim_{\alpha\to\infty} \rho^2 t - \frac{\rho^2(1 - e^{-\alpha t})}{\alpha} \longrightarrow \rho^2 t \tag{3.11}$$

$$\begin{aligned}\lim_{\alpha\to\infty} \mathrm{Var}[S_t|D_0, S_0] &= \lim_{\alpha\to\infty} \rho^2 t - \frac{\rho^2(1 - e^{-\alpha t})}{\alpha} - \frac{\rho^2(1 - e^{-\alpha t})^2}{2\alpha} + \frac{\sigma^2(1 - e^{-2\alpha t})}{2\alpha} \\ &\longrightarrow \rho^2 t\end{aligned} \tag{3.12}$$

where equation 3.11 is term (b) from equation 3.7. In this context of a subgroup of two animals, $e^{\boldsymbol{F}t}_{i,j} = 1 - e^{-\alpha t}$ when i is the subordinate and j is the dominant and $\boldsymbol{\Xi}_{\mathrm{dom}(i),j} = \boldsymbol{\Xi}_{\mathrm{j,\,j}} = \rho^2 t$. Equation 3.12 is term (d) from equation 3.7 where $\boldsymbol{\Xi}_{i,\mathrm{dom}(i)} = \boldsymbol{\Xi}_{i,j} = \mathrm{Cov}(S_t, D_t|D_0, S_0)$ and $e^{\boldsymbol{F}t}_{i,-l} = e^{\boldsymbol{F}t}_{i,i} = e^{-\alpha t}$. The derivations of the $e^{\boldsymbol{F}t}$ terms can be found in Appendix A. In this scenario, whilst the subordinate and dominant may be some distance apart at time 0, their movement processes are identical, meaning the densities of the subordinate's location relative to its own and its dominant's previous location will be one and the same. Similarly, the densities of the dominant's location relative to its own and its subordinate's previous location will be one and the same. This scenario may occur when the resolution of the data is too coarse to be able to capture enough information about this interaction and, as such, we cannot ascertain the ordering. It may also occur when the interaction doesn't have an ordering at all. This co-movement is a scenario which our modelling framework is not able to represent correctly. A final possibility is that the animals are not interacting — they are simply moving in a similar pattern due to some other factor such

as an environmental cue.

The effect of σ on a subordinate's movement is merely similar to ρ's effect on the dominant's — increasing the variance. The strength of the interaction is unaltered. α and σ both contribute to the speed of the subordinate and so, if $\sigma = \rho$, the subordinate can be thought of as travelling at a greater speed than the dominant, a feature which we discuss later in Section 3.5.

3.4 Behavioural States

The movement model we have described so far relies on some knowledge of which animal is subordinate to which and when. However, we are unlikely to know either of those pieces of information and so we treat them as unknown. To estimate them, we incorporate behavioural state switching where the states correspond to the animal's social behaviour. As we are operating in continuous time, we do so with a continuous-time Markov chain with a discrete state space.

The state space we want to explore is the space of all possible hierarchies as defined in Section 3.1. Ideally, the transition rates of our Markov chain would correspond to switching between those hierarchies. However, even for a group as small as four animals there are in excess of 100 of these structures, meaning that approach is not practical. The resulting transition matrix would be unwieldy and difficult to define and interpret.

We can, however, explore the same state space by defining our behavioural states to represent each individual animal's behaviour, as opposed to the group's. That is, an individual can be in a state that corresponds to being attracted to (i.e. subordinate to) a particular animal or be in a Brownian motion (i.e. leading or independent) state. The following generator matrix is then used from the point of view of an individual animal:

$$\Lambda = \begin{array}{c} \\ S_{A1} \\ \vdots \\ S_{An} \\ BM \end{array} \begin{array}{c} \begin{array}{cccc} S_{A1} & \dots & S_{An} & BM \end{array} \\ \begin{pmatrix} \lambda_{1,1} & \dots & \dots & \lambda_{1,n+1} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \lambda_{n+1,1} & \dots & \dots & \lambda_{n+1,n+1} \end{pmatrix} \end{array} \tag{3.13}$$

where n is the number of animals in the data and state S_{Ai} represents being subordinate to animal i. Here we use the parameterisation of a continuous-time Markov chain in which an animal stays in state u for a holding time, which is exponentially distributed with rate λ_u where λ_u is the rate of leaving state u ($\lambda_u = \sum_{u \neq v} \lambda_{uv}$). After this holding time, the animal switches to state v with probability λ_{uv}/λ_u. The diagonal elements of λ are therefore determined by the off-diagonal ones: $\lambda_{uu} = -\lambda_u$.

There are two types of heterogeneity to consider with this approach. Firstly, heterogeneity between

animals — where each animal would have their own set of transition rates. Secondly, heterogeneity within Λ. That is, the transition rates to different subordinate states are distinct. Both heterogeneity approaches will provide more granular insights into the animals' behaviours. However, both will also increase the complexity of the model. For the analysis in this chapter, we compromise through defining homogeneity between animals, but heterogeneity within Λ. With homogeneity between animals, the transition rate λ_{uv} is the rate of switching from state u to v averaged over all animals in the group. With heterogeneity within Λ, the rate of switching to S_{Ai} is distinct from the rate of switching to state S_{Aj} (where $i \neq j$) and thus the rates will be a more accurate representation of the influence of each animal.

With this approach, in the case of $n = 4$ as in the above example, the number of transition parameters is a more manageable 20 as well as being more intuitive than a transition matrix relating to group-level behaviours.

3.5 Model Formulation

As well as the above considerations for Λ, there are several other variations or extensions that can be made to the model described thus far to customise it for the analysis at hand. We will discuss them here.

3.5.1 Simultaneous or Sequential Switching

As we are dealing with multiple animals, the question arises as to how many animals can switch state at any one time. Should a change in the social behaviour of a group be restricted to the transition of a single animal, or should multiple animals be able to switch state simultaneously? It is perhaps more natural to think of a group of animals undergoing a structural transformation to be switching state sequentially. The transition of an animal may well be the result of the previous behavioural changes of its peers. However, this approach requires a greater number of switching times (compared to simultaneous switching to represent equivalent hierarchical changes) in the behavioural process that we are reconstructing. As such, sequential switching carries a heavier computational burden. The downside to simultaneous switching is that we may lose some information as to how a change in structure occurred (e.g. ascertaining which animal initiated the change in behaviour) — though it can still accommodate sequential switching in principal. The model which we use in this chapter allows simultaneous switching.

3.5.2 Options for Leading or Independent Movement

The model described so far assumes leading or independent animals are restricted to a single BM state and movement parameter ρ. To capture richer movement for these animals, various extensions to the model can be made.

Firstly, additional BM states with distinct noise parameters can be included to represent different

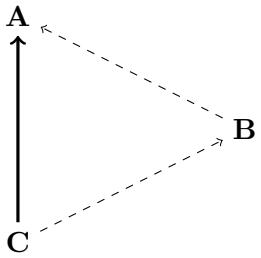


Figure 3.3: Our state estimation of C being subordinate to A implies a direct interaction between the two. However, we may not have complete data on a social group of animals and we may be missing links in the hierarchy. For example, B is missing from the data and in reality C is subordinate to B, who is in turn subordinate to A.

speeds of movement. In the results to be discussed, we use two BM states to loosely embody slow and fast movement. Secondly, we might assume that the non-stationary case is not sufficient and that leading or independent animals are themselves attracted to some location, a resource or nesting site for example. It would then be natural to model the movement of these animals with an OU process. In this case, the expressions that make up the conditional variance matrix $\boldsymbol{\Xi}$ (equation 3.7) would alter. These can be derived using the same method detailed by Niu *et al.* (2016), which we also used to derive equation 3.7 in Appendix A. We would need a more general form of e^{Ft} and Δ than those given in Niu *et al.* (2016) in order to model the movement of all possible hierarchies our social assumptions allow in Section 3.1. However, we assumed the movement of leading and independent animals contained some attraction during the e^{Ft} and Δ sections of our derivation of equation 3.7. The majority of the work required for this extension therefore already exists in Appendix A.

Both methods add movement parameter(s) and may add further row(s) and column(s) to the transition matrix, meaning some consideration is needed as to the costs and benefits of these extensions.

3.5.3 Direct and Indirect Following

Our definition of the social structure in Section 3.1, and the accompanying behavioural states in Section 3.4, assume that all the relevant animals have been tagged. Say we infer animal C as being subordinate to animal A, we are therefore implying a direct link between the two. In reality though, we may not have complete data on a social group of animals and we may be missing links in the hierarchy. See Figure 3.3 for an example.

This situation leads to a couple of issues. Firstly, we must be aware that our state labels and resulting hierarchy do not represent the de-facto social structure as there may be other interacting animals in the group that we don't have data on. Secondly, direct and indirect subordination would no doubt result in different α and σ values. It may therefore be prudent to reformat the state space to allow for 'strong' following (direct subordination) and 'weak' following (indirect subordination).

Our generator matrix would then take the following form:

$$\Lambda = \begin{array}{c} \\ IS_{A1} \\ DS_{A1} \\ \vdots \\ IS_{An} \\ DS_{An} \\ BM \end{array} \begin{array}{c} \begin{array}{cccccc} IS_{A1} & DS_{A1} & \dots & IS_{An} & DS_{An} & BM \end{array} \\ \begin{pmatrix} \lambda_{1,1} & \dots & \dots & \dots & \dots & \lambda_{1,2n+1} \\ & \ddots & & & & \vdots \\ \vdots & & \ddots & & & \vdots \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ \lambda_{2n+1,1} & \dots & \dots & \dots & \dots & \lambda_{2n+1,2n+1} \end{pmatrix} \end{array} \tag{3.14}$$

where n is the number of animals, IS_{Ai} is indirect subordination to animal i and DS_{Ai} is direct subordination to animal i.

Whilst we haven't performed a thorough analysis with this extension, our exploratory trials with simulated data suggest this approach may not provide any more detailed information than the Markov chain defined in Section 3.4. That is, both the indirect and direct subordination states often converged to similar behaviours. Furthermore, this extension greatly increases the complexity of an already complex state space.

An alternative method to gain some flexibility in the strength of an interaction could be through a statistically-hierarchical α. Aside from the above issue of incomplete data, Wittemyer *et al.* (2005) highlight another scenario when this approach might be fruitful as they show that different 'tiers' of a social structure of Africa elephants contain different levels of cohesion. For example, when family-based groups fuse together to form kinship groups, there is a greater level of cohesion within families than between them.

3.6 Inference

Markov chain Monte Carlo (MCMC) methods are used to infer both the behavioural and movement parameters. Each iteration of the MCMC algorithm consists of two parts. Firstly, we sample the behaviours of the animals in continuous time. This is done through simulating state switches between the observed data points and the acceptance of these behavioural trajectories is evaluated by a Metropolis-Hastings (MH) ratio. The exact algorithm for this step is detailed in Section 3.6.1. Secondly, the parameters of the Ornstein-Uhlenbeck and Brownian motion processes are sampled in accordance with the latest behavioural process reconstruction by means of a MH random walk. Further details of this, including restrictions we place on the parameters, are in Section 3.6.3. Moreover, due to the complexity of the state space we are exploring and the accompanying computational burden, we explore some preliminary analysis methods in section 3.6.4 with which we aim to aid this process.

3.6.1 Behavioural Parameters

To estimate the behavioural parameters, both the states and Λ, we reconstruct the animals' behavioural trajectories between the observed data points. Through this, we treat the behavioural process in continuous time and account for our uncertainty in what the animal is doing at unobserved times, which enables us to undertake exact inference in a Bayesian setting. For the analysis in this chapter, we allow multiple animals to switch state simultaneously at any one time. Therefore, our method to simulate these trajectories is a straightforward multi-animal extension of the 'kappa' method introduced by Blackwell *et al.* (2016) (and discussed in Chapter 2), where the observed data is augmented with sampled state switching times. Allowing multiple animals to switch state simultaneously will help keep the size of the augmented data set (and therefore the computational cost of fitting the model) relatively small compared to sequential switching — see Section 3.5.1. In order to obtain a good acceptance rate for these trajectories, we only simulate over a small segment of data in each proposal. The model and analysis in this chapter only concerns the spatially homogeneous case.

Say we are proposing a new behavioural trajectory for the interval $[t_a, t_b]$ where t_a and t_b are observed times. Let τ_o represent the observed data times for that interval; τ_s represent the switching times of the previously accepted trajectory through $[t_a, t_b]$; $\tau = \tau_o \cup \tau_s$ so that $\tau = \{\tau_1 < ... < \tau_p\}$ where p is the size of τ; β represent the behavioural states of all animals at times τ where β_j is a vector containing the states of all animals at time τ_j for $j = 1, ..., p$.

To propose a new trajectory in this interval, we discard τ_s, and the associated states in β, and sample new potential state switching times, $\hat{\tau}_s$. The new switching times are produced from a homogeneous Poisson process over the interval (t_a, t_b) with rate λ_{max}, where $\lambda_{max} \geq \max(\lambda_u)$ for all states u. As we have set $\lambda_{max} \geq \max(\lambda_u)$, the expected number of potential switching times sampled will be greater than the expected number when in the most volatile state. This guarantees we have enough switching times in order to reconstruct the behavioural process appropriately. Note, we are using the notation λ_{max} instead of κ and, throughout this thesis, we set λ_{max} as a fixed value before running any inference algorithm.

Let $\hat{\tau} = \tau_o \cup \hat{\tau}_s$. Having sampled the new potential switching times $\hat{\tau}_s$, we now simulate a new behavioural trajectory ($\hat{\beta}$) forwards through $\hat{\tau} = \{\hat{\tau}_1 < ... < \hat{\tau}_{\hat{p}}\}$ where $\hat{p}$ is the size of $\hat{\tau}$. This new trajectory is initialised with the current estimate of the behavioural states for all animals at time t_a. That is, $\hat{\beta}_1 = \beta_1$. To simulate the trajectory, we loop through each time in $\hat{\tau}$, where there are two scenarios to account for:

- If $\hat{\tau}_j \in \tau_o$ for $j = 2, ..., \hat{p}$, the behavioural states of all animals at $\hat{\tau}_j$ are carried forward from $\hat{\beta}_{j-1}$ as observations are not part of the behavioural process.
- If $\hat{\tau}_j \in \hat{\tau}_s$ for $j = 2, ..., \hat{p} - 1$, we sample a state switch for each animal in turn (randomly ordered). The probability of $\hat{\tau}_j$ being a switch for animal i is $\lambda_u / \lambda_{max}$ when i is in state u. If

it is, the new state is v with probability λ_{uv}/λ_u. Otherwise, the state of i at $\hat{\tau}_j$ remains the same as it was in $\hat{\beta}_{j-1}$.

We initialise $\hat{\beta}_1 = \beta_1$ in order to keep our new sample consistent with the data before t_a. Therefore, we also require $\hat{\beta}_{\hat{p}} = \beta_p$ (except for when $\hat{\tau}_p$ is the final observation in the data set). If this condition isn't met at $\hat{\tau}_{\hat{p}}$, we reject the simulation and return to our previously sampled trajectory. Furthermore, we also reject a trajectory if at any point a state is sampled which produces a cyclic hierarchy.

If these conditions are met, we accept or reject the new trajectory proposal with a Metropolis-Hastings ratio. As the behavioural trajectory is proposed from the model itself, the distribution of that trajectory cancels out with the proposal distribution and so the MH ratio simplifies to a ratio of likelihoods of the observed movement through the newly proposed state switches against our previous reconstruction:

$$\prod_{k=2}^{|\tau_o|} \frac{f(\boldsymbol{g}_{\tau_{o_k}} | \boldsymbol{g}_{\tau_{o_{k-1}}}, \hat{\beta}_{[\tau_{o_{k-1}}, \tau_{o_k}]}, \hat{\tau}_{[\tau_{o_{k-1}}, \tau_{o_k}]})}{f(\boldsymbol{g}_{\tau_{o_k}} | \boldsymbol{g}_{\tau_{o_{k-1}}}, \beta_{[\tau_{o_{k-1}}, \tau_{o_k}]}, \tau_{[\tau_{o_{k-1}}, \tau_{o_k}]})} \tag{3.15}$$

where $\boldsymbol{g}_{\tau_{o_k}}$ are the locations of the animals in a particular axis at time τ_{o_k}; $\hat{\beta}_{[\tau_{o_{k-1}}, \tau_{o_k}]}$ are the newly sampled states throughout the interval $[\tau_{o_{k-1}}, \tau_{o_k}]$ at times $\hat{\tau}_{[\tau_{o_{k-1}}, \tau_{o_k}]}$; $\beta_{[\tau_{o_{k-1}}, \tau_{o_k}]}$ are the previous state estimations in the same interval at times $\tau_{[\tau_{o_{k-1}}, \tau_{o_k}]}$. Blackwell (2003) details how these movement likelihood terms are calculated and more detail can be found in Appendix C. The above likelihood terms are also conditional on the movement parameters.

Simulations over longer segments of the data have the potential to provide a more holistic view of how the animals are behaving. That is, estimate how an animal traverses through behavioural changes, rather than obtaining a piecemeal view of the transitions through several smaller simulations. However, longer simulations are inherently more likely to deviate away from a reasonable proposal and fall foul of the above conditions — leading to a poor acceptance rate. Therefore, a mix of shorter and longer updates will facilitate good mixing of the behaviours.

We use the conjugate Dirichlet prior for the multinomial likelihood of the transition rates to obtain their full conditional distributions and resample them using Gibbs sampling at each iteration of the MCMC algorithm.

Individual Animal Updates

The above method of simulating collective trajectories, aside from the states at $\hat{\beta}_1$, doesn't use any information as to what behaviours have previously been deemed acceptable — each simulation is a completely independent attempt at proposing a likely trajectory. Furthermore, there is no consideration for the state in which the trajectory must end in. This naivety is compounded when simulating the trajectories for multiple animals simultaneously and it leads to a low acceptance rate

— even when only proposing trajectories over very small segments of data. Both the low acceptance rate and the tunnel vision of short simulations are not conducive to well mixing parameters.

To work around this, the trajectory updates happen in a two-step process. First, the trajectory of the entire group is reconstructed as detailed above (hereafter referenced as 'collective updates'). Secondly, we fine tune those trajectories an individual at a time ('individual updates'). That is, we resample the state switches of one animal through the currently accepted switching times whilst keeping the rest of the trajectory fixed. In effect, the role of the collective updates is to resample the number of switching points and the corresponding times, whilst the individual updates can further explore the state space at those times. The acceptance rate of individual updates is much higher than that of collective ones and the interplay between the two encourages better mixing of the behavioural states. During this individual update, we relax the initialisation of $\hat{\beta}_1 = \beta_1$ when this concerns the first observation in the data set and we sample the initial state uniformly.

Note, we don't propose a collective trajectory and then immediately fine tune it. Collective and individual updates are carried out separately and for each one (both collective and individual) we begin by randomly (uniformly) sampling the segment of data that will be updated.

3.6.2 Partial Observations

Partial observations, where we only have data on some of the animals we are tracking, are a potential obstacle when analysing data from multiple animals: tracking equipment may not be fully synchronised or a GPS tag may not have been able to transmit some data for example. However, the above movement process and simulation algorithm are naturally adaptable to take into account the uncertainty of any missing or asynchronous data using standard results for conditional multivariate normal distributions. During a trajectory proposal, say $\hat{\tau}_j \in \tau_o$ is partially observed. Using slightly simplified notation, the distribution of the movement at $\hat{\tau}_j$ is:

$$\begin{pmatrix} \boldsymbol{P}_{\hat{\tau}_j} \\ \boldsymbol{M}_{\hat{\tau}_j} \end{pmatrix} | \begin{pmatrix} \boldsymbol{P}_{\hat{\tau}_{j-1}} \\ \boldsymbol{M}_{\hat{\tau}_{j-1}} \end{pmatrix} \sim \text{MVN}(\begin{pmatrix} \boldsymbol{\mu_P} \\ \boldsymbol{\mu_M} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Xi}_{PP} & \boldsymbol{\Xi}_{PM} \\ \boldsymbol{\Xi}_{MP} & \boldsymbol{\Xi}_{MM} \end{pmatrix}) \tag{3.16}$$

where $\boldsymbol{P}_{\hat{\tau}_j}$ and $\boldsymbol{M}_{\hat{\tau}_j}$ are the locations of the observed and unobserved animals respectively at time $\hat{\tau}_j$; $\boldsymbol{\mu_P}$ and $\boldsymbol{\mu_M}$ are the expected values of the locations of the observed and unobserved animals respectively. $\boldsymbol{\Xi}_{PP}$ represents the covariance between the observed animals and similarly for the other $\boldsymbol{\Xi}$ terms. In this context of partial observations, 'unobserved' animals are animals which are tracked and we would normally expect to have data on, not animals which are not being tracked at all. The likelihood of the observed data is calculated from:

$$\boldsymbol{P}_{\hat{\tau}_j} \sim \text{MVN}(\boldsymbol{\mu_P}, \boldsymbol{\Xi}_{PP}) \tag{3.17}$$

and the locations for the unobserved animals are sampled from the following conditional distribution:

$$\boldsymbol{M}_{\hat{\tau}_j}|\boldsymbol{P}_{\hat{\tau}_j} = \boldsymbol{p} \sim \text{MVN}(\boldsymbol{\mu_M} + \boldsymbol{\Xi}_{MP}\boldsymbol{\Xi}_{PP}^{-1}(\boldsymbol{p} - \boldsymbol{\mu_P}), \boldsymbol{\Xi}_{MM} - \boldsymbol{\Xi}_{MP}\boldsymbol{\Xi}_{PP}^{-1}\boldsymbol{\Xi}_{PM}) \tag{3.18}$$

The sampled locations are then taken as fixed for calculating the likelihood of the movement to the next observation. This method is only applicable for observations in the interior of τ_o in order to be consistent with the data outside of τ_o.

3.6.3 Movement Parameters

Treating the current behavioural process reconstruction as fixed, we update the parameters of all Ornstein-Uhlenbeck and Brownian motion processes simultaneously through a Metropolis-Hastings random walk. We use independent, normally-distributed proposals for each movement parameter. Whilst we place certain bounds on the range of each one, the prior distribution within those bounds is flat, with samples outside of that range automatically rejected. As such, the Metropolis-Hastings ratio (for appropriate values) is again reduced to equation 3.15. In general, all parameters must be greater than 0 and we have added the restriction that $\sigma \leq \rho$, the justification being that an animal influenced by another should have less variability in their movement than an animal following a BM process. In the case of building multiple BM states into the model, this restriction is loosened to $\sigma \leq \rho_{max}$ where $\rho_{max} = \max(\rho_1, ..., \rho_m)$ for m BM states. This case also brings potential complications to the state labelling and so, in order to keep consistency, we define $\rho_1 < ... < \rho_m$.

3.6.4 Preliminary Pairwise Analysis

The state space of our model is large and complex, and it may take a substantial amount of time to fully explore it. However, aside from the case when all animals reside in a single subgroup, each animal will not exert some influence (directly or indirectly) on all of the others all of the time. If we knew which animals did not interact with each other and when, we could limit the area of the state space we need to explore and therefore mitigate some of the computational cost of our complex state space. This may be achievable by running a preliminary pairwise analysis and then feeding those results into a group-level analysis in order to guide it. This concept of building up the group-level social structure from dyadic analysis is found in other areas of social behaviour research such as social network analysis (SNA) (Hobson *et al.*, 2013).

One approach to this is to fit our model to each pair of animals in the data set, the result of which will establish which animals exert some influence on which others and when. As there are only three ways to arrange two animals in a hierarchy, it is a simpler and faster model to fit than when considering all animals in the data. Though the resulting computational effort is by no means trivial, this approach lends itself well to parallel computing as each pairwise analysis is independent of the others — although there may be some difficulty in finding tuning parameter values so that

each run mixes well. This method would also provide us with a reasonable initial value for both the behavioural and movement parameters, potentially reducing the burn-in period required of the group-level inference.

Dyadic metrics (Long *et al.*, 2014; Joo *et al.*, 2018) offer another route for this preliminary step, one that is likely to be substantially faster than the pairwise fitting of our model. Some consideration will be needed to determine which metric is suitable for the particular analysis at hand, and potentially some combination of metrics might provide a more comprehensive view of interactive behaviour. In addition, some metrics contain user-inputted parameters and some aren't well suited to capturing the dynamics of interaction.

There are a number of other routes we could take to undertake this preliminary step. However, all of the methods with which we can constrain the social space pose the risk of constraining it too much and prohibiting the exploration of the state space in the group-level analysis. In theory, fitting the same model in the preliminary analysis and the group-level analysis is the least likely to produce conflicting outputs as the modelling framework is common to both steps. Thus, in this chapter, I will only discuss and experiment with fitting our model to each pair of animals during a preliminary analysis.

There are multiple approaches one could take to then use this pairwise step in the group-level inference. The method we have used in our analyses is to add extra conditions into the trajectory updates when sampling new state switches. That is, when sampling a state switch, we can condition on whether or not our newly sampled state is a behaviour we have 'seen' in the pairwise analyses. If not, we automatically reject the switch and trajectory proposal. For example, during a trajectory proposal in the group-level inference at time t, we sample animal 4 to switch to state S_{A1} (subordinate to animal 1). This switch is deemed acceptable if during the preliminary analysis between animals 1 and 4 (which I will denote as pair $\{1, 4\}$) the posterior probability of animal 4 being subordinate to animal 1 at time t is high enough. If so, the rest of the proposal is continued. Otherwise, the state switch (and subsequently the whole trajectory proposal) is rejected. Of course, there is some ambiguity as to what 'high enough' constitutes. In our analysis below, we will examine different boundaries for this.

Furthermore, we only condition on the sampled times that contain an actual state switch. That is, if a sampled time during an update, say t, doesn't result in a state switch, the animal remains in the state it is currently in. We don't condition on whether or not this continued state was 'seen' at time t for this animal during the pairwise analyses. The pairwise analysis is only intended to act as a guide and adding in those further conditions may restrict the exploration of the state space to too greater an extent.

An alternative approach would be to sample state switches in the group-level inference from a proposal distribution that is derived from the pairwise analysis. For instance, say at time t in the pair $\{1, 4\}$ analysis, the posterior probability of animal 4 being in state S_{A1} is 0.3. If t is a potential

switching time during the group-level analysis, animal 4 will be proposed to switch to state S_{A1} with a probability of 0.3 (though, this would need scaling so that all probabilities summed to 1). As our state switches would not be sampled from Λ, we would not be able to cancel the likelihood of switches with the proposal terms in the MH ratio as in Section 3.6.1. This guided proposal distribution approach would potentially be more efficient in proposing acceptable trajectories than the conditioning method. However, there is likely to be little difference between the methods as the wasted computational effort from rejecting updates in the conditioning approach is small and will be somewhat cancelled out by the added complexity of the MH ratio in the guided proposal method.

The approach of deriving the proposal distributions of the state switches from the pairwise analysis could also be applied, in theory, to the locations of the animals when modelling in a spatially heterogeneous setting.

3.7 Simulated Data Analysis

We assessed our influence hierarchy model and inference algorithm as described above through fitting the model to simulated data. This data consists of five dynamically interacting 'animals' over 100 observations, with two units of time between each one. We randomly deleted 10% of the data (uniformly across all data) to provide us with the setting of having partial observations. The 'observed' times were augmented with switching times with a Poisson process as detailed in section 3.6.1; the movement of the group was simulated forwards through the augmented data set using a multivariate OU process as in section 3.2; and state switches were sampled using a continuous-time Markov chain as in section 3.4. With regards to the model formulation, we included two BM states for leading or independent animals to capture 'slow' and 'fast' movement and, as previously discussed, we are allowing multiple animals to switch state at any switching time. Whilst we simulated the movement and behaviours of the animals in continuous time, the model is only fitted to the discrete-time 'observed' data in order to mirror typical real data. The parameters used to simulate this data set are provided in Table 3.1.

Parameter	Value	Parameter	Value
α	0.5	λ_{OU-OU}	0.00 - 0.04
σ	0.7	λ_{OU-BM}	0.00 - 0.01
ρ_{slow}	0.4	λ_{BM-OU}	0.00 - 0.01
ρ_{fast}	1.8	λ_{BM-BM}	0.00 - 0.12

Table 3.1: Parameter values used to simulate data. The table on the left details the movement parameters whilst the table on the right details the transition rates. λ_{OU-OU} represents switching between Ornstein-Uhlenbeck (that is, subordinate) states, λ_{OU-BM} represents switching from a subordinate state to a Brownian motion state (that is, leading or independent) and similarly for λ_{BM-OU} and λ_{BM-BM}. The values of the transition rates indicate the range that the rates in each of those categories were uniformly sampled from.

We will also examine the practicality of the preliminary pairwise analysis. That is, does including this step improve the speed at which we can generate the posterior distributions for the model parameters? Or perhaps, will the inclusion of the extra conditions hamper the mixing of the MCMC process? We will compare model fittings on the same data set; one which utilises the pairwise step ('informed') and one which foregoes it ('uninformed'). Furthermore, the level of conditioning we introduce with the pairwise analysis will be investigated.

Both the simulation and inference methods (in this and subsequent chapters) were fully implemented in R (R Core Team, 2017). Aside from the *MCMCpack* package (Martin *et al.*, 2011), which is used for the Dirichlet distribution, and making some use of some of the ideas in the code published alongside Blackwell *et al.* (2016) and Niu *et al.* (2016), the code we have developed is original. For reference, when running times are discussed, all inference runs (in this and subsequent chapters) were performed on the University of Sheffield's HPC 'ShARC' where each core runs at 2.4GHz with 4.0GB of RAM (a single core being used for each run).

3.7.1 Inference vs Simulation Model

As described above, the inference algorithm can propose that an animal, say i at time t, switches to a state, say v, that results in a cyclic hierarchy (including subordination to itself). If that occurs, the whole trajectory is rejected ('method 1'). In effect, the rate of switching to these cycle-inducing states is set to 0 after establishing i will switch state and so the rates of switching to other states are inflated. That is, we still expect t to be a switching time for animal i with probability λ_u/λ_{max} when i is in state u, but it cannot switch to v or other cycle-inducing states.

An alternative approach ('method 2') would be to set the transition rates from u to cycle-inducing states to be 0 before sampling whether or not t is a switching time for i. Thus, i will simply have an increased probability of remaining in state u. Either approach is valid, but they each have different pros and cons.

With method 1, in a spatially homogeneous setting (which we are currently working in), each state switch is sampled independently from the exact same single-trial multinomial distribution. The likelihood of the transition rates is therefore an s-trial multinomial, where s is the total number of potential switching times, concerning the number of successful switches to each state. As mentioned above and in Chapter 2, we can then use the conjugate Dirichlet prior to obtain the full conditional distribution of the rates and resample them using Gibbs sampling. This offers some efficiency both from a computational (sampling from the full conditional distribution vs calculating all of the separate likelihood terms) and proposing reasonable rate samples (sampling from the full conditional distribution vs random-walk proposals for example) point of view. This efficiency opportunity isn't afforded to method 2 as each state switch is sampled from a single-trial multinomial distribution that depends on the social context at the time. However, method 2 is perhaps more biologically sound — that is, we aren't encouraging an animal to switch state until an acceptable state is

found.

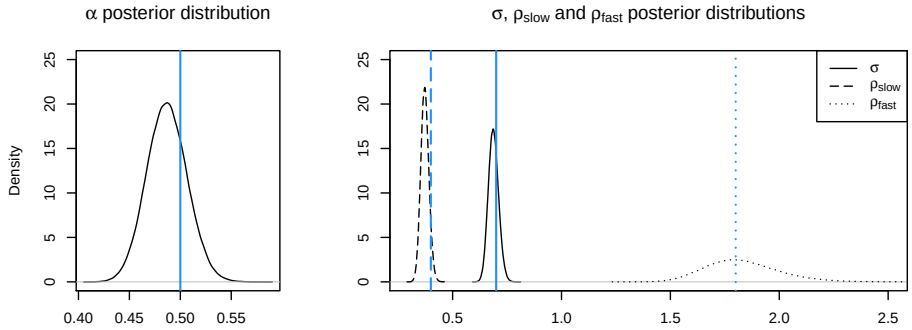
In this chapter (and in Chapter 4), as we are working in the spatially homogeneous case, we developed the inference algorithm using method 1 in order gain the aforementioned efficiencies. Furthermore, we are using a heterogeneous transition matrix in these chapters and so the benefit of being able to sample the rates from their full conditional distribution, as opposed to a random-walk MH approach, is substantial. Furthermore, the cost of rejecting trajectories with cyclic hierarchies is small as trajectory proposals only consist of computationally-simple state switches. However, method 1 is not easily translatable to simulated data. Simulating the state switches of five (or indeed more) interacting animals over a decent length of time without sampling any cyclic hierarchy is not likely. Therefore, the simulated data in this and all subsequent chapters was derived using method 2. This discrepancy is obviously not ideal (for example, analysis of how well the transition rate posteriors compare to the true values is not straightforward) but the simulation analysis in Chapters 3 and 4 will still serve as a useful sensor check as to whether the model 'works' or not.

In Chapter 5, as we move into a spatially heterogeneous context, we lose the ability to calculate the full conditional distributions of the transition rates (as the rates depend on the locations of the animals) and we move to a homogeneous transition matrix. The efficiencies of method 1 are now negated and therefore the inference model in Chapter 5 is the same as the one that the simulated data is derived from — the more biologically sound method 2.

3.7.2 Uninformed Analysis

To initialise the MCMC process, we randomised the behavioural parameters and started with over-dispersed movement parameter values. We set λ_{max} as 0.4 as that was sufficiently high for the transition rates used for the simulation and we updated segments of the behavioural trajectories ranging from 2 to 3 observations long for collective updates and 3 to 15 observations for individual updates. We simulated 50 new collective trajectories and 60 new individual trajectories per iteration. We ran the inference algorithm for 1 million iterations, of which 30k was burn-in, and we recorded every second output for the movement parameters and transition rates and every 20^{th} output for the trajectories for file size practicality — an approach we take for the remainder of this chapter. An uninformed Dirichlet prior was used for the transition rates.

Figure 3.4 displays the movement parameter posteriors against the true values used to derive the simulation. All are consistent with the true value and all looked to converge quickly (Figure 3.5). To check convergence, the Gelman-Rubin diagnostic is used to assess the potential improvement from running more or longer chains. The multivariate potential scale reduction factor (MPSRF) is 1, calculated from two separate MCMC runs using *coda* (Plummer *et al.*, 2006); this indicates that each chain is exploring the same posterior distribution, after burn-in. Further results of the Gelman-Rubin diagnostic are in Table 3.2. Due to the reasons discussed in Section 3.7.1 and the large number of transition rates (42 in a heterogeneous transition matrix regarding 7 states)



Parameter	True Value	Point Estimate	Standard Deviation	Effective Sample Size
α	0.5	0.487	0.0198	4645
σ	0.7	0.688	0.0233	22405
ρ_{slow}	0.4	0.373	0.0185	22187
ρ_{fast}	1.8	1.830	0.1680	6841

Figure 3.4: Top: posterior distributions for the four movement parameters from the uninformed simulation analysis. The blue vertical line indicates the true value used. Bottom: a summary of the movement parameter results. The effective sample size is calculated using using *coda* (Plummer *et al.*, 2006).

Movement Parameters	Min	1^{st} Quartile	Median	Mean	3^{rd} Quartile	Max
Uninformed — Point Est.	1.000	1.000	1.000	1.001	1.001	1.002
Uninformed — Upper C.I.	1.000	1.000	1.000	1.002	1.003	1.009

Transition Rates	Min	1^{st} Quartile	Median	Mean	3^{rd} Quartile	Max
Uninformed — Point Est.	1.000	1.000	1.001	1.003	1.003	1.030
Uninformed — Upper C.I.	1.000	1.000	1.002	1.006	1.006	1.061

Table 3.2: Summaries of the point estimates and upper confidence limits of the univariate potential scale reduction factors for the uninformed simulation analyses. The limits correspond to a coverage of 95%.

compared to the number of data, it is not straight forward or even meaningful to compare the inferred rates with the true values used in the simulation. For instance, it is unlikely all transitions set with a rate greater than 0 will be properly represented in a single simulation. However, from Table 3.2, we are able to at least see that the rates are estimated consistently across different runs. The multivariate PSRF for the rates is 1.02.

Figure 3.6 displays the behavioural state posteriors for animal 2, along with the true states from the simulation. We have presented the state posteriors at the discrete times of the observations. This is purely for simplicity of presentation and discussion — the posteriors can be calculated in continuous time for further analysis. Note that since this output refers to animal 2, the probability of it being in state S_{A2} (which would represent 'following itself') is necessarily zero. The posterior

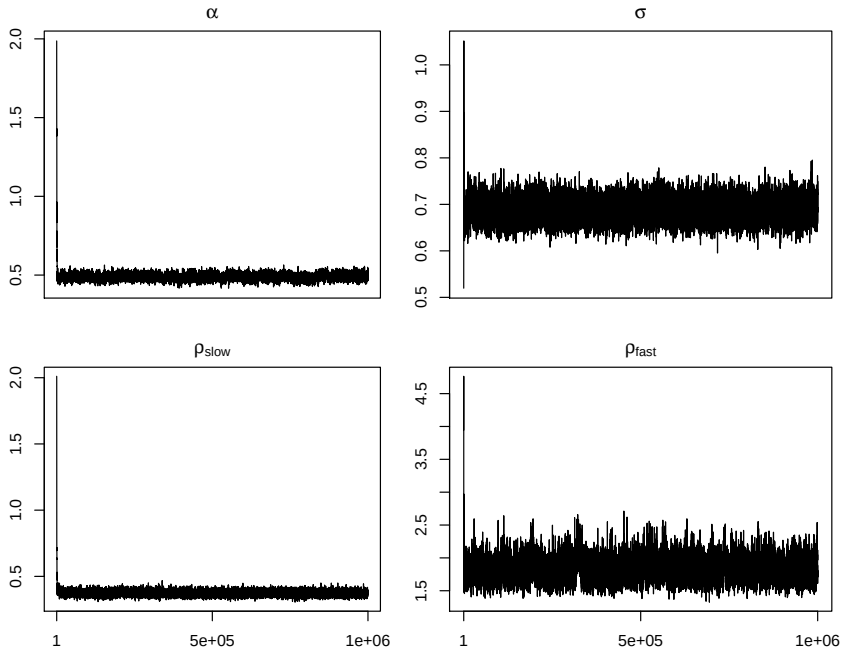


Figure 3.5: Trace plots for the four movement parameters from one of the uninformed simulation analyses. The chains have been thinned by a factor of 50.

estimates are broadly correct and confident. However, when the animals are in a close-knit group, there can be uncertainty as to the exact interaction. For example, around observations 36 to 41 in the simulation, animal 2 is subordinate to animal 5, who is in turn subordinate to animal 3. In this segment of data, the states S_{A3} and S_{A5} represent relatively similar behaviours from the perspective of animal 2 and thus we are uncertain as to which animal it is truly subordinate to.

Using Figure 3.6 alone, we cannot fully see in what capacity the animal interacts with its peers. In particular, whether it is a leading animal or not when it is in a BM state. We can reconfigure the state estimations for the whole group to estimate an animal's role as seen in Figure 3.7. An animal in a BM state is *leading* if they have a subordinate or *independent* if not. All animals in a subordinate state are *following*. In this instance, animal 2 largely interacts as a subordinate aside from the first 10 or so observations when it is leading others. We can now see that most of the time it is in a BM state, it is simply independent.

Thus, our model and inference approach can provide rich insight into the dynamics of the social behaviours that drive collective movement when our social assumptions (Section 3.1) are fair.

3.7.3 Informed Analysis

Preliminary Pairwise Analysis

In order to ascertain whether some preliminary pairwise analysis can aid the larger group-level analysis, we fit our model to each pair of animals in turn. As with the uninformed group analysis,

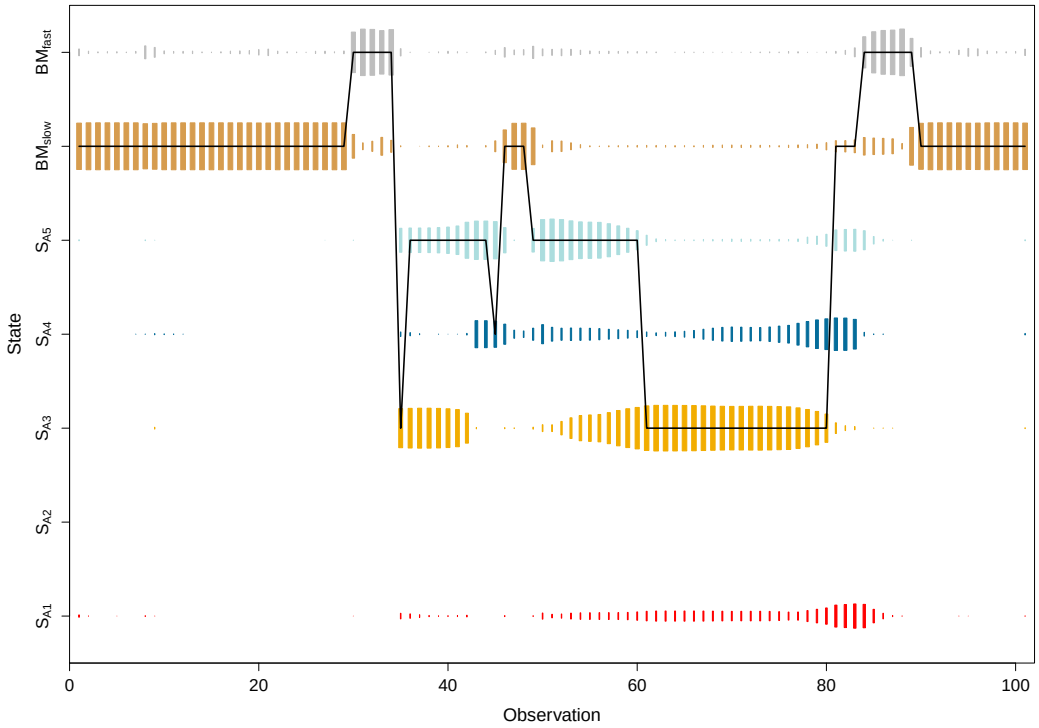


Figure 3.6: The state posterior distribution for animal 2 in the uninformed simulation analysis. There are seven states: the two BM speeds and five subordinate behaviours where state S_{Ai} indicates attraction to animal i. The area of each box represents the posterior probability of being in that state at that observation, from 0 to 1. The black line is the true state in the simulated data.

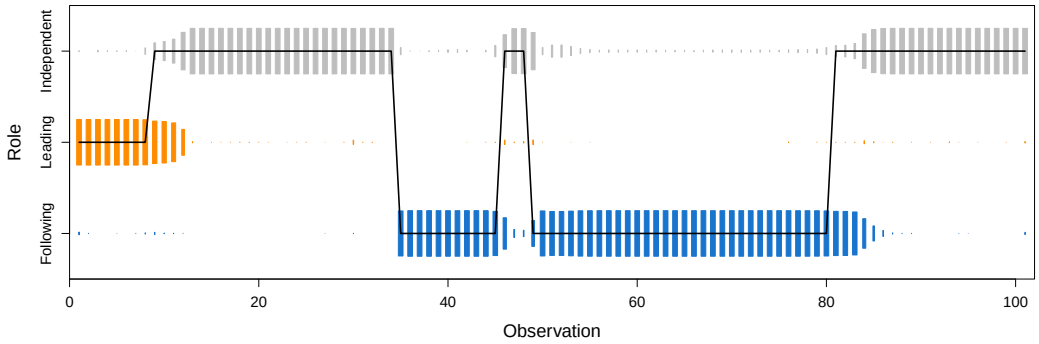


Figure 3.7: The role posterior distribution for animal 2 in the uninformed simulation analysis. The area of each box represents the posterior probability of being in that role at that observation, from 0 to 1. The black line is true role in the simulated data.

we initialise each of our pair analyses with a randomised behavioural process and over-dispersed movement parameter values. Due to the simpler setting of analysing just two animals, we can sample longer behavioural trajectories whilst still obtaining good acceptance rates. In this instance, collective (that is, pair) trajectories are updated over 2 to 7 observations whilst the individual updates are now carried out over 3 to 15 observations. During each iteration of the MCMC process, we undertake 17 collective and 24 individual updates. Again, the algorithm ran for 1 million iterations. We have chosen to use the longest burn-in period as a blanket rule for all 10 pairs, which was approximately 50k iterations.

Figure 3.8 displays the movement parameter posteriors for each of the pairwise analyses. Whilst the posterior distributions broadly align with the true values and are consistent with each other, there are some discrepancies. Largely, this will be caused by the missing information resulting from analysing just two animals out of a larger interactive group. That is, say the full data contains the social structure as shown in Figure 3.1a. Any interaction inferred in the pair {A, C} will (at best) relate to interaction that is indirect in the data. The estimations of α and σ will therefore not align with the true values as those parameters relate to direct interaction.

As mentioned earlier, we will use this pairwise analysis to add extra conditions whilst proposing new trajectories during the group-level analysis. That condition is: when sampling a state switch (in the group-level analysis) at time t, the new state must have a posterior probability of at least x in at least one of the pairwise analyses concerning that animal at time t. For example, say at time t in the group-level analysis we propose animal 2 to switch to state S_{A5} and we have set our acceptance boundary at 0.10. In the pair {2, 5} analysis, we require the posterior probability of state S_{A5} for animal 2 to be at least 0.10 at time t. If that's not the case, the state switch (and trajectory proposal) are rejected. Figure 3.9 showcases this example for all t in the data set.

The scale of the acceptance boundary is arbitrary. Though, we can see from Figure 3.10a that almost, if not, all of the true states will still be deemed acceptable if the boundary is drawn at 0.20 or less. Beyond 0.20 we would begin to classify too many true states as unacceptable. Furthermore, the main potential benefit of this conditioning is that the state space that requires exploring in the group-level analysis will be reduced. We can see from Figure 3.10b that the majority of this reduction occurs with boundaries up to 0.20. That is, during the pairwise analysis, almost 80% of the states that would estimate a false direct subordination have a posterior probability of less than 0.20 and they will be ignored during the group-level inference with an acceptance boundary of 0.20. The proportion of states we can ignore only increases slightly beyond 0.20. Therefore, there is little to be gained from the boundary being set beyond 0.20. We will trial boundaries of 0.01, 0.05, 0.10 and 0.20 to ascertain where a good compromise may be drawn.

The collating of the pairwise states took approximately 60 minutes for each animal — but this again lends itself well to parallelisation. The pairwise output was further thinned by a factor of 50 (so thinned by 1000 overall) to enable this step to be done in a semi-reasonable time frame.

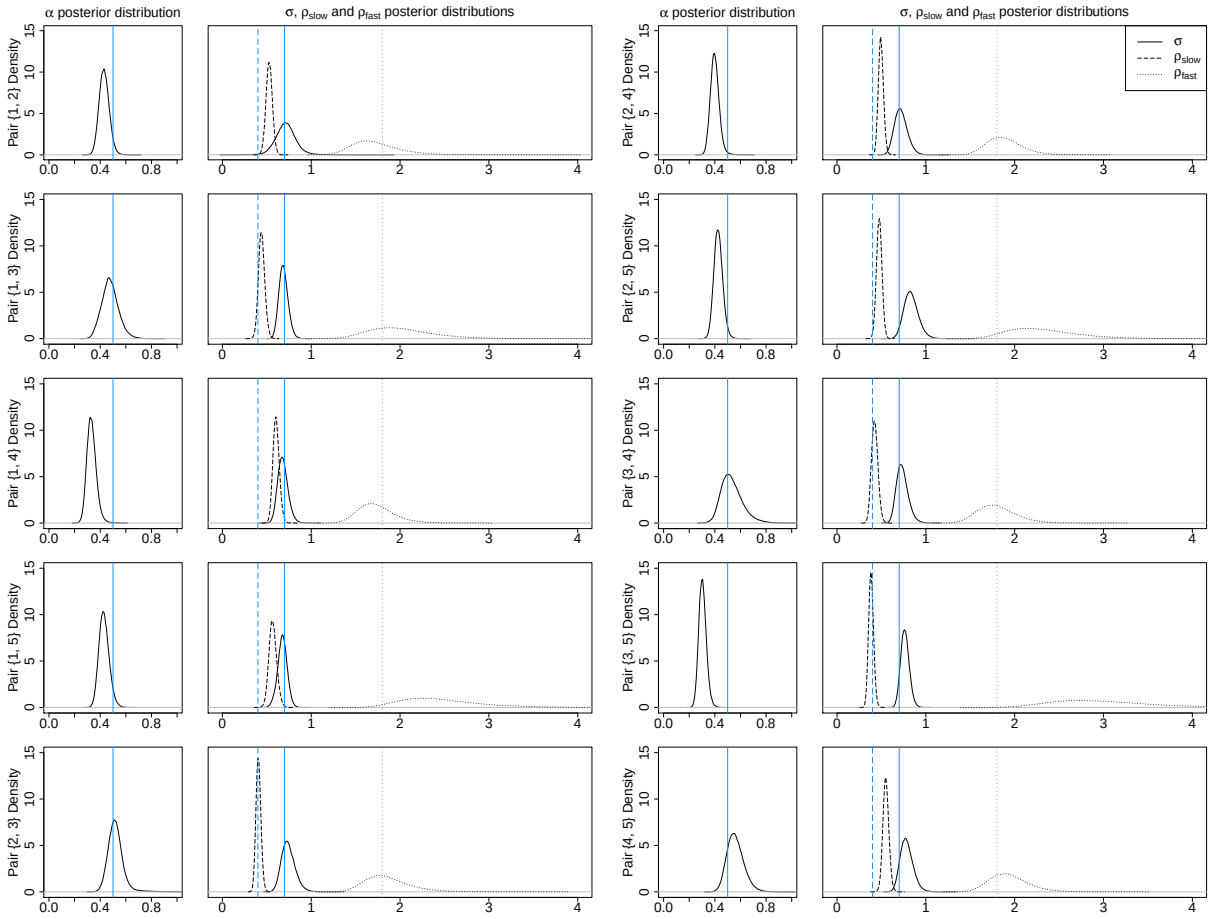


Figure 3.8: Posterior distributions of the four movement parameters for each of the pairwise analyses of the simulated data. The blue vertical lines indicates the true values used.

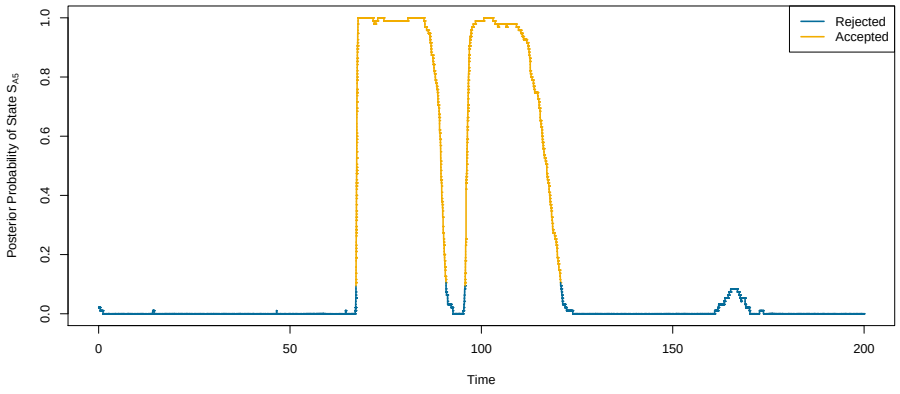
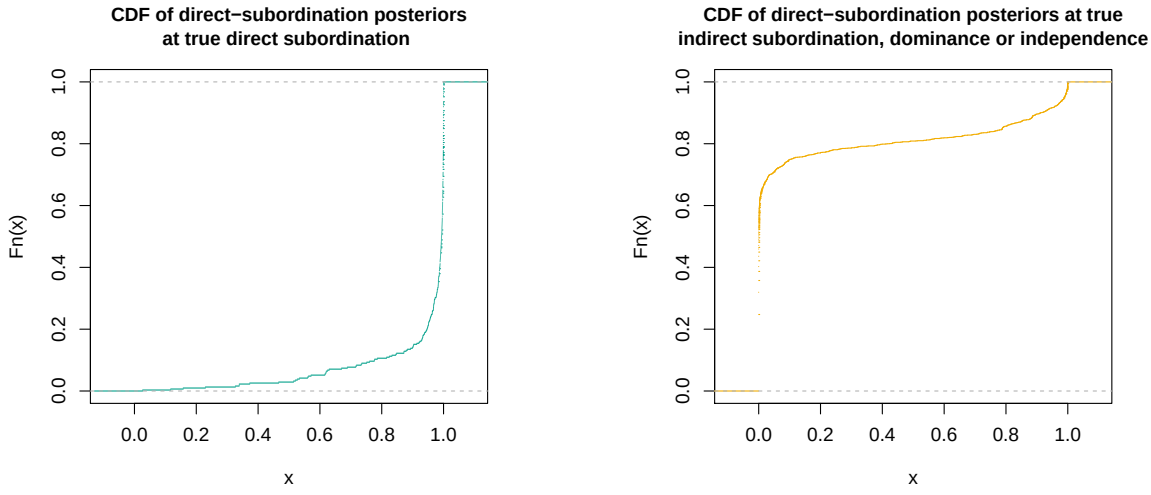


Figure 3.9: The posterior probability of state S_{A5} for animal 2 during the pair $\{2, 5\}$ analysis. The colour indicates when S_{A5} is deemed an acceptable state switch for animal 2 during the group-level analysis when the acceptance boundary is drawn at 0.10.



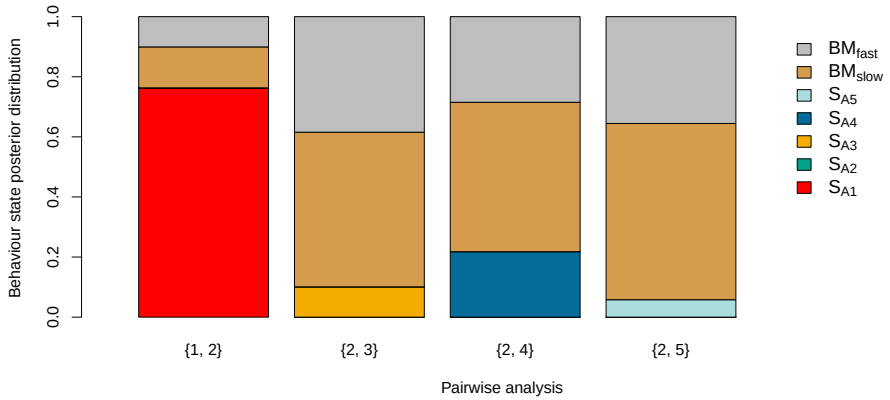
(a) (b)

Figure 3.10: CDFs of the direct-subordination posteriors (that is, the posterior probability that animal i is directly subordinate to animal j) from the preliminary pairwise analyses of the simulated data. (a) concerns the direct-subordination posteriors at times of true direct subordination of animal i to j in the simulation. (b) concerns the direct-subordination posteriors at all other times. The CDFs encompass all pairwise analyses and both subordination orderings.

Group-Level Analysis

We can use the pairwise analysis output to initialise the behavioural parameters of the group-level analysis, in particular, the initial state estimation at each observation. We could have estimated the behaviours in continuous time, along with the required switching times, however, the estimation of the behaviours at the discrete-time observations (along with some rudimentary switching times where required) should suffice as a good starting point. Whilst various methods could be utilised to do this, the method we have used seems simple and intuitive. That is, for each animal i at each observation j, we look at the posterior distribution of the behavioural states in each pairwise analysis involving that animal and simply select the most likely state (provided the resulting structure is not cyclic). However, as an animal can be in the BM states in each pairwise analysis, whereas it can only be in each subordinate state in a single analysis, the BM states can be over-represented. To workaround this, we treat the combined BM posterior probability from each pairwise run as independent and multiply them together to calculate the probability of animal i being in a BM state at observation j. The subordinate posteriors are then weighted so that all state probabilities sum to one. Figure 3.11 details an example.

The pairwise analysis also offers the opportunity to initialise the movement parameters. Figure 3.12 displays the movement parameter output (beyond burn-in) from all of the the pairwise analyses aggregated. To initialise α, the value of the largest peak was taken whilst σ, ρ_{slow} and ρ_{fast} are straightforward (although ρ_{slow} was heavily smoothed). Using these informed starting values for the movement parameters and behavioural states, the burn-in period was reduced to 10k iterations



	S_{A1}	S_{A2}	S_{A3}	S_{A4}	S_{A5}	BM
{1, 2}	0.76	0.00	0.00	0.00	0.00	0.24
{2, 3}	0.00	0.00	0.10	0.00	0.00	0.90
{2, 4}	0.00	0.00	0.00	0.22	0.00	0.78
{2, 5}	0.00	0.00	0.00	0.00	0.06	0.94
Overall	0.56	0.00	0.07	0.16	0.04	0.16

Figure 3.11: The bar chart and table show the posterior distributions of the behavioural states for animal 2 at observation 83 in all the relevant pairwise analyses. The *Overall* row in the table contains the weighted probabilities, which are used to initialise the group-level inference. If BM is the modal behaviour, the BM state with the greatest posterior is selected.

(from 30k in the uninformed analysis). Aside from the starting values and extra trajectory proposal conditions, the MCMC algorithm was set up exactly the same as in the uninformed analysis. This includes the tuning parameter values, which were also kept consistent between all informed analyses (that is, for all acceptance boundary runs).

Tables 3.3, 3.4 and 3.5 detail the performance of each of the boundaries we have trialled. In Table 3.3, a simple proxy for how restrictive these boundaries are is found in the *State Space* row. To obtain these percentages, we looked at each observation for each animal to see what proportion of states met that particular boundary condition. The *True States* row indicates if the boundary rejected any states that we know to be true from the simulated data, with 100% indicating that all true states are still accepted in the group-level analysis. The 0.01 boundary manages to halve the state space whilst still allowing all true states. The 0.05 and 0.10 boundaries further reduce the state space whilst only rejecting one true state between all of the animals. That rejected state was a state visit that only persisted through a single observation and one that also wasn't picked up in the uninformed analysis. The acceptance boundary of 0.20 appears to be a too demanding line to draw. More true states are being discounted whilst only making marginal gains in the amount it reduces the state space.

Whilst the uninformed run converges to the same movement posteriors as all of the informed runs

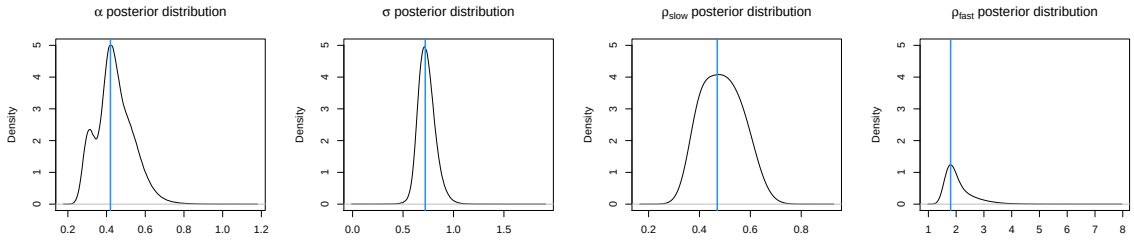


Figure 3.12: The densities of the movement parameters after the output for all 10 pairwise analyses was aggregated. The blue vertical lines indicates the values we used as the starting point for the informed group-level analysis.

	Boundary				
	Uniformed	0.01	0.05	0.10	0.20
Group-Level Burn-In	30k	10k	10k	10k	10k
State Space (%)	100	49.6	42.2	38.8	36.0
True States (%)	100	100	99.8	99.8	99.2
Pairwise Time (HH:MM)	00:00	12:20	12:20	12:20	12:20
Group Time (HH:MM)	26:18	25:55	26:23	22:46	25:32
Total Time (HH:MM)	26:18	39:15	39:43	36:06	38:52
Median Movement ESS	14515	14594	16912	17174	18010
Median Rate ESS	4861	4397	5278	7053	7143
Median Movement ESS/hr	551	371	425	475	463
Median Rate ESS/hr	184	112	132	195	183

Table 3.3: The *State Space* percentage row indicates how much of the state space can still to be explored in the group-level analysis for each of the acceptance boundaries examined as well as the uninformed case. The *True States* row indicates the percentage of true states that are deemed acceptable by each of the boundaries. The *Pairwise Time* is the runtime of the analysis of one pair as this step is easily parallelisable. The *Group Time* is the runtime of the group-level analysis whilst the *Total Time* is the above two rows added together (along with the time it took to collate the pairwise analysis — 60 minutes). The bottom four rows summarise the effective sample sizes (ESS) of the movement parameters and transition rates and the rate at which they were created.

		Boundary				
Movement Parameters		Uninformed	0.01	0.05	0.10	0.20
	Uninformed	1.00	1.01	1.01	1.02	1.03
	0.01	-	-	1.02	1.01	1.03
Boundary	0.05	-	-	-	1.01	1.02
	0.10	-	-	-	-	1.01
	0.20	-	-	-	-	-

Table 3.4: The Gelman-Rubin diagnostic (multivariate PSRF) of the movement parameters for each combination of the uninformed and informed simulation analyses.

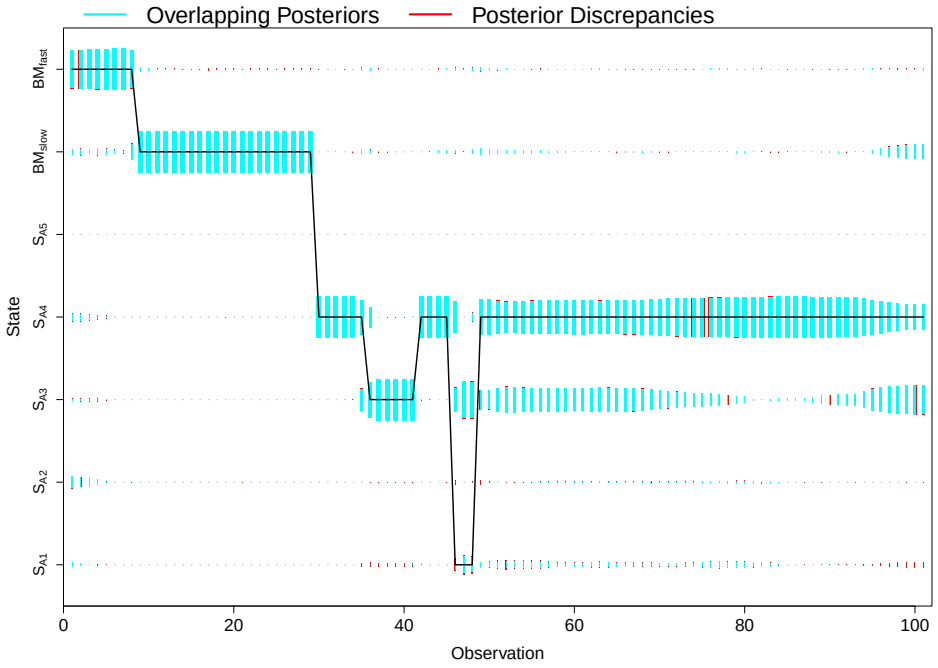
Transition Rates		Boundary				
		Uninformed	0.01	0.05	0.10	0.20
Boundary	Uninformed	1.02	1.08	1.38	1.41	1.53
	0.01	-	-	1.29	1.32	1.45
	0.05	-	-	-	1.03	1.08
	0.10	-	-	-	-	1.06
	0.20	-	-	-	-	-

Table 3.5: The Gelman-Rubin diagnostic (multivariate PSRF) of the transition rates for each combination of the uninformed and informed simulation analyses.

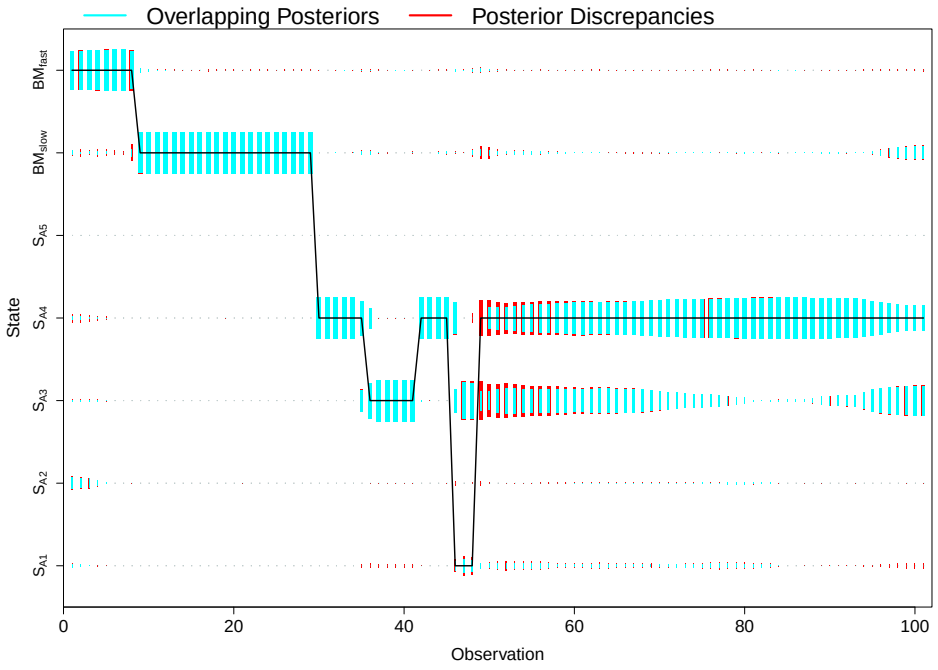
(Table 3.4), the same can not be said for the transition rates (Table 3.5) — indicating that all boundaries prohibitively impact the exploration of the state space to some extent. The evaluation of the uninformed and 0.01 boundary analyses suggests a good level of convergence, but it is nonetheless a downgrade from when comparing two uninformed analyses in Section 3.7.2. Beyond the 0.01 boundary though, there is clear divergence from the uninformed posteriors. This is a result of the more demanding boundaries restricting the exploration of the true state space, leading to discrepancies in the estimated states and subsequently the transition rates. Note, we did not apply the Rubin-Gelman diagnostic across runs with the same acceptance boundary. We were testing to see if each boundary converged to the same posteriors in order to examine the importance of the boundary value. Not only do the informed analyses produce different posteriors to the uninformed case, they also do not provide a consistent improvement in the rate at which we generate effective samples (Table 3.3).

Figure 3.13 explicitly shows the impact of the state space reduction. The state posteriors for animal 5 from an uninformed analysis are overlapped with those from the informed analyses with a 0.01 acceptance boundary (Figure 3.13a) and 0.20 acceptance boundary (Figure 3.13b). The posteriors of the 0.01 boundary are highly consistent with the posteriors of the uninformed case. However, as the 0.20 boundary restricts the state space to a greater extent, the posteriors deviate further from the (more accurate) distributions that are the result of a freely explored state space.

Overall, we were able to initialise the movement and behaviour parameters for the informed group-level analysis from the pairwise output. Doing so enabled us to reduce the burn-in period required from 30k iterations to 10k (Table 3.3). However, these benefits are negligible compared to both the extra runtime required for the pairwise analysis and the restrictions more demanding conditions have on state space exploration. Though, the 0.01 boundary may be useful in scenarios where group-level model fittings struggle to move through the burn-in period by providing sensible initial parameter values.



(a)



(b)

Figure 3.13: The state posteriors for animal 5 from an uninformed analysis overlapped with those from an informed analysis with a 0.01 acceptance boundary (a) and a 0.20 acceptance boundary (b). The black line is true state in the simulated data.

3.8 Zebra Data Analysis

We fitted the model to GPS data of four wild Burchell's zebra (*Equus burchellii*) in northern Botswana to examine its capability when applied to real data. The data is available on Movebank (Bartlam-Brooks & Harris, 2013) and was originally analysed by Bartlam-Brooks *et al.* (2013) to predict effects of environmental change on zebra migration, during which the movement of each zebra was analysed independently. Data was collected for seven adult mares at hourly intervals over 20 months, but not all were tagged simultaneously for the entire time span.

The subset of data that we have analysed encompasses approximately 11 days for four zebra that are simultaneously tracked (tag IDs 6399, 6402, 6405 and 6407). This period was purely selected due to this synchronicity occurring for a large enough number of observations that would allow us to suitably trial our model. In total, this subset contains 1056 observations over the four animals. Whilst most of these (1021) were recorded at hourly intervals, some slightly deviated by a minute (30) or two (5). For the purpose of this analysis, we rounded the slightly asynchronous observations to the nearest hour as little information would be gained from keeping those slight deviations, and it enabled us to reduce the data set from 298 asynchronous to 264 synchronous observations. We were informed by the original collectors of the data that the zebra were not from the same harem. Whilst that will limit the amount of social interaction in the data, the zebras may be thought of as a proxy for their harem and so we can potentially gain insight into how the harems interact when they join together to form larger herds (see Wittemyer *et al.* (2007) for a similar tactic with African elephants).

We have included two BM states for leading or independent animals when formulating the specific model to be fitted. In a brief attempt to fit the model with only a single BM state to the data, the subordination states translated to a pseudo-BM process (that is, an OU process with extremely weak attraction) that represented a different speed of movement to that of our actual BM state. It therefore became necessary to include an additional BM state to better model that diversity of movement and allow the subordination states to represent the social behaviours. We will again compare analyses when being 'informed' by a preliminary pairwise analysis and when not ('uninformed').

3.8.1 Uninformed Analysis

As this is an uninformed analysis, the behaviour and movement parameters were initialised randomly and/or over-dispersed. After some tuning runs it became apparent we required a greater λ_{max} than in the simulated data, this time it was set to 1, to enable an appropriate level of state switching ($\max(\lambda_u) = 0.91$ in the resulting analysis). As a result of this increased level of switching, trajectories were only reconstructed over tiny segments of data in order to obtain decent acceptance rates. The collective trajectories were simulated over only a single interval whilst individual updates were undertaken over 3 to 5 observations. Even then, the acceptance rates were only 14%

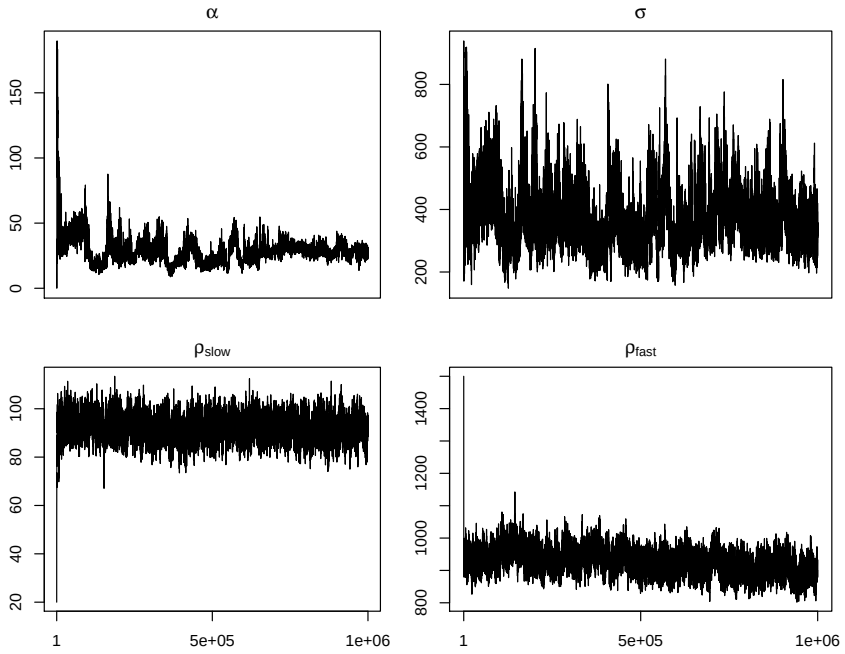
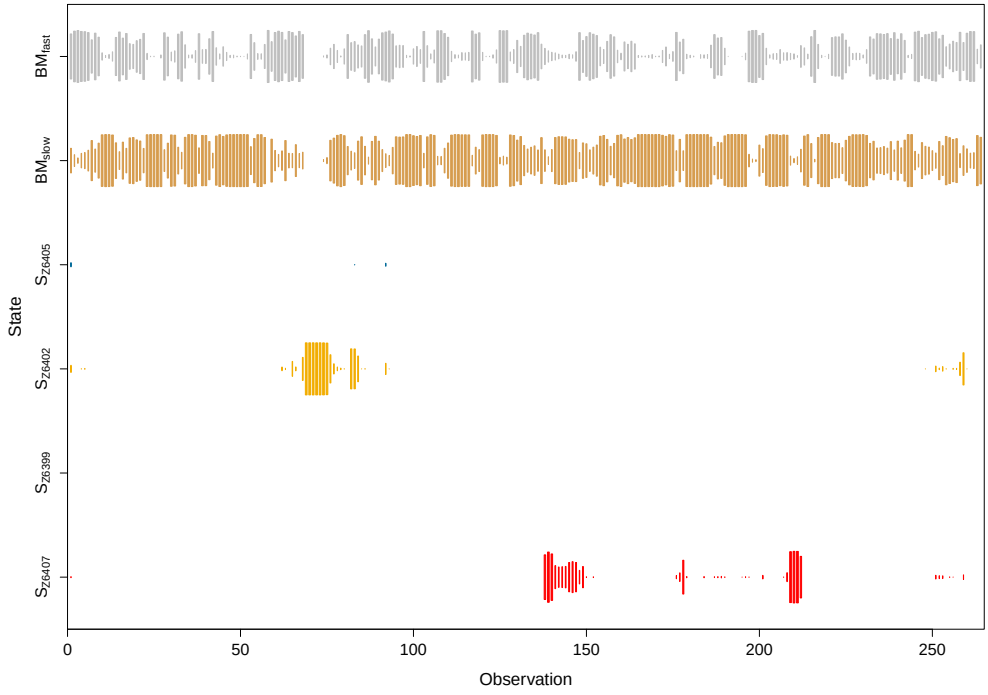


Figure 3.14: Trace plots for the four movement parameters from one of the uninformed model fittings to the zebra data. The chains have been thinned by a factor of 50.

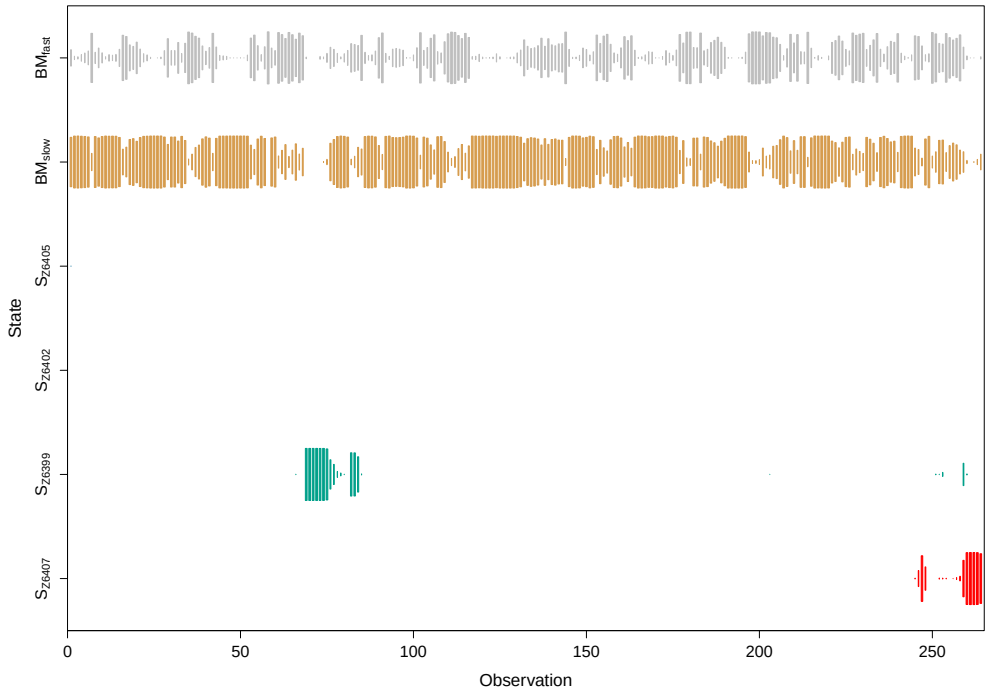
and 30% for collective and individual updates respectively. We simulated 263 new collective and individual trajectories per iteration. We ran the MCMC algorithm for 1 million iterations and an uninformed Dirichlet prior was used for the transition rates.

We ran the uninformed inference twice, neither of which reached convergence. Figure 3.14 shows the movement parameter trace plots for one of the runs. The α and σ samples struggle to settle in a particular range and the ρ_{fast} samples look like they are still shifting slightly — a stark contrast with the trace plots of the uninformed simulation analysis in Figure 3.5. In order to analyse these runs, we have taken the burn-in to be 700k iterations as the chains beyond that look relatively stable. Not only do the two runs not converge separately, they do not approach the same posteriors. This is most explicit at times when the state posteriors from the different runs estimate opposing behaviours. Figure 3.15 shows an example of this with the state posteriors for two zebra (IDs 6399 and 6402) where, through observations 69-83, the posteriors from the different analyses estimate opposing interaction. That is, in Figure 3.15a (from uninformed run 1) zebra 6399 is subordinate to 6402, whilst in Figure 3.15b (from uninformed run 2) zebra 6402 is subordinate to 6399. The difference in behaviours is then accompanied by differing posteriors for the movement parameters as shown in Figure 3.16.

From Figure 3.16, we can see that in both analyses α is quite high in the context of a subordinate's movement distribution (see equations 3.8 — 3.12 in Section 3.3 with $t = 1$ in this case). As discussed in that section, this potentially alludes to: the resolution of the data is too coarse to



(a)



(b)

Figure 3.15: The state posteriors for zebras 6399 (a) and 6402 (b). These are from two different uninformed runs. Through observations 69-83 the state posteriors estimate opposing interaction.

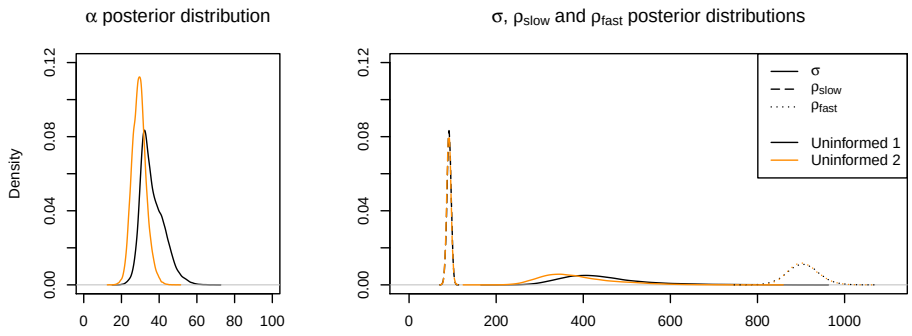


Figure 3.16: The posterior distribution of the four movement parameters from the two uninformed zebra analyses.

accurately capture the interaction, or the interaction of these animals is not suitably modelled by our assumption of there being some social order. The opposing interaction estimations may be born out of the scenario where the animals are interacting in a mutually informed sense but we are forcing the state labels to declare a social order — with either order being equally likely if not a good representation of the interaction. Alternatively, both animals could be moving according to some environmental cue that we are not observing and so we are falsely inferring some interaction.

All of the above reasons suggest our model is not well suited to modelling the movement described by this data set. Whilst that might be the case, the opposing state estimations might also be born out of the poor mixing that such short trajectory updates will result in. It may be that undertaking a pairwise step before the group-level analysis can help alleviate some of that difficulty.

3.8.2 Informed Analysis

Preliminary Pairwise Analysis

For the pairwise analysis we have again fitted our model to each pair of animals. Whilst the motivation for this is to ascertain the interaction between each pair in a simpler context, the rate of accepting trajectory proposals was still low and we were restricted to updating them over the same lengths of data as in the uninformed group-level analysis. The movement and behavioural parameters were initialised with random and/or over-dispersed values and we ran each pair for 1 million iterations.

For this pairwise analysis, we had to use different proposal distributions for the movement parameters for each pair (i.e. the standard deviations were different pair to pair). We initially trialled using the same standard deviations but they didn't universally facilitate good mixing. This necessity was mostly driven by the attraction parameters α and σ. As we can see from Figure 3.17, which contains the posterior distributions of the movement parameters for each pair, those two parameters took vastly different values in each analysis. For this particular data set, this extra tuning requirement

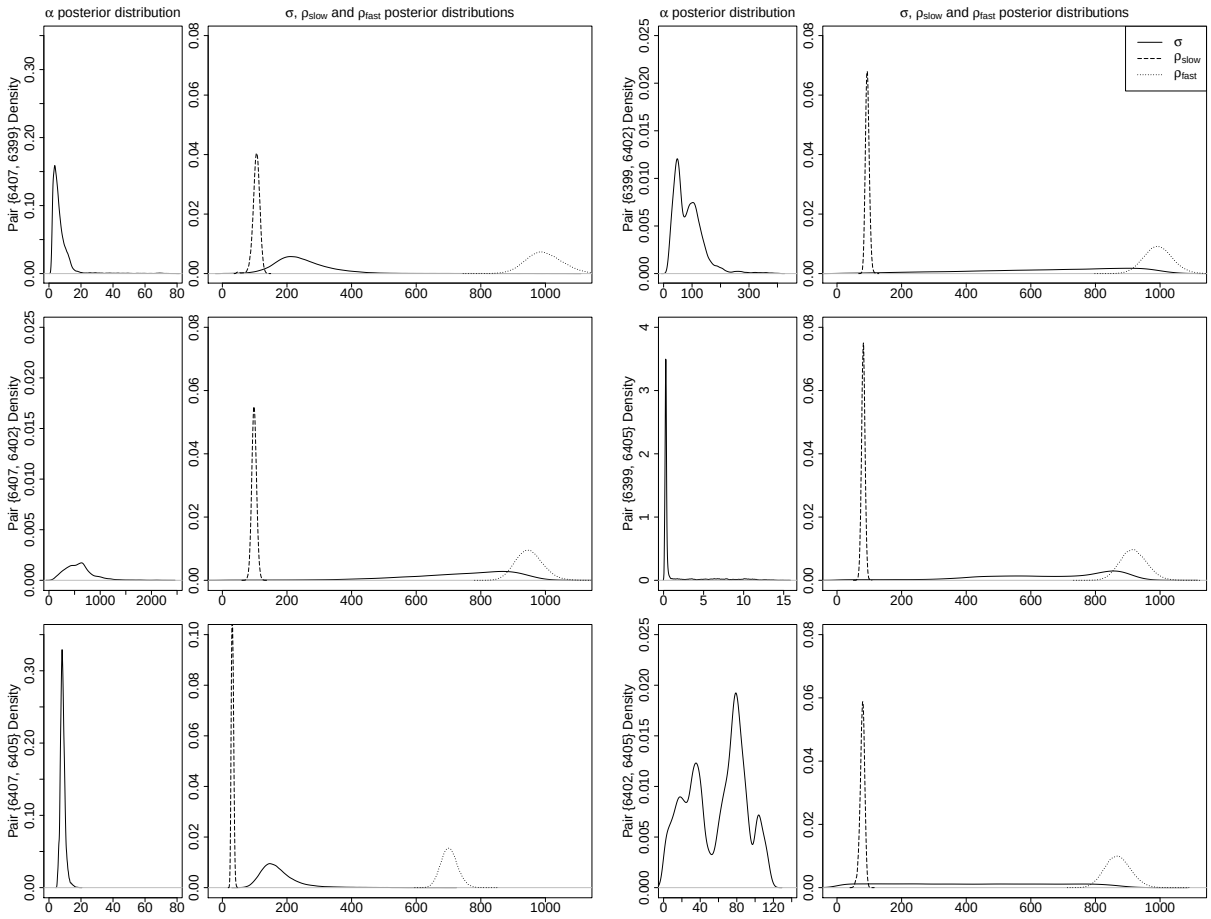


Figure 3.17: Posterior distributions of the four movement parameters for each of the zebra data pairwise analyses. Note the different scales of the α plots and pair {6407, 6405} noise coefficients plot.

was merely a nuisance as there are only six pairwise analyses to tune for a group of four animals. However, for larger groups, this could be a substantial hurdle as one may need to separately tune $\binom{n}{2}$ runs for a group of n animals.

Even with this customisation, we were not able to obtain entirely satisfactory results. For example, there appears to be some bi-modality in α for pair {6399, 6402} (as shown in Figures 3.17 and 3.19) that was not fully explored in the set number of iterations. That posterior, as well as the posterior of α for pair {6407, 6402}, covers a huge range of values. There is little interaction inferred in those two pairs (see Figure 3.18) and so we would expect a large degree of uncertainty in the posterior of α. Though, we suspect in these cases these posteriors are a consequence of the issues discussed in the uninformed zebra analysis. That is, this model does not accurately capture the true interaction (if any) between these animals and so there is little to distinguish between higher values of α. Pair {6399, 6405} also contains little estimated interaction (Figure 3.18), but this time the interaction inferred is very weak when compared to the other pairs (Figure 3.17).

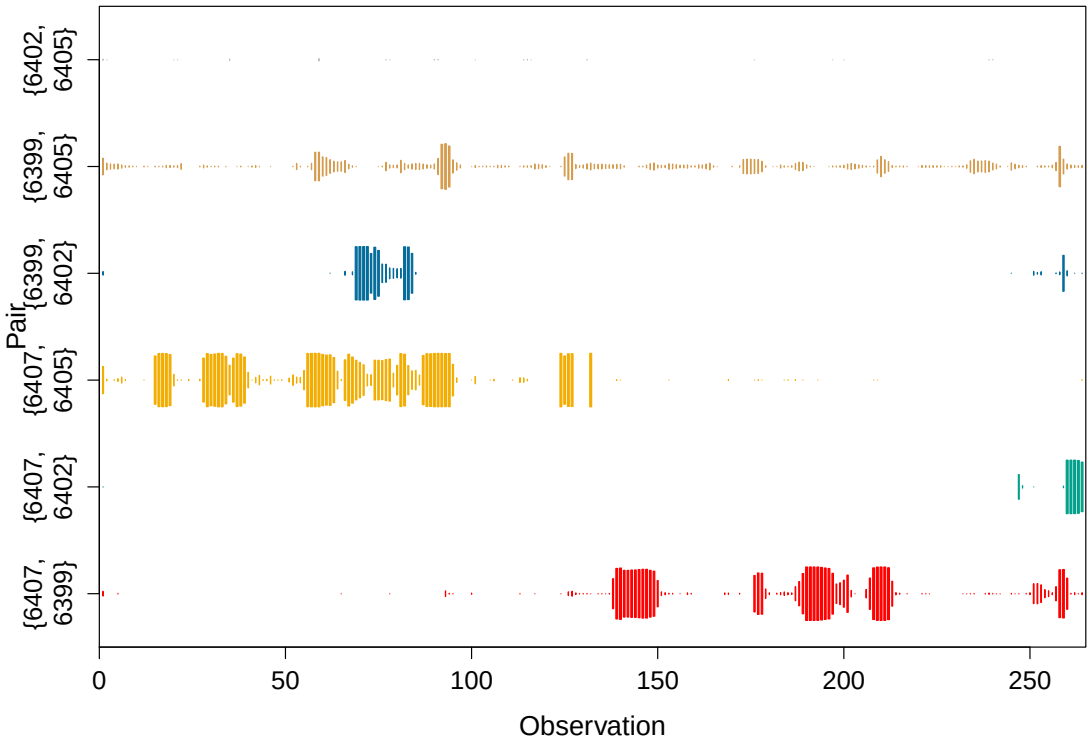


Figure 3.18: The posterior distributions of a pair of zebras interacting from their respective pairwise analysis. The posteriors encompass either interaction ordering.

As only a small of amount of interaction is inferred in all three of the aforementioned pairs, we would also expect a lot of uncertainty with regards to σ — which is exactly what we have. However, each σ posterior for those pairs has a peak overlapping with the posterior of ρ_{fast}. Rather than the σ peaks indicating a larger degree evidence for those values, we suspect they are a result of our restriction that $\sigma < \rho_{max} = \rho_{fast}$ (Section 3.6.3), the large standard deviations we used for the noise coefficients proposal distributions in order to cover the wide range of possible values and the fact that we update all movement parameters simultaneously. As such, when the current samples of σ and ρ_{fast} are similar, newly proposed candidates for those parameters often fall foul of the above restriction and so the chains momentarily get stuck. Additionally, in cases of large α values, as in pairs {6407, 6402} and {6399, 6402}, the restriction that $\sigma < \rho_{max} = \rho_{fast}$ loses some intuition and purpose as the two animals tend towards the same movement process (see Section 3.3). The peak of σ in those cases may therefore be beyond the peak of ρ_{fast}. These three pairs highlight how tricky it can be to tune this pairwise step.

The movement posteriors for pairs {6407, 6399} and {6407, 6405} are more straightforward — these pairs also contain the most amount of interaction. For pairs that contain no interaction, the α and σ samples are effectively a random walk — as seen for pair {6402, 6405}. Despite the above challenges in running this pairwise step, we do not require the pairwise analyses to have

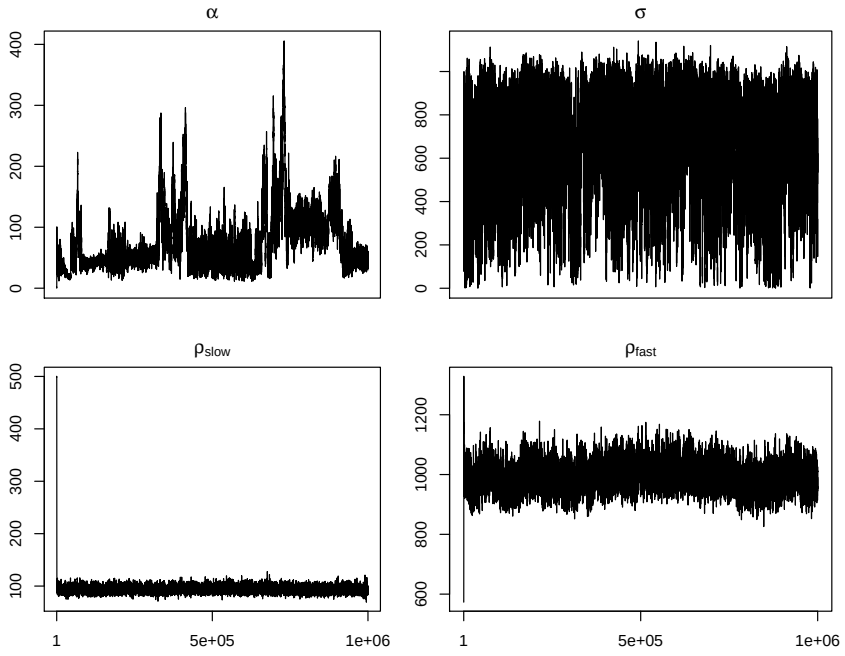
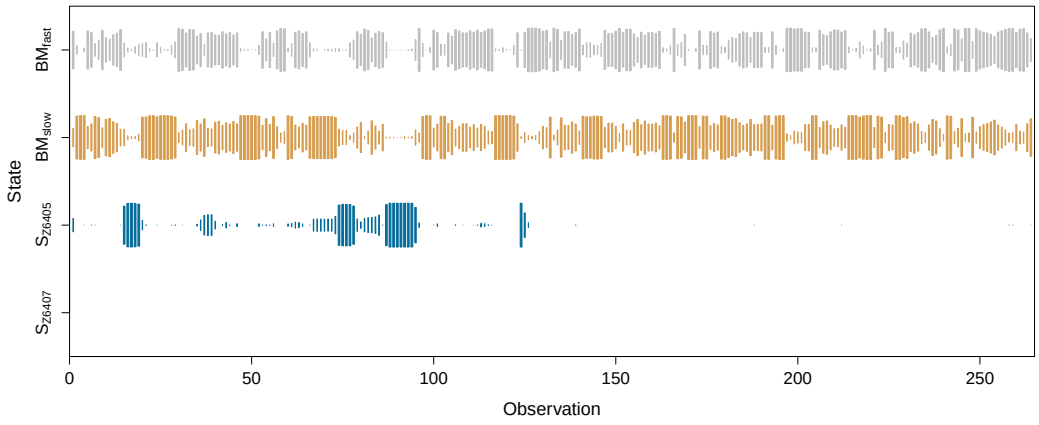


Figure 3.19: Trace plots for the four movement parameters from the zebra pair {6399, 6402} analysis. The chains have been thinned by a factor of 50.

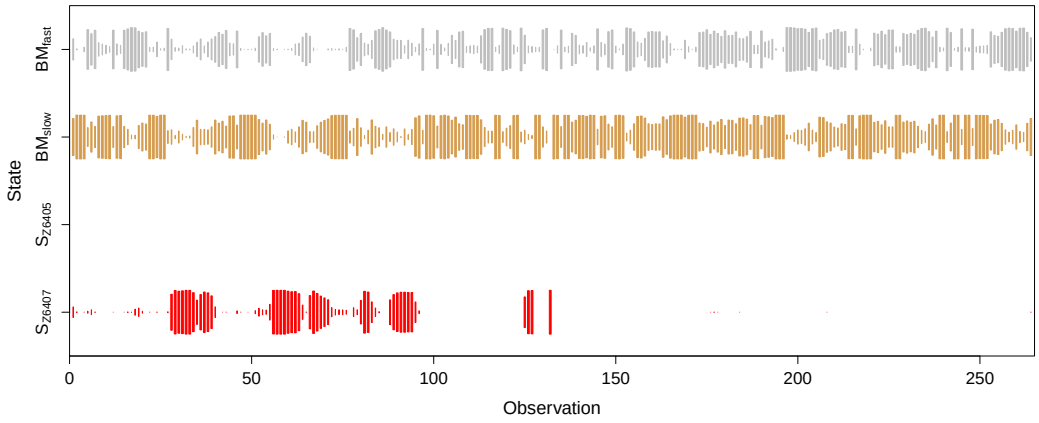
converged in order to guide the group-level analysis, merely that they have surpassed the burn-in period — particularly when using lenient acceptance boundaries. We have taken the burn-in to be 500k iterations for all pairs, which seems sufficient in each case.

Similar to the uninformed zebra analysis, the pairwise runs also estimate conflicting interaction. For example, in pair {6407, 6405}, there is a segment of data for which the modal state estimation in one run classifies zebra 6407 as being subordinate to 6405, whereas in another run zebra 6405 is subordinate to zebra 6407. This can be seen in Figure 3.20 through observations 89-94.

In order to use this pairwise information to guide the group-level inference, similarly to Section 3.7.3, we deploy an acceptance boundary when determining which states are acceptable during the group-level trajectory proposals. Based on the results from Section 3.7.3, we will only run a 0.01 acceptance boundary as more demanding conditioning impaired the ability of the group-level inference to explore the necessary state space whilst offering little benefits. However, as we have seen from two different runs of the pairwise analysis in Figure 3.20, the interaction between two animals has not converged to the same ordering. Which analysis we use to guide the group-level conditions will then dictate the state space we can explore. We have therefore decided to use a 'mutual' acceptance boundary for the informed zebra analysis. That is, when using an acceptance boundary of x, if the posterior of zebra i being subordinate to zebra j is at least x in the pair $\{i, j\}$ run, not only will this behaviour be acceptable in the group-level analysis but so too will zebra j being subordinate to i.



(a)



(b)

Figure 3.20: The state posteriors for pair {6407, 6405}. (a) are the posteriors for zebra 6407 and (b) are the posteriors for zebra 6405 but from a separate analysis. The two analyses estimate a conflicting modal state (and therefore interaction) through observations 89-94.

Group-Level Analysis

We initialise the behavioural parameters of the group-level analysis much like in Section 3.7.3 for the informed simulation analysis. However, similar to the acceptance boundary, there is an obvious issue with this method when our pairwise analysis has not converged. Or at least, we know that multiple pairwise analyses do not converge to the same posteriors. Nevertheless, we still implemented this approach as it is just intended to be a sensible starting point and a guide for the group-level MCMC run. Furthermore, we have relaxed the conditioning on the pairwise output, as mentioned above, which enables both contrasting interactions to be proposed in the group-level updates.

When looking to initialise the movement parameters of the group-level analysis using the information we have gained from the pairwise step, we can see from Figure 3.17 that this step is not as

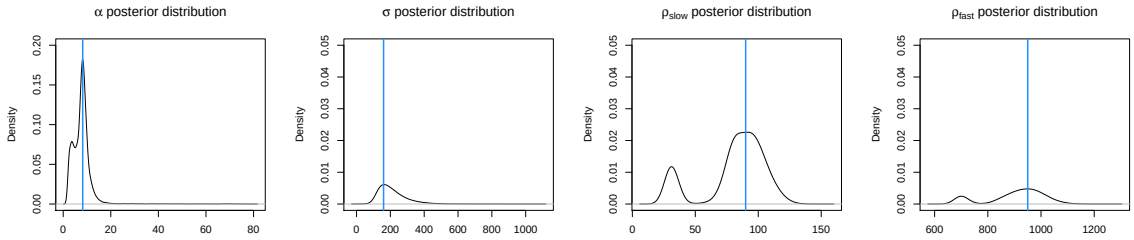


Figure 3.21: The densities of the movement parameters after the output for all six pairwise analyses was aggregated for the ρ parameters and the output of pairs {6407, 6399} and {6407, 6405} was aggregated for α and σ. The blue vertical lines indicates the values we used as the starting point for the group-level zebra analysis.

straightforward as it was for the simulated data. In particular, this is difficult for the attraction parameters α and σ. There are wide ranges of values covered by the pairwise analyses and so selecting some arbitrary point in some combined distribution will unlikely offer any benefit. However, as mentioned above, pairs {6407, 6402} and {6399, 6402} (the cause of most of the diversity in the α values) are perhaps not interacting in a way that is suitably modelled by our social framework. Furthermore, pair {6399, 6405} only contains a minuscule amount of inferred interaction whilst {6402, 6405} doesn't contain any. Therefore, we only utilised the output from pairs {6407, 6399} and {6407, 6405} to select an appropriate starting value for α and σ (Figure 3.21). Regarding the ρ parameters, the smaller peaks in Figure 3.21 are due only to pair {6407, 6405} whilst all other pairs contribute to the higher-valued peaks — we have therefore selected the higher-value peaks as the initial values.

The state posteriors for the informed run are highly consistent with one of the uninformed runs (uninformed run 2) as shown in Figure 3.22. However, that inherently means they are inconsistent with uninformed run 1 due to the discrepancies between the two uninformed analyses. The method we used to initialise the behavioural process in the group-level inference produced a starting point that was more aligned to uninformed run 2. Whilst the inference algorithm allows for this interaction to reverse during the trajectory updates, in practice this would be an extremely unlikely event given the poor mixing resulting from having to propose new trajectories over very small segments of data in order to obtain a reasonable acceptance rate.

Figure 3.23 compares the posteriors of the movement parameters from multiple informed runs. These runs were initialised with the same movement parameters and behavioural process estimated from the pairwise step. Despite this, the posteriors clearly have not converged as further indicated by the Gelman-Rubin diagnostic — the multivariate PSRF of these parameters is 1.25. Additionally, the multivariate PSRF for the transition rates is 1.99. However, Figure 3.24 does highlight the benefit of being able to initialise the model parameters with sensible values as within-chain convergence was achieved with fewer iterations. The greatly reduced burn-in period in turn facilitated a greater rate of generating effective samples — even when taking into account the sizeable

Figure 3.22: The state posteriors for zebra 6405 from the informed analysis and uninformed run 2 overlapped.

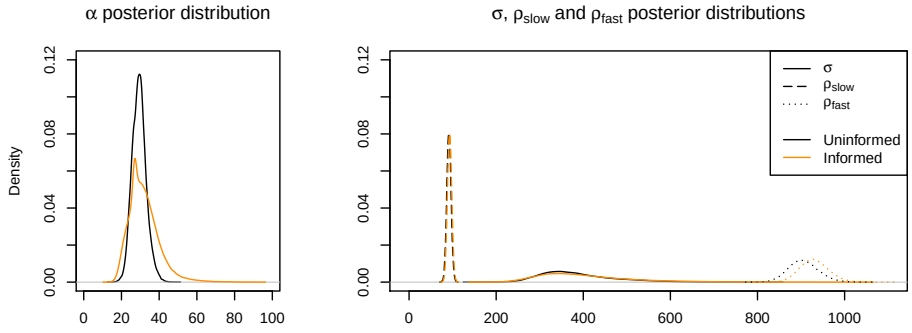


Figure 3.23: Movement parameter posteriors from uninformed run 2 and the informed zebra analysis.

run time of the pairwise analysis (Table 3.6). Given that the acceptance rates of the trajectory updates remain consistent between the uninformed (13.9% (collective) and 30.2% (individual)) and informed analyses (14.4% and 30.6%), it appears that the ESS improvement is from the burn-in improvement alone and not through better mixing of the behaviours. It is interesting that the pairwise analysis doesn't run much quicker than the group-level analysis — though this is perhaps a blessing from the point of view of increasing the number of animals in future analyses.

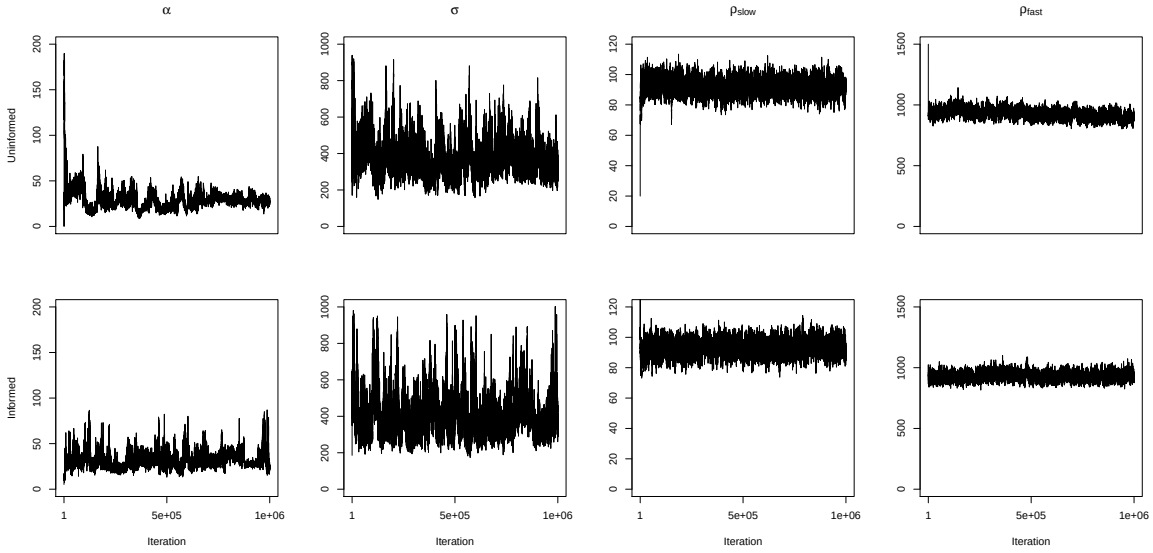


Figure 3.24: The trace outputs of the movement parameters from uninformed run 2 (top row) and the informed zebra analysis (bottom row). The chains have been thinned by a factor of 50.

3.9 Discussion

Our influence hierarchy model offers a flexible method in continuous time with which we can capture social interactions in animal movement. Modelling the state switches in continuous time means we avoid approximation from using discrete data and our social behaviour analysis is not bound to the temporal scale of the data. Through the influence hierarchy framework, a wide range of social constructs can be captured: from despotic leadership to fission-fusion dynamics. Furthermore, social hierarchies are a simple and common concept meaning a certain amount of the abstraction of statistical models is stripped away.

Our formulation of the behavioural states allows us to capture rich information with regards to the social behaviours in a group's movement. Animals consistently influencing the movement of other animals may be considered 'keystone' animals and identifying such individuals can have productive applications in conservation and management decisions (King *et al.*, 2018; Westley *et al.*, 2018). Obtaining a picture of a group's social structure may also help us understand how resilient or adaptive they are to change (King *et al.*, 2018) and we can monitor how anthropogenic activity might be impacting them (Westley *et al.*, 2018).

We restrict the social structure to the hierarchies defined in section 3.1 in order to keep the model tractable. However, this does mean we omit certain interactive behaviours. For instance, the 'double-subordinate' as in McDonald & Shizuka (2013) and discussed in Strandburg-Peshkin *et al.* (2018), where an animal is equally influenced by two others (or more). Cyclic structures are also not accounted for in our model, but analysis from McDonald & Shizuka (2013) suggests social structures tend to be highly orderly and so this may not be problematic.

	Uniformed	Informed
Group-Level Burn-In	700k	75k
State Space (%)	100	37.7
Pairwise Time (HH:MM)	00:00	54:45
Group Time (HH:MM)	76:51	78:08
Total Time (HH:MM)	76:51	133:33
Median Movement ESS	502	1550
Median Rate ESS	1054	5680
Median Movement ESS/hr	6.5	11.6
Median Rate ESS/hr	13.7	42.5

Table 3.6: The *State Space* percentage row indicates how much of the state space is still to be explored in the group-level analysis given the acceptance boundary used. The *Pairwise Time* is the runtime of the analysis of one pair as this step is easily parallelisable. The *Group Time* is the runtime of the group-level analysis whilst the *Total Time* is the above two rows added together (along with the time it took to collate the pairwise analysis — 40 minutes). The bottom four rows summarise the rate at which effective movement parameter and transition rate samples were created.

Whilst the collective movement model we have detailed above is inspired by the one presented by Niu *et al.* (2016), there are some key differences to both that model and the one subsequently presented by Niu *et al.* (2020). Our followers can be attracted to any animal, not just a leader, and so we can model a wider range of social structures. Through the behavioural switching process, all animals can switch between being a follower or a leader (or independent). As such, we can capture changes in influence over time. It is these differences that provide our modelling framework with greater flexibility. For instance, allowing all animals the possibility to be leaders means that we can represent multiple subgroups, which will be required to model fission-fusion dynamics. Though, whilst these changes have the potential to capture complex social behaviours, they will not always be warranted.

Our social interaction assumption is one based on order — that is, there is a dominant and subordinate in each edge of the hierarchy. However, that will not always be a fair assumption — such as the case where two or more animals are mutually influenced or informed by each other. This is evident in Section 3.8 when we fit the model to zebra movement data as separate analyses inferred opposing interaction with similar certainty. Furthermore, through this definition of the behavioural states, any interaction we do infer is direct. This is susceptible to falsely declaring direct influence when not all relevant animals have been tagged. Again, this is demonstrated within the zebra data analysis. The data set we have analysed contains only four zebra and each one was from a different harem. There are many other animals they are likely to be interacting with that we don't have data on and so we have to be conscious not to over-interpret our results as the de-facto social behaviour. Even if we took the approach of viewing each zebra as a proxy for their harem, it is unlikely only four harems have collated to form a larger herd at times of interaction. Niu *et al.* (2016) and Niu *et al.* (2020) use a moving central point that acts as the focus of a group's movement which can

navigate the case of having incomplete data on a social group. Whilst that approach doesn't allow the finer details of social behaviour to be examined, treating the group as a mutually-informed collective may be a more suitable route to model the social drivers of zebra movement.

Whilst probing further into the social interactions can offer richer information, the sheer size of the state space can be problematic as a considerable amount of time is needed to fully explore it. We offer a potential solution to this problem in the form of the preliminary pairwise analysis, which can simplify the state space of the group-level analysis. However, the results from both the simulation and zebra analyses suggest this pairwise step does not improve the rate at which we can explore it. Furthermore, the conditioning on the pairwise analysis can be quite a sensitive process and can easily lead to 'correct' areas of the state space being ignored. The pairwise step does potentially have some benefit through reducing the burn-in period of the group-level analysis by enabling us to initialise a complex inference algorithm with sensible parameter values. Though, even that is not without its risks — particularly in the context of a slowly mixing algorithm. That is, the modelling fitting algorithms may struggle to move on from the initial, reasonable parameter values that might just be local maxima.

Overall, whilst the model has the potential to gain detailed insight into the social drivers of animal movement when our modelling assumptions are fair, it is currently limited by the inference methods. The current method of proposing trajectories is naive as we simulate forwards through the data whilst ignoring the required end state and the previously accepted trajectory. Furthermore, we currently require collective updates in order to resample the switching times as we allow all animals to switch simultaneously. The naivety of forwards simulation is compounded when doing so with respect to multiple animals and so acceptable trajectories are rarely sampled — limiting the wider MCMC process. If we are to extract the potential of this influence hierarchy framework, and further extensions such as spatial heterogeneity, we require a much more efficient means of proposing new trajectories. This is explored in Chapter 4.

Chapter 4

An Improved Model Fitting Algorithm and Model Evaluation

The influence hierarchy model we outlined in Chapter 3 has the potential to capture the social drivers of animal movement. The results in that chapter begin to show that potential, particularly when the model was fitted to simulated data. However, it is clear that the inference approach previously outlined is not efficient enough in order to fit such a complex model. We introduce a newly developed algorithm in Section 4.1, which we then use to fit the influence hierarchy model to simulated data (Section 4.2) and baboon data (Section 4.3). In Section 4.4, we explore the robustness of our model, which is followed by an investigation in Section 4.5 into whether data sets can be thinned in order to ease the computational cost of the model fitting process whilst still obtaining useful results. Finally, we compare our approach to modelling the social element of animal movement to some existing methods in Section 4.6. The majority of the material in this chapter was introduced in Milner *et al.* (2021).

4.1 New Model Fitting Algorithm

Much like in Chapter 3, Markov chain Monte Carlo (MCMC) methods are used to infer both the behavioural and movement parameters. Again, each iteration of the new algorithm consists of two parts. Firstly, we sample the behavioural process of the animals in continuous time. Secondly, the parameters of the Ornstein-Uhlenbeck (OU) and Brownian motion (BM) processes are sampled in accordance with the latest behaviour trajectory reconstruction by means of a Metropolis-Hastings (MH) random walk. The updates to the inference algorithm are entirely contained in the first part — simulating the behavioural trajectories of the animals. As such, the element of the inference process regarding the movement parameters is exactly the same as in Section 3.6.3.

In Chapter 3, we allowed all animals to switch state simultaneously at each sampled switching

time in order to reduce the size of the augmented data (see Section 3.5.1). This approach proved inefficient as it meant we had to update the trajectories of all animals simultaneously. Collective updates involve discarding the entirety of the previously accepted trajectory and performing a new forwards simulation for the entire group. The likelihood of a forwards simulation for all animals meeting all of the relevant conditions (such as completing in the required state), as well as proposing reasonable behaviours for each of the animals, is slim. As such, the rate of acceptance of the proposals is small. Even after adding in individual-level updates to improve the exploration of the behaviours, the previous algorithm just didn't mix well.

We have therefore decided to restrict each switching time to contain at most one switch. That is, at most one animal can switch state at any given time. With this restriction in place, we have developed an algorithm which only proposes a trajectory for a single animal in each update, whilst treating the current trajectory estimates of the other animals as fixed. Note, this is different from the 'individual-level' updates in Chapter 3 as we can now resample the switching times an animal at a time as each switching time only relates to a single animal — thus negating the need for collective updates. Through this approach, we explore the social behaviours of the group in a stepwise fashion by making relatively small modifications to the overall group-level trajectory. This approach will require numerous updates to result in wholesale changes to the social structure estimated (if required), but each update will be based on a trajectory that is already deemed plausible and will be beholden to fewer conditions (as it is only simulating a single animal). Therefore, in theory, newly proposed trajectories will be accepted more often, which in turn will mean they can be simulated over longer segments of data — both of which will improve the mixing of the MCMC process. This approach will also allow us to see how changes in hierarchy evolve over time (as opposed to just spontaneously occurring), and it is a feature/assumption shared with other social network approaches such as the stochastic actor-based model by Snijders *et al.* (2010). This new algorithm is detailed in Section 4.1.1.

Note that we are still using the transition matrix defined in Section 3.4. That is, the states of the Markov chain relate to the behaviour of an individual animal. Whilst the restriction that only one animal can switch state at any one time would mean a transition matrix relating to a group switching between hierarchies would be mostly populated with 0s, it would still be a large, unwieldy and difficult to define matrix. Thus we are persisting with the individualistic approach. Additionally, the work in this chapter is still contained within a spatially homogeneous context.

4.1.1 Behavioural Parameters

Similarly to the previous algorithm (Section 3.6.1), we estimate the behavioural parameters (both the states and Λ) by sampling the animals' behavioural trajectories between the observed data points. That is, the observed data is augmented with sampled state switching times in order to treat the behaviours in continuous time. As before, we only simulate a small section of the

trajectory with each proposal.

Let τ_o represent the observed data times for the interval $[t_a, t_b]$; τ_{s_i} represent the current switching times in the same interval for animal i; $\tau = \tau_o \cup \tau_s$ so that $\tau = \{\tau_1 < ... < \tau_p\}$ where p is the size of τ and $\tau_s = \bigcup_i \tau_{s_i}$ for all animals i. Finally, let β represent the behavioural states of all animals at times τ where β_j is a vector containing the states of all animals at time τ_j for $j = 1, ..., p$.

To propose a new trajectory for a given animal, say i, in the interval $[t_a, t_b]$, we discard the current switching times for i, τ_{s_i}, and sample new ones, $\hat{\tau}_{s_i}$. The new potential switching times are produced from a homogeneous Poisson process over the interval (t_a, t_b) with rate λ_{max}, where $\lambda_{max} \geq \max(\lambda_u)$ for all states u. $\hat{\tau}_{s_i}$ is combined with $\tau_{s_{-i}}$, the remainder of τ_s, to produce a new set of switching times, $\hat{\tau}_s$, and augmented times, $\hat{\tau} = \tau_o \cup \hat{\tau}_s$.

We then simulate the behavioural process forwards through $\hat{\tau} = \{\hat{\tau}_1 < ... < \hat{\tau}_{\hat{p}}\}$, where $\hat{p}$ is the size of $\hat{\tau}$, to obtain our new behavioural trajectory $\hat{\beta}$. We initialise $\hat{\beta}_1 = \beta_1$, after which there are three scenarios to account for:

- If $\hat{\tau}_j \in \tau_o$ for $j = 2, ..., \hat{p}$, the behavioural states of all animals at $\hat{\tau}_j$ are carried forward from $\hat{\beta}_{j-1}$ as observations are not part of the behavioural process.
- If $\hat{\tau}_j \in \tau_{s_{-i}}$ for $j = 2, ..., \hat{p} - 1$, we use our previously sampled states for all animals except i at these times whilst the behavioural state of i is carried forward from $\hat{\beta}_{j-1}$.
- If $\hat{\tau}_j \in \hat{\tau}_{s_i}$ for $j = 2, ..., \hat{p} - 1$, the behavioural states of all animals except i at $\hat{\tau}_j$ are carried forward from $\hat{\beta}_{j-1}$. The probability of $\hat{\tau}_j$ being a switch for i is λ_u/λ_{max} when i is in state u. If so, the new state is v with probability λ_{uv}/λ_u, otherwise, the state of i at $\hat{\tau}_j$ is also carried forward from $\hat{\beta}_{j-1}$.

As with the old algorithm in Section 3.6.1, we require $\hat{\beta}_1 = \beta_1$, $\hat{\beta}_{\hat{p}} = \beta_p$ (except for when $\hat{\tau}_p$ is the final observation in the data set) and the new trajectory must not create a cyclic hierarchy at any point. If any of these conditions aren't met, we reject the trajectory and return to our previously accepted trajectory for i.

If the above conditions are met, we accept or reject the new trajectory with a Metropolis-Hastings step. The behavioural trajectories are once again proposed from the model itself (like in Chapter 3), and so the MH ratio simplifies to a ratio of likelihoods of the observed movement through the newly proposed state switches against our previous reconstruction:

$$\prod_{k=2}^{|\tau_o|} \frac{f(\boldsymbol{g}_{\tau_{o_k}} | \boldsymbol{g}_{\tau_{o_{k-1}}}, \hat{\beta}_{[\tau_{o_{k-1}}, \tau_{o_k}]}, \hat{\tau}_{[\tau_{o_{k-1}}, \tau_{o_k}]})}{f(\boldsymbol{g}_{\tau_{o_k}} | \boldsymbol{g}_{\tau_{o_{k-1}}}, \beta_{[\tau_{o_{k-1}}, \tau_{o_k}]}, \tau_{[\tau_{o_{k-1}}, \tau_{o_k}]})} \tag{4.1}$$

where $\boldsymbol{g}_{\tau_{o_k}}$ are the locations of the animals in a particular axis at time τ_{o_k}; $\hat{\beta}_{[\tau_{o_{k-1}}, \tau_{o_k}]}$ are the newly sampled states throughout the interval $[\tau_{o_{k-1}}, \tau_{o_k}]$ at times $\hat{\tau}_{[\tau_{o_{k-1}}, \tau_{o_k}]}$; $\beta_{[\tau_{o_{k-1}}, \tau_{o_k}]}$ are the previous

state estimations in the same interval at times $\tau_{[\tau_{o_{k-1}},\tau_{o_k}]}$. The above terms are also conditional on the movement parameters.

We again use the conjugate Dirichlet prior for the multinomial likelihood of the transition rates to obtain their full conditional distributions and resample them using Gibbs sampling during each iteration of the MCMC algorithm.

Note, in both the old and new trajectory simulation algorithms, we sample new switching times with a Poisson process with rate λ_{max}. Though, in the old algorithm that Poisson process relates to the switching times for all n animals in the data, whilst in the new algorithm it relates to the switching times for just a single animal. Let's say we use the same value of λ_{max} in both algorithms to update the interval $[t_a, t_b]$. In the old algorithm, the expected number of switching times is $(t_b - t_a)\lambda_{max}$ and so each animal has the potential to switch $(t_b - t_a)\lambda_{max}$ times. In the new algorithm, the expected number of switching times is also $(t_b - t_a)\lambda_{max}$, but only animal i can switch state at these times. Overall though, with the new algorithm, the behaviours in the interval $[t_a, t_b]$ are simulated for all n animals during their own updates and so there will, on average, be $n(t_b - t_a)\lambda_{max}$ sampled switching times in this interval. Through these, each animal can switch state $(t_b - t_a)\lambda_{max}$ times as before, but now we have n times as many switching times.

4.2 Simulated Data Analysis

We analysed simulated data to show the theoretical capabilities of both our model and inference approach. Here, we analyse a single data set to examine the outputs of the model in detail, whilst in Section 4.4, we analyse 400 different simulations to provide insight into the robustness of the model in different scenarios.

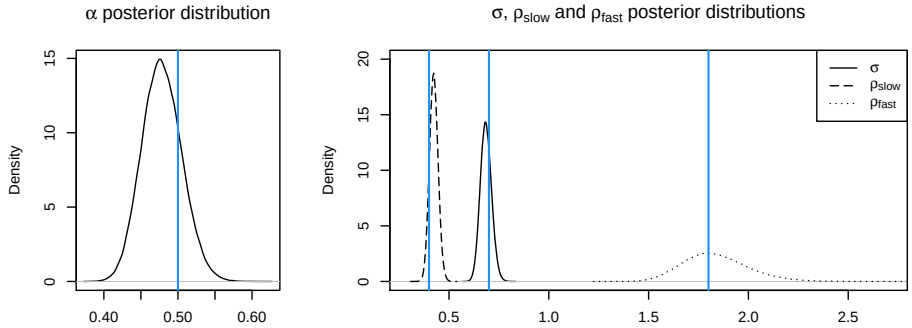
As before, the simulation consists of five dynamically interacting animals for 100 discrete-time intervals, with each interval two units of time, and we randomly deleted 10% of the data to simulate partial observations. Though this time, the 'observed' times were augmented with switching times obtained through a Poisson process as detailed in Section 4.1.1 (i.e. each animal has their own switching times). The exact parameters used to simulate this data set are provided in Table 4.1. We again incorporated two BM states for leading/independent animals to allow for different speeds of movement, meaning we have seven behavioural states in total.

To initialise the MCMC algorithm we randomised the behavioural parameters and started with over-dispersed movement parameter values. We set λ_{max} as 0.2 as that was sufficiently high for the transition rates used for the simulation and we updated sections of an animal's trajectory ranging from 3 to 12 observations long. We proposed a new trajectory 70 times per iteration of the MCMC and randomly selected the focal animal each time. We ran the algorithm for 1.4 million iterations, of which 50,000 was burn-in, and we recorded every second iteration for the movement parameters and every 20^{th} iteration for the behavioural parameters. We used an uninformative Dirichlet prior

Parameter	Value
α	0.5
σ	0.7
ρ_{slow}	0.4
ρ_{fast}	1.8

Parameter	Value
λ_{OU-OU}	0.00 – 0.04
λ_{OU-BM}	0.00 – 0.01
λ_{BM-OU}	0.00 – 0.01
λ_{BM-BM}	0.00 – 0.12

Table 4.1: Parameter values used to simulate data. The table on the left details the movement parameters whilst the table on the right details the transition rates. λ_{OU-OU} represents switching between Ornstein-Uhlenbeck (that is, subordinate) states, λ_{OU-BM} represents switching from a subordinate state to a Brownian motion state (that is, leading or independent) and similarly for λ_{BM-OU} and λ_{BM-BM}. The values of the transition rates indicate the range that the rates in each of those categories were uniformly sampled from.



Parameter	True Value	Point Estimate	Standard Deviation	Effective Sample Size
α	0.5	0.478	0.0269	4755
σ	0.7	0.685	0.0281	15075
ρ_{slow}	0.4	0.424	0.0216	10175
ρ_{fast}	1.8	1.840	0.1640	8358

Figure 4.1: Top: posterior distributions of the four movement parameters for our simulation analysis. The blue vertical line indicates the true value used. Bottom: a summary of the movement parameter results. The effective sample size is calculated using using *coda* (Plummer *et al.*, 2006).

for the transition rates.

Figure 4.1 shows the movement parameter posteriors against the true values used. All are consistent with the true value and the multivariate potential scale reduction factor (MPSRF) of the Gelman-Rubin diagnostic is 1 (calculated from two separate MCMC runs using *coda* (Plummer *et al.*, 2006)). Further details of the Gelman-Rubin diagnostic (for these and subsequent analyses in this chapter) are in Table 4.2. Due to the reasons discussed in Sections 3.7.1 and 3.7.2 it is not meaningful to compare the inferred rates with the true values used in the simulation. However, from Table 4.2, we are able to at least see that the rates converged to the same posterior distributions across different runs. The multivariate PSRF for these is 1.03.

Figure 4.2 shows our behavioural state posteriors for animal 3, along with the true states from

Movement Parameters	Min	1^{st} Quartile	Median	Mean	3^{rd} Quartile	Max
Simulation — Point Est.	1.000	1.000	1.001	1.001	1.001	1.001
Simulation — Upper C.I.	1.000	1.001	1.002	1.003	1.003	1.005
Baboon — Point Est.	1.001	1.003	1.005	1.006	1.008	1.014
Baboon — Upper C.I.	1.003	1.016	1.021	1.028	1.033	1.067

Transition Rates	Min	1^{st} Quartile	Median	Mean	3^{rd} Quartile	Max
Simulation — Point Est.	1.000	1.000	1.001	1.006	1.004	1.046
Simulation — Upper C.I.	1.000	1.001	1.003	1.011	1.009	1.101
Baboon — Point Est.	1.000	1.003	1.012	1.031	1.031	1.380
Baboon — Upper C.I.	1.000	1.005	1.046	1.123	1.104	3.226

Table 4.2: Summaries of the point estimates and upper confidence limits of the univariate potential scale reduction factors for both the simulated and baboon data analyses. The limits correspond to a coverage of 95%.

the simulation. The posterior estimates are broadly correct and confident, though quick, relatively nuanced switches as around observations 67 and 68 can be smoothed over. Through this depiction, we are able to see that this animal was strongly influenced by its peers throughout the data, particularly by animal 2 until a change of behaviour around observation 80 when it then followed animal 5.

The uninformed simulation analysis in Chapter 3 (Section 3.7.2) took approximately 26 hours to run 1 million iterations whilst the above analysis with the new inference algorithm took approximately 36 hours to run 1.4 million iterations — which equates to a similar number of iterations per hour. Whilst we cannot directly compare the old and new algorithms as the data sets analysed are different, at first glance it appears that the model fitting undertaken with the old algorithm produced effective samples at a faster rate (roughly twice as fast). However, the fundamental issue with the old algorithm is the requirement of collective updates and the extent to which they are limited by the number of animals in the data set. Recall that collective updates are required to resample switching times and that these were carried out over only 2 to 3 observations for a group of 5 animals in order to obtain a decent acceptance rate. Increasing the number of animals increases the challenge of simulating an acceptable trajectory for the entire group, which in turn decreases the length of data we can reasonably simulate over, and so there is not a lot of room to study a larger group. The new algorithm does not suffer such a drastic limit as we are able to carry out trajectory updates (that resample both the times and behaviours) up to 12 observations long with ease for an analysis containing the same number of animals. So whilst we would still likely have to decrease the length of those simulations when analysing a greater number of animals, we have more capability to do so. Overall, moving from simultaneous to sequential switching has increased the computational effort of fitting the model to data, but we have greatly reduced the biggest limitation of the previous approach.

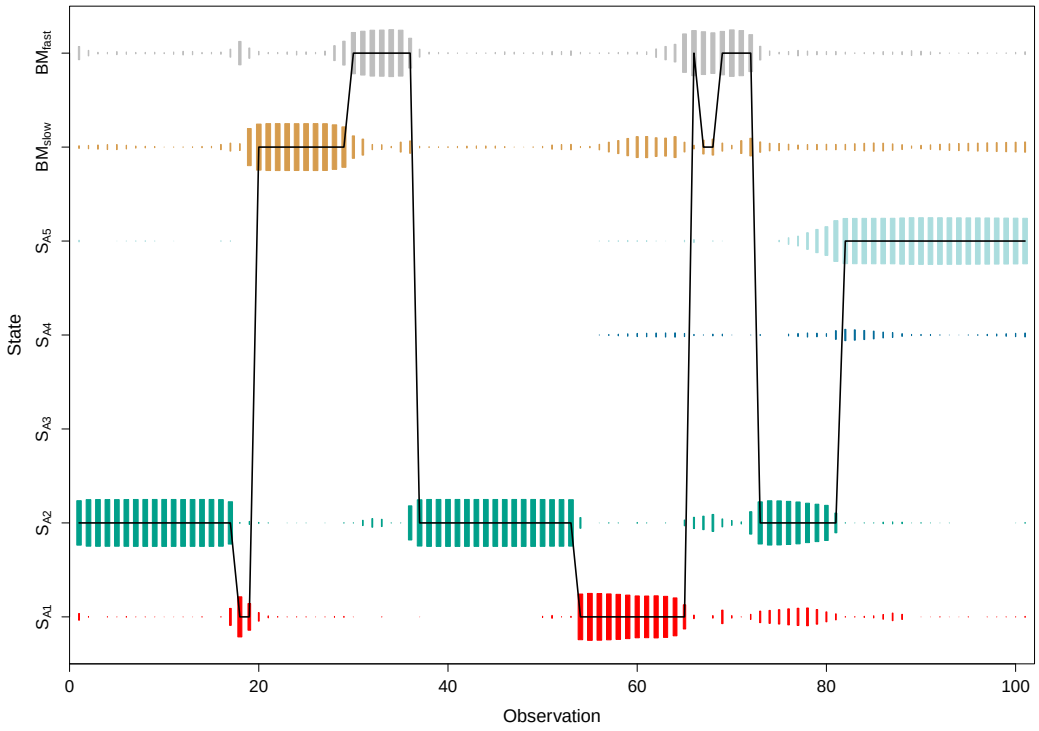


Figure 4.2: The state posterior distribution for animal 3 in the simulation data. There are seven states: the two Brownian motion speeds and five subordinate behaviours where state S_{Ai} indicates attraction to animal i. The area of each box represents the posterior probability of being in that state at that observation, from 0 to 1. The black line is true state in the simulated data.

4.3 Baboon Data Analysis

To test our new method 'in the field', we have taken a subset of wild olive baboon data that was originally analysed by Strandburg-Peshkin *et al.* (2015), which is available on Movebank (Crofoot *et al.*, 2015). The GPS data was collected at the Mpala Research Centre in Kenya for 26 baboons in a single troop. The data was recorded at a frequency of 1Hz, for 12 hours a day (06:00 – 18:00) over 30 days. We took a subset of this data for five baboons (ID's 3, 4, 5, 11, 9) for 15 minutes (899 observations) to act as a test for the model. We chose this time period to contain some directional conflict as in movie S2 in the supplementary materials of Strandburg-Peshkin *et al.* (2015). We converted the GPS coordinates to UTM zone 37N easting-northing using *sp* (Pebesma & Bivand, 2005; Bivand *et al.*, 2013).

For similar reasons to the zebra data analysis in Chapter 3, we have included two BM states for leading or independent animals. The MCMC algorithm was initialised as in Section 4.2. After some tuning runs, λ_{max} was again set as 0.2, which seems sufficient given the resulting transition rates. Updates of an animal's trajectory are performed over lengths of 3 to 40 observations and we proposed 210 trajectories in each iteration. The MCMC ran for 500,000 iterations, of which

	Independent (%)	Leading (%)	Following (%)
baboon 3	47.2	37.2	15.7
baboon 4	54.1	10.0	35.9
baboon 5	20.4	38.0	41.6
baboon 11	28.7	4.0	67.3
baboon 9	47.7	21.4	30.9

Table 4.3: The percentage of observations each baboon spent in each role based on the modal state.

	baboon 3 (%)	baboon 4 (%)	baboon 5 (%)	baboon 11 (%)	baboon 9 (%)
baboon 3	-	0.2	7.8	0.1	7.2
baboon 4	7.8	-	18.4	6.5	3.0
baboon 5	28.5	10.0	-	1.3	1.9
baboon 11	16.1	0.0	34.6	-	15.9
baboon 9	13.9	0.0	9.8	7.2	-

Table 4.4: The percentage of observations each baboon is subordinate to another based on the modal state. Cell ij in the table corresponds to baboon i being subordinate baboon j.

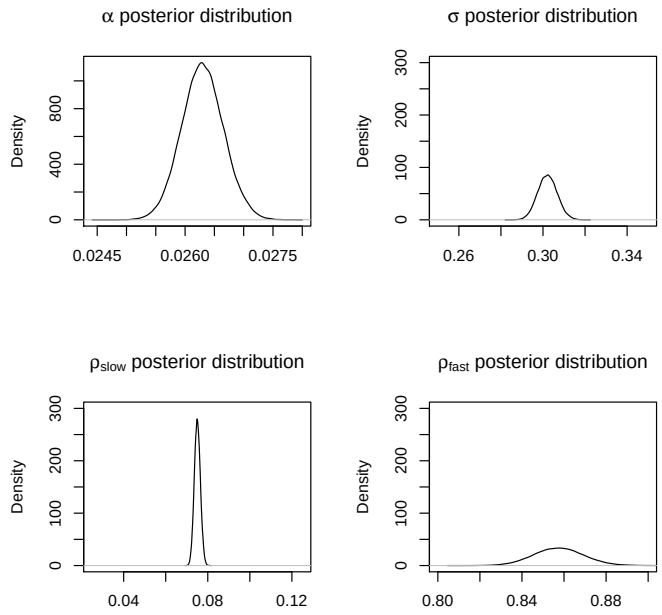
100,000 was burn-in, and we recorded every second iteration for the movement parameters and every 20^{th} iteration for the behavioural parameters. An uninformative Dirichlet prior was used for the transition rates.

Figure 4.3 shows the posterior distributions for the movement parameters. The posterior for α shows strong evidence that there is indeed interaction between the five baboons in the data analysed. The Gelman-Rubin diagnostic is again used to check convergence over two MCMC runs and the multivariate potential scale reduction factor of the movement parameters is 1.02.

The state posteriors of baboon 5 are shown in Figure 4.4, through which we can observe the dynamics of this baboon's social behaviour in a subordinate sense. That is, it mostly alternates being attracted to baboons 3 and 4, though there is some uncertainty as to which baboon it is subordinate to in the last 200 observations. We can obtain a clearer picture of the dynamics of an animal's influence on the group by looking at the role posteriors as shown in Figure 4.11a (for baboon 9 in this case). For this baboon, we estimate it largely interacts as a subordinate until around observation 700, which is when it takes on a consistent leadership role. Looking at the data, this time corresponds to a change in the direction that the baboons are moving in.

These results also allow us to see the long-term manner in which the animals interact, both in the sense of their role (Table 4.3) and with each other (Table 4.4). Though both of these tables only show us a static overview of the social interaction, they highlight which animals seem to have persistently high levels of influence in the group (baboons 3 and 5) and potentially strong bonds within it (baboons 3, 5 and 11 for example).

It is worth noting that these baboon runs have not completely converged as indicated by the



Parameter	Point Estimate	Standard Deviation	Effective Sample Size
α	0.0263	0.000359	1033
σ	0.3020	0.004620	1479
ρ_{slow}	0.0751	0.001430	1620
ρ_{fast}	0.8580	0.011900	5625

Figure 4.3: Top: posteriors distributions of the four movement parameters for the baboon data analysis. Note the different scales of the *Density* axes. Bottom: a summary of the movement parameter inference. The effective sample size is calculated using using *coda* (Plummer *et al.*, 2006).

Gelman-Rubin diagnostic for some of the transition rates (Table 4.2) and the corresponding trace plots (Figure 4.6). The trace plots for the majority of the transition rates (and all of the movement parameters — Figure 4.5) indicate the chains quickly moved past the burn-in period and converged. However, this is not the case for a small number of rates — see the rate of switching from S_{B11} to BM_{slow} in Figure 4.6 for example. Similarly, whilst the univariate PSRF point estimate for the vast majority of rates was less than 1.04, the highest was 1.38 — resulting in a multivariate PSRF of 1.4. It may be we could have aided the inference of the rates by increasing the number of trajectory updates per iteration. Improving the mixing of the behaviours will inherently improve the mixing of the transition rates. Or we simply could have run a greater number of iterations. Either way, exact inference by means of simulating when behavioural switches occur is a computationally costly task. The results described above were obtained over approximately 2.8 days and so we have not yet overcome the computational hurdle that can come with continuous-time models, particular ones with a complex state space. Doing so would not only enable quicker convergence, but facilitate the analyse of larger data sets (both in terms of the number of animals and the time period) and

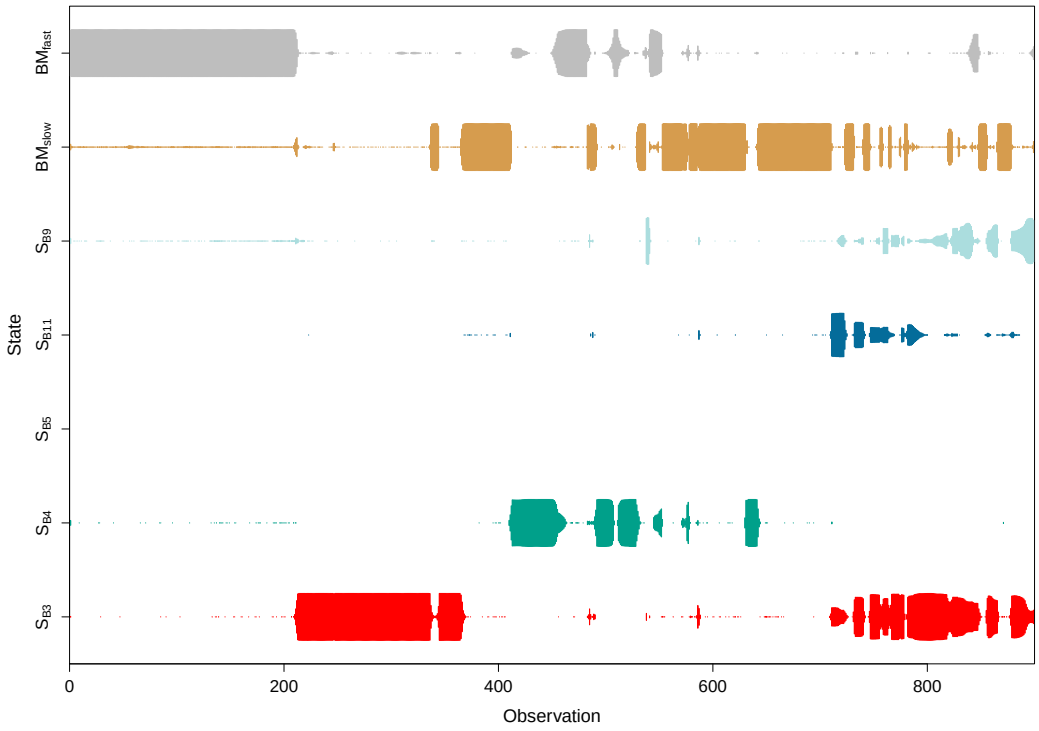


Figure 4.4: The state posterior distribution for baboon 5. There are seven states: the two Brownian motion speeds and five subordinate behaviours where state S_{Bi} indicates attraction to baboon i. The area of each box represents the posterior probability of being in that state at that observation, from 0 to 1.

investigation of more long-term and biological questions.

This data set has previously been analysed in a number of studies (Strandburg-Peshkin *et al.*, 2015; Farine *et al.*, 2016, 2017; Strandburg-Peshkin *et al.*, 2017). It was originally analysed by Strandburg-Peshkin *et al.* (2015) to investigate how a troop of baboons collectively make their movement decisions. 'Movement initiations' were extracted from the data of 25 baboons through a method based on minima and maxima distances between a pair of animals. Through this approach, Strandburg-Peshkin *et al.* quantified the probability of a baboon following a movement initiation in the context of the number of initiators and their consensus in direction — situation-based covariates that would not be trivial to implement in our model.

Though, whilst their interaction assumption of a following animal moving towards the initiator is conceptually similar to ours, it is perhaps more restrictive. Our OU approach models the subordinate's movement as being distributed around the location of its dominant. As a result, we do not have to constrain our definition of influence to being the 'initiator' or the animal at the front of the group (Strandburg-Peshkin *et al.*, 2015) for example. Along with the state switching we have defined, this will also help to smooth out erroneous interactions in the data.

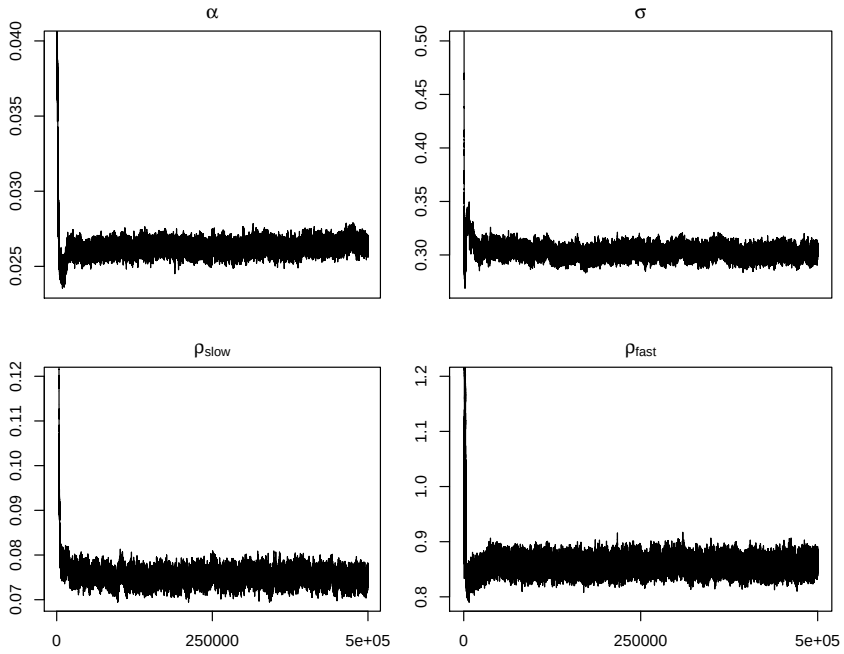


Figure 4.5: Trace plots for the four movement parameters from one of the model fittings to the baboon data.

This data set was subsequently analysed to investigate how cohesion is then maintained during movement (Farine *et al.*, 2016). The movement of a baboon was best predicted when the locations of 4 to 6 of its peers were taken into account — though this is open to individual variation (Farine *et al.*, 2017). In the short term (approximately less than 10 minutes), those 4 to 6 animals consisted of local neighbours. Beyond 10 minutes, they consisted of an animals top 'spatial affiliates' — those with whom it is most often the nearest neighbour with. Distilling interaction down to the most causal, as we do, discards the opportunity to allow movement to be informed (or indeed, predicted) by several others. However, this spatial affiliate approach does not take into account any changes in social behaviour as affiliation is calculated over the entire study time frame. Whilst that might be a reasonable approach to take for a stable social animal such as baboons, it might not be in other cases.

4.4 Reliability

Whilst the above results are encouraging, there is scope for the model to infer false positives (interaction where there is none) and false negatives (no interaction where there is some). To investigate this, we have analysed 400 different simulations derived from four different sets of parameters — 100 from each set. The parameters of set 1 are identical to those used to create the simulated data in Section 4.2; set 2 has been derived to simulate similar movement and social behaviours as inferred from the baboon data; set 3's parameters were chosen to represent different

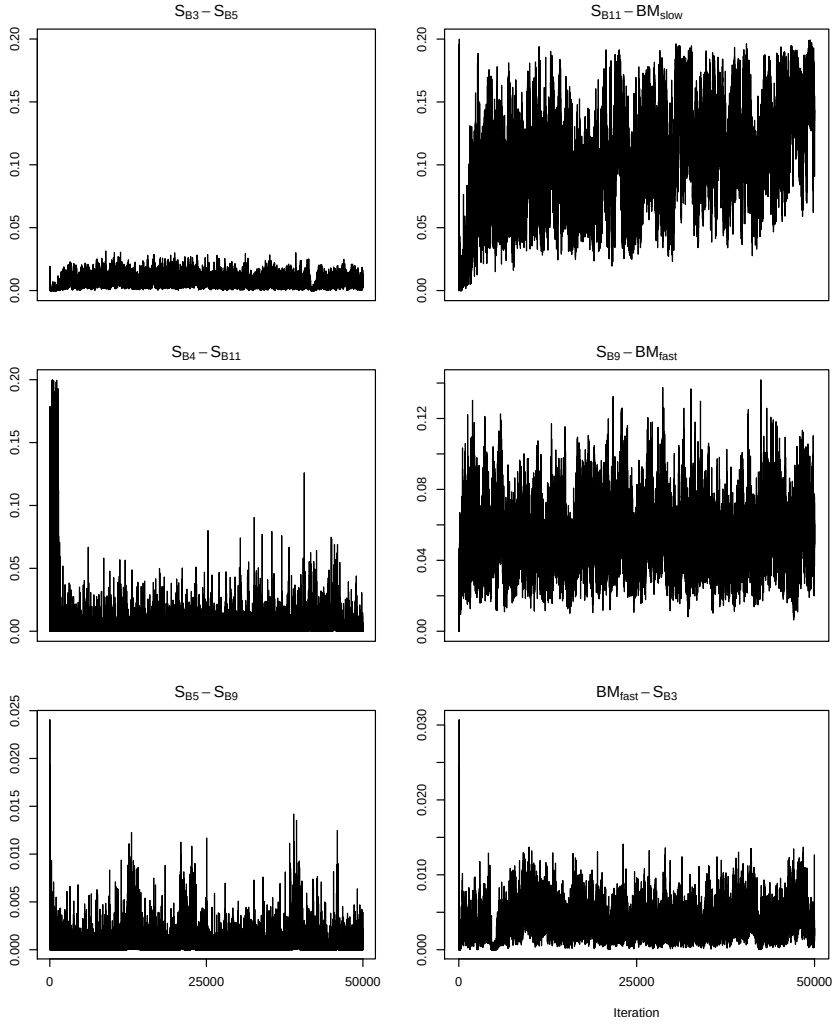


Figure 4.6: Trace plots for a sample of transition rates from one of the model fittings to the baboon data.

behaviours from those of sets 1 and 2. For instance, the three noise parameters are less distinct than in the other sets to provide a tougher inference scenario (see Table 4.5). Set 4 consists of animals solely moving in Brownian motion in order to analyse the rate at which the model introduces false positives. Whilst the movement parameters and Λ were kept constant across all simulations for that parameter set, the initial behavioural states at time 0 were sampled from the stationary distribution of Λ in order for each data set to contain different behaviours. All simulations allow for two BM states and consist of 100 movement steps.

Generally, each run was initialised much like those in Section 4.2. Each set was treated to its own movement parameter proposal distributions in order to encourage good mixing, with the same configuration being used for all runs corresponding to that set. Additionally, to encourage false positives in set 4, we initialised α to be low on the assumption that any false positives inferred would point to weak attraction. Each run was carried out for 1 million iterations with a burn-in

Parameter Set	α	σ	ρ_{slow}	ρ_{fast}
1	0.5	0.7	0.4	1.8
2	0.0266	0.297	0.0751	0.855
3	0.2	0.7	0.5	1.4
4			0.4	1.8

Parameter Set	λ_{OU-OU}	λ_{OU-BM}	λ_{BM-OU}	λ_{BM-BM}
1	0.000 – 0.040	0.000 – 0.010	0.000 – 0.010	0.000 – 0.120
2	0.000 – 0.031	0.001 – 0.087	0.000 – 0.019	0.000 – 0.028
3	0.000 – 0.015	0.000 – 0.015	0.000 – 0.015	0.000 – 0.020
4	0.000 – 0.000	0.000 – 0.000	0.000 – 0.000	0.000 – 0.075

Table 4.5: The four sets of values used to simulate data. For sets 1, 3 and 4, the values of the transition rates indicate the range that the rates in each of those categories were uniformly sampled from. The values for set 2 were chosen (within the above ranges) to simulate similar behaviours inferred from the baboon data.

of 500,000 to be confident we had indeed surpassed the actual burn-in period without manually checking each output. For these runs, we only recorded every fifth iteration for the movement parameters and every 50^{th} iteration for the behavioural parameters in order to keep the total size of the output files manageable.

In order to evaluate the rate of false positives and false negatives in interaction, we use our state estimations of the whole group to calculate the probability of two animals interacting at a given time, whether directly or indirectly and regardless of dominance and subordination ordering. This probability is an *interaction posterior*. At times of true interaction in the simulations, we would expect the *interaction posterior* to be close to 1; at times of true non-interaction, we would expect the complimentary *non-interaction posterior* to be close to 1. The CDFs of these posteriors for each run are plotted in Figure 4.7.

The CDFs relating to sets 1 and 3 form the desired curve, though those of set 3 display more uncertainty. This is intuitive as α is smaller in set 3 compared to set 1 (that is, the attraction is weaker) and the noise parameters were set to provide more of a challenge. The anomalous result in the *non-interaction posteriors* for set 3 comes from a data set where there is very little non-interaction — each pair of animals don't interact for only approximately 8.7% of the data on average. Uncertain state estimations during some of those segments of non-interaction are a blot on otherwise reasonable results for that data set. With regards to set 4, there are no plots concerning true interaction as there isn't any in those simulations — a feature confidently estimated in the *non-interaction posteriors*. However, there are clearly two erroneous results. The most spurious of those is the consequence of not tuning each simulation separately and the resulting acceptance rate of the movement parameters was extremely low (approximately 1%). Whilst this is clearly not a desirable outcome, ordinarily we would be able to tune the MCMC differently to navigate this issue; as it is, the problem is evident from the most cursory diagnostics, so would not mislead

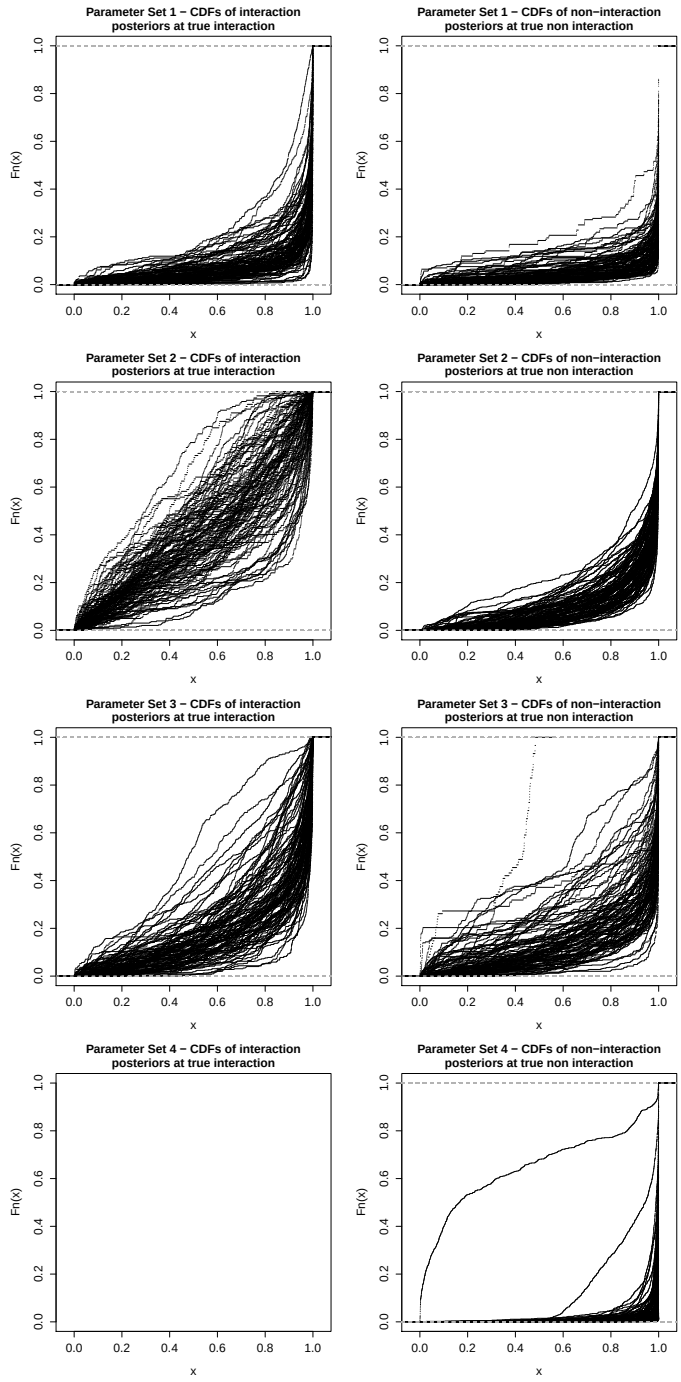


Figure 4.7: The plots in the left column contain the CDFs of the *interaction posteriors* at times of true interaction. The plots in the right column contain the CDFs of the *non-interaction posteriors* at times of true non-interaction. Each row of plots corresponds to a parameter set and each CDF is derived from the posteriors for all pair combinations during a single run.

in an actual analysis. The second erroneous result simply inferred some interaction. A small number of segments of the data are estimated, with a large degree of uncertainty, to contain weak interaction.

Figure 4.7 indicates a greater degree of uncertainty in the *interaction posteriors* of set 2 compared to those of sets 1 and 3. In part, this effect will be due to set 2's simulations containing comparatively little interaction and so false negatives are more pronounced in their *interaction posterior* CDFs than those of sets 1 and 3. Similarly, they will also be more pronounced than false positives in set 2's corresponding *non-interaction posterior* CDFs. More importantly though, this uncertainty highlights a limitation of modelling the movement of leading/independent animals only through Brownian motion. The difference in certainty of the state posteriors in Section 4.3 and those of the simulations derived from Section 4.3's results will in part be due to there being some feature of the real movement that is not captured by BM, such as persistence. Without that feature, for example, independent animals and subgroups in the simulations will frequently overlap movement paths. It therefore becomes a challenge to identify the exact interactions taking place. This produces scenarios in the simulations that are perhaps not biologically relevant and another process, such as Ornstein-Uhlenbeck, may be better suited to model the movement of the leading and independent animals.

We can examine this uncertainty further with Figures 4.8 and 4.9. The *role posteriors* compared against the true role show that we are generally certain whether an animal is a subordinate or not at any given time — though there are some instances of independence inferred at times of true subordination. The consistency of the movement parameter posteriors with the true values are further evidence of this reliability. However, there is some uncertainty whilst ascertaining exactly which animal a subordinate is following. Depending on the resulting social structure, a number of misclassifications might occur. For example, the true dominant might now be estimated to be independent instead of leading and a true independent might now be estimated to be the subordinates leader. The effect of this uncertainty is more noticeable in the CDFs relating to true leadership as there is comparatively little true leadership in the data. This scenario also ties in with the simulations containing jumbled subgroups as discussed above. Overall though, these simulation results display a good level of robustness.

4.5 Data Thinning

Alongside our model being slow to fit to data, a recurring question in animal tracking studies is: at what frequency should data be collected at for the analysis at hand (Hughey *et al.*, 2018; Scharf & Buderman, 2020; Williams *et al.*, 2020)? We have therefore experimented with thinning the same baboon data analysed above (henceforth referenced as the 'full analysis') to investigate how the inference speed can be improved by fitting the model to a thinned data set, and how the results compare to the full analysis. We experimented by thinning the data by a factor of five and of 20.

We will also discuss our choices of λ_{max} for these experiments. $\lambda_{max}\delta t$ is the mean number of potential switching points sampled from the Poisson process between sequential observations over

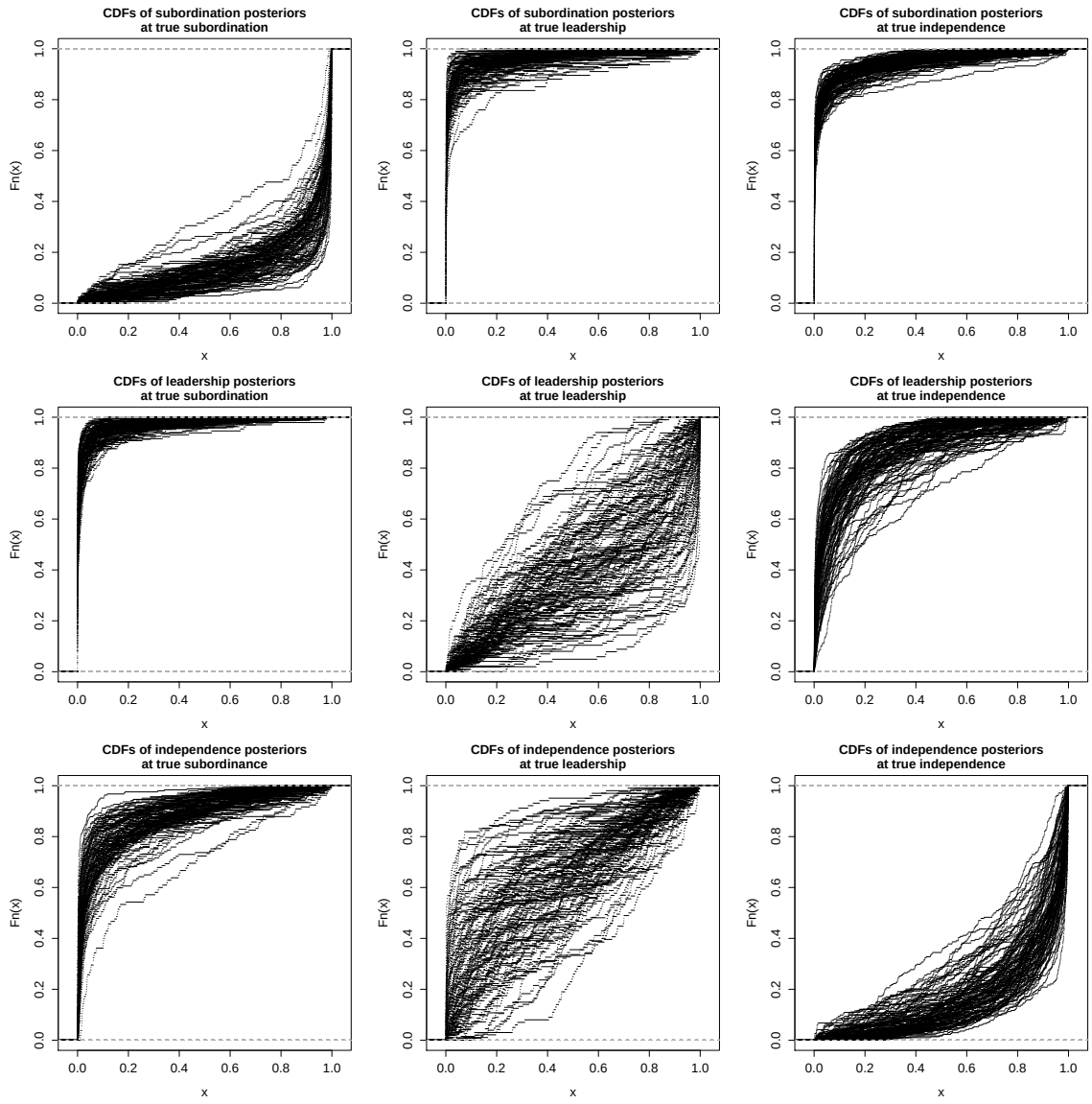


Figure 4.8: CDFs of the *role posteriors* compared to the true role in the simulated data. Each CDF is derived from the posteriors of all animals of a single run. All graphs relate to parameter set 2. The graphs on the diagonal contain the posteriors of being in the correct role and so they are expected to be close to 1. The off-diagonal graphs contain the posteriors of being in an incorrect role and so they expected to be close to 0.

a time span of δt. Whilst we have stated previously that $\lambda_{max} \geq \max(\lambda_u)$ to ensure all state switches can be sampled appropriately, if $\lambda_{max}\delta t$ is high, say five, there is little information to be gained from sampling that many state switches between sequential observations. Reducing λ_{max} will reduce the size of the augmented data set and therefore the number of computations needed.

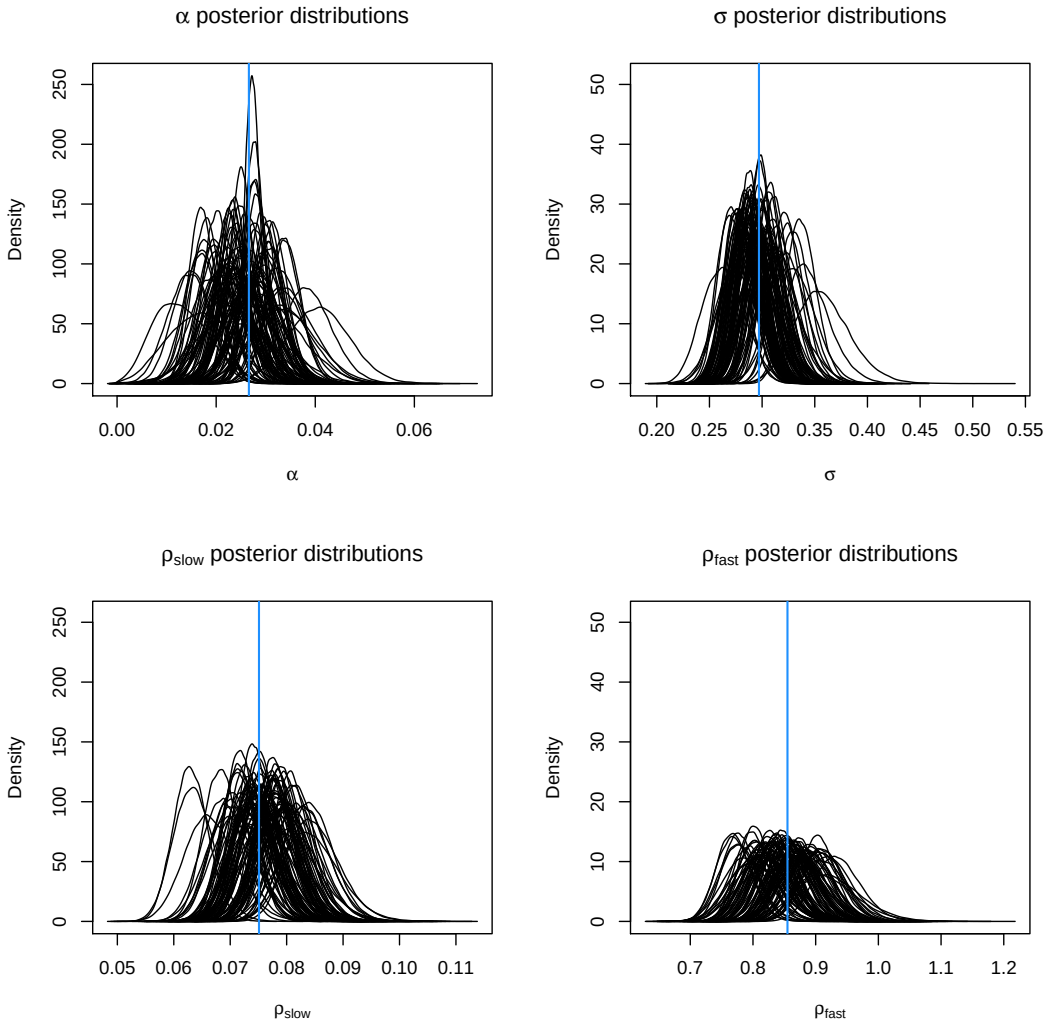


Figure 4.9: Posterior distributions of the four movement parameters. Each line is the posterior for a single run and they all relate to parameter set 2. The blue horizontal line indicates the true value used for the simulations. Note the different scales of the *Density* axes.

4.5.1 Thinning by a Factor of Five

We took every fifth observation of the baboon data detailed above, leaving us with 180 observations (at 0.2Hz) instead of 899 (at 1Hz). We ran the inference again for 500K iterations to provide a comparable analysis. λ_{max} was kept at 0.2 and so $\lambda_{max}\delta t = 1$ for sequential observations.

As a result of thinning the data, the uncertainty of the movement parameter posteriors increased as can be seen in Figure 4.10. We also see from those plots that the posteriors shift. The differences in the estimations will most likely be for two reasons. Firstly, any discrepancies in state estimations between the two analyses and secondly, the model we have chosen to fit may not be sufficient to represent some of the movement behaviours. That is, we chose to model the movement of leading/independent animals through Brownian motion. However, if there is some persistence in

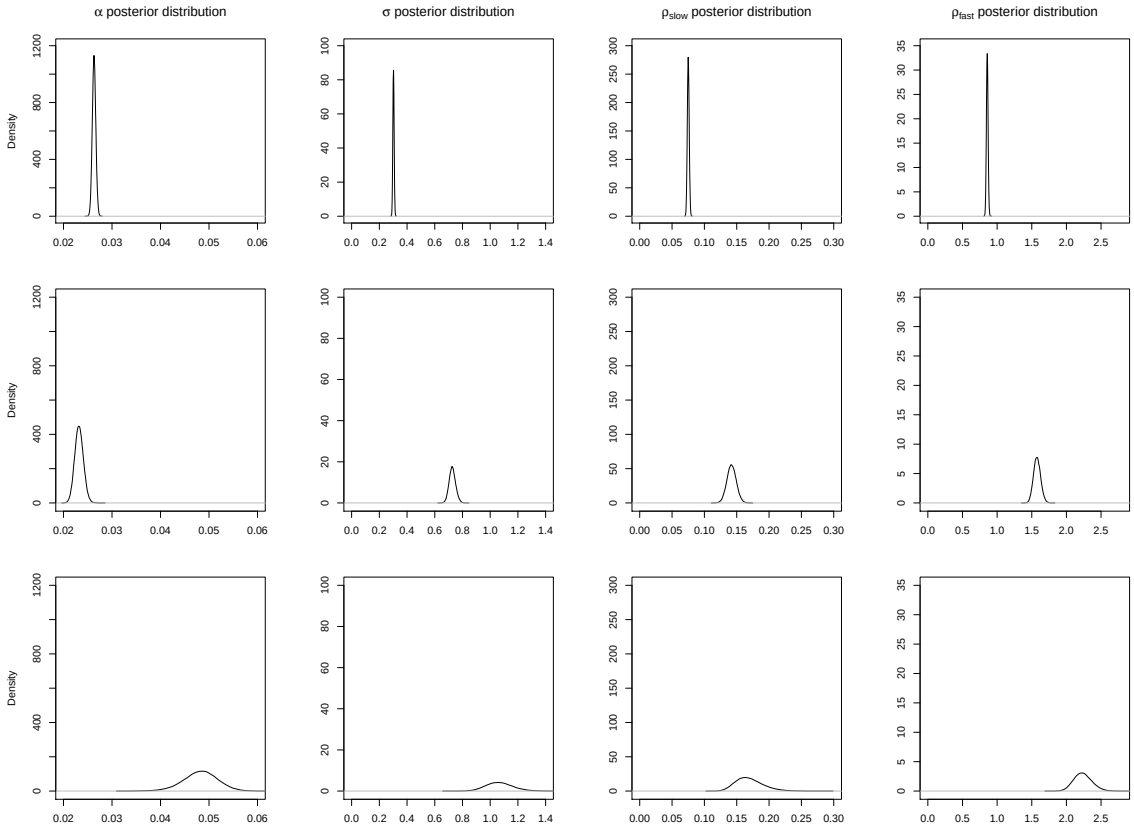


Figure 4.10: Posterior distributions of the four movement parameters for the full analysis (top row), thinned by a factor of five (middle) and thinned by a factor of 20 (bottom).

their movement in reality, the noise coefficient of BM will not remain constant as the data is thinned.

To illustrate this, say an animal is moving in one dimension through three observations and each of the two intervals span time δt. Let's denote the movement through the intervals as m_1 and m_2. If there is no persistence in the animal's movement, the correlation between m_1 and m_2 is 0. Therefore, if we removed the middle observation, the variance of the movement over time $2\delta t$ is: $\text{Var}(m_1 + m_2) = \text{Var}(m_1) + \text{Var}(m_2) = 2\text{Var}(m_1)$ as $\text{Var}(m_1) = \text{Var}(m_2)$. Thus, the diffusion coefficient of the BM process will remain constant regardless of the data thinning. However, if in the true process there is some persistence, then $\text{Corr}(m_1, m_2) = c$ where $c > 0$. Removing the middle observation then gives: $\text{Var}(m_1 + m_2) = \text{Var}(m_1) + \text{Var}(m_2) + 2\text{Cov}(m_1, m_2) = 2\text{Var}(m_1) + \frac{2c}{\text{Var}(m_1)\text{Var}(m_2)}$. The variance of $m_1 + m_2$ in the persistent case is therefore greater than that in the non-persistent case as all $c, \text{var}(m_1), \text{var}(m_1) > 0$. Thus, the estimated diffusion coefficient of the BM process will increase as we thin the data. Section 3.5.2 detailed how this model may be extended to capture any persistence.

When comparing the role posteriors of baboon 9, this thinned analysis (Figure 4.11b) smooths

over the state estimations from the full analysis (Figure 4.11a). To investigate this smoothing, we looked at the modal state estimations at the 1Hz observation times in both analyses and quantified the periods of behaviour that were present in the full but absent from the thinned analysis. To classify a period of behaviour as being smoothed over, we require that at least 80% of that period to have a different modal state estimated in the thinned analysis. Or rather, we require at least 20% of that period to have the same modal state estimated in the thinned analysis in order for it to be accounted for. The 80% threshold is arbitrary, but it prevents a period from being accounted for in the thinned analysis by an overly short visit to that state whilst still capturing significant smoothing. Figure 4.12 shows that the smoothed-over periods are predominately small in length — mostly under five seconds which is our new temporal scale.

This smoothing is likely to be a result of our assumption that the animal's behaviour is Markovian. If the full data suggests the short state transitions as seen in Figures 4.4 and 4.11a, the Markovian behavioural process will readily accommodate them. The constant transition rates incorporate no presumption against very short visits, as reflected in the exponentially distributed holding times. Through thinning the data, the 'evidence' of these transitions is no longer present, though they may remain with much lower probability.

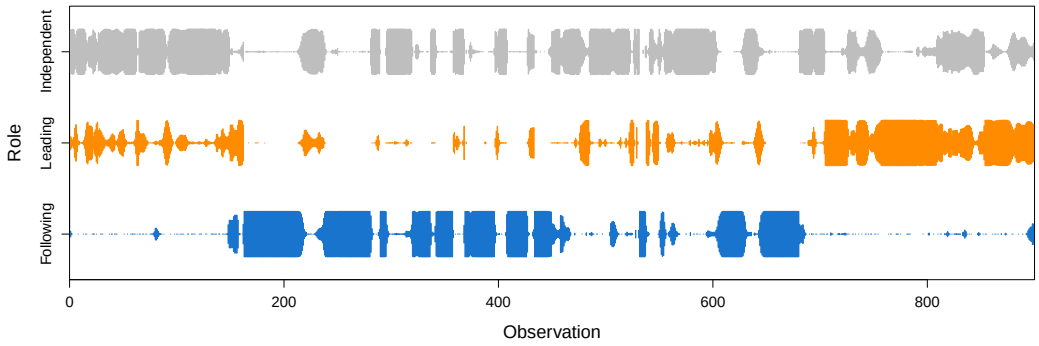
4.5.2 Thinning by a Factor of 20

We took every 20^{th} observation of the data set used for the full analysis, leaving us with 45 observations (at 0.05Hz). We ran the inference again for 500K iterations but for this analysis we reduced λ_{max} to 0.05. As discussed above, the motivation for this was to keep $\lambda_{max}\delta t$ small.

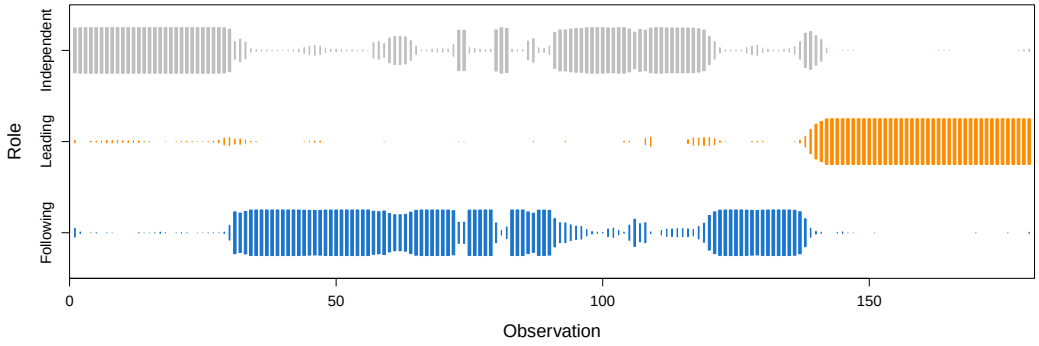
Whilst this amount of thinning can be thought of as extreme smoothing, where we'd expect to get results largely in agreement to those of the full analysis, Figure 4.11c shows the state/role posteriors can be drastically different. For example, the thinned data no longer supports the two periods of subordination that baboon 9 undertakes between observations 600-700 in Figure 4.11a. Thus, the resulting social structure has a different picture. These different behavioural estimations then have further impacts on the movement parameter posterior distributions (Figure 4.10).

In the full analysis, $\max(\lambda_u) = 0.153$ and so we wouldn't expect an animal to switch state much more frequently than every 7 seconds. We therefore suspect thinning the data to 0.2Hz keeps enough information in order to estimate similar behaviours. Thinning the data further, and reducing λ_{max}, reduces the scope of social behaviours that can be captured. That is, short-term interactions can't be seen and are unlikely to be sampled. Whilst the thinning-by-20 results aren't wrong per se, they offer a different temporal perspective of the social behaviours of the baboons. These thinning experiments backup the view that the results of an analysis will be most informative (and efficient) if the temporal resolution of the data collected is informed by the nature of the behaviours that are to be investigated (Pinter-Wollman *et al.*, 2013).

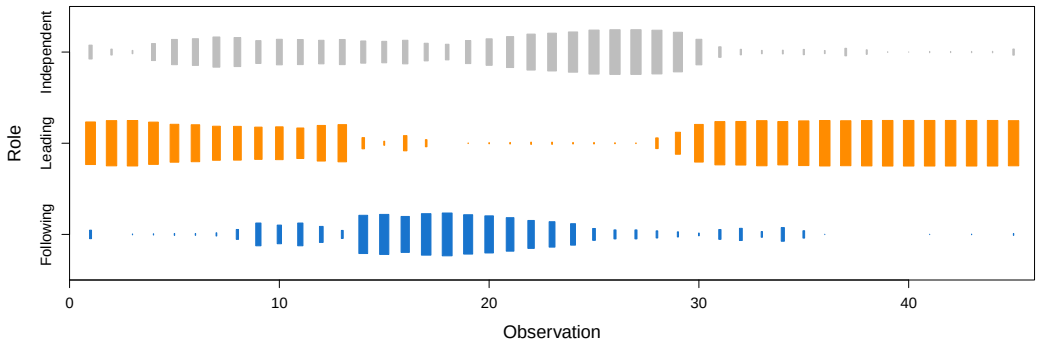
With regards to the speed of the inference, the thinned-by-five analysis was completed in approx-



(a)



(b)



(c)

Figure 4.11: The role posterior distribution for baboon 9 for the full analysis (a), thinned by a factor of five (b) and thinned by a factor of 20 (c). The area of each box represents the posterior probability of being in that role at that observation, from 0 to 1.

imately 24 hours, as opposed to approximately 67 hours for the full analysis. Whilst the number of observations was reduced to a fifth, the number of simulated data was equivalent as we kept the same λ_{max}. Though, there is scope to use a smaller λ_{max} during the thinned-by-five analysis as the resulting $\max(\lambda_u)$ was 0.111. The thinned-by-20 analysis was completed in approximately 7 hours.

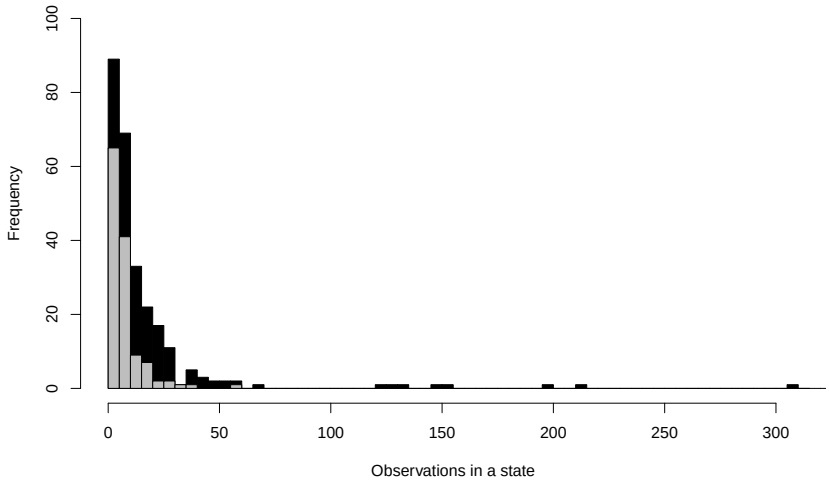


Figure 4.12: A histogram of the number of observations in a state before switching in the full analysis (black) and which of these periods were smoothed over in the thinned analysis (grey).

4.6 Comparison of Methods

To compare our approach with other methods, we have examined our results from both the simulation and full baboon analyses alongside those obtained when applying dyadic metrics (Long *et al.*, 2014; Joo *et al.*, 2018) to the same data. Much like our approach, dyadic metrics utilise movement data from multiple animals to investigate any interdependence and in turn better understand their collective behaviours. As their name suggests, the dyadic metrics are built to analyse the interdependence of two animals.

The metrics we have chosen to implement are *proximity* and *dynamic interaction* using the *wildlifeDI* R package from Long *et al.* (2014). Joo *et al.* (2018) indicate each metric offers an informative view into a specific element of interaction. *Proximity* evaluates whether two animals are within a distance threshold defined by the user; *dynamic interaction*, when split into its *displacement* and *direction* components, offers insight into whether two animals are moving at a similar speed or in a similar orientation respectively. Furthermore, whilst most other metrics are restricted to producing a single output to define the interaction over the whole data set, the ones we chosen have 'local' variations that allow us to evaluate the interaction throughout time.

In order to compare results, we use the same *interaction posterior* approach that we used in Section 4.4 as that is comparable to the interpretation of the dyadic metrics. This is done for every data point in which both animals are simultaneously observed as this is a necessary criterion for the metrics. The scaling parameter of *dynamic interaction in displacement*, β in Joo *et al.* (2018), is 1.

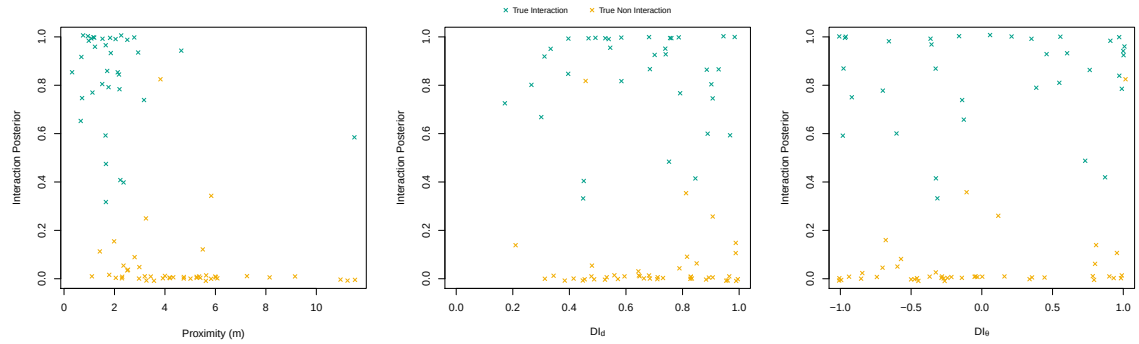
Figure 4.13a plots our *interaction posterior* against each of the three metrics for animals 1 and 3. These animals were chosen as ideal candidates to compare methods as they undertook periods of both direct and indirect interaction. Whilst the *interaction posterior* is largely (and correctly) concentrated at 0 and 1, the *displacement* and *direction* metrics are uniform across their ranges regardless of whether there was true interaction or not. *Proximity* fares better as most simultaneous data less than 2m apart correspond to true interaction and most simultaneous data greater than 4m apart correspond to true non-interaction. However, the distances in between highlight the difficulty in determining interaction from *proximity* alone.

Figure 4.13b displays the same comparison for baboons 4 and 9. Whilst the *interaction posterior* is again largely concentrated towards 0 and 1, the *proximity* metric does not offer a discernible pattern. It is uninformative when the *interaction posterior* is close to 0 and interaction appears to be grouped at distinct ranges of proximity — at approximately 60m and less than 20m. For both *displacement* and *direction*, there is moderate concentration at $(1, 1)$ — indicating a consensus between our model and these metrics on some moments of interaction. However, both metrics are quite uniform when the *interaction posterior* is certain there is no interaction. In particular, *direction* is just as likely to suggest interaction as it is non-interaction.

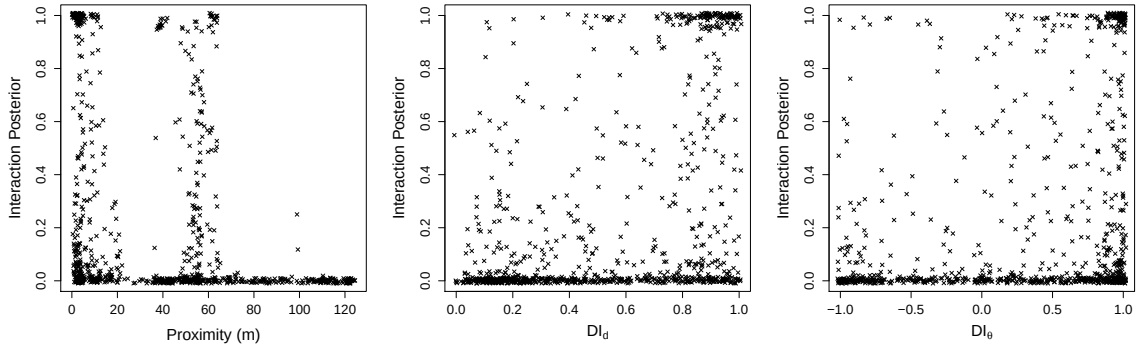
Overall, there is some consistency of the metrics with our model in estimating when two animals interact. However, when the animals are not interacting (either known from the simulated data or estimated from our model), the dyadic metrics are generally not in agreement and are fairly uniform across their ranges. We suspect this is because the metrics can have quite relaxed definitions of interaction. For instance, *dynamic interaction* can hint at interaction without any concern as to whether the animals are reasonably proximate to one another. That is in contrast to our perhaps more defined notion of interaction where we are not likely to introduce false positives (see Section 4.4). Though, that definition comes with its own limitations such as not being able to capture co-movement. From a practical point of view, the dyadic metrics are much faster to run than our approach and they will require less tuning. However, our model can more naturally handle incomplete or unsynchronised data. The metrics require essentially simultaneous data to evaluate the cohesion of two animals at a given time whereas our model is able provide an estimate of interaction in continuous time across the temporal range of the data. Furthermore, our model has the capability to jointly analyse the data of larger social groups.

4.7 Discussion

The new trajectory proposal algorithm provides a huge improvement in performance compared with the one developed in Chapter 3. We initially chose, in Chapter 3, to allow all animals the possibility of switching state at each switching time in order to reduce the number of data points required. This decision proved to be a false economy. The previous algorithm required a two-step process where, in effect, collective updates resampled the switching times whilst individual updates



(a)



(b)

Figure 4.13: The *interaction posteriors* (the posterior probability of two animals being in the same subgroup) from the simulation analysis (Section 4.2) are plotted against the *proximity*, *dynamic interaction in displacement* (DI_d) and *dynamic interaction in direction* (DI_θ) dyadic metrics for the same data set (a). The results used are for animals 1 and 3 and each point corresponds to a simultaneous observation. Similarly, the same approach has been taken for baboons 4 and 9 from Section 4.3 (b). Both axes have been jittered in order to help display the density of the points.

resampled the state switches at those times. Whilst the individual updates improved the mixing of the behaviours, the whole process was slow as it hinged on the sluggish collective ones. In the new approach, through the restriction of only one animal being allowed to switch state at each switching time, a single update is now able to resample both the behaviours and the times at which they occur for a single animal. Thus, we can now propose new trajectories over sizeable segments of data and still obtain a good acceptance rate. So whilst the new method requires a greater number of switching times than in the previous method in order to capture the same behaviours, it facilitates much better mixing.

With the new algorithm, we were able to examine the robustness of our model through 400 model fittings, a task that would not have been desirable with the previous algorithm. Whilst this analysis highlighted the limitations of modelling self-driven animals with a BM process, the model was proven to be highly reliable in limiting the amount of false positive and false negative interactions

estimated. We were also able to fit the model to high-frequency baboon data, if only for a short segment of it. Through this analysis, we are able to see clear evidence of interaction between the baboons during their movement and obtain both a dynamic and static view of their social behaviours. The dynamic view enables us to see how the social behaviour evolves over time in terms of specific relationships, the role of individuals and the overall structure. The static view can help pin point long-term influential animals and bonds within the group.

However, despite the improvements in this chapter, a considerable amount of time is still required to explore such a complex state space. The current method is still naive, as we simulate forwards through the data whilst ignoring the end state of the trajectory. One potential route to remedy this is to explore techniques based on the Forward Algorithm (widely used for efficient implementation of hidden Markov models) as proposed by Blackwell (2018).

In order to ease the computational burden of the model fitting process, we can scale the frequency of the data to be analysed to the nature of the behaviours under investigation. The scaling could either take place at the data collection stage or through data thinning. Our data thinning investigation in Section 4.5 showed that sizeable computational gains can be made whilst still being able to infer the social and movement behaviours at the desired temporal scale. However, this analysis also showed that our assumption that the behavioural process is Markovian may not be biologically sound. An alternative approach may be to model the behavioural process as semi-Markovian in order for the times spent in a state to be more realistically distributed.

Various elements of the model can be treated as either heterogeneous or homogeneous. In the analyses in this chapter (and in Chapter 3), we used a heterogeneous Λ. An alternative approach would be to treat it as homogeneous, potentially resulting in just four transition rate parameters relating to being in an OU (subordinate state) or a BM (leading/independent) state. Whilst this will provide coarser results, the number of rate parameters can quickly become undesirable in a heterogeneous Λ and it should improve the mixing of the inference. The model and inference we have detailed is spatially homogeneous. However, more rich and beneficial information could be obtained through including spatial covariates such as environmental data or the positions of the animals relative to each other. These may affect the social behaviour, or independent and/or leading animals may be affected by a static (Blackwell *et al.*, 2016) or dynamic (Wang *et al.*, 2019) environment rather than following Brownian motion. The algorithm detailed in Section 4.1.1 to propose new trajectories is easily extended to the spatially heterogeneous case, though it slows it down considerably — a problem we examine and look to solve in Chapter 5. Furthermore, we could add some heterogeneity into the animals themselves. For example, we could use characteristic information such as the animal's age or sex to ascertain how they interplay with social dominance. Or independent transition rates could be estimated for each animal in a model that is hierarchical in the statistical sense, but some consideration will be needed to weigh up the benefits of substantially increasing the number of parameters.

Data Availability

The R code required to run the inferences in Sections 4.2, 4.3, 4.4 and 4.5, along with the simulated data, is available at the following GitHub repository:
https://github.com/jemilner/influenceHierarchy (Milner *et al.*, 2020, doi:10.5281/zenodo.3972134). The baboon data used in Section 4.3 is available on Movebank (Crofoot *et al.*, 2015). The particular subset that we used relates to baboons 3, 4, 5, 9 and 11 between the times 05:51:38 and 06:06:36 on 2012-08-03 (YYYY-MM-DD).

Additionally, in the GitHub repository, there is a brief README file that instructs on how to use the code in R. This includes guidance on what format the data is required to be in, what tuning parameters need consideration and altering the number of states which model leading and independent movement. There is also information for reproducing the analysis in this chapter.

Chapter 5

Towards Spatial Heterogeneity

Previous work on developing our model has concentrated on the spatially homogeneous case and we discuss the merits of extending it to the spatially heterogeneous one in Section 5.1. In order to unlock those merits, we first present a new method with which we can sample animal movement and behaviour in heterogeneous space in continuous time in Section 5.2. The accompanying model fitting algorithm is detailed in Section 5.3 and we examine the capability of this new approach in Section 5.4. Whilst we don't utilise this new functionality to undertake analysis that incorporates spatial covariates, we do utilise it to introduce a *zone of interaction* into our model — a proximity that animals must be within in order to interact. The justification for doing so, along with the implementation details, are presented in Sections 5.5 and 5.6. The zonal model is tested and refined in Section 5.7 through being fitted to simulated data, after which we revisit the baboon troop from Chapter 4 in Section 5.8.

5.1 Spatial Heterogeneity

A spatially heterogeneous movement model has the capacity to provide more insight than its spatially homogeneous counterpart. As the spatial context within which an animal (or group of animals) is moving changes, a spatially heterogeneous model is able to capture how the animal's movement behaviour responds to that change. For example, a species of schooling fish (golden shiners) modulates their movement speed depending on the level of light (Berdahl *et al.*, 2013). An extended model could also capture information on spatial utilisation. For instance, roads and sleeping sites are important predictors of baboon movement (Strandburg-Peshkin *et al.*, 2017).

Studies have also shown there is an important interplay between social structure and the environment. For example, group composition has an impact on the effectiveness of obstacle avoidance (Croft *et al.*, 2015); returning to the golden shiners, collective dynamics enhance their environmental gradient-tracking ability (Berdahl *et al.*, 2013); the ranking of African elephant families is correlated with spatial behaviours such as time spent near water and in protected areas (Wittemyer

et al., 2007). Such insights can provide useful tools in monitoring the resilience of (or impact on) social groups (due) to environmental change (King *et al.*, 2018).

Extending our influence hierarchy approach to the spatially heterogeneous case and capturing the environmental context in which the social behaviours occurred therefore provides a greater opportunity to learn about and monitor social animals.

5.2 Spatially Heterogeneous Movement

In previous chapters, which are entirely concerned with the spatially homogeneous case, we used a forwards simulation approach for simulating an animal's behavioural trajectory between two observations, say at times t_a and t_b. That is, we sample their state switches forwards in time from t_a. The new trajectory must finish in the (currently estimated) behavioural state at t_b so that it is consistent with the trajectory outside of those two observations.

In order to extend the model to the spatially heterogeneous case though, we also need to sample the animals locations between t_a and t_b. A potential candidate with which to do this is a spatially heterogeneous extension of the forwards simulation approach as detailed by Blackwell *et al.* (2016). However, as the name suggests, forwards simulation is naive in that it does not take into account the animal's next known location at t_b. As a result, quite wayward movement trajectories can be simulated (relative to the end location at t_b) and acceptance rates of these can be low. We have therefore developed a new movement proposal method which is conditional on the animal's next known location in order to facilitate good mixing of our model in a spatially heterogeneous context. We will first provide details of this method, along with information on how to use it in the inference process, and then examine it with regards to tuning and how it performs against forwards simulation.

5.2.1 Latent Diffusion Bridge

Say we are proposing the location of an animal at a switching time, t_s, between observations t_a and t_b (where $t_a < t_b$). Conceptually, conditioning the proposal distribution of the movement from t_a to t_s on the location at t_b is simple enough. For example, for animals moving under Brownian motion (BM) processes this would translate to a Brownian bridge. However, as t_s is a switching time and we are now operating in heterogeneous space, we must propose an animal's location at t_s before any potential state switch as the transition may depend on the spatial context. Therefore, we don't actually know (or have not yet estimated) the state the animal is in from t_s to t_b at the moment of proposing its location. This lack of knowledge is heightened when there are multiple switching times between two observations as the animal may switch state multiple times before t_b.

In order to satisfy our requirement of conditioning on the next known location whilst not knowing the behavioural state(s) of the remaining movement, we introduce a latent diffusion state. That is,

a BM process separate from any included in the movement model and we will refer to this state as the latent diffusion process (LDP). This method in general will be referred to as the latent diffusion bridge (LDB). So when proposing the location of an animal at time t_s, conditional on its location at time t_b, we derive that proposal distribution on the basis that the animal is moving according to one of the diffusion processes in the movement model in the interval $[t_a, t_s)$ and according to the LDP in the interval $[t_s, t_b]$.

We use BM for the LDP as we know (or assume) the animal is moving according to some diffusion process, in particular either BM or OU, and BM can be thought of as a special case of an OU process. It also offers a simple interpretation. A small diffusion coefficient of the LDP will result in location samples at t_s that are streamlined towards the location at t_b. On the other hand, a large coefficient will provide little guidance — much like forwards simulation. In the context of Markov chain Monte Carlo (MCMC) sampling, this coefficient is much like any other tuning parameter.

Note that the LDP is solely part of the proposal process and it is not part of the movement model. That is, the declaration that the focal animal moves under the LDP from t_s to t_b is merely temporary. Once the location at t_s is sampled, we can then sample the state the animal will be in for their next movement step as normal, given their new spatial context.

5.2.2 Latent Diffusion Bridge Distribution

In order to use it, we need to derive the proposal distribution of the LDB. Similarly to Chapter 4, we propose new trajectories (now including movement) one animal at a time (the focal animal) whilst treating those of its peers as fixed. Thus, there are two different scenarios when we utilise the LDB and require the corresponding proposal distribution. Firstly, when t_s is a switching time for the focal animal, we sample the locations of all the animals. We require this complete information on the group as the focal animal may end up interacting with any of its peers at t_s. Secondly, when t_s is a switching time for another animal, we resample the focal animal's location.

Switching Time of the Focal Animal

For a potential switching time for animal i between two observations, t_a and t_b, we need to sample the location of the whole group. Let $t_0 = t_a$ and let a potential switching time, t_s, for animal i be the s^{th} switching time in the interval $(t_0, t_s]$, with $t_s < t_b$. In general, the times of the $1^{st}, ..., s^{th}$ switches correspond to times $t_1, ..., t_s$. Let $\tau_c = \{t_{s-1}, t_s, t_b\}$, $\beta_c = \{\beta_{s-1}, \beta_s\}$ (where β_s is a vector containing the states of all animals at time t_s) and Ω represent all movement and transition parameters and the coefficient of diffusion of the LDP. The location of the group at time

t_s in a particular axis, given their location at time t_{s-1}, is distributed as follows:

$$\boldsymbol{G}_s|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \tau_c, \beta_c, \Omega \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Xi}) \tag{5.1}$$

$$\boldsymbol{G}_s|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \tau_c, \beta_c, \Omega = \boldsymbol{A}\boldsymbol{g}_{s-1} + \boldsymbol{B}\boldsymbol{\Theta}_{s-1} + \boldsymbol{C} \tag{5.2}$$

where equation 5.2 is simply equation 5.1 written out more explicitly. Here we have used a simplified notation, compared to equation 3.5, where $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{G}_{s-1}, \boldsymbol{F}_{s-1}, \delta t)$, $\boldsymbol{\Xi} = \boldsymbol{\Xi}(\boldsymbol{F}_{s-1}, \delta t)$ and $\delta t = t_s - t_{s-1}$. $\boldsymbol{A} = e^{\boldsymbol{F}_{s-1}(t_s - t_{s-1})}$, $\boldsymbol{B} = 1 - \boldsymbol{A}$, $\boldsymbol{C} \sim \text{MVN}(0, \boldsymbol{\Xi})$ and $\boldsymbol{F}_{s-1}$ is the attraction matrix at time t_{s-1}. The respective leader's locations in $\boldsymbol{\Theta}_{s-1}$ are taken from $\boldsymbol{g}_{s-1}$. Note, here we are conditioning on β_s (which we will not have sampled yet) and t_b even though $\boldsymbol{G}_s$ is independent of both. However, it is useful to include them to provide consistency with other conditional distributions during the derivation of equation 5.6. With the forwards simulation method, the locations of the animals at t_s are proposed with the distribution in equation 5.1.

We, however, want to propose the locations at t_s conditionally also on the next known locations, $\boldsymbol{G}_b$. Firstly, we note:

$$\boldsymbol{G}_b|\boldsymbol{G}_s = \boldsymbol{g}_s, \tau_c, \beta_c, \Omega \sim \text{MVN}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Xi}}) \tag{5.3}$$

$$\boldsymbol{G}_b|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \tau_c, \beta_c, \Omega \sim \text{MVN}(\boldsymbol{\mu}^R, \boldsymbol{\Xi}^R) \tag{5.4}$$

$$\boldsymbol{G}_b|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \tau_c, \beta_c, \Omega = \tilde{\boldsymbol{A}}\boldsymbol{G}_s + \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s + \tilde{\boldsymbol{C}} \tag{5.5}$$

where equation 5.3 is the distribution of the movement from t_s to t_b. Equation 5.4 is the recursive distribution of the movement through the times t_{s-1}, t_s and t_b. The recursive method with which we obtain this distribution is the same one used in Chapters 3 and 4 and more details can be found in Appendix C and Blackwell (2003). Equation 5.5 is the distribution of equation 5.4 explicitly written out where $\tilde{\boldsymbol{A}} = e^{\boldsymbol{F}_s(t_b - t_s)}$, $\tilde{\boldsymbol{B}} = 1 - \tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{C}} \sim \text{MVN}(0, \tilde{\boldsymbol{\Xi}})$ and the leader's locations in $\boldsymbol{\Theta}_s$ are taken from the expectation of $\boldsymbol{G}_s$, $\boldsymbol{\mu}$. We can use the above and the following result to derive $\boldsymbol{G}_s|\boldsymbol{G}_{s-1}, \boldsymbol{G}_b$:

$$\begin{pmatrix} \boldsymbol{G}_s \\ \boldsymbol{G}_b \end{pmatrix} |\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \tau_c, \beta_c, \Omega \sim \text{MVN}(\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}^R \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Xi} & (\tilde{A}\boldsymbol{\Xi})^T \\ \tilde{A}\boldsymbol{\Xi} & \boldsymbol{\Xi}^R \end{pmatrix} \tag{5.6}$$

where $(\tilde{\boldsymbol{A}}\boldsymbol{\Xi})^T$ is the transpose of $\tilde{\boldsymbol{A}}\boldsymbol{\Xi}$. The derivation of $\text{Cov}[\boldsymbol{G}_b, \boldsymbol{G}_s|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \tau_c, \beta_c, \Omega] = \tilde{\boldsymbol{A}}\boldsymbol{\Xi}$ is provided in Appendix D. Having derived equation 5.6, we can then use standard results for a conditional multivariate normal distribution such that:

$$\text{E}[\boldsymbol{G}_s|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \boldsymbol{G}_b = \boldsymbol{g}_b, \tau_c, \beta_c, \Omega] = \boldsymbol{\mu} + (\tilde{A}\boldsymbol{\Xi})^T(\boldsymbol{\Xi}^R)^{-1}(\boldsymbol{g}_b - \boldsymbol{\mu}^R) \tag{5.7}$$

$$\text{Car}[\boldsymbol{G}_s|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \boldsymbol{G}_b = \boldsymbol{g}_b, \tau_c, \beta_c, \Omega] = \boldsymbol{\Xi} - (\tilde{A}\boldsymbol{\Xi})^T(\boldsymbol{\Xi}^R)^{-1}(\tilde{A}\boldsymbol{\Xi}) \tag{5.8}$$

Note that the behavioural states at t_s (β_s) that feed in to $\boldsymbol{F}_s$ are the same as those at t_{s-1} but with

the state/movement process of the focal animal changed to the LDP. Consequently, as the LDP has no attraction term, if the focal animal is in a subordinate state from t_{s-1} to t_s, it will lose its social link to its dominant (and the rest of the subgroup that is 'above' it) for the bridge step from t_s to t_b. Whilst that may not seem ideal, we do not know if that animal will retain that social link when we come to sample its state at t_s.

Switching Time of Another Animal

Whilst resampling the location of animal i at the switching time of another animal, we treat the locations and behaviours of the other animals as fixed. Say t_s, still within (t_a, t_b), is now a switching time for one of the other animals. We can use the information of the group's location at time t_b to propose more reasonable samples for animal i's location at t_s. First, note the following:

$$G_{s_i}|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \boldsymbol{G}_{s_{-i}} = \boldsymbol{g}_{s_{-i}}, \tau_c, \beta_c, \Omega \sim \mathrm{N}(\mu, \Xi) \tag{5.9}$$

$$\boldsymbol{G}_s|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \boldsymbol{G}_{s_{-i}} = \boldsymbol{g}_{s_{-i}}, \tau_c, \beta_c, \Omega = (g_{s_1}, ..., g_{s_{i-1}}, G_{s_i}, g_{s_{i+1}}, ..., g_{s_n})^T \tag{5.10}$$

$$\boldsymbol{G}_b|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \boldsymbol{G}_{s_{-i}} = \boldsymbol{g}_{s_{-i}}, \tau_c, \beta_c, \Omega \sim \mathrm{MVN}(\boldsymbol{\mu}^R, \boldsymbol{\Xi}^R) \tag{5.11}$$

$$\boldsymbol{G}_b|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \boldsymbol{G}_{s_{-i}} = \boldsymbol{g}_{s_{-i}}, \tau_c, \beta_c, \Omega = \tilde{\boldsymbol{A}}\boldsymbol{G}_s + \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s + \tilde{\boldsymbol{C}} \tag{5.12}$$

where $\boldsymbol{G}_{s_{-i}}$ are the locations of all animals except i at time t_s. μ and Ξ in equation 5.9 are derived from equation 5.1 when we 'know' the locations of all animals except i at t_s and therefore we can condition on them. Equation 5.10 concatenates this univariate random variable with the fixed locations of the other animals at time t_s, where n is the number of animals in the group in total. We wrap the univariate distribution of equation 5.9 in this vector to create consistent dimensions in equation 5.12 and further derivations below. $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{B}}$, $\tilde{\boldsymbol{C}}$ and $\boldsymbol{\Theta}_s$ in equation 5.12 are the same as in equation 5.5 above.

Similarly to using equation 5.6 in the case of animal i's switching time, we can use equation 5.13 here:

$$\begin{pmatrix} G_{s_i} \\ \boldsymbol{G}_b \end{pmatrix} |\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \boldsymbol{G}_{s_{-i}} = \boldsymbol{g}_{s_{-i}}, \tau_c, \beta_c, \Omega \sim \mathrm{MVN}(\begin{pmatrix} \mu \\ \boldsymbol{\mu}^R \end{pmatrix}, \begin{pmatrix} \Xi & (\tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\Xi)^T \\ \tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\Xi & \boldsymbol{\Xi}^R \end{pmatrix} \tag{5.13}$$

where $\mathbf{1}_{\boldsymbol{i}}$ represents a column vector of 0s with a 1 at the i^{th} position, the length of which is the number of animals in the group. The derivation of $\mathrm{Cov}[\boldsymbol{G}_b, G_{s_i}|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \boldsymbol{G}_{s_{-i}} = \boldsymbol{g}_{s_{-i}}, \tau_c, \beta_c, \Omega]$ is once again supplied in Appendix D. Having derived equation 5.13, we can then use standard results for a conditional multivariate normal distribution such that:

$$\mathrm{E}[G_{s_i}|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \boldsymbol{G}_b = \boldsymbol{g}_b, \boldsymbol{G}_{s_{-i}} = \boldsymbol{g}_{s_{-i}}, \tau_c, \beta_c, \Omega] = \mu + (\tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\Xi)^T(\boldsymbol{\Xi}^R)^{-1}(\boldsymbol{g}_b - \boldsymbol{\mu}^R) \tag{5.14}$$

$$\mathrm{Var}[G_{s_i}|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \boldsymbol{G}_b = \boldsymbol{g}_b, \boldsymbol{G}_{s_{-i}} = \boldsymbol{g}_{s_{-i}}, \tau_c, \beta_c, \Omega] = \Xi - (\tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\Xi)^T(\boldsymbol{\Xi}^R)^{-1}(\tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\Xi) \tag{5.15}$$

Note, in the above algorithm, we propose the locations at t_s conditionally on the next observed

locations at t_b. An alternative approach would be to propose them conditionally on the previously accepted trajectory. That is, we could condition on the locations at the time, in the previously accepted trajectory, with the smallest interval after t_s, say $g_{\hat{s_+}}$ at time $t_{\hat{s_+}}$. That time $t_{\hat{s_+}}$ could be an observation, but it could also be a switching time for one of the other animals. There are pros and cons between using the next observed locations or $g_{\hat{s_+}}$. Whilst the $g_{\hat{s_+}}$ approach will sometimes result in conditioning on estimates, those estimates will have been deemed 'acceptable'. However, the time difference between t_s and $t_{\hat{s_+}}$ will be relatively small (compared to the difference between t_s and t_b) and so conditioning on $g_{\hat{s_+}}$ will propose new locations that do not stray far from our previous estimate (as the variance of the LDB distribution is a function of that time interval). That will restrict the trajectories from mixing well and exploring the spatial possibilities. Conditioning on the next observed locations will still 'guide' the distribution of g_s (or g_{s_i}), but, through containing a greater variance, will restrict it less. Furthermore, conditioning on the previous trajectory will create a dependency in the proposal distribution, causing complications when calculating the Metropolis-Hastings (MH) ratio.

5.2.3 Transition Rates

With this extension to spatial heterogeneity, we can no longer use Gibbs sampling for the transition rates. Previously, in the spatially homogeneous case, all the state switches are drawn from the same single-trial multinomial distribution (derived from the current estimate of the transition rates). The likelihood of the rates was therefore an s-trial multinomial, where s is the total number of potential switching times, and so, using a Dirichlet prior, we were able to derive their full conditional distribution from which to sample from. However, in the spatially heterogeneous case, each (potential) state switch may have been sampled from a different distribution as the transition rates depend on location. We therefore lose the above ability and resort to a Metropolis-Hastings random walk.

Consequently, we now formulate our transition matrix to be homogeneous, whereas previously we have used a heterogeneous one with each transition rate being distinct. Random-walk proposals would not have been desirable with the 42 transition rates used in previous chapters. This change is also partly motivated from the suspicion that a homogeneous Λ will mix better than a heterogeneous one, even if the end results are less rich. This improvement in mixing will be valuable now we have moved to the trickier context of spatial heterogeneity.

There are several ways we could categorise the rates; the one we have chosen results in only four transition rates: subordinate-to-subordinate (λ_{OU-OU}), subordinate-to-leader/independent (λ_{OU-BM}), leader/independent-to-subordinate (λ_{BM-OU}) and lastly leader/independent-to-leader/independent (λ_{BM-BM}). This grouping significantly reduces the number of parameters (four compared to the previous 42 in a seven state example) whilst still providing a good level of granularity in the Markov chain. This categorisation is quite general, but it would also be quite simple to categorise the rates specifically for the analysis at hand. For instance, dividing the above categories by age or sex.

An example of what form this homogeneous transition matrix might take in this model is given below, where there are three animals in the data set, two BM states in the model, $\lambda_{OU-OU} = 0.04$, $\lambda_{OU-BM} = 0.02$, $\lambda_{BM-OU} = 0.03$ and $\lambda_{BM-BM} = 0.02$. Let's say at a particular switching time for animal 2: animal 1 is subordinate to animal 2, who is in turn subordinate to animal 3. Animal 2 cannot switch to states S_{A2} (subordinate to itself) or S_{A1} (producing a cyclic structure). Therefore, rates $\lambda_{S_{A3},S_{A1}} = \lambda_{S_{A3},S_{A2}} = 0$ and $\lambda_{S_{A3}}$ for this particular animal/scenario is 0.04 (from the possibility of switching to either BM state).

$$\Lambda = \begin{array}{c} \\ S_{A1} \\ S_{A2} \\ S_{A3} \\ BM_1 \\ BM_2 \end{array} \begin{array}{c} \begin{array}{ccccc} S_{A1} & S_{A2} & S_{A3} & BM_1 & BM_2 \end{array} \\ \begin{pmatrix} 0 & \lambda_{OU-OU} & \lambda_{OU-OU} & \lambda_{OU-BM} & \lambda_{OU-BM} \\ \lambda_{OU-OU} & 0 & \lambda_{OU-OU} & \lambda_{OU-BM} & \lambda_{OU-BM} \\ \lambda_{OU-OU} & \lambda_{OU-OU} & 0 & \lambda_{OU-BM} & \lambda_{OU-BM} \\ \lambda_{BM-OU} & \lambda_{BM-OU} & \lambda_{BM-OU} & 0 & \lambda_{BM-BM} \\ \lambda_{BM-OU} & \lambda_{BM-OU} & \lambda_{BM-OU} & \lambda_{BM-BM} & 0 \end{pmatrix} \end{array} \tag{5.16}$$

Note, even though the transition rates depend on the locations of the animals in the spatially heterogeneous case, we have not explicitly accounted for that dependency in the notation. That is, instead of using $\lambda_{uv}(g_t)$ to denote the rate of switching from state u to state v when an animal is at location g_t, we are using the simplified notation λ_{uv}. All analysis that we undertake in this chapter concerns hard spatial boundaries — an animal is either in a particular region or it isn't. As such, each transition rate to state v, say, at any given switching time is either its full value (e.g. λ_{uv}) when an animal is in an appropriate region for behaviour v, or 0 when it isn't. Therefore, when we discuss a transition rate, it is on the basis that the animal in question is in an appropriate region. When we discuss the rate of leaving a state (e.g. λ_u), that rate is simply a sum of all the transition rates (from u) to states that are possible given the spatial context at that time.

5.3 Inference with the Latent Diffusion Bridge

Once again, Markov chain Monte Carlo methods are used to infer both the behavioural and movement parameters. Each iteration of the MCMC algorithm consists of three parts. Firstly, we propose new movement and behavioural trajectories in continuous time, with help of the LDB, as detailed in Section 5.3.2. Secondly, we propose new transition rates in accordance with the current behavioural trajectories by means of a Metropolis-Hastings random walk. Finally, we propose new movement parameters in accordance with the current movement trajectories in exactly the same fashion as in Section 3.6.3. That is, with a Metropolis-Hastings random walk.

5.3.1 Fitted Model Structure

As discussed in Section 3.7.1, there has thus far been a discrepancy between the exact model used to simulate data and the one fitted to it. The inference model developed in this chapter now aligns with that used to simulate data. That is, transition rates that result in a cyclic hierarchy are set to 0 before sampling a state switch. The reasons for this realignment are presented in Section 3.7.1.

5.3.2 Movement and Behavioural Trajectories

The algorithm with which we propose an animal's movement and behavioural trajectory is largely the same as that in Chapter 4 (Section 4.1.1) — though we now sample the animal's location at switching times too. Using the same notation as in that section, there are again three scenarios to account for:

- If $\hat{\tau}_j \in \tau_o$ for $j = 2, ..., \hat{p}$, the locations are known. If $\hat{\tau}_j$ is a partial observation (that is, i's location was not observed at this time), the missing location can be proposed using the LDB method as this scenario is equivalent to updating animal i at the switching time of another animal. That is, we can condition on the locations of all animals except i at time $\hat{\tau}_j$ and the locations of all animals at the next observation (equations 5.14 and 5.15). Here though, we 'know' what state i is in in both sections of the bridge, as it cannot switch at an observation. So whilst we don't need the LDP, the derivation of the movement distribution is the same as that for the LDB. The behavioural states of all animals at $\hat{\tau}_j$ are carried forward from $\hat{\beta}_{j-1}$.

- If $\hat{\tau}_j \in \tau_{s_{-i}}$ for $j = 2, ..., \hat{p}-1$, we propose animal i's location at $\hat{\tau}_j$ from a distribution defined by equations 5.14 and 5.15, which conditions on the locations of all animals except i at time $\hat{\tau}_j$ and the locations of all animals at the next observation. We use our previously sampled states for all animals except i at these times whilst the behavioural state of i is carried forward from $\hat{\beta}_{j-1}$.

- If $\hat{\tau}_j \in \hat{\tau}_{s_i}$ for $j = 2, ..., \hat{p}-1$, we propose the locations of all animals with a distribution defined by equations 5.7 and 5.8, which conditions on their next known locations. The behavioural states of all animals except i at $\hat{\tau}_j$ are carried forward from $\hat{\beta}_{j-1}$. The probability of $\hat{\tau}_j$ being a switch for i is λ_u/λ_{max} when i is in state u. If so, the new state is v with probability λ_{uv}/λ_u, otherwise, the state of i at $\hat{\tau}_j$ is also carried forward from $\hat{\beta}_{j-1}$.

In all three above scenarios, we carry out any location sampling before any state switching. As with Chapter 4, we require $\hat{\beta}_{\hat{p}} = \beta_p$ (except for when $\hat{\tau}_p$ is the final observation in the data set) and that i's new behavioural trajectory does not conflict with the previously accepted states of all of the other animals at their own switching times. Recall that $\hat{p}$ is the length of the new trajectory (both observed and sampled switching times) whilst p is the length of the previous trajectory. Finally, similarly to previous chapters, state estimations for the first observation are sampled uniformly.

When using forwards simulation to sample movement, the distribution of that trajectory (up to the final switching time) cancels out with the proposal distribution — greatly simplifying the MH ratio (see Chapter 2). This is not the case with the LDB algorithm as our proposal distribution is different from that of a movement step in our model. However, as in previous chapters, the behavioural process is still proposed from the model and so the resulting MH ratio is based entirely on movement:

$$\frac{\prod_{j=2}^{\hat{p}} \hat{\Psi}_j}{\prod_{j=2}^{p} \Psi_j} \tag{5.17}$$

where

$$\Psi_j = \begin{cases} P(\boldsymbol{G}_j \mid \boldsymbol{G}_{j-1} = \boldsymbol{g}_{j-1}), & \tau_j \in \tau_o \\ \frac{P(G_{j_i} \mid \boldsymbol{G}_{j-1}=\boldsymbol{g}_{j-1},\, \boldsymbol{G}_{j_{-i}}=\boldsymbol{g}_{j_{-i}})\; P(\boldsymbol{G}_{j_{-i}} \mid \boldsymbol{G}_{j-1})}{Q(G_{j_i} \mid \boldsymbol{G}_{j-1}=\boldsymbol{g}_{j-1},\, \boldsymbol{G}_{j_{-i}}=\boldsymbol{g}_{j_{-i}},\, \boldsymbol{G}_{j+}=\boldsymbol{g}_{j+})} & \tau_j \in \tau_{s_{-i}} \\ \frac{P(\boldsymbol{G}_j \mid \boldsymbol{G}_{j-1}=\boldsymbol{g}_{j-1})}{Q(G_j \mid \boldsymbol{G}_{j-1}=\boldsymbol{g}_{j-1},\, \boldsymbol{G}_{j+}=\boldsymbol{g}_{j+})} & \tau_j \in \tau_{s_i} \end{cases} \tag{5.18}$$

and each term is also conditioned on $\tau_c = \{\tau_{j-1}, \tau_j, \tau_{j+}\}$, $\beta_c = \{\beta_{j-1}, \beta_j\}$, all movement and transition parameters and the coefficient of diffusion of the LDP. τ_{j+} is the time of the next observation after τ_j and $\hat{\Psi}_j$ takes an identical form but for the new simulation.

In order to keep the notation as simple as possible in the above MH ratio, we have assumed complete observations. So for $\tau_j \in \tau_o$, Ψ_j is simply the likelihood of the movement of the whole group from τ_{j-1} to τ_j. For $\tau_j \in \tau_{s_{-i}}$, Ψ_j consists of the likelihood of our sample of animal i at τ_j given the locations of all others at τ_j. This is multiplied by the marginal of all animals except i of the movement step from τ_{j-1} (obtained from equation 5.1). For $\tau_j \in \tau_{s_i}$, the likelihood is again simply the movement of the whole group from τ_{j-1} to τ_j

The proposal distributions, Q, are normally distributed and defined by equations 5.14 and 5.15 for $\tau_j \in \tau_{s_{-i}}$ and equations 5.7 and 5.8 for $\tau_j \in \tau_{s_i}$. In both scenarios, the proposal for the latest trajectory is independent of the previous trajectory (and vice versa).

5.3.3 Transition Rates

Treating the current sample of the animals' behavioural and movement trajectories as fixed, we update all of the transition rates simultaneously through a Metropolis-Hastings random walk. Independent, normally-distributed proposals are used for each rate and so the MH ratio is reduced to a ratio of likelihoods:

$$\frac{\prod_{j \in \hat{\tau}_s} \hat{\mathrm{P}}(\hat{\beta}_{j,i} = v \mid \hat{\beta}_{j-1,i} = u)}{\prod_{j \in \tau_s} \mathrm{P}(\beta_{j,i} = v \mid \beta_{j-1,i} = u)} \tag{5.19}$$

where τ_s are all the switching times of all the animals, $\beta_{j,i}$ is the behavioural state of animal i (the animal which switching time j concerns) at switching time j and u and v are states. Therefore,

$$\hat{\mathrm{P}} = \begin{cases} \frac{\hat{\lambda}_{uv}}{\lambda_{max}}, & u \neq v \\ 1 - \frac{\hat{\lambda}_u}{\lambda_{max}}, & u = v \end{cases} \tag{5.20}$$

and

$$\mathrm{P} = \begin{cases} \frac{\lambda_{uv}}{\lambda_{max}}, & u \neq v \\ 1 - \frac{\lambda_u}{\lambda_{max}}, & u = v \end{cases} \tag{5.21}$$

where $\hat{\lambda}_{uv}$ is the newly sampled rate of switching from state u to state v and $\hat{\lambda}_u$ is the new rate of leaving state u. λ_{uv} and λ_u relate to the previously accepted rates. We require each rate to be greater than or equal to 0 and the rate of leaving each state must be less than or equal to λ_{max}. If $\hat{\lambda}_u > \lambda_{max}$, the probability of remaining in a state would be negative.

5.4 Latent Diffusion Bridge Evaluation

5.4.1 Trajectory Simulation Amendment

The trajectory proposal algorithm we have described above samples each potential behaviour switch from the current Λ estimate. Doing so means all terms relating to state switching (or remaining in a state) in the likelihood cancel with the corresponding proposal terms in the MH ratio. However, our complex state space means that the condition that the final switch must result in animal i being in the required state plays an overwhelming role in the trajectory updates. That is, a prohibitive amount of proposals get rejected because of it.

To investigate this effect, we looked at the percentage of trajectory proposals that are rejected before MH evaluation due to each condition. To do so, we fitted the model to 10 simulations, each containing three animals and two BM states (so five states in total) for 100 discrete-time intervals of two units of time each. The 10 simulated data sets were derived from a single set of parameter values (Table 5.1) but they each contained a different realisation of the behaviours possible. A value of 0.2 was used for λ_{max}, a more than sufficient value for the transition rates used, and we undertook trajectory updates over 2 to 7 observations.

Similarly to previous chapters, the behavioural process is initialised in the MCMC process randomly and initial transition rates and movement parameter values are well dispersed. We carried out enough trajectory updates per iteration so that each interval for each animal will be resampled (not necessarily accepted) on average. Mostly motivated by the size of the output files, but also with an eye on autocorrelation, we only recorded every second iteration for the movement and transition parameters and every 20^{th} iteration for the behavioural process. The above is consistent

Parameter	Value	Parameter	Value
α	0.2	λ_{OU-OU}	0.04
σ	1.1	λ_{OU-BM}	0.02
ρ_{slow}	0.4	λ_{BM-OU}	0.03
ρ_{fast}	2.5	λ_{BM-BM}	0.02

Table 5.1: Parameter values used for the 10 simulations used in the LDB analysis.

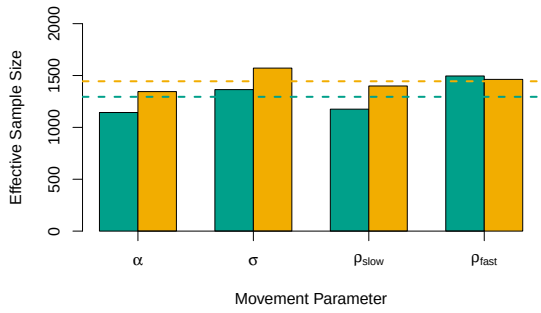
Condition	Rejected (%)	
	Standard	Forced
End State	36.8	0.0
Cyclic Hierarchy	1.9	2.3
No Switches	7.6	7.5

Table 5.2: The percentage of trajectory proposals that are rejected before before MH evaluation during the first 2 hours of the standard and 'forced' methods. *End state* is where the proposal doesn't finish in the required state; *cyclic hierarchy* is where a sampled state results in a cyclic hierarchy; *no switches* is where an update must contain a switch (the start and end states are different) but no switches are sampled. These percentages are averages from fitting each method to 10 simulations, each containing three interacting animals.

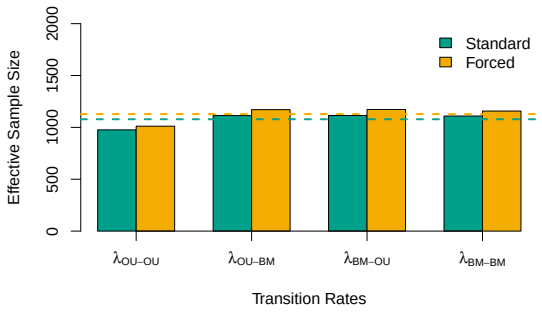
in all analyses in this chapter unless otherwise stated.

Table 5.2 displays the percentage of trajectory proposals that are rejected due to each condition averaged across the 10 simulations. A considerable 36.8% are rejected due to the required end state condition. In previous chapters, throwing away trajectory proposals in a spatially homogeneous context was not a massive computational waste as they consisted purely of computationally simple state switches. However, we are now also proposing the animal's movement in the trajectories with the LDB, which involves more complex computations (multiple conditional variance matrices, samples from multivariate normal distributions and inverse matrices for example). It would therefore be prudent to investigate if there is a way of avoiding all of that wasted effort.

A potential route around this hurdle is to 'force' the last potential switch to result in the required state. Table 5.2 contains the percentage of trajectory proposals that are rejected before MH evaluation with this method (from fitting the model to the same 10 simulations as above with the same tuning parameters and initial parameter values). Figure 5.1 contains further comparisons between the two approaches over the 10 runs. In order to provide a fair comparison, each run was carried out for 24 hours to allow for the standard method being computationally quicker as more trajectory proposals require evaluating with the forced method. That is, in 24 hours the standard method worked through an average of 305k iterations whilst the force method worked through 228k. Though, we did not delve into the exact burn-in period for each run and the ESS was calculated from the second half of the output for simplicity (a more-than sufficient burn-in period in all cases). It is clear this forced approach is not a silver bullet. The acceptance rate of the proposals has improved, but maybe not quite to the extent one might expect given 36.8% would



(a)



(b)

	Standard	Forced
Trajectory Acceptance Rate — Overall (%)	20.4	25.2
Trajectory Acceptance Rate — First 2k Iterations (%)	16.8	24.6

Figure 5.1: Comparison between the standard and forced methods for sampling the final switch in a trajectory update. Results are averaged over 10 runs each, the length of which was dictated by time, as opposed to number of iterations. (a) displays the effective sample size (ESS) on the movement parameters, whilst (b) contains the ESS of the transition rates. The dashed horizontal lines represent the average across the corresponding parameters. ESS is calculated using *coda* (Plummer *et al.*, 2006).

normally fail the end state condition. Whilst it ensured that all trajectories finished in the required state, it may be that many of those that would normally fail the end state condition were already off-piste and unreasonable. However, it does improve the rate at which we generate effective samples as there is an increase in the amount produced over the same time period compared with the standard method. This improvement in the effective sample size (ESS) will be a consequence of the improvement in the mixing of the behaviours caused by the increase in the trajectory acceptance rate. Additionally, it appears the forced approach has a more pronounced effect early on in the inference process. As we can see in Figure 5.1, there is a greater discrepancy in the trajectory acceptance rates during the first 2k iterations. This will help encourage good mixing in the early stages of the inference process and help the parameters approach the correct area of the parameter space without getting stuck in local maxima. Whilst we didn't examine the burn-in period for all of these runs, it was heavily reduced during a separate exploratory analysis on data containing five animals (from approximately 60k to 10k).

Using the 'forced' method requires us to keep track of the last switching time in the update (regardless of whether a switch occurs or not) in the MH ratio. The proposal for this is 1 as there is only a single state that would meet the condition. The term in the likelihood is the probability of that switch ($\lambda_{uv}/\lambda_{max}$) or stay ($1 - \lambda_u/\lambda_{max}$). Thus, the practicalities of this method, both in terms of implementation and running cost, are minimal.

All subsequent analyses in this chapter are conducted with the 'forced' method.

5.4.2 Tuning the Bridge

The LDB brings with it a tuning parameter — the diffusion coefficient of the LDP we are using to represent the focal animal's movement from t_s to t_b. We will use the notation of ρ_b for this parameter. In this section we will discuss how to interpret and consider this tuning parameter.

First, we will discuss the limits of the coefficient. As $\rho_b \longrightarrow \infty$, this method will tend towards forwards simulation. To explain, consider the case when we are sampling the locations of the entire group (Section 5.2.2). The term $(\tilde{A}\boldsymbol{\Xi})^T(\boldsymbol{\Xi}^R)^{-1}$ in equations 5.7 and 5.8 will tend towards 0, as the recursive covariance ($\boldsymbol{\Xi}^R$) will tend to ∞, and so the distribution of $\boldsymbol{G}_s|*$ will converge to MVN($\boldsymbol{\mu}, \boldsymbol{\Xi}$). This is the same distribution as in equation 5.1 — that of sampling a forwards movement step. A similar calculation unfolds for the case when we are sampling the location of just the focal animal.

We require the coefficient to be greater than or equal to 0. To show how the bridge behaves as $\rho_b \longrightarrow 0$, consider an individual animal which is moving in a BM state with coefficient ρ_s from t_a to t_s. The univariate equivalents of equations 5.7 and 5.8 result in the following distribution:

$$G_s|* \sim \mathrm{N}(\mu + (\rho_s^2(t_s - t_a)/(\rho_s^2(t_s - t_a) + \rho_b^2(t_b - t_s)))(g_b - \mu^R), \tag{5.22}$$

$$\rho_s^2(t_s - t_a) - ((\rho_s^2(t_s - t_a))^2/(\rho_s^2(t_s - t_a) + \rho_b^2(t_b - t_s)))) \tag{5.23}$$

As $\rho_b \longrightarrow 0$ and $\mu = \mu^R$, this distribution tends to $\mathrm{N}(g_b, 0)$. The location of this animal at time t_s will therefore be proposed as its location at time t_b with absolute certainty. Whilst the calculation is more complex for the multivariate case, the result will be the same for the focal animal — their sample for G_s will tend to g_b as $\rho_b \longrightarrow 0$.

The value used for the LDP diffusion coefficient will clearly affect the acceptance rate of the trajectory samples. Therefore, the usual method of observing the acceptance rates can be used to tune this parameter. Aside from that, it appears a rule of thumb for a reasonable value will be within the range of the diffusion coefficients of the BM processes in the movement model. Though, this does present a chicken-or-the-egg style problem. However, it also appears that the tuning of the LDP coefficient is not overly sensitive. To showcase this, for several values of the coefficient, we fit an otherwise identical model (tuned as detailed in Section 5.4.1) to 10 simulations (also as detailed in Section 5.4.1).

Table 5.3 shows that for quite a wide range of values (in the context of the movement model's BM coefficients being 0.4 and 2.5) we obtain good mixing. The acceptance rates all remain reasonable, with a slight peak from 1.5 to 2.5 and the only significant decrease occurring at 0.5. Due to this decreased level of mixing, the runs with a coefficient of 0.5 did have to be tuned separately. The mean ESS of both the movement parameters and transition rates follow a similar pattern. Applying the Gelman-Rubin diagnostic (multivariate potential scale reduction factor, MPSRF) on each pair of analyses, we can see (in Tables 5.4 and 5.5) that they all converge to the same posteriors, with the

	LDP Coefficient					
	0.5	1.0	1.5	2.0	2.5	3.0
Trajectory acceptance rate (%)	17.7	22.7	25.2	26.2	26.0	23.6
Movement parameters ESS	1253	1241	1444	1515	1463	1339
Transition rate ESS	856	980	1128	1082	948	904
Augmented data length	201	217	219	219	219	219

Table 5.3: The trajectory acceptance rate, ESS output and augmented data length for each of the LDP coefficients. Each number is the mean calculated over the 10 runs for each coefficient.

Movement Parameters		LDP Coefficient					
Multivariate PSRF		0.5	1.0	1.5	2.0	2.5	3.0
LDP Coefficient	0.5	-	1.95	1.70	1.70	1.67	1.68
	1.0	-	-	1.18	1.20	1.19	1.19
	1.5	-	-	-	1.02	1.02	1.02
	2.0	-	-	-	-	1.01	1.01
	2.5	-	-	-	-	-	1.01
	3.0	-	-	-	-	-	-

Table 5.4: The Gelman-Rubin diagnostic (multivariate PSRF) of the movement parameters between each different tuning run. For example, using a coefficient of 0.5 to fit the model to simulation 1 compared to using a coefficient of 1.0. These are the means over the 10 simulations.

Transition Rates		LDP Coefficient					
Multivariate PSRF		0.5	1.0	1.5	2.0	2.5	3.0
LDP Coefficient	0.5	-	2.30	2.49	2.49	2.48	2.49
	1.0	-	-	1.19	1.18	1.18	1.18
	1.5	-	-	-	1.01	1.01	1.01
	2.0	-	-	-	-	1.01	1.01
	2.5	-	-	-	-	-	1.01
	3.0	-	-	-	-	-	-

Table 5.5: The Gelman-Rubin diagnostic (multivariate PSRF) of the transition rates between each different tuning run. For example, using a coefficient of 0.5 to fit the model to simulation 1 compared to using a coefficient of 1.0. These are the means over the 10 simulations.

exception of the 0.5 and 1.0 coefficient runs. Note, we did not apply the Gelman-Rubin diagnostic across runs with the same coefficient. We were testing to see if each tuning converged to the same posteriors in order to examine the importance of the coefficient value.

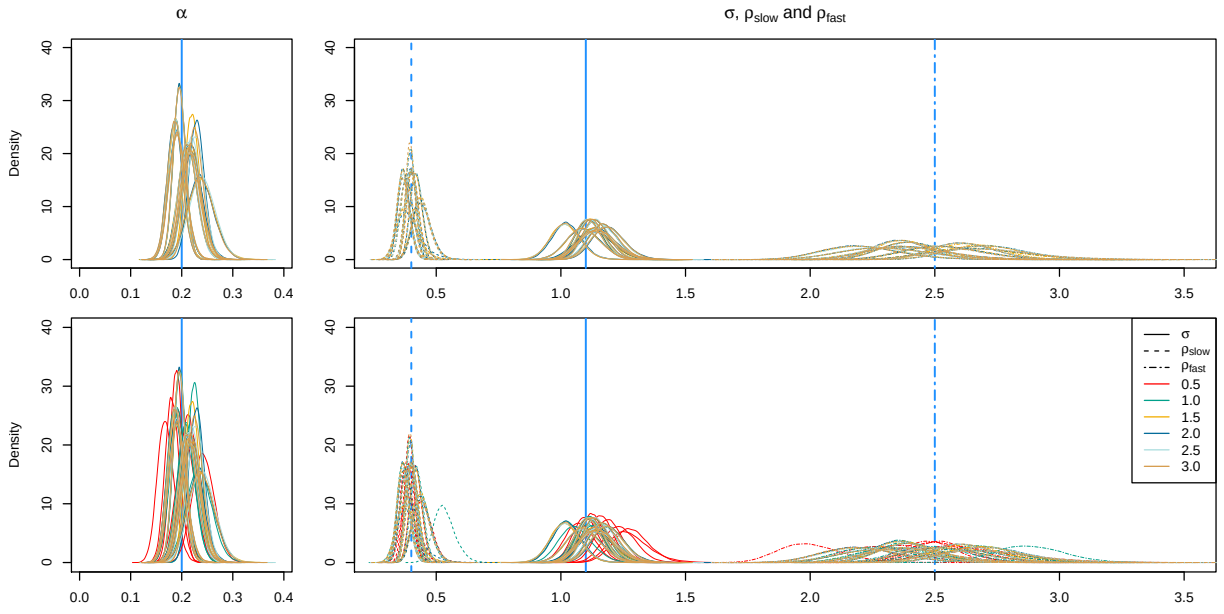
These results can be seen further in Figures 5.2a and 5.2b. The posteriors for coefficients 1.5 to 3.0 are largely in agreement with each other and the true parameter values. The posteriors from these coefficient runs overlap to such an extent we can clearly see 10 distinct distributions — one for each of the simulations. Some of the transition rate posteriors don't align strongly with the true value. However, this is just symptomatic of the variation that will naturally occur between 10 simulations. For example, the group of λ_{BM-BM} posteriors that are close to 0 relate to a simulation containing

only one switch between BM states. The posteriors relating to coefficients 0.5 and 1.0 (particularly 0.5) do not align strongly with the others and produce spurious results.

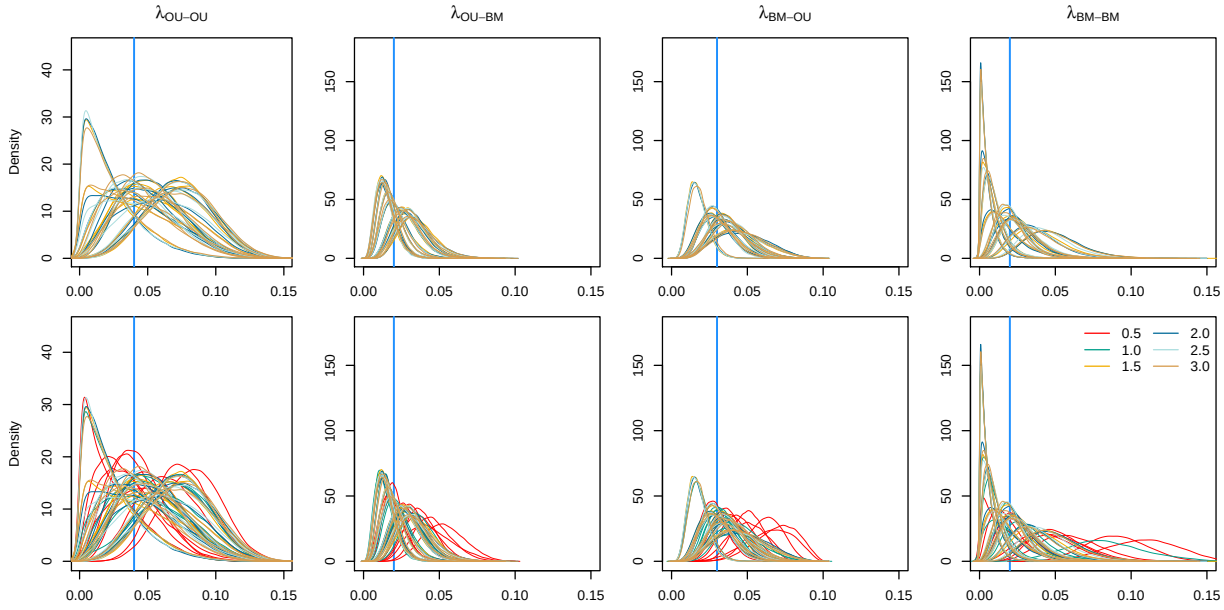
The multivariate PSRFs (concerning a coefficient of 0.5) are much poorer with regards to the transition rates than they are for the movement parameters. The cause of both of these discrepancies will partly come from the poor mixing using this coefficient, but, for the transition rates, it also stems from the quirk that using a too-small coefficient results in a smaller than expected augmented data set. Over an interval (t_a, t_b), the expected number of switching times that will be sampled for an individual is $(t_b - t_a)\lambda_{max}$. Therefore, for simulated data consisting of 101 observations for 3 animals over 200 time units, the expected length of the augmented data will be $101 + (200 - 0) \times 0.2 \times 3 = 221$ when using $\lambda_{max} = 0.2$. From Table 5.3, the mean data length for coefficients 1.5 to 3.0 is 219, whilst for 0.5 it is 201. As explained above, as $\rho_b \longrightarrow 0$, new locations for an animal at time t_s are proposed increasingly towards its location at t_b and with increasing certainty, potentially producing erratic and unlikely movement steps. Additionally, subsequent movement steps (towards switching times between t_s and t_b) will be highly unlikely as the focal animal remains rooted to g_b, particularly in the case when the focal animal is in a subordinate state and some persistent movement towards their dominant is expected. Therefore, it will be unlikely that a new trajectory proposal containing a relatively high number of switching times (compared to the previously accepted trajectory) will contain a more acceptable series of movement steps. A knock-on effect is that the transition rate samples are evaluated over a smaller number of switching times than would normally be expected for the λ_{max} used. This leads to the transition rates being inflated if the behaviours estimated are broadly correct. This is because the more-or-less correct number of state switches now consume a greater proportion of the switching times than if the augmented data set length was closer to 221. That is, the rate of leaving a state will inflate towards λ_{max}, which is still 0.2.

A similar story to the movement and transition parameters unfolds with the state estimations (Figure 5.3). All the coefficient runs align with each other and (mostly) with the true state, but with notable exceptions for coefficients 0.5 and 1.0. The example in Figure 5.3 was chosen as the results of coefficients 1.5 to 3.0 closely aligned both with each other and the true state. This enables the plots to clearly show the runs for coefficients 0.5 and 1.0 (0.5 in particular) deviating even when all others are correct. For a more complete view, Figure 5.4 shows the CDFs of the correct state posteriors for each run under each tuning. The posteriors from coefficients 1.5 to 3.0 are similarly confident — in Figure 5.4a the CDFs from these coefficient runs are once again heavily overlapping. In Figure 5.4b, we can then see the state estimations are less accurate for coefficient 0.5.

The tuning of the diffusion coefficient of the LDP is therefore quite insensitive, only the too-small or too-large extremes need to be avoided. Coefficients 1.5, 2.0 and 2.5 all look optimum for these simulations and so, without much to choose from, a value of 1.5 is used in the subsequent section comparing the LDB method against the forwards simulation approach (as well as in Section 5.4.1).



(a)



(b)

Figure 5.2: (a): posterior distributions of the movement parameters for each of the LDP coefficient runs. Both rows are identical except that the top row omits the posteriors for coefficients 0.5 and 1.0. (b): posterior distributions of the transition rates for each of the LDP coefficient runs. Both rows are identical except that the top row omits the posteriors for coefficients 0.5 and 1.0. The blue vertical lines indicate the true value used.

5.4.3 LDB vs Forwards Simulation

In order to examine any efficiency gained from the LDB method, we ran both the LDB and forwards simulation algorithms on the same 20 simulated data sets. These simulations are deriving exactly

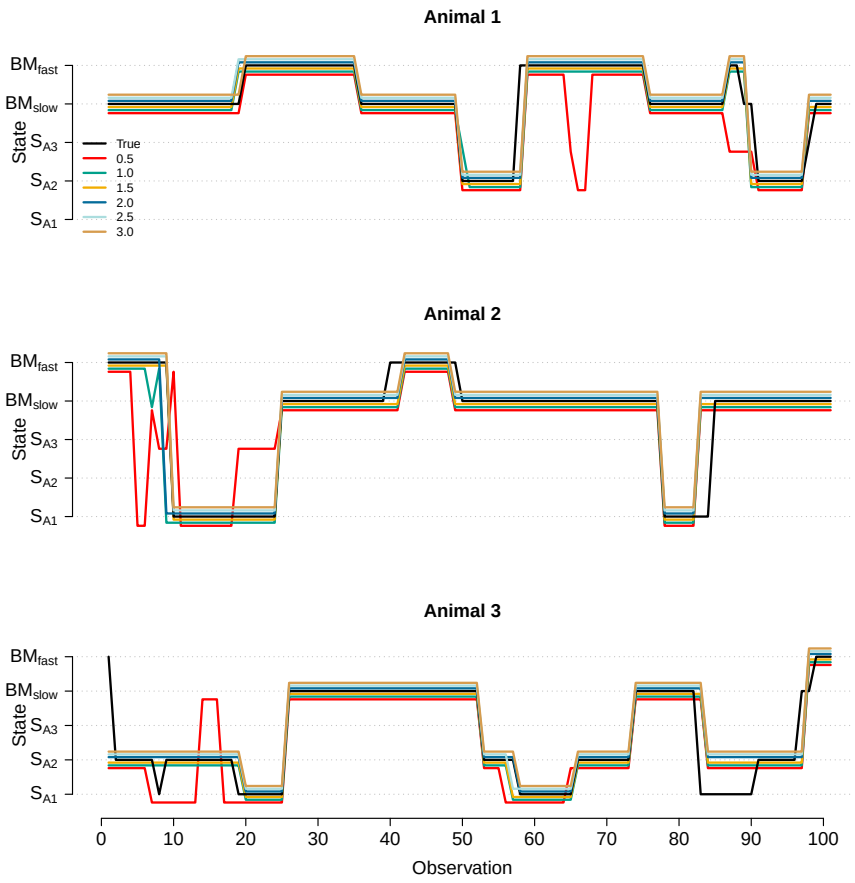
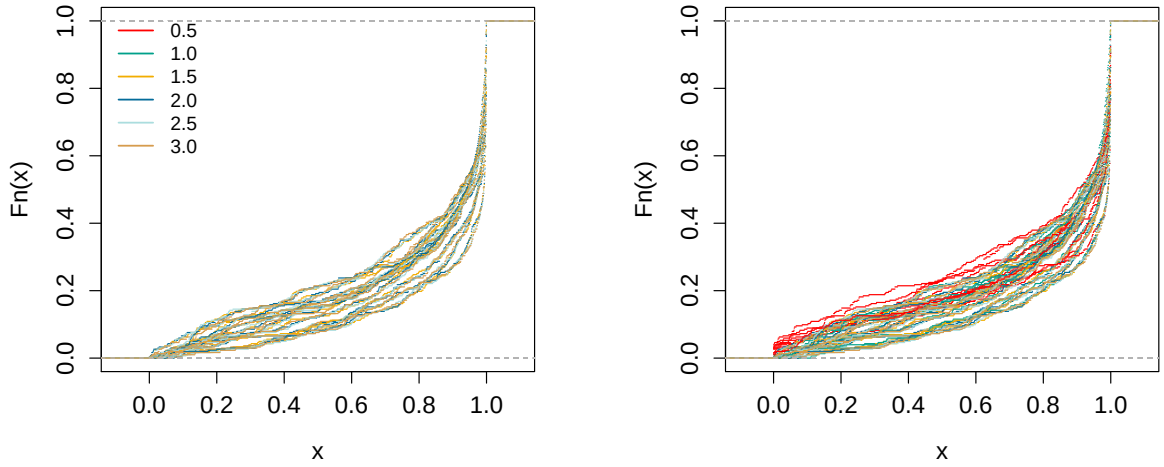


Figure 5.3: The modal state estimation for each of the LDP coefficients for one of the simulations. The lines have been shifted to avoid overlap.

as described in Section 5.4.1. We ran the inference for a set amount of time (24 hours), as opposed for a set number of iterations, to allow for the computational differences between the two methods. That is, forwards simulation will complete an iteration quicker than the LDB algorithm due to a more simplified movement proposal process. Again, we analysed the second half of these chains, rather than precisely locating the burn-in period for each run. Whilst this is always sufficient for the LDB runs, it is not always for the forwards simulation output as we will shortly discuss.

Forwards simulation, due to its rudderless movement proposals, does not mix well. Reasonable trajectories of the animal's movement are not proposed frequently and therefore the behavioural states are not resampled frequently. Subsequently, the movement and transition parameters do not mix well as they are beholden to the current, stubborn trajectory realisation. The movement and transition rate samples may well be jumping around their parameter space sufficiently enough in order to aid good mixing and exploration of the behavioural states, but the trajectory proposals are still likely to go awry and so the opportunity to explore is lost. The movement and transition rate samples will then persist in an area of the parameter space corresponding to a particular trajectory realisation for some time.

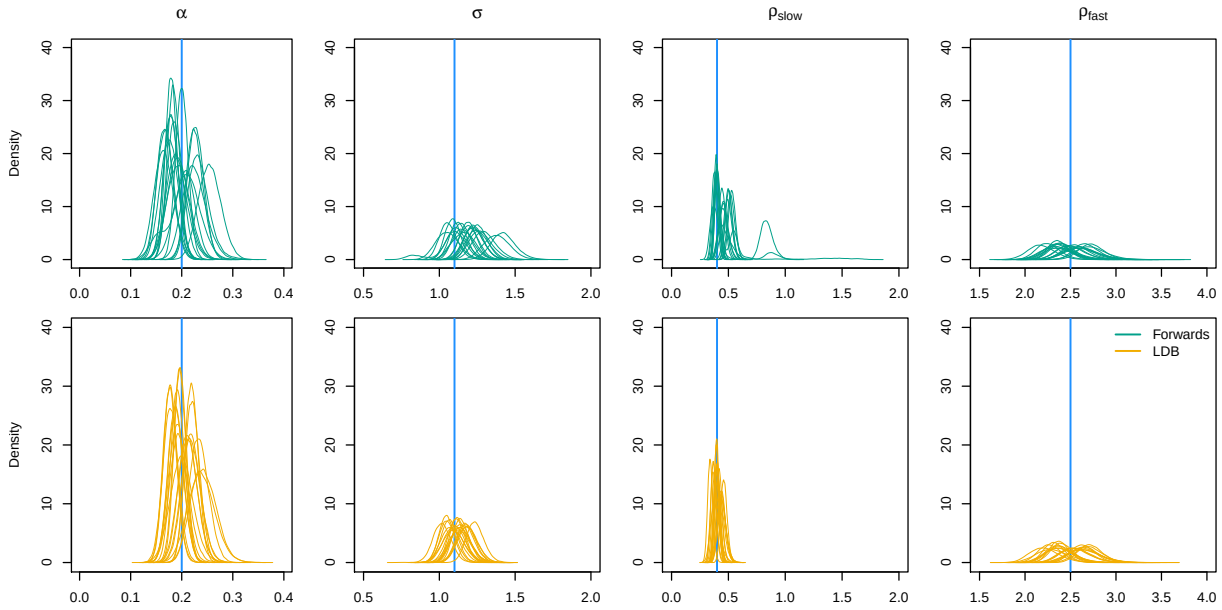


(a) (b)

Figure 5.4: Each line represents the CDF of the true state posterior. That is, the probability an animal is estimated to be in the true state. Each CDF is derived from the posteriors for all animals during a single run at the times of the observed data. (a) contains the CDFs for each run for each coefficient 1.5 to 3.0. (b) adds in the CDFs for coefficients 0.5 and 1.0 for comparison.

The naivety of forwards simulation is compounded when updating longer segments of the data. Whilst the movement trajectories essentially reset at each observation in the segment being updated, each interval (between two observations) has the potential to contain a wildly unreasonable sample. We propose new trajectories in the LDB runs over 2 to 7 observations. Whilst this provides a good level of mixing for that algorithm, it is too ambitious for forwards simulation. For example, using the LDB, updates over 4 observations are accepted approximately 23% of the time. With forwards simulation, this acceptance rate drops to 8%. As a result, trajectory updates with forwards simulation are undertaken over 2 to 4 observations. However, updating over such a small section of data, although it increases the acceptance rate, harms the mixing in another way if used in isolation (that is, not in conjunction with longer updates). It is akin to not being able to see the forest for the trees.

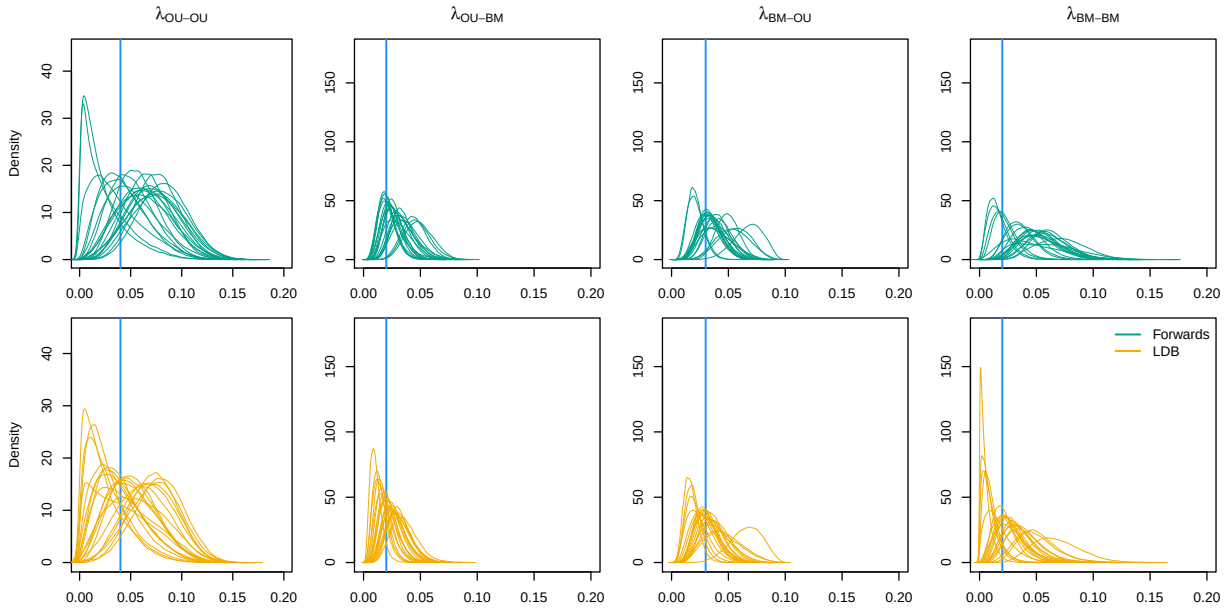
Figures 5.5 and 5.6 show more concrete evidence for this improvement in mixing when using the LDB method. The ESS of the movement parameters increase by over a third on average, whilst that of the transition rates increases by roughly a quarter. The ESS is calculated over the second half of the chains and thus is calculated over 12 hours in each case. Whilst this suggests an improvement, the ESS comparison underplays the advantage gained. Not only does the ESS increase, but it is also drawn from a posterior that more closely resembles the true values used. The LDB runs average out to be extremely consistent with the true value for both the movement parameters and transition rates, whilst the forwards simulation runs are more prone to spurious results.



Parameter	True Value	Forwards		LDB	
		Point Est.	ESS	Point Est.	ESS
α	0.2	0.196	1144	0.204	1407
σ	1.1	1.193	960	1.118	1577
ρ_{slow}	0.4	0.465	807	0.402	1306
ρ_{fast}	2.5	2.459	1360	2.468	1467

Figure 5.5: Posterior distributions from 20 runs for the four movement parameters for both the forwards simulation (top row) and LDB (bottom row) methods. The blue vertical line indicates the true value used. The table contains a summary of the movement parameter results. Point estimates are averages of the density peaks; the effective sample size is the mean over the 20 runs.

Through Figure 5.7 we can see the improved mixing in action to a greater extent. The forwards simulation chains move towards convergence in a stepped fashion (Figure 5.7a). This feature in the chains is due to the aforementioned struggle to resample the trajectories of the animals and thus the movement parameter samples spend considerable time exploring values corresponding to particular trajectory estimations, estimations which will be quite wayward at the start of the inference process. Whilst the movement parameter trace plots indicate the forwards simulation runs reach convergence, the transition rate trace plots tell a different story. We suspect, particularly for simulations 1 and 16, we are still observing the burn-in period for these runs as the transition rate samples wrestle towards the correct area of the parameter space. The LDB runs, on the other hand, converge quickly. It is hard to tell from the transition rate trace plots, but all starting parameter values were initiated to be well dispersed from the eventual posteriors. Additionally, these trace plots indicate that if we had taken into account the burn-in periods, the ESS comparison in Figures 5.5 and 5.6 would be in the favour of the LDB algorithm to an even greater extent.



Parameter	True Value	Forwards		LDB	
		Point Est.	ESS	Point Est.	ESS
λ_{OU-OU}	0.04	0.053	640	0.041	1032
λ_{OU-BM}	0.02	0.028	940	0.021	1223
λ_{BM-OU}	0.03	0.038	1107	0.030	1234
λ_{BM-BM}	0.02	0.024	898	0.024	1001

Figure 5.6: Posterior distributions from 20 runs for the four transition rates for both the forwards simulation (top row) and LDB (bottom row) methods. The blue vertical line indicates the true value used. The table contains a summary of the transition rate results. Point estimates are averages of the density peaks; the effective sample size is the mean over the 20 runs.

Finally, we can observe the difference between the two methods on the state estimations of the animals in Figure 5.8. The LDB produces state estimations that are more certain and more consistent with the true state. We would normally expect the two results to align, even with the improved mixing with the LDB method. However, as seen above, this improved mixing facilitates convergence, something which has yet to be reached in some of the forwards simulation runs. It will also help the MCMC process avoid getting stuck in local maxima and approach the most likely estimate.

Overall, whilst the spatially heterogeneous version of the forwards simulation is a perfectly sensible approach in other contexts (Blackwell *et al.*, 2016), it struggles to propose reasonable movement trajectories in the context of our model. The naivety of forwards simulation is compounded by the fact we are jointly modelling the movement of multiple animals and so wayward simulations are not only unlikely with respect to the focal animal, but also to any others in their subgroup. In order to examine the robustness of our model and inference methods, we like to provide challenging

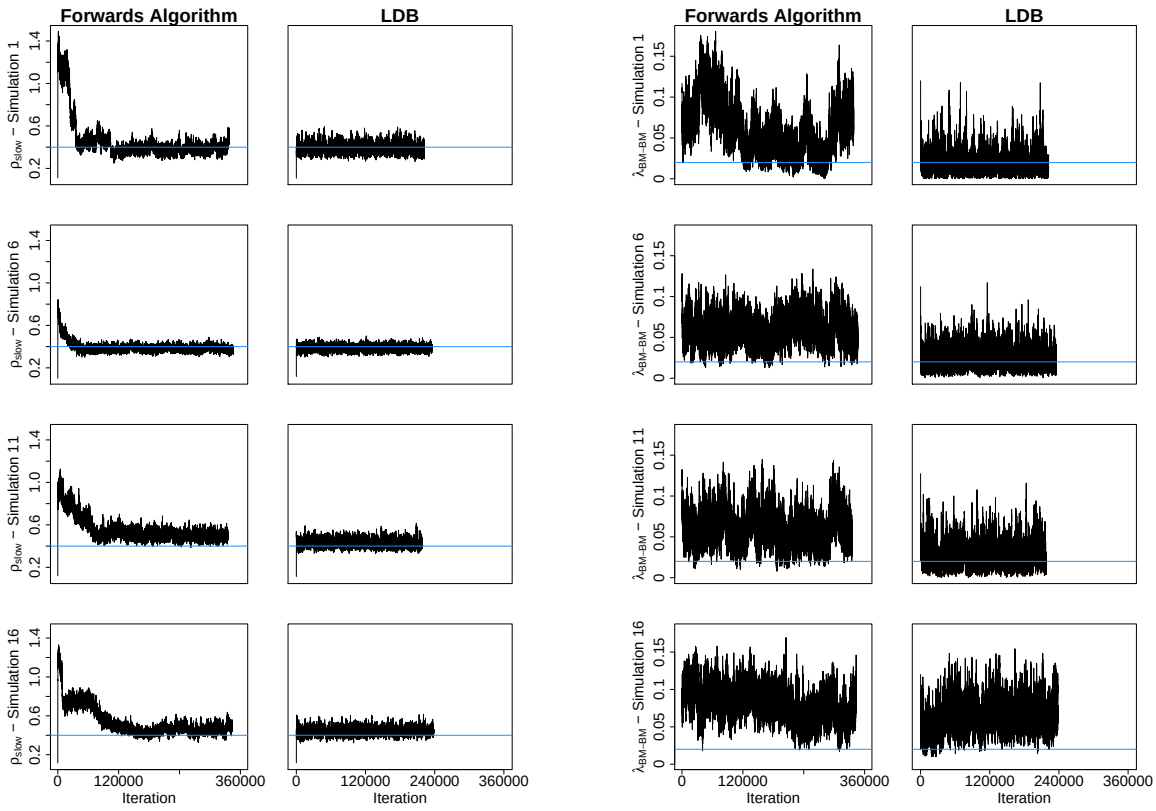


Figure 5.7: (a) contains the trace plots of the movement parameter ρ_{slow} (the diffusion coefficient of state BM_{slow}) for a sample of runs of both the forwards simulation and LDB methods. (b) contains the trace plots of the transition rate λ_{BM-BM} in the same context as (a). The horizontal blue lines indication the true value used in the simulation and the chains have been thinned.

scenarios. Therefore, we initialise the behaviours randomly — meaning the animals are potentially switching state between every observation initially. This, in conjunction with the fact that we could only reasonably propose new trajectories over four or fewer observations, means the inference process when using forwards simulation struggles to get going, as seen in Figure 5.7. Even when it does, the poor mixing has a persistent effect. Figures 5.7 and 5.8 suggest the parameter samples can get close to the true value, but they struggle to make the leap to the global maxima.

The above analysis points towards the LDB providing a sizeable mixing improvement over forwards simulation, an advantage that far outweighs the added computational cost. The advantage of the LDB would be further highlighted in the scenario where we had a reduced computational ability or time to undertake the analysis. For example, say we were only able to run a half or even a third of the number of iterations currently completed. As indicated by the trace plots in Figure 5.7, we would still obtain good parameter estimations when using the LDB. However, the forwards simulation runs would most likely still be within the burn-in period. Furthermore, some exploratory analysis indicates the above improvements are amplified when the data set contains more animals.

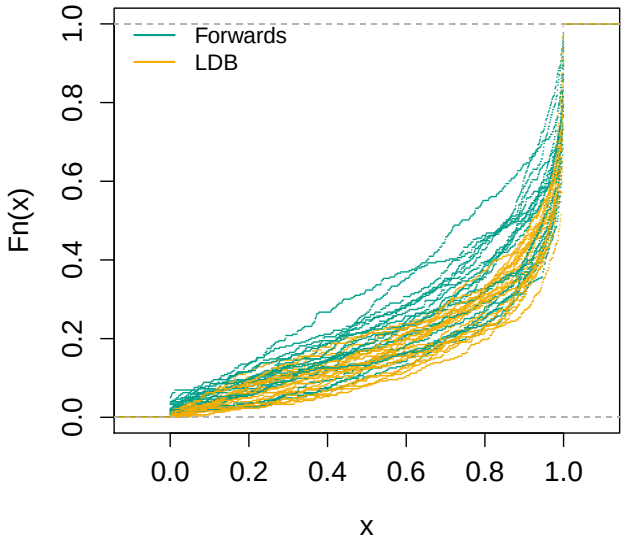


Figure 5.8: Each line represents the CDF of the true state posterior. That is, the posterior probability an animal is estimated to be in the true state. Each CDF is derived from the posteriors for all animals during a single run at the times of the observed data.

5.5 Proximity

The distance between two animals has frequently been utilised in various approaches to analysing social behaviour in animal groups. In its various guises: interactive ranges add biological realism to self-propelled particle (SPP) models for mass-migrating animals (Couzin *et al.*, 2002; Buhl *et al.*, 2006; Codling *et al.*, 2007); distance to a group has been used to define membership of it (Haydon *et al.*, 2008); proximity based approaches are used to define social networks (Wittemyer *et al.*, 2005; Castles *et al.*, 2014; Farine & Whitehead, 2015; Davis *et al.*, 2018), which can then be used for network-based diffusion analysis (Franz & Nunn, 2009); and proximity can be used as metric to quantify dyadic interaction (Long *et al.*, 2014; Joo *et al.*, 2018). Outside of the animal kingdom, proximity is built into latent space models, which are used to investigate social networks in a range of contexts such as sociology and marketing (Sarkar & Moore, 2005). The motivation behind proximity being a staple of social analysis is intuitive: it is a prerequisite for an interaction to occur between animals (Castles *et al.*, 2014; Farine, 2015). Indeed, Garroway *et al.* (2013) concluded that spatial proximity was a better predictor for social associations than kinship — at least for their study animal, southern flying squirrels. Though, despite the above uses and intuition, proximity should not be considered an infallible tool with which to analysis social interaction. For example, other studies point to interaction in some species being driven by topological distance, rather than metric distance (Ballerini *et al.*, 2008; Camperi *et al.*, 2012). See Section 1.2.3 for further discussion. We will therefore discuss in the following section why adding some notion of proximity into our

model can complement our existing framework.

5.6 Zone of Interaction

Incorporating some sense of proximity into our social movement model was not possible in Chapters 3 and 4 as we did not sample the animals' locations at behavioural switching times. Though, now that our model has been extended to the spatially heterogeneous case, we can begin to utilise that capability. Here, we incorporate a *zone of interaction* into our modelling framework.

In a commonly used class of models, Couzin *et al.* (2002) split their interactive zone into ones of repulsion, alignment and attraction in order to simulate collective movement in self-propelled particles. Their work shows that, from simulation experiments, the radii of these zones is a determinant of the type of collective behaviour displayed by the particles, which may be thought to represent a school of fish or a flock of birds. This formulation of the interaction zone is different to ours, in that we are using a single range. That is, a pair of animals that are within some interaction radius can directly interact through our defined behavioural states (see Figure 5.9). These differences are born out of the different 'biological realism' we require the radii to represent. For Couzin *et al.* (2002), part of the motivation was that the zone of repulsion allowed collisions between the particles to be avoided. This feature is not required in our model as avoidance isn't a focus of our approach and we do not model the animals as self-propelled particles. Both approaches provide a boundary, beyond which animals cannot interact and so spurious, distant interactions are prevented. In our case, this has an added bonus in that OU attraction can lead to quite extreme movement behaviours if the distance between interacting animals is large. Having a radius of interaction will mitigate the worst of this effect. An additional perk is that it will reduce the state space that needs exploring (also discussed by Scharf *et al.* (2016)). That is, our behavioural states correspond to interacting with specific animals. If some of those animals are out of range, the corresponding states are effectively removed from the state space at that time.

It is worth noting that our zone of interaction only comes into effect at actual switching times. That is, in order to switch to a subordinate state, the dominant must be in range of the focal animal. However, at non-switching times, the animals can drift apart to distances greater than the interaction radius as it does not seem biologically realistic for the zone to act as a strict cut-off for interaction. Furthermore, the zones only relate to interactive behaviours. Thus, the animals can switch to a BM state regardless of the spatial context.

Much of the literature mentioned above requires user input on what constitutes proximity. Whilst that input can be based on expert advice, the exact nature of that advice can lead to significantly different social networks (Castles *et al.*, 2014; Davis *et al.*, 2018). Therefore, we explore methods with which to infer the interaction radius, both to circumvent that problem and to increase the richness of information we can obtain by analysing animal movement data.

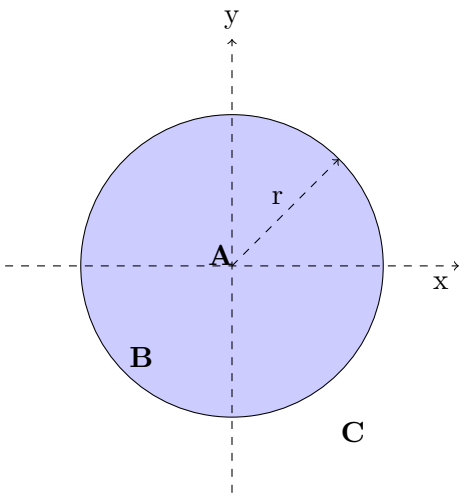


Figure 5.9: A diagram representing the zone of interaction for animal A. Animal B is within this zone and so can interact with A (and vice versa). Animal C is too far away in order to be influenced by A (and vice versa).

5.6.1 Interaction Radius

We explore the posterior of the interaction radius, r, with a random walk MH (with an independent, normally-distributed proposal). In order for new a radius sample to be consistent with the currently estimated movement and behavioural trajectories, we require that new sample, $\hat{r}$, to be greater than the maximum distance between directly interacting animals at the times when they start interacting (according to the current trajectories). If that's not the case, $\hat{r}$ is rejected as at least one state switch would not have been possible with this new radius and so the likelihood of it is 0.

To evaluate the new radius, we need to evaluate the likelihood of the currently sampled behaviours under both the new and old radii. Thus, we update the radius simultaneously with the transition rates. The MH ratio is exactly the same as in the non-zonal case (Section 5.3.3), except the transition rates are now dependent on the radius. That is, transition rates to subordinate states are now set to 0 if the potential dominants are out of range of the focal animal. As a result, there is some interplay between the samples of the interaction radius and the transition rates. For example, if $\hat{\lambda}_{OU-OU} > \lambda_{OU-OU}$ and $\hat{r} > r$ then $\hat{\lambda}_u \geq \lambda_u$ if u is a subordinate state as there is a greater rate of switching to other possible subordinate states and the focal animal may now be in the interaction radius of more animals. If, say, $\hat{\lambda}_{OU-OU} > \lambda_{OU-OU}$ but $\hat{r} < r$, the effect on $\hat{\lambda}_u$ depends on which change had the greater effect. For example, in the scenario where $\hat{\lambda}_{OU-OU}$ is only marginally greater than λ_{OU-OU} but $\hat{r}$ means the focal animal is now out of range of more animals than with r, $\hat{\lambda}_u$ would likely decrease.

5.6.2 New Trajectory Simulation Conditions

Including the zone of interaction only slightly alters the trajectory proposal process (Section 5.3.2). The MH ratio remains identical as detailed in that section, as does the simulation algorithm aside from two alterations. Firstly, we require an extra condition during the movement proposal part of the process. That is, the focal animal i may move to a location which causes another animal's switch (say animal j's switch) to be out of range. This situation can only occur at a time when animal j switches to state S_{Ai} and we resample i's location at this time to one that would've made that switch impossible. If this scenario does play out, we reject the trajectory update as it does not meet the condition that it must remain consistent with the data we are treating as fixed. Secondly, we must take into account which transitions are allowed by the interaction radius when we are sampling state switches. That is, transitions to states that correspond to subordination to an out-of-range animal are set to 0.

5.7 Zone of Interaction Evaluation

5.7.1 Inferring the Radius

In our first steps in exploring this approach, we look at how robustly we can infer the interaction radius. That is, we fit this model to a number of simulations that have interaction zones built in to them. Obviously we are interested in how well we can infer the radius of the zones, but we will cast a keen eye over the estimations of the transition rates as well, as these are now dependent on the radius.

We fit the model to 20 different simulations. As in previous sections in this chapter, these are derived from a single set of parameter values but each simulation contains a different realisation of the behaviours possible under those values. The values used (Table 5.6) were chosen to allow a broad mix of behaviours that are inspired by fission-fusion dynamics. That is, the animals do not form a stable social group. As usual, two BM states are included to represent different speeds of movement but for these simulations we have included four interacting animals. A greater number of animals would provide more scenarios to simulate, however, there is a trade off with computation speed when analysing a sizeable number of data sets. In principle though, a group of four allows for most scenarios — such as two distinct subgroups. Each simulation consists of 100 discrete intervals of two units of time. A value of 1.5 was used for the coefficient of the LDP.

In order to test the inference of the interaction radius, 10 of the runs were initialised with a radius that was too small, whilst the other 10 were initialised to be too big. We set $\lambda_{max} = 0.25$ as the maximum rate of leaving a state in the simulations is 0.18 (providing an animal is in all of the relevant zones). However, some of the rate posteriors (Figure 5.12) indicate that a slightly larger λ_{max} would've allowed the samples to explore more freely. Again though, some consideration was given to the computational needs of analysing 20 simulations. Trajectories were updated over 2 to

Parameter	Value
α	0.2
σ	1.1
ρ_{slow}	0.4
ρ_{fast}	2.5

Parameter	Value
λ_{OU-OU}	0.03
λ_{OU-BM}	0.02
λ_{BM-OU}	0.05
λ_{BM-BM}	0.03
r (interaction radius)	12.0

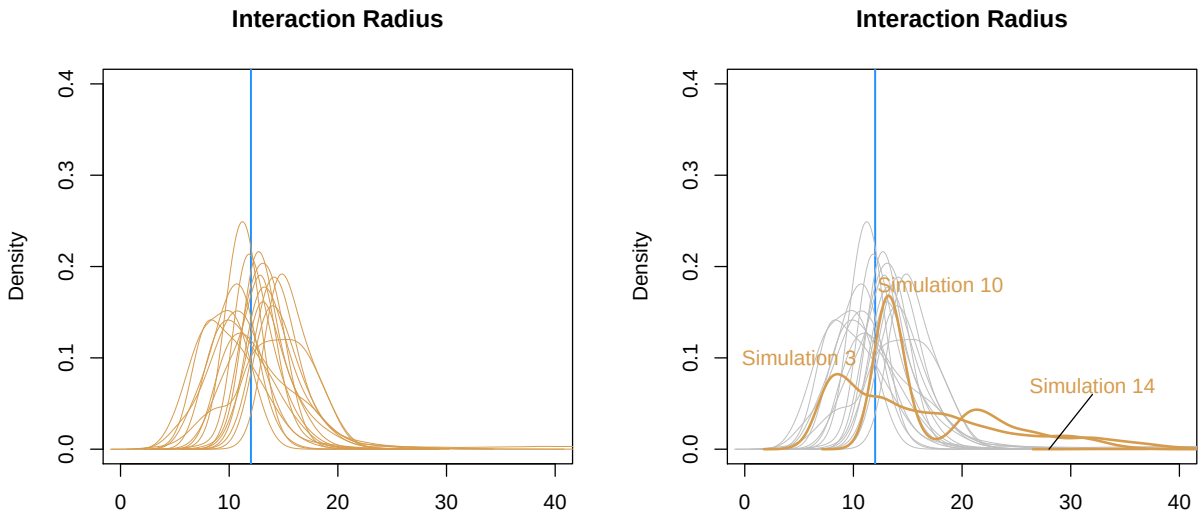
Table 5.6: Parameter values used for the simulations containing the zone of interaction.

6 observations.

Each run was carried out for 36 hours and, for all of the results in this section, we analyse the second half of the outputs. Whilst this was a sufficient burn-in for all cases, each run could have been tuned more effectively. Each run was tuned identically, rather than arduously tuning each one separately, which in theory is a suitable approach as all simulations are derived from the same parameter values. However, as the behaviours exhibited in each simulation are different, our one-size-fits-all approach is less than optimal. The results that follow could therefore, in theory, be enhanced (an increased ESS for example) under more optimal settings.

Figure 5.10 plots the posterior distribution of the radius from each of the simulation runs. The majority of them are consistent with the true value (Figure 5.10a) — the difference between the point estimates is the result of the long tails that occur in some of the posteriors. However, there are some notable exceptions (Figure 5.10b). The spurious results can be better explained with the help of the trace plots in Figure 5.11. Both the posterior and trace plot for simulation 10 indicate some evidence of bimodality. Whilst analysing the modal state estimations for iteration ranges 25k-50k ('early', larger radius), 150k-250k ('middle', smaller radius) and 300k-350k ('late', larger radius), we found identical interaction estimated in both the early and late iteration ranges between two animals — an interaction that is beyond the radius samples during the middle range. This interaction commences when those two animals are roughly 19 units of distance apart — a distance in line with larger radius samples. Thus, the evidence of those two animals interacting in the data (albeit, not true) led to the larger radius samples. The radius samples for simulation 3 cover a wide range of values. It may be that there is simply some evidence of interaction further afield in the data. However, the meandering trace plot for this simulation indicates that a tuning (with regards to the radius) more specific for this data set would prevent the samples wandering so freely.

Simulation 14 produces the most cause for concern. The trace plot indicates the samples are mixing well until around iteration 170k. This explosion of radius estimation highlights a flaw in how we had setup up the MCMC process. We used an uninformative, flat prior for the radius and so only the data contributes to the posterior. However, if the radius sample jumps above the maximum distance between any two animals in the data (the red line in the trace plots), the radius no longer has any impact on any likelihood terms and thus it is free to randomly walk off. In the case of



(a) (b)

Parameter	True Value	Mean	Median	Mode	ESS
r	12.0	12.8	12.0	11.4	231

Figure 5.10: Posterior distributions from 20 runs for the interaction radius. (a) contains all the posteriors except for the those of simulations 3, 10 and 14. (b) overlays those spurious results on top of the same posteriors in (a). The blue vertical line indicates the true value used. The point estimates and ESS are averages from the 20 runs, except for simulation 14.

simulation 14, it did. In the cases of simulations 1 and 16, we were fortunate that the radius samples walked back into a meaningful range. Other runs on simulation 14 with different seeds produced a posterior more consistent with those in Figure 5.10a. A flat prior in this instance is therefore not a sensible choice as not all of the parameter space is reasonable. A more sensible prior would discourage radius samples beyond that maximum proximity — an idea that we explore in Section 5.7.2.

The posteriors of the transition rates for all runs are shown in Figure 5.12. They are largely confident and consistent with the true value. λ_{OU-OU} appears to be the most difficult to infer as for this transition to be possible an animal must be in the interaction zone of two others (its current and prospective dominant) — a relatively rare occurrence in data sets of 4 animals. Therefore, some of the simulations will contain fewer subordinate-to-subordinate switches than we would like in order to confidently estimate that rate. The variation in the point estimates is indicative of this — with some wide, uncertain posteriors inflating the mean and median away from the mode.

A potential, unwanted side-effect of introducing an interaction radius is that it may hamper our ability to correctly infer the more distant true interaction. That is, as the radius samples bounce

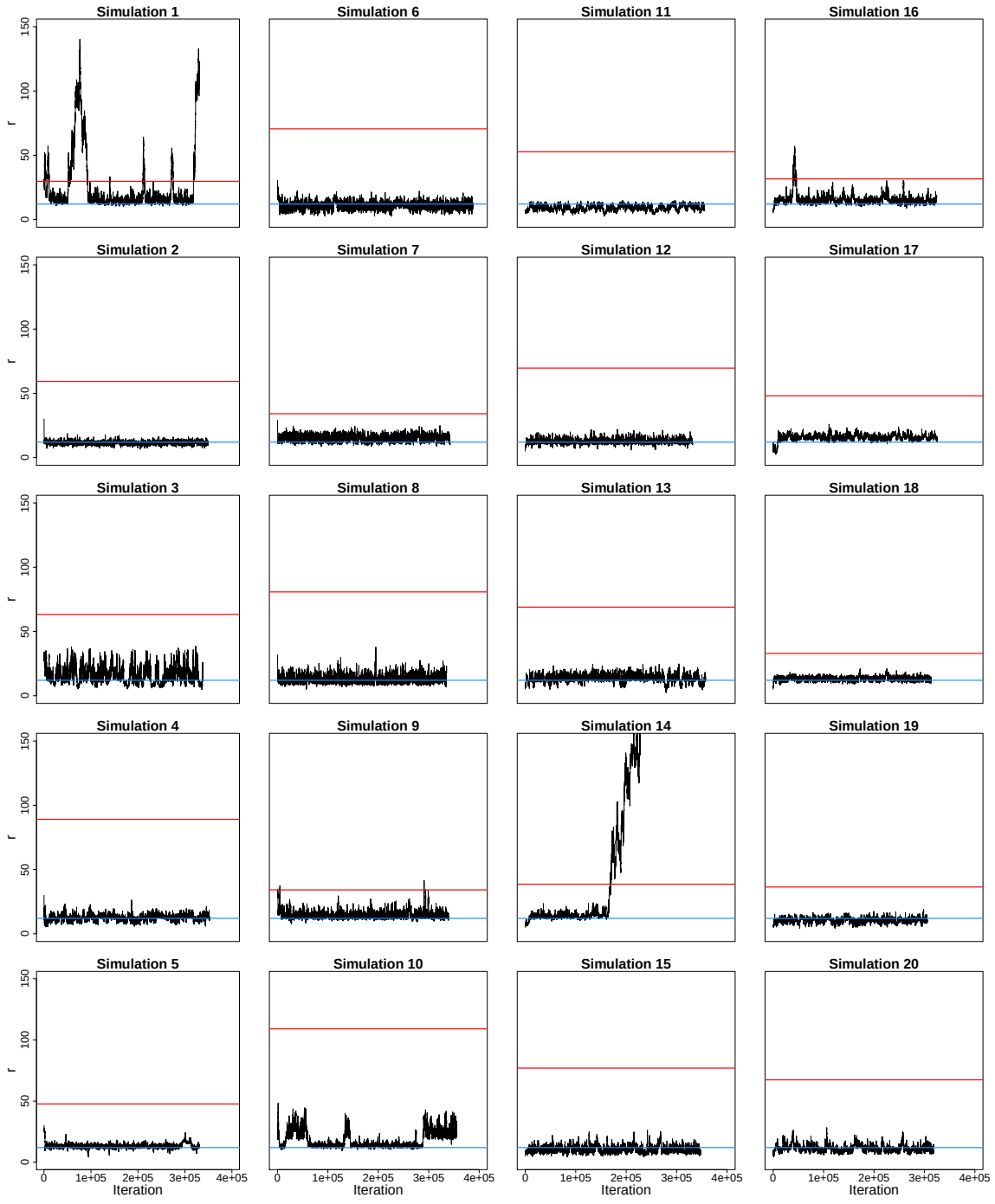
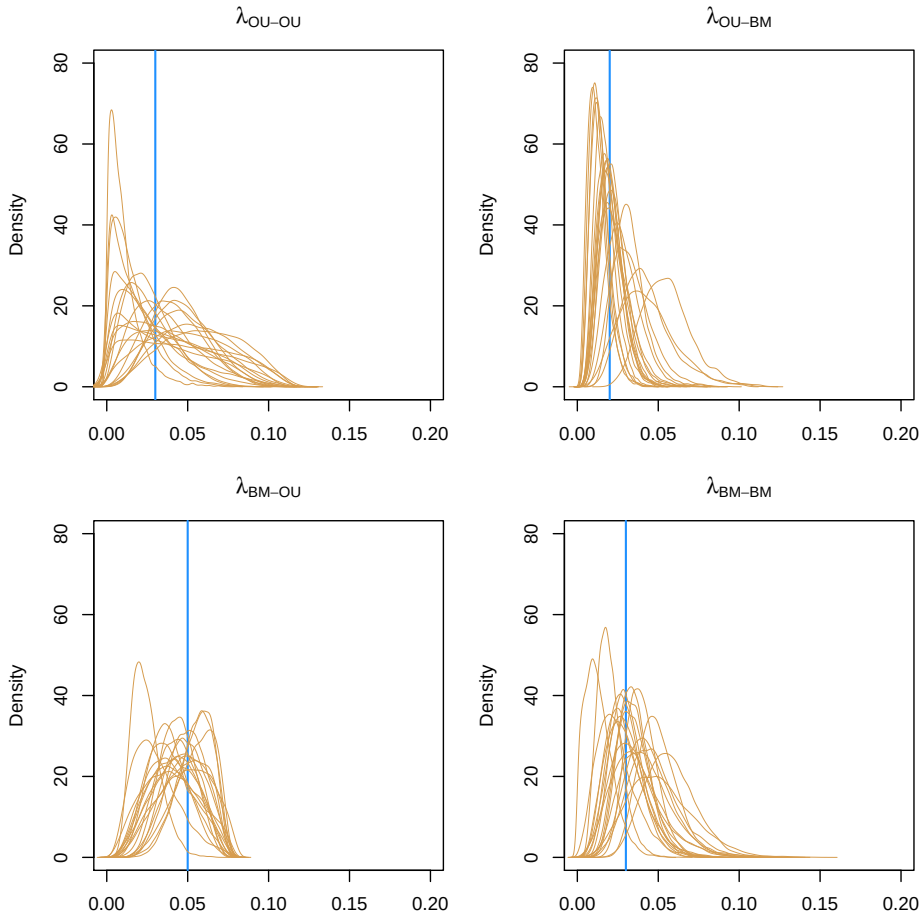


Figure 5.11: Trace plots of the interaction radius samples from each of the 20 simulation runs. The blue line indicates the true radius whilst the red line indicates the maximum distance between any two animals throughout that simulation. The y axis has been truncated to 150 as little information is contained beyond that — the radius samples in simulation 14 continue to randomly walk.

around, some of those samples may deem interaction that we know to be true in the simulation to be overly distant — thus, our ability to infer the true behaviour is restricted. Figure 5.13 displays the interaction posteriors from all simulation runs at each observed time (for simplicity). An



Parameter	True Value	Mean	Median	Mode	ESS
λ_{OU-OU}	0.03	0.037	0.034	0.022	432
λ_{OU-BM}	0.02	0.025	0.024	0.022	474
λ_{BM-OU}	0.05	0.044	0.044	0.044	617
λ_{BM-BM}	0.03	0.036	0.035	0.032	585

Figure 5.12: Top: posterior distributions from 20 runs for the four transition rates. The blue vertical line indicates the true value used. Bottom: a summary of the transition rate results. Point estimates and the ESS are averages from the 20 runs.

interaction posterior in this context is the posterior probability of two animals directly interacting at times of true direct interaction, regardless of ordering. We overlay these posteriors with the 20th, 40th, 60th and 80th percentiles for a number of proximity ranges: each third of the true radius and all distances beyond that. The aforementioned side-effect would present itself through smaller percentile values for the larger proximity ranges as the inference of true interaction was restricted. However, this is not the case as the percentiles selected are consistent across the ranges. Note, the zone of interaction only limits switching to a subordinate state — interaction can persist outside of it as seen in this figure. Whilst persisting interaction beyond the radius was rare (as subordinates

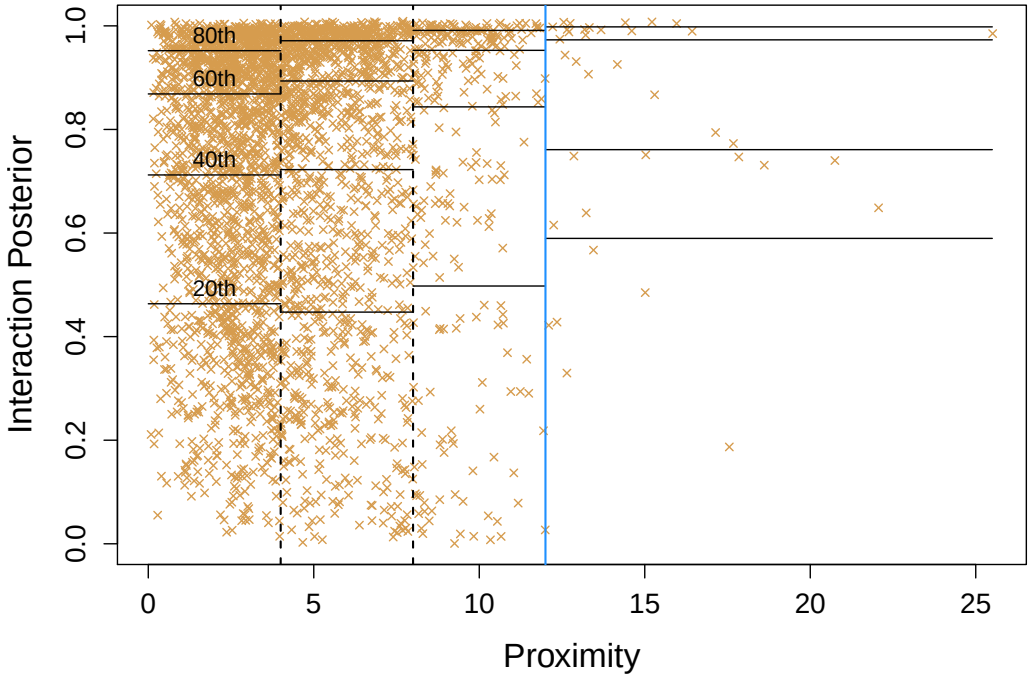


Figure 5.13: The *interaction posteriors* (the posterior probability of two animals directly interacting at true direct interaction) are plotted against their *proximity*. The results encompass all simulation runs and come from the state estimations at observations. The horizontal lines indicate the 20th, 40th, 60th and 80th percentiles for the corresponding proximity ranges. The blue line represents the true value of the interaction radius. Points beyond that proximity result from animals moving outside the zones during persistent interaction (the zones only limit switching). Both axes have been jittered in order to help display the density of the points.

either gravitate back to the dominant or eventually switch to a BM state), it is important to show that the radius does not act as strict cut-off point for interaction, and these wanderings are still confidently inferred.

5.7.2 Interaction Radius Prior

As mentioned above, a flat prior for the interaction radius is not a sensible choice. We have therefore trialled a prior distribution which is relatively flat for meaningful values of the radius, but whose density gradually reduces to 0 beyond that. Namely, a half-normal distribution with a sufficiently large standard deviation.

Figure 5.14 displays the trace of the radius samples whilst fitting this version of the model to the 20 simulations. The standard deviation was set to 110 units of distance, which encompasses all observed distances between any two animals in all of the simulations. Therefore, the half-

normal distribution will be relatively flat for meaningful values of the radius, but the radius chain is discouraged from wandering off too far. Ignoring simulation 1 for now, this figure shows fewer instances where the radius sample jumps above the maximum distance line for extended periods when compared to Figure 5.11. In particular, see the plots for simulations 14 and 16.

Simulation 1 represents a tricky case as the maximum distance between two animals (beyond which the radius doesn't impact any likelihood terms) is not much greater than the true radius and the range of values for which there is some evidence of. The radius sample can therefore quite easily jump above that maximum distance. This is true even with our new prior as, given the prior's large standard deviation in the context of this data set, the prior will have little impact on the evaluation of radius samples around that part of the parameter space. However, a number of other simulations represent a similar situation (simulations 7, 9, 16 and 18), but they didn't have this problem. Another explanation for the erratic radius samples of the simulation 1 run is that there just isn't strong evidence in the data for a zone of interaction. Nonetheless, we can see in Figure 5.14 the samples are now reluctant to wander off too far due to the prior. Through longer runs, we would therefore be in a better position to ascertain exactly how much evidence there is of an interaction radius as the samples have a greater chance of returning to meaningful values — thus avoiding the situation of simulation 14 in the flat prior case.

Figure 5.15 compares the posterior distribution of the radius when using a flat prior and when using the half-normal prior. Aside from the already discussed runs (regarding simulations 1 and 14), all posteriors are consistent — including the biomodality found in simulation 10. Furthermore, the prior did not negatively restrict the radius sampling as the ESS, averaged over all runs, remained consistent — 245 compared to 231 previously.

Strictly speaking, the standard deviation of the prior should not be informed by the data — which is broadly in line with our approach in setting the prior above. However, it may be reasonable in some circumstances to set the standard deviation of the prior to be a bit more restrictive than what we have used thus far. For most of our simulations, the standard deviation was much greater than any observed distance between animals. An alternative approach could be to use the maximum distance observed (or thereabouts) for the standard deviation. The half-normal prior will remain relatively flat for all values of the radius that impact the likelihood terms, but it will discourage the samples from wandering off to an even greater extent. We ran the simulation inferences again with exactly that approach for the prior. The results obtained (Figure 5.16) are consistent with previous results, except that this time the samples for simulation 1 are constrained further and thus we are able to see the radius values for which there is some evidence. However, this more restrictive prior may just be forcing the inference process to look for evidence of a zone of interaction, even if the data doesn't support it — as what might be happening with the simulation 1 run in Figure 5.16.

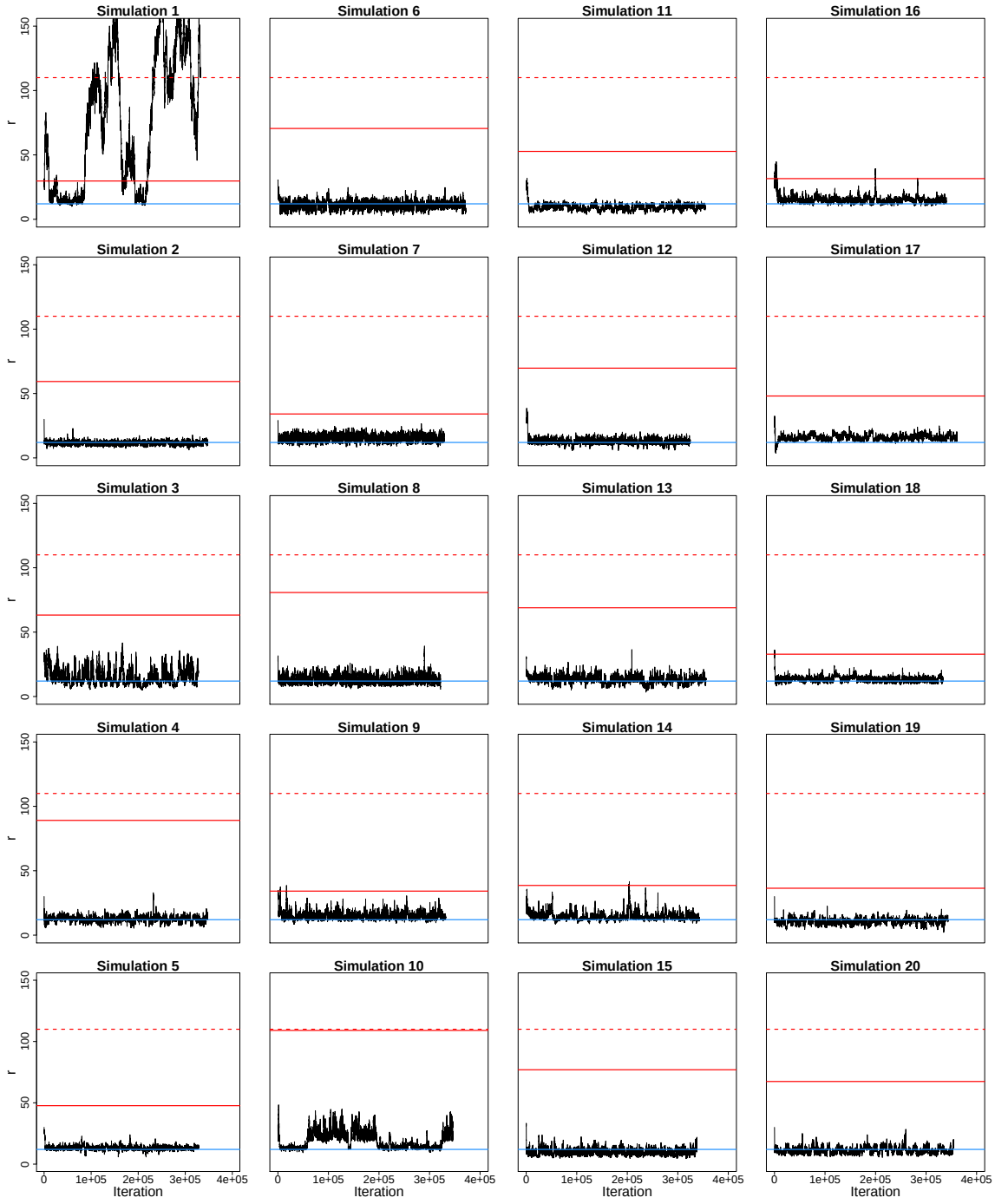


Figure 5.14: Trace plots of the interaction radius from each of the 20 simulation runs when using a half-normally distributed prior. The blue line indicates the true radius; the red line indicates the maximum distance between any two animals throughout that simulation; the dotted red line indicates the standard deviation of the prior distribution. The y axis has been truncated to 150 as little information is contained beyond that.

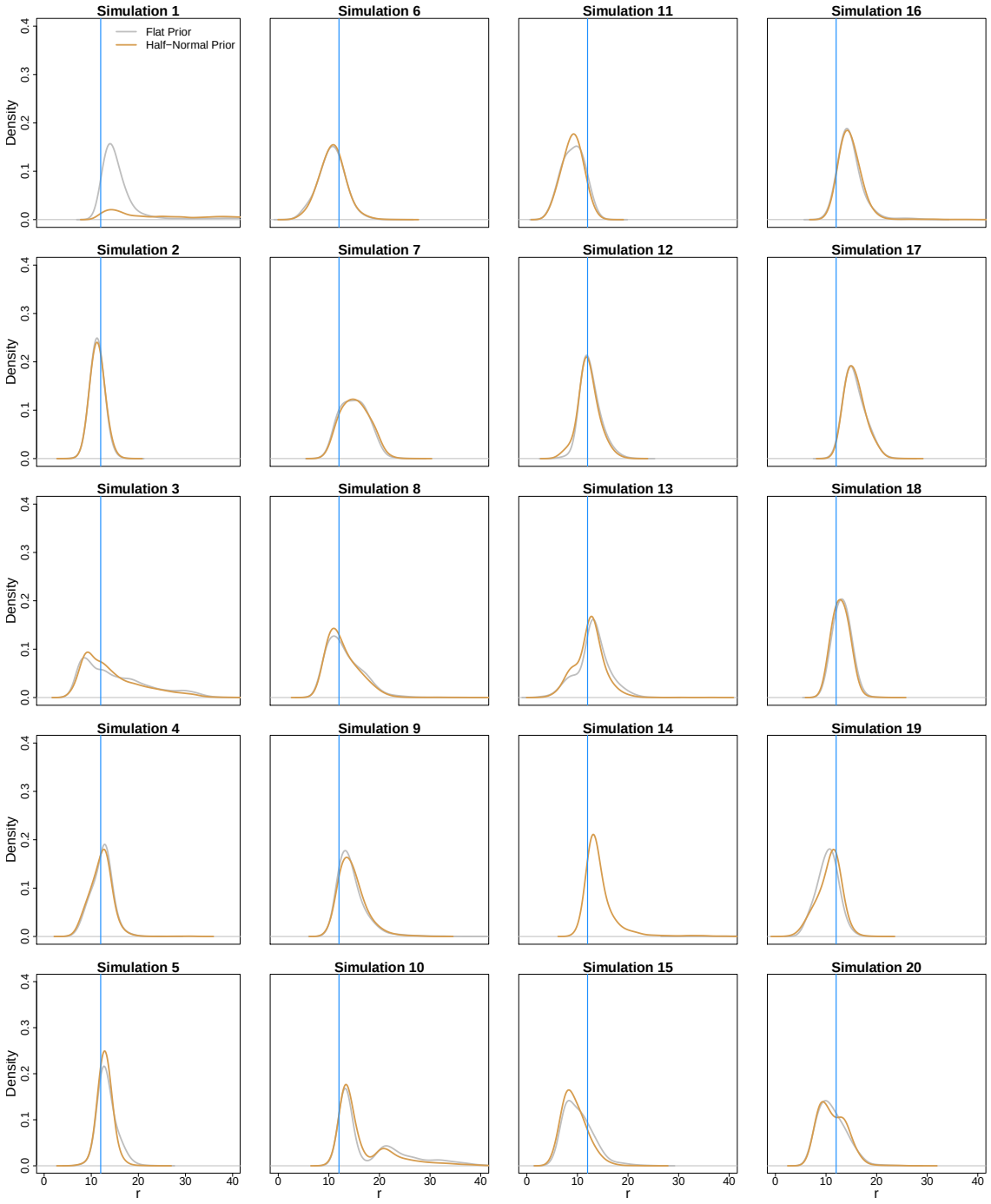


Figure 5.15: The posterior distributions of the interaction radius from each of the 20 simulation runs. The silver posteriors are the products of using a flat prior for the radius, whilst the gold posteriors are obtained when using the half-normally distributed prior with a standard deviation of 110. Note: the posterior for simulation 14 from the flat prior run does not appear in this range, neither does most of the posterior for simulation 1 from the half-normal run.

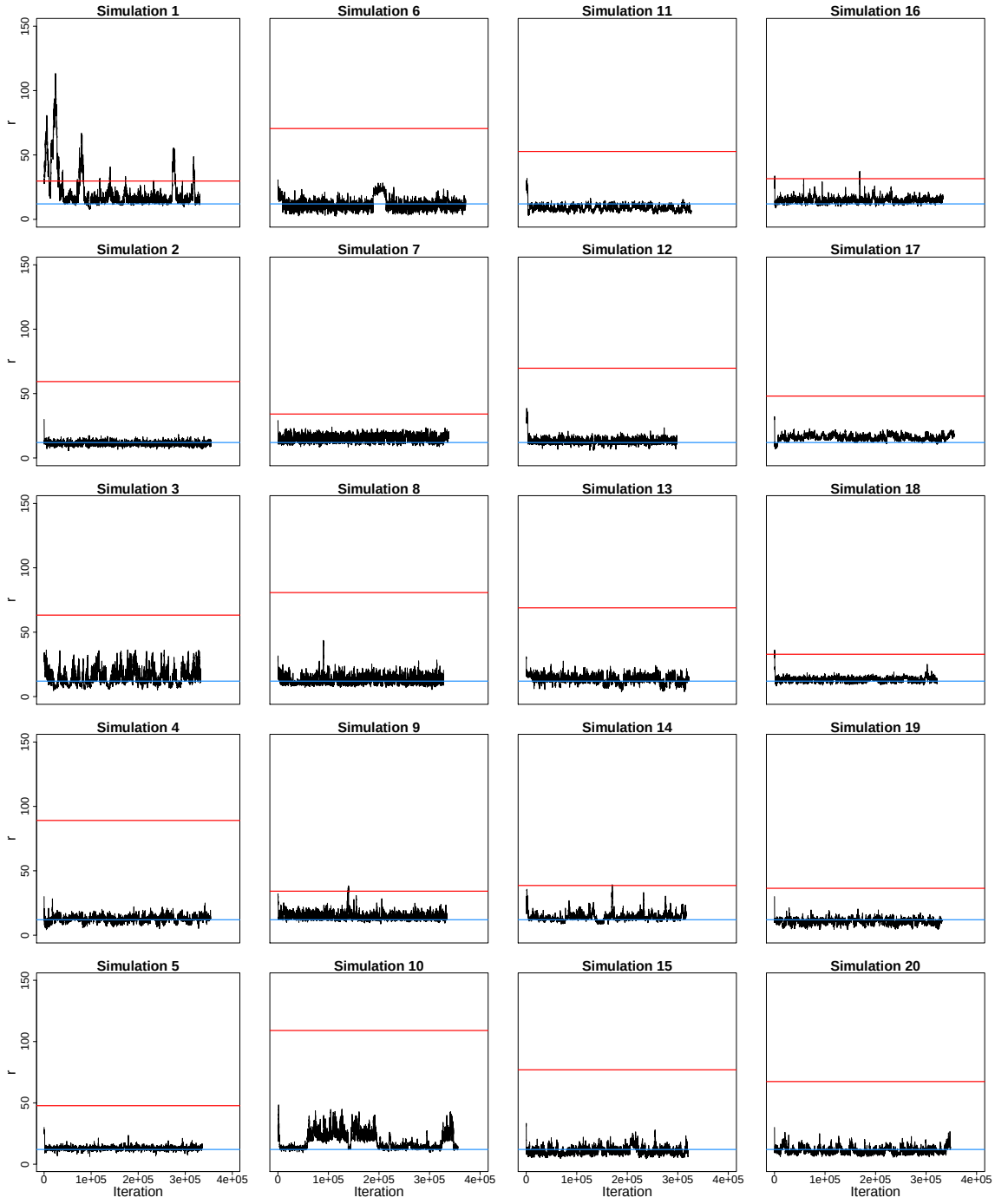


Figure 5.16: Trace plots of the interaction radius from each of the 20 simulation runs when using a half-normally distributed prior — the standard deviation of which is set by the maximum distance between any two animals throughout that simulation (the red line). The blue line indicates the true radius.

5.8 Revisiting the Baboons

5.8.1 Setup

In order to trial our extended model on real data, we revisit the baboon troop we analysed in Chapter 4. As the subset of data we analysed was chosen because it included some directional conflict, through which the group temporally splits up, it represents a useful case study for these extensions also. Whilst we are interested in the insight that can be gained from including the zone of interaction in the model, we have also carried out a number of runs without it so that we can evaluate its impact on the model fitting process.

Our results from Chapter 4 found that thinning the data still allowed us to obtain useful information from the model whilst reducing the computational effort of fitting it. However, the amount of thinning does change the scale of behaviours we can analyse. Furthermore, our results in Chapter 4 alluded to there being some persistence in the movement of leading and independent animals. That persistence is another factor to consider when deciding on how much data thinning to carry out. That is, the more data we retain, the smaller the average time interval will be between a sampled switching time and the next observed time. As a result, the LDB proposal distributions will be more 'guided' towards the next observed location, an effect that is particularly useful in our scenario of modelling the movement of leading/independent animals with BM despite the suggested persistence. We have therefore decided to thin the data by a factor of two in order to reduce the computational effort of the inference process, but whilst still retaining enough data to analyse fine-scale behaviours and provide a good level of guidance for the LDB. After a few tuning runs, a value of 1.1 was used for the diffusion coefficient of the LDP.

We are still modelling the movement of self-driven baboons with BM, even though previous analysis suggests it is insufficient, as the temporal range of the data set we are analysing (15 minutes) does not really lend itself to modelling that movement with an OU process. That is, it is not long-term enough to establish a fixed site of interest. Moreover, investigating that type of behaviour is not the focus of our analysis in this chapter.

During the tuning process, it became clear that the model was mixing slowly. This was evident from the radius trace plots as a single run would spend half the time in one range of values and half the time in a different range (see Figure 5.17). We briefly trialled using a greater standard deviation for the radius proposal distribution, and a distribution with heavier tails, in order to see if the limiting factor was the (lack of) ability of the proposal distribution to traverse the parameter space. However, neither method facilitated better mixing. In fact, both methods only hindered the inference process. Increasing the standard deviation of the proposal distribution (and to a lesser extent, increasing the heaviness of the tails) led to more radius samples being automatically rejected as they were more likely to fall foul of the condition that they must be large enough to facilitate all currently sampled interaction (recall Section 5.6.1).

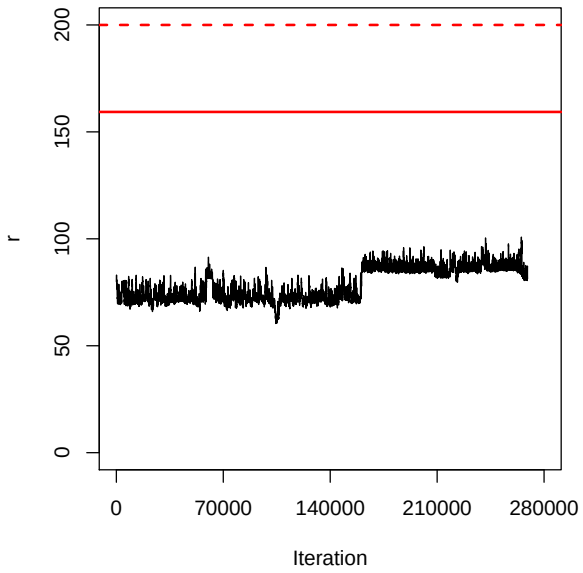


Figure 5.17: The trace plot of the interaction radius from a tuning run .The solid red line indicates the maximum distance observed between any two baboons whilst the dotted red line indicates the standard deviation of the prior distribution.

The radius samples appeared to mix well enough then, but the trajectories that would backup the new radius were not being proposed and accepted frequently enough — despite obtaining a good acceptance rate for trajectory updates over 2 to 10 observations. For instance, if the latest radius sample was large enough to allow some new interaction, that new interaction wouldn't be proposed and accepted quick enough before the radius value dropped back down again. A more fruitful change to our inference process was therefore to increase the number of trajectory updates per iteration. Thus far, we have tended to set the number of trajectory updates so that roughly every interval for every animal will be resampled per iteration (on average). Now though, with the added complexity of spatial heterogeneity and the zone of interaction, we have doubled that in order to improve the above situation. Whilst that does have a knock-on effect on the computational effort required per iteration, and so reducing the number of iterations possible, we can see from Figure 5.18 that increasing the number of updates per iteration has led to an improvement in mixing. However, it is also clear from these plots that many more iterations would be required in order to gain a complete picture of the radius posterior from a single run.

As the above mixing problem emanates from the trajectory updates, as opposed to the radius sampling, the same problem will occur with the behavioural state posteriors — more iterations will be required for any one run to be insightful on its own. During this section then, in order to obtain well-explored posteriors, we have collated the output of 10 runs that include the zone of interaction (*zonal* runs) and, separately, we have also collated the output of 10 runs that don't (*non-zonal* runs). That is, the radius and state chains (post burn-in) from the 10 zonal runs will be combined

to produce the respective posteriors (and similarly for the non-zonal runs).

5.8.2 Zonal vs Non-Zonal

Curiously, including the zone of interaction did not increase the burn-in period. Both the zonal and non-zonal runs had a burn-in period of approximately 15k iterations on average. Though, determining the burn-in period is not an exact science, and that is even more so the case here. For example, we judged the radius chains as having passed the burn-in period when they had spent an extended period at a range within 60m to 95m as all runs suggest that's where the evidence of a radius lies. A similar approach was taken for the movement and rate parameters as newly sampled behaviours can cause slight deviations in their traces — though to a far lesser extent than seen with the radius. Initialising the radius to be excessively small typically produced the longer burn-ins as larger samples of the radius need to be locked in by the coinciding behaviours.

Figure 5.19 displays a noticeable difference between the zonal and non-zonal transition rate posteriors. Namely, λ_{OU-OU} and λ_{BM-OU} are estimated to be greater when the zone of interaction is included. This is to be expected as, in the zonal runs, animals can only switch to an OU (subordinate) state when they are within range of a potential dominant. As such, the likelihood of these rates are no longer evaluated at times when switching to a subordinate state is unlikely (i.e. when an animal is beyond any reasonable interacting proximity with its peers).

The λ_{BM-BM} posteriors are the most erratic with regards to the transition rates. This is not an artefact of the zone of interaction as this is true for both the zonal and non-zonal runs. Nor does it appear to be an artefact of the model in general as the λ_{BM-BM} rate posteriors from a number of runs in Section 5.4.2 are highly consistent with each other (Figure 5.2b). It is not clear why the λ_{BM-BM} posteriors in Figure 5.19 are less aligned with each other than the posteriors of the other rates are. Some exploratory analysis suggests it is due to some of the runs containing differing levels of evidence for some quick BM to BM transitions (that is, a short period of time in a BM state immediately before or after a spell in another BM state). The varying levels of evidence for these transitions will most likely be a casualty of the slow mixing of the trajectory updates.

The movement parameter posteriors are very consistent between the zonal and non-zonal runs as seen in Figure 5.20. α is estimated to be slightly larger in the zonal runs, indicating a stronger level of interaction. Whilst the most distant interaction that is inferred in the non-zonal runs is not necessarily weak interaction, that is the case here. Eliminating that will bolster the average strength of the remaining interaction.

As mentioned previously, a single run will not provide us with the complete picture of the radius posterior. This can be seen in Figure 5.21. Whilst each run is consistent in that the peak of their posterior and that the majority of their evidence of the radius is within roughly 70m to 80m, they are not consistent outside of that range. It appears most runs also find some evidence for the radius to be within 80m to 100m. Due to the slow mixing of the trajectory updates, a single run is not

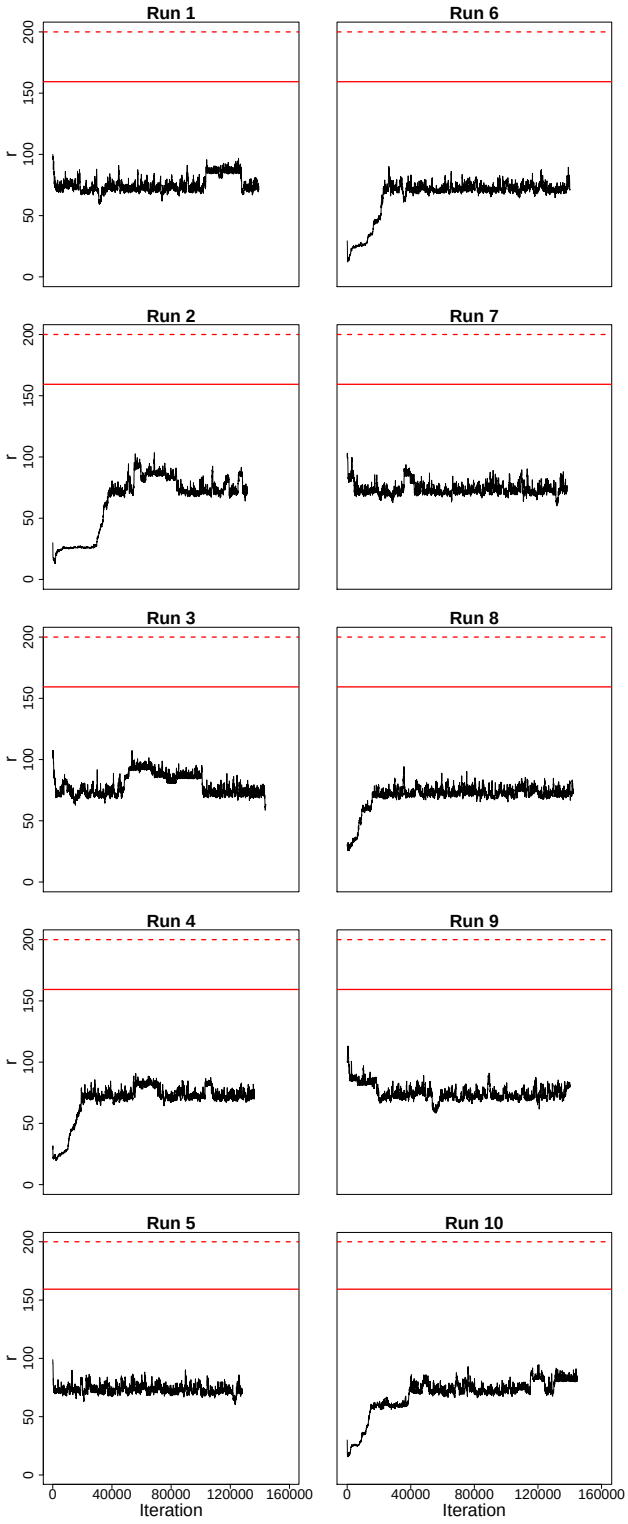
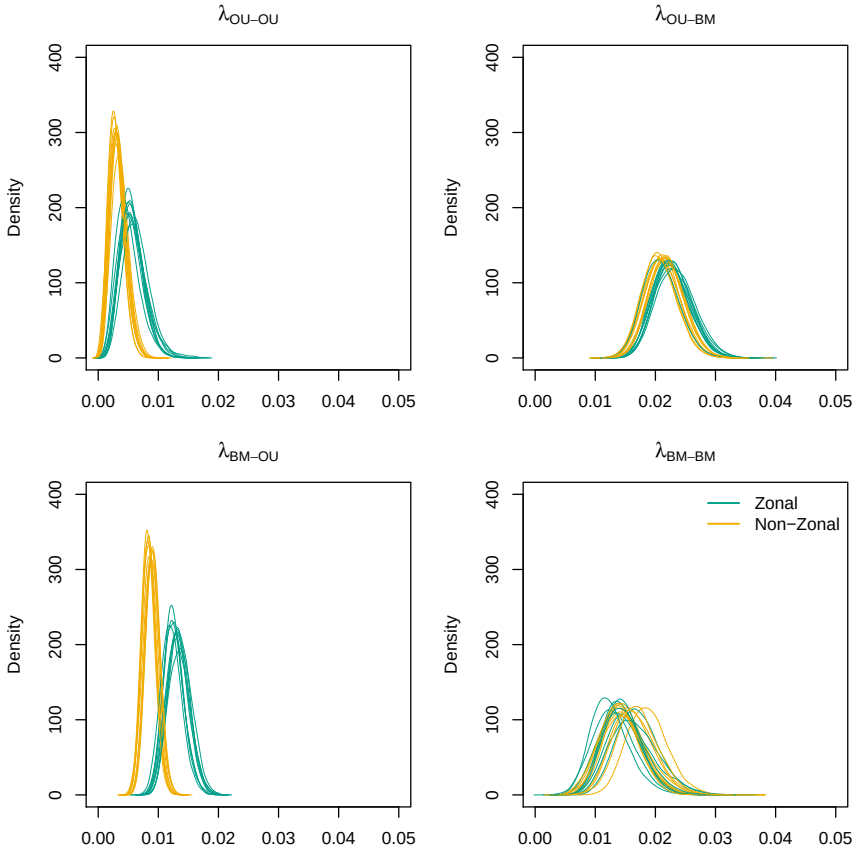


Figure 5.18: Trace plots of the interaction radius from each of the 10 baboon runs that included the zone of interaction. The solid red line indicates the maximum distance observed between any two baboons whilst the dotted red line indicates the standard deviation of the prior distribution.

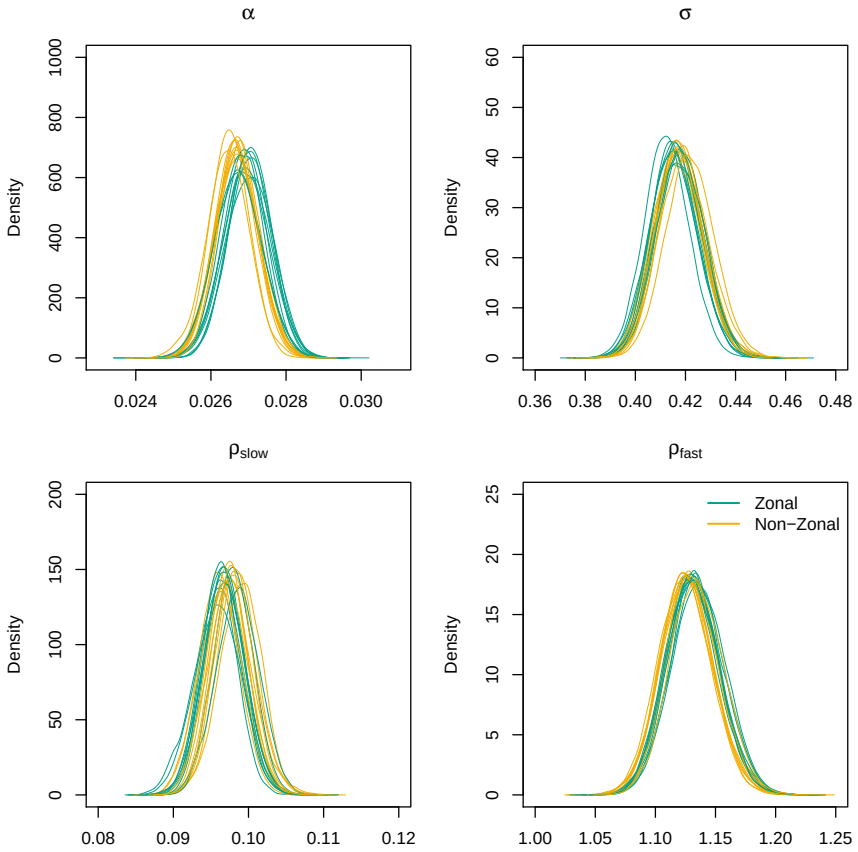


Parameter	Non-Zonal		Zonal	
	Point Est.	ESS	Point Est.	ESS
λ_{OU-OU}	0.0033	236	0.0057	253
λ_{OU-BM}	0.0214	520	0.0226	445
λ_{BM-OU}	0.0087	593	0.0130	406
λ_{BM-BM}	0.0155	327	0.0148	326

Figure 5.19: Posterior distributions for the four transition rates from each of the 10 non-zonal and 10 zonal baboon runs. The table contains a summary of the transition rate results. Point estimates and ESS are averages over the 10 runs — all averaging methods produce very similar point estimates.

able to properly explore the entirety of this range. Through combining the (post burn in) chains of the 10 zonal runs, we can see that the various smaller peaks in the individual posteriors in the 80m to 100m range simply result in a heavy tail for the combined distribution. Despite this heavy tail dragging up the mean, all points estimates are fairly consistent.

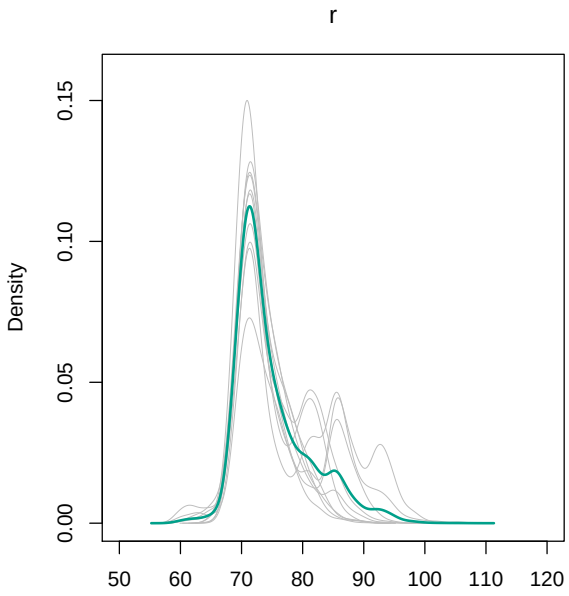
Figure 5.22 provides a glimpse of how the zone of interaction interplays with the behavioural states. For instance, baboon 5 moves within interaction range of baboon 3 sometime between observations 101 and 111 — after which they quickly begin to interact. Both the radius and behaviours shown are based on the modal estimations.



Parameter	Non-Zonal		Zonal	
	Point Est.	ESS	Point Est.	ESS
α	0.027	349	0.027	295
σ	0.418	410	0.416	427
ρ_{slow}	0.098	293	0.097	291
ρ_{fast}	1.128	1072	1.132	1000

Figure 5.20: Posterior distributions for the four movement parameters from each of the 10 non-zonal and 10 zonal baboon runs. The table contains a summary of the movement parameter results. Point estimates and ESS are averages over the 10 runs — all averaging methods produce very similar point estimates. Note the different densities for each of the parameters.

Figure 5.23 displays the modal state estimation for each baboon at each observed time for both the zonal and non-zonal analyses. As we can see, there is a large degree of consistency between the zonal and non-zonal mode. Whilst there is the occasional brief discrepancy or lag in alignment, the only major difference was at point A (around observation 270 for baboon 5). By major difference, we mean the the posterior probability of the modal state in the zonal analysis was not within 0.2 of the posterior probability of that same state in the non-zonal analysis (and vice versa). The 0.2 boundary is arbitrary, but it is to be expected that there will be some difference between the runs — particularly as the difference in the movement and rate parameter estimations will have a



Parameter	Mean	Median	Mode	ESS
r	74.9	73.0	71.1	533

Figure 5.21: The grey plots represent the posterior distributions of the radius from each of the 10 baboon runs that included the zone of interaction. The overlaid bolder, teal plot is the distribution resulting from combining the (post burn in) chains of those 10 runs. The point estimates and ESS are taken from the combined distribution.

knock-on effect on the behaviours proposed and accepted. However, discrepancies larger than that are worth investigating. For reference, most discrepancies were less than 0.09. It transpires that, in the non-zonal analysis, the brief period where baboon 5 is subordinate to baboon 9 (at point A) is the only interaction strongly inferred that occurs at a distance that is beyond the interaction radius as estimated in the zonal analysis. Or at least, it's the only interaction with a proximity (approximately 84m to 89m) that is within or beyond the tail of the radius posterior. The zonal analysis still infers baboon 5 is influenced by another, but this time it is baboon 3 — albeit with less certainty and the posterior probability of this is smaller than 0.5. The overall change in group structure is shown in Figure 5.24. It may seem odd that the zone of interaction has caused baboon 5 to change subgroup. However, there are numerous other baboons in the troop that we have not analysed. If we had analysed them all, we may have found that all of the relevant baboons are in fact in the same subgroup (e.g. baboons 3 and 9 share a dominant) and so we would have merely inferred a new dominant for baboon 5. Though, in this case, an exploratory look at the full data suggests that baboon 5 being subordinate to baboon 9 in the non-zonal case is a spurious result.

Whilst the non-zonal analysis didn't infer much interaction beyond the bulk of the interaction radius posterior (obtained in the zonal analysis) in terms of the modal state, Figure 5.25a shows that the

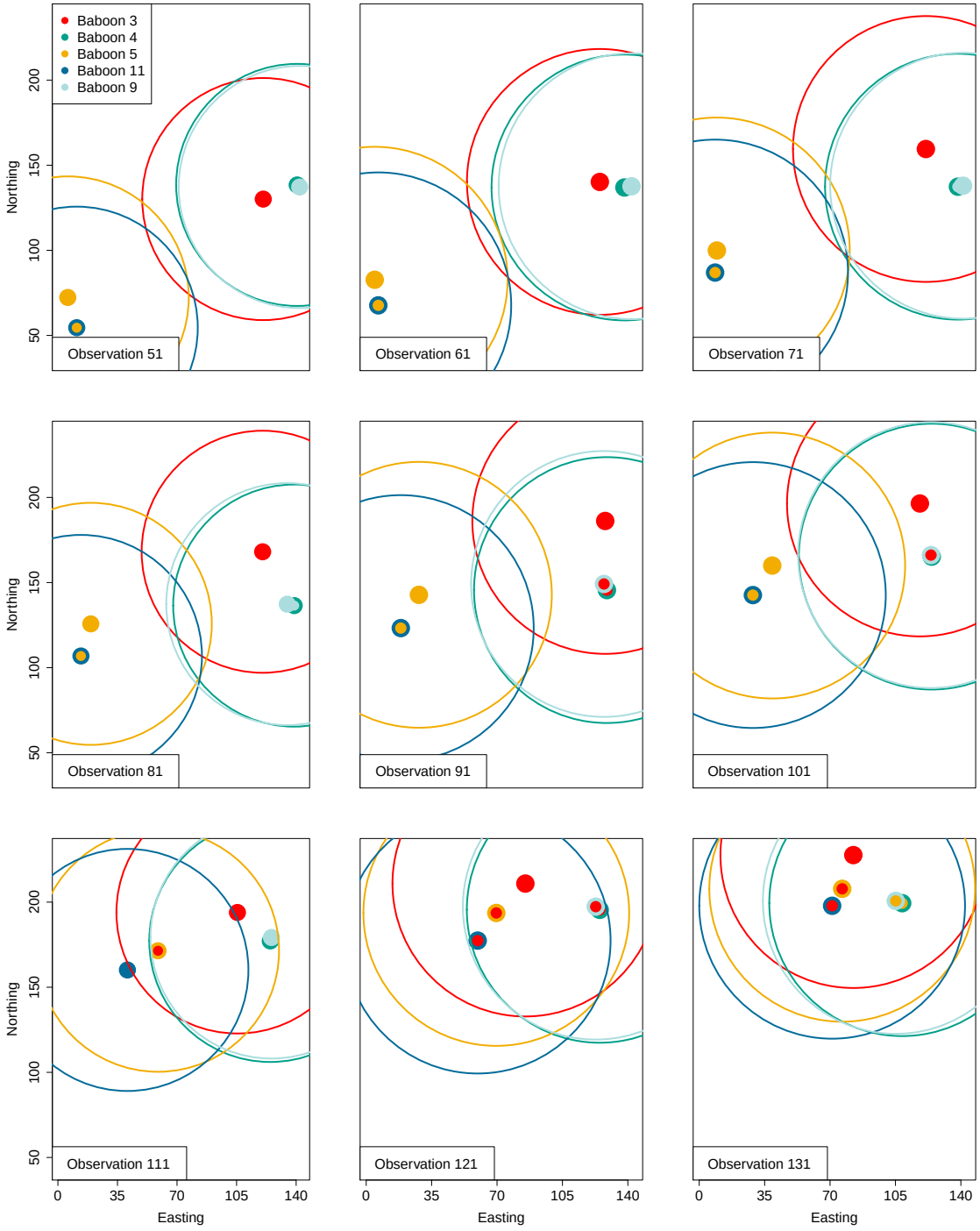


Figure 5.22: Each frame plots the locations of the five baboons at the indicated data point. The data has been annotated with the modal interaction radius and behavioural states at those times. The coloured in circles are the baboons. The border represents that baboon according to the legend. The colour of the filling represents which baboon it is subordinate to. If the filling is the same colour as the border, that baboon is in a BM state. The large, hollow circles represent the zone of interaction for the baboon of the same colour.

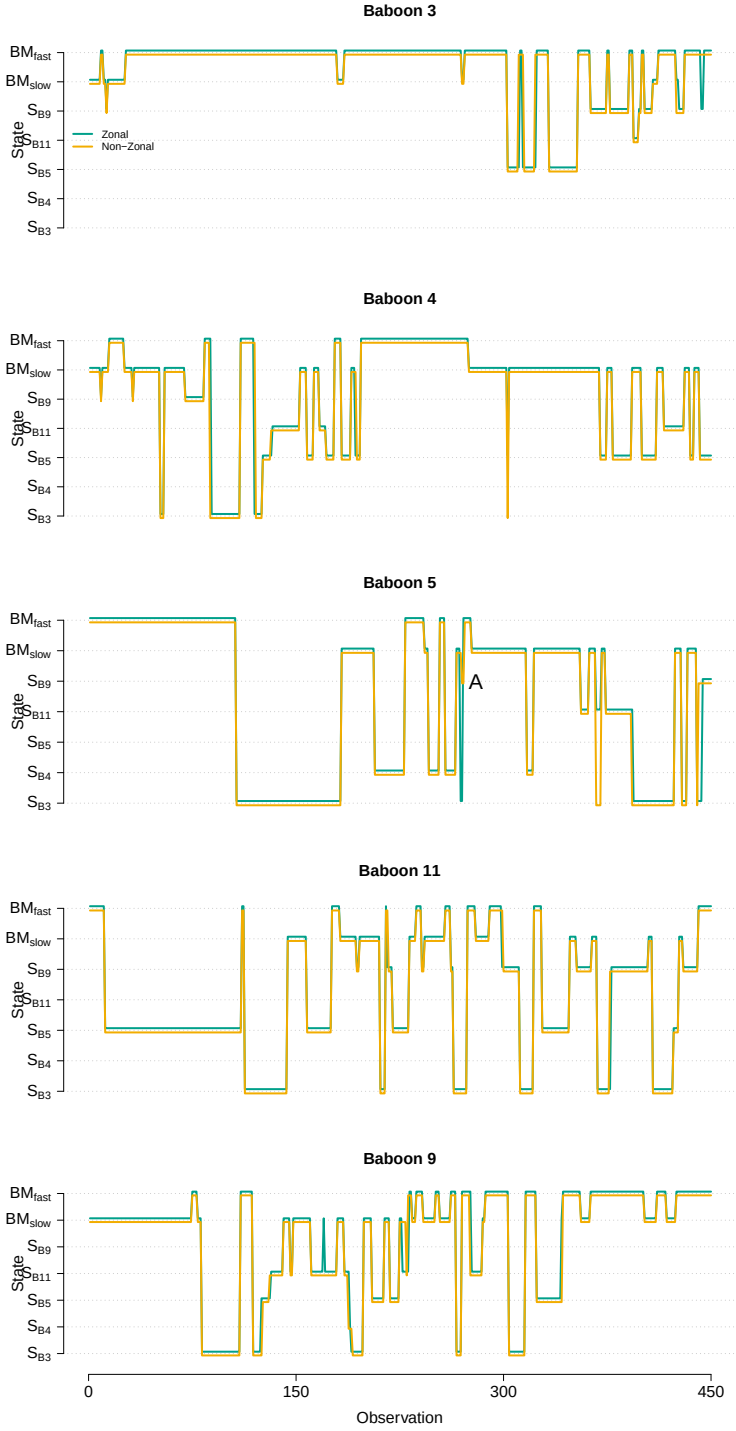


Figure 5.23: The modal state estimation for both the non-zonal and zonal baboon analyses. The modal state was calculated after the output from each of the 10 non-zonal and 10 zonal runs were collated. The lines have been shifted to avoid overlap. Label 'A' highlights the only major difference.

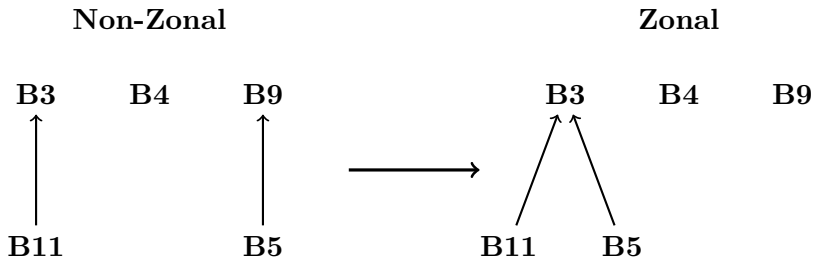


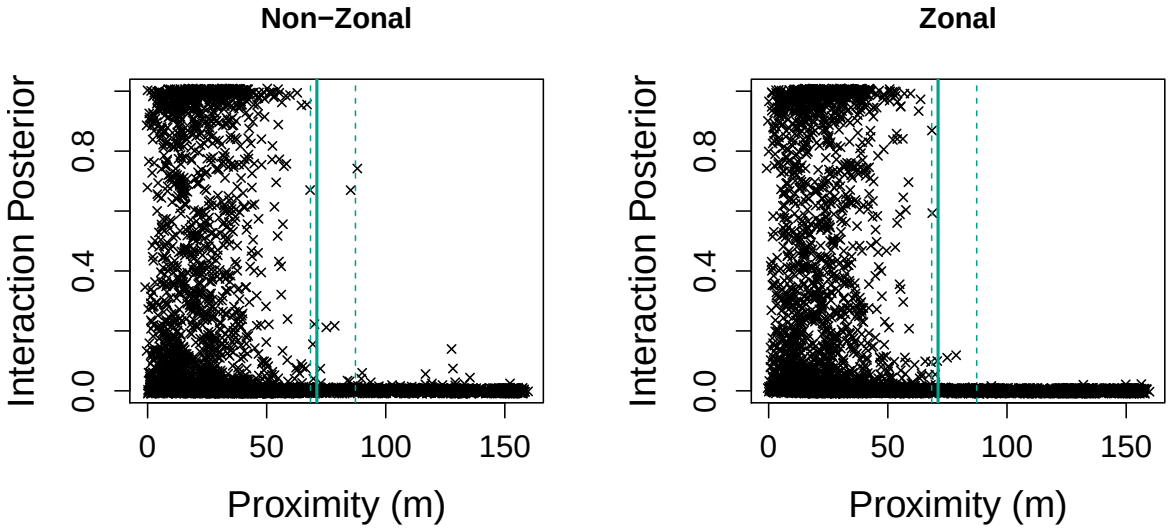
Figure 5.24: The difference in the estimated social structure between the non-zonal and zonal runs (based on modal state estimation) at point A in Figure 5.23. Each node represents a baboon and unconnected nodes are estimated to be independent at that time.

non-zonal runs still spent a decent amount of time considering quite distant interactions — up to 140m apart. Interestingly, the most distant interaction strongly inferred in the non-zonal analysis (i.e. baboons 5 and 9 as discussed above) is hardly considered a possibility in the zonal analysis (less than 0.01 posterior probability) even though the proximity of that interaction is within the heavy tail of the radius posterior. In conjunction with the radius, it appears there is not much evidence for that interaction.

On average, using the point estimates of the radius in Figure 5.21, a baboon was within the interaction zones of 3.15 (mode) to 3.19 (mean) others. Even when using the 95^{th} percentile (a radius of 87.3m), this only rose to 3.28 others. This is as opposed to effectively being in the zones of 4 other baboons in the non-zonal case. Therefore, although the value of the radius itself needs to be explored and this data set concerns a typically stable social animal (Farine *et al.*, 2016), the zone of interaction had a sizeable impact in reducing the state space by discouraging spurious interactions. Furthermore, the addition of the zone of interaction did not impede the mixing of the model. As mentioned above, the burn-in period did not increase and the non-zonal analyses only produced a marginally greater ESS of the movement and transition rate parameters. Looking at the Gelman-Rubin diagnostic (Table 5.7), there is again little difference between the non-zonal and zonal runs — with perhaps the zonal runs having a slight edge. Again though, this table highlights that most of the runs have not converged and so they would be of limited use on their own. Overall then, the zone of interaction offers the ability to gain more insight into the social behaviours of a group of animals, whilst also limiting perhaps biologically extreme behaviours, at little cost.

5.9 Discussion

Understanding how animals interact with their environment can play an important role in ecosystem management. For example, the ability of schooling fish (or any other species that behaves with a similar mechanism) to respond to environmental information may decline if populations or group sizes decrease (Berdahl *et al.*, 2013). Extending our influence hierarchy model to the spatially heterogeneous case therefore greatly increases the scope of its applications. Though, this extension



(a) (b)

Figure 5.25: The *interaction posteriors* (the posterior probability of two baboons directly interacting) plotted against their *proximity* (*wildlifeDI* R package). The plots encompass all possible pairs of baboons at each observation. (a) are the posteriors from the non-zonal analysis whilst (b) are the posteriors from the zonal analysis. The *interaction posteriors* were calculated after the output from each of the 10 non-zonal and 10 zonal runs were collated. The solid teal line represents the modal point estimate of the interaction radius, whilst the dashed lines indicate the 5th and 95th percentile. Note, the radius estimates are included in the non-zonal plot for illustrative and comparative purposes only as the radius does not feature in the non-zonal model. Both axes have been jittered in order to help display the density of the points.

	Movement Parameter		Transition Rates	
	Non-Zonal	Zonal	Non-Zonal	Zonal
MPSRF Minimum	1.000	1.020	1.020	1.010
MPSRF 1^{st} Quartile	1.060	1.050	1.070	1.070
MPSRF Median	1.090	1.110	1.150	1.120
MPSRF Mean	1.118	1.117	1.177	1.164
MPSRF 3^{rd} Quartile	1.170	1.160	1.220	1.200
MPSRF Maximum	1.350	1.310	1.730	1.580
Upper C.I. Minimum	1.000	1.000	1.000	1.000
Upper C.I. 1^{st} Quartile	1.010	1.010	1.040	1.040
Upper C.I. Median	1.090	1.075	1.125	1.145
Upper C.I. Mean	1.181	1.178	1.282	1.308
Upper C.I. 3^{rd} Quartile	1.230	1.220	1.330	1.482
Upper C.I. Maximum	2.790	2.620	4.380	3.910

Table 5.7: Summaries of the Gelman-Rubin diagnostic for the zonal and non-zonal baboon runs. The diagnostic was calculated for each pair of runs and the upper confidence limits correspond to a coverage of 95% for the individual potential scale reduction factors.

provided a hurdle. The spatially heterogeneous version of forwards simulation, as a means to propose an animal's trajectory, is naive with regards to the known and required end point of that trajectory. This naivety compounds in the context of jointly modelling multiple animals and, as a result, it proved to be an inefficient means of proposing new trajectories. We therefore developed the LDB, which conditions on the next known location and thus provides more reasonable samples of an animal's movement. The LDB proved easy to tune as it added only a single tuning parameter to the model, one which performed well for a large range of values. It also had a sizeable impact in reducing the burn-in period when compared to forwards simulation, even for a group as small as three animals. That reduction is symptomatic of the general improvement in mixing that the LDB provides. The acceptance rate of trajectory updates almost tripled and, subsequently, longer trajectories could be sampled — further aiding mixing of both the behaviours and movements. Whilst the LDB does come at a computational cost, the benefits it brings far outweigh that cost. Based on some rough analysis not presented here, the above effects are amplified for larger groups — greater computational cost but even greater benefits to the efficiency of the inference process.

The above extension allowed us to implement a zone of interaction in the model — a familiar concept in the world of social movement models. Our version consists of a single distance which two animals must be within of each other in order to directly interact. Typically in other implementations of this concept, the radius (or radii when distinct zones are used for attraction, repulsion or alignment) is an inputted parameter of the model, requiring some previous knowledge on the matter. Here though, the radius is a parameter that we can infer whilst fitting the model to data. Not only does this avoid the pitfall of inferring different social structures depending on the radius inputted (Castles *et al.*, 2014; Davis *et al.*, 2018) but it can also provide more insight into the social behaviour of the study species. The zone and the inference of the radius slot neatly into our modelling framework and only a simple prior is required in order to keep the radius samples honest if there is no prior knowledge.

Results from fitting zonal and non-zonal versions of the model to baboon data indicate that the increased complexity of the zonal version doesn't have a consequential impact on how well the model mixes, nor increase the burn-in period. The zones did however reduce the size of the state space that is deemed sensible — even for a stable social animal such as baboons. In this case study then, the increased complexity accompanying the zonal feature is cancelled out by this practical benefit. The two different versions of the model inferred largely the same social behaviours, with the zonal runs only eliminating a quick, distant interaction. The zone therefore can increase the biological realism of the model by discouraging spurious interactions at little cost. Furthermore, the boundary of interaction can serve to limit the scale of a movement step a subordinate can take. Without that boundary, unreasonably large movement steps are possible as the expected value of a subordinate's next location is a weighted mean between the current location of itself and its dominant. With the boundary, those overly large movement steps can be avoided as it limits the distance within which interaction can occur. However, it therefore introduces the possibility of ruling out interactions that

do in fact occur over relatively large distances (compared to a reasonable movement step) (Farine & Whitehead, 2015; Scharf *et al.*, 2016; Calabrese *et al.*, 2018). For instance, when interactions occur through acoustic signalling.

There are ways to add flexibility into the radius. In a model containing multiple BM states, one method could be to have a distinct radius value for each BM speed. For instance, an animal moving under fast BM could have a larger zone of interaction than an animal moving under a slower BM process, with subordinates inheriting the radius from their leader. This would take into account that groups may interact differently whilst stationary than whilst on the move (Farine, 2015). Another method would be to have a smooth boundary for the interaction zones. For example, the strength of interaction decays exponentially with distance (see Torney *et al.* (2018) for an example of this approach). Not only would this method not outright reject distant interaction, it would also remove a problem of inferring a hard-boundary radius. That is, in order for more distant interactions (than the current radius sample allows) to be sampled and evaluated in the trajectory updates, the radius sample needs to jump above that prospective distance. If the jump required is sizeable, that will limit how well the model can explore potential interactions. Using an exponentially decaying smooth boundary would circumvent that problem as we would need to infer the rate of decay, rather than the radius.

Despite the extensions and improvements detailed in this chapter, several caveats of previous iterations of this model persist. In particular, the assumption that all relevant animals are tagged and the increasing complexity of the Markov chain (and the accompanying increase in computational effort required to explore that) resulting from more animals in the data. In fact, the extension to spatial heterogeneity and adding space, alongside behaviour and time, to the sample space that requires exploring has only exacerbated the slowness with which the trajectory updates mix (as well as adding a substantial computational cost in proposing a trajectory). Additionally, in the scenario where some relevant animals haven't been tagged, the zone of interaction must carry an interpretation warning. Missing animals from the analysis will likely inflate the radius estimation as that radius will correspond to what may in reality be indirect interaction.

As before, for groups where there are likely to be missing animals, the approach of Niu *et al.* (2016) may be a more suitable option as missing animals have little impact on that model. Additionally, that approach has recently been developed on by Niu *et al.* (2020) to allow the animals to switch between membership of the group or independence. Thus, the model in Niu *et al.* (2020) can be thought of as a special case of our influence hierarchy approach where the social structure is either despotic (led by a fixed individual) or democratic ('led' by a central abstract point in the group). The model therefore does not grow in complexity as the number of animals in the data increases. This provides much needed computational respite, but alongside a much simpler view of the social behaviours of the study group.

Chapter 6

Discussion and Future Work

6.1 Review of the Thesis Aims

We have developed a flexible model with which we can obtain an in-depth view of the social drivers behind collective movement when it is fitted to tracking data. The influence hierarchy approach, coupled with the behavioural switching process we have defined, can capture a wide array of complex social behaviours. This framework is based on the common concept of social hierarchies and so, despite the complexity of the model and the inference algorithms, the output is simple to interpret. With the inclusion of behavioural state switching, we are able to capture the dynamism of social interaction. Through this, we are able to infer the most causal influence at any given time, which can then highlight 'keystone' animals or persistent relationships within the group. We are also able to observe how the social structure changes over time — a feature which is particularly useful when investigating how animal groups are impacted by external factors or how social behaviours are adapted to heterogeneity (e.g. environmental heterogeneity). Moreover, we can still aggregate our results to obtain a longer-term static view of the social structure — similar to that of social network analysis (SNA). That is, obtain a more network-based picture of the social structure that encompasses all interactions observed during the study period, which can then facilitate a range of network analysis (Sosa *et al.*, 2021).

The models we have developed are formulated in continuous time. As such, the behavioural process is not married to the temporal resolution of the data and there is a greater deal of flexibility with regards to the data collection process. For instance, we do not require the data for all animals to be synchronous. Moreover, we avoid the approximations introduced when using a discrete-time model as we are able to undertake exact Bayesian inference of the model parameters.

We have also developed new algorithms with which we can fit this class of models to data. This process culminated in the development of the latent diffusion bridge (LDB) in Chapter 5, which provides a more efficient means to extend our modelling framework to spatial heterogeneity than

existing methods. The LDB removes some of the naivety of forwards simulation approaches by taking into account where a trajectory reconstruction is required to end. In particular, it takes into account the next known location of both the focal animal and the animals it is estimated to be interacting with. As such, realistic trajectories are proposed more frequently, which in turn improves the mixing of the wider MCMC algorithm. However, our modelling framework harnesses a complex state space and the LDB still naively reconstructs the behavioural process. Therefore, a considerable amount of time is still required to fully explore the state space. We discuss potential solutions to this problem in Section 6.3.

We examined how robust our modelling framework is to the particular social behaviours being analysed in Section 4.4. We found that, after analysing 400 simulated data sets which covered a broad range of social behaviours, that our model does not frequently infer false positives or false negatives with regards to two animals interacting. One shortcoming highlighted in this reliability testing, as well as during the data thinning experiments in Section 4.5, was that modelling the movement of leading and independent animals with Brownian motion (BM) is often not sufficient. When that is the case, the model can be formulated so that the movement of those animals is modelled with an Ornstein-Uhlenbeck (OU) process, with some attraction towards a geographical location (see Section 3.5.2). However, non-stationary processes can model the scenario when there is no persistence in the movement and their inclusion can provide some flexibility into the model (Blackwell, 1997). For instance, we could formulate the model so that leading and independent animals can switch between OU and BM movement. The non-stationary multivariate diffusion process we defined in Section 3.2, whilst perhaps overly simplistic on its own, would therefore play an important role in a switching, but overall stationary, version. Note, much of the work required to derive the stationary version of the conditional covariance matrix was completed in the process of deriving the non-stationary form (see Section 3.5.2 and Appendix A).

The spatial heterogeneity extension developed in Chapter 5 will widen the applicability of our model and allow us to investigate the interplay between social behaviours and the environment. We took advantage of that increased functionality to implement a zone of interaction, a similar concept to that found in the local-interaction-based models discussed in Section 1.2.3. This feature reduces the possibility of overly-large movement steps being produced by the model whilst also shrinking the behavioural state space that needs to be explored. As we infer the radius of this zone, we also obtain more information on the social interactions of the study species. Furthermore, the implementation of the interaction zone showcased how easily individual heterogeneity (with regards to interaction preferences) can be incorporated into the model. That is, we just need to keep track of those preferences when proposing state switches and evaluating their likelihoods. As such, demographics such as age and sex can be readily built into the model to examine the role they have in the social structure.

Whilst the above suggests we have made considerable progress towards meeting the aims we outlined in Chapter 1, the following sections provide some discussion on the limitations of our approach and

recommendations for future developments that could circumvent them.

6.2 Modelling Assumptions

6.2.1 Social

Whilst our influence hierarchy approach, coupled with the behavioural process we have defined, provides a very flexible framework, it has its limitations. We restrict the structure to represent the most causal influence and so an animal can only have one 'dominant.' However, this fails to account for non-dyadic interaction, where an animal is influenced by multiple others (Strandburg-Peshkin *et al.*, 2018) (the 'double-subordinate' in terms of network triads (McDonald & Shizuka, 2013)). Behavioural states could be included in our framework to represent such influence. For example, a state for subordination to animals x and y where the focal animal has some level of attraction to both x and y. However, the movement model is not currently defined for such interaction and, perhaps more importantly, this approach would rapidly increase the number of states required for larger groups. As such, this method is unlikely to be practical for groups of more than three or four animals. Additionally, co-movement cannot be modelled with our current social framework. As we saw with the zebra data analysis in Chapter 3, our assumption of there always being some ordering to interaction will not always be a fair one. A movement process, similar to the one we defined in Chapter 3, can in theory be constructed to explicitly capture mutually-informed movement. Again though, incorporating such behaviour into our model (to cover both co-movement and subordination) would produce a prohibitively large state space. However, if the estimation of explicit interaction is not required, similar models already exist which can model a group of animals as a mutually-informed collective (Niu *et al.*, 2016, 2020).

We assume the behavioural process is Markovian. Whilst this assumption no doubt ignores some of the complexity of animal behaviour, it represents a good compromise between model complexity and tractability (Patterson *et al.*, 2017). In some scenarios though, we may want to 'improve' on this Markovian assumption as highlighted by the baboon data analysis in Chapters 4 and 5. As discussed in Section 4.5.1, the constant transition rates of a Markov chain incorporate no presumption against very short state visits, as reflected in the exponentially distributed holding times. When there is some prior belief that short state visits are unlikely, an alternative approach may be to model the behavioural process as semi-Markovian. That is, the time spent in a state is modelled with some other non-negative distribution, one which may be more appropriate for the analysis at hand.

One way in which a semi-Markov process could be implemented is to make the transition rates depend on the time since the last actual state switch. Such an approach is relatively simple to implement in our existing framework. In this scenario, λ_{max} bounds above the hazard function of the chosen holding time distribution, as opposed to the transition rates. A state switch occurs at a potential switching time with probability $h(t)/\lambda_{max}$, where $h(t)$ is the value of the hazard function

t time units after the last switch. Such an approach provides a large degree of flexibility as to the nature of the holding times. Alternatively, we could represent each behavioural state (that we believe shouldn't have exponentially distributed holding times) with multiple substates — each of which is associated with the same movement process of the original state. For example, we could replace state u with states u_a and u_b in the Markovian behavioural process, with the holdings times of the new states still exponentially distributed. That is, $h_{u_a} \sim \text{exponential}(\lambda_{u_a})$, where h_{u_a} is the holding time for state u_a and λ_{u_a} is the rate of leaving state u_a (and similarly for u_b). For ease of interpretation, we could dictate that the substates must be visited in sequence. For instance, if u is a subordinate state, u_a might then be considered to represent the earlier stages of interaction, whilst state u_b might be considered to represent the later stages. The total time spent in state u is then $h_u = h_{u_a} + h_{u_b}$. h_u, in this particular in-sequence example, is hypoexponentially distributed — the shape of which is dictated by the rates λ_{u_a} and λ_{u_b}. In general, u can be substituted with any number of substates and they don't have to be visited in sequence (or even limited to one visit each). As such, h_u has a phase-type distribution, which can approximate any positive distribution. See Langrock *et al.* (2012) for an example of this approach in the context of discrete-time hidden Markov models (HMMs). In the context of our model though, one which already contains a complex state space, this approach may not be suitable.

6.2.2 Data

Our social framework drills into the social structure to the point of dyadic interactions. This brings forth one of two scenarios: all of the animals have been tagged and so each dyadic interaction inferred is a direct interaction, or some of the relevant animals haven't been tagged and so some of the inferred relationships are indirect. Even as tracking tags become smaller and cheaper, it won't necessarily be possible to tag all of the interacting animals. For instance, group composition may change throughout a long-term study. Consideration must therefore be given to not over interpret the inferred relationships.

When all relevant animals haven't been tagged, the inferred interactions will be a mix of direct and indirect influence. The model we have described in this thesis will not be able to distinguish between direct and indirect interaction as each interaction is based on the same scalar attraction parameter, α. A route that would potentially be able to prize the two apart is to make α a statistically-hierarchical parameter. The strength of an interaction would therefore able to fluctuate. Not only would this enable direct and indirect interaction to be expressed through different α values, but it would also allow interactions to differ in strength due to other factors. For example, a parent-child interaction may be stronger than one between two adult males.

Tracking data is becoming increasingly accurate. As such, we have ignored measurement error during the development of our models as the error will generally by inconsequential compared to the movement of an animal. However, that won't always be the case as even modern devices can produce very spurious observations (Frair *et al.*, 2010). In that scenario, our model can be extended

to take into account measurement error. One method to do this would be to resample the observed locations during the trajectory proposals. For example, the proposed true location could simply be drawn from a distribution centred on the observed location, a distribution which corresponds to the known or estimated form of the measurement error.

Tracking tags offer a range of benefits as a means to collect animal movement data (see Chapter 1). In models such as ours though, that look to analyse social behaviours, the interactions inferred are then based entirely on spatial data. Whilst we use terms such as 'dominant' and 'subordinate' for ease of reference, we are unable to obtain the nature of an interaction from spatial data alone. As such, we refer to our social framework as an 'influence hierarchy' to cover the ambiguity regarding the type of interaction. Directly observing interaction would provide that further detail (again, see Chapter 1), but movement models are often utilised where direct observation is not feasible. Pairing tracking tags with other sensors, such as accelerometers, may prove to be a fruitful direction in this regard. Accelerometers are already used to analyse behaviour and energy use (Kays *et al.*, 2015) and can typically record data at a much fast rate than GPS devices (Hughey *et al.*, 2018). Therefore, they may be able to shed some light on the nature of an interaction. For instance, Fehlmann *et al.* (2017) use accelerometer data and supervised machine learning methods to determine when baboons are grooming others.

6.3 Running Time

Perhaps the main critique of our model is the length of time it takes to fit it to data. We originally planned to undertake some analysis on how increasing the number of animals in the data set effects the running time and uncertainty of the model. In particular, as the proportion of tagged animals within a group increases, do some of the subordinate states become indistinguishable? However, it is clear that, even with the improvement obtained with the LDB algorithm, the model is currently too slow to be practically useful for larger data sets (due to a larger social group and/or number of observations). It would therefore be more prudent to explore some of the options outlined below before performing any analysis on how the number of animals affects the model fitting.

A simple, but effective, means with which we can speed up the inference process is to thin the data if its original resolution is deemed surplus to requirements. In Chapter 4, we investigated the impact that data thinning had on the inference of the model parameters. We found that the parameter estimations ultimately reflected the behaviours that could still be captured with the new resolution. For instance, thinning by a factor of 20 meant that the evidence for the short term interactions that we saw in the original data was unlikely to still be present in the heavily-thinned data. Whilst we took a simplistic approach to thinning the data, merely taking every fifth or 20^{th} observation, there are more formal (and potentially more fruitful) methods available. For example, Drovandi *et al.* (2017) treat data thinning akin to experimental design, where the design choice is which subset of data to analyse. That is, deciding which (minimal) subset of data the analyst

would have ideally collected in the first place in order to carry out the analysis at hand. After selecting an initial subset of data, some utility function is defined so that a decision can be made as to whether or not to include some more of the data in the analysis. One approach to this could be to evaluate the increase in the precision of the parameter estimations the extra data may provide whilst taking into account the cost of analysing a larger data set. Alternatively, locating turning points in an animal's movement path could be a way of stripping back ultra-high-resolution data to its more informative observations (Potts *et al.*, 2018).

Our granular social framework is more informative as a greater proportion of the animals are tagged. This requirement quickly increases the complexity of the Markov chain we use to model the animals' behaviours, which in turn increases the computational effort required to explore it. Moreover, coupling this complex behavioural process with time (Chapters 4 and 5) and space (Chapters 5) creates a huge sample space that requires exploring. One approach to simplify that sample space is to restrict the behavioural state space — similarly to how the interaction radius does so in Chapter 5. For example, it may be deemed biologically spurious (or just uninformative) for a social structure to contain more than, say, four levels of hierarchy. Such a restriction could be easily implemented through the transition rates — at a potential switching time, the rates of transitions that would produce a fifth level of hierarchy would simply be set to 0 before a switch is proposed. Alternatively, demographics might play some role. For instance, different types of whale are not known to interact (Scharf *et al.*, 2016). Of course, this method is only suitable if there is some prior knowledge that suggests such a restriction.

Our method of proposing trajectory realisations is still largely naive with regards to the behavioural process. We still ignore the required end state of the trajectory aside from 'forcing' the simulation to switch to (or stay in) that state at the last potential switching time. As such, we are slow to explore the state space. Hobolth & Stone (2009) detail some methods that can be used to simulate a continuous-time Markov chain conditional on the end state. One of those methods could slot into our existing 'uniformization' approach where we sample potential switching times with a Poisson process with rate λ_{max}. Though, in that method when sampling a state switch, the transition rates also need to be known at each future switching point of the trajectory — information we won't yet have in the spatially heterogeneous case. Another route for a more efficient means of reconstructing the behavioural process may be to do so with *integrated continuous-time hidden Markov models* (InCh) (Blackwell, 2018). During a trajectory reconstruction, potential switching times are still sampled with a Poisson process but here, in a spatially heterogeneous case, the corresponding locations are sampled with Brownian bridges (as opposed to being sampled from the 'correct' movement process). Then, conditional on those times and locations, the behavioural process is effectively a time-inhomogeneous HMM (Blackwell, 2018). The likelihood of the trajectory can then be calculated with the Forward Algorithm (as used alongside HMMs), integrating out the behavioural states. The state sequence can then be decoded using the Viterbi algorithm. This use of the efficient Forward and Viterbi algorithms within a wider MCMC framework could provide an

economical means of obtaining the posterior distributions of the model parameters.

Perhaps the approach that could provide the greatest gains in terms of computational efficiency is to approximate the model in discrete time. One approach would be to only allow state switches to occur at observed times, removing the need to augment the observed data with potential switching times. The model would still be formulated in continuous time, but rather than reconstruct the behavioural process in continuous time, we would approximate it in discrete time. For example, the observed time t_i contains a switch out of state u with probability $1 - e^{\lambda_u(t_i - t_{i-1})}$ where t_{i-1} is the immediately previous observed time. If so, that switch is to state v with probability λ_{uv}/λ_u as before.

If the data for all of the animals is synchronous, the behavioural process can then be modelled with a HMM (a time-inhomogeneous one if the data is irregular) and we can utilise the efficient algorithms that are in place to implement them. Otherwise, some trajectory reconstruction will still be required in order to obtain synchronous data — which the multivariate diffusion process defined in Chapter 3 requires. However, proposing a trajectory in the discrete-time-approximation case will be a much simpler and quicker process than in the continuous-time case. Mainly, in the spatially heterogeneous case, that is because we will no longer be required to propose locations at potential switching times as we will not be augmenting the observed data with them. The calculation of those proposal distributions, and the accompanying likelihood evaluations, encompass the bulk of the computational effort required for the trajectory updates as described in Chapter 5. Additionally, as the data set won't be augmented with switching times, we will be able to undertake trajectory updates over larger segments of the observed data, which will improve the mixing of the MCMC process.

The above method introduces some approximation — which we were keen to avoid as stated in our thesis aims. However, that approximation will be limited in cases where the data has been collected at a high frequency, such as the baboon data used in Chapters 4 and 5. Such a scenario will only become more commonplace as tracking technology continues to evolve. Additionally, the model will still be formulated in continuous time and so neither the movement or behavioural process will be married to a particular time interval. Thus, it will still be time invariant, enabling comparisons between studies and data collected at different resolutions, and it will still work seamlessly with irregular data. When data can't be collected at a high frequency then (for example, marine animal diving data), a discrete-time approximation of our continuous-time model still affords benefits over a purely discrete-time approach.

Bibliography

Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., Viale, M. & Zdravkovic, V. (2008) Interaction ruling animal collective behavior depends on topological rather than metric distance: evidence from a field study. *Proceedings of the National Academy of Sciences* **105**, 1232.

Bartlam-Brooks, H. & Harris, S. (2013) Data from: In search of greener pastures: using satellite images to predict the effects of environmental change on zebra migration. Movebank Data Repository.

Bartlam-Brooks, H.L.A., Beck, P.S.A., Bohrer, G. & Harris, S. (2013) In search of greener pastures: using satellite images to predict the effects of environmental change on zebra migration. *Journal of Geophysical Research: Biogeosciences* **118**, 1427–1437.

Berdahl, A., Torney, C.J., Ioannou, C.C., Faria, J.J. & Couzin, I.D. (2013) Emergent sensing of complex environments by mobile animal groups. *Science* **339**, 574–576.

Bivand, R.S., Pebesma, E. & Gomez-Rubio, V. (2013) *Applied spatial data analysis with R, Second edition.* Springer, NY.

Blackwell, P.G. (1997) Random diffusion models for animal movement. *Ecological Modelling* **100**, 87 – 102.

Blackwell, P.G. (2003) Bayesian inference for Markov processes with diffusion and discrete domponents. *Biometrika* **90**, 613–627.

Blackwell, P.G. (2018) Integrated continuous-time hidden Markov models https://arxiv.org/abs/1807.11907.

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

Blonder, B., Wey, T.W., Dornhaus, A., James, R. & Sih, A. (2012) Temporal dynamics and network analysis. *Methods in Ecology and Evolution* **3**, 958–972.

Bode, N.W.F., Franks, D.W., Wood, A.J., Piercy, J.J.B., Croft, D.P. & Codling, E.A. (2012) Distinguishing social from nonsocial navigation in moving animal groups. *The American Naturalist* **179**, 621–632.

Bode, N.W.F., Wood, A.J. & Franks, D.W. (2010) Social networks and models for collective motion in animals. *Behavioral Ecology and Sociobiology* **65**, 117–130.

Bond, M.L., König, B., Lee, D.E., Ozgu, A. & Farine, D.R. (2020) Proximity to humans affects local social structure in a giraffe metapopulation. *The Journal of animal ecology* **90**, 212–221.

Brooks, S.P. & Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**, 434–455.

Buhl, J., Sumpter, D.J.T., Couzin, I.D., Hale, J.J., Despland, E., Miller, E.R. & Simpson, S.J. (2006) From disorder to order in marching locusts. *Science* **312**, 1402–1406.

Calabrese, J.M., Fleming, C.H., Fagan, W.F., Rimmler, M., Kaczensky, P., Bewick, S., Leimgruber, P. & Mueller, T. (2018) Disentangling social interactions and environmental drivers in multi-individual wildlife tracking data. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170007.

Camperi, M., Cavagna, A., Giardina, I., Parisi, G. & Silvestri, E. (2012) Spatially balanced topological interaction grants optimal cohesion in flocking models. *Interface Focus* **2**, 715–725.

Castles, M., Heinsohn, R., Marshall, H.H., Lee, A.E., Cowlishaw, G. & Carter, A.J. (2014) Social networks created with different techniques are not comparable. *Animal Behaviour* **96**, 59–67.

Chase, I.D. & Lindquist, W.B. (2016) The fragility of individual-based explanations of social hierarchies: a test using animal pecking orders. *PLOS ONE* **11**, 1–16.

Clutton-Brock, T. (2009) Cooperation between non-kin in animal societies. *Nature* **462**, 51—57.

Codling, E.A., Pitchford, J.W. & Simpson, S.D. (2007) Group navigation and the "many-wrongs principle" in models of animal movement. *Ecology* **88**, 1864–1870.

Couzin, I.D., Ioannou, C.C., Demirel, G., Gross, T., Torney, C.J., Hartnett, A., Conradt, L., Levin, S.A. & Leonard, N.E. (2011) Uninformed individuals promote democratic consensus in animal groups. *Science* **334**, 1578–1580.

Couzin, I.D., Krause, J., Franks, N.R. & Levin, S.A. (2005) Effective leadership and decision-making in animal groups on the move. *Nature* **433**, 513–516.

Couzin, I.D., Krause, J., James, R., Ruxton, G.D. & Franks, N.R. (2002) Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology* **218**, 1 – 11.

Crofoot, M., Kays, R. & Wikelski, M. (2015) Data from: Shared decision-making drives collective movement in wild baboons. Movebank Data Repository.

Croft, S., Budgey, R., Pitchford, J.W. & Wood, A.J. (2015) Obstacle avoidance in social groups: new insights from asynchronous models. *Journal of the Royal Society Interface* **12**, 20150178.

Davis, G.H., Crofoot, M.C. & Farine, D.R. (2018) Estimating the robustness and uncertainty of animal social networks using different observational methods. *Animal Behaviour* **141**, 29–44.

del Mar Delgado, M., Miranda, M., Alvarez, S.J., Gurarie, E., Fagan, W.F., Penteriani, V., di Virgilio, A. & Morales, J.M. (2018) The importance of individual variation in the dynamics of animal collective movements. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170008.

Delgado, M.d.M., Penteriani, V., Morales, J.M., Gurarie, E. & Ovaskainen, O. (2014) A statistical framework for inferring the influence of conspecifics on movement behaviour. *Methods in Ecology and Evolution* **5**, 183–189.

Drovandi, C.C., Holmes, C.C., McGree, J.M., Mengersen, K., Richardson, S. & Ryan, E.G. (2017) Principles of experimental design for big data analysis. *Statistical Science* **32**, 385–404.

Dunn, J.E. & Gipson, P.S. (1977) Analysis of radio telemetry data in studies of home range. *Biometrics* **33**, 85–101.

Ellis, S., Franks, D., Nattrass, S., Cant, M., Weiss, M., Giles, D., Balcomb, K. & Croft, D. (2017) Mortality risk and social network position in resident killer whales: sex differences and the importance of resource abundance. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20171313.

Farine, D.R. (2015) Proximity as a proxy for interactions: issues of scale in social network analysis. *Animal Behaviour* **104**, e1 – e5.

Farine, D.R. (2018) When to choose dynamic vs. static social network analysis. *Journal of Animal Ecology* **87**, 128–138.

Farine, D.R., Strandburg-Peshkin, A., Berger-Wolf, T., Ziebart, B., Brugere, I., Li, J. & Crofoot, M.C. (2016) Both nearest neighbours and long-term affiliates predict individual locations during collective movement in wild baboons. *Scientific Reports* **6**, 27704.

Farine, D.R., Strandburg-Peshkin, A., Couzin, I.D., Berger-Wolf, T.Y. & Crofoot, M.C. (2017) Individual variation in local interaction rules can explain emergent patterns of spatial organization in wild baboons. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20162243.

Farine, D.R. & Whitehead, H. (2015) Constructing, conducting and interpreting animal social network analysis. *Journal of Animal Ecology* **84**, 1144–1163.

Fehlmann, G., O'Riain, M., Hopkins, P.W., O'Sullivan, J., Holton, M.D., Shepard, E.L. & King, A.J. (2017) Identification of behaviours from accelerometer data in a wild social primate. *Animal Biotelemetry* **5**.

Frair, J.L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N.J. & 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: Biological Sciences* **365**, 2187–2200.

Franz, M. & Nunn, C.L. (2009) Network-based diffusion analysis: a new method for detecting social learning. *Proceedings of the Royal Society B: Biological Sciences* **276**, 1829–1836.

Garroway, C.J., Bowman, J. & Wilson, P.J. (2013) Complex social structure of southern flying squirrels is related to spatial proximity but not kinship. *Behavioral Ecology and Sociobiology* **67**, 113–122.

Gelman, A. & Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–472.

Gilks, W., Richardson, S. & Spiegelhalter, D. (1996) *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC.

Hamede, R.K., Bashford, J., Mccallum, H. & Jones, M. (2009) Contact networks in a wild Tasmanian devil (Sarcophilus harrisii) population: using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. *Ecology Letters* **12**, 1147–1157.

Hamilton, W. (1964a) The genetical evolution of social behaviour. I. *Journal of Theoretical Biology* **7**, 1–16.

Hamilton, W. (1964b) The genetical evolution of social behaviour. II. *Journal of Theoretical Biology* **7**, 17–52.

Hamilton, W.D. (1963) The evolution of altruistic behavior. *The American Naturalist* **97**, 354–356.

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

Haydon, D.T., Morales, J.M., Yott, A., Jenkins, D.A., Rosatte, R. & Fryxell, J.M. (2008) Socially informed random walks: incorporating group dynamics into models of population spread and growth. *Proceedings of the Royal Society B: Biological Sciences* **275**, 1101–1109.

Herbert-Read, J.E., Perna, A., Mann, R.P., Schaerf, T.M., Sumpter, D.J.T. & Ward, A.J.W. (2011) Inferring the rules of interaction of shoaling fish. *Proceedings of the National Academy of Sciences* **108**, 18726–18731.

Hobolth, A. & Stone, E.A. (2009) Simulation from endpoint-conditioned, continuous-time Markov chains on a finite state space, with applications to molecular evolution. *The Annals of Applied Statistics* **3**, 1204–1231.

Hobson, E.A., Avery, M.L. & Wright, T.F. (2013) An analytical framework for quantifying and testing patterns of temporal dynamics in social networks. *Animal Behaviour* **85**, 83–96.

Hooten, M.B., Scharf, H.R., Hefley, T.J., Pearse, A.T. & Weegman, M.D. (2018) Animal movement models for migratory individuals and groups. *Methods in Ecology and Evolution* **9**, 1692–1705.

Hughey, L.F., Hein, A.M., Strandburg-Peshkin, A. & Jensen, F.H. (2018) Challenges and solutions for studying collective animal behaviour in the wild. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170005.

Johnson, D.S., London, J.M., Lea, M.A. & Durban, J.W. (2008) Continuous-time correlated random walk model for animal telemetry data. *Ecology* **89**, 1208–1215.

Joo, R., Etienne, M.P., Bez, N. & Mahévas, S. (2018) Metrics for describing dyadic movement: a review. *Movement Ecology* **6**, 26.

Katz, Y., Tunstrøm, K., Ioannou, C.C., Huepe, C. & Couzin, I.D. (2011) Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences* **108**, 18720–18725.

Kays, R., Crofoot, M.C., Jetz, W. & Wikelski, M. (2015) Terrestrial animal tracking as an eye on life and planet. *Science* **348**.

King, A.J., Fehlmann, G., Biro, D., Ward, A.J. & Fürtbauer, I. (2018) Re-wilding collective behaviour: an ecological perspective. *Trends in Ecology & Evolution* **33**, 347 – 357.

Krause, J., Lusseau, D. & James, R. (2009) Animal social networks: an introduction. *Behavioral Ecology and Sociobiology* **63**, 967–973.

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

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

Lewis, J.S., Wartzok, D. & Heithaus, M.R. (2011) Highly dynamic fission–fusion species can exhibit leadership when traveling. *Behavioral Ecology and Sociobiology* **65**, 1061–1069.

Long, J.A., Nelson, T.A., Webb, S.L. & Gee, K.L. (2014) A critical examination of indices of dynamic interaction for wildlife telemetry studies. *Journal of Animal Ecology* **83**, 1216–1233.

Lukeman, R., Li, Y.X. & Edelstein-Keshet, L. (2010) Inferring individual rules from collective behavior. *Proceedings of the National Academy of Sciences* **107**, 12576–12580.

Mann, R.P. (2011) Bayesian inference for identifying interaction rules in moving animal groups. *PLOS ONE* **6**, e22827–e22827.

Martin, A.D., Quinn, K.M. & Park, J.H. (2011) MCMCpack: Markov chain Monte Carlo in R. *Journal of Statistical Software* **42**, 22.

Mbizah, M.M., Farine, D.R., Valeix, M., Hunt, J.E., Macdonald, D.W. & Loveridge, A.J. (2020) Effect of ecological factors on fine-scale patterns of social structure in African lions. *Journal of Animal Ecology* **89**, 2665–2676.

McClintock, B.T., Johnson, D.S., Hooten, M.B., Ver Hoef, J.M. & Morales, J.M. (2014) When to be discrete: the importance of time formulation in understanding animal movement. *Movement Ecology* **2**, 21–21.

McDonald, D.B. & Shizuka, D. (2013) Comparative transitive and temporal orderliness in dominance networks. *Behavioral Ecology* **24**, 511–520.

McKellar, A.E., Langrock, R., Walters, J.R. & Kesler, D.C. (2015) Using mixed hidden Markov models to examine behavioral states in a cooperatively breeding bird. *Behavioral Ecology* **26**, 148–157.

Milner, J.E., Blackwell, P.G. & Niu, M. (2020) jemilner/influenceHierarchy: Code for MEE article: Modelling and inference for the movement of interacting animals.

Milner, J.E., Blackwell, P.G. & Niu, M. (2021) Modelling and inference for the movement of interacting animals. *Methods in Ecology and Evolution* **12**, 54–69.

Morales, J.M. & Ellner, S.P. (2002) Scaling up animal movements in heterogeneous landscapes: the importance of behavior. *Ecology* **83**, 2240–2247.

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

Morales, J.M., Moorcroft, P.R., Matthiopoulos, J., Frair, J.L., Kie, J.G., Powell, R.A., Merrill, E.H. & Haydon, D.T. (2010) Building the bridge between animal movement and population dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 2289–2301.

Nagy, M., Ákos, Z., Biro, D. & Vicsek, T. (2010) Hierarchical group dynamics in pigeon flocks. *Nature* **464**, 890–893.

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

Niu, M., Frost, F., Milner, J.E., Skarin, A. & Blackwell, P.G. (2020) Modelling group movement with behaviour switching in continuous time. *Biometrics* In press.

Øksendal, B. (2003) *Stochastic differential equations: an introduction with applications*. Springer.

Parton, A. & 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.

Patterson, T.A., Basson, M., Bravington, M.V. & Gunn, J.S. (2009) Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology* **78**, 1113–1123.

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

Pebesma, E.J. & Bivand, R.S. (2005) Classes and methods for spatial data in R. *R News* **5**, 9–13.

Pinter-Wollman, N., Hobson, E.A., Smith, J.E., Edelman, A.J., Shizuka, D., de Silva, S., Waters, J.S., Prager, S.D., Sasaki, T., Wittemyer, G., Fewell, J. & McDonald, D.B. (2013) The dynamics of animal social networks: analytical, conceptual, and theoretical advances. *Behavioral Ecology* **25**, 242–255.

Plummer, M., Best, N., Cowles, K. & Vines, K. (2006) CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11.

Potts, J.R., Börger, L., Scantlebury, D.M., Bennett, N.C., Alagaili, A. & Wilson, R.P. (2018) Finding turning-points in ultra-high-resolution animal movement data. *Methods in Ecology and Evolution* **9**, 2091–2101.

Potts, J.R., Mokross, K. & Lewis, M.A. (2014) A unifying framework for quantifying the nature of animal interactions. *Journal of the Royal Society Interface* **11**, 20140333.

Psorakis, I., Roberts, S.J., Rezek, I. & Sheldon, B.C. (2012) Inferring social network structure in ecological systems from spatio-temporal data streams. *Journal of the Royal Society Interface* **9**, 3055–3066.

R Core Team (2017) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramos-Fernández, G., King, A.J., Beehner, J.C., Bergman, T.J., Crofoot, M.C., Di Fiore, A., Lehmann, J., Schaffner, C.M., Snyder-Mackler, N., Zuberbühler, K., Aureli, F. & Boyer, D. (2018) Quantifying uncertainty due to fission–fusion dynamics as a component of social complexity. *Proceedings of the Royal Society B: Biological Sciences* **285**, 20180532.

Ramos-Fernández, G. & Morales, J.M. (2014) Unraveling fission-fusion dynamics: how subgroup properties and dyadic interactions influence individual decisions. *Behavioral Ecology and Sociobiology* **68**, 1225–1235.

Reynolds, C.W. (1987) Flocks, herds and schools: a distributed behavioral model. *Computer Graphics* **21**, 25–34.

Rosenthal, S.B., Twomey, C.R., Hartnett, A.T., Wu, H.S. & Couzin, I.D. (2015) Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proceedings of the National Academy of Sciences - PNAS* **112**, 4690–4695.

Russell, J.C., Hanks, E.M. & Haran, M. (2016) Dynamic models of animal movement with spatial point process interactions. *Journal of Agricultural, Biological, and Environmental Statistics* **21**, 22–40.

Russell, J.C., Hanks, E.M., Modlmeier, A.P. & Hughes, D.P. (2017) Modeling collective animal movement through interactions in behavioral states. *Journal of Agricultural, Biological, and Environmental Statistics* **22**, 313–334.

Sarkar, P. & Moore, A. (2005) Dynamic social network analysis using latent space models. *SIGKDD Explorations* **7**, 31–40.

Scharf, H.R. & Buderman, F.E. (2020) Animal movement models for multiple individuals. *WIREs Computational Statistics* **12**, e1506.

Scharf, H.R., Hooten, M.B., Fosdick, B.K., Johnson, D.S., London, J.M. & Durban, J.W. (2016) Dynamic social networks based on movement. *The Annals of Applied Statistics* **10**, 2182–2202.

Scharf, H.R., Hooten, M.B., Johnson, D.S. & Durban, J.W. (2018) Process convolution approaches for modeling interacting trajectories. *Environmetrics* **29**, e2487.

Schlägel, U.E., Signer, J., Herde, A., Eden, S., Jeltsch, F., Eccard, J.A. & Dammhahn, M. (2019) Estimating interactions between individuals from concurrent animal movements. *Methods in Ecology and Evolution* **10**, 1234–1245.

Shizuka, D. & McDonald, D.B. (2015) The network motif architecture of dominance hierarchies. *Journal of the Royal Society Interface* **12**, 20150080.

Silk, M.J., Jackson, A.L., Croft, D.P., Colhoun, K. & Bearhop, S. (2015) The consequences of unidentifiable individuals for the analysis of an animal social network. *Animal Behaviour* **104**, 1–11.

Snijders, L., Blumstein, D.T., Stanley, C.R. & Franks, D.W. (2017) Animal social network theory can help wildlife conservation. *Trends in Ecology & Evolution* **32**, 567–577.

Snijders, T.A., van de Bunt, G.G. & Steglich, C.E. (2010) Introduction to stochastic actor-based models for network dynamics. *Social Networks* **32**, 44–60.

Sosa, S., Sueur, C. & Puga-Gonzalez, I. (2021) Network measures in animal social network analysis: their strengths, limits, interpretations and uses. *Methods in Ecology and Evolution* **12**, 10–21.

Strandburg-Peshkin, A., Farine, D.R., Couzin, I.D. & Crofoot, M.C. (2015) Shared decision-making drives collective movement in wild baboons. *Science* **348**, 1358–1361.

Strandburg-Peshkin, A., Farine, D.R., Crofoot, M.C. & Couzin, I.D. (2017) Habitat and social factors shape individual decisions and emergent group structure during baboon collective movement. *eLife* **6**, e19505.

Strandburg-Peshkin, A., Papageorgiou, D., C. Crofoot, M. & Farine, D. (2018) Inferring influence and leadership in moving animal groups. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170006.

Strandburg-Peshkin, A., Twomey, C.R., Bode, N.W., Kao, A.B., Katz, Y., Ioannou, C.C., Rosenthal, S.B., Torney, C.J., Wu, H.S., Levin, S.A. & Couzin, I.D. (2013) Visual sensory networks and effective information transfer in animal groups. *Current Biology* **23**, R709–R711.

Sueur, C., King, A.J., Conradt, L., Kerth, G., Lusseau, D., Mettke-Hofmann, C., Schaffner, C.M., Williams, L., Zinner, D. & Aureli, F. (2011) Collective decision-making and fission-fusion dynamics: a conceptual framework. *Oikos* **120**, 1608–1617.

Tang, J., Leu, G. & Abbass, H.A. (2018) Networking the boids is more robust against adversarial learning. *IEEE Transactions on Network Science and Engineering* **5**, 141–155.

Torney, C.J., Lamont, M., Debell, L., Angohiatok, R.J., Leclerc, L.M. & Berdahl, A.M. (2018) Inferring the rules of social interaction in migrating caribou. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170385.

Tucker, M.A., Böhning-Gaese, K., Fagan, W.F., Fryxell, J.M., Moorter, B.V., Alberts, S.C., ... & Mueller, T. (2018) Moving in the Anthropocene: global reductions in terrestrial mammalian movements. *Science* **359**, 466–472.

Tunstrøm, K., Katz, Y., Ioannou, C.C., Huepe, C., Lutz, M.J. & Couzin, I.D. (2013) Collective states, multistability and transitional behavior in schooling fish. *PLOS Computational Biology* **9**, e1002915.

van Ravenzwaaij, D., Cassey, P. & Brown, S.D. (2018) A simple introduction to Markov chain Monte-Carlo sampling. *Psychonomic Bulletin & Review* **25**, 143–154.

Voelkl, B. & Noë, R. (2008) The influence of social structure on the propagation of social information in artificial primate groups: a graph-based simulation approach. *Journal of Theoretical Biology* **252**, 77–86.

Wang, Y.S., Blackwell, P.G., Merkle, J.A. & Potts, J.R. (2019) Continuous time resource selection analysis for moving animals. *Methods in Ecology and Evolution* **10**, 1664–1678.

Westley, P.A., Berdahl, A.M., Torney, C.J. & Biro, D. (2018) Collective movement in ecology: from emerging technologies to conservation and management. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**.

Williams, H.J., Taylor, L.A., Benhamou, S., Bijleveld, A.I., Clay, T.A., Grissac, S., Demšar, U., English, H.M., Franconi, N., Gómez-Laich, A., Griffiths, R.C., Kay, W.P., Morales, J.M., Potts, J.R., Rogerson, K.F., Rutz, C., Spelt, A., Trevail, A.M., Wilson, R.P. & Börger, L. (2020) Optimizing the use of biologgers for movement ecology research. *Journal of Animal Ecology* **89**, 186–206.

Wittemyer, G., Douglas-Hamilton, I. & Getz, W. (2005) The socioecology of elephants: analysis of the processes creating multitiered social structures. *Animal Behaviour* **69**, 1357–1371.

Wittemyer, G., Getz, W.M., Vollrath, F. & Douglas-Hamilton, I. (2007) Social dominance, seasonal movements, and spatial segregation in African elephants: a contribution to conservation behavior. *Behavioral Ecology and Sociobiology* **61**, 1919–1931.

Appendix A

Conditional Covariance Derivation

As Niu *et al.* (2016) details, the conditional covariance $\Xi(\boldsymbol{F}, t)$, where $\boldsymbol{F}$ is the attraction matrix and t is the time length of the movement step, can be calculated using the following:

$$\Xi(F, t) = \Delta - e^{Ft}\Delta e^{F^T t} \tag{A.1}$$

where Δ is the stationary covariance matrix. The derivation of the terms in the Ξ matrix therefore consists of three steps: deriving e^{Ft}, Δ and then Ξ itself. However, Niu *et al.* (2016) only does this to the extent that is required for their model. That is, the social structure (represented in F) is limited to there being a single leader, with all of the remaining animals following it. Thus, we need to carry out the above three steps in the context of the wider range of social interactions that our influence hierarchy approach allows (see Section 3.1). We will derive a more general form of e^{Ft} in Section A.1, Δ in Section A.2 and Ξ in Section A.3. By general form, we mean general enough to represent all the possible social interactions our social assumptions allow in Section 3.1.

Note, all of e^{Ft}, Δ and Ξ are matrices and, in each case, each row and column relates to a specific animal. That is, row i relates to animal i and column j relates to animal j. As such, reference to i concerns both animal i and the i^{th} index in the relevant matrix (and similarly for j).

Whilst we only ever use Brownian motion (BM) in this thesis to model the movement of leading and independent animals, and so there is no attraction term for these animals, it useful to begin this derivation assuming that they do have some attraction to some location. That is, we model their movement with an Ornstein-Uhlenbeck (OU) process. This helps avoid division by 0 issues throughout Section A.2. We are able to set their attraction to 0 at the end of this process, in Section A.3, to obtain the conditional variance terms when leading and independent animals are modelled with BM.

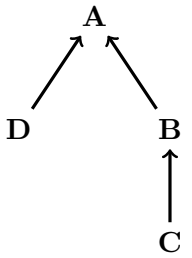


Figure A.1: A dominates B, B dominates C and A also dominates D. Animal A is rank 1, both B and D are rank 2 and C is rank 3. As such $rank(A) = 1$, $rank(B) = 2$ etc. An animal's 'direct' hierarchy just concerns the branch of the structure they are in. For example, for C, their direct hierarchy is A $\longleftarrow$ B $\longleftarrow$ C and D is in a different branch. The notation $i(r)$ represents the animal at the r^{th} rank in animal i's direct hierarchy. Still from the perspective of C, $i(1)$ is A, $i(2)$ is B and $i(3)$ is C itself. Then, A is 2 ranks directly above C and B is 1 rank directly above C. Even though D is ranked higher than C, it is in a different branch of the hierarchy and so it is not ranked directly above it. Animal i is 'lower' (or 'higher') in the hierarchy than animal j when it is of lower (or higher) rank than j in the same subgroup.

A.1 e^{Ft}

This section contains the derivation of the general form of the exponential matrix e^{Ft} when the movement of leading and independent animals contains some attraction. We will show the general form is:

$$e^{Ft}_{i,j} = \begin{cases} e^{-\beta t} & \text{E1 — } i = j \text{ and } i \text{ is a leading/independent animal} \\ e^{-\alpha t} & \text{E2 — } i = j \text{ and } i \text{ is a subordinate} \\ \frac{(\alpha t)^d e^{-\alpha t}}{d!} & \text{E3 — } i \neq j \text{ and } i \text{ is } d \text{ ranks directly below non-leading } j \\ \frac{\alpha}{\alpha-\beta}(e^{Ft}_{dom(i),j} - e^{Ft}_{i,i(2)}) & \text{E4 — } i \neq j \text{ and } i \text{ is lead by } j \\ 0 & \text{E5 — otherwise} \end{cases}$$

where $dom(i)$ is the dominant animal of animal i and $i(r)$ is the animal at the r^{th} 'rank' in animal i's direct hierarchy above it. See Figure A.1 for an explanation of the hierarchy terminology. α is the rate of attraction of subordinates and β is the rate of attraction of leading/independent animals.

To derive these terms, we will use the power series of e^{Ft}:

$$e^{Ft} = \sum_{n=0}^{\infty} \frac{(Ft)^n}{n!} = I + Ft + \frac{(Ft)^2}{2!} + \frac{(Ft)^3}{3!} + ... \tag{A.2}$$

where I is the identity matrix.

For all derivations in this section, without loss of generality, let F be ordered in accordance with the group's hierarchy. That is, the first row corresponds to an animal of rank 1 (i.e. a leading animal), the next row(s) correspond to the animal(s) of rank 2 in that same subgroup etc. If there is more than one subgroup, the first x rows correspond to a subgroup of size x, the following y rows correspond to another subgroup of size y etc, with the rows of each subgroup ordered as above. The upper triangle of F is therefore 0 (Section 3.2 details how F is constructed).

E5 — otherwise

We will first derive E5 as it makes the derivation of the other terms simpler. During this derivation, the subscript of F indicates the number of animals in the group. That is, F_k is the attraction matrix of a group of k animals.

This case relates to when i and j are in different subgroups, different branches of the same subgroup or when j is directly below i in a branch. The derivation for being in different subgroups or branches is identical. Say the $(k+1)^{st}$ animal is subordinate to animal $k-1$ and let $i = k+1$, $j = k$. i and j are therefore either in different subgroups or different branches of the same subgroup (due to the ordering of F_k). We will show that $(F_k t)^n_{i,j} = 0$ for all $n \geq 0$ by induction. For the base case, $n = 0$:

$$(F_{k+1}t)^0 = I_{k+1} \quad \Longrightarrow \quad (F_{k+1}t)^0_{k+1,k} = 0 \qquad \text{as (k+1, k) is off the diagonal}$$

We assume true for $n = p-1$. Then for $n = p$:

$$\begin{aligned}
(F_{k+1}t)^p &= (F_{k+1}t)(F_{k+1}t)^{p-1} \\
&= \left(\begin{array}{ccccc|c} & F_k t & & & & 0 \\ 0 & \dots & 0 & \alpha t & 0 & -\alpha t \end{array}\right)\left(\begin{array}{ccc|c} & (F_k t)^{p-1} & & 0 \\ (F_{k+1}t)^{p-1}_{k+1,1} & \dots & (F_{k+1}t)^{p-1}_{k+1,k} & (F_{k+1}t)^{p-1}_{k+1,k+1} \end{array}\right) \\
&= \left(\begin{array}{ccc|c} & (F_k t)^p & & 0 \\ \alpha t(F_{k+1}t)^{p-1}_{k-1,1} - \alpha t(F_{k+1}t)^{p-1}_{k+1,1} & \dots & \alpha t(F_{k+1}t)^{p-1}_{k-1,k} - \alpha t(F_{k+1}t)^{p-1}_{k+1,k} & -\alpha t(F_{k+1}t)^{p-1}_{k+1,k+1} \end{array}\right)
\end{aligned}$$

For the (k+1, k) index:

$$\begin{aligned}
(F_{k+1}t)^p_{k+1,k} &= \alpha t(F_{k+1}t)^{p-1}_{k-1,k} - \alpha t(F_{k+1}t)^{p-1}_{k+1,k} \\
&= 0
\end{aligned}$$

as the first term is in the upper triangle (which is 0) and the second term is 0 through our induction assumption. As $(F_k t)^n_{i,j} = 0$ for all $n \geq 0$, it therefore holds that $e^{Ft}_{i,j} = \sum_{n=0}^{\infty} \frac{(F_k t)^n_{i,j}}{n!} = 0$ when i and j are in different branches or subgroups of the hierarchy. Note, the above holds when the $(k+1)^{st}$ animal is subordinate to any animal other than animal k, not just $k-1$.

As $F_k t$ is a lower triangular matrix, $(F_k t)^n$ is also a lower triangular matrix for all $n \geq 0$. Therefore,

$(F_k t)^n_{i,j} = 0$ for all $n \geq 0$ and $e^{F_k t}_{i,j} = 0$ when j is directly below i.

As a result of E5, all the non-zero terms in row i of $e^{F_k t}$ relate to animals directly above animal i in the hierarchy. As such, we only have to consider hierarchies consisting of a single branch in the derivations of the remaining terms of e^{Ft}. This allows us to make use of the following as the $(k+1)^{st}$ animal will always be subordinate to the k^{th} animal:

$$
\begin{aligned}
(F_{k+1}t)^n &= \left(\begin{array}{ccccc} F_k t & & & & 0 \\ 0 & \dots & 0 & \alpha t & -\alpha t \end{array}\right)\left(\begin{array}{ccccc} (F_k t)^{n-1} & & & & 0 \\ (F_{k+1}t)^{n-1}_{k+1,1} & \dots & & (F_{k+1}t)^{n-1}_{k+1,k} & (F_{k+1}t)^{n-1}_{k+1,k+1} \end{array}\right) \\
&= \left(\begin{array}{ccc|c} & (F_k t)^n & & 0 \\ \hline \alpha t(F_{k+1}t)^{n-1}_{k,1} - & & \alpha t(F_{k+1}t)^{n-1}_{k,k} - & \\ \alpha t(F_{k+1}t)^{n-1}_{k+1,1} & \cdots & \alpha t(F_{k+1}t)^{n-1}_{k+1,k} & -\alpha t(F_{k+1}t)^{n-1}_{k+1,k+1} \end{array}\right)
\end{aligned} \tag{A.3}
$$

With this in mind, we introduce some notation simplifications:

- the row and column of e^{Ft} relating to the leader of the subgroup will be indexed with 1
- the row and column of e^{Ft} relating to $dom(i)$ will be indexed with $i-1$

E1 — $i = j$ and i is a leading/independent animal

We will show that $(Ft)^n_{1,1} = (-\beta t)^n$ for all $n \geq 0$ by induction. For the base case, $n = 0$:

$$(Ft)^0 = I_k \quad \Longrightarrow \quad (Ft)^0_{1,1} = 1 = (-\beta t)^0$$

We assume true for $n = p-1$. Then for $n = p$:

$$
\begin{aligned}
(Ft)^p_{1,1} &= (Ft)_{1,\bullet} \cdot (Ft)^{p-1}_{\bullet,1} \\
&= -\beta t (Ft)^{p-1}_{1,1} && \text{as a leaders row in } Ft \text{ only contains 0 except for } -\beta t \text{ at (1, 1)} \\
&= -\beta t (-\beta t)^{p-1} && \text{by our induction assumption} \\
&= (-\beta t)^p
\end{aligned}
$$

As $(Ft)^n_{1,1} = (-\beta t)^n$ for all $n \geq 0$, it therefore holds that $\sum_{n=0}^{\infty} \frac{(Ft)^n_{1,1}}{n!} = \sum_{n=0}^{\infty} \frac{(-\beta t)^n}{n!}$ as both sides of the equality have been modified identically. As such, $e^{Ft}_{1,1} = e^{-\beta t}$ (note, $\sum_{n=0}^{\infty} \frac{(Ft)^n_{1,1}}{n!} = (\sum_{n=0}^{\infty} \frac{(Ft)^n}{n!})_{1,1}$).

E2 — $i = j$ and i is a subordinate

We will show that $(Ft)^n_{i,i} = (-\alpha t)^n$ when i is a subordinate for all $n \geq 0$ by induction. For the base case, $n = 0$:

$$(Ft)^0 = I_k \quad \implies \quad (Ft)^0_{i,i} = 1 = (-\alpha t)^0$$

We assume true for $n = p - 1$. Then for $n = p$:

$$\begin{aligned}
(Ft)^p_{i,i} &= (Ft)_{i,\bullet} \cdot (Ft)^{p-1}_{\bullet,i} \\
&= -\alpha t (Ft)^{p-1}_{i,i} && \text{by equation A.3} \\
&= -\alpha t (-\alpha t)^{p-1} && \text{by our induction assumption} \\
&= (-\alpha t)^p
\end{aligned}$$

As $(Ft)^n_{i,i} = (-\alpha t)^n$ for all $n \geq 0$, it therefore holds that $\sum_{n=0}^{\infty} \frac{(Ft)^n_{i,i}}{n!} = \sum_{n=0}^{\infty} \frac{(-\alpha t)^n}{n!}$ as both sides of the equality have been modified identically. As such, $e^{Ft}_{i,i} = e^{-\alpha t}$.

E3 — $i \neq j$ and i is d ranks directly below non-leading j

We will show this in three parts. Firstly, that $(Ft)^n_{i,j} = 0$ for all $n = 0, ..., d-1$. Secondly, $(Ft)^n_{i,j} = \frac{n!(\alpha t)^d(-\alpha t)^{n-d}}{d!(n-d)!}$ for all $n = d$. Thirdly, the same as $n = d$ but for $n > d$. Each part will be done by induction. Note, E3 excludes the case where j is leading a subgroup — that scenario is covered by E4.

Part 1 - $n < d$

For the base case, $n = 0$:

$$(Ft)^0 = I_k \quad \implies \quad (Ft)^0_{i,j} = 0 \qquad \text{as (i, j) is off the diagonal for all } d > 0$$

The base case therefore holds for all d greater than 0. We assume true for $n = p - 1$ for all d such that $p - 1 < d - 1$. Then for $n = p$ for all d such that such that $p < d$:

$$\begin{aligned}
(Ft)^p_{i,j} &= (Ft)_{i,\bullet} \cdot (Ft)^{p-1}_{\bullet,j} \\
&= \alpha t (Ft)^{p-1}_{i-1,j} - \alpha t (Ft)^{p-1}_{i,j} \\
&= \alpha t (Ft)^{p-1}_{i-1,j} - 0 && \text{by our induction assumption} \\
&= 0 && \text{by our induction assumption}
\end{aligned}$$

The first use of our induction assumption follows as animal i is d ranks below j. As $d > d - 1$, the assumption holds. The second use follows as animal $i - 1$ ($dom(i)$) is $d - 1$ ranks below j.

Part 2 — $d = n$**.**

For the base case, $d = n = 1$ (as d must be greater than 0).

$$(Ft)^1_{i,j} = \alpha t = \frac{1!(\alpha t)^1(-\alpha t)^0}{1!0!} \qquad \text{as required}$$

We assume true for $d = n = p - 1$. Then for $d = n = p$:

$$\begin{aligned}
(Ft)^p_{i,j} &= (Ft)_{i,\bullet} \cdot (Ft)^{p-1}_{\bullet,j} \\
&= \alpha t(Ft)^{p-1}_{i-1,j} - \alpha t(Ft)^{p-1}_{i,j} \\
&= \alpha t(Ft)^{p-1}_{i-1,j} - 0 && \text{by part 1} \\
&= \frac{\alpha t(p-1)!(\alpha t)^{d-1}(-\alpha t)^{p-1-d+1}}{(d-1)!(p-1-d+1)!} && \text{by our assumption} \\
&= \frac{(p-1)!(\alpha t)^d(-\alpha t)^{p-d}}{(d-1)!(p-d)!} \\
&= \frac{p!(\alpha t)^d(-\alpha t)^{p-d}}{d!(p-d)!} && \text{by multiplying by } p = d
\end{aligned}$$

We substitute in part 1 as d is greater than $p-1$. The last line follows by multiplying the numerator by p and the denominator by d (as $p = d$ in this case).

Part 3 — $d < n$

For the base case, $n = 2$ and $d = 1$ as d must be greater than 0:

$$\begin{aligned}
(Ft)^2_{i,j} &= \alpha t(Ft)_{i-1,j} - \alpha t(Ft)_{i,j} \\
&= \alpha t(-\alpha t) - \alpha t \alpha t && \text{as } i - 1 = j \\
&= \frac{2!(\alpha t)^1(-\alpha t)^{2-1}}{1!(2-1)!} \\
&= \frac{n!(\alpha t)^d(-\alpha t)^{n-d}}{d!(n-d)!} && \text{as required.}
\end{aligned}$$

We assume true for $n = p-1$ for all $d > 0$ such that $d-1 < p-1$. Then for $n = p$ for all $d > 0$ such that such that $d < p$:

$$
\begin{aligned}
(Ft)^p_{i,j} &= (Ft)_{i,\bullet} \cdot (Ft)^{p-1}_{\bullet,j} \\
&= \alpha t (Ft)^{p-1}_{i-1,j} - \alpha t (Ft)^{p-1}_{i,j} \\
&= \alpha t (Ft)^{p-1}_{i-1,j} - \frac{\alpha t (p-1)! (\alpha t)^d (-\alpha t)^{p-1-d}}{(d)!(p-1-d)!} && \text{by our assumption or part 2 (*)} \\
&= \frac{\alpha t (p-1)! (\alpha t)^{d-1} (-\alpha t)^{p-1-d+1}}{(d-1)!(p-1-d+1)!} - \frac{\alpha t (p-1)! (\alpha t)^d (-\alpha t)^{p-1-d}}{d!(p-1-d)!} && \text{by our assumption or E2 (**)} \\
&= \frac{(p-1)! (\alpha t)^d (-\alpha t)^{p-d}}{(d-1)!(p-d)!} - \frac{\alpha t (p-1)! (\alpha t)^d (-\alpha t)^{p-1-d}}{d!(p-1-d)!} \\
&= (p-1)! (\alpha t)^d (-\alpha t)^{p-d} \left(\frac{1}{(d-1)!(p-d)!} + \frac{1}{d!(p-1-d)!} \right) \\
&= (p-1)! (\alpha t)^d (-\alpha t)^{p-d} \frac{d!(p-d-1)! + (d-1)!(p-d)!}{d!(d-1)!(p-d)!(p-d-1)!} \\
&= (p-1)! (\alpha t)^d (-\alpha t)^{p-d} \frac{(p-d-1)!(d! + (d-1)!(p-d))}{d!(d-1)!(p-d)!(p-d-1)!} \\
&= (p-1)! (\alpha t)^d (-\alpha t)^{p-d} \frac{(d-1)!d + (d-1)!(p-d)}{d!(d-1)!(p-d)!} \\
&= (p-1)! (\alpha t)^d (-\alpha t)^{p-d} \frac{p}{d!(p-d)!} \\
&= \frac{p! (\alpha t)^d (-\alpha t)^{p-d}}{d!(p-d)!} && \text{as required.}
\end{aligned}
$$

(*) as $d < p$, either $d = p-1$ or d is still less than $p-1$. We can then substitute in part 2 or our assumption respectively. (**) when $2 \leq d < p$ (animal $i-1$ is at least one rank below j) we can substitute in our assumption. When $d = 1$, animal $i-1$ is animal j. We can then substitute in E2, which can be reconfigured to take on the required form: $(Ft)^{p-1}_{j,j} = (-\alpha t)^{p-1} = \frac{(p-1)!(\alpha t)^0 (-\alpha t)^{p-1}}{0!(p-1)!}$.

Altogether, the above three parts show that:

$$
\frac{(Ft)^n_{i,j}}{n!} = \begin{cases} 0 & \text{when } 0 \leq n \leq d-1 \\ \frac{(\alpha t)^d (-\alpha t)^{n-d}}{d!(n-d)!} & \text{when } n \geq d \end{cases}
$$

Thus,

$$\begin{aligned}\sum_{n=0}^{\infty}\frac{(Ft)_{i,j}^{n}}{n!} &= 0 + \frac{(\alpha t)^d}{d!}\sum_{n=d}^{\infty}\frac{(-\alpha t)^{n-d}}{(n-d)!} \\ &= \frac{(\alpha t)^d}{d!}\sum_{n=0}^{\infty}\frac{(-\alpha t)^{n}}{n!} \\ &= \frac{(\alpha t)^d e^{-\alpha t}}{d!}\end{aligned}$$

E4 — $i \neq j$ and i is lead by j

We will show $(Ft)_{i,1}^{n} = \frac{\alpha}{\alpha-\beta}((Ft)_{i-1,1}^{n} - (Ft)_{i,i(2)}^{n})$ for all $n \geq 0$ by induction for two scenarios: when i is subordinate to j ($d = 1$) and when i is further down the hierarchy ($d > 1$). Let's consider the former first. Note, in this scenario, $i - 1 = j = 1$. For the base case, $n = 0$:

$$\begin{aligned}(Ft)^0 = I_k \quad \implies \quad (Ft)_{i,j}^{0} &= 0 && \text{as } i \neq j \\ &= \frac{\alpha}{\alpha-\beta}(1-1) \\ &= \frac{\alpha}{\alpha-\beta}((Ft)_{i-1,1}^{0} - (Ft)_{i,i(2)}^{0}) && \text{as } i-1 = 1 \text{ and } i(2) = i\end{aligned}$$

We assume true for $n = p - 1$. Then for $n = p$:

$$\begin{aligned}(Ft)_{i,1}^{p} &= (Ft)_{i,\bullet}\cdot(Ft)_{\bullet,1}^{p-1} \\ &= \alpha t(Ft)_{i-1,1}^{p-1} - \alpha t(Ft)_{i,1}^{p-1} \\ &= \alpha t(Ft)_{i-1,1}^{p-1} - \alpha t(\frac{\alpha}{\alpha-\beta}((Ft)_{i-1,1}^{p-1} - (Ft)_{i,i(2)}^{p-1})) && \text{by our assumption} \\ &= \alpha t(Ft)_{i-1,1}^{p-1} - \frac{\alpha^2 t}{\alpha-\beta}(Ft)_{i-1,1}^{p-1} + \frac{\alpha^2 t}{\alpha-\beta}(Ft)_{i,i(2)}^{p-1} \\ &= \alpha t(Ft)_{i-1,1}^{p-1}(1 - \frac{\alpha}{\alpha-\beta}) + \frac{\alpha^2 t}{\alpha-\beta}(Ft)_{i,i(2)}^{p-1} \\ &= \alpha t(Ft)_{i-1,1}^{p-1}\frac{-\beta}{\alpha-\beta} + \frac{\alpha^2 t}{\alpha-\beta}(Ft)_{i,i(2)}^{p-1} \\ &= \frac{\alpha}{\alpha-\beta}(Ft)_{i-1,1}^{p} + \frac{\alpha^2 t}{\alpha-\beta}(Ft)_{i,i(2)}^{p-1} && \text{by E1 and } i-1 = 1 \\ &= \frac{\alpha}{\alpha-\beta}(Ft)_{i-1,1}^{p} - \frac{\alpha}{\alpha-\beta}(Ft)_{i,i(2)}^{p} && \text{by E2 and } i(2) = i \\ &= \frac{\alpha}{\alpha-\beta}((Ft)_{i-1,1}^{p} - (Ft)_{i,i(2)}^{p})\end{aligned}$$

Now let's consider the scenario when i is further down the hierarchy. For the base case, $n = 0$:

$$\begin{aligned}(Ft)^0 = I_k \quad \implies \quad (Ft)^0_{i,j} &= 0 && \text{as } i \neq j \\ &= \frac{\alpha}{\alpha-\beta}(0-0) \\ &= \frac{\alpha}{\alpha-\beta}((Ft)^0_{i-1,1} - (Ft)^0_{i,i(2)})\end{aligned}$$

which holds for all $d > 1$. We assume true for $n = p-1$ for all $d > 1$. Then for $n = p$:

$$\begin{aligned}(Ft)^p_{i,j} &= (Ft)_{i,\bullet} \cdot (Ft)^{p-1}_{\bullet,1} \\ &= \alpha t(Ft)^{p-1}_{i-1,1} - \alpha t(Ft)^{p-1}_{i,1} \\ &= \alpha t(Ft)^{p-1}_{i-1,1} - \alpha t(\frac{\alpha}{\alpha-\beta}((Ft)^{p-1}_{i-1,1} - (Ft)^{p-1}_{i,i(2)})) && \text{by our assumption} \\ &= \alpha t(\frac{\alpha}{\alpha-\beta}((Ft)^{p-1}_{i-2,1} - (Ft)^{p-1}_{i-1,i(2)})) - \alpha t(\frac{\alpha}{\alpha-\beta}((Ft)^{p-1}_{i-1,1} - (Ft)^{p-1}_{i,i(2)})) && \text{by our assumption} \\ &= \frac{\alpha^2 t}{\alpha-\beta}(Ft)^{p-1}_{i-2,1} - \frac{\alpha^2 t}{\alpha-\beta}(Ft)^{p-1}_{i-1,i(2)} - \frac{\alpha^2 t}{\alpha-\beta}(Ft)^{p-1}_{i-1,1} + \frac{\alpha^2 t}{\alpha-\beta}(Ft)^{p-1}_{i,i(2)} \\ &= \frac{\alpha}{\alpha-\beta}[(\alpha t(Ft)^{p-1}_{i-2,1} - \alpha t(Ft)^{p-1}_{i-1,1})) - (\alpha t(Ft)^{p-1}_{i-1,i(2)} - \alpha t(Ft)^{p-1}_{i,i(2)}))] \\ &= \frac{\alpha}{\alpha-\beta}((Ft)^p_{i-1,1} - (Ft)^p_{i,i(2)}) && \text{by equation A.3}\end{aligned}$$

The second use of our induction assumption isn't applicable when $d = 2$ as the rank difference between $i-1$ and j is 1. However, in that case, we can substitute in the result proved above relating to that scenario. As $(Ft)^n_{i,1} = \frac{\alpha}{\alpha-\beta}((Ft)^n_{i-1,1} - (Ft)^n_{i,i(2)})$ for all $n \geq 0$ and $d \geq 1$, it therefore holds that $\sum_{n=0}^{\infty} \frac{(Ft)^n_{i,1}}{n!} = \frac{\alpha}{\alpha-\beta}(\sum_{n=0}^{\infty} \frac{(Ft)^n_{i-1,1}}{n!} - \sum_{n=0}^{\infty} \frac{(Ft)^n_{i,i(2)}}{n!})$ as both sides of the equality have been modified identically. As such, $e^F_{i,1} = \frac{\alpha}{\alpha-\beta}(e^{Ft}_{i-1,1} - e^{Ft}_{i,i(2)})$.

A.2 Δ

This section contains the derivation of the general form (in terms of our social framework) of the stationary covariance matrix, Δ. Again, we are still considering the case when there is an attraction term for leading/independent animals. This general form is:

$$\Delta_{i,j} = \begin{cases} \frac{\rho^2}{2\beta} & \text{D1 — when } i = j \text{ and } i \text{ is a leading/independent animal} \\ \frac{\alpha}{\alpha+\beta}\Delta_{dom(i),j} & \text{D2 — when } i \neq j \text{ and } i \text{ is lead by } j \\ \frac{\Delta_{i,dom(j)}+\Delta_{dom(i),j}}{2} & \text{D3 — when } i \neq j,\ i \text{ and } j \text{ are in the same subgroup but neither are the leader} \\ \frac{\sigma^2}{2\alpha} + \Delta_{dom(i),i} & \text{D4 — when } i = j \text{ and } i \text{ is a subordinate} \\ 0 & \text{D5 — when } i \neq j \text{ and } i \text{ and } j \text{ are in different subgroups} \end{cases}$$

where ρ and σ are the noise coefficients for leading/independent animals and subordinate animals respectively.

In Niu *et al.* (2016), this matrix is derived from:

$$\Sigma^T\Sigma = -(F\Delta + \Delta F^T) \tag{A.4}$$

where Σ is the diagonal matrix containing the relevant noise coefficients. Thus, $\Sigma^T\Sigma$ is a diagonal matrix with ρ^2 at leading/independent animal indices and σ^2 at subordinate animal indices. We too will use equation A.4 to derive Δ.

D1 — $i = j$ and i is a leading/independent animal

$$(\Sigma^T\Sigma)_{i,i} = -(F_{i,\bullet}\Delta_{\bullet,i} + \Delta_{i,\bullet}F^T_{\bullet,i})$$

$$(\Sigma^T\Sigma)_{i,i} = -((-\beta)\Delta_{i,i} + \Delta_{i,i}(-\beta)) \quad \text{as the } i \text{ row of F and } i \text{ column of } F^T \text{ are 0 except for } -\beta \text{ at the } i \text{ index}$$

$$\rho^2 = 2\beta\Delta_{i,i} \quad \text{as the } i \text{ index of } \Sigma^T\Sigma \text{ is } \rho^2 \text{ when } i \text{ is leading/independent}$$

$$\Delta_{i,i} = \frac{\rho^2}{2\beta} \quad \text{as required}$$

D2 — $i \neq j$ and i is lead by j

$$(\Sigma^T\Sigma)_{i,j} = -(F_{i,\bullet}\Delta_{\bullet,j} + \Delta_{i,\bullet}F^T_{\bullet,j})$$

$$(\Sigma^T\Sigma)_{i,j} = -(\alpha\Delta_{dom(i),j} - \alpha\Delta_{i,j} + \Delta_{i,\bullet}F^T_{\bullet,j}) \quad \text{as the } i \text{ row of F is 0 except for } \alpha \text{ at the } dom(i) \text{ index and } -\alpha \text{ at the } i \text{ index}$$

$$(\Sigma^T\Sigma)_{i,j} = -(\alpha\Delta_{dom(i),j} - \alpha\Delta_{i,j} + \Delta_{i,j}(-\beta)) \quad \text{as the } j \text{ column of } F^T \text{ is 0 except for } -\beta \text{ at the } j \text{ index}$$

$$0 = \Delta_{i,j}(\alpha + \beta) - \alpha\Delta_{dom(i),j} \quad \text{as the (i, j) index of } \Sigma^T\Sigma \text{ is off-diagonal}$$

$$\Delta_{i,j} = \frac{\alpha}{\alpha+\beta}\Delta_{dom(i),j} \quad \text{as required}$$

D3 — $i \neq j$, i and j are in the same subgroup but neither are the leader

$$(\Sigma^T\Sigma)_{i,j} = -(F_{i,\bullet}\Delta_{\bullet,j} + \Delta_{i,\bullet}F^T_{\bullet,j})$$

$$(\Sigma^T\Sigma)_{i,j} = -(\alpha\Delta_{dom(i),j} - \alpha\Delta_{i,j} + \Delta_{i,\bullet}F^T_{\bullet,j})$$ as the i row of F is 0 except for α at the $dom(i)$ index and $-\alpha$ at the i index

$$(\Sigma^T\Sigma)_{i,j} = -(\alpha\Delta_{dom(i),j} - \alpha\Delta_{i,j} + \Delta_{i,dom(j)}\alpha + \Delta_{i,j}(-\alpha))$$ as the j column of F^T is 0 except for α at the $dom(j)$ index and $-\alpha$ at the j index

$$0 = 2\Delta_{i,j} - (\Delta_{dom(i),j} + \Delta_{i,dom(j)})$$ as the (i, j) index of $\Sigma^T\Sigma$ is off-diagonal

$$\Delta_{i,j} = \frac{\Delta_{dom(i),j} + \Delta_{i,dom(j)}}{2}$$ as required

D4 — $i = j$ and i is a subordinate

$$(\Sigma^T\Sigma)_{i,i} = -(F_{i,\bullet}\Delta_{\bullet,i} + \Delta_{i,\bullet}F^T_{\bullet,i})$$

$$(\Sigma^T\Sigma)_{i,i} = -(\alpha\Delta_{dom(i),i} - \alpha\Delta_{i,i} + \Delta_{i,\bullet}F^T_{\bullet,i})$$ as the i row of F is 0 except for α at the $dom(i)$ index and $-\alpha$ at the i index

$$(\Sigma^T\Sigma)_{i,i} = -(\alpha\Delta_{dom(i),i} - \alpha\Delta_{i,i} + \Delta_{i,dom(i)}\alpha + \Delta_{i,i}(-\alpha))$$ as the i column of F^T is 0 except for α at the $dom(i)$ index and $-\alpha$ at the i index

$$\sigma^2 = -2\alpha\Delta_{dom(i),i} + 2\alpha\Delta_{i,j}$$ as the (i, i) index of $\Sigma^T\Sigma$ is σ^2 when i is a subordinate, $i = j$ and Δ is symmetric

$$\Delta_{i,j} = \frac{\sigma^2}{2\alpha} + \Delta_{dom(i),i}$$ as required

D5 — $i \neq j$ and i and j are in different subgroups

This term relates to the scenario when i and j are in different subgroups. Subgroups are independent of each other and thus the covariance of animals from two different subgroups is 0.

A.3 Ξ

This section contains the derivation of the general form of the conditional covariance matrix Ξ and it follows on from establishing the general form of e^{Ft} and the stationary matrix Δ. As shown in the algorithm for computing Ξ (Appendix B), when calculating the Ξ terms for a particular animal we only consider the terms that relate to animals ranked higher than itself in the same subgroup and animals with the same rank as itself (again, in the same subgroup) who we have already iterated through. The covariance terms with animals lower than itself in the hierarchy will be calculated

when the algorithm iterates through those animals.

As mentioned earlier, the form of Ξ can be derived using equation A.1. We will therefore derive the various terms of Ξ using that equation and then manipulate out β for the case when we model the movement of leading/independent animals with BM.

Using equation A.1, the terms of Ξ consist explicitly of:

$$\Xi_{i,j} = \Delta_{i,j} - \sum_{n=1}^{\text{rank}(i)} \sum_{m=1}^{\text{rank}(j)} e(i, i(n))e(j, j(m))\Delta_{i(n),j(m)} \tag{A.5}$$

where $e(i, i(n))$ is shorthand for $e^{Ft}_{i,i(n)}$. Note, the full version of equation A.5 is:

$$\Xi_{i,j} = \Delta_{i,j} - \sum_{n=1}^{k} \sum_{m=1}^{k} e(i, n)e(j, m)\Delta_{n,m} \tag{A.6}$$

where k is the number of animals and therefore the number of rows and columns of e^{Ft}. However, the terms $e(i, n)$ and $e(j, m)$ are 0 when animals n and m are not directly above animals i and j respectively. Thus, we can reduce the full equation down to equation A.5.

The summation in equation A.1 contains two exponential terms. The first one corresponds to animal/row i and the second to animal/row j. These will be referenced to as i-exponential terms and j-exponential terms respectively.

Furthermore, we will make use of the feature of e^{Ft} where terms that correspond to the same level of relationship in the hierarchy are equivalent (except when the leader is concerned). That is, if animal i is d ranks below j in the hierarchy and k is also d ranks below l, then $e(i, j) = e(k, l)$. This is particularly useful when we want $e(i, i(3))$ say, but we only know $e(dom(i), dom(i)(2))$.

The following sections will derive the five covariance expressions in turn:

$$\Xi(F,t)_{i,j} = \begin{cases} \rho^2 t, & \text{C1} \\ \Xi_{\text{dom}(i),j} - \frac{\rho^2 e^{\boldsymbol{F}t}_{i,j}}{\alpha}, & \text{C2} \\ \frac{\Xi_{\text{dom}(i),j} + \Xi_{i,\text{dom}(j)}}{2} - \frac{\rho^2 e^{\boldsymbol{F}t}_{i,l} e^{\boldsymbol{F}t}_{j,l}}{2\alpha} - \frac{\sigma^2 (e^{\boldsymbol{F}t}_{i,-l} \cdot e^{\boldsymbol{F}t}_{j,-l})}{2\alpha}, & \text{C3} \\ \Xi_{i,\text{dom}(i)} - \frac{\rho^2 e^{\boldsymbol{F}t}_{i,l} e^{\boldsymbol{F}t}_{i,l}}{2\alpha} + \frac{\sigma^2 (1 - e^{\boldsymbol{F}t}_{i,-l} \cdot e^{\boldsymbol{F}t}_{i,-l})}{2\alpha}, & \text{C4} \\ 0, & \text{C5} \end{cases}$$

C1 — $i = j$ and i is a leading/independent animal

$$
\begin{aligned}
\Xi_{i,i} &= \Delta_{i,i} - \sum_{n=1}^{1}\sum_{m=1}^{1} e(i,i(n))e(i,i(m))\Delta_{i(n),i(m)} && \text{the rank of } i \text{ is 1}\\
&= \frac{\rho^2}{2\beta} - e(i,i(1))e(i,i(1))\frac{\rho^2}{2\beta} && \text{substitute in } \Delta \text{ terms and } i(1)=i\\
&= \frac{\rho^2}{2\beta} - e^{-2\beta t}\frac{\rho^2}{2\beta} && \text{substitute in } e^{Ft} \text{ terms}\\
&= \frac{\rho^2}{2\beta}(1-(1+(-2\beta t)+\frac{(-2\beta t)^2}{2!}+\frac{(-2\beta t)^3}{3!}+...)) && \text{expand out the exponential}\\
&= \frac{\rho^2}{2\beta}(2\beta t-\frac{(-2\beta t)^2}{2!}-\frac{(-2\beta t)^3}{3!}...))\\
&= \frac{\rho^2 2\beta t}{2\beta}(1+\frac{(-2\beta t)}{2!}+\frac{(-2\beta t)^2}{3!}...))\\
&= \rho^2 t(1+\frac{(-2\beta t)}{2!}+\frac{(-2\beta t)^2}{3!}...))\\
&= \rho^2 t && \text{when } \beta = 0
\end{aligned}
$$

C2 — $i \neq j$ and j is the leader of i's subgroup

Taking a cue from the stationary covariance, we might expect the conditional covariance between animal i and it's leader to relate to the conditional covariance between i's dominant and their mutual leader. In order to find the difference between $\Xi_{i,j}$ and $\Xi_{dom(i),j}$ we want $\Xi_{dom(i),j}$ to contain as many similar terms to $\Xi_{i,j}$ as possible. That is, transform

$$\Xi_{dom(i),j} = \Delta_{dom(i),j} - \sum_{n=1}^{\text{rank}(dom(i))} e(dom(i),dom(i)(n))e(j,j)\Delta_{dom(i)(n),j}$$

to look like

$$\Xi_{i,j} = \Delta_{i,j} - \sum_{n=1}^{\text{rank}(i)} e(i,i(n))e(j,j)\Delta_{i(n),j}$$

Note, $rank(j) = 1$ and therefore there is a single j-exponential term. We do the following to $\Xi_{dom(i),j}$:

- the $dom(i)$-exponential term relating to the subgroup leader, $e(dom(i),dom(i)(1))$, is substituted with $e(dom(i),dom(i)(1)) = \frac{\alpha-\beta}{\alpha}e(i,i(1)) + e(i,i(2))$ using results from e^{Ft} (see E4). Note, $dom(i)(1) = i(1) = j$. This almost provides us with the first two required i-exponential

terms.

- all other $dom(i)$-exponential terms, $e(dom(i), dom(i)(2)), ..., e(dom(i), dom(i)(rank(dom(i)))$, can be rewritten in terms of i. For example, $e(dom(i), dom(i)(2)) = e(i, i(3))$ as $dom(i)$ and $dom(i)(2)$ have the same rank difference as i and $i(3)$ (see E3 in e^{Ft}). This provides us with the remaining $e(i, i(3)), ..., e(i, i(rank(i))$ terms with $rank(i) = rank(dom(i)) + 1$.
- multiply all terms by $\frac{\alpha}{\alpha+\beta}$
- note $\Delta_{dom(i)(n),j} = \Delta_{i(n),j}$ and, from D2 in Δ, $\Delta_{i(n),j} = \frac{\alpha}{\alpha+\beta}\Delta_{i(n-1),j}$

Now, we have:

$$\frac{\alpha}{\alpha+\beta}\Xi_{dom(i),j} = \frac{\alpha}{\alpha+\beta}\Delta_{dom(i),j} - \sum_{n=1}^{\text{rank}(i)} e(i, i(n))e(j, j)\Delta(n)$$

where

$$\Delta(n) = \begin{cases} \Delta_{i(1),j}\frac{\alpha-\beta}{\alpha+\beta} & \text{when } n = 1 \\ \Delta_{i(n),j} & \text{otherwise} \end{cases}$$

When subtracting $\frac{\alpha}{\alpha+\beta}\Xi_{dom(i),j}$ from $\Xi_{i,j}$, we are left with the following:

$$\begin{aligned}
\Xi_{i,j} - \frac{\alpha}{\alpha+\beta}\Xi_{dom(i),j} &= -e(i, i(1))e(j, j)\Delta_{i(1),j} + \\
&\quad \frac{\alpha-\beta}{\alpha+\beta}e(i, i(1))e(j, j)\Delta_{i(1),j} && \text{as all other terms cancel out} \\
&= -(1 - \frac{\alpha-\beta}{\alpha+\beta})e(i, j)e(j, j)\Delta_{j,j} && \text{as } i(1) = j \\
&= -\frac{2\beta}{\alpha+\beta}e(i, j)e^{-\beta t}\frac{\rho^2}{2\beta} && \text{substitute in from } \Delta \text{ and } e^{Ft} \\
&= -\frac{\rho^2}{\alpha+\beta}e(i, j)e^{-\beta t} \\
\Xi_{i,j} - \Xi_{dom(i),j} &= -\frac{\rho^2}{\alpha}e(i, j) && \text{when } \beta = 0 \\
&= -\frac{\rho^2}{\alpha}e^{Ft}_{i,j}
\end{aligned}$$

C3 — $i \neq j$, i and j are in the same subgroup and neither are the leader

C3 and C4 follow a similar process to term C2. Taking a cue from the stationary covariance, we might expect the conditional covariance between animal i and another in the same subgroup, say j, to be related to the conditional covariance between i and j's dominant and the conditional covariance between j and i's dominant.

Similar to C2, we want to make the terms that make up $\Xi_{i,dom(j)}$ and $\Xi_{dom(i),j}$ as similar as possible

to those of $\Xi_{i,j}$. That is, transform

$$\Xi_{i,dom(j)} = \Delta_{i,dom(j)} - \sum_{n=1}^{\text{rank}(i)} \sum_{m=1}^{\text{rank}(dom(j))} e(i,i(n))e(dom(j),dom(j)(m))\Delta_{i(n),dom(j)(m)}$$
$$\Xi_{dom(i),j} = \Delta_{dom(i),j} - \sum_{n=1}^{\text{rank}(dom(i))} \sum_{m=1}^{\text{rank}(j)} e(dom(i),dom(i)(n))e(j,j(m))\Delta_{dom(i)(n),j(m)}$$

to look like

$$\Xi_{i,j} = \Delta_{i,j} - \sum_{n=1}^{\text{rank}(i)} \sum_{m=1}^{\text{rank}(j)} e(i,i(n))e(j,j(m))\Delta_{i(n),j(m)}$$

To start with, we do the following to $\Xi_{i,dom(j)}$:

- the $dom(j)$-exponential term relating to the subgroup leader, $e(dom(j),dom(j)(1))$, is substituted with $e(dom(j),dom(j)(1)) = \frac{\alpha-\beta}{\alpha}e(j,j(1)) + e(j,j(2))$ using results from e^{Ft} (see E4). Note, $dom(j)(1) = j(1)$. This almost provides us with the first two required j-exponential terms.
- all other $dom(j)$-exponential terms, $e(dom(j),dom(j)(2)),...,$ $e(dom(j),dom(j)(rank(dom(j))))$, can be rewritten in terms of j. For example, $e(dom(j),dom(j)(2)) = e(j,j(3))$ as $dom(j)$ and $dom(j)(2)$ have the same rank difference as j and $j(3)$ (see E3 in e^{Ft}). This provides us with the remaining $e(j,j(3)),...,e(j,j(rank(j))$ terms with $rank(j) = rank(dom(j)) + 1$.

So now,

$$\Xi_{i,dom(j)} = \Delta_{i,dom(j)} - \sum_{n=1}^{\text{rank}(i)} \sum_{m=1}^{\text{rank}(j)} e(i,i(n))e(j,j(m))\Delta(m)$$

where

$$\Delta(m) = \begin{cases} \Delta_{i(n),j(1)}\frac{\alpha-\beta}{\alpha} & \text{when } m = 1 \\ \Delta_{i(n),j(m-1)} & \text{otherwise} \end{cases}$$

Similarly, for $\Xi_{dom(i),j}$:

- the $dom(i)$-exponential term relating to the subgroup leader, $e(dom(i),dom(i)(1))$, is substituted with $e(dom(i),dom(i)(1)) = \frac{\alpha-\beta}{\alpha}e(i,i(1)) + e(i,i(2))$ using results from e^{Ft} (see E4). Note, $dom(i)(1) = i(1)$. This almost provides us with the first two required i-exponential terms.
- all other $dom(i)$-exponential terms, $e(dom(i),dom(i)(2)),...,e(dom(i),dom(i)(rank(dom(i))),$

can be rewritten in terms of i. For example, $e(dom(i), dom(i)(2)) = e(i, i(3))$ as $dom(i)$ and $dom(i)(2)$ have the same rank difference as i and $i(3)$ (see E3 in e^{Ft}). This provides us with the remaining $e(i, i(3)), ..., e(i, i(rank(i))$ terms with $rank(i) = rank(dom(i)) + 1$.

So now,

$$\Xi_{dom(i),j} = \Delta_{dom(i),j} - \sum_{n=1}^{\text{rank}(i)} \sum_{m=1}^{\text{rank}(j)} e(i, i(n))e(j, j(m))\Delta(n)$$

where

$$\Delta(n) = \begin{cases} \Delta_{i(1),j(m)} \frac{\alpha-\beta}{\alpha} & \text{when } n = 1 \\ \Delta_{i(n-1),j(m)} & \text{otherwise} \end{cases}$$

Adding the new representations of $\Xi_{i,dom(j)}$ and $\Xi_{dom(i),j}$ together produces the following:

$$\Xi_{i,dom(j)} + \Xi_{dom(i),j} = 2\Delta(i, j) - \sum_{n=1}^{\text{rank}(i)} \sum_{m=1}^{\text{rank}(j)} f(n, m)$$

where

$$f(n,m) = \begin{cases} 2\frac{\alpha-\beta}{\alpha} e(i, i(1))e(j, j(1))\Delta_{i(1),i(1)} & (1) \text{ when } n = m = 1 \\ e(i, i(1))e(j, j(m))[\Delta_{i(1),j(m-1)} + \frac{\alpha-\beta}{\alpha}\Delta_{i(1),j(m)}] & (2) \text{ when } n = 1, m > 1 \\ e(i, i(n))e(j, j(1))[\Delta_{i(n-1),j(1)} + \frac{\alpha-\beta}{\alpha}\Delta_{i(n),j(1)}] & (3) \text{ when } n > 1, m = 1 \\ 2e(i, i(n))e(j, j(m))\Delta_{i(n),j(m-1)} & (4) \text{ when } n, m > 1 \text{ and } i(n) = j(m) \\ e(i, i(n))e(j, j(m))[\Delta_{i(n),j(m-1)} + \Delta_{i(n-1),j(m)}] & (5) \text{ otherwise} \end{cases}$$

Term (2) can be rearranged as follows, using D2 from Δ in both the first and last step:

$$\begin{aligned} e(i, i(1))e(j, (j(m))[\Delta_{i(1),j(m-1)} + \frac{\alpha-\beta}{\alpha}\Delta_{i(1),j(m)}] &= e(i, i(1))e(j, (j(m))\Delta_{i(1),j(m-1)}(1 + \frac{\alpha-\beta}{\alpha+\beta}) \\ &= 2\frac{\alpha}{\alpha+\beta} e(i, i(1))e(j, (j(m))\Delta_{i(1),j(m-1)} \\ &= 2e(i, i(1))e(j, (j(m))\Delta_{i(1),j(m)} \end{aligned}$$

and similarly for term (3). Term (4) is obtained as when $i(n) = j(m)$ then subsequently $i(n-1) = j(m-1)$ (note, $\Delta_{i(n),j(m-1)}$ could also be written as $\Delta_{i(n-1),j(m)}$). Term (5) is simplified using D3 from Δ:

$$e(i, i(n))e(j, j(m))[\Delta_{i(n),j(m-1)} + \Delta_{i(n-1),j(m)}] = 2e(i, i(n))e(j, j(m))\Delta_{i(n),j(m)}$$

Then, when subtracting $\frac{\Xi_{i,dom(j)} + \Xi_{dom(i),j}}{2}$ from $\Xi_{i,j}$, terms (2), (3) and (5) cancel out with their

counterpart in $\Xi_{i,j}$ and we are left with:

$$\begin{aligned}
\Xi_{i,j} - \frac{\Xi_{i,dom(j)} + \Xi_{dom(i),j}}{2} &= -e(i,i(1))e(j,j(1))\Delta_{i(1),j(1)}(1 - \frac{\alpha-\beta}{\alpha}) \\
&\quad - \sum_{n=2}^{\text{rank}(i)} \sum_{m=2}^{\text{rank}(j)} \mathbf{1}_{i(n)=j(m)} \left(e(i,i(n))e(j,j(m))[\Delta_{i(n),j(m)} - \Delta_{i(n),j(m-1)}]\right) \\
&= -e(i,i(1))e(j,j(1))\frac{\rho^2}{2\beta}\frac{\beta}{\alpha} \\
&\quad - \sum_{n=2}^{\text{rank}(i)} \sum_{m=2}^{\text{rank}(j)} \mathbf{1}_{i(n)=j(m)} \left(e(i,i(n))e(j,j(m))[\Delta_{i(n),j(m)} - \Delta_{i(n),j(m-1)}]\right) \\
&= -\frac{\rho^2}{2\alpha}e(i,i(1))e(j,j(1)) \\
&\quad - \sum_{n=2}^{\text{rank}(i)} \sum_{m=2}^{\text{rank}(j)} \mathbf{1}_{i(n)=j(m)} \left(e(i,i(n))e(j,j(m))[\Delta_{i(n),j(m)} - \Delta_{i(n),j(m-1)}]\right)
\end{aligned}$$

As i-exponential terms are 0 when relating to animals not directly above it in the hierarchy (and similarly for the j-exponential terms), the double summation equates to a dot product of the i^{th} and j^{th} rows of e^{Ft} with the exception of the $i(1) = j(1)$ column. Furthermore, as $n, m > 1$ and $i(n) = j(m)$, the difference $\Delta_{i(n),j(m)} - \Delta_{i(n),j(m-1)}$ is always $\frac{\sigma^2}{2\alpha}$ as $j(m-1) = dom(j(m)) = dom(i(n))$ (see D4 in Δ). Thus,

$$\Xi_{i,j} = \frac{\Xi_{\text{dom}(i),j} + \Xi_{i,\text{dom}(j)}}{2} - \frac{\rho^2 e^{\boldsymbol{F}t}_{i,l} e^{\boldsymbol{F}t}_{j,l}}{2\alpha} - \frac{\sigma^2(e^{\boldsymbol{F}t}_{i,-l} \cdot e^{\boldsymbol{F}t}_{j,-l})}{2\alpha}$$

C4 — $i = j$ and i is a subordinate

Once again, taking a cue from the stationary covariance, we might expect the conditional variance of a subordinate animal to relate to the conditional covariance between that animal and it's dominant. Similar to previous derivations, we want to make the terms that make up $\Xi_{i,dom(j)}$ as similar as possible to those of $\Xi_{i,j}$. That is, transform

$$\Xi_{i,dom(i)} = \Delta_{i,dom(i)} - \sum_{n=1}^{\text{rank}(i)} \sum_{m=1}^{\text{rank}(dom(i))} e(i,i(n))e(dom(i),dom(i)(m))\Delta_{i(n),dom(i)(m)}$$

to look like

$$\Xi_{i,i} = \Delta_{i,i} - \sum_{n=1}^{\text{rank}(i)} \sum_{m=1}^{\text{rank}(i)} e(i,i(n))e(i,i(m))\Delta_{i(n),i(m)}$$

We do the following to $\Xi_{i,dom(i)}$:

- the $dom(i)$-exponential terms relating to the subgroup leader, $e(dom(i), dom(i)(1))$, is substituted with $e(dom(i), dom(i)(1)) = \frac{\alpha-\beta}{\alpha}e(i, i(1)) + e(i, i(2))$ using results from e^{Ft} (see E4). Note, $dom(i)(1) = i(1)$. This almost provides us with the first two required i-exponential terms.
- all other $dom(i)$-exponential terms, $e(dom(i), dom(i)(2)), ..., e(dom(i), dom(i)(rank(dom(i)))$, can be rewritten in terms of i. For example, $e(dom(i), dom(i)(2)) = e(i, i(3))$ as $dom(i)$ and $dom(i)(2)$ have the same rank difference as i and $i(3)$ (see E3 in e^{Ft}). This provides us with the remaining $e(i, i(3)), ..., e(i, i(rank(i))$ terms with $rank(i) = rank(dom(i)) + 1$.

So now,

$$\Xi_{i,dom(i)} = \Delta_{i,dom(i)} - \sum_{n=1}^{\text{rank}(i)} \sum_{m=1}^{\text{rank}(i)} e(i, i(n))e(i, i(m))\Delta(m)$$

where

$$\Delta(m) = \begin{cases} \Delta_{i(n),i(1)}\frac{\alpha-\beta}{\alpha} & \text{when } m = 1 \\ \Delta_{i(n),i(m-1)} & \text{otherwise} \end{cases}$$

When subtracting $\Xi_{i,dom(i)}$ from $\Xi_{i,i}$ we are left with the following:

$$\Xi_{i,i} - \Xi_{i,dom(i)} = \Delta_{i,i} - \Delta_{i,dom(i)} + \sum_{n=1}^{\text{rank}(i)} \sum_{m=1}^{\text{rank}(i)} f(n, m)$$

where

$$f(n,m) = \begin{cases} -e(i,i(1))e(i,i(1))\Delta_{i(1),i(1)}(1 - \frac{\alpha-\beta}{\alpha}) & (1) \text{ when } n = m = 1 \\ -e(i,i(1))e(i,i(m))[\Delta_{i(1),i(m)} - \Delta_{i(1),i(m-1)}] & (2) \text{ when } n = 1, m > 1 \\ -e(i,i(n))e(i,i(1))[\Delta_{i(n),i(1)} - \frac{\alpha-\beta}{\alpha}\Delta_{i(n),i(1)}] & (3) \text{ when } n > 1, m = 1 \\ -e(i,i(n))e(i,i(m))[\Delta_{i(n),i(m)} - \Delta_{i(n),i(m-1)}] & (4) \text{ when } n = m > 1 \\ -e(i,i(n))e(i,i(m))[\Delta_{i(n),i(m)} - \Delta_{i(n),i(m-1)}] & (5) \text{ otherwise} \end{cases}$$

$\Delta_{i,i} - \Delta_{i,dom(i)}$ simplifies to $\frac{\sigma^2}{2\alpha}$ from D4 in Δ. Term (1) simplifies as follows:

$$\begin{aligned} -e(i,i(1))e(i,i(1))\Delta_{i(1),i(1)}(1 - \frac{\alpha-\beta}{\alpha}) &= -e(i,i(1))e(i,i(1))\frac{\rho^2}{2\beta}(1 - \frac{\alpha-\beta}{\alpha}) \\ &= -\frac{\rho^2}{2\alpha}e(i,i(1))e(i,i(1)) \end{aligned}$$

Term (2) simplifies after making use of D2 from Δ and setting $\beta = 0$ for the last step:

$$\begin{aligned}
-e(i,i(1))e(i,i(m))[\Delta_{i(1),i(m)} - \Delta_{i(1),i(m-1)}] &= -e(i,i(1))e(i,i(m))\Delta_{i(1),i(m-1)}(\frac{\alpha}{\alpha+\beta} - 1) \\
&= -e(i,i(1))e(i,i(m))\frac{\rho^2}{2\beta}(\frac{\alpha}{\alpha+\beta})^{m-2}\frac{-\beta}{\alpha+\beta} \\
&= \frac{\rho^2}{2\alpha}e(i,i(1))e(i,i(m))
\end{aligned}$$

Term (3) simplifies in a similar fashion, again using D2 from Δ and setting $\beta = 0$ for the last step:

$$\begin{aligned}
-e(i,i(n))e(i,i(1))\Delta_{i(n),i(1)}(1 - \frac{\alpha-\beta}{\alpha}) &= -e(i,i(n))e(i,i(1))\Delta_{i(n),i(1)}\frac{\beta}{\alpha} \\
&= -e(i,i(n))e(i,i(1))\frac{\rho^2}{2\beta}(\frac{\alpha}{\alpha+\beta})^{n-1}\frac{\beta}{\alpha} \\
&= -\frac{\rho^2}{2\alpha}e(i,i(n))e(i,i(1))
\end{aligned}$$

Terms (2) and (3) cancel each other out as there will be corresponding terms for each n and m. Term (4) simplifies using D4 from in Δ:

$$-e(i,i(n))e(i,i(m))[\Delta_{i(n),i(m)} - \Delta_{i(n),i(m-1)}] = -\frac{\sigma^2}{2\alpha}e(i,i(n))e(i,i(m))$$

when $n = m > 1$. Term (5) will eventually cancel out. For each combination of n and m (when $n \neq m$ and $n, m > 1$), there is a corresponding term with the values of n and m switched (after making use of the symmetry of Δ). We can group these together and make use of D3 from Δ as follows:

$$\begin{aligned}
-e(i,i(n))e(i,i(m))[\Delta_{i(n),i(m)} - \Delta_{i(n),i(m-1)} + \Delta_{i(m),i(n)} - \Delta_{i(m),(n-1)}] \\
= -e(i,i(n))e(i,i(m))[2\Delta_{i(n),i(m)} - 2\Delta_{i(n),i(m)}] \\
= 0
\end{aligned}$$

We are then left with the following:

$$\begin{aligned}
\Xi_{i,i} - \Xi_{i,dom(i)} &= \frac{\sigma^2}{2\alpha} - \frac{\rho^2}{2\alpha}e(i,i(1))e(i,i(1)) - \frac{\sigma^2}{2\alpha}\sum_{n=2}^{\text{rank}(i)}\sum_{m=2}^{\text{rank}(i)}\mathbf{1}_{n=m}\left(e(i,i(n))e(i,i(m))\right) \\
&= -\frac{\rho^2}{2\alpha}e(i,i(1))e(i,i(1)) + \frac{\sigma^2}{2\alpha}(1 - \sum_{n=2}^{\text{rank}(i)}\sum_{m=2}^{\text{rank}(i)}\mathbf{1}_{n=m}\left(e(i,i(n))e(i,i(m))\right))
\end{aligned}$$

As the double summation equates to a dot product of the i^{th} row of e^{Ft} and itself with the exception

of the $i(1)$ column, then:

$$\Xi_{i,i} = \Xi_{i,\mathrm{dom}(i)} - \frac{\rho^2 e^{\boldsymbol{F}t}_{i,l} e^{\boldsymbol{F}t}_{i,l}}{2\alpha} + \frac{\sigma^2(1 - e^{\boldsymbol{F}t}_{i,-l} \cdot e^{\boldsymbol{F}t}_{i,-l})}{2\alpha}$$

C5 — $i \neq j$ and i and j are in different subgroups

$$\Xi_{i,j} = \Delta_{i,j} - \sum_{n=1}^{\mathrm{rank}(i)} \sum_{m=1}^{\mathrm{rank}(j)} e(i, i(n)) e(j, j(m)) \Delta_{i(n),j(m)}$$

For all n and m, animals $i(n)$ and $j(m)$ will be in different subgroups and so $\Delta_{i(n),j(m)} = 0$ for all n and m. The resulting expression is therefore 0.

Appendix B

Conditional Covariance Computation

input : *ord*: order the animals in the hierarchy as found in $\boldsymbol{F}$ from top to bottom. Multiple animals in the same level of the hierarchy (or of the same rank) can be in any order.
output: $\boldsymbol{\Xi}(\boldsymbol{F}, t)$
for *i in ord* **do**
  if *i is a leader* **then**
    calculate $\boldsymbol{\Xi}(\boldsymbol{F}, t)_{i,i}$ using (a)
  else
    calculate $\boldsymbol{\Xi}(\boldsymbol{F}, t)_{i,l(i)} = \boldsymbol{\Xi}(\boldsymbol{F}, t)_{l(i),i}$ using (b)
    before $\leftarrow$ *ord*$[2 : (i-1)]$
    for *j in before* **do**
      calculate $\boldsymbol{\Xi}(\boldsymbol{F}, t)_{i,j} = \boldsymbol{\Xi}(\boldsymbol{F}, t)_{j,i}$ using (c)
    end
    calculate $\boldsymbol{\Xi}(\boldsymbol{F}, t)_{i,i}$ using (d)
end

Algorithm 1: How to compute $\boldsymbol{\Xi}(\boldsymbol{F}, t)$. $l(i)$ is the leader of i and *before* corresponds to all the animals before i in *ord* except for $l(i)$.

Appendix C

Recursive OU distribution

Between two observations, there are $m \geq 2$ time points consisting of the two observations and $m-2$ sampled switching times. To calculate the likelihood of the movement between the observations, say at times t_1 and t_m, we perform a recursive calculation through t_i (for $i = 1, ..., m$) to find the distribution of the movement at t_m. The following equations are conditional on $t_1, ..., t_m, \beta_1, ..., \beta_m$ and the movement parameters, where β_i are the behaviour states of all animals at time t_i.

For $i = 2, ..., m$, we can write

$$\boldsymbol{G}_{t_i} = \tilde{\boldsymbol{A}}_{t_i}\boldsymbol{G}_{t_{i-1}} + \tilde{\boldsymbol{B}}_{t_i}\boldsymbol{\Theta}_{t_{i-1}} + \tilde{\boldsymbol{C}_{t_i}} \tag{C.1}$$

from the definition of the OU process. Here, $\boldsymbol{G}_{t_i}$ is the location of the animals at time t_i in a particular axis; $\tilde{\boldsymbol{A}}_{t_i} = e^{\boldsymbol{F}_{t_{i-1}}(t_i - t_{i-1})}$; $\tilde{\boldsymbol{B}}_{t_i} = (1 - e^{\boldsymbol{F}_{t_{i-1}}(t_i - t_{i-1})})$; $\tilde{\boldsymbol{C}_{t_i}} \sim \text{MVN}(0, \boldsymbol{\Xi}(\boldsymbol{F}_{t_{i-1}}, t_i - t_{i-1}))$; $\boldsymbol{\Theta}_{t_{i-1}}$ is a vector which contains the location of each animal's leader at time t_{i-1}; $\boldsymbol{F}_{t_{i-1}}$ is the configuration of the attraction matrix at time t_{i-1}.

Through writing $\boldsymbol{G}_{t_{i-1}}$ as follows (which we can do for all i)

$$\boldsymbol{G}_{t_{i-1}} \sim \text{MVN}(\boldsymbol{\eta}_{i-1}, \boldsymbol{\xi}_{i-1}) \tag{C.2}$$

then it follows that $\boldsymbol{G}_{t_i}$ is also a multivariate normal, with mean

$$\boldsymbol{\eta}_i = \tilde{\boldsymbol{A}}_{t_i}\boldsymbol{\eta}_{i-1} + \tilde{\boldsymbol{B}}_{t_i}\boldsymbol{\Theta}_{t_{i-1}} \tag{C.3}$$

and variance-covariance matrix

$$\boldsymbol{\xi}_i = \boldsymbol{\Xi}(\boldsymbol{F}_{t_{i-1}}, t_i - t_{i-1}) + \tilde{\boldsymbol{A}}_{t_i}\boldsymbol{\xi}_{i-1}\tilde{\boldsymbol{A}}_{t_i}^T \tag{C.4}$$

Since we know $\boldsymbol{G}_{t_1}$, $\boldsymbol{G}_{t_1} = \boldsymbol{g}$ say, then the recursive process starts with

$$\boldsymbol{\eta}_1 = \boldsymbol{g} \tag{C.5}$$

$$\boldsymbol{\xi}_1 = 0 \tag{C.6}$$

Appendix D

LDB Distribution Derivation

Switching Time of Animal i

For a potential switching time for animal i between two observations, t_a and t_b, we need to sample the location of the whole group. Let $t_0 = t_a$ and let a potential switching time, t_s, for animal i be the s^{th} switching time in the interval $(t_0, t_s]$, with $t_s < t_b$. In general, the times of the $1^{st}, ..., s^{th}$ switches correspond to times $t_1, ..., t_s$. Let $\tau_c = \{t_{s-1}, t_s, t_b\}$, $\beta_c = \{\beta_{s-1}, \beta_s\}$ (where β_s is a vector containing the states of all animals at time t_s) and Ω represent all movement and transition parameters and the coefficient of diffusion of the LDP. The location of the group at time t_s in a particular axis, given their location at time t_{s-1}, is distributed as follows:

$$\boldsymbol{G}_s|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \tau_c, \beta_c, \Omega \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Xi}) \tag{D.1}$$

$$\boldsymbol{G}_s|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \tau_c, \beta_c, \Omega = \boldsymbol{A}\boldsymbol{g}_{s-1} + \boldsymbol{B}\boldsymbol{\Theta}_{s-1} + \boldsymbol{C} \tag{D.2}$$

where equation D.2 is simply equation D.1 written out more explicitly. Here we have used a simplified notation, compared to equation 3.5, where $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{G}_{s-1}, \boldsymbol{F}_{s-1}, \delta t)$, $\boldsymbol{\Xi} = \boldsymbol{\Xi}(\boldsymbol{F}_{s-1}, \delta t)$ and $\delta t = t_s - t_{s-1}$. $\boldsymbol{A} = e^{\boldsymbol{F}_{s-1}(t_s - t_{s-1})}$, $\boldsymbol{B} = 1 - \boldsymbol{A}$, $\boldsymbol{C} \sim \text{MVN}(0, \boldsymbol{\Xi})$ and $\boldsymbol{F}_{s-1}$ is the attraction matrix at time t_{s-1}. The respective leader's locations in $\boldsymbol{\Theta}_{s-1}$ are taken from $\boldsymbol{g}_{s-1}$. Note, here we are conditioning on β_s (which we will not have sampled yet) and t_b even though $\boldsymbol{G}_s$ is independent of both. However, it is useful to include them to provide consistency with other conditional distributions during the derivation of equation D.6. With the forwards simulation method, the locations of the animals at t_s are proposed with the distribution in equation D.1.

We, however, want to propose the locations at t_s conditionally also on the next known locations,

$\boldsymbol{G}_b$. Firstly, we note:

$$\boldsymbol{G}_b|\boldsymbol{G}_s = \boldsymbol{g}_s, \tau_c, \beta_c, \Omega \sim \text{MVN}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Xi}}) \tag{D.3}$$

$$\boldsymbol{G}_b|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \tau_c, \beta_c, \Omega \sim \text{MVN}(\boldsymbol{\mu}^R, \boldsymbol{\Xi}^R) \tag{D.4}$$

$$\boldsymbol{G}_b|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \tau_c, \beta_c, \Omega = \tilde{\boldsymbol{A}}\boldsymbol{G}_s + \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s + \tilde{\boldsymbol{C}} \tag{D.5}$$

where equation D.3 is the distribution of the movement from t_s to t_b. Equation D.4 is the recursive distribution of the movement through the times t_{s-1}, t_s and t_b. The recursive method with which we obtain this distribution is the same one used in Chapters 3 and 4 and more details can be found in Appendix C and Blackwell (2003). Equation D.5 is the distribution of equation D.4 explicitly written out where $\tilde{\boldsymbol{A}} = e^{\boldsymbol{F}_s(t_b - t_s)}$, $\tilde{\boldsymbol{B}} = 1 - \tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{C}} \sim \text{MVN}(0, \tilde{\boldsymbol{\Xi}})$ and the leader's locations in $\boldsymbol{\Theta}_s$ are taken from the expectation of $\boldsymbol{G}_s$, $\boldsymbol{\mu}$.

We can use the above and the following result to derive $\boldsymbol{G}_s|\boldsymbol{G}_{s-1}, \boldsymbol{G}_b$:

$$\begin{pmatrix} \boldsymbol{G}_s \\ \boldsymbol{G}_b \end{pmatrix} |\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \tau_c, \beta_c, \Omega \sim \text{MVN}(\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}^R \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Xi} & (\tilde{A}\boldsymbol{\Xi})^T \\ \tilde{A}\boldsymbol{\Xi} & \boldsymbol{\Xi}^R \end{pmatrix} \tag{D.6}$$

where $(\tilde{\boldsymbol{A}}\boldsymbol{\Xi})^T$ is the transpose of $\tilde{\boldsymbol{A}}\boldsymbol{\Xi}$. The $\text{Cov}[\boldsymbol{G}_b, \boldsymbol{G}_s|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \tau_c, \beta_c, \Omega]$ is derived from the following:

$$\begin{aligned}
\text{Cov}[\boldsymbol{G}_b, \boldsymbol{G}_s|*] &= \text{E}[\boldsymbol{G}_b\boldsymbol{G}_s|*] - \text{E}[\boldsymbol{G}_b|*]\text{E}[\boldsymbol{G}_s|*] && \text{(D.7)}\\
&= \text{E}[(\tilde{\boldsymbol{A}}\boldsymbol{G}_s + \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s + \tilde{\boldsymbol{C}})\boldsymbol{G}_s|*] - \boldsymbol{\mu}^R\boldsymbol{\mu} && \text{(D.8)}\\
&= \text{E}[\tilde{\boldsymbol{A}}\boldsymbol{G}_s\boldsymbol{G}_s|*] + \text{E}[\tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s\boldsymbol{G}_s|*] - \boldsymbol{\mu}^R\boldsymbol{\mu} && \text{(D.9)}\\
&= \tilde{\boldsymbol{A}}\text{E}[\boldsymbol{G}_s^2|*] + \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s\boldsymbol{\mu} - \boldsymbol{\mu}^R\boldsymbol{\mu} && \text{(D.10)}\\
&= \tilde{\boldsymbol{A}}\text{Var}[\boldsymbol{G}_s|*] + \tilde{\boldsymbol{A}}\text{E}[\boldsymbol{G}_s|*]^2 + \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s\boldsymbol{\mu} - \boldsymbol{\mu}^R\boldsymbol{\mu} && \text{(D.11)}\\
&= \tilde{\boldsymbol{A}}\text{Var}[\boldsymbol{G}_s|*] + (\tilde{\boldsymbol{A}}\boldsymbol{\mu} + \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s)\boldsymbol{\mu} - \boldsymbol{\mu}^R\boldsymbol{\mu} && \text{(D.12)}\\
&= \tilde{\boldsymbol{A}}\text{Var}[\boldsymbol{G}_s|*] + \text{E}[\boldsymbol{G}_b|*]\boldsymbol{\mu} - \boldsymbol{\mu}^R\boldsymbol{\mu} && \text{(D.13)}\\
&= \tilde{\boldsymbol{A}}\text{Var}[\boldsymbol{G}_s|*] && \text{(D.14)}\\
&= \tilde{\boldsymbol{A}}\boldsymbol{\Xi} && \text{(D.15)}
\end{aligned}$$

To simplify notation, $*$ is used as shorthand to represent being conditional on $\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \tau_c, \beta_c$ and Ω. Equation D.9 omits $\text{E}[\tilde{\boldsymbol{C}}\boldsymbol{G}_s|*] = 0$, which is true as $\tilde{\boldsymbol{C}}|*$ and $\boldsymbol{G}_s|*$ are independent and $\text{E}[\tilde{\boldsymbol{C}}|*] = 0$.

Having derived equation D.6, we can then use standard results for a conditional multivariate normal

distribution such that:

$$\mathrm{E}[\boldsymbol{G}_s|\boldsymbol{G}_{s-1}=\boldsymbol{g}_{s-1},\boldsymbol{G}_b=\boldsymbol{g}_b,\tau_c,\beta_c,\Omega]=\boldsymbol{\mu}+(\tilde{A}\boldsymbol{\Xi})^T(\boldsymbol{\Xi}^R)^{-1}(\boldsymbol{g}_b-\boldsymbol{\mu}^R) \tag{D.16}$$

$$\mathrm{Var}[\boldsymbol{G}_s|\boldsymbol{G}_{s-1}=\boldsymbol{g}_{s-1},\boldsymbol{G}_b=\boldsymbol{g}_b,\tau_c,\beta_c,\Omega]=\boldsymbol{\Xi}-(\tilde{A}\boldsymbol{\Xi})^T(\boldsymbol{\Xi}^R)^{-1}(\tilde{A}\boldsymbol{\Xi}) \tag{D.17}$$

Note that the behavioural states at t_s (β_s) that feed in to $\boldsymbol{F}_s$ are the same as those at t_{s-1} but with the state/movement process of the focal animal changed to the LDP. Consequently, as the LDP has no attraction term, if the focal animal is in a subordinate state from t_{s-1} to t_s, it will lose its social link to its dominant (and the rest of the subgroup that is 'above' it) for the bridge step from t_s to t_b. Whilst that may not seem ideal, we do not know if that animal will retain that social link when we come to sample its state at t_s.

Switching Time of Another Animal

Whilst resampling the location of animal i at the switching time of another animal, we treat the locations and behaviours of the other animals as fixed. Say t_s, still within (t_a, t_b), is now a switching time for one of the other animals. We can use the information of the group's location at time t_b to propose more reasonable samples for animal i's location at t_s. First, note the following:

$$G_{s_i}|\boldsymbol{G}_{s-1}=\boldsymbol{g}_{s-1},\boldsymbol{G}_{s_{-i}}=\boldsymbol{g}_{s_{-i}},\tau_c,\beta_c,\Omega\sim\mathrm{N}(\mu,\Xi) \tag{D.18}$$

$$\boldsymbol{G}_s|\boldsymbol{G}_{s-1}=\boldsymbol{g}_{s-1},\boldsymbol{G}_{s_{-i}}=\boldsymbol{g}_{s_{-i}},\tau_c,\beta_c,\Omega=(g_{s_1},...,g_{s_{i-1}},G_{s_i},g_{s_{i+1}},...,g_{s_n})^T \tag{D.19}$$

$$\boldsymbol{G}_b|\boldsymbol{G}_{s-1}=\boldsymbol{g}_{s-1},\boldsymbol{G}_{s_{-i}}=\boldsymbol{g}_{s_{-i}},\tau_c,\beta_c,\Omega\sim\mathrm{MVN}(\boldsymbol{\mu}^R,\boldsymbol{\Xi}^R) \tag{D.20}$$

$$\boldsymbol{G}_b|\boldsymbol{G}_{s-1}=\boldsymbol{g}_{s-1},\boldsymbol{G}_{s_{-i}}=\boldsymbol{g}_{s_{-i}},\tau_c,\beta_c,\Omega=\tilde{\boldsymbol{A}}\boldsymbol{G}_s+\tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s+\tilde{\boldsymbol{C}} \tag{D.21}$$

where $\boldsymbol{G}_{s_{-i}}$ are the locations of all animals except i at time t_s. μ and Ξ in equation D.18 are derived from equation D.1 when we 'know' the locations of all animals except i at t_s and therefore we can condition on them. Equation D.19 concatenates this univariate random variable with the fixed locations of the other animals at time t_s, where n is the number of animals in the group in total. We wrap the univariate distribution of equation D.18 in this vector to create consistent dimensions in equation D.21 and further derivations below. $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{B}}$, $\tilde{\boldsymbol{C}}$ and $\boldsymbol{\Theta}_s$ in equation D.21 are the same as in equation D.5 above.

Similarly to using equation D.6 in the case of animal i's switching time, we can use equation D.22 here:

$$\begin{pmatrix}G_{s_i}\\ \boldsymbol{G}_b\end{pmatrix}|\boldsymbol{G}_{s-1}=\boldsymbol{g}_{s-1},\boldsymbol{G}_{s_{-i}}=\boldsymbol{g}_{s_{-i}},\tau_c,\beta_c,\Omega\sim\mathrm{MVN}(\begin{pmatrix}\mu\\ \boldsymbol{\mu}^R\end{pmatrix},\begin{pmatrix}\Xi & (\tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\Xi)^T\\ \tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\Xi & \boldsymbol{\Xi}^R\end{pmatrix} \tag{D.22}$$

where $\mathbf{1}_{\boldsymbol{i}}$ represents a column vector of 0s with a 1 at the i^{th} position, the length of which is the number of animals in the group. The $\mathrm{Cov}[\boldsymbol{G}_b, G_{s_i}|\boldsymbol{G}_{s-1}=\boldsymbol{g}_{s-1},\boldsymbol{G}_{s_{-i}}=\boldsymbol{g}_{s_{-i}},\tau_c,\beta_c,\Omega]$ is derived

from the following:

$$\begin{aligned}
\mathrm{Cov}[\boldsymbol{G}_b, G_{s_i}|*] &= \mathrm{E}[\boldsymbol{G}_b G_{s_i}|*] - \mathrm{E}[\boldsymbol{G}_b|*]\mathrm{E}[G_{s_i}|*] && \text{(D.23)}\\
&= \mathrm{E}[(\tilde{\boldsymbol{A}}\boldsymbol{G}_s + \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s + \tilde{\boldsymbol{C}})G_{s_i}|*] - \boldsymbol{\mu}^R\mu && \text{(D.24)}\\
&= \mathrm{E}[(\tilde{\boldsymbol{A}}(g_{s_1}, ..., g_{s_{i-1}}, G_{s_i}, g_{s_{i+1}}, ..., g_{s_n})^T + \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s + \tilde{\boldsymbol{C}})G_{s_i}|*] - \boldsymbol{\mu}^R\mu && \text{(D.25)}\\
&= \mathrm{E}[\tilde{\boldsymbol{A}}(0, ..., 0, G_{s_i}, 0, ..., 0)^T G_{s_i} + \tilde{\boldsymbol{A}}(g_{s_1}, ..., g_{s_{i-1}}, 0, g_{s_{i+1}}, ..., g_{s_n})^T G_{s_i} + \\
&\quad \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s G_{s_i}|*] - \boldsymbol{\mu}^R\mu && \text{(D.26)}\\
&= \tilde{\boldsymbol{A}}\mathrm{E}[(0, ..., 0, G_{s_i}^2, 0, ..., 0)^T|*] + \tilde{\boldsymbol{A}}(g_{s_1}, ..., g_{s_{i-1}}, 0, g_{s_{i+1}}, ..., g_{s_n})^T\mathrm{E}[G_{s_i}|*] + \\
&\quad \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s\mathrm{E}[G_{s_i}|*] - \boldsymbol{\mu}^R\mu && \text{(D.27)}\\
&= \tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\mathrm{E}[G_{s_i}^2|*] + \tilde{\boldsymbol{A}}(g_{s_1}, ..., g_{s_{i-1}}, 0, g_{s_{i+1}}, ..., g_{s_n})^T\mu + \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s\mu - \boldsymbol{\mu}^R\mu && \text{(D.28)}\\
&= \tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\mathrm{Var}[G_{s_i}|*] + \tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\mu^2 + \tilde{\boldsymbol{A}}(g_{s_1}, ..., g_{s_{i-1}}, 0, g_{s_{i+1}}, ..., g_{s_n})^T\mu + \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s\mu - \\
&\quad \boldsymbol{\mu}^R\mu && \text{(D.29)}\\
&= \tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\mathrm{Var}[G_{s_i}|*] + (\tilde{\boldsymbol{A}}(g_{s_1}, ..., g_{s_{i-1}}, \mu, g_{s_{i+1}}, ..., g_{s_n})^T + \tilde{\boldsymbol{B}}\boldsymbol{\Theta}_s)\mu - \boldsymbol{\mu}^R\mu && \text{(D.30)}\\
&= \tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\mathrm{Var}[G_{s_i}|*] + \mathrm{E}[\boldsymbol{G}_b|*]\mu - \boldsymbol{\mu}^R\mu && \text{(D.31)}\\
&= \tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\Xi && \text{(D.32)}
\end{aligned}$$

To simplify notation, $*$ is used as shorthand to represent being conditional on $\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \boldsymbol{G}_{s_{-i}} = \boldsymbol{g}_{s_{-i}}, \tau_c, \beta_c$ and Ω. Equation D.26 omits $\mathrm{E}[\tilde{\boldsymbol{C}}\boldsymbol{G}_{s_i}|*] = 0$, which is true due to independence and $\mathrm{E}[\tilde{\boldsymbol{C}}|*] = 0$.

Having derived equation D.22, we can then use standard results for a conditional multivariate normal distribution such that:

$$\mathrm{E}[G_{s_i}|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \boldsymbol{G}_b = \boldsymbol{g}_b, \boldsymbol{G}_{s_{-i}} = \boldsymbol{g}_{s_{-i}}, \tau_c, \beta_c, \Omega] = \mu + (\tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\Xi)^T(\boldsymbol{\Xi}^R)^{-1}(\boldsymbol{g}_b - \boldsymbol{\mu}^R) \quad \text{(D.33)}$$

$$\mathrm{Var}[G_{s_i}|\boldsymbol{G}_{s-1} = \boldsymbol{g}_{s-1}, \boldsymbol{G}_b = \boldsymbol{g}_b, \boldsymbol{G}_{s_{-i}} = \boldsymbol{g}_{s_{-i}}, \tau_c, \beta_c, \Omega] = \Xi - (\tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\Xi)^T(\boldsymbol{\Xi}^R)^{-1}(\tilde{\boldsymbol{A}}\mathbf{1}_{\boldsymbol{i}}\Xi) \quad \text{(D.34)}$$