Stochastic models of animal movement and habitat selection

The University Of Sheffield.

Théo Michelot

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

This dissertation is submitted for the degree of
Doctor of Philosophy

April 2019

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This 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 in the following paragraph.

Parts of Chapters 3 and 4 were published in Michelot et al. (2019), and in the preprint Michelot et al. (2018a), co-authored by Paul Blackwell, Jason Matthiopoulos, and Simon Chamaillé-Jammes. Chapter 5 was collaborative work with Pierre Gloaguen and Marie-Pierre Étienne, and is covered in the preprint Michelot et al. (2018b). I describe their contributions in the introduction of that chapter. Chapter 6 extends work presented in Michelot and Blackwell (2019), written with Paul Blackwell. Paul Blackwell provided a supervisory role for all chapters. Michelot et al. (2018a) and Michelot et al. (2018b) are currently under review.

Théo Michelot
April 2019

Acknowledgements

I would like to express my sincere gratitude to my supervisor Paul Blackwell, for his support and his patience during the last three years. His unalterable calm has been a great source of comfort in the face of stress! I am very grateful to Jason Matthiopoulos, for his excellent ideas and suggestions, on which much of the work presented in this thesis is based.

I would like to thank others who have helped me in my work, during these past few years. In particular, I couldn't have done this without the help of Roland Langrock, who introduced me to academic research and to statistical models of animal movement. I would also like to thank Theoni Photopoulou and Vianey Leos Barajas, who have been great academic role models (and amazing friends). Thanks to Ruth King and Toby Patterson, for giving me great opportunities, and for their mentorship.

I am grateful to Debbie Russell and Esther Jones, who provided the grey seal data used in Chapter 6, and to Simon Chamaillé-Jammes, who provided the zebra data used in Chapter 4. I thank the editors and referees who have provided feedback on parts of this work. I am very grateful to Pierre Gloaguen for providing feedback on Chapter 5, and Alison Parton for her feedback on Chapter 6.

Thanks to my friends, who gave me many excuses to think about something else other than work: Claire, Gabriel, Corentin, Julien, Micka'il. Most of all, thanks to Bastien, for his continued remote support, and thanks to Magda, for making Sheffield such a great place to live.

Finally, I would like to thank my parents and my brother, Lucas, for everything else.

Abstract

The analysis of animal movement reveals important features of habitat preferences and behaviours, and informs environmental conservation decisions. In this thesis, we present new statistical methods, to tackle the problem of scale dependence in models of animal movement. The inferences obtained from most existing approaches are tied to a particular spatio-temporal scale, which makes the interpretation and comparison of results difficult.

We first focus on models of habitat selection, which combine tracking data and environmental data, to understand the drivers of animal movement. The two most popular approaches describe habitat selection on two different scales, and their parameters have different interpretations. We propose a time series approach to integrate local and global habitat selection. We explain how stochastic processes with known stationary distributions can be used, to describe both the short-term transition density and the long-term equilibrium distribution of the movement. The proposed approach captures both the short-term and long-term habitat selection. We suggest using Markov chain Monte Carlo (MCMC) algorithms to model animal movement. A MCMC algorithm describes transition rules, which lead to a limiting distribution: its target distribution. We also suggest the Langevin diffusion process as a continuous-time model of movement with known stationary distribution. We describe methods of estimation, to obtain habitat selection and movement parameters from tracking data.

We then turn to the problem of the time formulation in models of animal movement and behaviour. Most widely-used models describe movement in discrete time, and their results are tied to the time scale of the observed data. We extend a popular continuous-time model of movement, to include behavioural heterogeneity. The approach can be used to identify behavioural phases from movement data collected at irregular intervals, and with measurement error. We describe a framework of Bayesian inference, to estimate movement parameters and behavioural phases from tracking data.

‘Les hommes ? Il en existe, je crois, six ou sept. Je les ai aperçus il y a des années. Mais on ne sait jamais où les trouver. Le vent les promène. Ils manquent de racines, ça les gêne beaucoup.’

Antoine de Saint-Exupéry, *Le Petit Prince*

Table of contents

List of figures

List of tables

Nomenclature

List of Abbreviations

CTCRW	Continuous-time correlated random walk
ESS	Effective sample size
GPS	Global Positioning System
HMM	Hidden Markov model
MCMC	Markov chain Monte Carlo
pdf	Probability density function
RSF	Resource selection function
SSF	Step selection function

Chapter 1

Biological motivation

In this thesis, we explore several methods of statistical modelling and inference, to analyse ecological data. In this chapter, we present the ecological data, and we introduce some of the statistical problems arising from their analysis. In Section 1.4, we present the aim of this thesis, and we outline its structure.

1.1 Animal movement data

A common theme in all chapters of this thesis is the use of statistics to study the movements of wild animals. Understanding how and why animals move is crucial to learn about their behaviour, their habitat preferences, and their patterns of space use. Here, we describe the type of data that we will consider throughout.

Electronic tags have been used for several decades to monitor the location of wild animals through time (Cagnacci et al., 2010; Block et al., 2011; Kays et al., 2015). An animal is equipped with a tag (e.g. a collar), and its locations are recorded at subsequent time points, until the battery runs out or the tag falls off the animal. Most telemetry devices now use satellite positioning systems, such as the Global Positioning System (GPS) or the Argos system. The accuracy of their measurements may vary greatly. The location error of GPS tags is usually of the order of 50m or less and, for many studies, it is small compared with the scale of the movement (Frair et al., 2010). However, Argos devices are less accurate, and their error ranges between several hundred metres and several kilometres (Patterson et al., 2010).

Improvements in battery duration have made it possible to collect data at ever-increasing temporal resolutions, and over longer periods of time. Modern data sets may for example include locations observed at the time scale of every minute, over several weeks or several months (e.g. Grecian et al., 2018). Recording each position reduces the battery life, and there still exists a trade-off between the sampling frequency and the duration of the study (Brown et al., 2012). The choice of this trade-off should be informed by the aim of the study. For example, it may be useful to have high-frequency data to learn about small-scale behaviour, but it may be more useful to cover a long period of time to learn about the spatial extent of an animal's territory.

Telemetry devices most often record the locations of an animal in two dimensions: longitude and latitude. The data then consist of a bivariate time series of locations, indexed by the times of observation. In studies of animal movement, it is sometimes useful to think of the atomic unit of data as the displacement between two successive locations, that we will call "step". Note that planar geometry cannot be applied in longitude-latitude space, because it describes a sphere. Instead, spherical geometry must be used to describe movement, e.g. using great circle distances and bearings. Alternatively, the raw locations can be projected onto a plane, to make such calculations simpler. In animal tracking studies, the most common method of projection is the Universal Transverse Mercator (UTM) coordinate system, which divides the globe into 60 rectangles (White and Garrott, 2012). Unless stated otherwise, we will always consider locations projected onto a plane in this thesis, for mathematical convenience.

In the case of marine animals and birds, the coordinate along the third dimension (depth or altitude) can sometimes be measured, e.g. using pressure sensors. Other types of data can also be collected on wild animals using telemetry tags, such as accelerometer data (Brown et al., 2013; Leos-Barajas et al., 2017), information about the diving activity of marine mammals (DeRuiter et al., 2017; McClintock et al., 2017), or physiological variables (Hooten et al., 2018). Animal locations can also be recorded in mark-recapture studies, as described e.g. by Ovaskainen (2004). In that setting, an animal is captured at the start of the study, it is individually marked (so that it can be recognized), and then released. In subsequent capture occasions, the study region is surveyed and, if the marked animal is observed, its location is recorded. That procedure often results in sparse movement data sets, because an animal may only be observed directly by the scientists, during a survey. However, it is convenient for animals on which telemetry devices cannot be attached, e.g. insects (Ovaskainen, 2004). Throughout this thesis, we focus on two-dimensional locations, as measured by GPS and Argos systems. We may refer to "telemetry data", "tracking data", "location data", and "movement data" interchangeably.

Modern telemetry data present a challenge for movement ecologists. They contain a wealth of information about animal behaviour and space use, that could be used to improve conservation. They are often large and complex data sets, with strong patterns of autocorrelation, sometimes measurement error, and many different factors affecting the movement decisions of animals. This motivates the development of adequate statistical methodology to tackle these problems, and deepen our understanding of animal ecology.

1.2 Statistical models of animal movement

The study of animal movement is important to tackle many ecological problems. For the simplest question, "Where did the animal go?", descriptive statistics or visualisation tools may be sufficient. Statistical methods have been developed to extract more information from tracking data (Hooten et al., 2017). In particular, time series models, based on stochastic processes, offer a natural framework to describe the dynamics of animal movement, and to capture the strong serial correlation that is often present in such data (Patterson et al., 2017).

Movement metrics The observed data are two-dimensional locations. In a time series model, the location of the animal can be represented by a stochastic process $(\boldsymbol{X}_t)_{t\geq 0}$, where t denotes the time. Tracking data may be obtained with measurement error, and state-space models have been proposed to separate the observation process from the dynamics of the movement (Anderson-Sprecher and Ledolter, 1991; Jonsen et al., 2003, 2005; Patterson et al., 2008, 2010). In that framework, the locations are assumed to be observed with noise, i.e. we consider the observation process

$$\tilde{\boldsymbol{X}}_t = \boldsymbol{X}_t + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t$ captures the measurement error, and is typically assumed to be Gaussian. The goal is to "filter" the noise, and infer the true locations of the animal. The Kalman filter can be implemented as a computationally efficient method to estimate the true trajectory from noisy observations (Jonsen et al., 2003; Johnson et al., 2008a; Fleming et al., 2017). This procedure is often used as the first stage of an analysis: the track is first processed to estimate the true locations, and the filtered trajectory is then used to learn about some characteristics of the animal's movement.

Modern telemetry data are often collected at a high temporal resolution, and contain higher-order autocorrelation. In particular, autocorrelation is often found in the speed and in the direction of movement, and not only in the location. Various metrics of movement can be derived from the locations, to capture this persistence. In particular, the first-order difference

$\boldsymbol{D}_t = \boldsymbol{X}_{t+1} - \boldsymbol{X}_t$ is a vector of the step taken by the animal at time t, and it measures the length and direction of a displacement (Jonsen et al., 2005). Alternatively, the step length $L_t = \|\boldsymbol{X}_{t+1} - \boldsymbol{X}_t\|$ and the turning angle $\varphi_t = \arctan(D_t^y, D_t^x) - \arctan(D_{t-1}^y, D_{t-1}^x)$ can be used, where $\boldsymbol{D}_t = (D_t^x, D_t^y)$, and where 'arctan' is the arctangent function with two arguments. The step length $L_t \geq 0$ is the distance between two successive locations, and it is often used as a measure of the speed of movement. The turning angle $\varphi_t \in (-\pi, \pi]$ is the angle between three successive locations, and it measures a change of direction. Figure 1.1 illustrates the definition of the step length and turning angle. Time series models have been developed to describe the dynamics of the processes $(\boldsymbol{D}_t)$, (L_t), or (φ_t), rather than those of the location process $(\boldsymbol{X}_t)$ itself (Bovet and Benhamou, 1988; Morales et al., 2004; Jonsen et al., 2005). They are often called "correlated random walks", because they can capture the serial correlation in the speed and in the direction of the animal's movement. Analogous continuous-time models of persistent movement can be formulated in terms of the velocity (Johnson et al., 2008a; Gurarie et al., 2009, 2017), or in terms of the continuous-time speed and bearing (Parton and Blackwell, 2017).

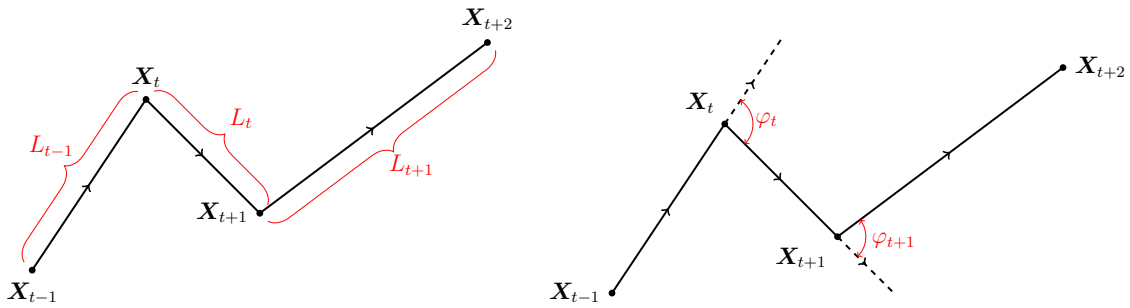


Fig. 1.1 Illustration of step lengths and turning angles.

Scales of movement Several characteristics of the process $(\boldsymbol{X}_t)$ may be of interest, depending on the aim of the study. In most of this work, we will focus on two of them: the *step distribution* (or step density), and the *utilisation distribution*.

The step density is the probability density of an animal's location at time $t+1$, given its location at time t, and we will denote it by $p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t)$. This notation omits the possible conditioning on previous locations $(\boldsymbol{X}_{t-1}, \dots)$, or on external influences, that may also be included. The step density describes the short-term dynamics of the movement model. Different specifications of the step density may be used to capture the speed of movement, or the persistence in the direction of movement. The (isotropic) Gaussian random walk is perhaps the simplest movement model; its transition density is a normal distribution centred

on the current location,

$$\boldsymbol{X}_{t+1}|\{\boldsymbol{X}_t = \boldsymbol{x}_t\} \sim N(\boldsymbol{x}_t, \sigma^2 \boldsymbol{I}_2).$$

In the Gaussian random walk, the transition density only has one parameter: the variance σ^2. In the context of animal movement, the variance parameter is a measure of the speed of movement of the animal. More complex models can be formulated to describe more detailed features of animal movement.

The utilisation distribution π is the long-term (equilibrium) distribution of the location process $(\boldsymbol{X}_t)$. It is the probability density function of the animal's location in geographical space in the long run, and it can be defined by

$$\Pr(\boldsymbol{X}_t \in \mathcal{A}) = \int_{\mathcal{A}} \pi(\boldsymbol{x}) d\boldsymbol{x},$$

where $\mathcal{A}$ is a spatial region.

The utilisation distribution may for example be uniform, if the animal is equally likely to occupy any location in space. More realistically, it could be an irregular surface, reflecting the tendency of the animal to spend more time in certain areas than others. It can be expressed as a measure of habitat suitability and habitat preference, which we describe in Section 1.3.1.

The step density and the utilisation distribution describe animal movement on two different scales. This exemplifies the observation of Turchin (1998), that animal movement can be viewed from two perspectives, termed the "Lagrangian" and the "Eulerian" points of view. The Lagrangian approach describes the movement from the point of view of the individual animal. The step density fits into the Lagrangian framework, in that it captures the movement as a process, and can be viewed as a model for the movement decisions made by the animal. On the other hand, the Eulerian approach describes the movement from the point of view of a point in space. The utilisation distribution fits into the Eulerian framework: for each point in space, it gives the probability density of the animal's presence.

The step density and the utilisation distribution are closely related, because the long-term distribution of space use arises from the accumulation of short-term displacements. However, most statistical models that have been developed to analyse animal movement have only focused on one scale. Models of individual animal movement often do not capture overall space use, and models of space use often overlook the dependence inherent to movement trajectories. Bridging the gap between these two scales is one of the challenges tackled in this thesis (Chapters 3-5).

Movement and behaviour Recent telemetry studies often cover long durations, of several months or more (e.g. Michelot et al., 2017). Over such long periods of time, most animals display behavioural heterogeneity, which is reflected in their movement patterns. Various statistical models of movement have been used to infer animal behaviour from tracking data. In particular, a large body of literature is dedicated to the classification of location data into several discrete behavioural states, e.g. "foraging", "resting", "exploring" (Blackwell, 1997; Morales et al., 2004; Jonsen et al., 2005; Patterson et al., 2017). In those multistate models, the behavioural states are assumed to differ in terms of some metrics of movement, usually linked to the speed or tortuosity of the movement, such as step lengths and turning angles. Multistate models were developed to capture behavioural heterogeneity, and also to model the serial correlation in the behaviours.

The most common formulations are based on the use of an unobserved (discrete-time or continuous-time) Markov process $(S_t)_{t\geq 0}$, which models the behavioural component of movement. In those models, S_t can take a finite number of values $\{1, \ldots, N\}$, and each is associated with specific movement characteristics. Models of movement and behaviour can often be written in the form of a hidden Markov model (Patterson et al., 2009; Langrock et al., 2012). This has greatly contributed to their growing popularity in the past few years, because efficient computational methods are available to apply hidden Markov models to large data sets (Zucchini et al., 2016; Michelot et al., 2016).

Time formulation In a time series movement model, the time can be formulated as a discrete or continuous variable. For the analysis of animal movement, the choice between discrete-time and continuous-time models is often difficult, and the relative merits of the two approaches have been debated (McClintock et al., 2014; Patterson et al., 2017). Although animals move in continuous time, their location may only be observed at discrete intervals (e.g. every minute or every hour). Discrete-time approaches are based on the assumption that the underlying movement process can be appropriately modelled at the time scale of the observations. However, most metrics derived from movement data strongly depend on the sampling rate, such that the inference of discrete-time models is tied to the temporal scale of the observations (Codling and Hill, 2005; Schlägel and Lewis, 2016).

One of the consequences of the scale dependence of discrete-time methods is that they require locations to be collected at regular time intervals through the period of the study. This is often inconvenient, because many telemetry data sets are collected at irregular time intervals. The irregularity may be fortuitous; for example, some marine mammals may only be observed when they surface. In other cases, the data may be collected at irregular intervals deliberately.

For example, the sampling frequency of some telemetry tags varies through time based on the level of activity of the animal, to save battery time when it is not moving (Brown et al., 2012). To use a discrete-time model with irregular locations, it would be necessary to interpolate the data points on a regular time grid, introducing approximation uncertainty.

On the other hand, continuous-time models consider that telemetry observations arise from a continuous movement process. As such, they can naturally accommodate different temporal scales, and irregular sampling rates (Patterson et al., 2017). Various approaches have been used to model animal movement in continuous time, most of them based on diffusion processes. These include Ornstein-Uhlenbeck processes (Dunn and Gipson, 1977; Blackwell, 1997, 2003; Blackwell et al., 2016), Brownian bridges (Horne et al., 2007), and more complex processes based on potential functions (Brillinger et al., 2001; Preisler et al., 2013).

Discrete-time models may seem like a less natural representation of animal movement, but they have been widely used because they are often simpler to formulate and to implement than their continuous-time counterparts. The time formulation can have important theoretical and practical implications, that we will discuss at several points in this thesis. In particular, in Chapters 5 and 6, we will focus on continuous-time models of animal movement, and present some of their advantages.

1.3 Habitat selection models

The study of animal movement often involves combining tracking data with other types of data. In particular, the integration of environmental data and telemetry data is promising to understand how movement decisions by animals are affected by characteristics of their habitat.

1.3.1 Resource selection functions

An important topic in animal ecology is the relationship between the distribution of space use by an animal (or population) and the distribution of some resources of interest (Manly et al., 2002). This motivates the study of resource selection, that is the lack of proportionality between resource availability and resource use by the animal. Resource availability is quantified as the distribution of resources that are available to the animal, i.e. that can be accessed and used by the animal. The distributions of resources are usually measured at regular points in space, either in a census, or using geographic information systems (GIS). Resource use is observed from location data collected on animals, e.g. using telemetry

devices such as GPS tags, or in surveys. This framework is used to measure preference or avoidance of particular resources by animals, and therefore to link the distribution of the animals' space use to habitat characteristics. Although we refer to the environmental variables as the densities of "resources", they can be any types of spatial covariates. In particular, we could investigate the relative preference of an animal for discrete habitat types, or the effect of continuous habitat features such as elevation and distance to water.

In initial habitat selection analyses, the aim was usually to compare preferences for discrete (categorical) habitat types (Neu et al., 1974). For such a categorical variable, the availability of a habitat type may be measured as the proportion of the study region that it covers, and its use is the proportion of time that the animal spends in that habitat type. Neu et al. (1974) suggested using a chi-square test to compare the availability and use in this context, and detect a potential lack of proportionality. If a lack of proportionality is detected, then confidence intervals can be defined for the expected proportion of use of each habitat. Observed values that lie outside the confidence intervals reveal preference or avoidance of the corresponding habitat type. However, many recent studies are based on continuous-valued environmental variables, for which the comparison of availability and use is more problematic.

In the context of resource selection analyses, a "spatial unit" refers to a point in geographical space, and a "resource unit" refers to a point in environmental space. This terminology implies that both spaces are discrete, which can usually be assumed because habitat covariates are measured on a discrete spatial grid. A spatial unit is a grid cell, and a resource unit is the habitat composition of a grid cell. We say that a (spatial or resource) unit is used if an animal is present in that unit, and we say that it is unused otherwise.

In cases where all used and unused resource units are known over the study region, it is possible to estimate the parameters of the resource selection probability function, which gives the probability of use of a resource unit $\boldsymbol{c} = (c_1, \ldots, c_J)$ (where J is the number of spatial covariates). It is usually modelled with a logistic model, where the response variable y is 1 if a resource unit is used, and 0 otherwise (Manly et al., 2002). The resource selection probability function is of the form

$$\Pr(y = 1|\boldsymbol{c}) = \frac{\exp(\beta_0 + \boldsymbol{\beta}'\boldsymbol{c})}{1 + \exp(\beta_0 + \boldsymbol{\beta}'\boldsymbol{c})},$$

where $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_J)$ is a vector of unknown resource selection parameters to estimate.

In most studies, we do not know which resource units were not used (Pearce and Boyce, 2006). In such cases, the intercept parameter β_0 of the resource selection probability function is not identifiable (Manly et al., 2002, Chapter 5). It is however still possible to estimate the

relative probability of use of a resource unit, called the resource selection function (RSF). The resource selection function $w(\boldsymbol{c})$ gives the relative probability of use of a resource unit $\boldsymbol{c}$, and it is usually formulated as a log-linear model,

$$w(\boldsymbol{c}) = \exp(\boldsymbol{\beta}'\boldsymbol{c}). \tag{1.1}$$

Each coefficient β_i measures the strength of selection of the animal for the covariate c_i. A positive value indicates preference, whereas a negative value indicates avoidance. The resource selection parameters $\{\beta_1, \ldots, \beta_J\}$ are most commonly estimated by logistic regression, from use-availability data (Boyce and McDonald, 1999; Johnson et al., 2006). The used sample consists of the locations $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ at which the animal was observed. The available sample $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\}$ is generated at random from a distribution of availability, and can contain used and unused units. The sample of availability is sometimes called a sample of "pseudo-absences", because it is not known whether they were actual absences (Pearce and Boyce, 2006). In most cases, the distribution of availability is assumed to be uniform over the study region, but it does not need to be (Lele and Keim, 2006; Johnson et al., 2008b). The challenge of specifying resource availability has been recognised for a long time (Johnson, 1980; Porter and Church, 1987), and different inferences can be obtained for different definitions of the available region (Beyer et al., 2010). Northrup et al. (2013) gave practical guidance for the choice of the number of pseudo-absences, and showed that estimates can be biased if the availability sample is too small.

Other approaches of inference have been proposed to estimate the RSF coefficients. In particular, Aarts et al. (2012) suggested treating the observed locations as the output of a spatial point process. They binned the telemetry observations on a discrete spatial grid, to obtain a count for each grid cell. They fitted a Poisson generalized linear model to the counts, with the resources as explanatory variables, and obtained estimates of the β_i parameters. For large sample sizes, this is equivalent to the logistic regression approach. Marzluff et al. (2004) and Millspaugh et al. (2006) proposed an alternative two-stage approach, where the utilisation distribution of the animal is first estimated using non-parametric density estimation (e.g. kernel density estimation). The non-parametric utilisation distribution, termed the "resource utilisation function" in this context, is then evaluated in each cell of the resource rasters, and a regression is used to link the utilisation to the densities of resources. Hooten et al. (2013) demonstrated that the approach is equivalent to the standard RSF under some assumptions, but that resource utilisation functions are preferable to standard RSFs when locations are observed with measurement error.

Resource selection functions are used to model the utilisation distribution of an animal or population (Johnson et al., 2006). The utilisation distribution is the long-term distribution of space use, as defined in Section 1.2. It can be modelled over the study region Ω by the normalized resource selection function,

$$\pi(\boldsymbol{x}) = \frac{w(\boldsymbol{c}(\boldsymbol{x}))}{\int_{\boldsymbol{z}\in\Omega} w(\boldsymbol{c}(\boldsymbol{z}))d\boldsymbol{z}} = \frac{\exp(\boldsymbol{\beta}'\boldsymbol{c}(\boldsymbol{x}))}{\int_{\boldsymbol{z}\in\Omega} \exp(\boldsymbol{\beta}'\boldsymbol{c}(\boldsymbol{z}))d\boldsymbol{z}}, \tag{1.2}$$

where $\boldsymbol{c}(\boldsymbol{x})$ associates a spatial location $\boldsymbol{x}$ to its habitat composition. The denominator is a normalizing constant, to ensure that π is a probability density over Ω. Based on Equation 1.2, resource selection functions rely on the assumption that the long-term distribution of space use by the animal can be written as a function of environmental covariates. In most applications, the covariates are treated as constant in time, such that data may need to be split into several periods over which the environmental (e.g. seasonal) variations can be assumed to be small (see e.g. Nielsen et al., 2003).

Resource selection functions constitute a valuable framework to describe space use as the result of habitat selection. They are most often used to identify habitats that are important for a population, to inform management and conservation efforts (Johnson et al., 2004). They have also been used to measure the effects of anthropogenic disturbances on space use by animals (Hebblewhite and Merrill, 2008), and to study animal interactions, in particular predator-prey dynamics (Hebblewhite et al., 2005).

1.3.2 Step selection functions

Step selection functions (SSFs) provide another method to estimate habitat selection from telemetry and environmental data. They were developed to tackle the difficulty of defining availability in resource selection models (Arthur et al., 1996). In resource selection analyses, availability is typically assumed to cover the whole study region. However, in reality, the choice by the animal of a location is also strongly constrained by its movement. Over a short time interval, only a small neighbourhood of its current location is available, and it must select from this restricted area. Ignoring the effect of movement can lead to misinterpretations. For example, if an animal never uses a particular habitat type, it could either mean that it is avoiding it (the typical inference from a RSF analysis), or that the animal was too far from that habitat type to perceive it. The definition of the region of availability can therefore have a strong impact on the estimated selection parameters (Beyer et al., 2010). In SSF models, the region of availability is defined by a movement model, and it varies in time.

Following Rhodes et al. (2005) and Forester et al. (2009), we define a SSF model by the likelihood of a displacement from $\boldsymbol{x}$ to $\boldsymbol{y}$, given by

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{w(\boldsymbol{c}(\boldsymbol{y}))\phi(\boldsymbol{y}|\boldsymbol{x})}{\int_{\boldsymbol{z}\in\Omega} w(\boldsymbol{c}(\boldsymbol{z}))\phi(\boldsymbol{z}|\boldsymbol{x})d\boldsymbol{z}}, \tag{1.3}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are any two points in the study region Ω. The function w is called the step selection function, and it takes the same log-linear form as the RSF, i.e. $w(\boldsymbol{c}) = \exp(\boldsymbol{\beta}'\boldsymbol{c})$. However, unlike the RSF, it is not used as a model for long-term habitat selection and space use, and it does not capture the utilisation distribution of the animal. The SSF is a model of the short-term habitat selection, at the scale of the time step of observation. The function $\phi(\boldsymbol{y}|\boldsymbol{x})$ defines the movement component of the model, and characterises the scale of availability. It gives the probability density of a step from $\boldsymbol{x}$ to $\boldsymbol{y}$ in the absence of habitat selection, and it is called the resource-independent movement kernel (Rhodes et al., 2005; Forester et al., 2009). Examples of resource-independent movement kernels that have been used in SSF analyses include a uniform distribution on a disc (Arthur et al., 1996; Rhodes et al., 2005), or a distribution of step lengths and turning angles to include persistence in direction (Fortin et al., 2005).

Fortin et al. (2005) and Potts et al. (2014a) suggest a more general formulation, in which the SSF is a function of both the origin and the endpoint of the step (rather than only the endpoint, as written in Equation 1.3). In their model, the SSF can be written $w(\boldsymbol{c}(\boldsymbol{x}, \boldsymbol{y}))$. That is, the values of the covariates depend on both $\boldsymbol{x}$ and $\boldsymbol{y}$. This is a useful extension to account for the habitat composition over the whole step, rather than only at the endpoint. For example, Fortin et al. (2005) considered the proportion of the step covered by conifers, and the minimum distance to a road over the step, as covariates in their analysis of elk movement. In the following, we will mostly focus on the simpler case where the SSF is only a function of the endpoint of the step, as described in Equation 1.3.

Conditional logistic regression is used to estimate the parameters of a SSF (Forester et al., 2009). For each observed step from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t+1}$, K random steps starting from $\boldsymbol{x}_t$ are simulated from the resource-independent movement kernel. The habitat characteristics of the observed endpoint $\boldsymbol{x}_{t+1}$ can then be contrasted to those of the potential endpoints $\{\boldsymbol{z}_{t+1}^{(1)}, \ldots, \boldsymbol{z}_{t+1}^{(K)}\}$. This procedure estimates habitat selection at the scale of the step, by comparing habitat features within a neighbourhood of the animal's location (defined by the movement kernel ϕ). If ϕ is chosen as a uniform distribution over the study region, the locations are assumed to be temporally independent, and this becomes equivalent to the RSF framework.

Increasingly complex models of animal movement have been integrated into the SSF approach. Duchesne et al. (2015) showed that a step selection model defines a movement model equivalent to a biased correlated random walk. Biased correlated random walks are routinely used in ecology as a flexible basis for models of individual movement (Turchin, 1998; Codling et al., 2008). Avgar et al. (2016) extended the step selection approach to allow simultaneous inference on habitat selection and on the movement process, making it a very attractive framework to estimate habitat preference from movement data. Nicosia et al. (2017) further extended the framework to the case where movement arises from several behavioural states, with different movement characteristics. Step selection models have been used to analyse the impact of landscape features and human presence on animal space use (e.g. Coulon et al., 2008; Roever et al., 2010; Prokopenko et al., 2017; Scrafford et al., 2018), as well as animal interactions (Potts et al., 2014b). Thurfjell et al. (2014) presented a review of SSF analyses in the ecological literature.

1.3.3 Spatially-structured diffusion models

Spatially-structured diffusion models provide another related framework to describe animal movement and habitat selection (Ovaskainen and Cornell, 2003; Ovaskainen, 2004, 2008). Those models describe animal movement as a continuous-time diffusion process, among discrete habitat types. The movement characteristics of the animal can depend on the habitat type; e.g. the animal may move faster in certain types of habitats than in others. In that formulation, the time-varying density of the animal's location in space can be discontinuous at the edges of habitat patches. This captures habitat selection at the level of the short-term movement, because the discontinuity allows for preference for one habitat over another when the animal is near the edge of a habitat patch. Ovaskainen (2008) explains how various biologically-relevant quantities can be estimated from spatially-structured diffusion models. In particular, they derive the quasi-stationary distribution of the model, which is analogous to the utilisation distribution of the animal.

Spatially-structured diffusion models and SSF models both describe animal movement in a heterogeneous environment, and they can both capture the effect of the habitat on the animal's movement decisions. One difference is that SSF models typically try to describe the fine-scale movement patterns, whereas diffusion models usually involve simpler movement assumptions. Besides, diffusion models require discrete habitat types, whereas SSFs can also be used with continuous environmental covariates.

1.4 Aim of this thesis

In this work, we develop several new statistical methods to analyse animal movement data. In particular, we focus on the problem of scale dependence in models of animal movement and habitat selection. The main contents of this thesis, presented in Chapters 3-6, can roughly be divided into three parts. First, Chapters 3 and 4 develop a new class of models of animal movement and habitat selection. These models work across spatio-temporal scales, and describe both short-term and long-term properties of animal movement. Second, Chapter 5 also tackles the problem of reconciling the local and global of habitat selection, but it is based on a different modelling approach. Third, Chapter 6 develops a new model of animal movement and behaviour, for which inference does not depend on the temporal scale of the observed data.

Chapter 3 introduces a new family of habitat selection models, that combines advantages from the two main existing frameworks: resource selection functions (RSFs) and step selection functions (SSFs). RSFs and SSFs can be useful to infer habitat preferences from location data, but the two methods are fundamentally incompatible because they describe selection at different scales. As a solution, we propose an analogy between the movement of animals and the transitions of a Markov chain Monte Carlo (MCMC) algorithm in its parameter space. This formulation defines a movement model with a known long-term distribution: the target distribution of the MCMC sampler. It captures the movement at the scale of the individual displacement, but also at the scale of the long-term use of space. We model the target distribution with a RSF, such that short-term movement and long-term space use are both described in response to environmental features. We introduce a particular MCMC algorithm, that we call the local Gibbs sampler, to describe realistic patterns of animal movement. We demonstrate the scaling properties of the proposed model in simulation studies.

In Chapter 4, we describe a method of inference for the MCMC movement model introduced in Chapter 3. The model is expressed in terms of habitat selection parameters and movement parameters. We derive the likelihood function in several special cases of the MCMC movement model, including a model with Gaussian transition density. We explain how the likelihood can be used to estimate all model parameters, from telemetry and environmental data. We use simulations to investigate the performance of the estimation method in different scenarios. We illustrate its use with the analysis of the movement track of a plains zebra.

In Chapter 5, we propose another modelling framework for the analysis of animal movement and habitat selection. Like in Chapters 3 and 4, we use a stochastic process with a known stationary distribution, to capture both the short-term movement and the long-term space

use. However, in Chapter 5, we consider a *continuous-time* process: the Langevin diffusion process. We describe a method of (approximate) inference for the Langevin diffusion movement model, to estimate habitat selection and movement parameters. Because it is formulated in continuous time, inferences obtained from this model do not depend on the time scale of observation. In a simulation study, we determine the conditions under which the method can recover the model parameters. We showcase the approach in the analysis of three tracks of Steller sea lions.

Chapter 6 covers a different research question, and presents a continuous-time model of animal movement and behaviour. Its formulation is based on an autocorrelated velocity process, to model movement persistence. The behaviour component is modelled with a continuous-time Markov chain, switching between a finite number of states, treated as proxies for behaviours of the animal. We describe a method of inference, to estimate movement parameters in each behavioural state, and to estimate the underlying state process from observed tracking data. This method can be applied to telemetry data collected at irregular time intervals and/or observed with measurement error. This flexible and widely applicable framework is a step forward to make the underutilised continuous-time movement models more accessible to ecologists. We present several simulation studies, to verify that parameter estimation works under different sampling scenarios, and to compare the proposed method to its discrete-time analogue. As an illustration, we analyse a movement track of grey seal. Chapter 6 is independent of Chapters 3-5.

Chapter 2

Statistical background

In this chapter, we introduce some of the statistical methods that will frequently be used, to avoid redundancy in the rest of the thesis. Methodological results that are only used in one part of this thesis are described in the relevant chapters.

2.1 Markov chains

A stochastic process is an ordered collection of random variables $(X_t)_{t\in\mathcal{T}}$ defined on a set $\mathcal{S}$, where $\mathcal{T}$ is a set of indices (Guttorp, 1995). In most applications, including all those covered in this thesis, the index is time, and we have $\mathcal{T} = \mathbb{N}$ (discrete time) or $\mathcal{T} = \mathbb{R}^+$ (continuous time). Different stochastic processes are characterised by different dependence structures.

In discrete time, a stochastic process $(X_t)_{t\in\mathbb{N}}$ is called a Markov chain, or Markov process, if it satisfies

$$p(X_{t+1}|X_t, X_{t-1}, \ldots, X_1) = p(X_{t+1}|X_t).$$

This property is called the Markov property. It states that the distribution of the process at time $t+1$, conditional on its value at time t, does not depend on the past (i.e. on its value at times $1, \ldots, t-1$). A Markov chain is sometimes called a "memoryless" process, because the future of the process does not depend on its past, given the present.

The set $\mathcal{S}$ over which the process is defined is called the state space, and its elements are called "states". When the state space is finite, this conditional distribution $p(X_{t+1}|X_t)$ can be written as the collection of transition probabilities $\Pr(X_{t+1} = j|X_t = i)$, for any two states $i, j \in \mathcal{S}$. Here, we focus on homogeneous Markov chains, for which the transition probabilities do not depend on t. If we denote by $\mathcal{S} = \{1, 2, \ldots, N\}$ the finite state space,

the Markov process is fully specified by its initial distribution,

$$\boldsymbol{\pi}^{(1)} = (\Pr(X_1 = 1), \ldots, \Pr(X_1 = N)),$$

and by its transition probability matrix,

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1N} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{N1} & \gamma_{N2} & \cdots & \gamma_{NN} \end{pmatrix},$$

where

$$\gamma_{ij} = \Pr(X_{t+1} = j | X_t = i), \quad i, j = 1, \ldots, N.$$

The initial distribution describes the state of the chain at time $t = 1$, and the transition probability matrix gives its dynamics from t to $t + 1$. For a Markov chain with transition probability matrix $\boldsymbol{\Gamma}$, we call *stationary distribution* a probability distribution $\boldsymbol{\pi}$ which satisfies $\boldsymbol{\pi}\boldsymbol{\Gamma} = \boldsymbol{\pi}$. Under some mild conditions, a unique stationary distribution exists, and it is the limiting distribution of the Markov chain (Guttorp, 1995). This means that, for any initial distribution $\boldsymbol{\pi}^{(1)}$, a Markov chain will converge to its stationary distribution as t increases. This property is central to the development of Markov chain Monte Carlo methods, that we will describe in Section 2.3.

In the general case of a continuous state space $\mathcal{S}$, $p(X_{t+1}|X_t)$ is called the transition density, or transition kernel, of the process. It is a continuous probability density function over $\mathcal{S}$, and it cannot be written in matrix form like in the case of a discrete state space. In this case, the distribution π is a stationary distribution of the Markov chain (X_t) if and only if, for any $y \in \mathcal{S}$,

$$\int_{x \in \mathcal{S}} \pi(x) p(X_{t+1} = y | X_t = x) dx = \pi(y). \tag{2.1}$$

Because π is a probability density over $\mathcal{S}$, it is also subject to the constraint $\int_{\mathcal{S}} \pi(x) dx = 1$.

The Gaussian random walk described in Section 1.2 is an example of a Markov chain on a continuous state space. In that special case, the transition density is a normal distribution with mean X_t.

2.2 Monte Carlo integration

Monte Carlo integration is a general method, based on random sampling, to approximate intractable multidimensional integrals (see for example Robert and Casella, 2013). In general, Monte Carlo integration can be summarised as follows. Consider that we want to evaluate the following integral,

$$I = \int_{\boldsymbol{x} \in \mathcal{E}} f(\boldsymbol{x}) g(\boldsymbol{x}) d\boldsymbol{x}, \tag{2.2}$$

where f is a probability density function over $\mathcal{E}$. The integral can be approximated by

$$\hat{I} = \frac{1}{n} \sum_{i=1}^{n} g(\boldsymbol{X}_i),$$

where

$$\forall i \in \{1, \ldots, n\},\ \boldsymbol{X}_i \sim f.$$

We can verify that $\hat{I}$ is an unbiased estimator of I,

$$E[\hat{I}] = E\left[\frac{1}{n} \sum_{i=1}^{n} g(\boldsymbol{X}_i)\right] = \frac{1}{n} \sum_{i=1}^{n} E[g(\boldsymbol{X}_i)] = E[g(\boldsymbol{X})],$$

where $\boldsymbol{X} \sim f$. So, we find

$$E[\hat{I}] = I$$

and, from the strong law of large numbers, $\hat{I}$ converges to I with probability 1 as n goes to infinity.

2.3 Markov chain Monte Carlo

Markov chain Monte Carlo methods are commonplace in statistics, and they are used in several different contexts in this thesis. Here, we provide an overview of some of their important properties. Markov chain Monte Carlo (MCMC) is a general method to sample from any probability distribution π (Gilks et al., 1996). In this context, the distribution of interest π is called the target distribution. The core idea of MCMC is to define a Markov chain $(X_t)_{t \in \mathbb{N}}$ with stationary distribution the target distribution, and to simulate from the chain until the samples give a good approximation of π. This simulation procedure is called an MCMC algorithm, or MCMC sampler.

The Markov chain of an MCMC algorithm is specified by its transition kernel $p(X_{t+1}|X_t)$, as defined in Section 2.1. In this section, and in later chapters, we will sometimes write the transition kernel as $p(y|x) = p(X_{t+1} = y|X_t = x)$, for convenience. A sufficient condition on the transition kernel for the process to have the target distribution π as stationary distribution, is the detailed balance condition, given by

$$\forall x, y \in \mathcal{S},\ \pi(x)p(y|x) = \pi(y)p(x|y).$$

This directly follows from the definition of the stationary distribution, given in Equation 2.1. Indeed, if the transition kernel satisfies the detailed balance condition, for all $y \in \mathcal{S}$, we have

$$\begin{aligned}\int_{x\in\mathcal{S}} \pi(x)p(y|x)dx &= \int_{x\in\mathcal{S}} \pi(y)p(x|y)dx \\ &= \pi(y)\int_{x\in\mathcal{S}} p(x|y)dx \\ &= \pi(y),\end{aligned}$$

because $p(x|y)$ is a probability density for x over $\mathcal{S}$, i.e. it integrates to 1. This proves that π is a stationary distribution for a Markov chain if it satisfies detailed balance.

The detailed balance condition has been used to construct Markov chains with a known stationary distribution, to use as the basis for an MCMC algorithm. As an example, we consider the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970). In the Metropolis-Hastings algorithm, each transition of the Markov chain is composed of two parts: a proposal step, and an acceptance-rejection step. The algorithm starts from an initial state $X_0 = x_0$. Then, at each iteration $t = 1, 2, \ldots$, a new sample X^* is generated from a proposal distribution $q(X^*|X_{t-1})$. The proposed sample is accepted with a probability given by the acceptance ratio,

$$r(X_{t-1}, X^*) = \min\left\{1,\ \frac{\pi(X^*)q(X_{t-1}|X^*)}{\pi(X_{t-1})q(X^*|X_{t-1})}\right\}.$$

If the update is accepted, we take $X_t = X^*$, else $X_t = X_{t-1}$.

We can verify that the Metropolis-Hastings algorithm satisfies the detailed balance condition. For all $x, y \in \mathcal{S}$,

$$\pi(x)p(y|x) = \pi(x)q(y|x)r(x, y)$$

$$
\begin{aligned}
&= \pi(x)q(y|x) \min\left\{1,\ \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)}\right\} \\
&= \min\left\{\pi(x)q(y|x),\ \pi(y)q(x|y)\right\}
\end{aligned}
$$

By symmetry, we find

$$
\begin{aligned}
\pi(x)p(y|x) &= \min\left\{\pi(y)q(x|y),\ \pi(x)q(y|x)\right\} \\
&= \pi(y)p(x|y),
\end{aligned}
$$

as required. The distribution π is therefore a stationary distribution for the Metropolis-Hastings sampler.

Note that the Metropolis-Hastings algorithm does not directly require any evaluations of the target distribution, but only *ratios* of evaluations of the target distribution. This is a great advantage because, in most applications, the target distribution is only known up to a multiplicative constant. The constant cancels out in the acceptance ratio of the Metropolis-Hastings sampler, so it does not need to be determined to run the algorithm.

If the proposal distribution q is symmetric, i.e. if $\forall x, y \in \mathcal{S},\ q(y|x) = q(x|y)$, the acceptance ratio of the Metropolis-Hastings algorithm simplifies to

$$
r(X_{t-1}, X^*) = \min\left\{1,\ \frac{\pi(X^*)}{\pi(X_{t-1})}\right\}.
$$

In the special case of a symmetric proposal, the algorithm is often called the Metropolis sampler.

Markov chain Monte Carlo algorithms have mostly been employed to solve problems of Bayesian inference (Gelman et al., 2013). In a Bayesian analysis, the objective is often to obtain the posterior distribution of the parameters of the model, given the observed data and the model assumptions. In many applications, the posterior distribution is not analytically tractable, and MCMC samplers can be used to approximate it numerically, through simulations.

2.4 Stochastic differential equations

2.4.1 Definition

A diffusion process is a continuous-time stochastic process, defined as the solution to a stochastic differential equation (Øksendal, 2003). A stochastic differential equation is an equation of the form

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dW_t, \tag{2.3}$$

where (X_t) is the unknown process, and where (W_t) is a Wiener process. The Wiener process is an almost-surely continuous stochastic process, with stationary and independent increments, which satisfy

$$W_{t+\delta} - W_t \sim N(0, \delta),$$

for any time interval $\delta > 0$.

In Equation 2.3, the function b is called the drift term, and σ is called the diffusion term. They are usually chosen as parametric functions, and their formulation determines the dynamics of the diffusion process. Under some boundedness conditions on b and σ, and given an initial condition $X_0 = x_0$, Equation 2.3 has a unique solution (X_t) (Øksendal, 2003, Chapter 5).

The diffusion process (X_t) is a continuous-time Markov process, i.e. it satisfies

$$p(X_{t_{n+1}}|X_{t_n}, X_{t_{n-1}}, \ldots, X_{t_1}) = p(X_{t_{n+1}}|X_{t_n}),$$

where the times are ordered, i.e. $0 \leq t_1 \leq \cdots \leq t_{n-1} \leq t_n \leq t_{n+1}$. This is the continuous-time analogue of the Markov property presented in Section 2.1.

The transition density of a diffusion process is the distribution $p(X_{t+\delta}|X_t)$, for a time interval $\delta > 0$, and it is analogous to the transition density of the discrete-time Markov chains described in Section 2.1. Because it is Markovian, the diffusion process is fully specified by its initial value x_0 (or initial distribution $p(x_0)$) and its transition density.

2.4.2 A few relevant examples

The simplest diffusion process is the Brownian motion, defined by $b(X_t, t) = 0$ and $\sigma(X_t, t) = \sigma$. The Brownian motion is the continuous-time analogue of the Gaussian random walk, and it was first formulated to describe the movements of interacting particles (Brown, 1828; Einstein, 1905). The process has independent normal increments,

$$X_{t+\delta}|\{X_t = x_t\} \sim N(x_t, \sigma^2\delta),$$

for $\delta > 0$. The Brownian motion has only one parameter, the variance σ^2 of the transition density. This parameter is a measure of the speed of the process.

The Brownian motion *with drift* is a slightly more general process, for which the drift function is set to a non-zero constant, i.e. $b(X_t, t) = b$. Increments of the Brownian motion with drift are described by

$$X_{t+\delta}|\{X_t = x_t\} \sim N(x_t + b\delta, \sigma^2\delta). \tag{2.4}$$

As the name suggests, this introduces a tendency for the process to drift towards a particular direction, specified by the parameter b.

Uhlenbeck and Ornstein (1930) introduced a stochastic process to model the velocity of a particle under friction. In one dimension, the Ornstein-Uhlenbeck process $(X_t)_{t\geq 0}$ is defined as the solution to the stochastic differential equation

$$dX_t = \beta(\mu - X_t)dt + \sigma dW_t,$$

where W_t is a standard Wiener process, and with initial condition $X_0 = x_0$. Here, the drift term is a linear function of the current value X_t of the process, and the diffusion term is a constant. The process has three parameters: $\mu \in \mathbb{R}$ is the long-term mean, $\beta > 0$ is the strength of the reversion to the mean, and $\sigma > 0$ is the spread around the mean. The Ornstein-Uhlenbeck process is called *mean-reverting*, as it will tend to be attracted towards its long-term mean μ, at a rate measured by β. Indeed, it has a closed-form Gaussian transition density with mean

$$E(X_{t+\delta}|X_t = x_t) = e^{-\beta\delta}x_t + (1 - e^{-\beta\delta})\mu.$$

That is, the mean moves from x_t (at time t) to μ (as $\delta \to \infty$). The Brownian motion is a limiting case of the Ornstein-Uhlenbeck process, when $\beta \to 0$, i.e. when the reversion to the mean tends to zero.

2.4.3 Simulation and inference

A consequence of the Markovian property of diffusion processes is that the transition density can directly be used for simulation or for likelihood-based inference.

Simulation We denote by $\{t_1, \ldots, t_n\}$ the times at which the process should be simulated, and x_i the simulated value at time t_i. We initialise $X_1 = x_1$ and, for $i = 1, \ldots, n-1$, we sample x_{i+1} from the transition density $p(X_{t_{i+1}}|X_{t_i} = x_i)$.

Inference Say that we observed the diffusion process at discrete times $\{t_1, \ldots, t_n\}$, and we denote by $\{x_1, \ldots, x_n\}$ the observations. The transition density of the process, $p(X_{t_{i+1}} = x_{i+1} | X_{t_i} = x_i)$, defines the likelihood of the transition from x_i to x_{i+1}. The process is Markovian, and so the transitions are independent. Therefore, the likelihood of all steps is given by the product of the likelihoods of the individual transitions. The first observation is typically assumed to be deterministic but, more generally, we may denote by $p(x_1)$ the likelihood of the first observation x_1. Then, the likelihood of the n observations is

$$p(x_1, \ldots, x_n) = p(x_1) \prod_{i=1}^{n-1} p(X_{t_{i+1}} = x_{i+1} | X_{t_i} = x_i).$$

The transition density is often a parametric function, and likelihood-based methods of inference can then be used to learn about the parameters. For example, this approach could be used to estimate the variance σ^2 and the drift b of the Brownian motion with drift.

In many cases, the drift and diffusion functions are complex expressions, and the transition density may not be analytically tractable. In such cases, it is generally not possible to implement exact simulation and inference. Numerical methods can be employed to approximate the diffusion process by a simpler process, for which the transition density is known (Iacus, 2009). We present one such method in the context of the Langevin diffusion process, in Chapter 5.

Chapter 3

Markov chain Monte Carlo step selection models

In this chapter, we describe a new family of models of animal movement and habitat selection, based on the properties of Markov chain Monte Carlo (MCMC) algorithms. Section 3.1 presents the existing habitat selection models, and motivates the development of a new approach. In Sections 3.2 and 3.3, we discuss the use of MCMC as a basis for models of step selection. We describe a new MCMC sampler to be used as a model of animal movement in Section 3.4, and we present several special cases relevant to step selection in Section 3.5. We investigate the properties of those models with simulations in Section 3.7. Some of the material covered in this chapter was introduced in Michelot et al. (2019), and in the preprint Michelot et al. (2018a).

3.1 The discrepancy between resource selection and step selection

Habitat selection models, described in Section 1.3, combine telemetry data and environmental data to estimate habitat preferences, and to predict long-term patterns of space use by animals. They can roughly be classified into resource selection models, which treat telemetry data as a collection of independent points, and step selection models, which account for the serial correlation due to the movement of the animal.

Note that, unlike the resource selection function (RSF), the step selection function (SSF) is not used as a model for the utilisation distribution of the animal. Indeed, although they

usually take the same functional (exponential) form, they capture habitat selection at two different scales: the global scale of the study region for the RSF, and the local scale of the step for the SSF. In the two approaches, the habitat selection parameters $\boldsymbol{\beta}$ therefore measure a different type of selection. In particular, a fitted SSF cannot readily be used to predict long-term space use. Indeed, Barnett and Moorcroft (2008) showed that the stationary distribution of a SSF model can be written

$$\pi(\boldsymbol{x}) = \frac{w(c(\boldsymbol{x})) \int w(c(\boldsymbol{y}))\phi(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y}}{\int w(c(\boldsymbol{y})) \int w(c(\boldsymbol{z}))\phi(\boldsymbol{z}|\boldsymbol{y})d\boldsymbol{z}d\boldsymbol{y}}, \tag{3.1}$$

using the notation introduced in Section 1.3. This shows that, in general, the stationary distribution π of a SSF model is not proportional to the SSF w. The relationship crucially depends on the choice of the movement model, i.e. the choice of the movement kernel ϕ. The function given in Equation 3.1 does not reduce to the simple parametric expression used for the utilisation distribution in RSF models.

Equation 3.1 is valid when the SSF only depends on the habitat at the endpoint of the step, and not over the whole step (Equation 1.3). If the SSF depends on the whole step, the stationary distribution of a SSF model, i.e. the underlying utilisation distribution, can be approximated using numerical methods. Potts et al. (2014a) described an iterative procedure to approximate the long-term utilisation distribution of a SSF model on a discrete spatial grid. They wrote a "master equation" to describe the distribution of the animal's location in space at time $t + 1$, in terms of its distribution at time t. Starting from some initial distribution, and iterating over t, the equation eventually converges to a stationary distribution (analogous to the utilisation distribution). Similarly, Avgar et al. (2016) and Signer et al. (2017) argued that simulations from a fitted SSF model can be used to estimate its long-term distribution. A SSF model is fitted, and many locations are generated from the estimated movement kernel and step selection function. For long simulation runs, the distribution of sampled points approximates the stationary distribution of the model. However, the results of simulation-based methods will vary depending on the choice of the geographical region over which the simulations are run.

From a biological perspective, the relationship between the SSF and the utilisation distribution of the animal, given in Equation 3.1, can be seen as the result of local habitat selection. At the time scale of the observed movement steps, we consider that the animal has access to a neighbourhood of its current location, defined by the movement kernel ϕ of the model. Within this available region, it selects a location based on its habitat suitability, determined by the SSF w. Because the region of availability typically does not cover the whole study

region, the animal tends to favour habitat units which are both suitable and close. This causes the disproportion shown in Equation 3.1 between the SSF and the utilisation distribution. A special case was described by Moorcroft and Barnett (2008), who showed that, under some assumptions, the utilisation distribution of the animal is proportional to the *squared* SSF. A possible interpretation is that, if an animal is in a suitable habitat patch, it will tend to remain there longer than would be suggested by local habitat selection, merely because it is close and therefore easily accessible. Step selection models can also incorporate the effect of the habitat directly on the movement of the animal (e.g. different movement speeds in different habitat types). Potts et al. (2014a) describe how the long-term distribution of space use of the animal arises from models of movement and local habitat selection.

There is a strong ecological incentive to bring the RSF and SSF frameworks together, and to model individual animal movement and long-term habitat selection jointly. In particular, this would allow for the integration of telemetry data and survey data. Telemetry data are typically analysed with models of individual movement that do not capture the equilibrium distribution, such as SSF models. Survey data are independent observations from a population, and they are analysed with species distribution models that often do not capture the movement of individuals. This motivates the development of a step selection model with an explicit (and tractable) stationary distribution. In this chapter, we introduce a new framework to jointly model animal movement, space use, and habitat selection, to overcome this limitation of existing approaches.

3.2 The analogy between MCMC and animal movement

It is natural to think of the utilisation distribution of an animal as the long-term consequence of its short-term movement decisions. However, most movement models do not have an explicit stationary distribution, and fail to capture space use by the animal (although see Ovaskainen, 2008). Following the notation introduced in Section 1.2, we define $(\boldsymbol{X}_t)_{t\in\mathbb{N}}$ the two-dimensional location process of an animal, which takes values in $\Omega \subseteq \mathbb{R}^2$.

We propose an analogy between the movements of an animal in geographical space and the movements of a Markov chain Monte Carlo (MCMC) sampler in a two-dimensional parameter space. As described in Section 2.3, an MCMC sampler is defined by its transition kernel, constructed such that the long-term distribution of samples is the target distribution. This is similar to an animal moving in the short term according to a movement model, such that the long-term distribution of its location is the utilisation distribution.

Following this analogy, we can view an MCMC algorithm as a movement model, that we will call an "MCMC movement model". The transition kernel $p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t)$ of the algorithm determines how the next sample $\boldsymbol{X}_{t+1}$ is generated conditional on $\boldsymbol{X}_t$. It must satisfy certain properties (e.g. the detailed balanced condition, see Section 2.3) to ensure that the samples $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots\}$ converge to the given target distribution. In an MCMC movement model, the transition kernel models the step density of the location process $(\boldsymbol{X}_t)$ of the animal, i.e. the density of its next location, conditional on its current location. Due to the properties of the underlying MCMC algorithm, the movement process is stationary, and its equilibrium distribution is guaranteed to coincide with a given target distribution. This target distribution is the long-term distribution of the animal's location in space, i.e. its utilisation distribution. This approach thus describes a model which links the short-term movement rules (or movement decisions) of the animal to the long-term distribution in space. We summarise the correspondence made between the terminology of MCMC algorithms and that of movement models in Table 3.1.

MCMC	Animal movement	Notation
Parameter space	Geographical space	$\Omega \subseteq \mathbb{R}^2$
Iteration	Time step	$t = 1, 2, \dots$
Sampled point	Location	$\boldsymbol{X}_t \in \Omega$
Transition kernel	Step density	$p(\boldsymbol{X}_{t+1}\|\boldsymbol{X}_t)$
Target distribution	Utilisation distribution	$\pi(\boldsymbol{x})$

Table 3.1 Summary of the analogy between MCMC and animal movement.

Over the past 65 years, many different MCMC algorithms have been developed for the purpose of sampling from probability distributions (Metropolis-Hastings algorithms, Gibbs samplers, Hamiltonian Monte Carlo...) In the analogy presented here, each MCMC algorithm describes a different movement model, with specific transition rules. However, not all algorithms are good representations of animal movement. This can most clearly be illustrated by the example of the Gibbs sampler, which moves in the parameter space one component at a time (Gelfand and Smith, 1990). In the context of animal movement, this corresponds to the implausible scenario of an animal alternating between displacements along the longitude and the latitude axes. For example, Figure 3.1(A) shows the first 100 iterations of a Gibbs sampler on a bivariate normal target distribution π. The systematic perpendicular turns seen in the output of the Gibbs simulation do not resemble any feature of real animal movement. For this reason, the Gibbs sampler would generally be unsuitable to construct a model of animal movement.

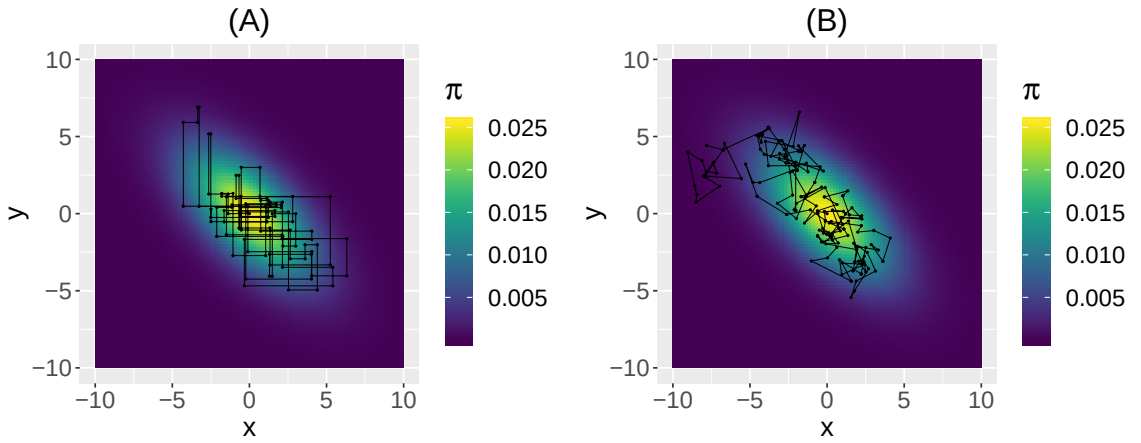


Fig. 3.1 Example runs of a Gibbs sampler (A) and Metropolis sampler (B). The background colour shows the (bivariate normal) target distribution π.

Another widely-used MCMC algorithm is the Metropolis-Hastings sampler, that we described in Section 2.3 (Metropolis et al., 1953; Hastings, 1970). Figure 3.1(B) shows the first 100 iterations of a Metropolis-Hastings sampler with normal proposal distribution, on the same target distribution π. The movement patterns of this algorithm resemble those of a random walk, and could appear to be a plausible (albeit simplistic) model of animal movement. However, the generated path includes rejections, when the sampler did not accept the proposed update, but these cannot be seen in the plot (because identical successive samples overlap). Because it is based on an acceptance-rejection step, we will argue that the Metropolis-Hastings algorithm is generally not a good candidate to construct a model of animal movement. Indeed, although it can happen than an animal stays at the same location over several time steps, we do not believe that this phenomenon is analogous to the systematic acceptance-rejection step that underlies Metropolis-Hastings (and other rejection-based) algorithms. Moreover, many telemetry data sets do not include any steps of length zero, and would therefore be a very implausible output for a rejection-based MCMC movement model. We discuss this point in more detail in Section 4.9. In Section 3.4, we develop a *rejection-free* MCMC algorithm, as a more realistic basis for a model of animal movement.

Most MCMC samplers are defined as discrete-time Markov processes. Indeed, the transition kernel $p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t)$ is defined over a (fixed) discrete interval, corresponding to one iteration of the algorithm. This framework therefore describes animal movement in discrete time. Discrete-time models of animal movement are widely used, but they have several conceptual limitations. In particular, they cannot accommodate irregular time intervals in observed tracking data, and their inferences cannot be compared across different temporal scales. In

Section 3.6.3, we propose an approach to relax this constraint for one particular model within the class described here. We discuss this problem in more detail in Chapter 5, where we introduce a continuous-time formulation that circumvents the limitations of discrete-time approaches.

3.3 MCMC as a model of step selection

In the framework introduced in Section 3.2, we consider a movement model based on the MCMC algorithm with transition kernel $p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t)$ and target distribution π. It can be extended to describe habitat selection, in addition to movement. This can be done by formulating the target distribution of the algorithm as a normalized resource selection function, i.e.

$$\forall \boldsymbol{x} \in \Omega, \quad \pi(\boldsymbol{x}) \propto \exp(\boldsymbol{\beta}' \boldsymbol{c}(\boldsymbol{x})),$$

as described in Equation 1.2, where $\boldsymbol{\beta}'$ denote the transpose of $\boldsymbol{\beta}$. The target distribution models the utilisation distribution of the animal, and so the coefficients $\boldsymbol{\beta}$ here capture the global long-term habitat selection (similarly to the coefficients of a standard RSF). The covariates $\boldsymbol{c}$ can encompass a wide range of spatial variables. They can be categorical variables, if the study region is divided in patches of different habitats (e.g. different vegetation types). They can also be continuous variables, such as the distributions of resources of interest (e.g. food items), or other continuous features of the landscape (e.g. elevation, slope). An interesting special case of a continuous covariate is the distance to habitat features, such as roads or rivers, to capture effects of attraction or repulsion. The distance to the centre of the animal's territory can also be included as a covariate, to model its tendency to remain within a home range.

This model relies on the assumption that there exists a long-term distribution of space use by the animal (the utilisation distribution), and that it can be written as a function of the distributions of some resources, or other spatial covariates. This assumption is also made in standard RSF models, and in other approaches such as resource utilisation functions (described in Section 1.3). Based on this assumption, we consider that the short-term movement rules of an animal are affected by this long-term distribution. The intuition behind this modelling choice is that the utilisation distribution represents a measure of habitat suitability, and that the movement decisions of an animal are determined by the habitat suitability within its perception range. Unlike standard SSF approaches, the selection is here independent of the time scale, i.e. independent of the scale of the movement model.

Regardless of the time scale, the animal's displacements are always determined by the same long-term distribution of space use.

Note that this is not a special case of the standard SSF model described e.g. by Forester et al. (2009), and formulated in Equation 1.3. Indeed, the likelihood of a step under an MCMC algorithm is given by its transition kernel, which is generally not separable into a product of a step selection function and a habitat-independent movement kernel (Equation 1.3). We will call the model presented here an "MCMC step selection model" because, although it is not formulated as a SSF, it combines movement and habitat selection.

The MCMC step selection model combines advantages from the standard RSF and SSF frameworks. In RSF analyses, the observed locations are assumed to be independent samples from a spatial point process, and the aim of the study is to model its steady-state distribution. On the other hand, SSFs treat the locations as the output of a (possibly second-order) Markov process, and the focus is on its transition density. We combine the two approaches: in an MCMC step selection model, the transition kernel describes the short-term movement rules of the animal, and the target distribution captures long-term habitat selection and space use.

This model is formulated in terms of two sets of parameters: the tuning parameters of the underlying MCMC algorithm, and the parameters $\boldsymbol{\beta}$ of the target distribution. The tuning parameters of the algorithm are the parameters of its transition kernel $p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t)$ (e.g. proposal variance of a Metropolis sampler). In an MCMC step selection model, they become movement parameters, as the transition kernel describes the step density of the animal. The parameters $\boldsymbol{\beta}$ of the target distribution are habitat selection parameters.

This modelling framework can be used to analyse animal movement and habitat data, in the following steps.

1. Choose an MCMC algorithm, to be used as a model of animal movement and habitat selection.
2. Write the likelihood of the model. Under an MCMC step selection model, the likelihood of an observed displacement from a location $\boldsymbol{x}_t$ to a location $\boldsymbol{x}_{t+1}$ is given by the transition kernel $p(\boldsymbol{X}_{t+1} = \boldsymbol{x}_{t+1}|\boldsymbol{X}_t = \boldsymbol{x}_t)$. It is a function of the habitat selection parameters $\boldsymbol{\beta}$, and of the other parameters of the sampler (i.e. the movement parameters of the transition kernel).
3. Use maximum likelihood estimation, or other likelihood-based methods, to estimate all model parameters.

With this method, we can jointly estimate movement and habitat selection characteristics. In this chapter, we focus on step 1, and introduce an MCMC algorithm which can be used as the basis of a flexible family of step selection models. We investigate steps 2 and 3, i.e. the possibility to carry out inference in this framework, in Chapter 4.

3.4 The local Gibbs algorithm

As explained in Section 3.2, rejection-based MCMC algorithms are generally not a good choice to construct a movement model. In this section, we introduce the local Gibbs algorithm, a new rejection-free MCMC sampler that better resembles animal movement.

3.4.1 Algorithm

We consider the target distribution $\pi : \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^d$. We focus on the case $d = 2$, because animal space use is most often studied in two-dimensional geographical space, but the algorithm can easily be extended to $d \neq 2$. In the following, we consider that Ω is bounded in $\mathbb{R}^2$, to ensure that π is integrable. This assumption is not strictly necessary, because π would be integrable over an unbounded region if it decreased to zero rapidly enough. We describe the case where $\Omega = \mathbb{R}^2$ in the discussion of Chapter 4.

The local Gibbs algorithm for the target distribution π is defined as follows. We choose $\phi : \Omega \to \mathbb{R}$ the density function of a symmetric distribution, i.e. such that $\forall \boldsymbol{x}, \boldsymbol{y} \in \Omega, \phi(\boldsymbol{y}|\boldsymbol{x}) = \phi(\boldsymbol{x}|\boldsymbol{y})$. We start from $\boldsymbol{X}_1 \in \Omega$; then, for $t = 1, 2, \dots,$

1. Sample a point $\boldsymbol{\mu}$ from $\phi(\cdot|\boldsymbol{X}_t)$.
2. For all $\boldsymbol{x} \in \Omega$, define
$$\tilde{\pi}(\boldsymbol{x}|\boldsymbol{\mu}) = \frac{\pi(\boldsymbol{x})\phi(\boldsymbol{x}|\boldsymbol{\mu})}{\int_{\boldsymbol{z}\in\Omega} \pi(\boldsymbol{z})\phi(\boldsymbol{z}|\boldsymbol{\mu})d\boldsymbol{z}}.$$
3. Sample $\boldsymbol{X}_{t+1}$ from $\tilde{\pi}(\cdot|\boldsymbol{\mu})$.

We can show that samples $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots\}$ generated with the local Gibbs algorithm converge to the target distribution π. We first demonstrate that π is a stationary distribution for this algorithm, by showing that the local Gibbs sampler satisfies the detailed balance condition (Section 2.3). Indeed, for any $\boldsymbol{x}, \boldsymbol{y} \in \Omega$, we have

$$\pi(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x}) = \pi(\boldsymbol{x}) \int_{\boldsymbol{\mu}\in\Omega} p(\boldsymbol{y}|\boldsymbol{\mu})p(\boldsymbol{\mu}|\boldsymbol{x})d\boldsymbol{\mu},$$

where $p(\boldsymbol{y}|\boldsymbol{x}) = p(\boldsymbol{X}_{t+1} = \boldsymbol{y}|\boldsymbol{X}_t = \boldsymbol{x})$. The intermediate point $\boldsymbol{\mu}$ is sampled from $\phi(\cdot|\boldsymbol{x})$, so $p(\boldsymbol{\mu}|\boldsymbol{x}) = \phi(\boldsymbol{\mu}|\boldsymbol{x})$. The endpoint $\boldsymbol{y}$ is sampled from $\tilde{\pi}(\cdot|\boldsymbol{\mu})$, so $p(\boldsymbol{y}|\boldsymbol{\mu}) = \tilde{\pi}(\boldsymbol{y}|\boldsymbol{\mu})$. We can make these substitutions and rewrite

$$\begin{aligned}\pi(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x}) &= \pi(\boldsymbol{x}) \int_{\boldsymbol{\mu}\in\Omega} \frac{\pi(\boldsymbol{y})\phi(\boldsymbol{y}|\boldsymbol{\mu})}{\int_{\boldsymbol{z}\in\Omega} \pi(\boldsymbol{z})\phi(\boldsymbol{z}|\boldsymbol{\mu})d\boldsymbol{z}} \phi(\boldsymbol{\mu}|\boldsymbol{x})d\boldsymbol{\mu} \\ &= \pi(\boldsymbol{x})\pi(\boldsymbol{y}) \int_{\boldsymbol{\mu}\in\Omega} \frac{\phi(\boldsymbol{y}|\boldsymbol{\mu})\phi(\boldsymbol{\mu}|\boldsymbol{x})}{\int_{\boldsymbol{z}\in\Omega} \pi(\boldsymbol{z})\phi(\boldsymbol{z}|\boldsymbol{\mu})d\boldsymbol{z}} d\boldsymbol{\mu}.\end{aligned}$$

By symmetry of ϕ, we have $\phi(\boldsymbol{y}|\boldsymbol{\mu}) = \phi(\boldsymbol{\mu}|\boldsymbol{y})$ and $\phi(\boldsymbol{\mu}|\boldsymbol{x}) = \phi(\boldsymbol{x}|\boldsymbol{\mu})$, and so

$$\begin{aligned}\pi(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x}) &= \pi(\boldsymbol{y})\pi(\boldsymbol{x}) \int_{\boldsymbol{\mu}\in\Omega} \frac{\phi(\boldsymbol{\mu}|\boldsymbol{y})\phi(\boldsymbol{x}|\boldsymbol{\mu})}{\int_{\boldsymbol{z}\in\Omega} \pi(\boldsymbol{z})\phi(\boldsymbol{z}|\boldsymbol{\mu})d\boldsymbol{z}} d\boldsymbol{\mu} \\ &= \pi(\boldsymbol{y})p(\boldsymbol{x}|\boldsymbol{y}),\end{aligned}$$

as required. The local Gibbs algorithm therefore satisfies the detailed balance condition, for the stationary distribution π.

To prove convergence to the stationary distribution, we must also verify that the Markov chain defined by the algorithm is π-irreducible, and aperiodic (Tierney, 1994). In the following, we make the additional assumption that the symmetric density $\phi(\cdot|\boldsymbol{x})$ is positive over a neighbourhood of $\boldsymbol{x}$, to avoid pathological cases. A Markov chain (with uncountable state space) is ψ-irreducible for a measure ψ if any subset $\mathcal{A} \subset \Omega$ of the state space over which $\psi > 0$ can be reached from any state $\boldsymbol{x} \in \Omega$ with positive probability. The transition kernel of the local Gibbs algorithm is π-irreducible over Ω, if ϕ is positive over a neighbourhood of its origin. Indeed, any region where $\pi > 0$ has a positive probability to be reached from any $\boldsymbol{x} \in \Omega$. A Markov chain is aperiodic if the number of steps to reach $\mathcal{A}$ from $\boldsymbol{x}$ is aperiodic. A sufficient condition for the chain to be aperiodic is that the transition kernel is positive over a neighbourhood of $\boldsymbol{x}$, because it can then remain in that neighbourhood arbitrarily long before moving to $\mathcal{A}$. This is clearly true for the local Gibbs sampler, if ϕ is positive over a neighbourhood of its origin. Finally, this proves that π is the only stationary distribution for the local Gibbs algorithm, and that π is also the limiting distribution of the chain (i.e. the algorithm will converge to π).

We call this algorithm "local Gibbs" to draw a parallel with the classic Gibbs sampler. In the classic Gibbs sampler, each parameter component $X^{(i)}$ is updated in turn, conditionally on the other components $\{X^{(1)}, \ldots, X^{(i-1)}, X^{(i+1)}, \ldots, X^{(n)}\}$, which are kept fixed. It is used when the conditional distribution $p(X^{(i)}|X^{(1)}, \ldots, X^{(i-1)}, X^{(i+1)}, \ldots, X^{(n)})$ is analytically

tractable. At each step, the Gibbs algorithm then samples from a one-dimensional slice of the target distribution (Gilks et al., 1996). The local Gibbs algorithm is not a special case of the classic Gibbs sampler, but it is based on a similar idea. That is, at each iteration, it samples from a restricted part of the target distribution. That restricted part, $\tilde{\pi}$, is defined by the choice of the density ϕ. An important feature shared by the Gibbs sampler and the local Gibbs sampler is that they are rejection-free. The process moves at every iteration, and an acceptance-rejection step is not needed to conserve the stationary distribution.

3.4.2 Local Gibbs movement model

Following the analogy introduced in Section 3.2, the local Gibbs algorithm can be used as the basis of a model of animal movement and habitat selection. The transition kernel $p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t)$ defines the movement component, and the target distribution π is modelled with a RSF (Equation 1.2). We will refer to this model as the "local Gibbs movement model", or "local Gibbs step selection model", or simply "local Gibbs model".

The intermediate point $\boldsymbol{\mu}$, sampled in Step 1 of the local Gibbs algorithm, does not have a direct biological interpretation. It is a sample from the density ϕ centred on the current location of the animal, and it is an intermediate point used in the construction of the transition density of the local Gibbs algorithm. We can integrate over $\boldsymbol{\mu}$ to obtain the step density of the local Gibbs model, i.e. the likelihood of a displacement from a point $\boldsymbol{x} \in \Omega$ to a point $\boldsymbol{y} \in \Omega$ under the model,

$$p(\boldsymbol{y}|\boldsymbol{x}) = \int_{\boldsymbol{\mu}\in\Omega} p(\boldsymbol{y}|\boldsymbol{\mu})p(\boldsymbol{\mu}|\boldsymbol{x})d\boldsymbol{\mu}$$

where $p(\boldsymbol{y}|\boldsymbol{\mu}) = \tilde{\pi}(\boldsymbol{y}|\boldsymbol{\mu})$ and $p(\boldsymbol{\mu}|\boldsymbol{x}) = \phi(\boldsymbol{\mu}|\boldsymbol{x})$. The step density becomes

$$p(\boldsymbol{y}|\boldsymbol{x}) = \pi(\boldsymbol{y}) \int_{\boldsymbol{\mu}\in\Omega} \frac{\phi(\boldsymbol{y}|\boldsymbol{\mu})\phi(\boldsymbol{\mu}|\boldsymbol{x})}{\int_{\boldsymbol{z}\in\Omega} \pi(\boldsymbol{z})\phi(\boldsymbol{z}|\boldsymbol{\mu})d\boldsymbol{z}} d\boldsymbol{\mu}. \tag{3.2}$$

The local Gibbs model is not a special case of the SSF models presented e.g. by Fortin et al. (2005) and Forester et al. (2009). In particular, the step density is not separable into the product of a habitat-independent movement component and a step selection function (as written in Equation 1.3 in the context of standard SSF models). However, we can still compute a habitat-independent step density, by fixing the distribution π to a constant in Equation 3.2. In the absence of covariate effects, i.e. if the utilisation distribution is flat, each step is the sum of two ϕ-distributed increments. For this reason, we call ϕ the half-step

density of the model. The habitat-independent step density is given by the convolution

$$p_0(\boldsymbol{y}|\boldsymbol{x}) = \int_{\boldsymbol{\mu}\in\Omega} \phi(\boldsymbol{y}|\boldsymbol{\mu})\phi(\boldsymbol{\mu}|\boldsymbol{x})d\boldsymbol{\mu}.$$

In the following, we will denote $p_0(\boldsymbol{y}|\boldsymbol{x})$ the *habitat-independent* step density of a model, to distinguish from the step density $p(\boldsymbol{y}|\boldsymbol{x})$, which generally depends on the habitat composition around $\boldsymbol{x}$.

The local Gibbs model is defined in terms of the symmetric distribution ϕ, that we call the half-step density. Note from the first step of the local Gibbs algorithm that we must be able to generate samples from ϕ. This is not a very strong constraint, and there are many candidate distributions. Each choice of ϕ leads to a different movement model, but it does not affect the equilibrium properties of the model. Regardless of ϕ, the long-term distribution of the movement process is the utilisation distribution π. We propose several special cases of ϕ in Section 3.5, corresponding to different movement models.

The spatial scale of ϕ is linked to the scale of perception of the animal, as we assume that the animal can perceive habitat features over $\phi(\cdot|\boldsymbol{\mu})$. It is also linked to the scale of its mobility, as it defines the size of the region that is accessible to the animal over one time interval. As the spatial scale of ϕ increases, the animal perceives a larger region, and may take longer steps. The animal's perception and mobility are therefore modelled jointly in this framework. This is a strong assumption, and it ignores characteristics such as memory. It is nevertheless a common limitation of almost all SSF models, in which the habitat-independent movement kernel also determines the area over which the animal can perceive its environment (Fortin et al., 2005; Forester et al., 2009).

In Section 3.3, we explained that, in the context of the MCMC movement model, the tuning parameters of the underlying MCMC algorithm can be viewed as movement parameters. This is in particular the case for the local Gibbs model, in which the "tuning" parameters are the parameters of the half-step density ϕ. In fact, the tuning parameters do not only describe characteristics of the movement, but also of the animal's perception, because the two are so closely linked in this framework. For many choices of ϕ, including those considered in the rest of this chapter and the next, the half-step density is formulated in terms of a "scale" parameter. Here, we use this term loosely to refer to a parameter that quantifies the spatial extent (i.e. the spatial scale) of the half-step density ϕ. It is convenient to qualify a parameter that describes both the scale of perception and the scale of movement.

3.4.3 Simulation from the local Gibbs algorithm

The local Gibbs algorithm given in Section 3.4.1 describes the steps to simulate from the local Gibbs movement model, provided a half-step density ϕ and a target distribution π. However, in step 2 of the algorithm, the integral $\int_{\boldsymbol{z}\in\Omega} \pi(\boldsymbol{z})\phi(\boldsymbol{z}|\boldsymbol{\mu})d\boldsymbol{z}$ cannot generally be derived analytically. Indeed, the function π depends on the distribution of spatial covariates, which cannot be expressed in a simple functional form. In practice, simulation from the local Gibbs sampler can be achieved by approximating that integral, using numerical methods. We propose a Monte Carlo method, based on the evaluation of the integrand at a large number of random points.

The simulation algorithm is described by the following steps. Start from a point $\boldsymbol{x}_1 \in \Omega$; then, at each iteration $t = 1, 2, \ldots,$

1. Sample a point $\boldsymbol{\mu}$ from $\phi(\cdot|\boldsymbol{x}_t)$.
2. Sample a large number of points $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K\}$ from $\phi(\cdot|\boldsymbol{\mu})$.
3. For $k \in \{1, \ldots, K\}$, define
$$p_k = \frac{\pi(\boldsymbol{z}_k)}{\sum_{l=1}^{K} \pi(\boldsymbol{z}_l)}.$$
4. Sample $\boldsymbol{x}_{t+1}$ from the $\boldsymbol{z}_k$, with probabilities given by the p_k.

Although this algorithm is not exact, because we sample from a numerical approximation of the transition kernel, it can be made arbitrarily accurate by increasing the value of K. In Section 3.7.2, we use simulations to investigate the effect of K on the convergence of the algorithm to the target distribution.

Note that the local Gibbs sampler would usually not be a good choice for the general purpose of simulation from a probability distribution (e.g. to carry out Bayesian inference). Indeed, although it does not require any rejections, the numerical integration requires K evaluation of the target distribution at each iteration. For this reason, it will almost always be more computationally expensive than, say, Metropolis-Hastings sampling, for which the optimal acceptance rate is around 23% (Roberts et al., 1997). In this work, we only consider the local Gibbs algorithm as the basis for a model of animal movement.

3.5 Special cases of the local Gibbs model

The local Gibbs model is very flexible, because the half-step density ϕ can be chosen from a wide range of probability distributions. We only made three assumptions on ϕ: it is

symmetric, it is positive over a neighbourhood of its origin, and it is possible to generate samples from it. In this chapter and the next, we turn our attention to a few special cases of the local Gibbs model, that are particularly relevant to the study of animal movement and habitat selection. In particular, we describe the case where ϕ is a normal distribution, and the case where ϕ is a uniform distribution over a disc. The former is a natural choice to model animal movement, reminiscent of a standard Gaussian random walk, and the latter is inspired by existing SSF models.

3.5.1 Normal kernel model

We first consider the case where the half-step density ϕ is chosen to be a bivariate normal density, centred on the origin, and we call the corresponding local Gibbs model the "normal kernel model". Let $\varphi(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be the bivariate normal pdf with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. This half-step density always satisfies the assumption of symmetry made in the local Gibbs algorithm, i.e. we have $\varphi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\Sigma}) = \varphi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\Sigma})$ for any symmetric covariance matrix $\boldsymbol{\Sigma}$. In the context of animal movement, however, a non-diagonal matrix would suggest a systematic drift in the movement, along an axis determined by the covariance structure. For example, a positive correlation would correspond to a tendency to move in either direction along the "South-West to North-East" axis. Likewise, if the horizontal and vertical variances were different, it would indicate different tendencies to move along the Northing axis and the Easting axis. These particular cases may be useful in some specific applications (perhaps when the movement of the animal is constrained to follow a natural corridor). However, in most studies, it is reasonable to assume that the density is radially symmetric, i.e. isotropic. We focus on this case, and we denote the covariance matrix by $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_2$, where $\boldsymbol{I}_2$ is the 2×2 identity matrix. The parameter σ^2 determines the spatial scale of the transition kernel of the algorithm, and thus of the step density of the corresponding movement model. Larger values of σ^2 lead to faster exploration of the target distribution, i.e. higher speed of movement for the animal.

Following from Equation 3.2, the step density from $\boldsymbol{x}$ to $\boldsymbol{y}$ is

$$p(\boldsymbol{y}|\boldsymbol{x}) = \pi(\boldsymbol{y}) \int_{\boldsymbol{\mu} \in \Omega} \frac{\varphi(\boldsymbol{y}|\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_2) \varphi(\boldsymbol{\mu}|\boldsymbol{x}, \sigma^2 \boldsymbol{I}_2)}{\int_{\boldsymbol{z} \in \Omega} \pi(\boldsymbol{z}) \varphi(\boldsymbol{z}|\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_2) d\boldsymbol{z}} d\boldsymbol{\mu}.$$

In the absence of covariate effects, each displacement from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t+1}$ is the sum of two normally-distributed increments. It can be shown that the sum of two normally-distributed random variables also follows a normal distribution. Indeed, in this case, the habitat-independent step density of the normal kernel model is a normal density, centred on the

current location of the animal, with variance $2\sigma^2 \boldsymbol{I}_2$. (A proof of this result is given in Appendix A.) This is the transition density of an isotropic Gaussian random walk.

In two dimensions and in the absence of covariate effects, the length of each displacement (the "step length") under the normal kernel model follows a Rayleigh distribution. Indeed, the Rayleigh distribution arises from the length of a two-dimensional vector with uncorrelated normal components of equal variances. Its density function is

$$p(l|\alpha) = \frac{l}{\alpha^2} \exp\left(-\frac{l^2}{2\alpha^2}\right),$$

where $\alpha > 0$ is its scale parameter, and $l \geq 0$ is the step length. The scale parameter is linked to the variance of the underlying normal distribution. In the local Gibbs model with normal half-step density, the scale of the Rayleigh (step length) distribution is $\alpha = \sqrt{2}\sigma$, with σ as defined above. Figure 3.2 shows the density function of the Rayleigh distribution for different values of the scale parameter α.

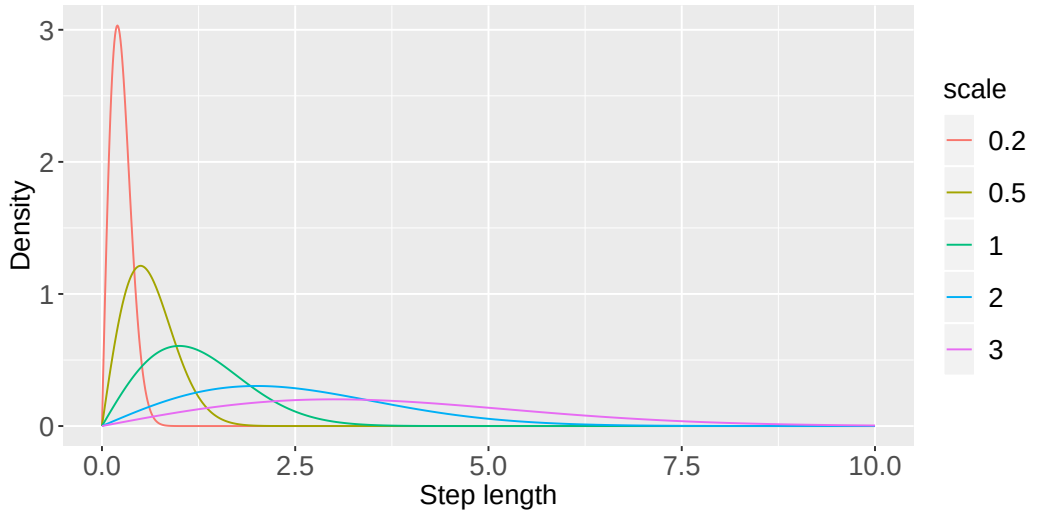


Fig. 3.2 Probability density functions of the Rayleigh distribution, with different scale parameters.

3.5.2 Availability radius model

We consider the case where the half-step density ϕ is uniform on a disc of radius r, centred on the origin. At each iteration, the intermediate point $\boldsymbol{\mu}$ is sampled from a disc of radius r centred on $\boldsymbol{X}_t$. Then, $\boldsymbol{X}_{t+1}$ is sampled from the target distribution π, truncated to the disc of radius r centred on $\boldsymbol{\mu}$. The notation is illustrated in Figure 3.3. In this formulation, the

half-step density only has one parameter: the radius r. Similarly to the variance of the normal kernel model, the radius parameter is a measure of the scale of movement and perception.

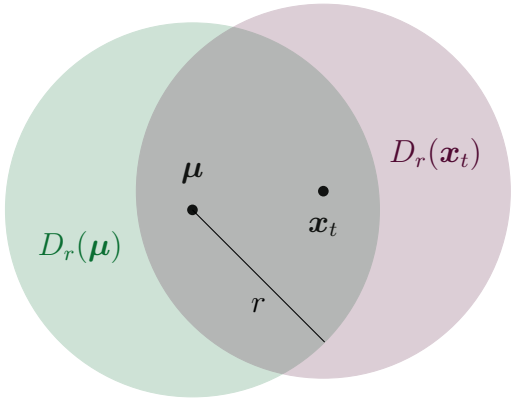


Fig. 3.3 Notation for the local Gibbs sampler with uniform half-step density on a disc (availability radius model). The next point $\boldsymbol{x}_{t+1}$ is sampled from the target distribution π truncated to $D_r(\boldsymbol{\mu})$.

In this formulation, the density ϕ is

$$\phi(\boldsymbol{y}|\boldsymbol{x}) = \frac{I_{D_r(\boldsymbol{x})}(\boldsymbol{y})}{\pi r^2},$$

where $D_r(\boldsymbol{x})$ is the disc of radius $r > 0$ and centre $\boldsymbol{x} \in \Omega$, and $I_A(\boldsymbol{x})$ is the indicator function of $A \subset \Omega$, i.e.

$$I_A(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{x} \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Plugging this expression of ϕ into Equation 3.2, the likelihood of a step from $\boldsymbol{x}$ to $\boldsymbol{y}$ is

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{x}) &= \pi(\boldsymbol{y}) \int_{\boldsymbol{\mu}\in\Omega} \frac{\pi r^2}{(\pi r^2)^2} \frac{I_{D_r(\boldsymbol{\mu})}(\boldsymbol{y}) I_{D_r(\boldsymbol{x})}(\boldsymbol{\mu})}{\int_{\boldsymbol{z}\in\Omega} \pi(\boldsymbol{z}) I_{D_r(\boldsymbol{\mu})}(\boldsymbol{z}) d\boldsymbol{z}} d\boldsymbol{\mu} \\ &= \frac{\pi(\boldsymbol{y})}{\pi r^2} \int_{\boldsymbol{\mu}\in D_r(\boldsymbol{x})\cap D_r(\boldsymbol{y})} \frac{1}{\int_{\boldsymbol{z}\in D_r(\boldsymbol{\mu})} \pi(\boldsymbol{z}) d\boldsymbol{z}} d\boldsymbol{\mu}. \end{aligned} \tag{3.3}$$

If the two points $\boldsymbol{x}$ and $\boldsymbol{y}$ are separated by a distance greater than $2r$, the intersection of $D_r(\boldsymbol{x})$ and $D_r(\boldsymbol{y})$ is empty. In that case, the outermost integral in Equation 3.3 is zero, and $p(\boldsymbol{y}|\boldsymbol{x}) = 0$. This is because, under this model, it is impossible to take a step longer than $2r$.

In the following, we will refer to r as the "availability radius" parameter, and we will call this model the "availability radius model". This is a reference to the eponymous model

used by Arthur et al. (1996) and Rhodes et al. (2005) in the context of standard SSFs. In their SSF model, the habitat-independent movement kernel is uniform over a disc centred on the current location. Note that the model we propose is different: in our formulation, it is the half-step density ϕ of the local Gibbs algorithm which is uniform on a disc, not the habitat-independent movement kernel. A key difference is that the model of Arthur et al. (1996) and Rhodes et al. (2005) does not have an explicit stationary distribution.

We obtain the habitat-independent step density of the availability radius model by fixing the utilisation distribution $\pi(\boldsymbol{x})$ to a constant,

$$\begin{aligned} p_0(\boldsymbol{y}|\boldsymbol{x}) &= \frac{1}{\pi r^2} \int_{\boldsymbol{\mu} \in D_r(\boldsymbol{x}) \cap D_r(\boldsymbol{y})} \frac{1}{\int_{\boldsymbol{z} \in D_r(\boldsymbol{\mu})} d\boldsymbol{z}} d\boldsymbol{\mu} \\ &= \frac{1}{\pi r^2} \frac{1}{A(D_r(\boldsymbol{\mu}))} \int_{\boldsymbol{\mu} \in D_r(\boldsymbol{x}) \cap D_r(\boldsymbol{y})} d\boldsymbol{\mu} \\ &= \frac{1}{(\pi r^2)^2} A(D_r(\boldsymbol{x}) \cap D_r(\boldsymbol{y})), \end{aligned} \tag{3.4}$$

where $A(\cdot)$ denotes the area. In the absence of covariate effects, the likelihood of a step between two points $\boldsymbol{x}$ and $\boldsymbol{y}$ is therefore proportional to the area of intersection of $D_r(\boldsymbol{x})$ and $D_r(\boldsymbol{y})$. In Appendix B.1, we show that the area $A(D_r(\boldsymbol{x}) \cap D_r(\boldsymbol{y}))$ can be written in terms of the radius r and of $d = \|\boldsymbol{y} - \boldsymbol{x}\|$, the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$. Indeed, we have

$$A(D_r(\boldsymbol{x}) \cap D_r(\boldsymbol{y})) = \begin{cases} 2r^2 \cos^{-1}\left(\frac{d}{2r}\right) - rd\sqrt{1 - \frac{d^2}{4r^2}} & \text{if } d \leq 2r, \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

Following from Equation 3.4, the habitat-independent step density of the availability radius model becomes

$$p_0(\boldsymbol{y}|\boldsymbol{x}) = \begin{cases} \frac{1}{(\pi r^2)^2} \left[2r^2 \cos^{-1}\left(\frac{d}{2r}\right) - rd\sqrt{1 - \frac{d^2}{4r^2}} \right] & \text{if } d \leq 2r, \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

This density is radially symmetric around $\boldsymbol{x}$, as it only depends on the distance d to the endpoint $\boldsymbol{y}$. Figure 3.4 shows a slice through the habitat-independent step density of the availability radius model, for different values of r.

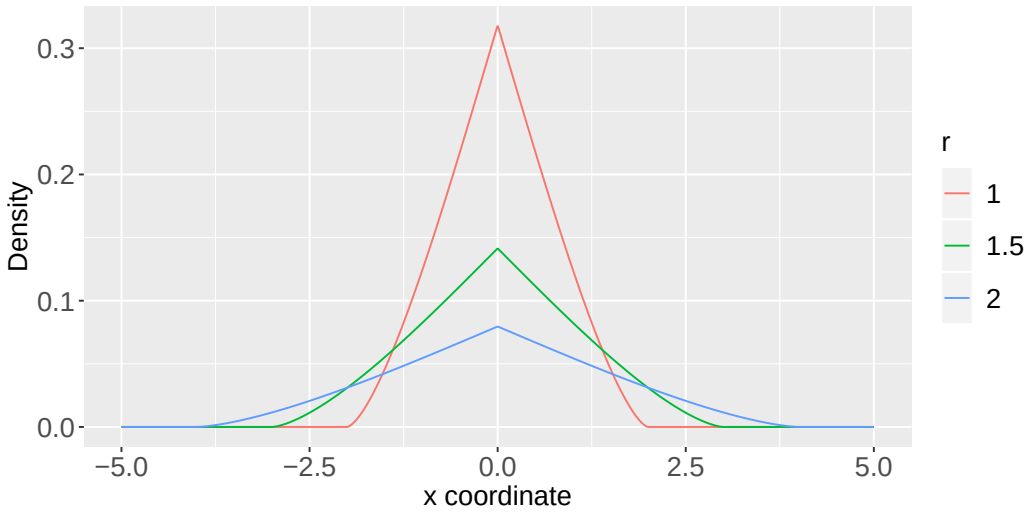


Fig. 3.4 Slice through the habitat-independent step density of the availability radius model, for different values of the radius r. The distribution is radially symmetric around the origin.

We can find the distribution of step lengths in the availability radius model, in the absence of covariate effects. The habitat-independent density of the step length $d = \|\boldsymbol{y} - \boldsymbol{x}\|$ is the integral of the density $p_0(\boldsymbol{y}|\boldsymbol{x})$ over an angular parameter $\theta \in [0, 2\pi]$. Without loss of generality, we consider that the origin of the step is $\boldsymbol{x} = (0, 0)$, and we omit the dependence on $\boldsymbol{x}$ in the following notation. We can write $\boldsymbol{y}$ in terms of its polar coordinates d and θ,

$$\boldsymbol{y}(d, \theta) = (d\cos\theta, d\sin\theta). \tag{3.7}$$

We calculate the habitat-independent step length density $p_0(d)$ as

$$p_0(d) = \int_{\theta=0}^{2\pi} p_0(\boldsymbol{y}(d, \theta))d\theta.$$

We apply a change of variable

$$p_0(d) = \int_{\theta=0}^{2\pi} p_0(d, \theta)|\boldsymbol{J}_y(d, \theta)|d\theta \tag{3.8}$$

where $\boldsymbol{J}_y(d, \theta)$ is the Jacobian matrix of the transformation given in Equation 3.7, and $|\cdot|$ is the determinant. We have

$$\boldsymbol{J}_y(d, \theta) = \begin{pmatrix} \cos\theta & -d\sin\theta \\ \sin\theta & d\cos\theta \end{pmatrix},$$

and therefore

$$|\boldsymbol{J}_y(d,\theta)| = d\cos^2(\theta) + d\sin^2(\theta) = d.$$

The density $p_0(d,\theta)$ is the expression of $p_0(\boldsymbol{y}|\boldsymbol{x})$ given in Equation 3.6 as a function of d. (It does not depend on θ because it is radially symmetric around $\boldsymbol{x}$.) For clarity, we denote $\mathcal{A}_r(d) = A(D_r(\boldsymbol{x}) \cap D_r(\boldsymbol{y}))$, as the area of intersection only depends on d (for a given r).

We make the substitutions in the integral of Equation 3.8,

$$\begin{aligned} p_0(d) &= \int_{\theta=0}^{2\pi} \frac{1}{(\pi r^2)^2} \mathcal{A}_r(d) \times d \times d\theta \\ &= \frac{d}{(\pi r^2)^2} \mathcal{A}_r(d) \int_{\theta=0}^{2\pi} d\theta \\ &= \frac{d}{(\pi r^2)^2} \mathcal{A}_r(d) \times 2\pi \\ &= \frac{2d}{\pi r^4} \mathcal{A}_r(d). \end{aligned}$$

Plugging in the expression of $\mathcal{A}_r(d)$, we find

$$p_0(d) = \begin{cases} \dfrac{4d}{\pi r^2} \cos^{-1}\left(\dfrac{d}{2r}\right) - \dfrac{2d^2}{\pi r^3}\sqrt{1 - \dfrac{d^2}{4r^2}} & \text{if } 0 \leq d \leq 2r \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

This result is actually well known, as the distribution of the distance between two points generated uniformly from a disc of radius r. It is a special case of Theorem 2.3.18 of Mathai (1999). In the availability radius model, $\boldsymbol{\mu}$ is uniformly distributed in $D_r(\boldsymbol{x})$, which is equivalent to saying that $\boldsymbol{x}$ is uniformly distributed in $D_r(\boldsymbol{\mu})$. In the absence of covariate effects, $\boldsymbol{y}$ is also uniformly distributed in $D_r(\boldsymbol{\mu})$. Finally, we see that both $\boldsymbol{x}$ and $\boldsymbol{y}$ are uniformly distributed in a disc of radius r, $D_r(\boldsymbol{\mu})$, which is why the formula holds.

Figure 3.5 shows the habitat-independent density of step lengths found in Equation 3.9, for different values of the radius parameter r.

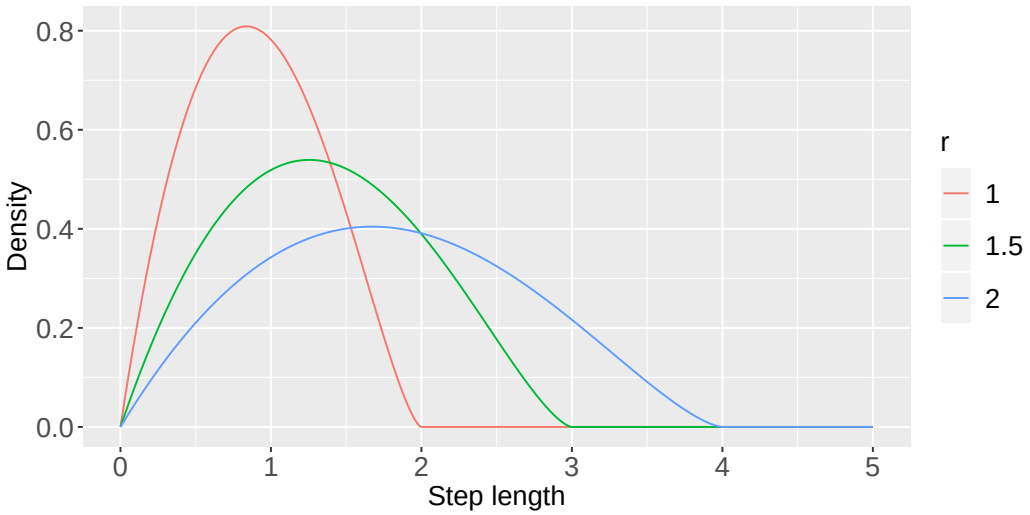


Fig. 3.5 Distributions of step lengths in the availability radius model, in the absence of covariate effects, for different values of the radius parameter r.

3.5.3 Other interesting formulations

3.5.3.1 Step length distributions

In the context of standard SSF models, Rhodes et al. (2005) explained how to choose the resource-independent movement kernel to yield a specific distribution of step lengths. They considered uniform turning angles (i.e. a radially symmetric movement kernel), and employed a method similar to the one we used in Section 3.5.2 to find the habitat-independent distribution of step lengths in the availability radius model. They showed that the habitat-independent movement kernel $\phi(\boldsymbol{y}|\boldsymbol{x})$ of the SSF model is linked to the habitat-independent distribution $p_0(d)$ of step lengths by the relationship

$$\phi(\boldsymbol{y}|\boldsymbol{x}) = \frac{p_0(d)}{2\pi d},$$

where d is the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, i.e. the step length. Rhodes et al. (2005) then suggested using an exponential distribution for the step lengths, i.e.

$$\phi(\boldsymbol{y}|\boldsymbol{x}) = \frac{\lambda \exp(-\lambda d)}{2\pi d},$$

with rate parameter $\lambda > 0$. In a similar framework, Forester et al. (2009) suggested a Weibull distribution of step lengths, i.e.

$$\phi(\boldsymbol{y}|\boldsymbol{x}) = \frac{\nu\lambda(\lambda d)^{\nu-1} \exp(-(\lambda d)^{\nu})}{2\pi d},$$

with scale parameter $\lambda > 0$ and shape parameter $\nu > 0$.

In the context of the local Gibbs model, we could use these functions as half-step densities (ϕ in Equation 3.2). They are radially symmetric, and therefore lead to valid MCMC algorithms. However, a (habitat-independent) step under the local Gibbs model is the sum of two ϕ-distributed increments, and so the habitat-independent step density and step length distribution do not generally have closed form expressions in this case.

3.5.3.2 Zero inflation

The half-step density ϕ of the local Gibbs model does not have to be a continuous function. In particular, a probability mass $m_0 \in (0, 1)$ can be assigned to the origin. When sampling from this zero-inflated half-step density, there is a probability m_0 that the sampled point is the origin, and a probability $1 - m_0$ that the sampled point is drawn from a continuous density (e.g. a normal density, or a uniform density on a disc, as described in the previous sections). Each step of the local Gibbs is the combination of two increments, so the probability of staying at the origin is m_0^2 in the zero-inflated model.

This is particularly useful to model the movements of animals that are susceptible to remain at the same location for more than one time interval. Note that the examples presented in Sections 3.5.1 and 3.5.2 do not forbid steps of length zero (i.e. immobility), because their densities at zero are positive, but zero inflation must be used for them to have a positive probability.

3.6 Mixture of local Gibbs steps

A mixture of MCMC algorithms, all with stationary distribution π, defines a valid MCMC algorithm for π. Tierney (1994) calls these mixtures "hybrid" algorithms. This is an interesting property for our application: an MCMC movement model can be defined by a combination of several transition kernels. Here, we introduce three extensions of the local Gibbs model, based on hybrid algorithms, to which we will return in Chapter 4.

3.6.1 Local Gibbs with random parameters

The local Gibbs algorithm is based on the choice of a symmetric distribution ϕ, the half-step density. Let $\boldsymbol{\theta}$ denote the vector of parameters of the half-step density, and $\phi(\cdot|\boldsymbol{x}, \boldsymbol{\theta})$ the half-step density around $\boldsymbol{x}$. (We make the dependence on $\boldsymbol{\theta}$ explicit in this section for notational clarity, but we will omit it for the rest of this chapter.)

An extension of the local Gibbs algorithm can be obtained by considering that the parameters $\boldsymbol{\theta}$ are themselves random, and are drawn at each iteration from a probability distribution $p(\boldsymbol{\theta}|\boldsymbol{\omega})$. This results in a hierarchical model, formulated in terms of the hyperparameters $\boldsymbol{\omega}$. In this case, the step density is obtained by integrating over $\boldsymbol{\theta}$, and it becomes

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\omega}) = \int_{\boldsymbol{\theta}} p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{\omega}) d\boldsymbol{\theta},$$

where $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$ is the step density, written in terms of the random parameters $\boldsymbol{\theta}$. From Equation 3.2, we obtain

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\omega}) = \pi(\boldsymbol{y}) \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\boldsymbol{\omega}) \int_{\boldsymbol{\mu} \in \Omega} \frac{\phi(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\theta}) \phi(\boldsymbol{\mu}|\boldsymbol{x}, \boldsymbol{\theta})}{\int_{\boldsymbol{z} \in \Omega} \pi(\boldsymbol{z}) \phi(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\theta}) d\boldsymbol{z}} d\boldsymbol{\mu} d\boldsymbol{\theta}. \tag{3.10}$$

This extension provides additional flexibility in the shape of the step density, to define a more realistic movement model. For example, the radius parameter r of the availability radius model could be treated as random (i.e. time-varying), to capture the variations in the scale of perception and movement of an animal through time. The radius parameter takes positive values, and could be modelled with a gamma distribution with shape parameter α and rate parameter ρ. In this example, $\boldsymbol{\theta} = r$, $\boldsymbol{\omega} = (\alpha, \rho)$, and $p(\boldsymbol{\theta}|\boldsymbol{\omega})$ is the probability density function of the gamma distribution. We explore this model formulation in more detail in Chapter 4.

Although the movement parameters can change randomly from one step to the next, they cannot depend on the current location of the process. For example, it could be tempting to make the scale parameter of the transition density a function of the covariate values at the current location. This could capture the tendency of an animal to move more slowly when the habitat is suitable, i.e. where the target distribution is high (perhaps because it is foraging). However, the resulting algorithm would generally not have π as its stationary distribution. In this example, the process would spend a disproportionate amount of time in areas where the target distribution is high. In the MCMC literature, samplers whose transition rules can depend on the current location of the process are called "adaptive" algorithms (Gilks et al., 1998). They have been studied because these "adaptations" can improve the mixing rate of the sampler, i.e. the speed of convergence to the target distribution. However, it has been shown that the convergence to the target distribution is not always guaranteed in adaptive algorithms (e.g. Andrieu and Thoms, 2008; Roberts and Rosenthal, 2009). In this work, we only consider the case where the transition rules change at random, and do not depend on the state of the process.

3.6.2 State-switching local Gibbs model

As mentioned in Section 1.2, patterns of animal movement are liable to vary through time, if the animal exhibits different behaviours. In particular, the speed of movement and the extent of the animal's perception may depend on the animal's behaviour. In the context of the local Gibbs model, failing to capture this change can lead to overestimating the scale of perception when the animal is moving slowly, and underestimating it when the animal is moving fast. This is because both types of movement would be modelled by the same, "averaged", movement model. This is a common problem in analyses of telemetry data collected over long durations, and state-switching models have been developed to capture the behavioural heterogeneity (Blackwell, 1997; Morales et al., 2004).

The local Gibbs model has two sets of parameters: the parameters $\boldsymbol{\beta}$ of the utilisation distribution, and the movement parameters $\boldsymbol{\theta}$ of the half-step density. For example, the movement parameters are the variance σ^2 in the normal kernel case, and the availability radius r in the availability radius model. This framework can be extended by considering that the animal can switch between N discrete states through time, each associated with a set of movement parameters $(\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(N)})$. We can model the switching behaviour with a latent process (S_t) defined on $\{1, \ldots, N\}$, which indicates which state is active at each time step t.

Each state $j \in \{1, \ldots, N\}$ is defined by a different step density $p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t, \boldsymbol{\beta}, \boldsymbol{\theta}^{(j)})$, i.e. by a different model of movement and perception. As an illustration, Figure 3.6 shows a trajectory simulated from a local Gibbs model with two states on an artificial utilisation distribution. Both states have a normal half-step density, but they have different scale parameters, $\sigma_1 = 0.2$ and $\sigma_2 = 1$. The scale parameter is larger in state 2, and so the region that is available to the animal at each time step is larger (wider perception and faster movement). State 2 may be used to represent an "exploratory" behavioural state, whereas state 1 is analogous to slower movement behaviours such as "area-restricted search".

The target distribution of the local Gibbs sampler does not depend on the movement parameters $\boldsymbol{\theta}$. It only depends on the selection parameters $\boldsymbol{\beta}$. In this multistate formulation, the movement process switches between N local Gibbs models, all with the same utilisation distribution π. (The parameters $\boldsymbol{\beta}$ are the same for all states.) The utilisation distribution of the state-switching model is therefore also π. The underlying MCMC algorithm can be seen as a hybrid algorithm, based on N transition kernels which share the same stationary distribution.

Roever et al. (2014) showed that ignoring animal behaviour in habitat selection studies could lead to incorrect conclusions. They argued for a two-stage modelling approach, in which

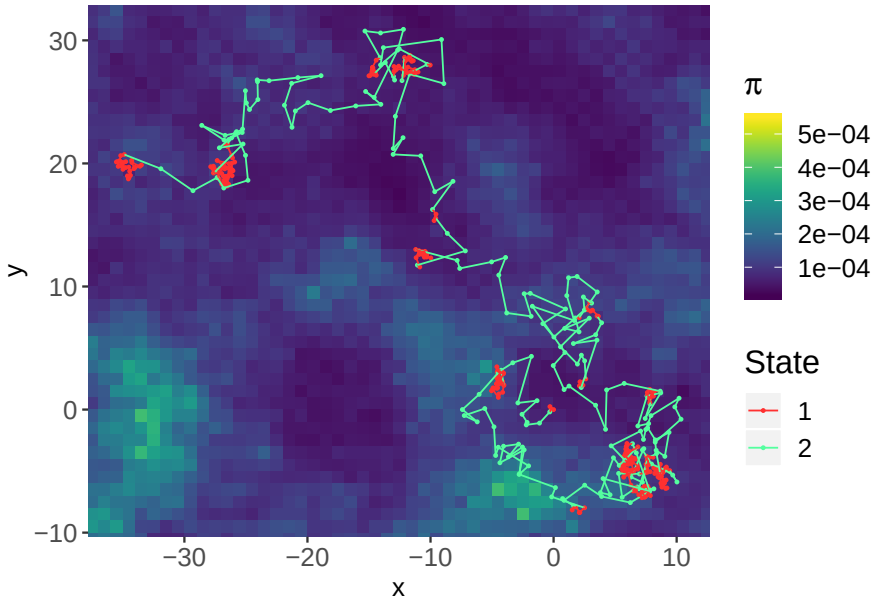


Fig. 3.6 Simulated trajectory from a 2-state local Gibbs model with normal half-step density, on an artificial utilisation distribution π. The two states are defined by two different movement parameters $\sigma_1 = 0.2$ and $\sigma_2 = 1$, leading to different speeds of exploration.

tracks would first be classified into behavioural states using a state-switching correlated random walk model (Morales et al., 2004), and a separate set of habitat selection parameters would then be estimated for each behavioural state. The state-switching local Gibbs model that we suggest here is different, because it estimates only one set of habitat selection parameters for all states. However, the scale of perception and movement can differ among the states, if they are characterised by different parameters $\boldsymbol{\theta}^{(j)}$. Then, the state-switching local Gibbs model does account for the behavioural heterogeneity in the scale of habitat selection.

3.6.3 Local Gibbs over irregular intervals

In Section 3.2, we argued that a movement model based on an MCMC algorithm is generally formulated in discrete time. Indeed, a time step of the model corresponds to an iteration of the algorithm. Therefore, although the habitat selection parameters of the model do not depend on the time scale (because they describe long-term selection), the movement parameters of an MCMC step selection model are generally scale-dependent.

This is in particular true of the local Gibbs sampler: the scale parameters of the half-step density are tied to a particular time scale. We can relax this constraint, by making an assumption on the relationship between the time interval and the scale of the half-step density.

There is no general scaling property for the parameters of the half-step density, but we can use the assumptions of Brownian motion to express this time dependence in the special case of the normal kernel model. The variance of the transition density of the Brownian motion is proportional to the length of the time interval (Einstein, 1905). Based on this assumption, we consider the local Gibbs model with half-step density $\phi(\cdot|\boldsymbol{x}) = \varphi(\cdot|\boldsymbol{x}, \Delta\sigma^2\boldsymbol{I}_2)$, where φ is the normal pdf and $\Delta > 0$ is the length of the time interval. The step density of this model over a time interval Δ can thus be written

$$p^{(\Delta)}(\boldsymbol{y}|\boldsymbol{x}) = \pi(\boldsymbol{y}) \int_{\boldsymbol{\mu}\in\Omega} \frac{\varphi(\boldsymbol{y}|\boldsymbol{\mu}, \Delta\sigma^2\boldsymbol{I}_2)\varphi(\boldsymbol{\mu}|\boldsymbol{x}, \Delta\sigma^2\boldsymbol{I}_2)}{\int_{\boldsymbol{z}\in\Omega} \pi(\boldsymbol{z})\varphi(\boldsymbol{z}|\boldsymbol{\mu}, \Delta\sigma^2\boldsymbol{I}_2)d\boldsymbol{z}} d\boldsymbol{\mu}. \tag{3.11}$$

A justification for this formulation can be found in the habitat-independent step density. In the absence of covariate effects, the step density over a time interval Δ is

$$p_0^{(\Delta)}(\boldsymbol{y}|\boldsymbol{x}) = \varphi(\boldsymbol{y}|\boldsymbol{x}, 2\Delta\sigma^2\boldsymbol{I}_2),$$

which is the step density of the Brownian motion with variance $2\sigma^2$.

This can be viewed as a special case of hybrid algorithm, where the transition kernel can change between iterations. Here, the transition kernel changes as a function of the time interval. This formulation is particularly valuable when the local Gibbs model is used to analyse telemetry data, which we will discuss in Chapter 4. The standard local Gibbs model could not be applied to estimate movement parameters from data collected at irregular time intervals, because those parameters are tied to a time scale. On the other hand, the model proposed in Equation 3.11 can accommodate irregular time intervals, and the scale parameter σ^2 is not tied to a particular discrete time step. We explore the potential of the normal kernel model to analyse irregular data in Section 4.7.3.

3.7 Simulation study

We presented a new type of MCMC algorithm, that we call the local Gibbs sampler, to model animal movement with a known utilisation distribution. In this section, we simulate from the local Gibbs sampler to verify that samples converge to the correct target distribution.

3.7.1 Simulation setup

We used two artificial spatial covariates c_1 and c_2 over $\Omega = [-20, 20] \times [-20, 20]$, with a resolution of 1 in each dimension. We generated them using the following procedure,

suggested by Avgar et al. (2016). We first generated a random uniform value in $[0, 1]$ for each grid cell. Then, we applied a (spatial) moving average with a circular window of radius $\rho = 10$ to each covariate field. That is, the value of each grid cell was updated to the mean of the values over a disc of radius ρ centred on that grid cell. A larger value of ρ leads to stronger spatial autocorrelation. Finally, we normalized the values of each covariate to be between 0 and 1.

We defined an artificial RSF in the exponential form (as described in Equation 1.1),

$$w(\boldsymbol{c}) = \exp(\beta_1 c_1 + \beta_2 c_2),$$

with $\beta_1 = 1$ and $\beta_2 = 3$. Plots of the simulated resources, and of the RSF, are shown in Figure 3.7.

We ran several simulations, described in Sections 3.7.2 and 3.7.3, to verify that the distribution of samples generated from the local Gibbs algorithm converge to the target distribution. In all experiments, the target distribution was the normalized RSF,

$$\pi(\boldsymbol{x}) = \frac{w(\boldsymbol{c}(\boldsymbol{x}))}{\int_{\boldsymbol{z}\in\Omega} w(\boldsymbol{c}(\boldsymbol{z}))d\boldsymbol{z}},$$

for any $\boldsymbol{x} \in \Omega$. We considered that the utilisation distribution was zero outside of Ω.

The empirical distribution of simulated points was defined by the count of sampled points in each cell of the covariate grid. We compared the normalized count of sampled points to the value of the target distribution in each grid cell, to assess the convergence of the simulated process to the correct distribution.

3.7.2 Effect of Monte Carlo sample size

It is generally not possible to simulate exactly from the local Gibbs sampler. Instead, we use Monte Carlo sampling to approximate the half-step density of the algorithm, as described in Section 3.4.3. At iteration t, we sample $\boldsymbol{\mu}$ from $\phi(\cdot|\boldsymbol{x}_t)$, and we generate K points $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K\}$ from $\phi(\cdot|\boldsymbol{\mu})$. The end point $\boldsymbol{x}_{t+1}$ is picked from the $\boldsymbol{z}_i$, with probabilities given by $\pi(\boldsymbol{z}_i)/\sum_j \pi(\boldsymbol{z}_j)$. This is a numerical approximation, and the size K of the Monte Carlo sample affects its accuracy. Here, we investigate the effect of the Monte Carlo sample size K on the long-term distribution of the simulated points.

We simulated from the local Gibbs sampler using different Monte Carlo sample sizes, and compared the distribution of simulated points to the true target distribution.

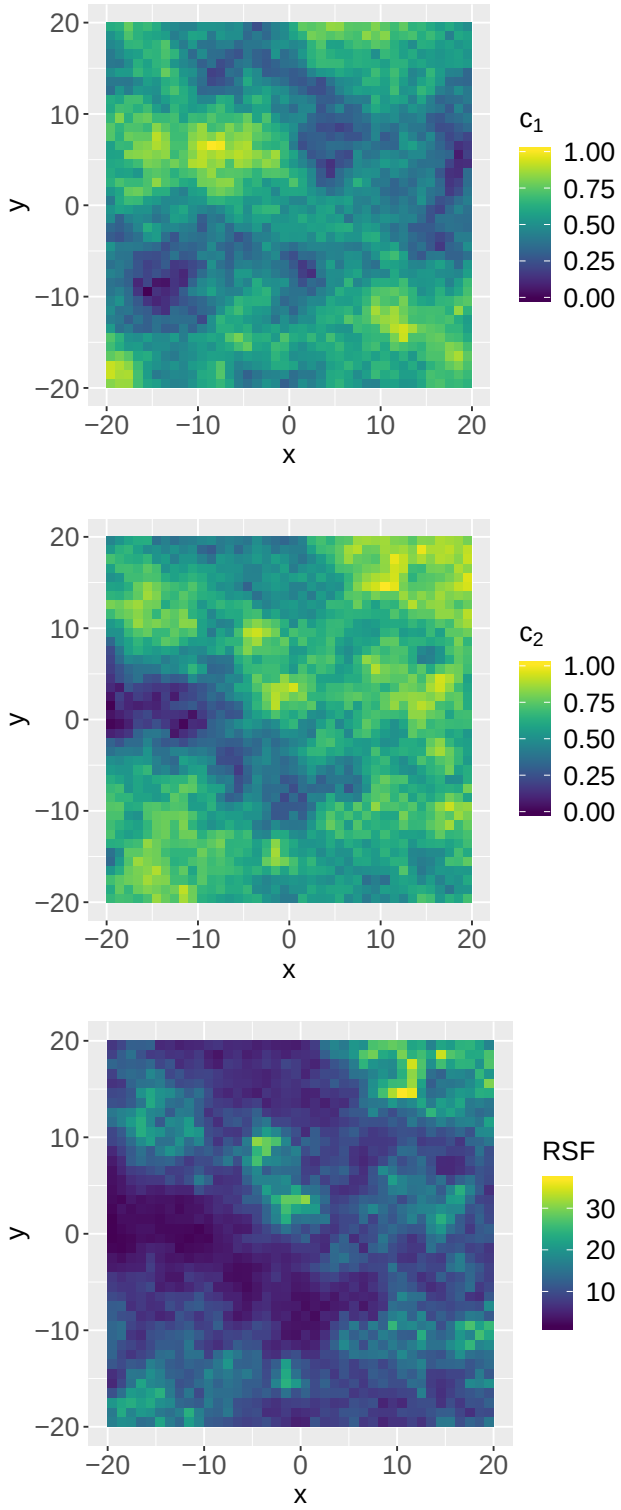


Fig. 3.7 Artificial covariates (top and middle), and resource selection function (bottom), used in the simulations.

3.7.2.1 Normal kernel model

We simulated 5×10^5 points from the local Gibbs model with normal half-step density described in Section 3.5.1. The normal kernel model has one scale parameter, which we set to $\sigma = 5$ in the simulations. (In Section 3.7.3, we investigate the distribution of samples generated for different values of σ.) We repeated the simulation with six different Monte Carlo sample sizes, $K = 2, 5, 20, 50, 200, 500$, to compare the distributions of sampled points to the target distribution π.

Figure 3.8 compares the value of the target distribution in each grid cell to the normalized count of simulated points, for each simulation scenario.

The results for the smaller Monte Carlo samples ($K = 2, 5$) display a lot of variance and a clear non-linear relationship. This indicates a discrepancy between the shapes of the distribution of simulated points and the target distribution, due to the very crude approximation made to evaluate the step density. For larger Monte Carlo samples ($K = 50$ or more), the empirical distribution captures the shape of the target distribution much more closely, as seen by the alignment with the identity line in Figure 3.8. There seems to be very little improvement between $K = 200$ and $K = 500$, which suggests that the step density of the algorithm is well approximated by a Monte Carlo sample of size $K = 200$, in this simulation scenario. Here, $K = 200$ thus seems like a good trade-off between the computational cost and the accuracy of the Monte Carlo approximation.

3.7.2.2 Availability radius model

In a similar experiment, we simulated 5×10^5 points from the local Gibbs model with uniform half-step density on a disc, described in Section 3.5.2. We set the radius parameter to $r = 5$ for the simulations, but we investigate simulations with different values of r in Section 3.7.3. We repeated this experiment for six different Monte Carlo sample sizes: $K = 2, 5, 20, 50, 200, 500$. The results are shown in Figure 3.9.

The results are very similar to those found for the normal kernel model. The approximation of the step density is very poor for small Monte Carlo samples ($K = 2, 5$). It improves for larger samples up to $K = 200$, beyond which little improvement seems to be achieved.

3.7.3 Effect of scale parameter

All formulations of the local Gibbs model that we consider in this work have a scale parameter, e.g. the variance σ^2 in the normal kernel model, and the radius r in the availability radius

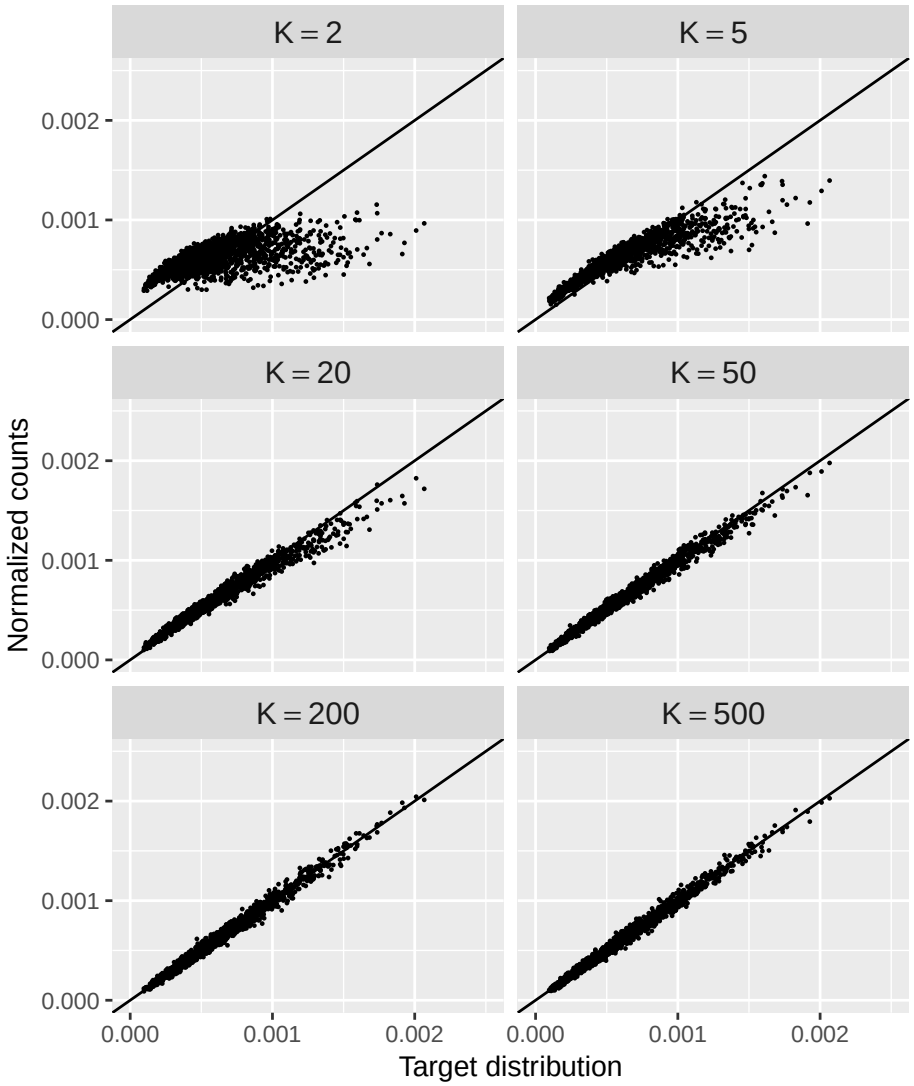


Fig. 3.8 Results of simulations from the normal kernel model, with different Monte Carlo sample sizes K. Alignment with the identity line indicate convergence of the simulated samples to the correct target distribution.

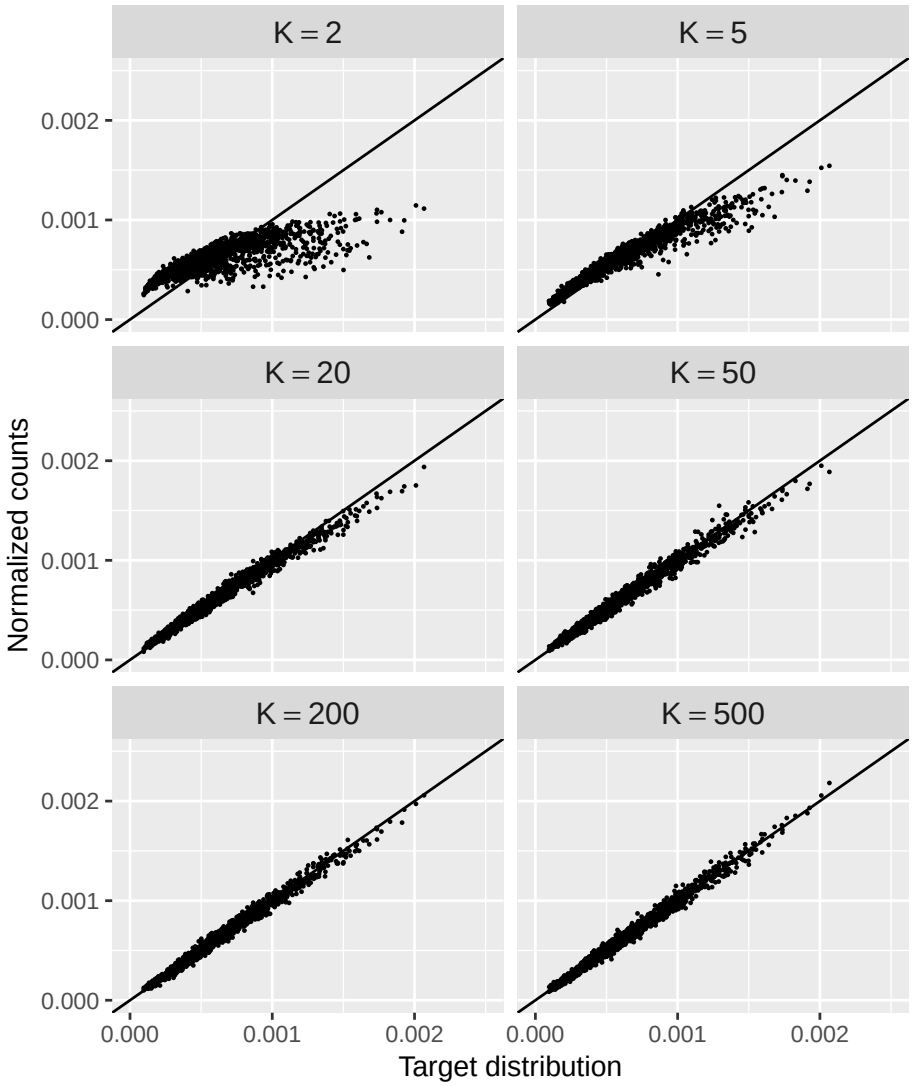


Fig. 3.9 Results of simulations from the availability radius model, with different Monte Carlo sample sizes K. Alignment with the identity line indicate convergence of the simulated samples to the correct target distribution.

model. This parameter determines the spatial extent of the half-step density, and is therefore related to the speed of exploration of the sampler. In this section, we investigate the effect of the scale parameter on the convergence of the samples to the target distribution.

3.7.3.1 Normal kernel model

We simulated 5×10^5 points from the normal kernel model, for the target distribution π. We used Monte Carlo samples of size $K = 200$, because little improvement was obtained for larger values of K in the simulations of Section 3.7.2. We ran this simulation for six different scale parameters, $\sigma = 0.25, 0.5, 1, 2, 4, 8$, and compared the distribution of samples to the target distribution in each case. Figure 3.10 displays the 300 first locations of each simulated data set, and clearly illustrates the different speeds of spatial exploration.

As in Section 3.7.2, we compared the empirical distribution of simulated points to the target distribution in each grid cell, to assess convergence of the sampler to the correct distribution. Plots of the normalized counts against values of the true target distribution are shown in Figure 3.11.

The plots of Figure 3.11 do not seem to display non-linearity, even for smaller values of the parameter σ. However, there is a lot of variance around the identity line for $\sigma = 0.25$ and, to a lesser extent, for $\sigma = 0.5$. This is because the sampler takes shorter steps (on average) for small values of σ, and is therefore slower to explore the target distribution. With $\sigma = 0.25$, it would take more iterations to obtain a good estimate of the target distribution. For values of σ larger than 2, the empirical distribution very accurately estimates the target distribution.

3.7.3.2 Availability radius model

We simulated 5×10^5 points from the availability radius model, with Monte Carlo sample size $K = 200$. We repeated the experiment for six different radius parameters: $r = 0.25, 0.5, 1, 2, 4, 8$. The results are shown in Figure 3.12.

The results are similar to those from the normal kernel model simulations. Figure 3.12 shows that there is a lot of variability in the estimation of the target distribution for small values of r, when the speed of spatial exploration of the sampler is low. The empirical distribution captures the true target distribution very well for larger r, in particular for $r \geq 4$.

Note that, for large enough values of r, the disc $D_r(\boldsymbol{\mu})$ would always cover the entire domain of definition Ω of the target distribution, and the algorithm would sample from the full target distribution directly.

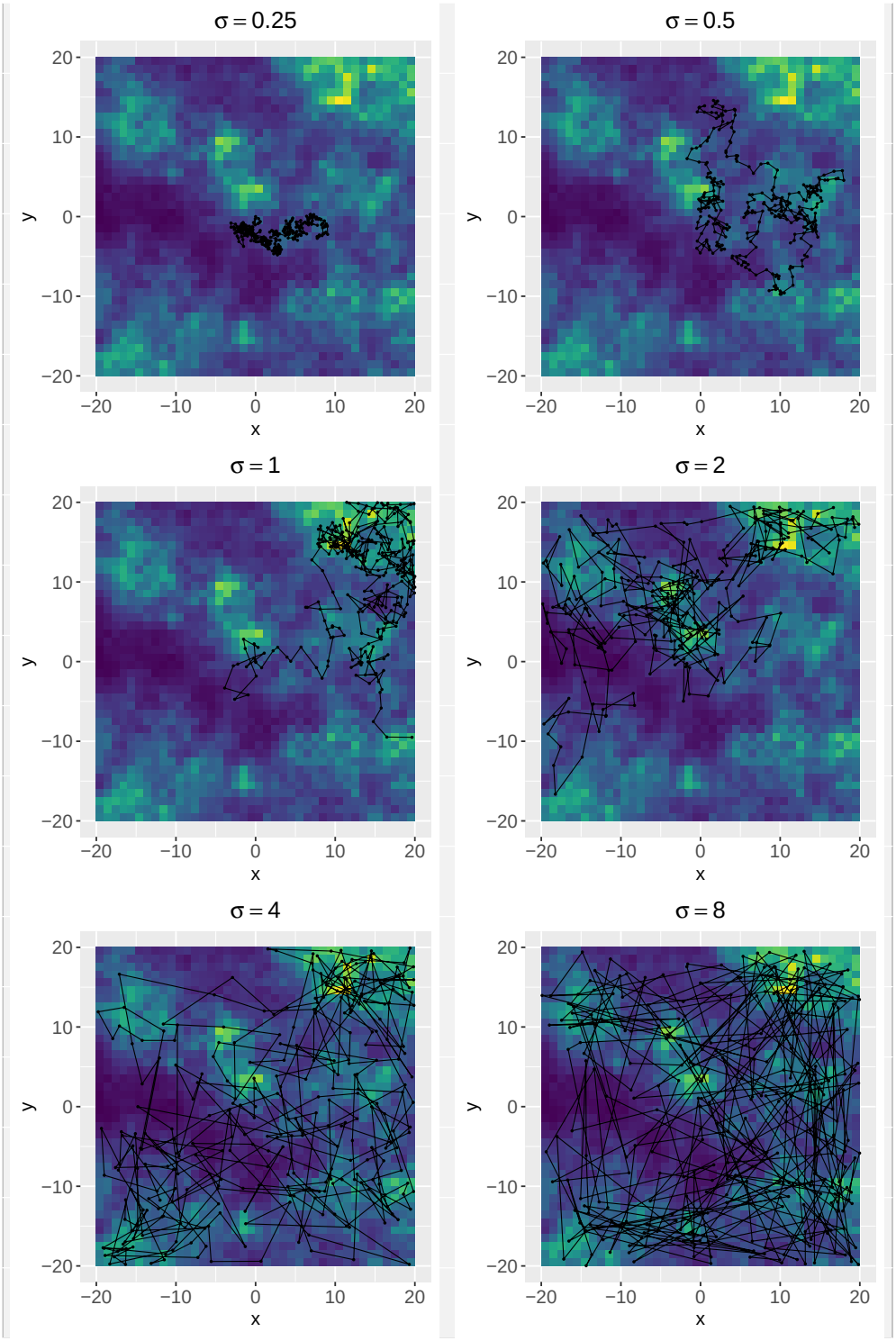


Fig. 3.10 Simulations from the local Gibbs movement model with normal half-step density, for different values of the scale parameter σ. Each plot shows 300 simulated locations, and the background is the target distribution.

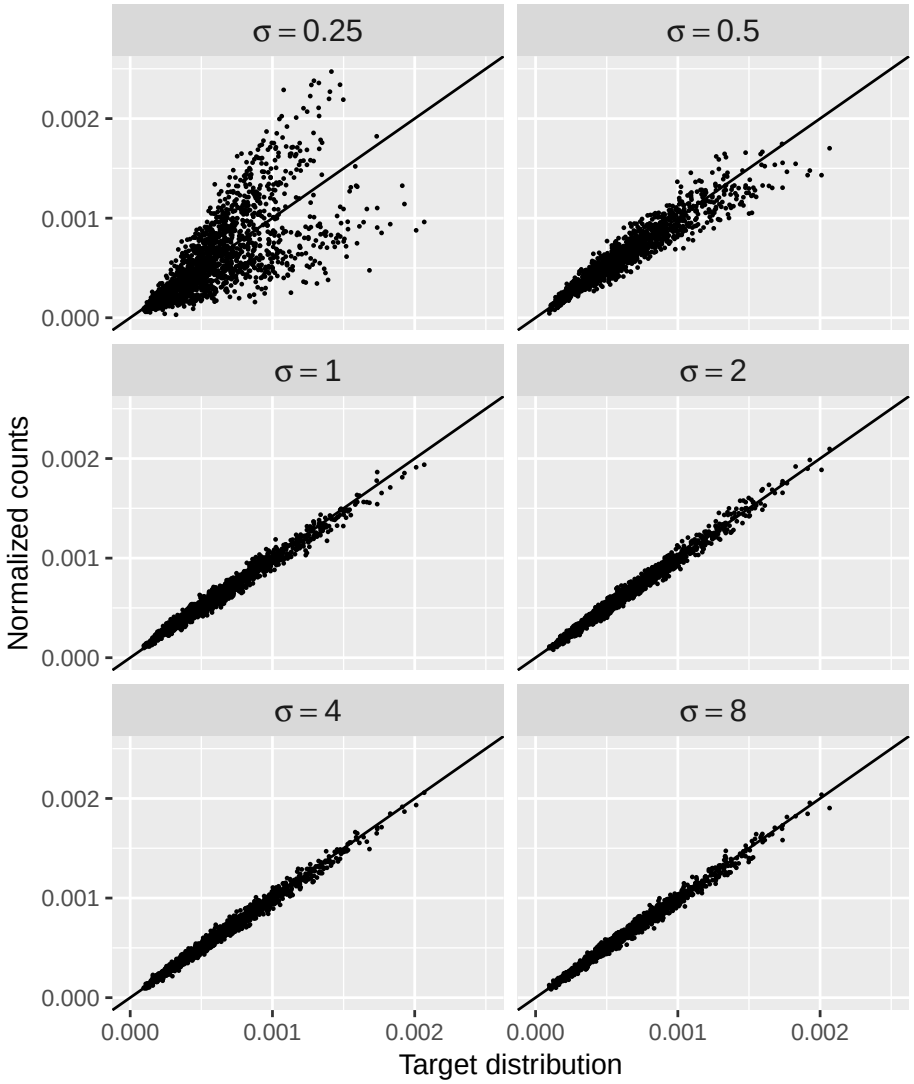


Fig. 3.11 Results of simulations from the normal kernel model, with different variance parameters, i.e. different speeds of spatial exploration. A small number of points were excluded from the plot for $\sigma = 0.25$, so that all results could be visualised on the same scale.

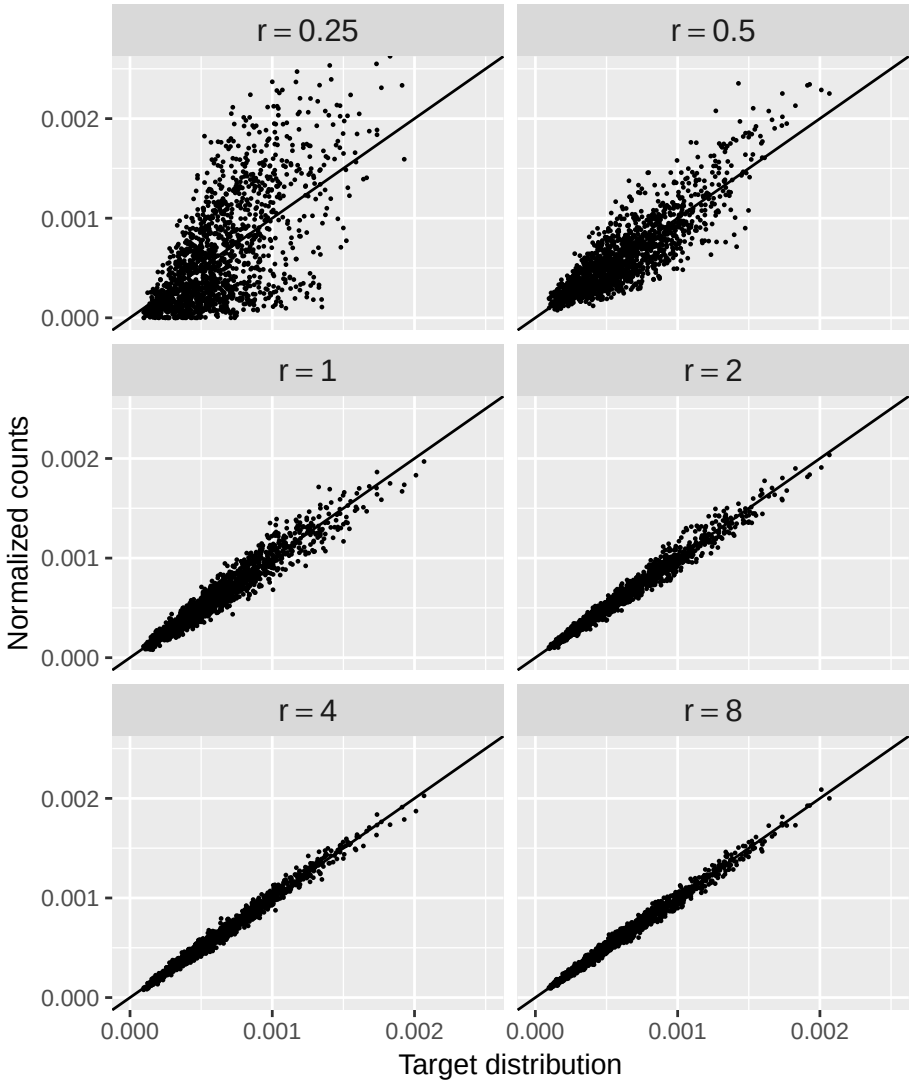


Fig. 3.12 Results of simulations from the availability radius model, with different availability radius parameters, i.e. different speeds of spatial exploration. A small number of points were excluded from the plot for $r = 0.25$, so that all results could be visualised on the same scale.

3.7.4 Discussion of simulations

The simulations confirmed empirically that the local Gibbs algorithm defines a valid MCMC sampler. Indeed, the distribution of samples from the algorithm converge to its target distribution. We also verified that the approximate sampling procedure proposed in Section 3.4.3 works for large enough Monte Carlo samples, i.e. a large enough K.

Note that the size of the Monte Carlo sample required to accurately approximate the step density may vary across scenarios. In particular, it may depend on the size of the region of availability, determined by the parameters of the half-step density (i.e. σ for the normal kernel model, and r for the availability radius model). If the region of availability is larger, the target distribution must be approximated over a larger domain in the simulations, and larger Monte Carlo samples may be required to achieve the same accuracy.

The accuracy of the approximation also depends on the spatial autocorrelation of the target distribution. If the autocorrelation is strong, i.e. if the target distribution is smooth, smaller Monte Carlo samples may be needed. On the other hand, the step density would be harder to approximate for a highly irregular target distribution (e.g. corresponding to very fragmented habitat), and larger Monte Carlo samples should be used.

3.8 Discussion

Linking models of animal movement and models of habitat selection is an important challenge in animal ecology. We suggested that Markov chain Monte Carlo algorithms can be used to construct a wide family of movement models, with a known stationary distribution. That stationary distribution is known in ecology as the utilisation distribution of the animal, i.e. the long-term probability density of the animal's location in space. If the utilisation distribution is modelled by a (normalized) resource selection function, the density of each step can be written in terms of the long-term habitat selection. The MCMC step selection model integrates advantages of the standard RSF and SSF approaches. The same parameters of habitat selection describe the movement at the scale of the local step and at the scale of global space use.

We focused on a specific MCMC algorithm, the local Gibbs sampler, to define the flexible local Gibbs movement model. It is specified by a half-step density ϕ, which determines the model of availability. Many different movement kernels can be achieved by different choices of ϕ. We made three assumptions on ϕ to define the local Gibbs algorithm: (1) it is symmetric, (2) we can generate samples from it, and (3) it is positive over a neighbourhood

of its origin. The third assumption is sufficient to ensure convergence of the algorithm to the stationary distribution π, but it is not a necessary condition, and it could be relaxed to obtain a wider family of samplers. For the purpose of our application to animal movement, however, we consider that it is always a sensible assumption.

The general framework presented here is not limited to the local Gibbs model, and other MCMC algorithms could be used. One limitation of the local Gibbs model is that it does not model movement persistence. Indeed, because the half-step density is required to be symmetric (to satisfy the detailed balance condition), all directions of movement are equally likely. For this reason, the local Gibbs model may fail to capture the autocorrelation in the direction of movement, which is often present in high-resolution animal movement data. Many movement models have been developed to capture this type of persistence, in particular the discrete- and continuous-time correlated random walks (Jonsen et al., 2005; Johnson et al., 2008a, respectively). In the context of MCMC movement models, this could be achieved with non-reversible MCMC samplers such as the one described by Michel and Sénécal (2017). The detailed balance condition is sufficient, but not necessary, to obtain convergence to a target distribution. Non-reversible MCMC algorithms do not satisfy detailed balance, but they nonetheless have a stationary distribution. The non-reversibility makes it possible to implement persistence in the direction of exploration of the sampler, which would become valuable in a model of animal movement.

In the local Gibbs model, the habitat selection only depends on the habitat at the endpoint $\boldsymbol{y}$ of a step, rather than over the whole step between $\boldsymbol{x}$ and $\boldsymbol{y}$. This is because the utilisation distribution of the model is formulated as a normalized RSF (Equation 1.1), defined for each point of the study region. Then, the covariates must be defined at points, rather than over steps. In this regard, the model presented here is less flexible than the SSF model of Fortin et al. (2005) and Potts et al. (2014a), which can describe habitat effects over movement steps. Similarly to RSF models, we assume that the utilisation distribution of the animal's location can be written in terms of the spatial distribution of environmental variables. In the definition of the utilisation distribution, we also consider that the covariates are constant in time, and so we cannot account for seasonality or other variations in their distributions. Because the local Gibbs model is formulated through the transition density of the sampler, one could in principle consider time-varying covariates which are updated at each iteration. However, this model would not have a stationary distribution, because π would change as the covariate distributions change.

In Section 3.6.3, we suggested that the properties of hybrid algorithms can be used to model movement over irregular time intervals. In particular, for the local Gibbs model with normal

half-step density, it is natural to scale the variance of the density proportionally to the length of the time interval. We presented this extension as a special case of the discrete-time hybrid algorithms. It could also be viewed as the basis for a continuous-time formulation of movement and habitat selection. Additional work is needed to formally describe the normal kernel model as a continuous-time model.

Others have proposed to use Markov processes to model animal movement and habitat selection. Hanks et al. (2015) used a continuous-time discrete-space Markov process to describe animal movement. In their method, the states of the Markov chain are the cells of a spatial grid, and the animal may transition from cell to cell through time. They allowed for the possibility to model the rate of transition from one cell to another in terms of their respective covariate values. This makes it possible to link the movement to habitat preferences, but they did not explicitly describe the long-term distribution of space use in their model. In a similar discrete-space approach, Whitehead and Jonsen (2013) suggested that the stationary distribution of the movement Markov chain can be used as an estimate of the utilisation distribution of the animal. As suggested by Wilson et al. (2018), it may be possible to combine the frameworks of Hanks et al. (2015) and Whitehead and Jonsen (2013), to capture space use and habitat selection within a discrete-space movement model. A disadvantage of their approaches, compared with the MCMC movement model proposed herein, is that they are formulated on a discrete spatial grid. As a consequence, the interpretation of their results is tied to the spatial scale of discretization. It is often chosen as the spatial grid on which the covariate were measured, which is arbitrary to describe the animal's movement.

If the spatial covariates used to construct the utilisation distribution take discrete values, the MCMC movement model presented in this chapter is similar to spatially-structured diffusion models (see Section 1.3.3 and Ovaskainen, 2004). For discrete-valued covariates, the target distribution of the MCMC movement model is discontinuous, with a different value for each habitat type. Then, the movement of the animal near the edge of an habitat patch is affected by this preferential selection. An advantage of spatially-structured diffusion models is that the parameters of movement (e.g. speed of diffusion) can depend on the current location of the process. As mentioned in Section 3.6.1, this is generally not possible in the MCMC movement model framework, because the stationary distribution would not be conserved. However, this limitation can be mitigated by a flexible time-varying movement kernel, such as the random availability radius model of Section 3.6.1, or the state-switching model of Section 3.6.2. The MCMC movement model is not limited to discrete-valued covariates, and it can be used for continuous variables. Another appeal of the method that we present here is

that the stationary distribution of the animal's location is formulated as a simple parametric function of the covariates, which links our framework to standard RSF models.

We presented a class of models, which constitutes the basis of a new framework for the analysis of animal movement and environmental data. In Chapter 4, we explain how the movement and habitat selection components of the local Gibbs model can be estimated for observed telemetry data.

Chapter 4

Inference in the local Gibbs model

In the previous chapter, we introduced a new framework to model animal movement and habitat selection, using Markov chain Monte Carlo to build a model of step selection. In this chapter, we show that this class of models can be fitted to animal telemetry data, to estimate parameters of movement and selection. In Section 4.2, we derive the general expression of the likelihood function of the local Gibbs model, and we explain how it can be calculated in practice using Monte Carlo integration. We focus on two special cases, the normal kernel model (Section 4.3) and the availability radius model (Section 4.4), for which we provide the analytical form of the likelihood, as well as details for its implementation. In Section 4.5, we describe maximum likelihood estimation for the local Gibbs model. We extend the framework to allow for switches between discrete behavioural states in Section 4.6. We use the hidden Markov model methodology to obtain inferences about the movement and selection parameters, as well as about the unobserved behaviour process. In Sections 4.7 and 4.8, we illustrate the estimation method in simulations and for the analysis of a zebra movement track, respectively. The research described in this chapter was presented in the preprint Michelot et al. (2018a).

4.1 Introduction

In Chapter 3, we described a new class of step selection models, based on the transition rules of MCMC algorithms. MCMC step selection models have two components: the movement model, and the habitat selection model. The movement component is defined by the parameters of the transition kernel of the underlying MCMC algorithm. The parameters of the transition kernel, treated as tuning parameters in the context of MCMC sampling, are

here viewed as parameters of the animal's movement and perception. The habitat selection component of the model is a resource selection function, written as a simple parametric function of spatial covariates of interest (Equation 1.1). Because it is based on a MCMC algorithm, the stationary distribution of the MCMC step selection model is guaranteed to be the (normalized) resource selection function. This stationary distribution is called the utilisation distribution of the animal, and gives the probability density of the animal's location in space.

The formulation of the MCMC step selection model was introduced in Section 3.3. Here, we modify the notation slightly to make the parametrisation of the model explicit. We denote by $(\boldsymbol{X}_t)_{t\in\mathbb{N}}$ the location process, and $p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t, \boldsymbol{\beta}, \boldsymbol{\theta})$ the step density of the MCMC step selection model between times t and $t+1$. Here, $\boldsymbol{\beta}$ is a vector of habitat selection parameters and $\boldsymbol{\theta}$ is a vector of movement parameters. The utilisation distribution $\pi(\boldsymbol{x}|\boldsymbol{\beta})$ of the model is a function of $\boldsymbol{\beta}$. The aim of this chapter is to establish a method to estimate $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ from telemetry and environmental data.

Animal location data and environmental data were presented in the general case in Sections 1.1 and 1.3.1 of the Introduction, respectively. In the following, we consider a sequence of T bivariate locations $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T)$ collected at regular time intervals on an animal, as a realisation of the location process $(\boldsymbol{X}_t)$. We consider J spatial covariates $\{c_1, c_2, \ldots, c_J\}$, defined over a study region $\Omega \subset \mathbb{R}^2$. We write $c_j(\boldsymbol{x})$ for the value of the covariate c_j at a spatial point $\boldsymbol{x} \in \Omega$. For notational convenience, we will often write $p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{\beta}, \boldsymbol{\theta}) = p(\boldsymbol{X}_{t+1} = \boldsymbol{x}_{t+1}|\boldsymbol{X}_t = \boldsymbol{x}_t, \boldsymbol{\beta}, \boldsymbol{\theta})$ the step density, when it is used as the likelihood of an observed step.

We consider the workflow highlighted in Section 3.3 for the analysis of telemetry and habitat data. This method requires the implementation of the likelihood function of the MCMC step selection model. The likelihood of a displacement is given by the transition kernel of the chosen MCMC algorithm, i.e. by the step density of the corresponding MCMC step selection model. Indeed, the step density $p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{\beta}, \boldsymbol{\theta})$ is the likelihood that the algorithm would take a step from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t+1}$, given the tuning parameters and the target distribution. Following from the Markov property, steps are independent under the MCMC step selection model. The likelihood of the whole trajectory is the product of the likelihood of the first location and the likelihoods of all the individual steps,

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T) = p(\boldsymbol{x}_1) \prod_{t=1}^{T-1} p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{\beta}, \boldsymbol{\theta}). \tag{4.1}$$

We will typically consider that the first observation is deterministic, and omit $p(\boldsymbol{x}_1)$ in the likelihood expression. We discuss the possibility to use the first observation to learn about the parameters at the end of the chapter, in Section 4.9.

In this chapter, we derive the likelihood of the local Gibbs model, introduced in Chapter 3, and we use it to estimate all model parameters. We first describe a general method to derive the likelihood of the local Gibbs model, for any half-step density ϕ (satisfying the assumptions described in Section 3.4). We then focus on some of the special cases introduced in the previous chapter, for illustrative purposes. These special cases are particularly relevant to the analysis of animal movement, and we apply them to a case study in Section 4.8.

4.2 Local Gibbs likelihood

The local Gibbs model can be used to infer movement and habitat selection parameters from tracking data. Here, we derive the likelihood of the local Gibbs model with general half-step density ϕ. We describe several special cases in Sections 4.3 and 4.4.

4.2.1 Exact likelihood

Under the local Gibbs movement model with half-step density ϕ, the likelihood of a step from $\boldsymbol{x}$ to $\boldsymbol{y}$ is given by the step density

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \pi(\boldsymbol{y}|\boldsymbol{\beta}) \int_{\boldsymbol{\mu}\in\Omega} \frac{\phi(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\theta})\phi(\boldsymbol{\mu}|\boldsymbol{x}, \boldsymbol{\theta})}{\int_{\boldsymbol{z}\in\Omega} \pi(\boldsymbol{z}|\boldsymbol{\beta})\phi(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\theta})d\boldsymbol{z}} d\boldsymbol{\mu}, \tag{4.2}$$

as shown in Section 3.4.

Consider T observed locations $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$, obtained at regular time intervals from a telemetry device. The local Gibbs model satisfies the Markov property, and each displacement from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t+1}$ is independent. Following from Equations 4.1 and 4.2, the likelihood of the track is

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T) = \prod_{t=1}^{T-1} \pi(\boldsymbol{x}_{t+1}|\boldsymbol{\beta}) \int_{\boldsymbol{\mu}\in\Omega} \frac{\phi(\boldsymbol{x}_{t+1}|\boldsymbol{\mu}, \boldsymbol{\theta})\phi(\boldsymbol{\mu}|\boldsymbol{x}_t, \boldsymbol{\theta})}{\int_{\boldsymbol{z}\in\Omega} \pi(\boldsymbol{z}|\boldsymbol{\beta})\phi(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\theta})d\boldsymbol{z}} d\boldsymbol{\mu}. \tag{4.3}$$

Likelihood-based methods can be used to estimate the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, from this expression. In Section 4.5, we describe maximum likelihood estimation for the local Gibbs model.

4.2.2 Monte Carlo approximation

The integrals of Equation 4.2 cannot generally be evaluated analytically, and numerical approximations must be used. We consider Monte Carlo integration, presented in Section 2.2, to obtain the approximate likelihood.

Monte Carlo integration can be applied to approximate the integrals in the local Gibbs likelihood of Equation 4.2. There are two separate integrals to estimate in this case: one integral over the intermediate centre $\boldsymbol{\mu} \in \Omega$, and one integral over the end point $\boldsymbol{z} \in \Omega$. We first consider the former, that we write as

$$I_\mu = \int_{\boldsymbol{\mu} \in \Omega} \phi(\boldsymbol{\mu}|\boldsymbol{x}, \boldsymbol{\theta}) \left(\frac{\phi(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\theta})}{\int_{\boldsymbol{z} \in \Omega} \pi(\boldsymbol{z}|\boldsymbol{\theta}) \phi(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\theta}) d\boldsymbol{z}} \right) d\boldsymbol{\mu}.$$

Here, $\phi(\boldsymbol{\mu}|\boldsymbol{x}, \boldsymbol{\theta})$ is a probability density function for $\boldsymbol{\mu}$ from which we can sample, corresponding to the function f in Equation 2.2. The expression in the brackets is the function g in Equation 2.2, that we want to integrate. We sample n_μ points $\{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{n_\mu}\}$ from $\phi(\cdot|\boldsymbol{x}, \boldsymbol{\theta})$, and calculate the Monte Carlo estimate as

$$\hat{I}_\mu = \frac{1}{n_\mu} \sum_{i=1}^{n_\mu} \frac{\phi(\boldsymbol{y}|\boldsymbol{\mu}_i, \boldsymbol{\theta})}{\int_{\boldsymbol{z} \in \Omega} \pi(\boldsymbol{z}|\boldsymbol{\theta}) \phi(\boldsymbol{z}|\boldsymbol{\mu}_i, \boldsymbol{\theta}) d\boldsymbol{z}}$$

We then consider the integral over $\boldsymbol{z}$ in the denominator,

$$I_z = \int_{\boldsymbol{z} \in \Omega} \pi(\boldsymbol{z}|\boldsymbol{\beta}) \phi(\boldsymbol{z}|\boldsymbol{\mu}_i, \boldsymbol{\theta}) d\boldsymbol{z}.$$

For a given $\boldsymbol{\mu}_i$, we sample n_z points $\{\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{i,n_z}\}$ from $\phi(\cdot|\boldsymbol{\mu}_i, \boldsymbol{\theta})$, and derive the Monte Carlo estimate,

$$\hat{I}_z = \frac{1}{n_z} \sum_{j=1}^{n_z} \pi(\boldsymbol{z}_{ij}|\boldsymbol{\beta}).$$

Finally, we combine the two approximations to obtain an estimate of the likelihood of a step from $\boldsymbol{x}$ to $\boldsymbol{y}$ under the local Gibbs model,

$$\hat{p}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \pi(\boldsymbol{y}|\boldsymbol{\beta}) \frac{n_z}{n_\mu} \sum_{i=1}^{n_\mu} \frac{\phi(\boldsymbol{y}|\boldsymbol{\mu}_i, \boldsymbol{\theta})}{\sum_{j=1}^{n_z} \pi(\boldsymbol{z}_{ij}|\boldsymbol{\beta})}. \tag{4.4}$$

In the following, we call $\{\boldsymbol{\mu}_i\}_i$ and $\{\boldsymbol{z}_{ij}\}_{i,j}$ the Monte Carlo samples. The approximation can be made arbitrarily accurate, by increasing the Monte Carlo sample sizes n_μ and n_z.

4.3 Normal kernel model

We first consider the normal kernel model, introduced in Section 3.5.1, wherein the half-step density ϕ is a normal density. It is a simple and natural formulation, leading to a standard Gaussian random walk in the absence of habitat selection. It is also a convenient model, because standard assumptions (from Brownian motion) can be used to apply the normal kernel model to data sampled at irregular time intervals. We describe that case in Section 4.3.3.

4.3.1 Likelihood

In the normal kernel model, the half-step density ϕ is a bivariate symmetric normal density, centred on the origin, and with variance parameter σ^2. The likelihood of a step from $\boldsymbol{x}$ to $\boldsymbol{y}$, given for the general case in Equation 4.2, becomes

$$p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\beta},\sigma) = \pi(\boldsymbol{y}|\boldsymbol{\beta}) \int_{\boldsymbol{\mu}\in\Omega} \frac{\varphi(\boldsymbol{y}|\boldsymbol{\mu},\sigma^2\boldsymbol{I}_2)\varphi(\boldsymbol{\mu}|\boldsymbol{x},\sigma^2\boldsymbol{I}_2)}{\int_{\boldsymbol{z}\in\Omega}\pi(\boldsymbol{z}|\boldsymbol{\beta})\varphi(\boldsymbol{z}|\boldsymbol{\mu},\sigma^2\boldsymbol{I}_2)d\boldsymbol{z}} d\boldsymbol{\mu},$$

where $\varphi(\cdot|\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the bivariate normal density with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{I}_2$ is the 2×2 identity matrix.

The integrals of the likelihood of a step under the normal kernel model cannot generally be calculated analytically. We use Monte Carlo integration to approximate the likelihood function, as follows. For $i \in \{1,\ldots,n_\mu\}$ and $j \in \{1,\ldots,n_z\}$, we sample

$$\begin{cases} \boldsymbol{\mu}_i \sim N(\boldsymbol{x},\sigma^2\boldsymbol{I}_2) \\ \boldsymbol{z}_{ij} \sim N(\boldsymbol{\mu}_i,\sigma^2\boldsymbol{I}_2) \end{cases}$$

Then, the approximate likelihood of a step from $\boldsymbol{x}$ to $\boldsymbol{y}$ is

$$\hat{p}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\beta},\sigma) = \pi(\boldsymbol{y})\frac{n_z}{n_\mu}\sum_{i=1}^{n_\mu}\frac{\varphi(\boldsymbol{y}|\boldsymbol{\mu}_i,\sigma^2\boldsymbol{I}_2)}{\sum_{j=1}^{n_z}\pi(\boldsymbol{z}_{ij})}. \tag{4.5}$$

The accuracy of the approximation is determined by the choice of n_μ and n_z, i.e. by the size of the Monte Carlo samples.

4.3.2 Implementation

We describe how the likelihood of a step from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t+1}$ can be computed under the normal kernel model, based on the Monte Carlo approximation given in Equation 4.5. The distribution of the Monte Carlo samples depends on the movement parameters, i.e. on the scale parameter σ of the half-step density. We propose to first generate standard normal samples, and then use the properties of the normal distribution to transform them and obtain the correct mean and variance.

We consider a Latin hypercube design to obtain the standard normal samples, to reduce the size of Monte Carlo samples needed to achieve a given accuracy and precision. Latin hypercube sampling is a method of variance reduction, described by McKay et al. (1979), which can be implemented as follows in two dimensions. We partition the domain $[0,1]\times[0,1]$ into $n_\mu \times n_\mu$ squares of equal dimensions. We select n_μ of these smaller squares such that there is exactly one square selected in each column and exactly one square selected in each row of the partition. There exist algorithms to obtain random squares that satisfy this condition (Jacobson and Matthews, 1996). Then, we sample one point $\boldsymbol{u}_i$ uniformly from each of the n_μ selected squares ($i = 1, \ldots, n_\mu$). The n_μ generated points are not independent, but they are uniformly distributed in $[0,1] \times [0,1]$, with the guarantee to cover each row and each column of the partition. Figure 4.1 shows one example of Latin hypercube sample for $n_\mu = 4$.

We transform each point $\boldsymbol{u}_i$ using the probit function in each dimension, to obtain $\boldsymbol{v}_i = \Phi^{-1}(\boldsymbol{u}_i)$, where Φ is the cumulative distribution function of the standard normal distribution. By property of the probability integral transformation, the points $\boldsymbol{v}_i$ are samples from the bivariate standard normal distribution. We obtain the Monte Carlo sample points by translation and scaling, as $\boldsymbol{\mu}_i = \boldsymbol{x}_t + \sigma \boldsymbol{v}_i$. By property of the normal distribution, $\boldsymbol{\mu}_i \sim N(\boldsymbol{x}_t, \sigma^2 \boldsymbol{I}_2)$, as required. A similar method is used to obtain $\boldsymbol{z}_{ij} \sim N(\boldsymbol{\mu}_i, \sigma^2 \boldsymbol{I}_2)$ from standard normal samples.

4.3.3 Normal kernel model with irregular intervals

In Section 3.6.3, we explained how the normal kernel model can be extended to describe movement over irregular time intervals. The idea is to scale the normal half-step density, and therefore the step density, with the length of the interval. In the normal kernel model, we suggest that the variance of the half-step density be proportional to the time interval, following the properties of Brownian motion. The framework of inference developed in this

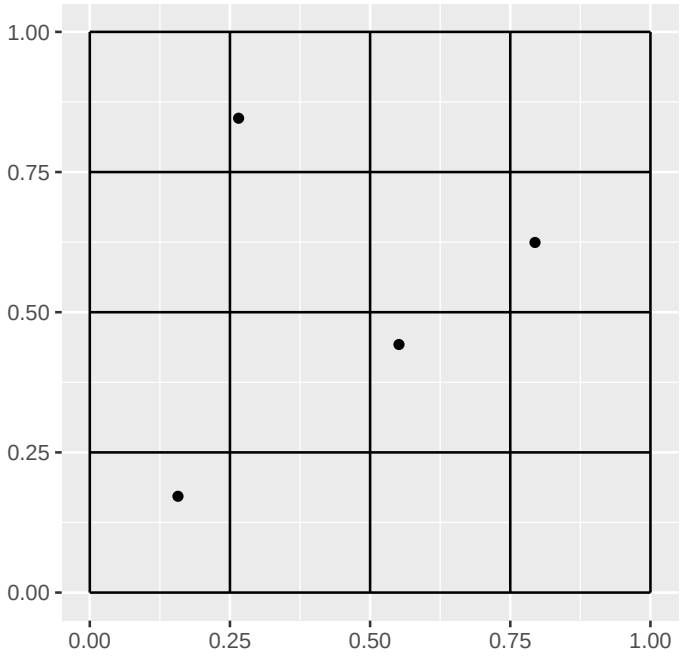


Fig. 4.1 Illustration of Latin hypercube design in two dimensions. There is exactly one sample in each row and in each column.

chapter can be used for the normal kernel model with irregular intervals, to analyse telemetry data collected irregularly.

The likelihood of a step from $\boldsymbol{x}$ to $\boldsymbol{y}$ over a time interval Δ is

$$p^{(\Delta)}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta}, \sigma) = \pi(\boldsymbol{y}) \int_{\boldsymbol{\mu}\in\Omega} \frac{\varphi(\boldsymbol{y}|\boldsymbol{\mu}, \Delta\sigma^2\boldsymbol{I}_2)\varphi(\boldsymbol{\mu}|\boldsymbol{x}, \Delta\sigma^2\boldsymbol{I}_2)}{\int_{\boldsymbol{z}\in\Omega} \pi(\boldsymbol{z})\varphi(\boldsymbol{z}|\boldsymbol{\mu}, \Delta\sigma^2\boldsymbol{I}_2)d\boldsymbol{z}} d\boldsymbol{\mu},$$

as we found in Section 3.6.3.

The Monte Carlo approximation is very similar to the regular case described in Section 4.3.1, but the variance σ^2 is scaled by the time interval Δ. So, for $i \in \{1, \ldots, n_\mu\}$ and $j \in \{1, \ldots, n_z\}$, we sample

$$\begin{cases} \boldsymbol{\mu}_i \sim N(\boldsymbol{x}, \Delta\sigma^2\boldsymbol{I}_2) \\ \boldsymbol{z}_{ij} \sim N(\boldsymbol{\mu}_i, \Delta\sigma^2\boldsymbol{I}_2) \end{cases}$$

Then, the approximate likelihood of a step from $\boldsymbol{x}$ to $\boldsymbol{y}$ over a time interval Δ is

$$\hat{p}^{(\Delta)}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\beta},\sigma) = \pi(\boldsymbol{y})\frac{n_z}{n_\mu}\sum_{i=1}^{n_\mu}\frac{\varphi(\boldsymbol{y}|\boldsymbol{\mu}_i,\Delta\sigma^2\boldsymbol{I}_2)}{\sum_{j=1}^{n_z}\pi(\boldsymbol{z}_{ij})}. \tag{4.6}$$

The approximate likelihood of the normal kernel model over irregular time intervals can be implemented using the method described in Section 4.3.2, substituting $\Delta\sigma^2$ for σ^2.

4.4 Availability radius model

We introduced the availability radius model in Section 3.5.2. In this section, we derive its likelihood and describe its implementation. In Section 4.4.3, we consider the case where the radius parameter r of the model is random, as an illustration of the hybrid algorithms described in Section 3.6.1.

4.4.1 Likelihood

In the availability radius model, the half-step density $\phi(\cdot|\boldsymbol{x})$ is uniform over the disc $D_r(\boldsymbol{x})$ of radius r, centred on the origin $\boldsymbol{x}$. Under this model, the likelihood of a step from $\boldsymbol{x}$ to $\boldsymbol{y}$ was found in Equation 3.3 to be

$$p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\beta},r) = \frac{\pi(\boldsymbol{y}|\boldsymbol{\beta})}{\pi r^2}\int_{\boldsymbol{\mu}\in D_r(\boldsymbol{x})\cap D_r(\boldsymbol{y})}\frac{1}{\int_{\boldsymbol{z}\in D_r(\boldsymbol{\mu})}\pi(\boldsymbol{z}|\boldsymbol{\beta})d\boldsymbol{z}}d\boldsymbol{\mu}.$$

The likelihood cannot generally be evaluated analytically, and we consider Monte Carlo integration as an approximation. For $i \in \{1,\dots,n_\mu\}$ and $j \in \{1,\dots,n_z\}$, we sample

$$\begin{cases}\boldsymbol{\mu}_i \sim U(D_r(\boldsymbol{x})\cap D_r(\boldsymbol{y}))\\ \boldsymbol{z}_{ij} \sim U(D_r(\boldsymbol{\mu}_i)),\end{cases}$$

where $U(\mathcal{E})$ denotes the multivariate uniform distribution over $\mathcal{E}$. Then, the approximate likelihood is

$$\hat{p}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\beta},r) = \frac{\pi(\boldsymbol{y}|\boldsymbol{\beta})}{\pi r^2}\cdot\frac{A(D_r(\boldsymbol{x})\cap D_r(\boldsymbol{y}))}{n_\mu}\cdot\sum_{i=1}^{n_\mu}\frac{n_z}{A(D_r(\boldsymbol{\mu}_i))}\frac{1}{\sum_{j=1}^{n_z}\pi(\boldsymbol{z}_{ij}|\boldsymbol{\beta})}$$

The term $A(D_r(\boldsymbol{\mu}_i))$ in the denominator is the area of a disc of radius r, i.e. πr^2. The term $A(D_r(\boldsymbol{x}) \cap D_r(\boldsymbol{y}))$ in the numerator is the area of intersection of the discs of radius r and of centres $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. In Section 3.5.2, we found that this area only depends on the distance $d = \|\boldsymbol{y} - \boldsymbol{x}\|$ between the centres of the two discs (Equation 3.5). If we denote $\mathcal{A}_r(d) = A(D_r(\boldsymbol{x}) \cap D_r(\boldsymbol{y}))$, the approximate likelihood can then be written

$$\hat{p}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta}, r) = \frac{\pi(\boldsymbol{y}|\boldsymbol{\beta})\mathcal{A}_r(d)}{(\pi r^2)^2} \frac{n_z}{n_\mu} \sum_{i=1}^{n_\mu} \frac{1}{\sum_{j=1}^{n_z} \pi(\boldsymbol{z}_{ij}|\boldsymbol{\beta})}. \tag{4.7}$$

4.4.2 Implementation

The approximate likelihood formula given in Equation 4.7 relies on the generation of two Monte Carlo samples, for each observed step: the intermediate centres $\{\boldsymbol{\mu}_i\}$, and the end points $\{\boldsymbol{z}_{ij}\}$. We describe one way to obtain these samples, to evaluate the likelihood $\hat{p}(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{\beta}, r)$ of a step from $\boldsymbol{x}_t = (x_t, y_t)$ to $\boldsymbol{x}_{t+1} = (x_{t+1}, y_{t+1})$. We write $d_t = \|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\|$ the distance between $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+1}$, i.e. the step length.

The intermediate centres For a radius r, the intermediate points $\boldsymbol{\mu}_i$ are sampled uniformly from the intersection of the discs $D_r(\boldsymbol{x}_t)$ and $D_r(\boldsymbol{x}_{t+1})$. A naive approach to generate them would be to sample points uniformly in $D_r(\boldsymbol{x}_t)$, and to keep those which are also in $D_r(\boldsymbol{x}_{t+1})$. However, the proportion of accepted points would greatly vary depending on the distance between $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+1}$. In particular, it would tend to zero when the step length approaches $2r$ (i.e. when the area of intersection of the two discs is close to zero). Instead, we propose to generate points uniformly from the smallest rectangle which includes the intersection of the discs, illustrated in Figure 4.2, and to reject those which do not lie within a distance r of both $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+1}$. In Appendix B.2, we show that, with this method, the expected acceptance rate is bounded above by $\pi/4 \approx 0.785$ (when $d_t = 0$), and bounded below by $2/3 \approx 0.667$ (when $d_t \to 2r$).

A point can be sampled uniformly from the intersection of the discs as follows. We sample a point (u_x, u_y) uniformly from the square of side 1 and centre $(0, 0)$,

$$\begin{cases} u_x \sim U(-0.5, 0.5) \\ u_y \sim U(-0.5, 0.5) \end{cases}$$

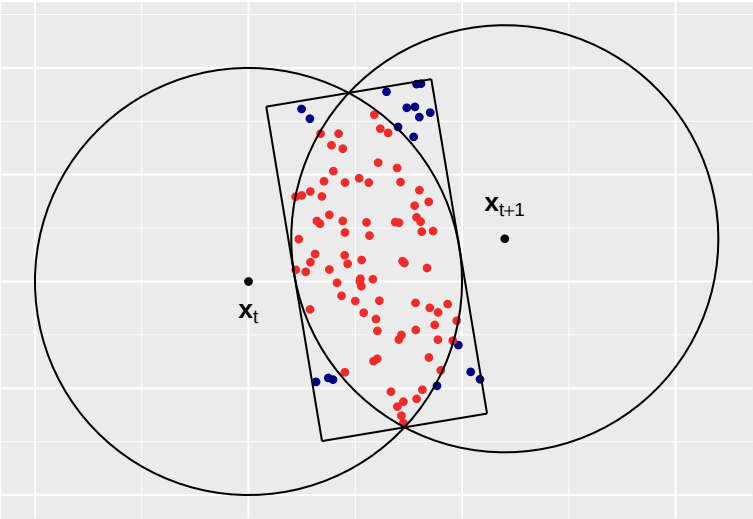


Fig. 4.2 Simulation of points in the intersection of two discs. The dots are sampled uniformly from the rectangle, and only the red dots are accepted.

Then, we scale both dimensions onto a rectangle of correct width and height,

$$\begin{cases} u'_x = (2r_i - d_t)u_x \\ u'_y = (2\sqrt{r^2 - d_t^2/4})u_y \end{cases}$$

We move to polar coordinates, to rotate the rectangle,

$$\begin{cases} l = \sqrt{(u'_x)^2 + (u'_y)^2} \\ \theta = \arctan(u'_y, u'_x) \end{cases}$$

where 'arctan' is the arctangent function of two arguments. We rotate the rectangle, to align it with the step from $\boldsymbol{x}$ to $\boldsymbol{x}_{t+1}$,

$$\theta' = \theta + \arctan(y_{t+1} - y_t, x_{t+1} - x_t)$$

Finally, we convert the sample back to Cartesian coordinates, and translate the centre of the rectangle to the mid-point between $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+1}$,

$$\begin{cases} u''_x = l\cos(\theta') + (x_t + x_{t+1})/2 \\ u''_y = l\sin(\theta') + (y_t + y_{t+1})/2 \end{cases}$$

The point (u''_x, u''_y) is sampled uniformly from the rectangle shown in Figure 4.2. We accept it—i.e. we take $\boldsymbol{\mu}_i = (u''_x, u''_y)$—if it is in $D_r(\boldsymbol{x}_t) \cap D_r(\boldsymbol{x}_{t+1})$, else we reject it. We repeat until n_μ points have been accepted. A Latin square design, similar to the one described in Section 4.3.2, can be used to generate the uniform samples (u_x, u_y) in the first step of the method.

The end points For each intermediate centre $\boldsymbol{\mu}_i$, the locations $\boldsymbol{z}_{ij}$ are sampled uniformly from the disc $D_r(\boldsymbol{\mu}_i)$, as follows. For each $j = 1, \ldots, n_z$, we take

$$\begin{cases} l_j \sim U(0, r^2) \\ \theta_j \sim U(-\pi, \pi) \end{cases}$$

Then, the point

$$\boldsymbol{z}_{ij} = \boldsymbol{\mu}_i + \sqrt{l_j} \begin{pmatrix} \cos\theta_j \\ \sin\theta_j \end{pmatrix}$$

is uniformly sampled from $D_r(\boldsymbol{\mu}_i)$.

The samples obtained from this procedure can be used to evaluate the approximate likelihood of a step under the availability radius model, as shown in Equation 4.7.

4.4.3 Random availability radius

As suggested in Section 3.6.1, additional flexibility can be achieved in the local Gibbs model by modelling the parameters of the half-step density as random. In the availability radius model, the half-step density has one parameter: the radius r. We can consider that r arises from the probability distribution $p(r|\boldsymbol{\omega})$, with (hyper-)parameters $\boldsymbol{\omega}$. For example, in the case study of Section 4.8, we investigate the case where r follows a gamma distribution with shape α and rate ρ, and we have $\boldsymbol{\omega} = (\alpha, \rho)$.

For a general distribution $p(r|\boldsymbol{\omega})$, following from Equation 3.10, the likelihood of a step from $\boldsymbol{x}$ to $\boldsymbol{y}$ is

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\omega}) = \frac{\pi(\boldsymbol{y}|\boldsymbol{\beta})}{\pi} \int_{r=d/2}^{\infty} \frac{p(r|\boldsymbol{\omega})}{r^2} \int_{\boldsymbol{\mu} \in D_r(\boldsymbol{x}) \cap D_r(\boldsymbol{y})} \frac{1}{\int_{\boldsymbol{z} \in D_r(\boldsymbol{\mu})} \pi(\boldsymbol{z}|\boldsymbol{\beta}) d\boldsymbol{z}} d\boldsymbol{\mu} dr. \quad (4.8)$$

The integral over the radius r starts at $d/2$, because a step cannot be longer than $2r$, and so the likelihood of a step is zero if $d > 2r$.

As before, we suggest Monte Carlo integration to approximate the likelihood. For $i \in \{1, \ldots, n_r\}$, $j \in \{1, \ldots, n_\mu\}$, and $k \in \{1, \ldots, n_z\}$, we sample

$$\begin{cases} r_i \sim p(r|\boldsymbol{\omega}) \\ \boldsymbol{\mu}_{ij} \sim U(D_{r_i}(\boldsymbol{x}) \cap D_{r_i}(\boldsymbol{y})) \\ \boldsymbol{z}_{ijk} \sim U(D_{r_i}(\boldsymbol{\mu}_{ij})) \end{cases}$$

Then, the Monte Carlo estimator of the likelihood is

$$\hat{p}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\omega}) = \frac{\pi(\boldsymbol{y}|\boldsymbol{\beta})}{\pi} \frac{1}{n_r} \sum_{i=1}^{n_r} \frac{1}{r_i^2} \frac{A(D_{r_i}(\boldsymbol{x}) \cap D_{r_i}(\boldsymbol{y}))}{n_\mu} \sum_{j=1}^{n_\mu} \frac{n_z}{A(D_{r_i}(\boldsymbol{\mu}_{ij}))} \frac{1}{\sum_{k=1}^{n_z} \pi(\boldsymbol{z}_{ijk})}.$$

We have $A(D_{r_i}(\boldsymbol{\mu}_{ij})) = \pi r_i^2$ for all j, and $A(D_{r_i}(\boldsymbol{x}) \cap D_{r_i}(\boldsymbol{y})) = \mathcal{A}_{r_i}(d)$. The expression simplifies to

$$\hat{p}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\omega}) = \frac{\pi(\boldsymbol{y}|\boldsymbol{\beta})}{\pi^2} \frac{n_z}{n_r n_\mu} \sum_{i=1}^{n_r} \frac{\mathcal{A}_{r_i}(d)}{r_i^4} \sum_{j=1}^{n_\mu} \frac{1}{\sum_{k=1}^{n_z} \pi(\boldsymbol{z}_{ijk})}. \tag{4.9}$$

The implementation of the approximate likelihood requires three sets of Monte Carlo samples: the availability radii $\{r_i\}$, the intermediate centres $\{\boldsymbol{\mu}_{ij}\}$, and the end points $\{\boldsymbol{z}_{ijk}\}$. The intermediate centres and end points can be generated as described in Section 4.4.2, similarly to the case where the radius parameter r is treated as constant.

The r_i are drawn from the probability density $p(r|\boldsymbol{\omega})$. Note that sampled radii that are smaller than $d/2$ do not contribute to the likelihood. Indeed, $A(D_{r_i}(\boldsymbol{x}) \cap D_{r_i}(\boldsymbol{y})) = 0$ when $r_i \leq d/2$. It is then more economical to sample the r_i from $p(r|\boldsymbol{\omega})$ truncated to $[d/2, \infty)$, and to multiply the likelihood by $1 - F_r(d/2|\boldsymbol{\omega})$, where F_r is the cumulative distribution function of the radius.

One way to sample from the truncated distribution is to generate a uniform sample $u_i \sim U(0, 1)$, and transform it with the quantile function $\tilde{F}_r$ of the truncated distribution:

$$r_i = \tilde{F}_r^{-1}(u_i) = F_r^{-1}[F_r(d/2|\boldsymbol{\omega}) + u_i(1 - F_r(d/2|\boldsymbol{\omega}))].$$

See Nadarajah et al. (2006) for details on the implementation of truncated distributions. The uniform samples u_i can be generated from a Latin hypercube design over the interval $[0, 1]$.

4.5 Parameter estimation in the local Gibbs model

In the previous sections, we explained how the likelihood of the local Gibbs model can be derived, and how it can be calculated in practice using Monte Carlo integration. In particular, we explained how the likelihood function can be implemented for the normal kernel model, and for the availability radius model. The likelihood function can then be used to estimate the parameters of the local Gibbs model, i.e. the parameters $\boldsymbol{\theta}$ of the half-step density, and the habitat selection parameters $\boldsymbol{\beta}$. We suggest maximum likelihood estimation, based on a numerical optimisation routine. Alternatively, the likelihood of the local Gibbs model could be used in a Bayesian context, e.g. to find the maximum a posteriori estimates, or to sample over the posterior distribution of the model parameters using an MCMC sampler.

The likelihood of T observations $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ under the local Gibbs model is a product of $T-1$ terms, corresponding to the $T-1$ observed steps. (Here, we assume that the first location is deterministic, and does not contribute to the likelihood.) In recent telemetry data sets, T may be very large, in the thousands or more. The likelihood can therefore become very large, if most terms are larger than 1, or very small, if most terms are smaller than 1. Due to limitations in computer precision, this can lead to numerical overflow (if the likelihood is too large to be encoded) or numerical underflow (if it is too small), respectively. A common solution is to compute the logarithm of the likelihood function, rather than the likelihood itself. The log-likelihood is much less susceptible to take extreme values. The logarithm function is strictly monotonic, so the same parameter values maximise the likelihood and the log-likelihood: the maximum likelihood estimates. In practice, we consider the approximate log-likelihood of the track under the local Gibbs model,

$$\hat{l}(\boldsymbol{\beta}, \boldsymbol{\theta} | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_T) = \sum_{t=1}^{T-1} \log(\hat{p}(\boldsymbol{x}_{t+1} | \boldsymbol{x}_t, \boldsymbol{\beta}, \boldsymbol{\theta})).$$

In the local Gibbs model with half-step density ϕ, the approximate likelihood of a step, $\hat{p}(\boldsymbol{x}_{t+1} | \boldsymbol{x}_t, \boldsymbol{\beta}, \boldsymbol{\theta})$, is given in Equation 4.4. The approximate log-likelihood of T observed locations is calculated as

$$\hat{l}(\boldsymbol{\beta}, \boldsymbol{\theta} | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_T) = \sum_{t=1}^{T-1} \Bigg[\log(\pi(\boldsymbol{x}_{t+1} | \boldsymbol{\beta})) + \log(n_z) - \log(n_\mu) + \log \left(\sum_{i=1}^{n_\mu} \frac{\phi(\boldsymbol{x}_{t+1} | \boldsymbol{\mu}_i^{(t)}, \boldsymbol{\theta})}{\sum_{j=1}^{n_z} \pi(\boldsymbol{z}_{ij}^{(t)} | \boldsymbol{\beta})} \right) \Bigg],$$

where $\{\boldsymbol{\mu}_i^{(t)}\}$ and $\{\boldsymbol{z}_{ij}^{(t)}\}$ are the Monte Carlo samples for the time step t, as defined in Section 4.2.2. To avoid storing a very large number of Monte Carlo samples, we can generate them only once, and scale them for each observed step. The procedure to translate and scale the uniform Monte Carlo samples to obtain $\boldsymbol{\mu}_i^{(t)}$ and $\boldsymbol{z}_{ij}^{(t)}$ is described in Section 4.3.2 for the normal kernel model, and in Section 4.4.2 for the availability radius model.

In all analyses presented in this chapter, we used the R function `optim` to maximise the log-likelihood with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, and obtain maximum likelihood estimates of the model parameters. The function `optim` is part of the base stats package, and it implements several numerical methods of optimisation. We used the default routine, based on the Nelder-Mead algorithm (Nelder and Mead, 1965). The optimiser takes starting parameter values as input, and iteratively evaluates the objective function (i.e. the log-likelihood) until the optimum is found.

The optimisation therefore requires the evaluation of the objective function (here, the log-likelihood) for many different sets of parameter values. If we randomly generated new Monte Carlo samples at each evaluation, the objective function would be slightly different at each iteration of the optimiser. This would cause numerical issues, with the optimum changing at each iteration. To circumvent this problem, we suggest generating the Monte Carlo samples only once, at the start, and using them for all the evaluations of the likelihood.

4.6 The state-switching local Gibbs model

In Section 3.6.2, we suggested that a model can be constructed as the mixture of several MCMC step selection models, corresponding to different behavioural states of an animal. We denote by S_t the behavioural state of the animal at time t. If the unobserved state process (S_t) is modelled as a Markov chain, the state-switching local Gibbs model becomes a hidden Markov model (HMM). HMMs are very popular tools to model behavioural heterogeneity in animal movement (Patterson et al., 2009; Langrock et al., 2012; Michelot et al., 2016). In particular, very efficient methods have been developed to evaluate the likelihood of a HMM, or to estimate the latent state sequence (Zucchini et al., 2016).

A HMM is a time series model composed of two stochastic processes: a state process (S_t), and an observation process $(\boldsymbol{Z}_t)$. The state process is defined as a Markov chain, as described in Section 2.1. We consider a N-state model, i.e. the behavioural state is defined in $\{1, \dots, N\}$. The state process satisfies the Markov property,

$$\Pr(S_{t+1} = j | S_t, S_{t-1}, \dots, S_1) = \Pr(S_{t+1} = j | S_t), \quad j = 1, \dots, N.$$

The future state of the process is independent of the past, conditional on the present state. The state process is fully specified by its initial distribution $\boldsymbol{\pi}^{(1)}$ and its transition probability matrix $\boldsymbol{\Gamma} = (\gamma_{ij})_{i,j=1}^{N}$, with

$$\gamma_{ij} = \Pr(S_{t+1} = j | S_t = i)$$

for any $i, j = 1, \ldots, N$.

The second process, $(\boldsymbol{Z}_t)$, models the observed variables. In the local Gibbs model, the observation at time t is the two-dimensional displacement ("step") from $\boldsymbol{X}_{t-1}$ to $\boldsymbol{X}_t$, i.e. $\boldsymbol{Z}_t = \boldsymbol{X}_t - \boldsymbol{X}_{t-1}$. In a HMM, the distribution of the observed variables is assumed to depend on the underlying state process, with the following property,

$$p(\boldsymbol{Z}_t | S_t, \ldots, S_1, \boldsymbol{Z}_{t-1}, \ldots, \boldsymbol{Z}_1) = p(\boldsymbol{Z}_t | S_t).$$

Conditionally on the current state, the observation $\boldsymbol{Z}_t$ is assumed to be independent of the past observations and the past states. The observation process is therefore fully specified by N state-dependent probability distributions, sometimes called emission distributions, defined as

$$p_j(\boldsymbol{Z}_t) = p(\boldsymbol{Z}_t | S_t = j).$$

In the local Gibbs model, the state-dependent probability distributions are the step densities of the model in each state. As described in Section 3.6.2, each state j is characterised by a different set of movement parameters $\boldsymbol{\theta}^{(j)}$, and we have

$$p_j(\boldsymbol{Z}_t = \boldsymbol{z}_t) = p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \boldsymbol{\beta}, \boldsymbol{\theta}^{(j)}),$$

where $\boldsymbol{z}_t = \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$.

The general dependence structure of a HMM is shown as a graph in Figure 4.3.

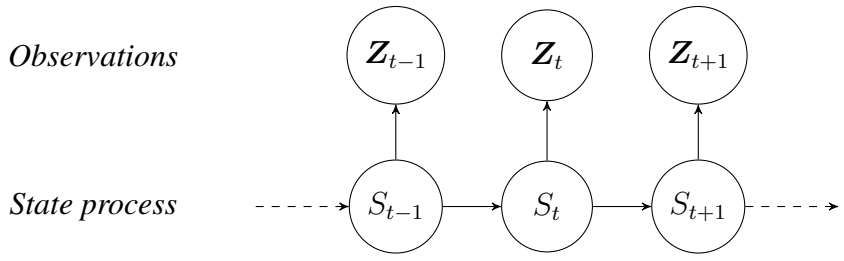


Fig. 4.3 Dependence graph of a hidden Markov model.

The dependence assumptions of HMMs greatly simplify the computation of the full model likelihood. However, the naive evaluation of the likelihood would still require T nested sums, where T is the length of the time series, to sum over all possible state sequences. The complexity of this evaluation is exponential in T, and it becomes computationally intractable for large numbers of observations. The so-called forward algorithm is a faster method that was developed to tackle this problem.

In the forward algorithm, the likelihood of T observations $\boldsymbol{z}_1, \dots, \boldsymbol{z}_T$ under a HMM is calculated as the following matrix product,

$$L(\boldsymbol{z}_1, \dots, \boldsymbol{z}_T) = \boldsymbol{\pi}^{(1)} \boldsymbol{P}(\boldsymbol{z}_1) \boldsymbol{\Gamma} \boldsymbol{P}(\boldsymbol{z}_2) \boldsymbol{\Gamma} \cdots \boldsymbol{P}(\boldsymbol{z}_T) \mathbf{1}',$$

where $\mathbf{1}$ is a column vector of N ones, and where $\boldsymbol{P}(\boldsymbol{z}_t)$ is the diagonal matrix of state-dependent densities,

$$\boldsymbol{P}(\boldsymbol{z}_t) = \begin{pmatrix} p_1(\boldsymbol{z}_t) & 0 & \cdots & 0 \\ 0 & p_2(\boldsymbol{z}_t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_N(\boldsymbol{z}_t) \end{pmatrix}.$$

It can be shown that the computational cost of the forward algorithm is linear in the number of observations T (Zucchini et al., 2016). Likelihood-based methods can therefore be used in many applications, even for long time series.

In the local Gibbs model, the state-dependent densities $p_j(\boldsymbol{z}_t)$ cannot be evaluated analytically, as explained in Section 4.2. To implement the state-switching local Gibbs model likelihood, we replace them by the Monte Carlo approximations of the step densities, $\hat{p}(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \boldsymbol{\beta}, \boldsymbol{\theta}^{(j)})$. We then obtain an approximation of the HMM likelihood, and it can be made arbitrarily accurate by increasing the size of the Monte Carlo samples. Direct maximisation of the (approximate) likelihood can be used to obtain estimates of all model parameters: the habitat selection parameters of the utilisation distribution $\boldsymbol{\beta}$, the state-dependent movement parameters of the step density $\{\boldsymbol{\theta}^{(j)}\}_{j=1}^N$, and the transition probabilities $(\gamma_{ij})_{i,j=1}^N$. The Viterbi algorithm is another computationally-efficient technique, used to obtain the most likely state sequence from a fitted HMM (Zucchini et al., 2016). In the context of animal movement modelling, the Viterbi algorithm classifies observed steps into behavioural phases.

In Section 4.7.4, we present a simulation study for the state-switching local Gibbs model with normal half-step density.

4.7 Simulation study

We consider simulations to investigate the performance of the method highlighted in Section 4.1 to recover the parameters of the local Gibbs model.

4.7.1 Simulation setup

We simulated movement trajectories from the local Gibbs sampler on an artificial target distribution, and used maximum likelihood estimation to recover all model parameters. We investigated the effect of the size of Monte Carlo samples in the approximation of the likelihood, and the effect of the number of locations, on the performance of the estimation. We considered three formulations: the normal kernel model, the state-switching normal kernel model, and the availability radius model with random radius. For each model, we implemented the approximate (log-)likelihood function, based on the Monte Carlo integration procedure described in Sections 4.2 to 4.4. In R, we used the numerical optimiser `optim` to maximise the log-likelihood, and obtain estimates of the parameters.

The artificial utilisation distribution was based on the environmental data from the case study of Section 4.8, to investigate the performance of the model in a realistic scenario. The data consist of a categorical raster of vegetation types, with four values: grassland, bushed grassland, bushland, and woodland. We defined the study region Ω as a square of 30km by 30km, and the raster was given at a resolution of 30m in each dimension. Because the covariate is categorical, we introduced three dummy covariates: c_1 (grassland), c_2 (bushed grassland), and c_3 (bushland). Each covariate c_i took the value 1 in the areas of the habitat type i, and 0 elsewhere. The fourth vegetation type (woodland) was the reference category. The utilisation distribution of the model, used as target distribution in the simulations, was formulated as

$$\pi(\boldsymbol{x}|\boldsymbol{\beta}) = \frac{1}{h(\boldsymbol{\beta})} \exp(\beta_1 c_1(\boldsymbol{x}) + \beta_2 c_2(\boldsymbol{x}) + \beta_3 c_3(\boldsymbol{x})),$$

for $\boldsymbol{x} \in \Omega$, where $h(\boldsymbol{\beta}) = \int_{\boldsymbol{z}\in\Omega} \exp(\beta_1 c_1(\boldsymbol{z}) + \beta_2 c_2(\boldsymbol{z}) + \beta_3 c_3(\boldsymbol{z}))d\boldsymbol{z}$ is a normalizing constant to ensure that π integrates to 1 over Ω. Like in the simulations of Chapter 3, we considered that the utilisation distribution was zero outside of Ω. For all simulations, we used the habitat selection parameters $\beta_1 = 3$, $\beta_2 = 2$, and $\beta_3 = 1$. Maps of the vegetation raster and of the (artificial) utilisation distribution are displayed in Figure 4.4.

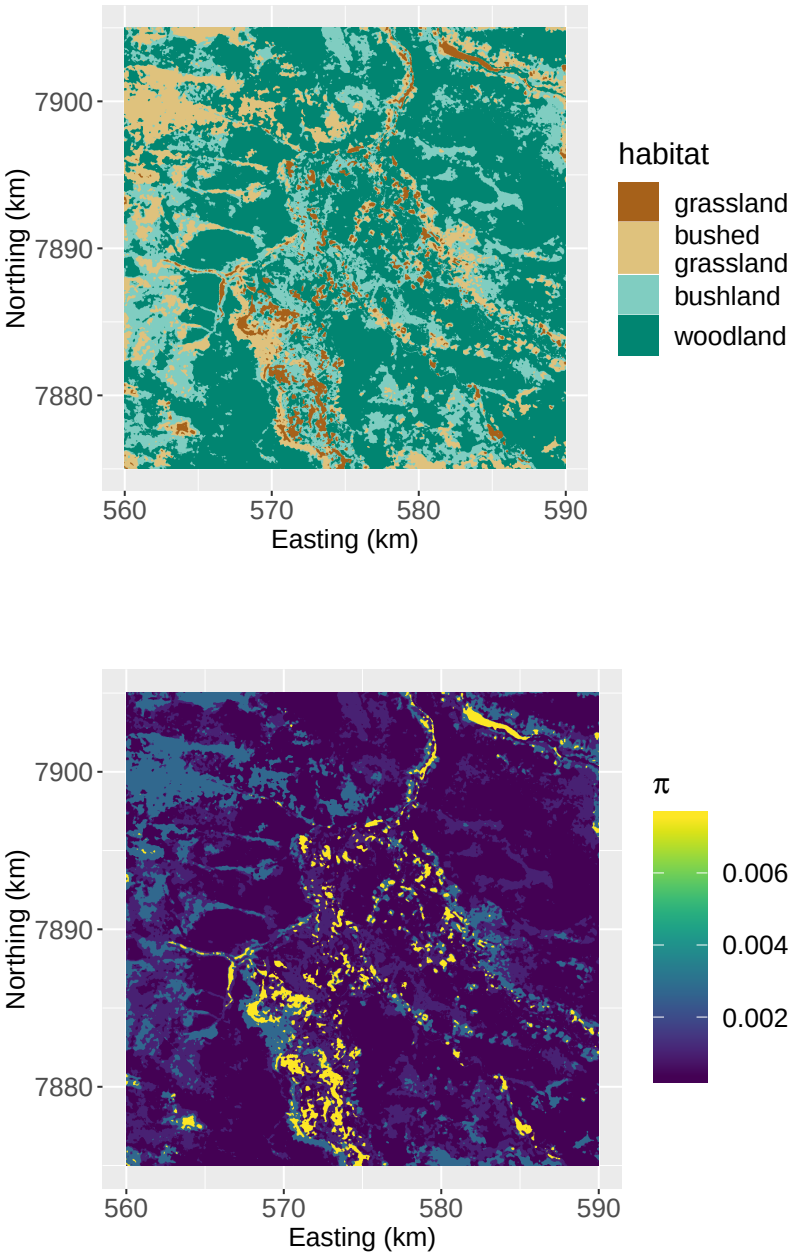


Fig. 4.4 Map of the vegetation raster and of the artificial utilisation distribution used in the simulations.

The spatial covariate is categorical, and so the utilisation distribution is constant over each habitat type. We can calculate the utilisation value π_i of habitat $i \in \{1, 2, 3, 4\}$ as

$$\pi_i = \frac{1}{h(\boldsymbol{\beta})} \exp(\beta_i),$$

with $h(\boldsymbol{\beta})$ as defined above. For each simulation study, we derived the estimated utilisation value $\hat{\pi}_i$ of each habitat, from the estimated habitat selection coefficients $\hat{\boldsymbol{\beta}}$, as

$$\hat{\pi}_i = \frac{1}{h(\hat{\boldsymbol{\beta}})} \exp(\hat{\beta}_i).$$

To assess the estimation, we compared the estimated utilisation values of each habitat type to the true values. This is a more interesting assessment than to compare the estimated habitat selection parameters $\hat{\boldsymbol{\beta}}$ to the true parameters $\boldsymbol{\beta}$. Indeed, a comparison of the utilisation values reflects how well the shape of the utilisation distribution was captured, and how well the relative habitat preferences were estimated.

For all simulations, the initial location was sampled from the artificial utilisation distribution π. The simulated track was accepted if it visited all four habitat types. Otherwise, it was rejected and another track was simulated, to ensure that all habitat selection parameters could be estimated.

4.7.2 Normal kernel model

We simulated tracks of length $T = 1000$ from the local Gibbs model with normal half-step density, with variance parameter $\sigma^2 = 1$, and target distribution π. We fitted the normal kernel model through numerical maximisation of the approximate log-likelihood, based on Equation 4.5. We tried several different sizes n_μ and n_z of the Monte Carlo samples, to investigate the effect of the approximation on the accuracy and precision of the estimation. For each scenario, the experiment was repeated 50 times. Results are shown in Figures 4.5.

The estimated utilisation values correctly captured the true utilisation distribution in most simulations. The precision of the utilisation estimates increased for larger Monte Carlo samples. The half-step density was very well estimated in almost all experiments. However, in one of the simulations with smaller Monte Carlo samples ($n_\mu = n_z = 20$), the variance of the half-step density was substantially overestimated.

We investigated the effect of the length of the track on the estimation. We simulated tracks of length $T = 100$, $T = 200$, and $T = 500$. To each, we fitted the normal kernel model with

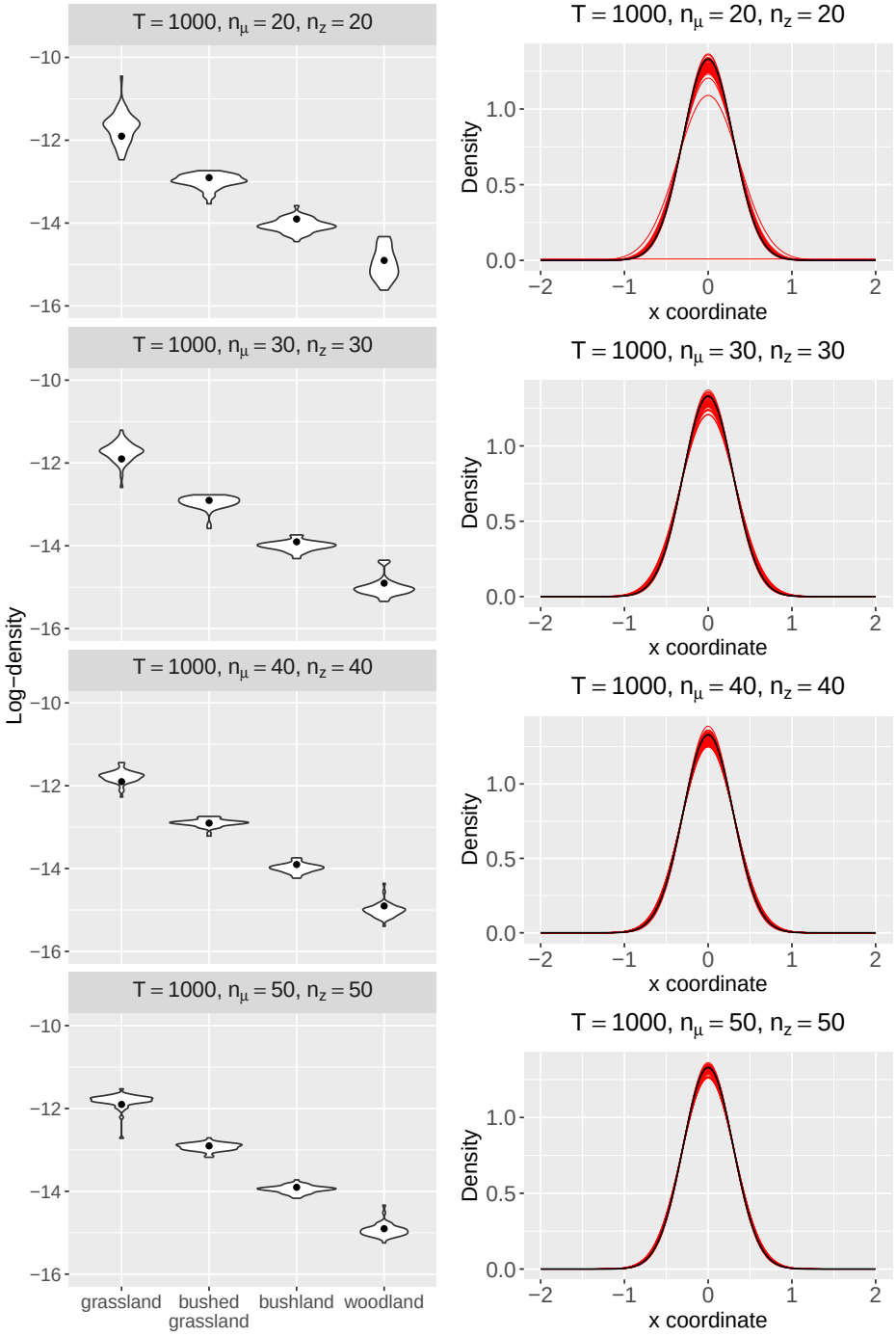


Fig. 4.5 Results of the simulation study for the normal kernel model, with $T = 1000$ and different sizes of Monte Carlo samples. Left column: Estimated utilisation values of the four habitat types, on the log scale. The violins show the 50 sets of estimates, and the dots are the true values. Right column: Estimated half-step densities. The red lines are the 50 estimated densities, and the black line is the true density used to simulate the tracks.

Monte Carlo samples of size $n_\mu = n_z = 40$. The experiment was repeated 50 times. The results are shown in Figures 4.6.

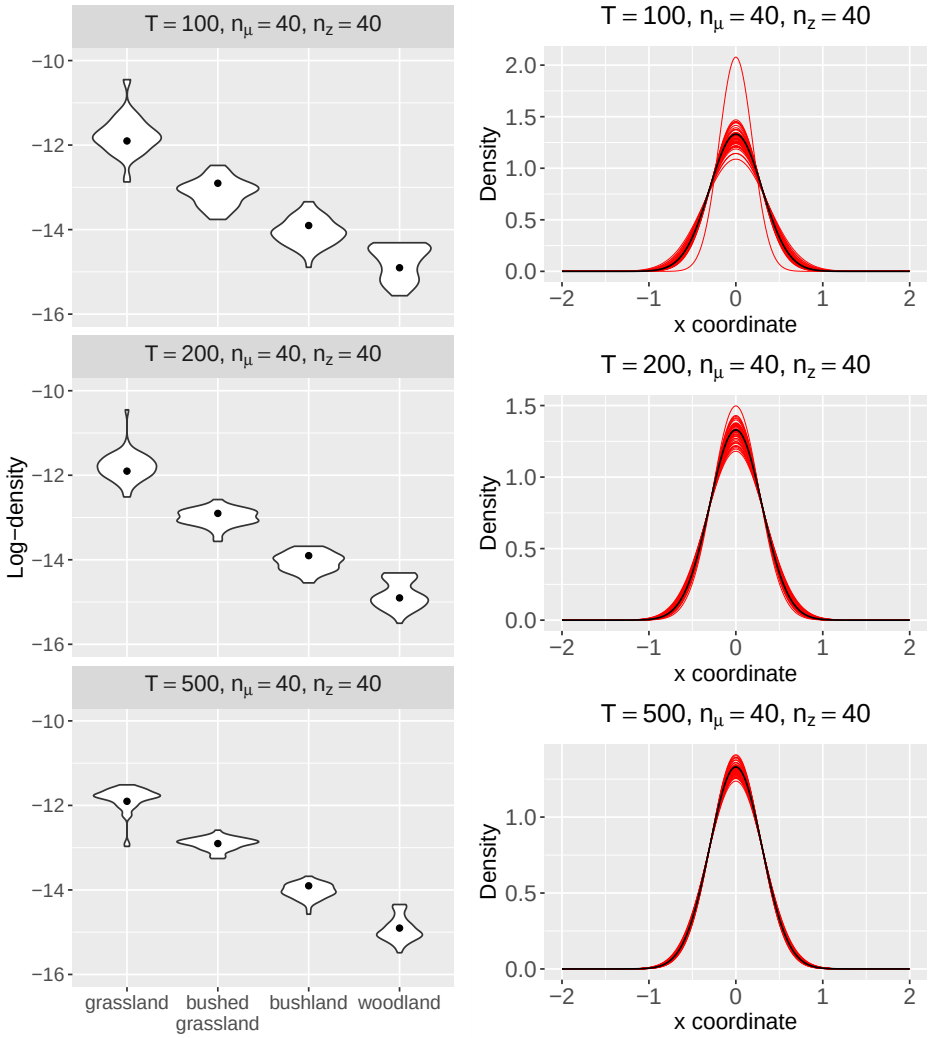


Fig. 4.6 Results of the simulation study for the normal kernel model, with $n_\mu = n_z = 40$ and different track lengths. Left column: Estimated utilisation values of the four habitat types, on the log scale. The violins show the 50 sets of estimates, and the dots are the true values. Right column: Estimated half-step densities. The red lines are the 50 estimated densities, and the black line is the true density used to simulate the tracks.

In all experiments, the utilisation distribution seemed to be captured in the estimation. The variability in the estimates was large for the shortest tracks, of length $T = 100$, and the precision increased for longer simulated tracks. The half-step density was adequately estimated in most simulations, although the variance was clearly underestimated in one of the experiments with $T = 100$.

There is a trade-off between the accuracy of the estimates and the computational cost of the method, determined by the choice of the size of Monte Carlo samples.

4.7.3 Normal kernel model with irregular time intervals

In Section 3.5.1, we suggested that the normal kernel model could be applied to irregular time intervals. In this formulation, the variance of the normal half-step density is expressed as $\Delta\sigma^2$, where $\Delta > 0$ is the length of the time interval, and σ^2 is an unknown scale parameter. We used simulations to investigate the performance of the method to recover model parameters from irregular location data.

We simulated 10^4 locations over time intervals of length $\Delta = 0.1$, with parameter $\sigma^2 = 1$. We then thinned the simulated trajectory by keeping 1000 locations at random, to imitate real data collected at irregular time interval. We implemented the approximate likelihood described in Section 4.3.3 to estimate the habitat selection and movement parameters, from the thinned track. We used Monte Carlo samples of sizes $n_\mu = n_z = 50$, and we repeated the experiment 50 times. The results are displayed in Figure 4.7.

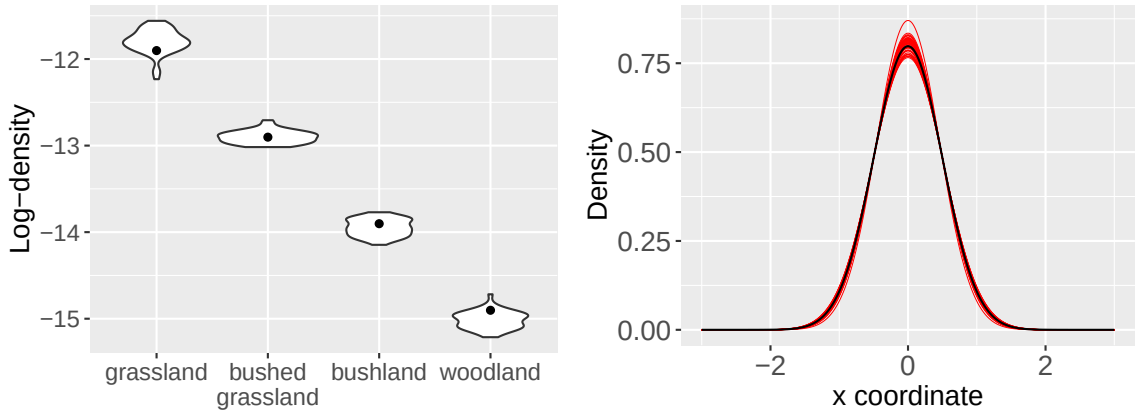


Fig. 4.7 Results of the simulation study with irregular time intervals. Left: Estimated utilisation values of the four habitat types, on the log scale. The violins show the 50 sets of estimates, and the dots are the true values. Right: Estimated half-step densities, for a time interval set to $\Delta = 1$. The red lines are the 50 estimated densities, and the black line is the true density used to simulate the tracks.

The habitat selection parameters, and the shape of the utilisation distribution, were well estimated in all 50 experiments. The true variance parameter of the half-step density was also captured well in the estimation. This illustrates the applicability of the normal kernel model to animal tracking data collected at irregular intervals.

4.7.4 State-switching normal kernel model

We considered the 2-state local Gibbs model with normal half-step density, on the target distribution π. In state $i \in \{1, 2\}$, the movement process was described by the normal kernel model with scale parameter σ_i. We chose the parameters

$$(\sigma_1, \sigma_2) = (0.5, 2), \quad \text{and } \mathbf{\Gamma} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}.$$

Large diagonal elements in the transition probability matrix $\mathbf{\Gamma}$ create state persistence, i.e. the process tends to persist in the current behaviour for several time steps. This is a desirable feature in real animal movement data, because a lack of state persistence would indicate that the behaviours happen at a finer time scale than that of the observations. The scale parameter affects the size of the area of perception, and the speed of movement of the simulated movement process. Here, $\sigma_1 < \sigma_2$, and so the process will tend to take longer steps in state 2. State 2 may represent fast exploratory movement, whereas state 1 may be analogous to slower area-restricted search behaviour.

We simulated a track of length $T = 1000$, and fitted the state-switching normal kernel model as a hidden Markov model (using the forward algorithm to implement the likelihood, Section 4.6). There were seven parameters to estimate: three habitat selection parameters (β_1, β_2, β_3), two movement parameters (σ_1, σ_2), and two transition probabilities (γ_{12}, γ_{21}). We used Monte Carlo samples of different sizes: $n_\mu = n_z = 20, 30, 40, 50$. We repeated each experiment 50 times, and the results are shown in Figure 4.8.

In the simulations with small Monte Carlo samples ($n_\mu = n_z = 20$), the utilisation estimates appear to be biased. The utilisation value of grasslands was overestimated, and the utilisation value of bushlands was underestimated, in most of those simulation runs. However, this bias decreased for larger Monte Carlo samples. The half-step densities seemed to be well estimated for both states, in almost all experiments. In one of the simulations with small Monte Carlo sample sizes $n_\mu = n_z = 20$, the variance in state 1 was largely overestimated, and all steps were classified in state 2. As a result, the estimated half-step density in state 2 was intermediate between the true half-step densities of states 1 and 2.

We used the Viterbi algorithm to estimate the most likely sequence of states, for each fitted model. Then, we calculated the proportion of states that were correctly estimated, and we averaged over the 50 replications (Table 4.1).

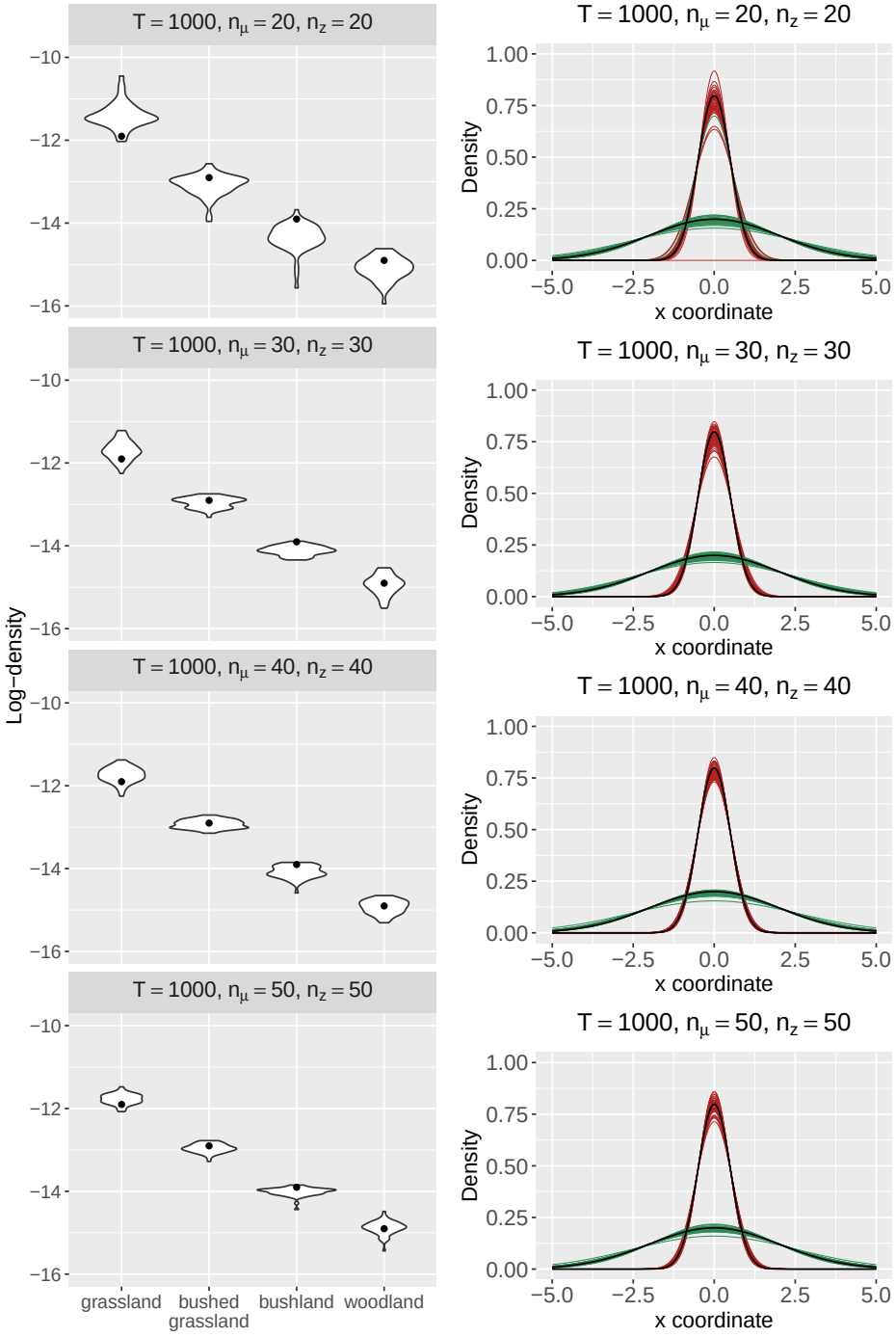


Fig. 4.8 Results of the simulation study for the state-switching normal kernel model, for Monte Carlo samples of different sizes. Left column: Estimated utilisation values of the four habitat types, on the log scale. The violins show the 50 sets of estimates, and the dots are the true values. Right column: Estimated normal half-step densities; the coloured lines show the 50 estimated densities, and the black lines are the true densities used in the simulations.

T	1000	1000	1000	1000	100	200	500
n_{MC}	20	30	40	50	40	40	40
Mean proportion	92.6%	94.1%	94.5%	94.4%	94.9%	94.7%	94.1%

Table 4.1 Mean proportions of states that were correctly estimated by the Viterbi algorithm, in the state-switching normal kernel model. Several simulations were conducted, with different sizes of Monte Carlo samples ($n_\mu = n_z = n_{\text{MC}}$) and simulated tracks of different lengths (T). The mean proportions are each calculated from 50 replications.

In all scenarios, a large majority of steps were classified in the correct state. The mean proportion of correctly classified states in simulations with $n_\mu = n_z = 20$ is pulled down by the experiment mentioned above, in which all observations were classified in state 2.

The main issue we encountered in simulations with small Monte Carlo samples ($n_\mu = n_r = 20$) was numerical instability. The optimiser often diverged to extreme parameter values, for which the likelihood function could not be evaluated, and crashed. In such cases, we started the optimisation again with different Monte Carlo samples. This problem did not arise in any of the simulations with $n_\mu = n_z = 30$ or more.

We investigated the effect of the number of observations (i.e. the length T of the track) on the estimation of the model parameters and of the underlying states. We simulated tracks of different lengths, $T = 100, 200, 500$, from the state-switching normal kernel model described above. We then fitted the model to each track with Monte Carlo samples of size $n_\mu = n_z = 40$. We repeated the procedure 50 times for each value of T, and the results are presented in Figure 4.9. We estimated the Viterbi state sequence for each fitted model, and calculated the average proportion of correctly decoded states for each simulation scenario (Table 4.1).

As in the single-state case, the precision of the habitat selection estimates increased for longer tracks. There was a lot of variability in the estimated half-step densities for the shortest tracks ($T = 500$), but the two states always captured clearly distinct movement speeds (slower movement in state 1 and faster movement in state 2). The half-step densities were well estimated for longer tracks. The Viterbi algorithm estimated the state sequences well in all scenarios (Table 4.1).

4.7.5 Random availability radius model

We ran a simulation study to investigate the performance of the estimation for the availability radius model with random radius parameter. We sampled the availability radius parameter r

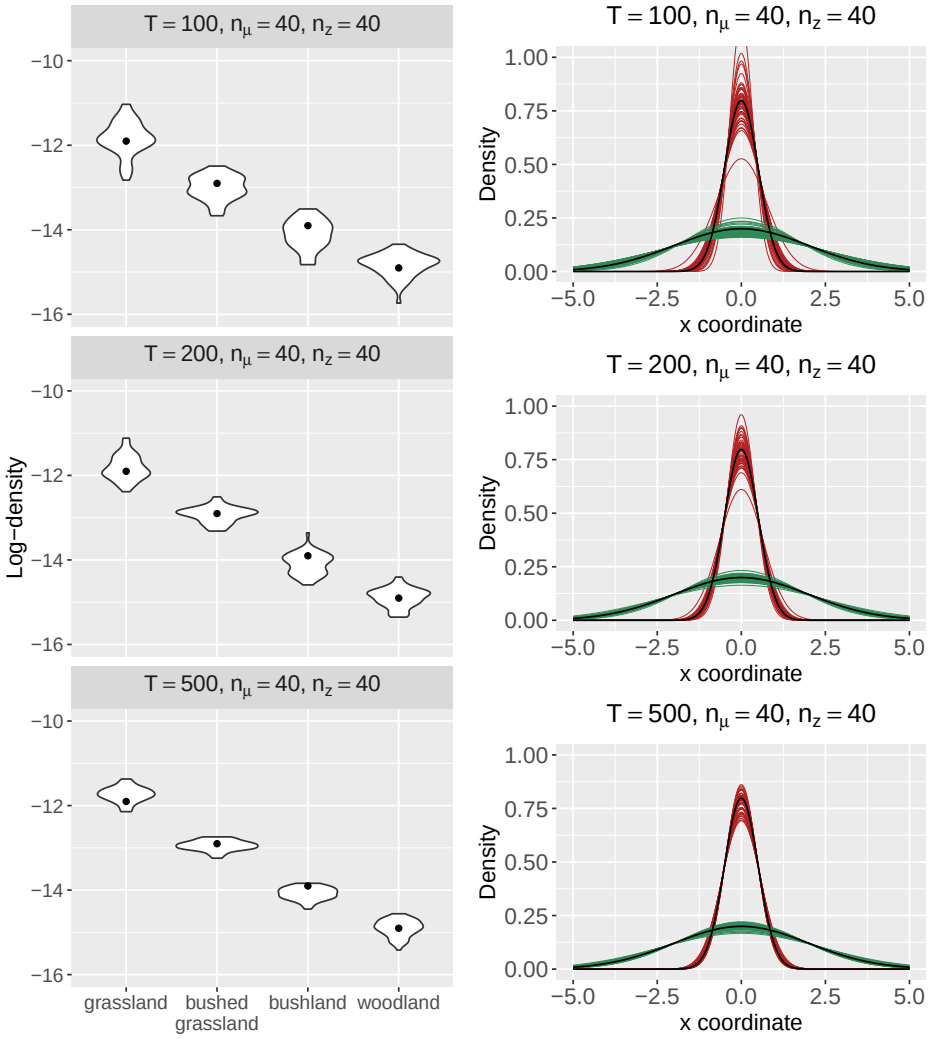


Fig. 4.9 Results of the simulation study for the state-switching normal kernel model, for simulated tracks of different lengths. Left column: Estimated utilisation values of the four habitat types, on the log scale. The violins show the 50 sets of estimates, and the dots are the true values. Right column: Estimated normal half-step densities; the coloured lines show the 50 estimated densities, and the black lines are the true densities used in the simulations.

from a gamma distribution with shape $\alpha = 0.7$ and rate $\rho = 3$. We chose these parameters to obtain speeds similar to those observed in the zebra data set of Section 4.8.

We simulated a track of length $T = 1000$, and fitted the model with Monte Carlo samples of different sizes. There were five parameters to estimate: three habitat selection parameters ($\beta_1, \beta_2, \beta_3$), and the shape α and rate ρ of the gamma distribution of the availability radius.

In the availability radius model with random radius parameter, the likelihood is written with three nested integrals (Equation 4.8). The approximate likelihood thus requires three nested sums, i.e. three Monte Carlo samples (Equation 4.9). We considered four cases, with Monte Carlo samples of sizes: (1) $n_r = 10$, $n_\mu = n_z = 20$, (2) $n_r = 15$, $n_\mu = n_z = 30$, (3) $n_r = 20$, $n_\mu = n_z = 40$, and (4) $n_r = 25$, $n_\mu = n_z = 50$. We used smaller values for n_r than for n_μ and n_z, because the integral over the radius r is one-dimensional, whereas the integrals over the intermediate point $\boldsymbol{\mu}$ and over the end point $\boldsymbol{z}$ are two-dimensional.

For each case, we ran 50 simulations. Figure 4.10 shows the estimated utilisation values, and the gamma distributions estimated for the availability radius. For two of the simulations with small Monte Carlo samples ($n_r = 10$, $n_\mu = n_z = 20$), the optimiser failed to converge, and those runs are excluded from the figure.

The utilisation distribution was well captured in all experiments, and the accuracy and precision of the utilisation estimates increased slightly for larger Monte Carlo samples. The gamma distribution of the availability radius was well estimated in almost all simulations. For smaller Monte Carlo samples, the mean availability radius was overestimated in a few cases.

We repeated the simulation experiments with tracks of different lengths. We simulated tracks of lengths $T = 100, 200, 500$ from the availability radius model with gamma-distributed radius parameter. We then estimated all model parameters with Monte Carlo samples of sizes $n_r = 20$ and $n_\mu = n_z = 40$. The estimated habitat selection parameters, and the estimated gamma densities of the radius, are shown in Figure 4.11.

The general shape of the utilisation distribution was roughly captured in all simulations, although the variability in the estimates was large for short tracks. The precision of the utilisation estimates and of the estimates of the availability radius distribution increased with the length of the tracks.

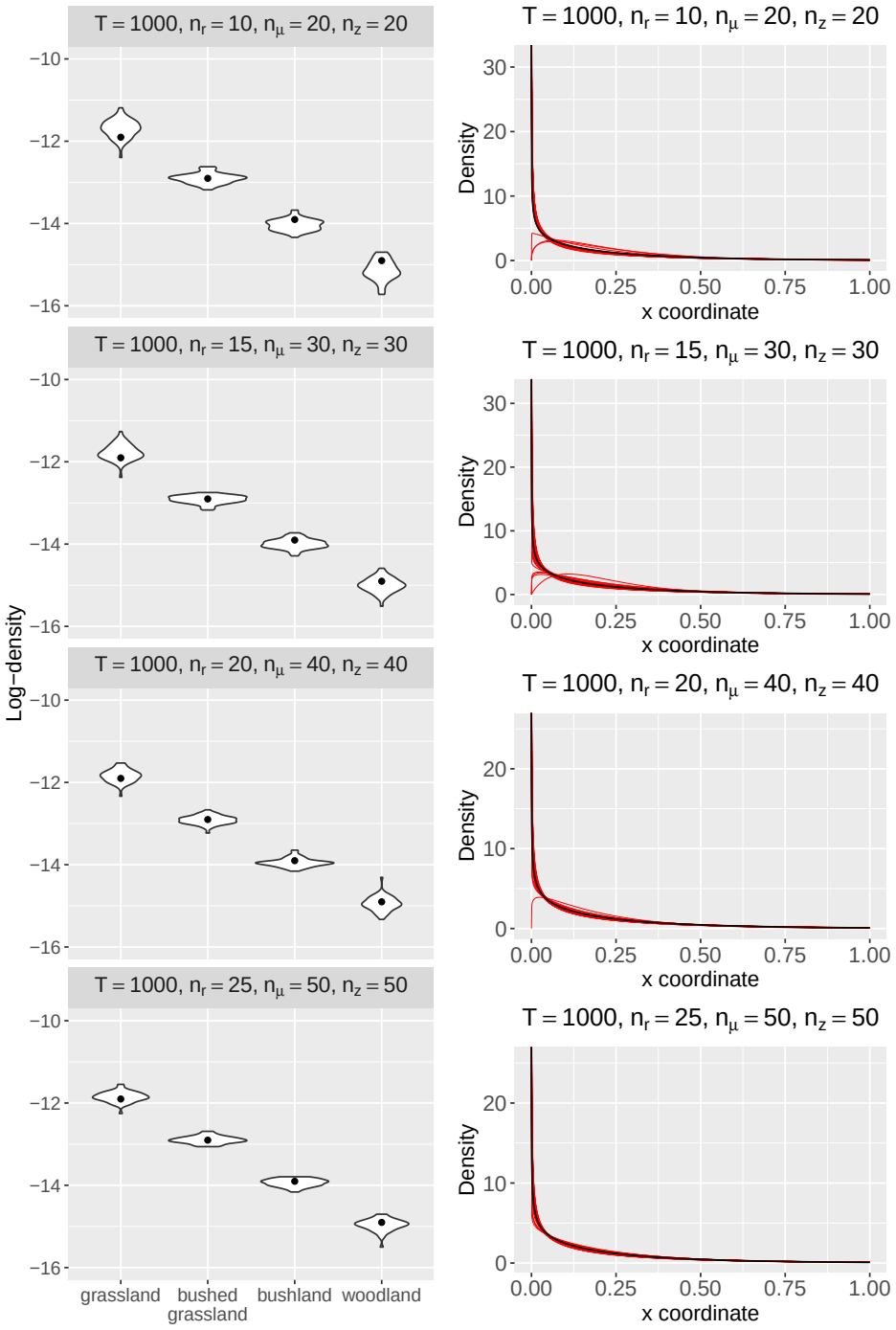


Fig. 4.10 Results of the simulation study for the availability radius model, for Monte Carlo samples of different sizes. Left column: Estimated utilisation values of the four habitat types, on the log scale. The violins show the 50 sets of estimates, and the dots are the true values. Right column: Estimated gamma densities; the red lines show the 50 estimated densities, and the black lines are the true densities used in the simulations.

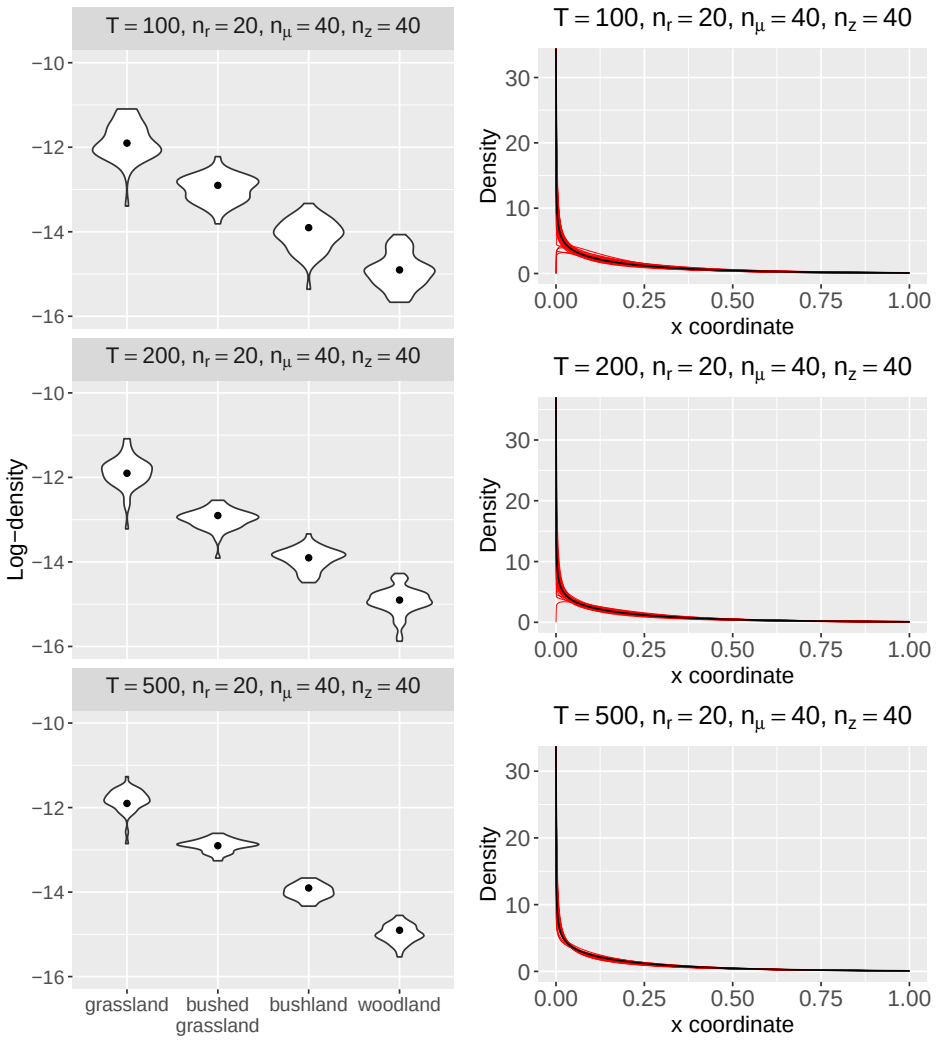


Fig. 4.11 Results of the simulation study for the availability radius model, for simulated tracks of different lengths. Left column: Estimated utilisation values of the four habitat types, on the log scale. The violins show the 50 sets of estimates, and the dots are the true values. Right column: Estimated gamma densities; the red lines show the 50 estimated densities, and the black lines are the true densities used in the simulations.

4.7.6 Investigation of bias for correlated movement

In the local Gibbs movement model, the half-step density ϕ is required to be symmetric (Section 3.4.1). For this reason, the model does not capture persistence in the movement direction, because all directions are equally likely in the absence of environmental effects. As mentioned in Section 1.2, modern tracking data can be collected at high temporal resolutions, and they often display some movement persistence. Here, we investigate the effect of this persistence on the accuracy of the estimates obtained from the local Gibbs model.

We considered a standard step selection function model, similar to those described by Fortin et al. (2005) and Forester et al. (2009), as a model that combines movement persistence and habitat selection. We simulated a track from the SSF model, on the same vegetation map used in the previous simulations. For the habitat-independent movement kernel of the SSF model, we used a distribution of step lengths and a distribution of turning angles. The shape of the distribution of turning angles determines the strength of the directional persistence. We considered a gamma distribution with shape 0.5 and rate 3 for the step lengths, to mimic the distribution of step lengths found in the zebra data analysed in Section 4.8. We used a von Mises distribution with mean zero for the turning angles. We chose several different values for the concentration parameter of the von Mises distribution, $\kappa = 0.5, 2, 5$, to induce different levels of persistence. (The larger the concentration, the more directed the movement.) The value $\kappa = 0.5$ was approximately the concentration of the distribution of turning angles in the zebra track of Section 4.8.

As mentioned in Section 3.1, the parameters of the SSF do not directly describe the utilisation distribution. To estimate the utilisation distribution of the SSF model, we generated a long trajectory of 10^5 locations, as suggested by Avgar et al. (2016) and Signer et al. (2017). We fitted a RSF to the simulated locations, to estimate the steady-state habitat selection parameters (to compare with the local Gibbs estimates).

We then divided the simulated data set into 50 tracks of 2000 observations each. For each subtrack, we fitted the local Gibbs model with normal half-step density, to check whether it was able to recover the habitat selection parameters. The estimated utilisation values of the four habitat types are shown for the 50 experiments in Figure 4.12.

In the scenario with concentration $\kappa = 0.5$, the habitat selection parameters were accurately estimated in all experiments. This confirms that, for moderate directional persistence, the estimates of the normal kernel model are not biased. However, when the concentration was larger (i.e. when the persistence was stronger), the estimates appeared to be biased. In Figure 4.12, the results for $\kappa = 2$ and $\kappa = 5$ indicate that the preference for grasslands

Fig. 4.12 Results of the simulation study for the correlated movement model, for different distributions of turning angles. The turning angles follow a von Mises distribution with mean zero, and with concentration κ (top: $\kappa = 0.5$, middle: $\kappa = 2$, bottom: $\kappa = 5$). The violin plots show the 50 sets of estimated utilisation values of the four habitat types, on the log scale. The dots are the "true" values, evaluated with a standard RSF from the 10^5 simulated locations.

and bushed grasslands was overestimated, and the preference for bushland and woodland was underestimated. This suggests that the local Gibbs model may produce biased habitat selection estimates when the movement is strongly autocorrelated, because this persistence cannot be accounted for by the formulation.

4.8 Zebra case study

We illustrate the use of the method described in this chapter to fit the local Gibbs model to animal movement and habitat data.

4.8.1 Data set

We considered a GPS track of a plains zebra (*Equus quagga*), with locations collected every 30 minutes from January to May 2014 in Hwange National Park (Zimbabwe). The data consist of 7246 locations observed at regular time intervals. Of the 7246 observations, 125 were missing at random. It is straightforward to deal with missing locations in this framework: the corresponding missing steps do not contribute to the likelihood.

We used a vegetation map of Hwange National Park, at a resolution of 30m in each dimension, to define a categorical habitat covariate. The vegetation covariate has four categories: grassland, bushed grassland, bushland, and woodland. The vegetation is increasingly dense in these four types of vegetation cover, and the aim of the study was to estimate the preference of the zebra for the different habitats. We truncated the vegetation raster to the study region, with a 5km margin around the zebra track. Over the study region, the four vegetation types were represented in very varying proportions: 1.3% of grassland, 17.4% of bushed grassland, 49.7% of bushland, and 31.6% of woodland.

Figure 4.13 shows two plots of the vegetation map, with and without the zebra movement track. Although the grassland habitats only cover a small proportion of the area, it is where the zebra spent most of its time. About 31% of the observed locations are in the grassland habitat, and 57% in the bushed grassland habitat.

For the analysis, we transformed the vegetation map into three dummy covariates, similarly to what we did in the simulation study of Section 4.7. The three dummy covariates corresponded to grassland (c_1), bushed grassland (c_2), and bushland (c_3), respectively. The fourth habitat type, woodland, was used as reference category.

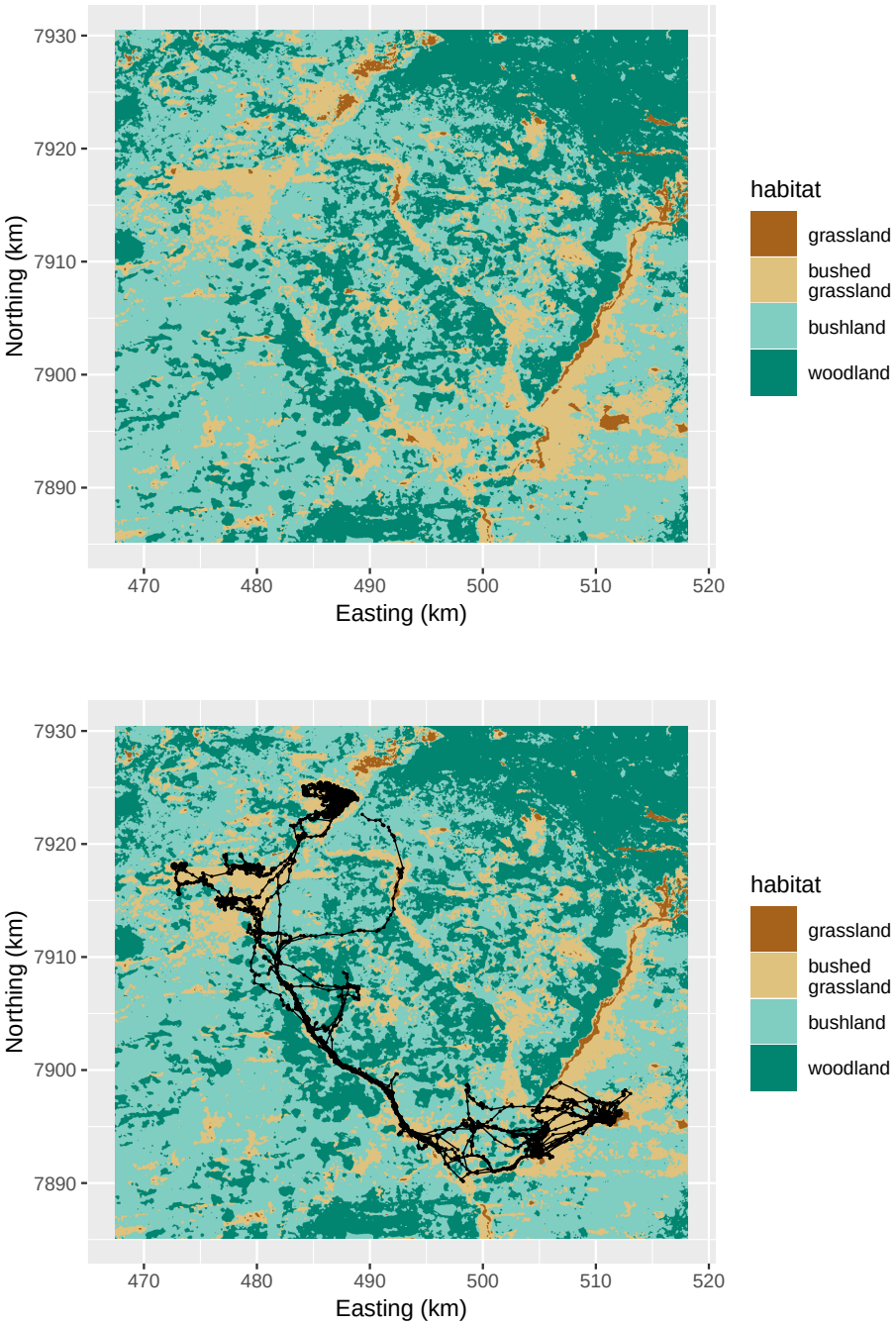


Fig. 4.13 Map of vegetation types in the study region (top), and zebra movement track (bottom).

4.8.2 Normal kernel model

We fitted the normal kernel model to the zebra track, to estimate four unknown parameters: the scale parameter σ, and the three habitat selection parameters $(\beta_1, \beta_2, \beta_3)$. We generated Monte Carlo samples of size $n_\mu = n_z = 40$, based on the results of the simulations of Section 4.7.

We used the numerical optimiser `optim` in R to maximise the approximate log-likelihood function of the model. Gradient-based optimisation techniques are susceptible to remain stuck in a local maximum of the objective function, and to fail to converge to its global maximum. The choice of starting values, which determine where the optimiser starts its search in parameter space, is important. Poorly-chosen starting values, e.g. very distant from the maximum likelihood estimates, are more likely to lead to numerical issues. To circumvent this problem, we ran the optimisation 50 times, starting from different sets of starting values (always with the same Monte Carlo samples). We selected the values randomly, each from a uniform distribution over an interval of plausible parameter values. We then selected the best of the 50 fitted models, i.e. the model with the largest likelihood. Each model fit took about 8 min on a 2GHz i5 CPU.

For the best fitting model, we calculated 95% confidence intervals of the parameter estimates, as follows. We computed the Hessian matrix of the approximate log-likelihood at the parameter estimates, using the function `hessian` of the R package numDeriv (Gilbert and Varadhan, 2016). Like for the point estimates, the accuracy of the estimated Hessian depends on the size of the Monte Carlo samples passed to the function. We tried several Monte Carlo sample sizes, and found that very little change occurred beyond $n_\mu = n_z = 50$. We then derived an estimate of the covariance matrix of the estimators, as the inverse of the Hessian matrix. Finally, we calculated standard errors as the square roots of the diagonal elements of the covariance matrix. We used assumptions of asymptotic normality to obtain 95% confidence intervals for the habitat selection parameters.

Point estimates and 95% confidence intervals of the habitat selection parameters are given in Table 4.2. We also present the corresponding (non-normalized) utilisation values of the four habitat types, which give a measure of the preference for each habitat, compared with the reference woodland habitat. Figure 4.14 shows a map of the estimated utilisation distribution of the zebra.

The estimates of the habitat selection parameters, and the estimated utilisation values of each type of vegetation, indicate that the zebra strongly selected the grassland habitat and, to a lesser extent, the bushed grassland habitat. Indeed, the utilisation value of the grassland

Habitat type	$\hat{\beta}_i$	95% CI	$\hat{\pi}_i$
Grassland	2.72	(2.53,2.92)	15.2
Bushed grassland	1.41	(1.23,1.59)	4.1
Bushland	0	(-0.18,0.18)	1
Woodland (reference)	0		1

Table 4.2 Habitat selection parameter estimates $\hat{\beta}_i$ for the normal kernel model, 95% confidence intervals, and corresponding (non-normalized) utilisation value $\hat{\pi}_i$ for each habitat type. Woodland is the reference category, and the corresponding parameter is not estimated but fixed to 0.

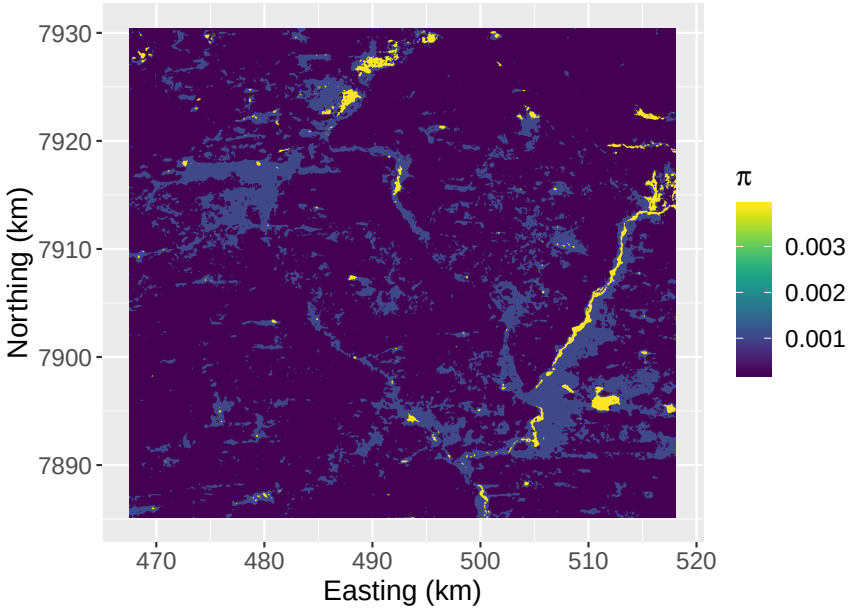


Fig. 4.14 Estimated utilisation distribution in the zebra case study, with the normal kernel model.

is about 15 times larger than the utilisation values of the bushland and woodland. This is consistent with our knowledge of zebra behaviour: they tend to favour open habitats, where they forage, and where the risk of predation is lower. It also confirms the results found by Courbin et al. (2016) in a similar study (although with a different data set).

The standard deviation of the half-step density was estimated to $\hat{\sigma} = 0.20$. As detailed in Section 3.5.1, the habitat-independent distribution of step lengths under the normal kernel model is a Rayleigh distribution with scale parameter $\lambda = \sqrt{2}\sigma$. In this application, we have $\hat{\lambda} = \sqrt{2}\hat{\sigma} = 0.28$. The estimated mean of the Rayleigh distribution, i.e. the habitat-independent mean step length, is $\sqrt{\pi/2}\hat{\lambda} = \sqrt{\pi}\hat{\sigma} = 0.35$km over 30-minute intervals.

To assess this movement model, we simulated 10^4 locations from the fitted model, on the same habitat map as the observations. We compared the distribution of step lengths observed in the zebra data set to the distribution of simulated step lengths (Figure 4.15). There is a clear discrepancy between the two distributions: the model fails to capture very short and very long step lengths, and overestimates the density of intermediate step lengths. The empirical distribution of step lengths has a mode at zero, and a long tail, which cannot be appropriately modelled by this formulation. We then considered the availability radius model with random radius parameter, for more flexibility.

4.8.3 Availability radius model

We fitted the local Gibbs model with random availability radius to the same track. We modelled the availability radius with a gamma distribution, and estimated its shape and rate parameters. We used Monte Carlo samples of size $n_r = 20$ and $n_\mu = n_z = 40$, following the simulation study of Section 4.7.5. Like in Section 4.8.2, we ran the numerical optimisation 50 times with random initial parameter values, and kept the model with the largest likelihood, to avoid numerical convergence issues. Each model fit took about 1.5 hour on a 2GHz i5 CPU. We used a method similar to the one described in the case of the normal kernel model to obtain standard errors for the estimated parameters.

The estimates of the habitat selection parameters, and of the utilisation values of all vegetation types, are given in Table 4.3. The parameter values are similar to those obtained with the normal kernel model, and the results confirm that the selection is stronger for open habitats (i.e. grassland and bushed grassland). The estimated shape of the gamma distribution of the availability radius was $\hat{\alpha} = 0.77$, and the rate was $\hat{\rho} = 3.56$. The estimated gamma distribution of the availability radius therefore had mean $\hat{E}(r) = \hat{\alpha}/\hat{\rho} = 0.22$km, and 95th percentile $\hat{P}_{0.95} = 0.71$km.

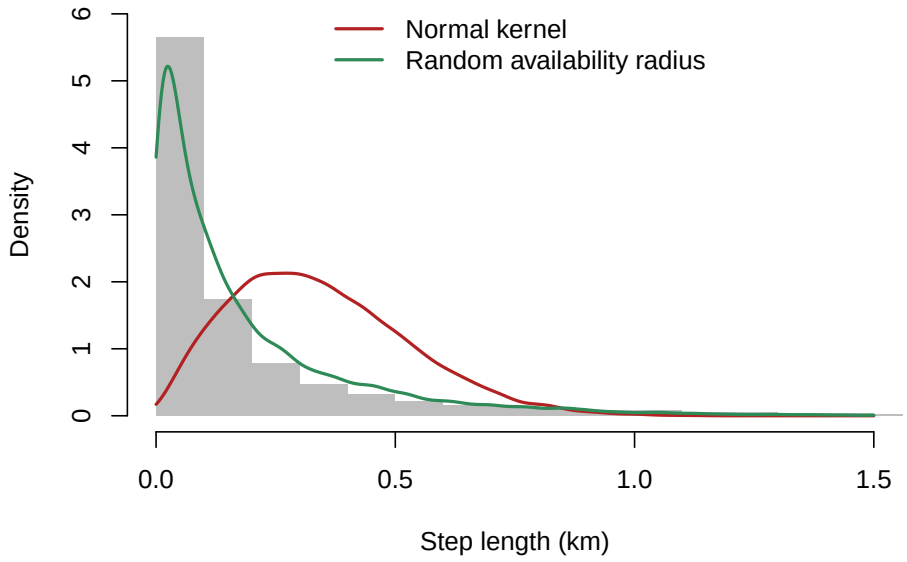


Fig. 4.15 Histogram of the observed step lengths in the zebra data set. The lines show the densities of simulated step lengths, obtained from two fitted models: the local Gibbs model with normal half-step density, and the local Gibbs model with gamma-distributed availability radius. We truncated the x-axis to $[0, 1.5]$ for better visualisation, but the maximum observed step length is around 3km.

Habitat type	$\hat{\beta}_i$	95% CI	$\hat{\pi}_i$
Grassland	2.38	(2.12,2.64)	10.8
Bushed grassland	1.37	(1.13,1.60)	3.9
Bushland	0.25	(0.02,0.48)	1.3
Woodland (reference)	0		1

Table 4.3 Habitat selection parameter estimates for the availability radius model, 95% confidence intervals, and corresponding (non-normalized) utilisation value for each habitat type. Woodland is the reference category, and the corresponding parameter is not estimated but fixed to 0.

To assess the random availability radius model, we cannot directly measure the goodness-of-fit for the estimated distribution of the availability radius, because its true value is never observed. Instead, like we did for the normal kernel model in Section 4.8.2, we simulated a track of length 10^4 from the fitted model, on the same habitat map. We compared the distributions of observed and simulated step lengths (Figure 4.15). The distribution of the simulated steps resembles that of the observed steps much more closely than with the normal kernel model. This indicates that the model was able to capture the speed of the zebra's movement. This is remarkable, as the step lengths or the speeds are never directly modelled: instead, we estimated the distribution of the unobserved radius of the relocation region.

There is a trade-off between realism of the movement model and computational speed: the random availability radius model was 15-20 times slower than the normal kernel model in this analysis, due to the additional nested integral in its likelihood (Equation 4.9). Here, the habitat selection estimates were very similar using both models. This suggests that the simpler one (normal kernel model) is sufficient to capture the RSF, even if the movement component is not flexible enough to capture the zebra's step lengths. However, we could not have known this before fitting the random availability radius model and, generally, model checking methods should be used to verify that features of the movement are appropriately captured by the model. Although the model does not capture the directional persistence that is found in the zebra track, the results of the simulations of Section 4.7.6 suggest that the habitat selection estimates are still accurate if the autocorrelation is moderate.

Standard model selection methods can also be used to choose the model formulation. In the zebra analysis, the Akaike Information Criterion (AIC) for the normal kernel model was -715.6, and the AIC for the availability radius model was -23401.2. This criterion thus strongly favours the latter, more complex, model formulation.

4.9 Discussion

We showed how a new class of step selection models, based on the simulation rules of MCMC algorithms, can be used to estimate an animal's habitat selection and movement characteristics. We provided a natural framework to model animal movement and space use, in which short-term step selection rules give rise to the long-term utilisation distribution. This approach connects standard RSF and SSF models, as the equilibrium distribution of the movement model is guaranteed to be proportional to the underlying resource selection function. We described maximum likelihood estimation for the local Gibbs sampler, a flexible

family of MCMC algorithms which can be used to model animal movement. Parameters of movement and habitat selection can be estimated jointly.

MCMC step selection models describe the animal's movement and space use within a consistent formulation. They can therefore be used to combine inference from telemetry and survey data, and to reconcile individual-based and population-based approaches. Survey data include direct observations of animals from a population, e.g. from capture-recapture or line transect surveys. Because they do not display the same spatio-temporal autocorrelation as telemetry data, survey data are often treated as independent observations from the utilisation distribution of the population. They can be integrated into a MCMC step selection analysis to provide additional information about the long-term habitat selection and space use. For example, consider that we have a sample of (dependent) telemetry observations $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$ from one or several individuals of a population, and a sample of independent survey locations $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$ from that population. The likelihood of the telemetry sample under the MCMC movement model, L_{mov}, is given in Section 4.1 as the product of the transition kernel of the model for all observed steps. The likelihood of the survey sample, L_{ind}, can be obtained as a function of the habitat selection parameters from the standard RSF methodology. For example, we can use logistic regression with a use-availability design, or a Poisson GLM on the location counts (Boyce and McDonald, 1999; Aarts et al., 2012). Then, the combined likelihood of the two data sets is simply the product of L_{mov} and L_{ind}. The resulting estimates of habitat selection are based on both sources of data.

We calculated the likelihood of an observed trajectory $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ as the product of the likelihoods of each step from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t+1}$, for $t = 1, \ldots, T-1$. In this approach, the first observation, $\boldsymbol{x}_1$, is treated as deterministic, and does not contribute to the inference. In some applications, we could however try to learn from the first observation. Blackwell (2003) suggested that the first observation of a telemetry data set could sometimes be assumed to arise from the equilibrium distribution of the movement model. This may in particular be a reasonable assumption if the first few days after tagging were discarded—a procedure analogous to the idea of burn-in in MCMC sampling. In the local Gibbs model, we could therefore consider that the first observed location is a sample from the utilisation distribution, and include it as such in the likelihood calculations. The improvement would be minimal in the analysis of a single track, but it could be valuable when many tracks are processed simultaneously, i.e. when many first locations are available.

Although the simulations and the case study were based on categorical habitat data, the method would also be applicable to continuous environmental covariates, such as elevation or distance to water. One could for example model territoriality or attraction towards a point in

space, with the inclusion of the distance to that point as a covariate. This idea also has some mathematical appeal. In the presentation of the local Gibbs model, we considered that the utilisation distribution was defined on a bounded set $\Omega \subset \mathbb{R}^2$. This is convenient to ensure that the RSF is integrable, and that the normalized RSF (i.e. the utilisation distribution) is well-defined. It is a reasonable assumption, because analyses of telemetry and environmental data are always restricted to a bounded geographical region. This can however be relaxed, and the utilisation distribution can be defined over $\Omega = \mathbb{R}^2$, if the improper integral $\int_{\boldsymbol{z} \in \mathbb{R}^2} w(\boldsymbol{c}(\boldsymbol{z}))d\boldsymbol{z}$ converges. In practice, one way to achieve this result is to include the distance to a point as a covariate, with a negative selection coefficient. The negative selection coefficient ensures that the RSF decreases as the distance increases, and that it is integrable over $\mathbb{R}^2$.

An important feature of the local Gibbs model is that the size of the region of availability does not need to be defined a priori. In habitat selection analyses based on use-availability designs, the choice of the spatial extent of the availability region is challenging, and can lead to biased selection estimates (Beyer et al., 2010; Northrup et al., 2013). Instead, we estimate it from the observed tracking data, with a movement model based on a symmetric half-step density. The scale of availability is for example measured by the variance parameter of the normal kernel model, and by the radius parameter in the availability radius model. One limitation of this method is that the animal's movement and perception are modelled jointly. The half-step density of the algorithm describes both the size of the region that the animal can perceive and the distance that it is susceptible to cover over one time interval. This is a strong assumption, that is made in most step selection models (Fortin et al., 2005; Forester et al., 2009), in which habitat selection is considered to take place at the scale of the movement kernel. Recently, Avgar et al. (2015) and Bastille-Rousseau et al. (2018) have proposed models to estimate the movement process and the perception on separate scales. Additional work is required to allow this flexibility within the local Gibbs model.

In Section 3.3, we suggested that rejection-based MCMC algorithms, such as Metropolis-Hastings, are not suitable to construct realistic MCMC step selection models. Here, we would like to give this idea more detailed consideration. In rejection-based algorithms, a proposed step is sampled at each iteration from some distribution, and it is either accepted or rejected, with some probability. The acceptance-rejection step is required to ensure that such samplers satisfy the detailed balance condition. In the framework described in Chapter 3, the proposal distribution of the algorithm could be considered as a model for availability, and the acceptance-rejection step could describe selection. The step is accepted, i.e. selected, if the habitat features of the end point are suitable, for example. If the step is not accepted, then it is assumed that the animal stays in its current location over one time interval. This is not a

problem in itself, as it can happen that an animal stays immobile over more than one time interval, resulting in a step of length zero. However, this phenomenon cannot generally be attributed to habitat selection. The animal may be immobile because it is resting, for example, and not as a result of "rejecting" a candidate relocation. There are also many telemetry data sets that do not include any steps of length zero. Such data sets are very improbable outputs for a step selection model based on a rejection-based MCMC algorithm: they can only arise if every proposed step was accepted. In preliminary experiments that are not presented in this thesis, we implemented a MCMC step selection model based on a Metropolis sampler with normal proposal distribution. We used it to estimate the habitat selection parameters and the variance parameter of the proposal distribution, from tracking data that did not include steps of length zero. The estimated selection parameters were all very close to zero, i.e. the estimated utilisation distribution was flat. This is because the most plausible explanation for a 100% acceptance rate in a Metropolis algorithm is that the target distribution is flat. We therefore believe that rejection-free algorithms (such as the local Gibbs sampler) should generally be preferred in the context of the method developed here, as they are more widely applicable to the analysis of animal movement data. Moreover, steps of length zero are not prohibited in rejection-free algorithms. In particular, as mentioned in Section 3.5.3.2, zero inflation can be used in the local Gibbs model to allow for steps of length zero with positive probability.

In the case study of Section 4.8, we fitted two different local Gibbs models to a zebra movement track, and compared them using simulations. We found that simulated step lengths from the local Gibbs model with random availability radius matched the observed step lengths more closely than those simulated from the local Gibbs model with normal half-step density. However, the computational effort was greatly increased in the former, because of the need to integrate over the random availability radius. It may be possible to find a local Gibbs formulation that combines the computational speed of the normal kernel model and the flexibility of the random availability radius model. For example, we could define the half-step density as the combination of a uniform distribution of angles and a given distribution for the distance to the origin, as suggested in Section 3.5.3.1. The uniform angles ensure that the half-step density is symmetric, and the shape of the distance distribution determines the habitat-independent movement model. It may be possible to achieve a distribution of step lengths with a mode close to zero, as in the zebra data set, with an exponential or Weibull distribution of distances.

In Section 1.1, we mentioned that there exist other types of animal movement data, in particular spatially-referenced mark recapture data (Ovaskainen, 2004). It is not clear that

the methods described in this chapter could be applied to such data. Indeed, mark-recapture data are sparse, and irregular in time. We could possibly apply the normal kernel model for irregular intervals, described in Section 3.6.3. For very sparse locations, however, the model would not capture the short-term movement of the animal, but rather a longer-term rate of diffusion. In the limit, if the time intervals between observations are very long, the scale of the movement kernel will cover the whole study region, and the observations will be treated as a random sample from the target utilisation distribution. Diffusion models such as those described by Ovaskainen and Cornell (2003) and Ovaskainen (2004) may be better suited to mark-recapture movement data.

Chapter 5

The Langevin movement model

In Chapters 3 and 4, we developed a class of models of animal movement and habitat selection, based on Markov chain Monte Carlo algorithms. That approach was formulated in discrete time, although we proposed an extension to analyse location data collected at irregular time intervals. In this chapter, we develop a continuous-time model of animal movement and habitat selection, to tackle this problem. We present the model in Section 5.2. In Sections 5.3 and 5.4, we develop a method of inference to estimate parameters of movement and habitat selection, from telemetry and environmental data. We investigate the performance of the method in a simulation study in Section 5.5, and we apply it to the analysis of three tracks of Steller sea lions in Section 5.6.

The material presented in this chapter was described in the preprint Michelot et al. (2018b). It is the result of a collaboration with Pierre Gloaguen and Marie-Pierre Étienne. Pierre Gloaguen originally suggested the use of the Langevin diffusion to model animal movement, he contributed to the statistical developments presented in Sections 5.3 and 5.4, and he contributed to the implementation of the models in R. Marie-Pierre Étienne contributed to the statistical developments presented in Sections 5.3 and 5.4.

5.1 Introduction

In Chapter 3, we presented a model of animal movement with an explicit stationary distribution, using the properties of MCMC algorithms. We used it to describe habitat selection as a scale-independent process, with the same parameters affecting the short-term movement and the long-term space use. The movement component of the MCMC step selection model, however, is tied to the scale of observation of the tracking data. Indeed, MCMC samplers are

defined as discrete-time Markov processes, where one time step corresponds to one iteration of the algorithm. In the framework described in Chapter 4, the movement parameters of MCMC step selection models are therefore scale-dependent. In some special cases, it is possible to make assumptions about the scaling properties of the parameters, as suggested for the variance of the normal kernel model (Section 3.6.3). However, in general, inferences made about the extent of the animal's perceptual range, and about its speed of movement, will depend on the time interval of the data. This is a general limitation of discrete-time models of animal movement, which can make their interpretation difficult (because it is tied to the arbitrary time scale of the observations). It also has practical implications, as mentioned in Section 1.2. In particular, the parameters of the model are defined over a fixed time interval, so discrete-time models cannot be used to analyse telemetry data collected at irregular intervals. Similarly, the movement parameters of the MCMC step selection model cannot be compared across studies if the data sets do not all have the same sampling frequency.

Spatially-structured diffusion models, described in Section 1.3.3, model movement and habitat preference in a continuous-time framework (Ovaskainen, 2004, 2008). Those methods focus on the analysis of animal movement in an environment composed of discrete habitat patches. In this chapter, we describe the movement of an animal, and its long-term use of space, in response to continuous environmental covariates.

Following the basic idea of MCMC movement models, we use a stochastic process for which both the short-term movement dynamics and the long-term stationary distribution are known. However, unlike in MCMC movement models, we now consider a *continuous-time* stochastic process, called the Langevin diffusion process, to model the animal's location.

5.2 Model formulation

5.2.1 The Langevin movement model

We consider a continuously differentiable distribution $\pi : \mathbb{R}^2 \to \mathbb{R}$, and we denote by $(\boldsymbol{X}_t)_{t\geq 0}$ the continuous-time location process of the animal. We model it with the Langevin diffusion on π, defined by Roberts et al. (1996) as the solution to the stochastic differential equation

$$d\boldsymbol{X}_t = \frac{1}{2}\nabla \log \pi(\boldsymbol{X}_t)dt + d\boldsymbol{W}_t, \tag{5.1}$$

where $\nabla = (\partial/\partial x, \partial/\partial y)$ is the spatial gradient operator, and $(\boldsymbol{W}_t)$ is a standard Wiener process. The process is also defined by an initial condition $\boldsymbol{X}_0 \sim p_0(\cdot)$. The initial location

is typically assumed to be deterministic, i.e. $X_0 = x_0$. In the following, we will make this assumption, and omit the initial condition. The drift term of Equation 5.1, $\nabla \log \pi(\boldsymbol{X}_t)/2$, gives the expected displacement. Here, the drift indicates that the process will tend to move towards the direction of increasing $\log \pi$, i.e. towards increasing π (because the logarithm is monotonic). The Wiener process adds stochasticity to the process.

This is conceptually similar to the movement models based on potential functions (Brillinger et al., 2001). The idea of a potential function is often used in physics to describe the spatial field of forces applied to an object. In the context of animal movement, the potential function is defined as a two-dimensional surface. It represents the "forces" of attraction and repulsion that drive the animal's movement. The animal will tend to be attracted by areas where the surface is low, and will tend to avoid areas of high potential. Preisler et al. (2004) suggested that the potential function could be formulated as the sum of terms of attraction and repulsion from habitat features, e.g. roads or food resources. In that framework, the animal's movement is thus modelled in response to the habitat characteristics of its surroundings. Similarly, the movement of the Langevin diffusion process is determined by the gradient of the function π at its current location. Note that the Langevin diffusion process reduces to an Ornstein-Uhlenbeck process when π is a normal distribution. (We defined the Ornstein-Uhlenbeck process in Section 2.4, and we will discuss its applications in movement ecology in Chapter 6.)

Under some mild regularity conditions on π, Roberts et al. (1996) showed that π is the stationary distribution of the process $(\boldsymbol{X}_t)$ defined in Equation 5.1. The Langevin diffusion process can therefore be viewed as a model of movement and space use, similarly to the MCMC movement model of Chapters 3 and 4. The (continuous-time) dynamics of the movement process are described in Equation 5.1, and π is the long-term utilisation distribution of the animal. This model satisfies the intuitive notion that long-term space use by animals is the consequence of their short-term movement decisions.

The process given in Equation 5.1 is not a flexible movement model, because the speed of the process is only determined by the gradient of the underlying distribution π. In particular, it does not allow for the possibility of two animals moving at different speeds on the same utilisation distribution. Therefore, we introduce a speed parameter, γ^2, and we consider the location process defined as the solution of

$$d\boldsymbol{X}_t = \frac{\gamma^2}{2} \nabla \log \pi(\boldsymbol{X}_t) dt + \gamma d\boldsymbol{W}_t. \tag{5.2}$$

This is inspired by a similar modification proposed by Roberts and Rosenthal (1998). This process is "speeded up" by a factor γ^2, compared with the process defined in Equation 5.1. The drift term is multiplied by γ^2, but the diffusion term is multiplied by γ rather than γ^2, because increments of the Brownian motion scale linearly with the *square root* of time (Einstein, 1905). The introduction of γ^2 only changes the speed of the process, but it does not change its stationary properties. Indeed, Xifara et al. (2014) described a similar extension of the Langevin diffusion, in the context of the development of gradient-based MCMC algorithms, and they showed that the process $(\boldsymbol{X}_t)$ defined in Equation 5.2 also has π as its stationary distribution. We use it as the basis for a model of movement and space use, that we call the Langevin movement model.

5.2.2 Modelling habitat selection

To link the movement of the Langevin process to environmental drivers, we model the utilisation distribution π with a resource selection function, like we did in Chapters 3 and 4. We consider J spatial covariates $c_1, \ldots, c_J$, and we use the exponential form for the resource selection function,

$$\pi(\boldsymbol{x}) = \frac{1}{C} \exp\left(\sum_{j=1}^{J} \beta_j c_j(\boldsymbol{x})\right) \tag{5.3}$$

where C is a normalizing constant (as in Equation 1.2). The drift term of the Langevin diffusion defined by Equation 5.2 can then be written

$$\begin{aligned}\frac{\gamma^2}{2}\nabla \log \pi(\boldsymbol{X}_t) &= \frac{\gamma^2}{2}\nabla \log\left[\frac{1}{C}\exp\left(\sum_{j=1}^{J}\beta_j c_j(\boldsymbol{X}_t)\right)\right] \\ &= \frac{\gamma^2}{2}\nabla\left(\sum_{j=1}^{J}\beta_j c_j(\boldsymbol{X}_t) - \log(C)\right).\end{aligned}$$

Furthermore, we have $\nabla \log(C) = \mathbf{0}$, and the gradient can be written inside the sum, i.e.

$$\frac{\gamma^2}{2}\nabla \log \pi(\boldsymbol{X}_t) = \frac{\gamma^2}{2}\sum_{j=1}^{J}\beta_j \nabla c_j(\boldsymbol{X}_t)$$

Finally, Equation 5.2 becomes

$$d\boldsymbol{X}_t = \frac{\gamma^2}{2} \sum_{j=1}^{J} \beta_j \nabla c_j(\boldsymbol{X}_t) dt + \gamma d\boldsymbol{W}_t. \tag{5.4}$$

Equation 5.4 requires the covariate functions c_j to be differentiable, to ensure that ∇c_j is well-defined. As mentioned before, spatial covariates are most commonly measured on a grid of points. They are then stored as a raster and, for most applications, they are considered to be piecewise-constant over each grid cell. The gradient of a piecewise-constant function is zero, except at the boundaries of each grid cell (where it is undefined). In the context of the Langevin movement model, it is therefore necessary to interpolate the observed covariates into differentiable functions. In Sections 5.5 and 5.6, we use bilinear interpolation for this purpose (e.g. Chang, 2018, Section 6.4.1). The advantage of bilinear interpolation is that the interpolated function takes a simple form, from which it is possible to derive the gradient analytically (Appendix C). If an other interpolation technique is used, it may be necessary to use tools of numerical differentiation to compute the gradient.

Note that, because the Langevin movement model is based on the gradient of the covariate fields, it cannot directly be used with discrete-valued covariates. Indeed, a discrete covariate function may be constant over large areas of the study region. Then, the gradient of the covariate is zero in most grid cells, except in the cells where its value changes. In particular, the model is not directly applicable to categorical covariates (e.g. habitat type), because it is unclear how the gradient should be measured within each category. In the following, we only consider continuous-valued covariates.

Equation 5.4 defines a continuous-time model of movement and habitat selection. In this approach, the movement of an animal is modelled in response to its local environment. At any instant, the animal tends to move towards better habitat, i.e. in the direction of the gradient of the RSF. This is formulated in continuous time, unlike other models of local habitat selection such as step selection functions (Forester et al., 2009) or the MCMC movement model of Chapters 3 and 4. Those models describe habitat selection at the scale of the time step of observations, whereas the Langevin movement model captures continuous-time habitat selection, independently of the time step of observations. Movement models based on potential functions also describe the instantaneous displacement of an animal as a response to local habitat features. Preisler et al. (2013) described a set of assumptions under which a potential-based model is stationary. In that context, the stationary distribution of the movement process can be derived directly from the potential function. The approach that

we present can be seen as an special case of the model of Preisler et al. (2013), with the formulation of the stationary distribution as a function of spatial covariates.

5.2.3 Relation to diffusion models

We presented the Langevin diffusion process as the solution of a stochastic differential equation, which describes the evolution of the process in time (Equation 5.2). Another possible representation of the process is through the Fokker-Planck equation, which describes the evolution of the probability density function of the process.

For simplicity, we consider the one-dimensional Langevin diffusion process, and we denote by $u(x,t)$ the probability density of the animal being at the location x at time t. The Fokker-Planck equation is a partial differential equation for the unknown function u and, for this model, it can be written

$$\frac{\partial u(x,t)}{\partial t} = -\frac{\gamma^2}{2}\frac{\partial\left[\frac{\partial[\log\pi(x)]}{\partial x}u(x,t)\right]}{\partial x} + \frac{\gamma^2}{2}\frac{\partial^2 u(x,t)}{\partial x^2}. \tag{5.5}$$

Similarly to the stochastic differential equation representation, the two terms of the right-hand side of Equation 5.5 are called the drift (or advection) term, and the diffusion term, respectively. From Equation 5.5, we can draw a parallel between the Langevin movement model and spatially-structured diffusion models (Section 1.3.3). Ovaskainen (2008) describes an animal's movement as the solution to a Fokker-Planck equation with general (and possibly habitat-dependent) drift and diffusion terms. In the Langevin movement model, the diffusion term is constant, and determined by the speed parameter γ^2. The drift term, however, is variable in space, because it depends on the spatial gradient of the utilisation distribution. By property of the Langevin diffusion process, the time limit of $u(x,t)$ is the stationary distribution $\pi(x)$.

5.3 Discretization of the Langevin process

The transition density of a diffusion process $(\boldsymbol{X}_t)$ is the probability density $p(\boldsymbol{X}_{t+\delta}|\boldsymbol{X}_t)$, for any time interval $\delta > 0$. As we explained in Section 2.4, diffusion processes are Markovian, and the transition density can therefore be used for simulation and inference. For simple processes, it may be analytically tractable, e.g. for the Brownian motion, or the Ornstein-Uhlenbeck process (see Chapter 6). In those cases, we can directly derive the rules of simulation of the process, and a likelihood function for inference. However, there is in

general no closed-form expression for the transition density of the Langevin diffusion process (Roberts et al., 1996). As a result, we cannot carry out exact simulation from the Langevin process, or exact inference for the Langevin movement model. Nevertheless, numerical methods can be used to obtain an approximation of the process and of the transition density (see e.g. Iacus, 2009).

In this section, we describe one such numerical method, the Euler-Maruyama discretization scheme, and we derive the transition density of an approximation of the Langevin diffusion process. In Section 5.4, we explain how the transition density can be used to obtain estimators of all the parameters of the model.

5.3.1 Euler-Maruyama discretization

If the transition density of a diffusion process is analytically intractable, it can be approximated by discretizing the process in time. The most common method is the Euler-Maruyama discretization scheme, which is analogous to the Euler method for ordinary differential equations. The Euler-Maruyama method can be applied to the general stochastic differential equation

$$d\boldsymbol{X}_t = b(\boldsymbol{X}_t)dt + \sigma(\boldsymbol{X}_t)d\boldsymbol{W}_t, \tag{5.6}$$

where b is the drift function, and σ is the diffusion function.

Discretization methods are used to approximate the process $(\boldsymbol{X}_t)$ by a simpler (tractable) process, over each interval $[t_i, t_{i+1})$ of a discrete time grid $(t_1, \ldots, t_n)$. Under the Euler-Maruyama scheme, for $t_i \leq t < t_{i+1}$, the solution process to Equation 5.6 is approximated by the solution to

$$d\boldsymbol{X}_t = b(\boldsymbol{x}_i)dt + \sigma(\boldsymbol{x}_i)d\boldsymbol{W}_t, \tag{5.7}$$

where $\boldsymbol{X}_{t_i} = \boldsymbol{x}_i$. The drift function b is approximated by a piecewise-constant function, with value $b(\boldsymbol{x}_i)$ over $[t_i, t_{i+1})$, for all $i \in \{1, \ldots, n-1\}$. Similarly, the diffusion term is fixed to a constant value, $\sigma(\boldsymbol{x}_i)$, over each interval. In other words, the diffusion process is approximated by a Brownian motion with drift between t_i and t_{i+1}.

Because the drift and diffusion are piecewise-constant, the approximate diffusion process has a closed-form Gaussian transition density over each interval $[t_i, t_{i+1})$, given by the transition density of the Brownian motion with drift,

$$p(\boldsymbol{x}_{i+1}|\boldsymbol{x}_i) = \varphi(\boldsymbol{x}_{i+1}|\boldsymbol{x}_i + \Delta_i b(\boldsymbol{x}_i), \sigma(\boldsymbol{x}_i)^2 \Delta_i \boldsymbol{I}_2)$$

where $\varphi(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the bivariate normal density with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{I}_2$ is the 2×2 identity matrix, and $\Delta_i = t_{i+1} - t_i$ is the length of the time interval.

In the case of the Langevin diffusion, the drift and diffusion terms are

$$b(\boldsymbol{X}_t) = \frac{\gamma^2}{2} \nabla \log \pi(\boldsymbol{X}_t), \quad \text{and } \sigma(\boldsymbol{X}_t) = \gamma.$$

Following from Equation 5.7, the Euler-Maruyama discretization of the Langevin diffusion is therefore defined by

$$d\boldsymbol{X}_t = \frac{\gamma^2}{2} \nabla \log \pi(\boldsymbol{x}_i) dt + \gamma d\boldsymbol{W}_t,$$

for $t_i \leq t < t_{i+1}$.

As in the general case, this equation defines a Brownian motion with drift, and the transition density of the discretized process between t_i and t_{i+1} is

$$p(\boldsymbol{x}_{i+1}|\boldsymbol{x}_i) = \varphi\left(\boldsymbol{x}_{i+1} \middle| \boldsymbol{x}_i + \frac{\gamma^2 \Delta_i}{2} \nabla \log \pi(\boldsymbol{x}_i),\ \gamma^2 \Delta_i \boldsymbol{I}_2\right). \tag{5.8}$$

5.3.2 Simulation from the Langevin diffusion process

5.3.2.1 Approximate simulation

The Euler-Maruyama approximation of the Langevin diffusion can be used for simulations. We consider $(t_1, \ldots, t_n)$ the (possibly irregular) times at which the process should be simulated, and denote by $\boldsymbol{X}_i$ the value of the process at time t_i. We initialise the simulation with $\boldsymbol{X}_0 = \boldsymbol{x}_0$, and simulate the process forward using the approximate transition density found in Equation 5.8. For $i = 0, \ldots, n-1$,

$$\boldsymbol{X}_{i+1}|\{\boldsymbol{X}_i = \boldsymbol{x}_i\} \sim N\left(\boldsymbol{x}_i + \frac{\gamma^2 \Delta_i}{2} \nabla \log \pi(\boldsymbol{x}_i),\ \gamma^2 \Delta_i \boldsymbol{I}_2\right).$$

This procedure does not simulate exactly from the Langevin diffusion process, and the accuracy of the approximation depends on the time step of discretization (Δ_i). We investigate the effect of the discretization in the simulations of Section 5.5.1.

5.3.2.2 Metropolis-adjusted Langevin algorithm

The simulation method described in Section 5.3.2.1 is not exact, because it is based on the transition density of the Euler-Maruyama discretization of the Langevin diffusion process. Roberts et al. (1996) called this simulation algorithm the "unadjusted Langevin algorithm", and they showed that it may not converge to the correct stationary distribution π. In the context of MCMC sampling, they described a "corrected" version of the discretized Langevin diffusion, to sample exactly from the target distribution. They considered the transition density of the discretized Langevin process as the proposal distribution for a Metropolis-Hastings algorithm. In their Metropolis-adjusted Langevin algorithm, a sample is proposed at each iteration from the approximate transition density, and it is either accepted or rejected, with some probability. This is a special case of Metropolis-Hastings, such that the limiting distribution of samples is the correct target distribution. Although the Metropolis-adjusted algorithm is guaranteed to sample from the stationary distribution π, it does not simulate from the exact transition density of the Langevin diffusion. It is useful to construct a MCMC algorithm that converges to the correct distribution, but it is not a good basis for a model of animal movement. Indeed, it requires a Metropolis acceptance-rejection step, to correct for the approximation of the transition density. For the reasons discussed in Sections 3.2 and 4.9, rejection-based algorithms are generally not good representations of animal movement. However, we propose to use the Metropolis-adjusted Langevin algorithm indirectly, not to model animal movement, but to assess the accuracy of the Euler-Maruyama approximation.

We suggest using the acceptance rate of the Metropolis-adjusted simulation algorithm to measure the discrepancy between the true Langevin diffusion and the discretized process. As the time step of discretization decreases, the discretized process becomes a better approximation, and the acceptance rate of the algorithm increases. In Section 5.5.2, we simulate from the Metropolis-adjusted Langevin algorithm at different time steps of discretization, to compare the rates of rejections. This criterion becomes very valuable to assess a model fitted to real data. In the case of real data, the time step of discretization is given by the time step of observation, and it cannot be adjusted to improve the approximation. Then, the problem is to determine whether the time step of observation leads to a good approximation of the process, in the context of the analysis. This may depend on the speed of the process (i.e. the speed of movement of the animal), and on the spatial autocorrelation structure of the target distribution (i.e. of the covariates when modelled with a RSF). In Section 5.6, we use the acceptance rate of simulations from the Metropolis-adjusted algorithm to assess a Langevin movement model fitted to tracking data from three Steller sea lions.

5.4 Inference

The parameters of the Langevin movement model are the habitat selection parameters $\{\beta_1, \ldots, \beta_J\}$, and the speed parameter γ^2. In this section, we develop a method to estimate those parameters from telemetry and environmental data. We consider $n+1$ observed locations $(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_n)$, as a realisation of the location process $(\boldsymbol{X}_t)$ at times $(t_0, \ldots, t_n)$, and J covariates $c_1, \ldots, c_J$. We describe an approach to inference based on the Euler-Maruyama discretization of the Langevin diffusion process.

5.4.1 Approximate likelihood of the Langevin movement model

The likelihood of the Langevin movement model is not analytically tractable. Instead, we consider the likelihood of the approximate process obtained from the Euler-Maruyama discretization scheme, given in Equation 5.8. Under the discretized process, the likelihood of a step from $\boldsymbol{x}_i$ to $\boldsymbol{x}_{i+1}$ is given by its transition density over the interval $[t_i, t_{i+1}]$. The steps are independent, and the likelihood of $n+1$ observed locations $(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_n)$ is the product of the likelihood of the individual steps,

$$L(\boldsymbol{\beta}, \gamma^2 | \boldsymbol{x}_0, \ldots, \boldsymbol{x}_n) = \prod_{i=0}^{n-1} \varphi \left(\boldsymbol{x}_{i+1} \middle| \boldsymbol{x}_i + \frac{\gamma^2 \Delta_i}{2} \nabla \log \pi(\boldsymbol{x}_i),\ \gamma^2 \Delta_i \boldsymbol{I}_2 \right).$$

Substituting the exponential RSF expression of Equation 5.3 for the utilisation distribution π, the right-hand side of the equation becomes a function of the habitat selection parameters $\boldsymbol{\beta}$. Maximum likelihood estimation, or other likelihood methods, can thus be used to estimate all model parameters and associated standard errors. However, in Section 5.4.2, we show that the maximum likelihood estimators of the parameters can be written explicitly in this context.

5.4.2 Least square estimation

The transition equations of the Langevin movement model can be written under the form of a linear model. Here, we show the derivation of the linear model equation, and we give closed-form estimators of the model parameters.

5.4.2.1 Linear model formulation

From the transition density of the Euler-Maruyama discretization of the Langevin process, given in Equation 5.8, we can write

$$\boldsymbol{X}_{i+1} = \boldsymbol{X}_i + \frac{\gamma^2 \Delta_i}{2} \nabla \log \pi(\boldsymbol{X}_i) + \gamma \sqrt{\Delta_i} \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{I}_2).$$

We rearrange the equation and divide both sides by $\sqrt{\Delta_i}$, so that the variance of the noise term is γ^2,

$$\frac{\boldsymbol{X}_{i+1} - \boldsymbol{X}_i}{\sqrt{\Delta_i}} = \frac{\gamma^2 \sqrt{\Delta_i}}{2} \nabla \log \pi(\boldsymbol{X}_i) + \gamma \boldsymbol{\varepsilon}_i.$$

The utilisation distribution π is modelled by a RSF, so we substitute its expression (Equation 5.3), and we have

$$\frac{\boldsymbol{X}_{i+1} - \boldsymbol{X}_i}{\sqrt{\Delta_i}} = \frac{\gamma^2 \sqrt{\Delta_i}}{2} \sum_{j=1}^{J} \beta_j \nabla c_j(\boldsymbol{X}_i) + \gamma \boldsymbol{\varepsilon}_i. \tag{5.9}$$

We denote by $\boldsymbol{y}_i = (\boldsymbol{x}_i - \boldsymbol{x}_{i-1})/\sqrt{\Delta_{i-1}}$ the observed normalized increments, and we stack them in a vector $\boldsymbol{Y}$,

$$\boldsymbol{Y} = \begin{pmatrix} y_{1,1} \\ y_{2,1} \\ \vdots \\ y_{n,1} \\ y_{1,2} \\ \vdots \\ y_{n,2} \end{pmatrix}$$

where $\boldsymbol{y}_i = (y_{i,1}, y_{i,2})$. We define the matrices $\boldsymbol{T}_\Delta$ and $\boldsymbol{D}$ as follows,

$$\boldsymbol{T}_\Delta = \begin{pmatrix} \sqrt{\Delta_0} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & \sqrt{\Delta_{n-1}} & 0 & \cdots & \vdots \\ 0 & \cdots & 0 & \sqrt{\Delta_0} & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \sqrt{\Delta_{n-1}} \end{pmatrix},$$

$$\boldsymbol{D} = \frac{1}{2}\begin{pmatrix} \frac{\partial c_1}{\partial x}(\boldsymbol{x}_0) & \frac{\partial c_2}{\partial x}(\boldsymbol{x}_0) & \cdots & \frac{\partial c_J}{\partial x}(\boldsymbol{x}_0) \\ \vdots & \vdots & & \vdots \\ \frac{\partial c_1}{\partial x}(\boldsymbol{x}_{n-1}) & \frac{\partial c_2}{\partial x}(\boldsymbol{x}_{n-1}) & \cdots & \frac{\partial c_J}{\partial x}(\boldsymbol{x}_{n-1}) \\ \frac{\partial c_1}{\partial y}(\boldsymbol{x}_0) & \frac{\partial c_2}{\partial y}(\boldsymbol{x}_0) & \cdots & \frac{\partial c_J}{\partial y}(\boldsymbol{x}_0) \\ \vdots & \vdots & & \vdots \\ \frac{\partial c_1}{\partial y}(\boldsymbol{x}_{n-1}) & \frac{\partial c_2}{\partial y}(\boldsymbol{x}_{n-1}) & \cdots & \frac{\partial c_J}{\partial y}(\boldsymbol{x}_{n-1}) \end{pmatrix}.$$

In the matrices $\boldsymbol{Y}$, $\boldsymbol{T}_\Delta$ and $\boldsymbol{D}$, the first n rows correspond to the x coordinate, and the last n rows correspond to the y coordinate. We define $\boldsymbol{Z} = \boldsymbol{T}_\Delta \boldsymbol{D}$. Finally, the transition equations given in Equation 5.9 can be written in matrix notation as

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\nu} + \boldsymbol{E}, \tag{5.10}$$

where $\boldsymbol{\nu} = \gamma^2\boldsymbol{\beta}$, and $\boldsymbol{E}$ is a $2n$-vector of normal terms with mean 0 and variance γ^2. The matrices $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are known, as they can be derived from the data. In this system, the response variable is the observed displacement from $\boldsymbol{X}_i$ to $\boldsymbol{X}_{i+1}$, and the explanatory variables are the gradients of the covariate functions at the point $\boldsymbol{X}_i$. Equation 5.10 is in the form of a linear model, and standard results can be used to derive estimators of the model parameters.

5.4.2.2 Estimators of $\boldsymbol{\beta}$ and γ^2

From the linear model given in Equation 5.10, ordinary least squares can be used to obtain the maximum likelihood estimator of $\boldsymbol{\nu}$,

$$\hat{\boldsymbol{\nu}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y},$$

where $\boldsymbol{Z}'$ denotes the transpose of the matrix $\boldsymbol{Z}$. The distribution of the estimator is

$$\hat{\boldsymbol{\nu}} \sim N(\boldsymbol{\nu},\ \gamma^2(\boldsymbol{Z}'\boldsymbol{Z})^{-1}). \tag{5.11}$$

The variance of the error vector $\boldsymbol{E}$ (defined in Equation 5.10) is the speed parameter γ^2. An estimator of γ^2 is

$$\hat{\gamma}^2 = \frac{1}{k}\|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\|^2,$$

where $k = 2n - J$ is the number of degrees of freedom. Here, $\hat{\boldsymbol{Y}} = \boldsymbol{Z}\hat{\boldsymbol{\nu}}$ is the predicted value of $\boldsymbol{Y}$, and $\|\boldsymbol{Y} - \hat{\boldsymbol{Y}}\|^2$ is therefore the sum of squared residuals. From Cochran's theorem, the estimator $\hat{\gamma}^2$ is independent of $\hat{\boldsymbol{\nu}}$, and it satisfies

$$\hat{\gamma}^2 = \frac{\gamma^2}{k} \times G, \text{ where } G \sim \chi^2_k. \tag{5.12}$$

We found an estimator of $\boldsymbol{\nu} = \gamma^2\boldsymbol{\beta}$ and of γ^2, that we can use to derive an estimator of $\boldsymbol{\beta}$. We first define $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\nu}}/\hat{\gamma}^2$, and we show that $\tilde{\boldsymbol{\beta}}$ is not an unbiased estimator of $\boldsymbol{\beta}$. Indeed, its expectation is

$$\begin{aligned} E[\tilde{\boldsymbol{\beta}}] &= E\left[\frac{\hat{\boldsymbol{\nu}}}{\hat{\gamma}^2}\right] \\ &= E[\hat{\boldsymbol{\nu}}]E\left[\frac{1}{\hat{\gamma}^2}\right], \end{aligned}$$

by independence of $\hat{\boldsymbol{\nu}}$ and $\hat{\gamma}^2$. From Equation 5.11,

$$E[\hat{\boldsymbol{\nu}}] = \boldsymbol{\nu} = \gamma^2\boldsymbol{\beta},$$

and, from Equation 5.12,

$$E\left[\frac{1}{\hat{\gamma}^2}\right] = E\left[\frac{k}{\gamma^2} \times \frac{1}{G}\right] = \frac{k}{\gamma^2} \times E\left[\frac{1}{G}\right]. \tag{5.13}$$

The inverse of G follows an inverse-chi-squared distribution with k degrees of freedom, with expectation

$$E\left[\frac{1}{G}\right] = \frac{1}{k-2}.$$

Following from Equation 5.13,

$$E\left[\frac{1}{\hat{\gamma}^2}\right] = \frac{k}{\gamma^2} \times \frac{1}{k-2} = \frac{k}{(k-2)\gamma^2}. \tag{5.14}$$

Finally, we obtain the expectation of $\tilde{\boldsymbol{\beta}}$, as

$$\begin{aligned} E[\tilde{\boldsymbol{\beta}}] &= E[\hat{\boldsymbol{\nu}}]E\left[\frac{1}{\hat{\gamma}^2}\right] \\ &= \gamma^2\boldsymbol{\beta} \times \frac{k}{(k-2)\gamma^2} \end{aligned}$$

$$= \frac{k}{k-2}\boldsymbol{\beta},$$

so $\tilde{\boldsymbol{\beta}}$ is not an unbiased estimator of $\boldsymbol{\beta}$. However, it is easy to see that

$$E\left[\left(\frac{k-2}{k}\right)\tilde{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}.$$

Finally, we define an estimator of $\boldsymbol{\beta}$,

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\frac{k-2}{k}\right)\tilde{\boldsymbol{\beta}} \\ &= \frac{(k-2)\hat{\boldsymbol{\nu}}}{k\hat{\gamma}^2}.\end{aligned}$$

Note that the estimators $\hat{\gamma}^2$ and $\hat{\boldsymbol{\beta}}$ defined here are unbiased for the linear model given in Equation 5.10. However, that linear model relies on an approximation of the Langevin diffusion process. The estimators are therefore biased estimators for the parameters of the Langevin diffusion, and they are only unbiased in the limit, as the time step of discretization tends to zero. In the simulation study of Section 5.5.3, we investigate the accuracy of this method for different time steps of discretization.

5.4.2.3 Uncertainty quantification

Confidence intervals for the parameters $\boldsymbol{\beta}$ and γ^2 can also be derived, from the distribution of the estimators. From Equation 5.12, it follows that

$$\gamma^2 = \frac{k}{G} \times \hat{\gamma}^2, \text{ where } G \sim \chi^2_k.$$

So, a confidence interval for γ^2, of order $\alpha \in (0, 0.5)$, is given by

$$CI_\alpha(\gamma^2) = \left[\frac{k}{q_{1-\alpha/2,k}} \times \hat{\gamma}^2,\ \frac{k}{q_{\alpha/2,k}} \times \hat{\gamma}^2\right],$$

where $q_{\alpha,k}$ is the quantile of order α of the chi-squared distribution with k degrees of freedom.

To quantify uncertainty in the habitat selection parameters $\boldsymbol{\beta}$, we first calculate the covariance structure of $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\nu}}/\hat{\gamma}^2$,

$$\begin{aligned}
\text{Cov}(\tilde{\beta}_i, \tilde{\beta}_j) &= E[\tilde{\beta}_i\tilde{\beta}_j] - E[\tilde{\beta}_i]E[\tilde{\beta}_j] \\
&= E\left[\frac{\hat{\nu}_i\hat{\nu}_j}{\hat{\gamma}^4}\right] - E\left[\frac{\hat{\nu}_i}{\hat{\gamma}^2}\right] E\left[\frac{\hat{\nu}_j}{\hat{\gamma}^2}\right] \\
&= E\left[\frac{1}{\hat{\gamma}^4}\right] E[\hat{\nu}_i\hat{\nu}_j] - E\left[\frac{1}{\hat{\gamma}^2}\right]^2 E[\hat{\nu}_i]E[\hat{\nu}_j]. \qquad (5.15)
\end{aligned}$$

By definition of the variance, we have

$$E\left[\frac{1}{\hat{\gamma}^4}\right] = \text{Var}\left[\frac{1}{\hat{\gamma}^2}\right] + E\left[\frac{1}{\hat{\gamma}^2}\right]^2.$$

We found $E(1/\hat{\gamma}^2)$ in Equation 5.14 and, from Equation 5.12,

$$\begin{aligned}
\text{Var}\left[\frac{1}{\hat{\gamma}^2}\right] &= \text{Var}\left[\frac{k}{\gamma^2} \times \frac{1}{G}\right] \\
&= \left(\frac{k}{\gamma^2}\right)^2 \text{Var}\left[\frac{1}{G}\right],
\end{aligned}$$

where $G \sim \chi^2_k$. We substitute the variance of an inverse-chi-squared distribution with k degrees of freedom,

$$\begin{aligned}
\text{Var}\left[\frac{1}{\hat{\gamma}^2}\right] &= \left(\frac{k}{\gamma^2}\right)^2 \frac{2}{(k-2)^2(k-4)} \\
&= \frac{2k^2}{(k-2)^2(k-4)\gamma^4}.
\end{aligned}$$

So, we have

$$\begin{aligned}
E\left[\frac{1}{\hat{\gamma}^4}\right] &= \frac{2k^2}{(k-2)^2(k-4)\gamma^4} + \left(\frac{k}{(k-2)\gamma^2}\right)^2 \\
&= \frac{k^2}{(k-2)^2\gamma^4}\left(1 + \frac{2}{k-4}\right).
\end{aligned}$$

We also have

$$\text{Cov}(\hat{\nu}_i, \hat{\nu}_j) = E[\hat{\nu}_i\hat{\nu}_j] - E[\hat{\nu}_i]E[\hat{\nu}_j],$$

and so, from Equation 5.11,

$$\begin{aligned} E[\hat{\nu}_i\hat{\nu}_j] &= \text{Cov}(\hat{\nu}_i, \hat{\nu}_j) + E[\hat{\nu}_i]E[\hat{\nu}_j] \\ &= \gamma^2\Lambda_{ij} + \gamma^4\beta_i\beta_j, \end{aligned}$$

where $\boldsymbol{\Lambda} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}$.

From Equation 5.11, we have $E[\hat{\nu}_i] = \nu_i = \gamma^2\beta_i$ and $E[\hat{\nu}_j] = \nu_j = \gamma^2\beta_j$. Plugging these expressions into Equation 5.15, we find

$$\begin{aligned} \text{Cov}(\tilde{\beta}_i, \tilde{\beta}_j) &= \frac{k^2}{(k-2)^2\gamma^4}\left(1 + \frac{2}{k-4}\right)(\gamma^2\Lambda_{ij} + \gamma^4\beta_i\beta_j) - \frac{k^2}{(k-2)^2\gamma^4}\gamma^4\beta_i\beta_j \\ &= \frac{k^2}{(k-2)^2\gamma^4}\left[\left(1 + \frac{2}{k-4}\right)(\gamma^2\Lambda_{ij} + \gamma^4\beta_i\beta_j) - \gamma^4\beta_i\beta_j\right] \\ &= \frac{k^2}{(k-2)^2\gamma^4}\left[\gamma^4\beta_i\beta_j\left(\frac{2}{k-4}\right) + \gamma^2\Lambda_{ij}\left(1 + \frac{2}{k-4}\right)\right] \\ &= \frac{k^2}{(k-2)^2}\left[\frac{2\beta_i\beta_j}{k-4} + \frac{\Lambda_{ij}}{\gamma^2}\left(1 + \frac{2}{k-4}\right)\right]. \end{aligned}$$

From this, we deduce the covariance structure of $\hat{\boldsymbol{\beta}}$,

$$\begin{aligned} \text{Cov}(\hat{\beta}_i, \hat{\beta}_j) &= \text{Cov}\left(\frac{k-2}{k}\tilde{\beta}_i, \frac{k-2}{k}\tilde{\beta}_j\right) \\ &= \left(\frac{k-2}{k}\right)^2 \text{Cov}(\tilde{\beta}_i, \tilde{\beta}_j) \\ &= \frac{2\beta_i\beta_j}{k-4} + \frac{\Lambda_{ij}}{\gamma^2}\left(1 + \frac{2}{k-4}\right). \end{aligned}$$

From the asymptotic normality of the estimators, we can derive an approximate confidence interval of order α for β_i,

$$CI_\alpha(\beta_i) = \left[\hat{\beta}_i - z_{\alpha/2}\sqrt{\text{Var}(\hat{\beta}_i)},\ \hat{\beta}_i + z_{\alpha/2}\sqrt{\text{Var}(\hat{\beta}_i)}\right],$$

where $z_{\alpha/2}$ is the quantile of level $\alpha/2$ of the standard normal distribution. In the equation above, the variance of the estimator is

$$\text{Var}(\hat{\beta}_i) = \text{Cov}(\hat{\beta}_i, \hat{\beta}_i)$$

$$= \frac{2\beta_i^2}{k-4} + \frac{\Lambda_{ii}}{\gamma^2}\left(1 + \frac{2}{k-4}\right).$$

5.5 Simulation study

5.5.1 Effect of time discretization on the target distribution

The transition density of the Euler-Maruyama discretization of the Langevin diffusion can be used for simulation. Roberts et al. (1996) explained that there is no guarantee that the discretized process has the same stationary distribution as the true process. Here, we investigate the effect of the time step of discretization on the long-term distribution of samples, and we verify that the target distribution is well approximated when the discretization is fine enough.

We considered three spatial covariates. We generated two random covariates over $[-50, 50] \times [-50, 50]$ at a resolution of 1, with the function `RMmatern` from the R package RandomFields (Schlather et al., 2015). The third covariate was the squared distance to the centre of the map, and was included to obtain an effect of attraction towards that point. This was used to ensure that the simulations did not come too close to the boundaries of the study region, where the gradient of the covariate functions is undefined. We defined the target distribution π as a normalized RSF based on these three covariates (Equation 5.3), with the coefficients $\boldsymbol{\beta} = (1, 1, -1)$. We simulated 10^6 locations from the Euler-Maruyama discretization of the Langevin diffusion process, with six different (regular) time steps of discretization: $\Delta \in \{5, 2, 1, 0.5, 0.2, 0.1\}$. The first 200 locations of each track, and the target distribution π, are shown in Figure 5.1.

For each simulated track of length 10^6, we counted the number of points in each cell of the spatial grid of resolution 1 (over which the covariates, and therefore the target distribution, are defined). We normalized the counts to obtain the empirical distribution of simulated points, and compared it to the true target distribution π. Figure 5.2 shows a comparison of the true and empirical distributions in the six experiments. Figure 5.3 shows the same comparison on the log scale, to visualise the smaller values. In all plots, we excluded the grid cells in which there were less than 20 simulated points, i.e. grid cells that were not (or very rarely) visited.

The top plots of Figures 5.2 and 5.3 indicate a non-linear relationship between the normalized counts and the values of the target distribution in the grid cells, in the simulations with long time intervals ($\Delta = 5$, $\Delta = 2$). For time intervals of $\Delta = 1$ and shorter, both figures

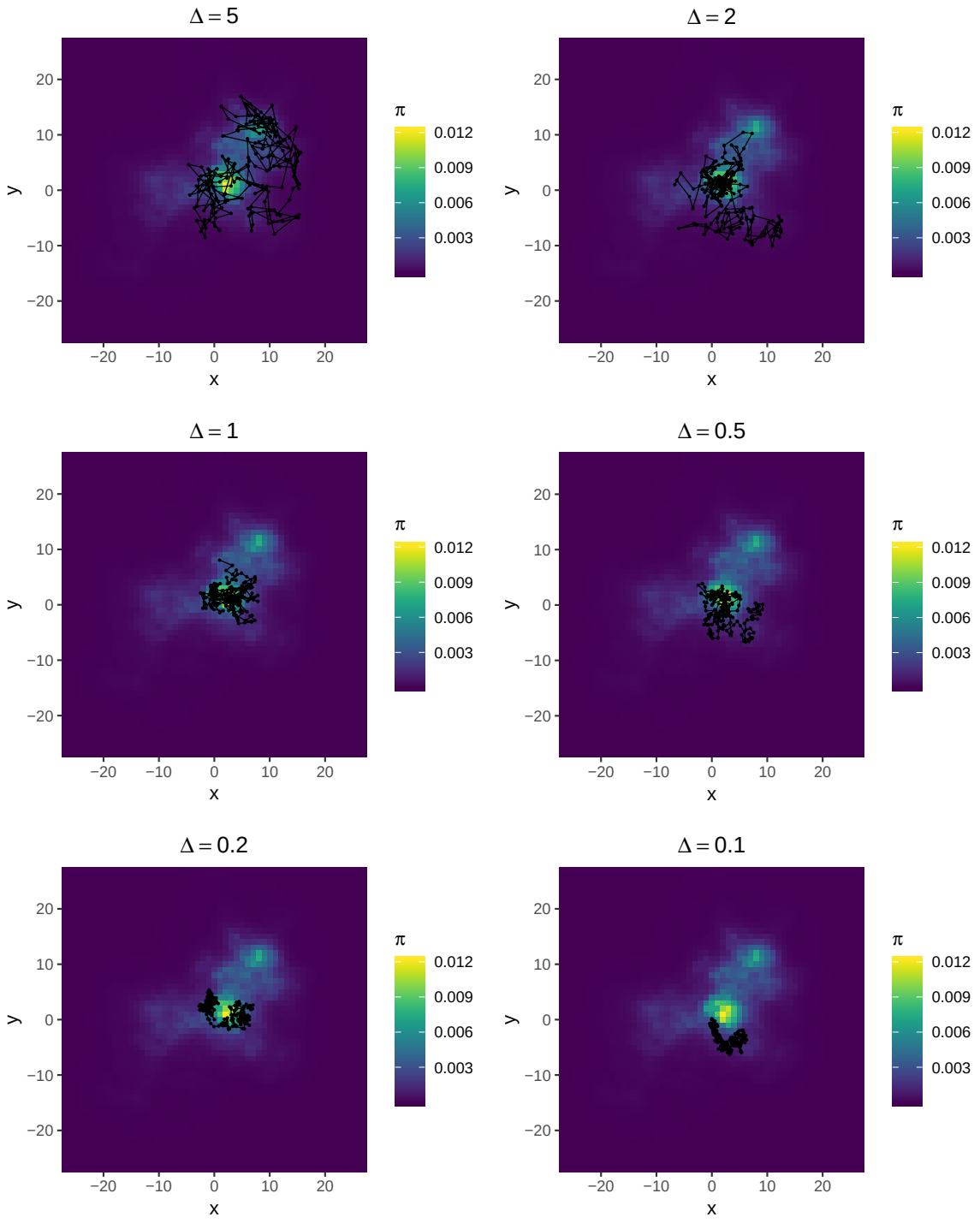


Fig. 5.1 First 200 locations of tracks simulated with different discretization steps from the Langevin movement model, on an artificial target distribution π.

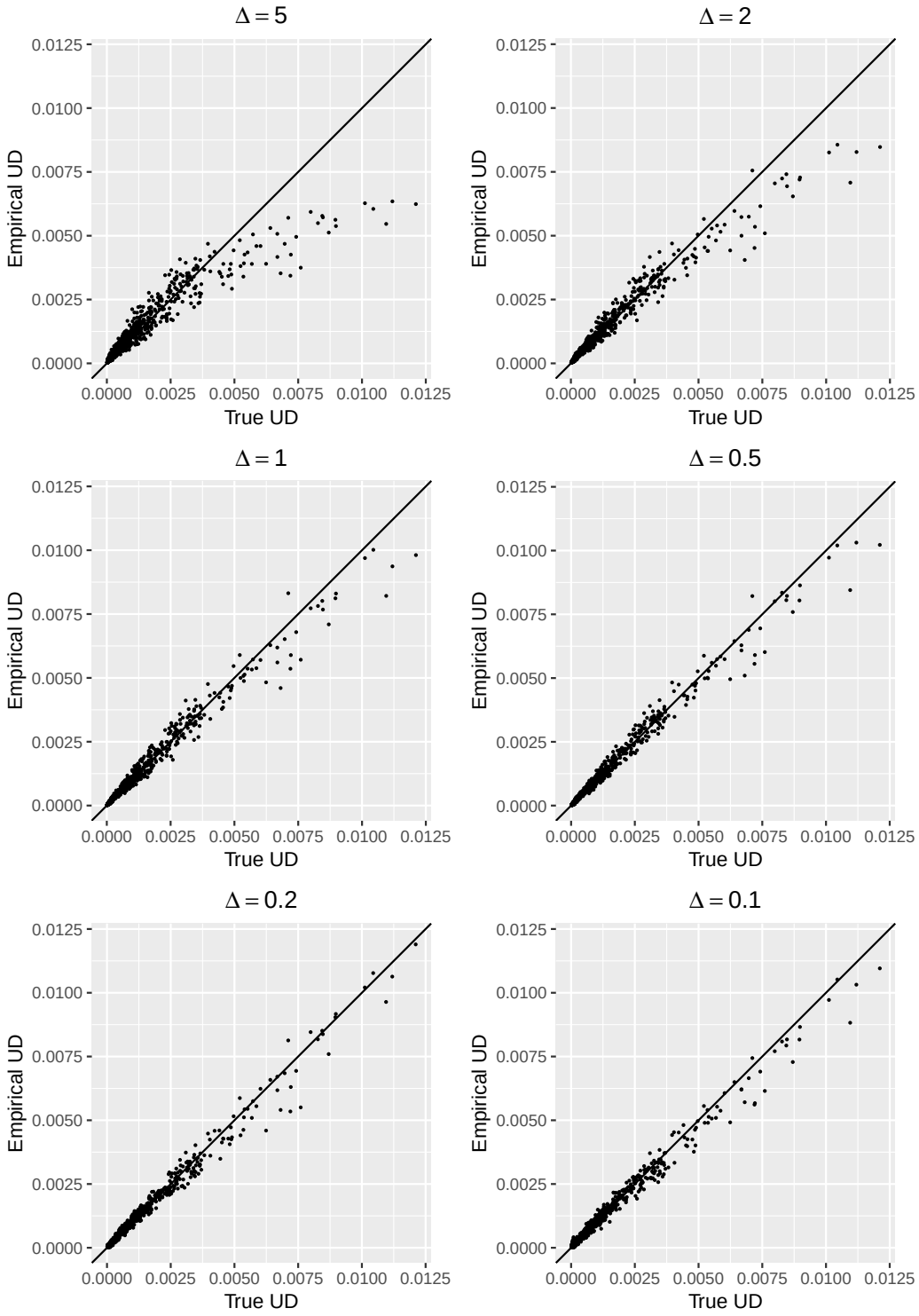


Fig. 5.2 Comparison of empirical distribution of simulated points and true target distribution, for different discretization steps.

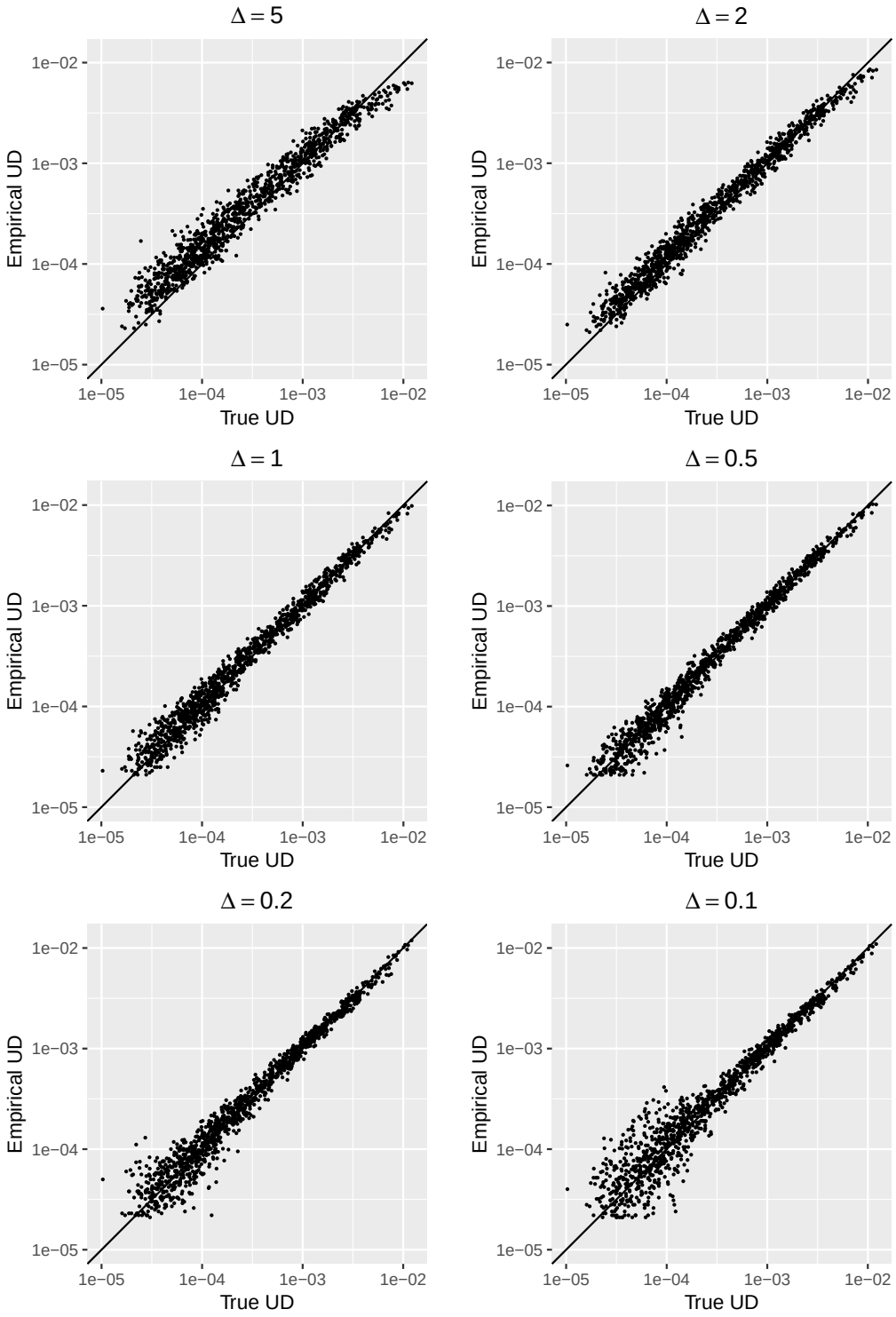


Fig. 5.3 Comparison of empirical distribution of simulated points and true target distribution, for different discretization steps. Both axes are on the log scale.

show a closer alignment with the identity line, suggesting that the simulated samples give a good approximation of the target distribution. The bottom plots of Figure 5.3 display more variability for grid cells where the target distribution takes small values. This is because the simulations with short time intervals covered a shorter period of time overall (the number of samples being fixed to 10^6 in all cases). Therefore, tracks simulated on a finer time grid were overall less likely to visit grid cells where the target distribution is low.

In this experiment, we simulated from the transition density of the Euler-Maruyama approximation of the Langevin diffusion process. The results confirm that, when the time step of discretization is short enough, the distribution of samples is a good estimate of the target distribution.

5.5.2 Metropolis acceptance probability

Simulations based on a discretization of the Langevin diffusion process are not exact. In Section 5.3.2.2, we suggested that the acceptance rate of the Metropolis-adjusted Langevin algorithm could be used to measure the discrepancy between the true and the approximated processes. In the Metropolis-adjusted algorithm, the transition density of the discretized Langevin diffusion is used as the proposal distribution of a Metropolis sampler. At each iteration, a point is drawn from the proposal distribution, and it is accepted or rejected with some probability.

Like in the previous experiment, we considered two random habitat covariates (c_1 and c_2), and the squared distance to the centre of the study region (c_3), to construct a RSF. This time, the covariates and the RSF were generated over $[-100, 100] \times [-100, 100]$, at a resolution of 1. The target distribution was defined as the normalized RSF, with the habitat selection parameters $\boldsymbol{\beta} = (4, 2, -0.1)$. The negative effect of the squared distance to the centre ensured that simulated tracks would remain within the limits of the study region. Plots of the two random covariates and of the target distribution are shown in Figure 5.4.

We simulated from the Metropolis-adjusted Langevin algorithm, with the target distribution defined above. For the simulations, we considered nine different (regular) time steps of discretization,

$$\Delta \in \{0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5\}.$$

For each, we simulated ten tracks of length $T = 1000$ from the Metropolis-adjusted Langevin algorithm, with speed parameter $\gamma^2 = 5$. We counted the average proportion of rejected steps over the ten tracks. Figure 5.5 shows the acceptance rates for the different time discretizations.

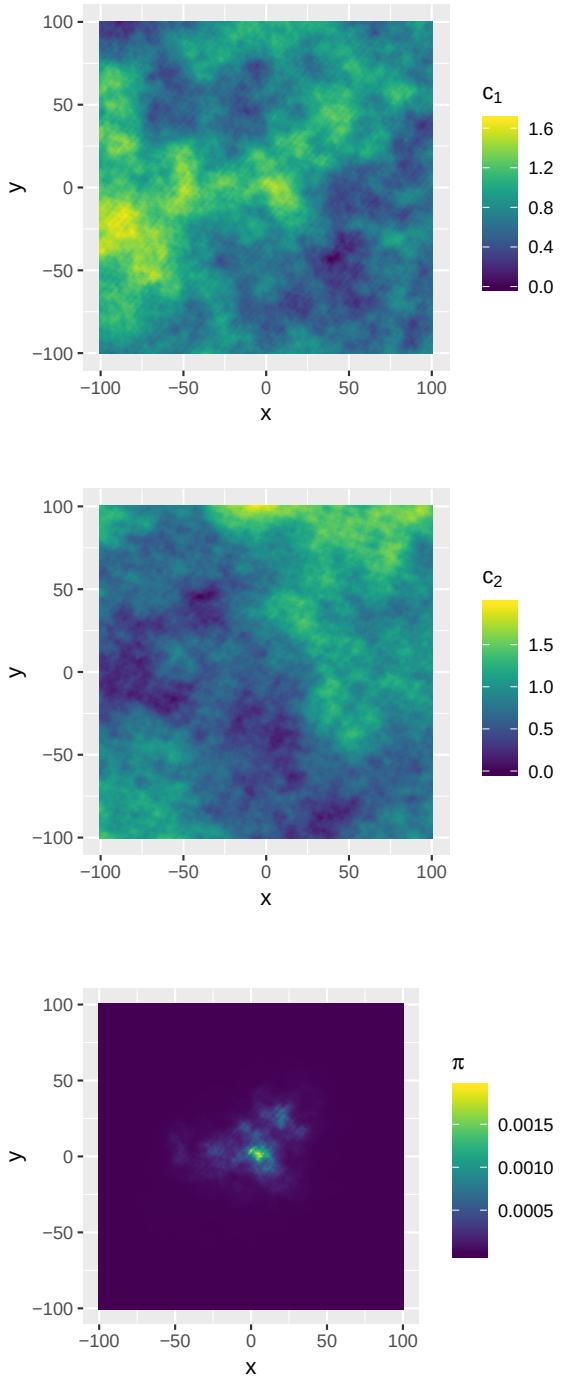


Fig. 5.4 Covariates and target distribution used in the simulations of Section 5.5.2 and 5.5.3. The target distribution π also includes the effect the squared distance to the centre of the map, not shown here.

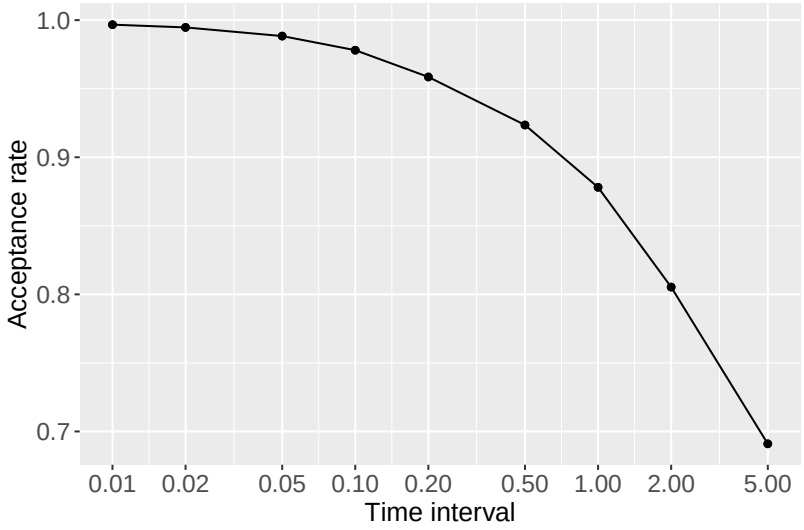


Fig. 5.5 Acceptance rates in simulations from the Metropolis-adjusted Langevin algorithm. The x axis is on the log scale.

As expected, the acceptance rate decreased as the time interval of discretization increased. This is because the Euler-Maruyama discretization is only valid at a fine time resolution, and the approximation becomes worse on a coarse time scale. The acceptance rate tends to 1 when the step of discretization decreases, i.e. when the transition density becomes a better approximation of the Langevin diffusion process. In this simulation scenario, the average acceptance rate was around 99.7% for $\Delta = 0.01$, and around 69.1% for $\Delta = 5$.

5.5.3 Estimation from thinned tracks

In Sections 5.5.1 and 5.5.2, we investigated the effect of the Euler-Maruyama discretization when used to simulate (approximately) from the Langevin diffusion process. Here, we look into its effect when the discretization scheme is used to carry out inference, as described in Section 5.4.

We considered the utilisation distribution defined in Section 5.5.2, and shown in Figure 5.4. We simulated 200 tracks from the Euler-Maruyama discretization of the Langevin diffusion process, as described in Section 5.3.2.1. The time step of discretization was Δ=0.01 in the simulations, which led to an acceptance rate of about 99.7% in Section 5.5.2. We then thinned the simulated tracks, to emulate real telemetry data, and estimated the model parameters from the thinned data. In different experiments, we thinned the tracks by 1 (i.e. time step of simulation), 2, 5, 10, 20, 50, and 100. Each time, we kept the first 2000 (thinned) locations of each of the 200 tracks, resulting in a total of 40000 locations. Note that, with this procedure,

the same number of observations was used in all experiments, but they covered different periods of time, because of the different thinning factors. For example, a track of 2000 locations covers 2000 time units if the time resolution is $\Delta = 1$, but it only covers 20 time units if the time resolution is $\Delta = 0.01$. We then applied the method described in Section 5.4 to fit the Langevin movement model separately to the 200 tracks of each thinned data set. Estimates from all experiments are shown in Figure 5.6.

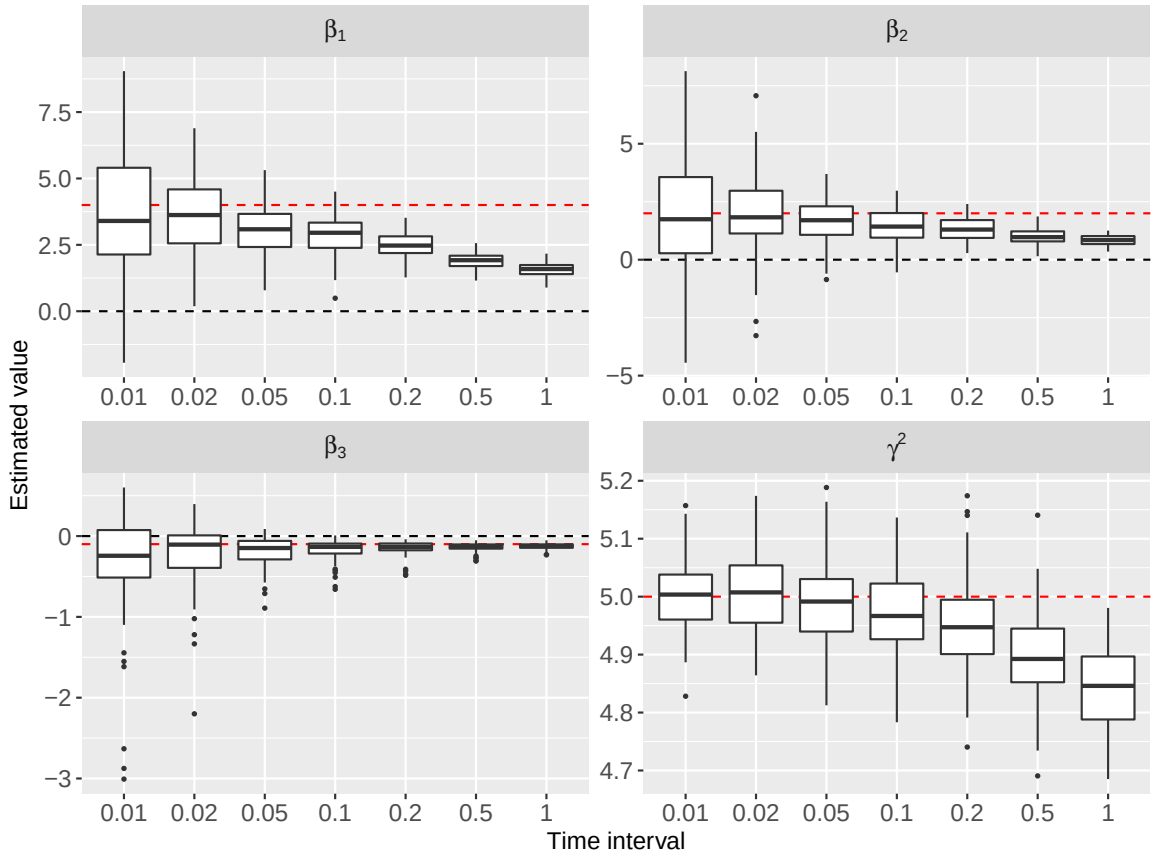


Fig. 5.6 Boxplots of estimates of the habitat selection parameters β_1, β_2 and β_3, and of the speed parameter γ^2, for different time intervals of observation. The red dotted lines show the true values of the parameters used in the simulation. The x axis is on the log scale.

The standard errors of the habitat selection parameters ($\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$) decreased as the time interval of observation increased. This is because the same number of (thinned) observations were used in all experiments, such that the tracks thinned to longer time intervals explored a larger area. The latter thus covered a larger spatial extent, and also a wider range of covariate values. Like in a standard linear regression, the uncertainty on the coefficients decreases when the variance of the covariates increases. To offset this effect, we considered a second scenario, in which all tracks covered the same period of time. We thinned each of the 100

tracks as before, but we then kept the locations over the time period from $t = 0$ to $t = 500$, regardless of the time interval of observation. At a resolution of $\Delta = 0.01$, each track comprised 50000 locations; for $\Delta = 1$, each track comprised 500 locations. We fitted the Langevin movement model to each track separately, for each time resolution. The estimates are shown in Figure 5.7.

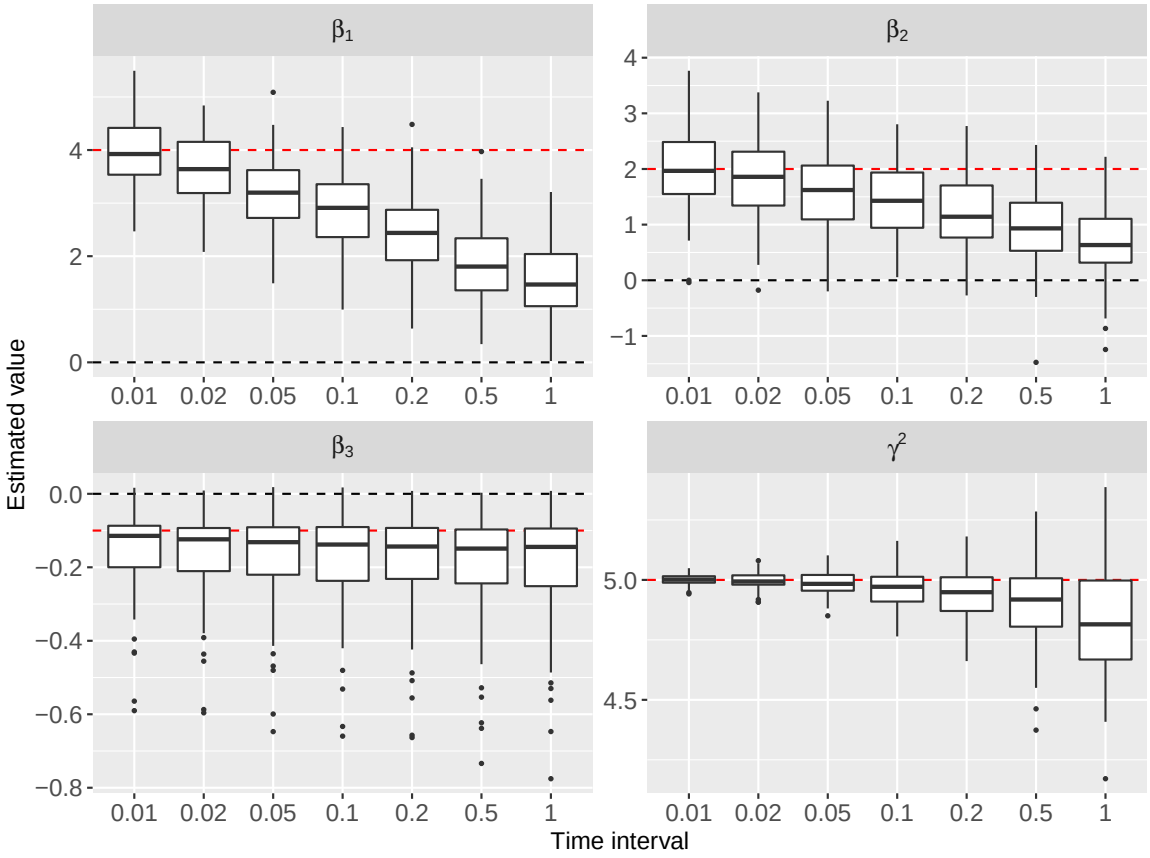


Fig. 5.7 Boxplots of estimates of the habitat selection parameters β_1, β_2 and β_3, and of the speed parameter γ^2, for different time intervals of observation. The red dotted lines show the true values of the parameters used in the simulation. The x axis is on the log scale.

When the tracks were always truncated to the same interval of time, the variability of the estimates of the habitat selection parameters was the same for all time intervals. This suggests that the uncertainty on the estimates of the habitat selection parameters depends on the extent of spatial exploration, rather than on the number of observations. However, in this case, the variance in the estimates of the speed parameter γ^2 increased as the number of observations decreased (i.e. in this case, as the time interval increased).

In both Figures 5.6 and 5.7, the mean estimates of β_1 and β_2 were close to the true values for the high-resolution tracks ($\Delta = 0.01$, $\Delta = 0.02$). For longer time intervals of observation,

however, they tended to decrease as the time interval increased. As a consequence, the true parameter values were not well captured by the estimates, for longer time intervals. This illustrates the deterioration of the Euler-Maruyama approximation for long time steps of discretization. There are two sources of bias here, because the estimators are only unbiased when the number of observations tends to infinity, and when the time step of observation tends to zero (see e.g. Kessler et al., 2012, Section 1.5.3). Based on these observations, we would expect the selection parameter of the third covariate (squared distance to the centre) to be underestimated in absolute value, for long time intervals. That is, we would expect it to be overestimated, because it is negative. However, it does not appear to be the case here: in fact, the mean estimates of β_3 are very similar for all time intervals. This may be because this covariate is a very smooth function, and its effect can be detected even over long time steps. Note that, although the strength of the selection was underestimated for the first two covariates, its sign (attraction or repulsion) was correctly captured in the estimation.

A possible interpretation of the bias on β_1 and β_2, in the context of the application to animal space use, is the following. As the time intervals between observed locations increases, the short-term effect of habitat features on the movement is less visible in the data. There is therefore more uncertainty about the selection process, and about the overall space use. As a consequence, the estimated utilisation distribution is flatter when the time intervals are longer. In the limit, if the locations were very sparse in time, we would have no information about the selection, and the estimated distribution of space use would be uniform over the study region. For this reason, the methods presented in this chapter are not adapted to low-frequency movement data, such as spatially-referenced mark-recapture data (Ovaskainen, 2004), because the Langevin diffusion process captures short-term local habitat selection.

The speed parameter γ^2 was well recovered in most experiments. However, for the longer time intervals ($\Delta = 0.5$ and $\Delta = 1$), it was slightly underestimated. The bias is not as severe as that observed for the habitat selection parameters, because the speed parameter appears in the diffusion term of the process, rather than only in the drift (Tang and Chen, 2009). The speed is underestimated because the location process tends to revert to a point in space, due to the inclusion of a force of attraction (with the squared distance to the centre). As a consequence, over long time intervals, the displacements tend to "cancel out", and the observed distance is much shorter than the actual distance travelled.

5.6 Steller sea lion case study

We illustrate the method described in Section 5.4, with a data set previously analysed by Wilson et al. (2018). Three Steller sea lions (*Eumetopias jubatus*) were equipped with Argos tags, off the coast of Alaska, and monitored for periods of 150 days, 46 days, and 153 days, respectively. The data set comprises a total of 2672 locations, indexed by time and each associated with an estimate of the measurement uncertainty (the Argos location class; Hays et al., 2001). The time intervals between observations were highly irregular, with a median interval of 1.28h, and 2.5% and 97.5% quantiles $P_{0.025}$ = 6min and $P_{0.975}$ = 17.4h. The locations were projected to UTM coordinates for the analysis.

The measurement error of Argos tags can be large, and cannot be neglected in this application. Following Wilson et al. (2018), we considered a two-stage approach to address this issue. We first fitted a state-space model to obtain estimates of the true locations of the animals, and we then analysed the filtered tracks with the Langevin movement model. In the first stage, we used the continuous-time correlated random walk (CTCRW) model, which was described by Johnson et al. (2008a), and is implemented in the R package crawl (Johnson and London, 2018). It models the movement and the observation process separately in a state-space model formulation, and the Kalman filter can be used to estimate the true location of the animal at the times of the observations. We describe the CTCRW in more detail in Chapter 6. The three filtered tracks are shown in Figure 5.8, on a map of the Northern Pacific Ocean.

Wilson et al. (2018) also provided four environmental variables over the study region, at a resolution of 1km: bathymetry (i.e. ocean depth), slope, distance to sites of interest, and distance to continental shelf. The sites of interest were either haul-out sites, where the sea lions rest, or rookeries (i.e. colonies). The continental shelf mainly consisted of the Aleutian Islands, on which the Steller sea lions live, and which can be seen on the map of Figure 5.8. The four covariates are shown in Figure 5.9.

Using the methodology described in Section 5.4, we fitted the Langevin movement model to the three filtered tracks. Most of the computation time was needed to evaluate the gradient of each covariate at all observed locations, which took about 1 sec on a 2GHz i5 CPU. Table 5.1 gives estimates and confidence intervals of the parameters, and the estimated utilisation distribution is displayed in Figure 5.10.

The 95% confidence interval of the estimate of β_2 includes zero, suggesting that that there is no clear effect of the slope covariate on the sea lions' movements. However, the model identified an effect for the other three covariates. The coefficient associated with bathymetry was estimated to be positive, which indicates that the sea lions were attracted by shallow areas.

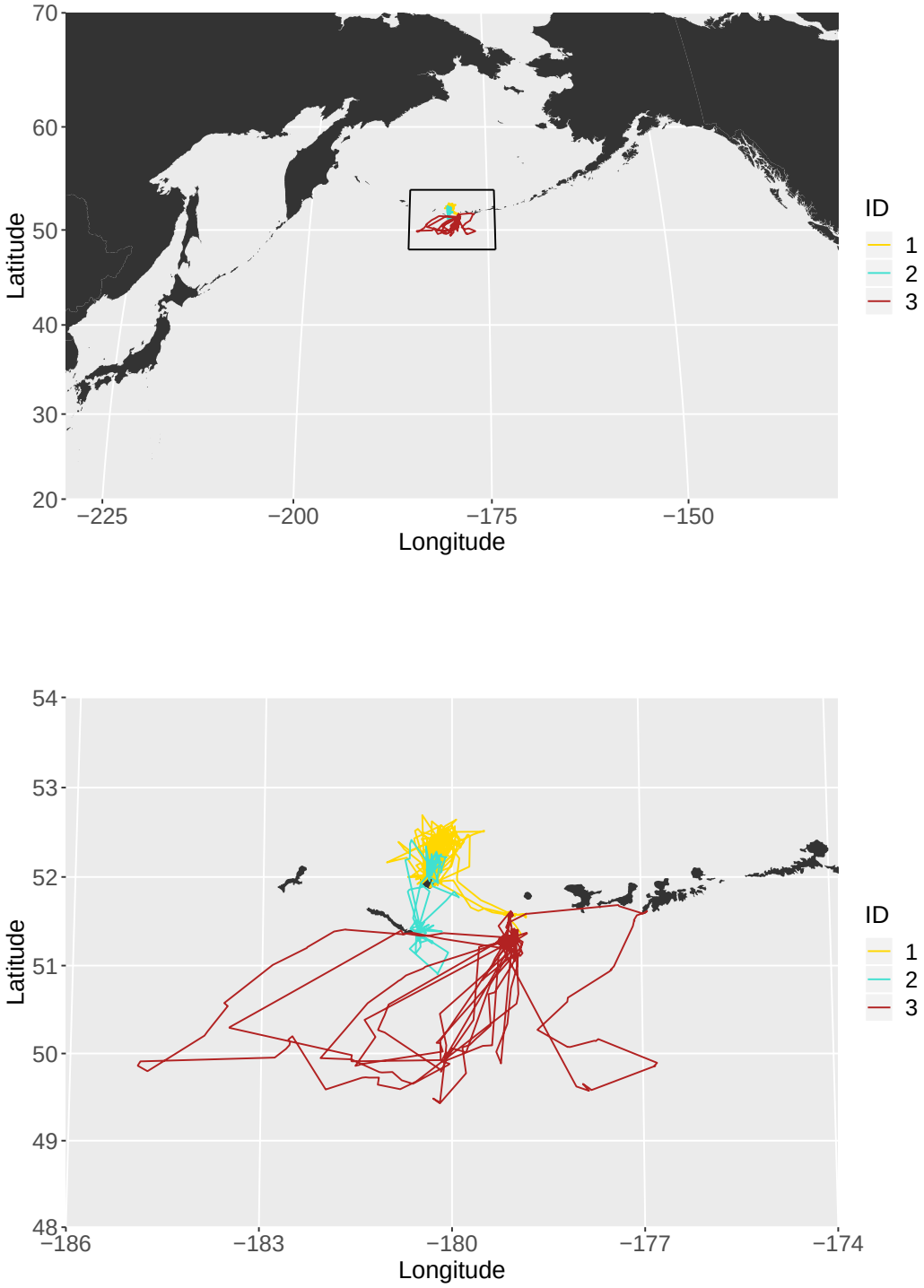


Fig. 5.8 Steller sea lion tracks in the Northern Pacific Ocean, after filtering with the R package crawl. The bottom plot magnifies the region delimited by a rectangle in the top plot.

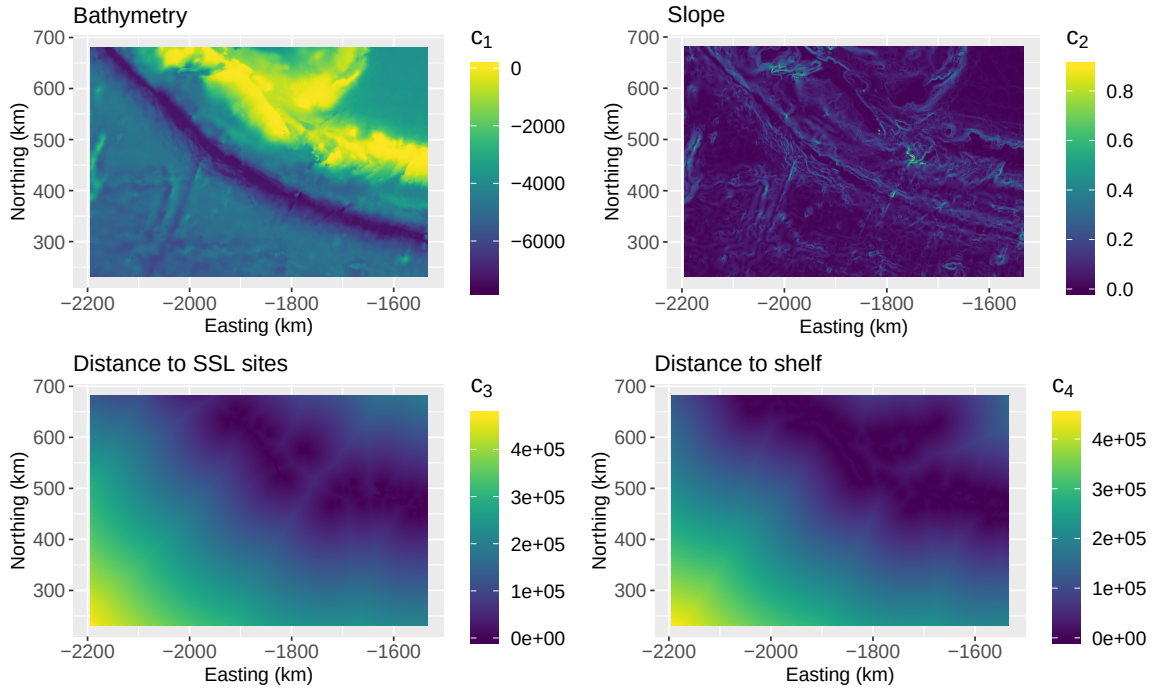


Fig. 5.9 Environmental covariates in the Steller sea lion analysis.

	Estimate	95% CI
β_1	$1.39 \cdot 10^{-4}$	$(4.40 \cdot 10^{-7}, 2.77 \cdot 10^{-4})$
β_2	$9.21 \cdot 10^{-2}$	$(-1.58 \cdot 10^{-1}, 3.42 \cdot 10^{-1})$
β_3	$-2.47 \cdot 10^{-5}$	$(-3.55 \cdot 10^{-5}, -1.39 \cdot 10^{-5})$
β_4	$3.57 \cdot 10^{-6}$	$(3.07 \cdot 10^{-7}, 6.83 \cdot 10^{-6})$
γ^2	12.4	$(11.9, 12.8)$

Table 5.1 Maximum likelihood estimates and 95% confidence intervals, from the Steller sea lion analysis.

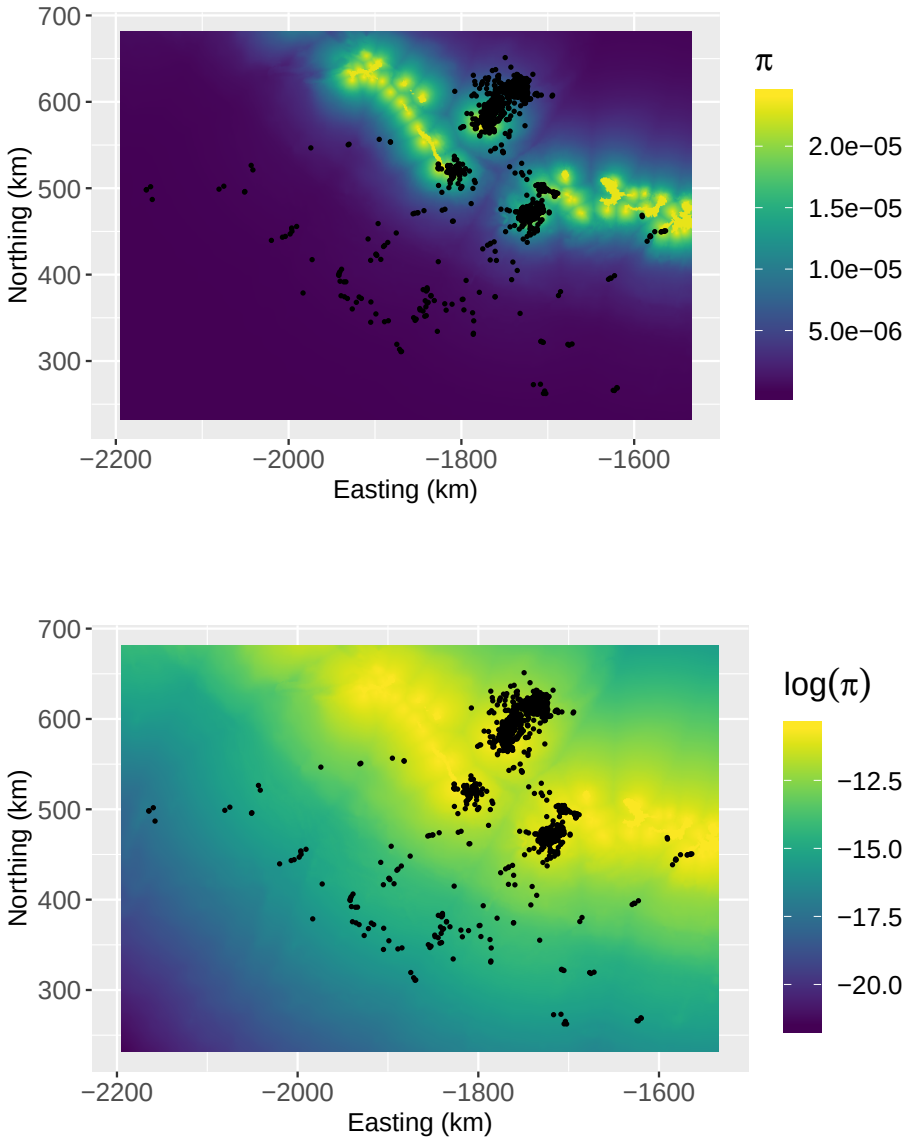


Fig. 5.10 Estimated utilisation distribution for the Steller sea lions, on the natural scale (top) and the log scale (bottom, for comparison with Wilson et al. 2018). The black dots are the filtered telemetry locations.

The estimate of the coefficient associated with the distance to sites of interest was negative, i.e. the sea lions tend to move towards sites of interest. These two observations are consistent, because sites of interest are haul-out sites and rookeries, present in areas of shallow waters, on or close to the islands. However, the estimate of the coefficient associated with the distance to the continental shelf was positive, which suggests the sea lions tend to move away from the shelf. This seems to contradict the observation that the seal lions spend most of their time close to the shelf. A possible cause is the strong collinearity between the distance to sites of interest and the distance to the shelf (and, to a lesser extent, bathymetry). Because the two covariates are related, it may not be feasible to estimate their effects separately. In practice, we could for example use forward stepwise selection to determine which environmental variables to keep in the model.

With the method presented in Sections 5.4.2.2 and 5.4.2.3, we also derived an estimate and confidence bounds for the speed parameter γ^2, shown in Table 5.1. The value of the estimate is not readily interpretable as a measure of the speed of movement of the animals. Rather, it is a relative measure of speed, compared with the basic Langevin diffusion process (i.e. with $\gamma^2 = 1$; Equation 5.1). From the estimate, we can also derive a habitat-independent distribution of step lengths for any chosen time step. Indeed, under the assumption that the covariates have no effect, the utilisation distribution is flat, and the Langevin movement model becomes a simple Brownian motion. This Brownian motion has variance γ^2, so its transition density is

$$\boldsymbol{X}_{t+\Delta} | \{\boldsymbol{X}_t = \boldsymbol{x}_t\} \sim N(\boldsymbol{x}_t,\ \gamma^2 \Delta \boldsymbol{I}_2),$$

for $\Delta > 0$. In the absence of covariate effects, the distribution of step lengths over a time interval Δ is thus a Rayleigh distribution with scale parameter $\gamma\sqrt{\Delta}$. (The Rayleigh distribution, and its relation to the bivariate normal distribution, were presented in Section 3.5.1). Note that this distribution does not scale linearly with the time interval, but with its square root, due to the properties of the Brownian motion. It is not a distribution of speeds, but of step lengths, tied to a particular time interval. Figure 5.11 shows the estimated habitat-independent distribution of step lengths over 1-hour time intervals, from the model fitted to the sea lion tracks.

As illustrated in the simulations of Section 5.5.3, the estimation method that we presented for the Langevin movement model might not always perform well. In particular, information about habitat selection may be lost when the time intervals between observations become very long, and the selection parameters may be poorly estimated. To assess the model fitted to the sea lion tracks, we propose to use the Metropolis-adjusted Langevin algorithm outlined in Section 5.3.2.2. We simulated from the Metropolis-adjusted algorithm on the estimated

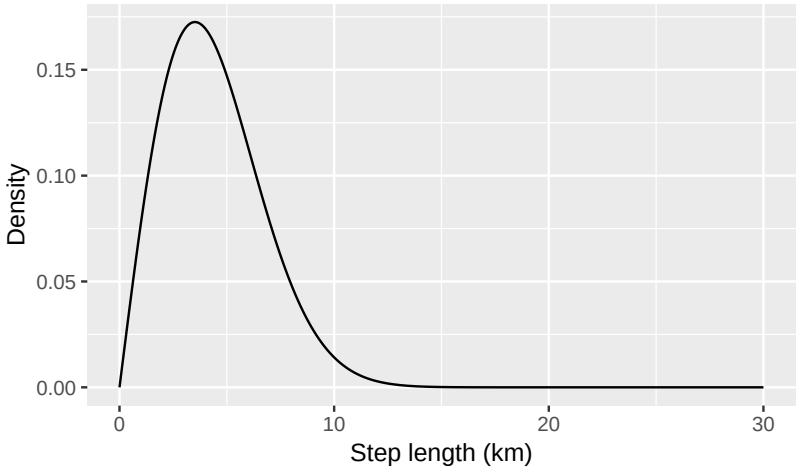


Fig. 5.11 Estimated distribution of step lengths for the sea lions, over 1-hour time intervals, in the absence of covariate effects.

utilisation distribution of the Steller sea lions, with the estimated speed parameter, and at the times of the observations. We then considered the acceptance rate of the algorithm as a measure of the accuracy of the approximation of the Langevin diffusion, in the context of this analysis. An acceptance rate close to 100% would suggest that the observed process captures the dynamics of a Langevin diffusion, whereas a low acceptance rate would suggest that the time scale of observation is too coarse to provide a good approximation of the process.

For each observed track, we simulated 50 trajectories from the Metropolis-adjusted Langevin algorithm on the time grid of the observations. Each simulation started from the initial location of the track, and we rejected simulations that left the study region (where the gradient of the covariate functions is undefined). Overall, the acceptance rate of the algorithm ranged between 93.2% and 99.1%, with a mean of 97.4%. This high acceptance rate seems to indicate that the Euler-Maruyama discretization is a good approximation of the Langevin process, at the time step of the observations. Further work is needed to develop a more formal framework of model checking, based on the acceptance probability of the Metropolis-adjusted Langevin algorithm.

5.7 Discussion

We have presented a new application for the Langevin diffusion process, to model animal movement and habitat selection in continuous time. The Langevin diffusion process is defined as the solution to a stochastic differential equation, and it has a known stationary distribution. The Langevin movement model is a continuous-time analogue of the MCMC

step selection model described in Chapters 3 and 4. Indeed, it is based on a stochastic process with known stationary distribution, and it can therefore describe long-term space use by animals as the consequence of short-term movements.

The continuous-time formulation of the Langevin movement model gives it an advantage over the discrete-time MCMC movement models. Indeed, all its parameters are independent of the temporal scale of observation. The habitat selection parameters $\boldsymbol{\beta}$ of the Langevin model describe a continuous-time process, rather than selection from one observed location to the next. Similarly, the movement parameter γ^2 quantifies instantaneous speed, and it is independent of the sampling interval. It is unclear whether the habitat selection estimates obtained from the two methods would be different. Indeed, although the MCMC movement model describes selection at the scale of the step (which depends on the time step of observation), the same parameters also model long-term habitat selection (which is independent of the time scale). The main limitation of the MCMC movement model, compared with its continuous-time counterpart, is that the movement and perception of the animal is modelled at the scale of the observed steps. This limitation is circumvented in the normal kernel model for irregular observations, that we presented in Section 3.6.3. The time-varying normal kernel model and the Langevin movement model could be compared in simulations, to investigate whether they lead to similar estimates of habitat selection.

We used the Euler-Maruyama discretization scheme to approximate the Langevin diffusion process, because the transition density of the process is not analytically tractable. This scheme is the most widely-used method to carry out inference for intractable diffusion processes, due to its simplicity (for applications in ecology, see e.g. Preisler et al., 2004; Russell et al., 2018). However, there exist other discretization schemes, which may better approximate the true underlying process. In the context of potential-based movement models, Gloaguen et al. (2018) showed that higher-order schemes could be used to obtain more reliable results. A particularly interesting alternative is the Ozaki discretization scheme. In the Ozaki discretization, the drift function of the diffusion process is approximated by a linear function between observations, rather than a by constant as in the Euler-Maruyama scheme (Equation 5.7). The Ozaki approximation of the drift term requires the evaluation of its partial derivatives, to obtain a linear approximation. In the context of the Langevin movement model, the drift function is proportional to $\nabla \log \pi$, such that we need to derive the second derivatives of $\log \pi$. If π is formulated as an exponential function, then the second derivatives of the covariate functions must be computed, following from Equation 5.4. As explained in Section 5.2, spatial covariates are usually measured on a discrete grid of regularly-spaced points, and they must be interpolated to obtain continuous functions. The derivatives and

second derivatives of the interpolated covariate functions must then be computed, possibly using numerical methods (although analytical formulas exist if bilinear interpolation is used). The second derivatives of the interpolated functions are susceptible to contain a greater error than the first derivatives, for simple interpolation methods, because the approximation of the curvature is worse than the approximation of the slope. This introduces an additional source of approximation in the Ozaki discretization, that is not present in the Euler-Maruyama approach, and which may cancel out some of the theoretical advantages of the more complex method. It is also worth noting that the computational cost would be greatly increased by the need to compute numerically the second derivatives of the covariate functions, and by the more complex transition equations (Gloaguen et al., 2018).

A major challenge for the application of the Langevin movement model is to assess the error of approximation due to the time discretization. Exact inference is not possible because the transition density of the Langevin diffusion process is not analytically tractable. We described the Euler-Maruyama discretization scheme, to define a simpler process that approximates the Langevin diffusion. The accuracy of the approximation degrades as the time intervals of discretization become longer. However, in practice, it is difficult to know how fine the time grid of observation must be to obtain accurate estimates. This may depend on the speed of movement of the animal, and on the shape of the target (utilisation) distribution. For example, high-frequency observations would be needed to detect environmental effects on the movements of an animal moving fast in a highly fragmented habitat. This is because information about local selection deteriorates at longer time scales. We suggested that the acceptance rate of the Metropolis-adjusted Langevin algorithm could provide a partial solution, to assess the discretization in the context of a real analysis. Indeed, we can simulate from the Metropolis-adjusted Langevin algorithm on the time grid of the observations, with the estimated speed parameter, and with the estimated utilisation distribution. The acceptance rate of the algorithm is a measure of similarity between the discretized process and the true Langevin process. It lies between 0 and 1, and a large value indicates that the discretized process is a good approximation of the true process. However, we have not determined the exact relationship between the acceptance rate and the accuracy of the estimates. Is an acceptance rate of 95% good enough? Or 99%? To answer this question, we must link the acceptance rate to the discrepancy between the estimated utilisation distribution and the true utilisation distribution. To achieve this, one possibility would be to combine the first two simulation studies presented in this chapter (Sections 5.5.1 and 5.5.2). We could simulate from the Metropolis-adjusted Langevin algorithm and from the "unadjusted" discretized Langevin diffusion, on various artificial utilisation distributions. Then, we could determine

whether there is a consistent relationship between the acceptance rate of the former and the accuracy of the empirical utilisation distribution of the latter.

In the analysis of Section 5.6, we used a two-stage approach to deal with the measurement error present in the Argos observations, similarly to Wilson et al. (2018). We first preprocessed the sea lion tracks with the continuous-time correlated random walk, to estimate the true locations of the animals. We then fitted the Langevin movement model to the filtered locations, i.e. we assumed that they were the true locations. This workflow is very common in analyses of noisy and irregular telemetry data (Patterson et al., 2010). However, there are several drawbacks to the two-stage approach. Indeed, it is difficult to propagate the uncertainty from the measurement error to the final parameter estimates (although multiple imputation could be used; see e.g. Scharf et al., 2017). Besides, the preprocessing stage does not account for the environmental effects that are modelled with the Langevin movement model in the second stage. A better approach would be to integrate the two stages into a state-space model that models measurement error and habitat selection jointly. We can combine the model of movement described in this chapter with an observation model. The state process of the full state-space model is the true location of the animal, and its dynamics are given by the transition density of the Langevin movement model, or a discretization of it (Equation 5.8). The observation process captures the measurement error, which is often assumed to be Gaussian. In this formulation, the observed location $\tilde{\boldsymbol{X}}_i$ and the true location $\boldsymbol{X}_i$ are related by the observation equation, $\tilde{\boldsymbol{X}}_i = \boldsymbol{X}_i + \boldsymbol{\eta}_i$, where $\boldsymbol{\eta}_i \sim N(\mathbf{0}, \sigma^2_{\text{obs}}\boldsymbol{I}_2)$ models the measurement error. The Euler-Maruyama discretization of the Langevin diffusion has a normal transition density, and a Kalman filter can be used to filter the locations and derive the likelihood of this hierarchical state-space model.

In the context of the MCMC step selection model, we found that misspecifications of the movement model can cause bias in the estimation of the habitat selection parameters (Section 4.7.6). A similar problem may arise for the Langevin movement model, if the observed movement does not satisfy the model assumptions. For example, in the Steller sea lion analysis, the Langevin diffusion process does not directly model the structure of the foraging trips observed in the data. More generally, the assumption that an animal's mean direction of movement (i.e. drift) is determined only by the local gradient of its utilisation distribution might be overly simplistic. A simulation study similar to that presented in Section 4.7.6 could be considered, to investigate the performance of the Langevin movement model when its assumptions are violated.

Chapter 6

State-switching continuous-time correlated random walks

This chapter explores a different modelling problem, related to the analysis of animal telemetry data. Chapters 3-5 addressed the estimation of habitat selection from correlated location data, to answer the questions "How does the animal use space?", and "Does the animal select particular habitats?". In this chapter, we describe a flexible continuous-time model to infer behavioural states from location data, to answer the question "What is the animal doing?". The work presented in this chapter is mostly independent from the rest of this thesis.

In Section 6.2, we introduce the model formulation, and we describe the Kalman filter for that model in Section 6.3. We describe a Bayesian method of inference in Section 6.4, to estimate movement characteristics and behavioural states from telemetry data. We assess the performance of the method in a simulation study in Section 6.5, and we showcase its use with the analysis of a grey seal movement track in Section 6.6. The material presented in this chapter is described in Michelot and Blackwell (2019).

6.1 Background

The collection of large high-resolution animal tracking data sets has motivated the development of a wide range of statistical methods (Patterson et al., 2017; Hooten et al., 2017). Many of them are based on the random walk and its extensions (Turchin, 1998; Codling et al., 2008). In its simplest formulation, the random walk is a discrete-time Markov process with independent normal increments. This model only has one parameter, the variance of

the transition density, which determines the speed of movement. Modern telemetry data are collected at high temporal resolutions, and often display autocorrelation in the speed and direction of movement. Correlated random walk models were developed to capture this movement persistence. They are most often described by a distribution of "step lengths", the distances between successive locations, and of "turning angles", the angles between successive directions (see Section 1.2, and Siniff and Jessen, 1969; Kareiva and Shigesada, 1983; Bovet and Benhamou, 1988; Morales et al., 2004). The distribution of step lengths captures the speed of movement, and the distribution of turning angles measures the autocorrelation in the direction of movement. For example, if the distribution of turning angles is very concentrated around zero, successive directions will be very similar, corresponding to persistent movement. This model is usually formulated in discrete time, and the distributions of these metrics of movement strongly depend on the sampling rate (Codling and Hill, 2005; Schlägel and Lewis, 2016). As discussed in Section 1.2 and in Chapter 5, results obtained from discrete-time movement models are tied to the time scale of the data.

Johnson et al. (2008a) introduced the continuous-time correlated random walk (CTCRW), in which the velocity of the animal is formulated as an Ornstein-Uhlenbeck process. Through the velocity, this model incorporates autocorrelation into both the speed and the direction of the movement, similarly to discrete-time correlated random walks based on step lengths and turning angles. It describes the velocity and the movement in continuous time, and therefore allows for scale-free inference. Johnson et al. (2008a) formulated the CTCRW as a state-space model, making fast inference possible through the Kalman filter, and made it available in the R package crawl (Johnson and London, 2018). Fleming et al. (2017) extended this implementation to a wider family of diffusion processes, including their "OUF" model of correlated movement around a centre of attraction.

Random walks have been used as "building blocks" for more complex, multistate, models. These state-switching models describe animal movements as the outcome of several distinct behaviours, e.g. "foraging", "resting", "exploring", based on the notion that the behavioural states of the animal differ noticeably in terms of some metrics of the movement, e.g. speed or sinuosity (Blackwell, 1997; Morales et al., 2004). The advantage of multistate time series models over simpler clustering methods is that they account for the temporal autocorrelation in the movement behaviours, and provide a mechanistic description of the movement process (Edelhoff et al., 2016). Although multistate models have received a lot of attention in discrete-time formulations, with the growing popularity of hidden Markov models (Patterson et al., 2009; Langrock et al., 2012), they have been underutilised in continuous-time approaches. Blackwell (1997) introduced a continuous-time multistate movement model, where the

location of the animal is modelled with an Ornstein-Uhlenbeck process. That model does not directly capture the movement persistence in speed and direction, which makes its application limited for high-frequency tracking data. More recently, Parton and Blackwell (2017) described a multistate approach in which the speed and the bearing of the animal are modelled in continuous time, with diffusion processes, analogously to discrete-time models based on step lengths and turning angles. However, their method requires computationally-costly numerical simulations to reconstruct the movement path at a fine time scale, a disadvantage in dealing with large tracking data sets. McClintock et al. (2014) presented a multistate analysis based on the CTCRW, but they constrained the state process to be constant over each time interval between two observations. Therefore, they did not carry out exact inference from the continuous-time model. Gurarie et al. (2017) recently reviewed the use of the CTCRW model, and related velocity-based models, for the analysis of animal tracking data. They proposed a method based on change point analysis to segment movement tracks into behavioural phases. Although they can be a useful tool of classification, change point approaches do not provide a mechanistic understanding of the behavioural state process.

Alternatively, McClintock (2017) suggested a two-stage approach based on multiple imputation methods, to estimate behavioural states from irregular or noisy tracking data. In that approach, a one-state continuous-time model (such as the CTCRW) is first fitted to the data. A large number, m, of possible realisations of the movement process are simulated from the model on a regular time grid. Then, a hidden Markov model is fitted to each realisation, to estimate the state-switching dynamics. The m sets of estimates are pooled, such that the resulting model takes into account the uncertainty in the locations. Note that, since the realisations are generated without taking into account the possible behaviours, this is not fully equivalent to fitting a multistate CTCRW model. In particular, if the one-state model fails to capture the behavioural heterogeneity in the movement, the simulated realisations may not correctly reflect the uncertainty in the continuous trajectory.

Here, we extend the framework of Johnson et al. (2008a) to incorporate behavioural states directly into the CTCRW framework, with an underlying continuous-time Markov process. The state-switching CTCRW offers a rigorous approach to model behavioural heterogeneity and movement persistence, two common features in modern telemetry data collected at high frequency over long periods of time. It can be used with irregularly-sampled movement data without the need to interpolate the locations, and can incorporate measurement error. We present the model formulation and describe a Bayesian estimation method to infer hidden states and movement parameters in this framework. We investigate the performance of the method in a simulation study, and demonstrate that it can be used to recover estimates of the

states and movement parameters, from irregular location data. We analyse a trajectory of grey seal (*Halichoerus grypus*) with a 2-state CTCRW model, and obtain posterior samples of the state-dependent movement parameters and of the unobserved state sequence. We explain how the movement parameters differ in the two states, and how they can be interpreted as measures of the animal's speed and movement persistence.

6.2 Model formulation

In the continuous-time correlated random walk (CTCRW), the velocity of an animal is modelled with an Ornstein-Uhlenbeck model. We described the Ornstein-Uhlenbeck process in the general case in Section 2.4. Here, we derive the transition density of the Ornstein-Uhlenbeck process, and of the CTCRW process. The transition density of the CTCRW was already described by Johnson et al. (2008a), and we only include its derivation here for clarity of presentation.

6.2.1 Ornstein-Uhlenbeck process

We consider a one-dimensional Ornstein-Uhlenbeck process $(V_t)_{t\geq 0}$, defined as the solution to

$$dV_t = \beta(\gamma - V_t)dt + \sigma dW_t, \tag{6.1}$$

where W_t is a standard Wiener process.

6.2.1.1 Solution

We can solve for the process (V_t) in Equation 6.1. If we define $Y_t = V_t - \gamma$, we can write this equation as

$$dY_t = -\beta Y_t dt + \sigma dW_t, \tag{6.2}$$

and note that we have

$$d(e^{\beta t}Y_t) = \beta e^{\beta t}Y_t dt + e^{\beta t}dY_t. \tag{6.3}$$

Then, from Equations 6.2 and 6.3,

$$d(e^{\beta t}Y_t) = e^{\beta t}\sigma dW_t.$$

Integrating both sides between t and $t+\delta$, we find

$$e^{\beta(t+\delta)}Y_{t+\delta} - e^{\beta t}Y_t = \sigma \int_{s=t}^{t+\delta} e^{\beta s} dW_s.$$

So, the process (Y_t) satisfies

$$Y_{t+\delta} = e^{\beta\delta}Y_t + \sigma \int_{s=t}^{t+\delta} e^{-\beta(t+\delta-s)} dW_s.$$

Finally, substituting $V_t - \gamma$ for Y_t, we obtain the solution of the Ornstein-Uhlenbeck equation,

$$V_{t+\delta} = \gamma + e^{-\beta\delta}(V_t - \gamma) + \sigma \int_{s=t}^{t+\delta} e^{-\beta(t+\delta-s)} dW_s. \tag{6.4}$$

6.2.1.2 Transition density

We define

$$\zeta(\delta) = \sigma \int_{s=t}^{t+\delta} e^{-\beta(t+\delta-s)} dW_s. \tag{6.5}$$

By properties of the Itô integral, $\zeta(\delta)$ is a Gaussian random variable with mean 0. The transition density of the Ornstein-Uhlenbeck process is therefore normal, with mean

$$E(V_{t+\delta}|V_t = v_t) = \gamma + e^{-\beta\delta}(v_t - \gamma).$$

The variance of the transition density is given by the variance of the error term $\zeta(\delta)$,

$$\text{Var}(\zeta(\delta)) = E\left[\zeta(\delta)^2\right] = \sigma^2 E\left[\left(\int_{s=t}^{t+\delta} e^{-\beta(t+\delta-s)} dW_s\right)^2\right].$$

Using the Itô isometry (Øksendal, 2003), this becomes

$$\begin{aligned}
\text{Var}(\zeta(\delta)) &= \sigma^2 \int_{s=t}^{t+\delta} e^{-2\beta(t+\delta-s)} ds \\
&= \sigma^2 \left[\frac{e^{-2\beta(t+\delta-s)}}{2\beta}\right]_{s=t}^{t+\delta} \\
&= \frac{\sigma^2}{2\beta}(1 - e^{-2\beta\delta}).
\end{aligned}$$

We can write the transition density of the Ornstein-Uhlenbeck process as

$$V_{t+\delta}|\{V_t = v_t\} \sim N\left(\gamma + e^{-\beta\delta}(v_t - \gamma),\ \frac{\sigma^2}{2\beta}(1 - e^{-2\beta\delta})\right). \tag{6.6}$$

This equation can be used to simulate forward from the Ornstein-Uhlenbeck process. The long-term distribution of (V_t) is the limit of Equation 6.6 as the time interval δ goes to infinity. We find

$$V_t \sim N\left(\gamma,\ \frac{\sigma^2}{2\beta}\right). \tag{6.7}$$

6.2.1.3 Multivariate isotropic case

These results can directly be extended to the multivariate isotropic case, which will be of interest in what follows. The isotropic Ornstein-Uhlenbeck process in k dimensions is the solution to the equation

$$d\boldsymbol{V}_t = \beta(\boldsymbol{\gamma} - \boldsymbol{V}_t)dt + \sigma d\boldsymbol{W}_t, \tag{6.8}$$

where $(\boldsymbol{V}_t)$ and $(\boldsymbol{W}_t)$ are k-dimensional processes, and $\boldsymbol{\gamma}$ is the k-dimensional mean of the process. The process is isotropic because β and σ are still scalar parameters (rather than matrices). The multivariate process can be regarded as k univariate Ornstein-Uhlenbeck processes, and Equations 6.4 and 6.6 hold in each dimension. The univariate and bivariate (isotropic) Ornstein-Uhlenbeck processes are illustrated in Figure 6.1.

The bivariate Ornstein-Uhlenbeck process has been considered to model the location of a moving animal in space (Dunn and Gipson, 1977; Blackwell, 1997). In particular, it is useful to capture home range behaviour, due to its attraction towards a point in space. The process is stationary, with long-term distribution given in Equation 6.7. In that model, the stationary distribution of the Ornstein-Uhlenbeck process can be viewed as the utilisation distribution of the animal. However, a limitation of the Ornstein-Uhlenbeck to model the position process of an animal is that it does not capture movement persistence. Similarly to Brownian motion, it does not describe "smooth" trajectories. As a consequence, it is not a good model for modern high-resolution telemetry data, which often display strong persistence in the speed and the direction of movement. This motivates the introduction of the continuous-time correlated random walk, which describes persistent movement.

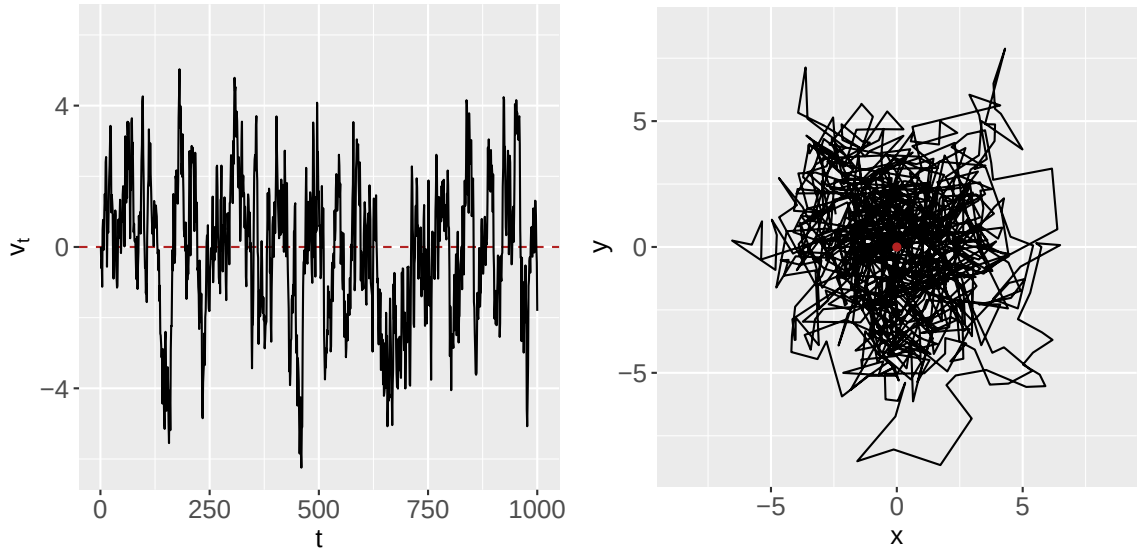


Fig. 6.1 Illustration of the Ornstein-Uhlenbeck process in one dimension (left) and two dimensions (right). The long-term mean is shown by the red dashed line for the one-dimensional process, and by the red dot for the two-dimensional process.

6.2.2 Continuous-time correlated random walk

6.2.2.1 Definition

The continuous-time correlated random walk (CTCRW) builds on the Ornstein-Uhlenbeck process, and was introduced by Johnson et al. (2008a) to model persistent animal movement. We denote by $\boldsymbol{X}_t = (X_t, Y_t)^\top$ the location of the animal at time t, and $\boldsymbol{V}_t = (V_t^x, V_t^y)^\top$ its velocity, linked by the equation

$$d\boldsymbol{X}_t = \boldsymbol{V}_t dt. \tag{6.9}$$

In the CTCRW, the velocity of the animal is modelled by a bivariate Ornstein-Uhlenbeck process, defined as the solution of Equation 6.8. The Ornstein-Uhlenbeck process is here used to model the autocorrelation in the velocity, and the reversion to a mean velocity. We will sometimes refer to the location process $(\boldsymbol{X}_t)$ as an integrated Ornstein-Uhlenbeck process, to indicate that its derivative (with respect to time) is an Ornstein-Uhlenbeck process.

In practice, the mean velocity parameter $\boldsymbol{\gamma}$ is usually taken to be zero. If $\boldsymbol{\gamma}$ is not zero, it introduces a systematic drift in the animal's movement, towards a particular bearing. For example, Johnson et al. (2008a) modelled drift in the trajectories of migrating northern fur seals. In the following, we will focus on the case where $\boldsymbol{\gamma} = \mathbf{0}$. We consider that β and σ are scalar parameters, corresponding to the isotropic case, but they could be taken as matrices for

a more general formulation (Blackwell, 2003; Gurarie et al., 2017). Johnson et al. (2008a) developed a method to estimate the movement parameters β and σ of the CTCRW from observed telemetry data, that we discuss and extend in Section 6.3.

The parameters of the CTCRW can be linked to the speed and sinuosity of the animal's movement, and therefore used for the biological interpretation of tracking data. Indeed, Gurarie et al. (2017) presented an alternative parametrisation of the CTCRW, defined with $\tau = 1/\beta$ and $\nu = \sqrt{\pi}\sigma/(2\sqrt{\beta})$. In this formulation, the parameter $\tau > 0$ is the time interval over which the autocorrelation function of the velocity process decreases by a factor e, and it is sometimes called the "relaxation time" of the process (Gillespie, 1996). Larger values of τ (corresponding to smaller values of β) indicate longer-term persistence in the speed and direction of the animal's movement. The autocorrelation function of the velocity process decreases to 0.05 over a time interval of length 3τ, because $e^{-3} \approx 0.05$. For most practical purposes, $\boldsymbol{V}_t$ and $\boldsymbol{V}_{t+3\tau}$ can therefore be regarded as approximately independent (Johnson et al., 2008a). The parameter $\nu > 0$ is the long-term mean of the speed of movement of the animal. The derivation of τ and ν is given in Appendix D. With this convenient formulation, the model offers a very useful framework to quantify movement characteristics of animals. Gurarie and Ovaskainen (2011) argued that the important biological features of animal movement could generally be summarised by a "characteristic temporal scale" and a "characteristic spatial scale" of movement. In our approach, these are given by τ and ν, respectively.

6.2.2.2 Solution

The CTCRW formulation is very convenient because it is possible to derive the transition densities of the velocity process $(\boldsymbol{V}_t)$ and of the position process $(\boldsymbol{X}_t)$ analytically (Johnson et al., 2008a). We found the transition density of the velocity process in Section 6.2.1, and we can now solve for the location process $(\boldsymbol{X}_t)$. As we will see, the transition density can be used to jointly simulate from the velocity and location processes, and for inference.

As before, the process is isotropic, and it is sufficient to solve for the one-dimensional process, that we denote $(X_t)_{t\geq 0}$. Integrating both sides of $dX_t = V_t dt$ between t and $t+\delta$, and using the solution found in Equation 6.4, we have

$$\begin{aligned} X_{t+\delta} - X_t &= \int_{s=t}^{t+\delta} V_s ds \\ &= \int_{s=t}^{t+\delta} \left(e^{-\beta(s-t)} V_t + \sigma \int_{u=t}^{s} e^{-\beta(s-u)} dW_u \right) ds. \end{aligned}$$

Thus,

$$\begin{aligned} X_{t+\delta} &= X_t + V_t \int_{s=t}^{t+\delta} e^{-\beta(s-t)} ds + \sigma \int_{s=t}^{t+\delta} \int_{u=t}^{s} e^{-\beta(s-u)} dW_u ds \\ &= X_t + V_t \left[-\frac{e^{-\beta(s-t)}}{\beta} \right]_{s=t}^{t+\delta} + \sigma \int_{u=t}^{t+\delta} \int_{s=u}^{t+\delta} e^{-\beta(s-u)} ds dW_u \\ &= X_t + \left(\frac{1 - e^{-\beta\delta}}{\beta} \right) V_t + \sigma \int_{u=t}^{t+\delta} \left[-\frac{e^{-\beta(s-u)}}{\beta} \right]_{s=u}^{t+\delta} dW_u. \end{aligned}$$

Finally, we find the solution to the integrated Ornstein-Uhlenbeck equation,

$$X_{t+\delta} = X_t + \left(\frac{1 - e^{-\beta\delta}}{\beta} \right) V_t + \frac{\sigma}{\beta} \int_{u=t}^{t+\delta} (1 - e^{-\beta(t+\delta-u)}) dW_u.$$

6.2.2.3 Transition density

Define $\xi(\delta)$ as the Gaussian error term

$$\xi(\delta) = \frac{\sigma}{\beta} \int_{u=t}^{t+\delta} (1 + e^{-\beta(t+\delta-u)}) dW_u, \tag{6.10}$$

with mean zero. The transition density of the integrated Ornstein-Uhlenbeck process is also Gaussian, with mean

$$E(X_{t+\delta} | X_t = x_t, V_t = v_t) = x_t + \left(\frac{1 - e^{-\beta\delta}}{\beta} \right) v_t,$$

and with variance given by $\text{Var}(\xi(\delta))$,

$$\text{Var}(\xi(\delta)) = \left(\frac{\sigma}{\beta} \right)^2 E \left[\left(\int_{u=t}^{t+\delta} (1 - e^{-\beta(t+\delta-u)}) dW_u \right)^2 \right].$$

Using the Itô isometry,

$$\begin{aligned} \text{Var}(\xi(\delta)) &= \left(\frac{\sigma}{\beta} \right)^2 \int_{u=t}^{t+\delta} (1 - e^{-\beta(t+\delta-u)})^2 du \\ &= \left(\frac{\sigma}{\beta} \right)^2 \int_{u=t}^{t+\delta} (1 + e^{-2\beta(t+\delta-u)} - 2e^{-\beta(t+\delta-u)}) du \end{aligned}$$

$$= \left(\frac{\sigma}{\beta}\right)^2 \left[u + \frac{e^{-2\beta(t+\delta-u)}}{2\beta} - \frac{2e^{-\beta(t+\delta-u)}}{\beta}\right]_{u=t}^{t+\delta}$$
$$= \left(\frac{\sigma}{\beta}\right)^2 \left(\delta + \frac{1-e^{-2\beta\delta}}{2\beta} - \frac{2(1-e^{-\beta\delta})}{\beta}\right).$$

Finally, we obtain the transition density of the integrated Ornstein-Uhlenbeck process,

$$X_{t+\delta}|\{X_t = x_t, V_t = v_t\} \sim$$
$$N\left(x_t + \left(\frac{1-e^{-\beta\delta}}{\beta}\right)v_t,\ \left(\frac{\sigma}{\beta}\right)^2\left(\delta + \frac{1-e^{-2\beta\delta}}{2\beta} - \frac{2(1-e^{-\beta\delta})}{\beta}\right)\right).$$

6.2.2.4 Covariance of location and velocity

For simulation and inference, we need the joint transition density of (V_t) and (X_t), and so we also need the covariance of the location process and the velocity process, i.e. the covariance of the error terms $\xi(\delta)$ and $\zeta(\delta)$.

$$\begin{aligned}\mathrm{Cov}(\zeta(\delta), \xi(\delta)) &= E(\zeta(\delta)\xi(\delta)) \\ &= E\left[\left(\sigma\int_{s=t}^{t+\delta} e^{-\beta(t+\delta-s)}dW_s\right)\left(\frac{\sigma}{\beta}\int_{s=t}^{t+\delta}(1-e^{-\beta(t+\delta-s)})dW_s\right)\right].\end{aligned}$$

Using the Itô isometry,

$$\begin{aligned}\mathrm{Cov}(\zeta(\delta), \xi(\delta)) &= \frac{\sigma^2}{\beta}\int_{s=t}^{t+\delta} e^{-\beta(t+\delta-s)}(1-e^{-\beta(t+\delta-s)})ds \\ &= \frac{\sigma^2}{\beta}\int_{s=t}^{t+\delta}(e^{-\beta(t+\delta-s)} - e^{-2\beta(t+\delta-s)})ds \\ &= \frac{\sigma^2}{\beta}\left[\frac{e^{-\beta(t+\delta-s)}}{\beta} - \frac{e^{-2\beta(t+\delta-s)}}{2\beta}\right]_{s=t}^{t+\delta} \\ &= \frac{\sigma^2}{\beta}\left(\frac{1-e^{-\beta\delta}}{\beta} - \frac{1-e^{-2\beta\delta}}{2\beta}\right).\end{aligned}$$

We find the covariance,

$$\text{Cov}(\zeta(\delta), \xi(\delta)) = \frac{\sigma^2}{2\beta^2}\left(1 - 2e^{-\beta\delta} + e^{-2\beta\delta}\right). \tag{6.11}$$

6.2.2.5 Simulation from the CTCRW

From the equations found above, we can simulate from the integrated Ornstein-Uhlenbeck process (i.e. from the CTCRW). Here, we describe the one-dimensional case, and the multivariate isotropic case can be obtained by simulating independently in each dimension.

We jointly simulate the velocity and location processes at n time points $\{t_1, t_2, \ldots, t_n\}$, and denote $\Delta_i = t_i - t_{i-1}$ the time intervals. We denote V_i and X_i the velocity and the location at time t_i, respectively. We initialise $V_1 = v_1$ and $X_1 = x_1$ and, for $i = 2, \ldots, n$, we simulate from the CTCRW as follows

$$\begin{pmatrix} V_i \\ X_i \end{pmatrix} \sim N\left[\boldsymbol{T}_i \begin{pmatrix} v_{i-1} \\ x_{i-1} \end{pmatrix}, \boldsymbol{Q}_i\right], \tag{6.12}$$

conditionally on $V_{i-1} = v_{i-1}$ and $X_{i-1} = x_{i-1}$, and where

$$\boldsymbol{T}_i = \begin{pmatrix} 1 & (1 - e^{-\beta\Delta_i})/\beta \\ 0 & e^{-\beta\Delta_i} \end{pmatrix}, \quad \boldsymbol{Q}_i = \begin{pmatrix} \text{Var}(\xi(\Delta_i)) & \text{Cov}(\xi(\Delta_i), \zeta(\Delta_i)) \\ \text{Cov}(\xi(\Delta_i), \zeta(\Delta_i)) & \text{Var}(\zeta(\Delta_i)) \end{pmatrix}.$$

Figure 6.2 displays trajectories simulated from the univariate and bivariate integrated Ornstein-Uhlenbeck processes, based on Equation 6.12. Unlike the Ornstein-Uhlenbeck process, the integrated Ornstein-Uhlenbeck process is not stationary, and it does not have a centre of attraction. It is smoother, however, as it captures autocorrelation in the velocity (in addition to autocorrelation in the location).

6.2.3 Multistate model

6.2.3.1 Formulation

In this work, we use the CTCRW as a building block for more complex and realistic movement models. Multistate models of animal movement have been developed to account for behavioural heterogeneity. In the most common formulation, a (discrete- or continuous-time) Markov process models switches between discrete "behavioural" states, on which depend the parameters of the movement process (Blackwell, 1997; Morales et al., 2004).

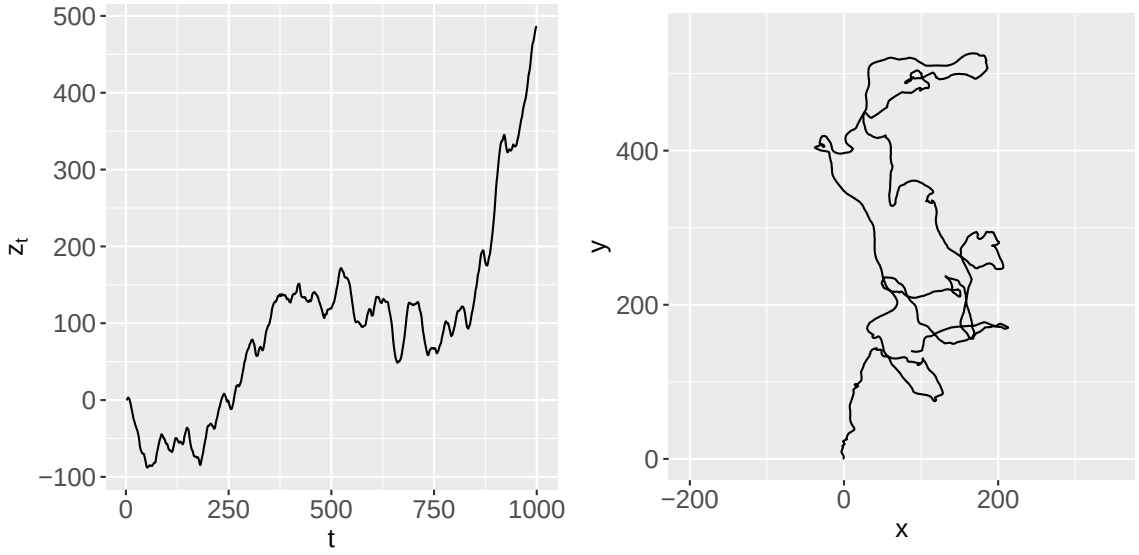


Fig. 6.2 Illustration of the integrated Ornstein-Uhlenbeck process (or continuous-time correlated random walk), in one dimension (left) and two dimensions (right).

Following this idea, we introduce a N-state continuous-time Markov process $(S_t)_{t\geq 0}$, defined by its infinitesimal generator matrix,

$$\mathbf{\Lambda} = \begin{pmatrix} -\lambda_1 & \lambda_{12} & \cdots & \lambda_{1N} \\ \lambda_{21} & -\lambda_2 & \cdots & \lambda_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{N1} & \lambda_{N2} & \cdots & -\lambda_N \end{pmatrix},$$

where $\forall i \in 1, \ldots, N,\ \lambda_i = \sum_{j\neq i} \lambda_{ij}$. At any time $t \geq 0$, the discrete state S_t takes one of N values $\{1, \ldots, N\}$, typically used as proxies for the behavioural states of the animal (e.g. "foraging", "exploring"). In the following, we find it convenient to parameterise the generator matrix as

$$\mathbf{\Lambda} = \begin{pmatrix} -\lambda_1 & \lambda_1 p_{12} & \cdots & \lambda_1 p_{1N} \\ \lambda_2 p_{21} & -\lambda_2 & \cdots & \lambda_2 p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_N p_{N1} & \lambda_N p_{N2} & \cdots & -\lambda_N \end{pmatrix} \tag{6.13}$$

where $\lambda_i > 0$ is the rate of transition out of state i, and the $p_{ij} \in [0, 1]$ are the transition probabilities out of state i. For each state i, they satisfy $\sum_j p_{ij} = 1$. The generator matrix is the continuous-time analogue of the transition probability matrix used in hidden Markov models, and its entries determine the state-switching dynamics. In particular, as

a consequence of the Markov property, the dwell times in state i follow an exponential distribution with rate λ_i.

We now consider that the parameters of the CTCRW model (β and σ in Equation 6.8) are state-dependent, so that each behavioural state can be associated with a different type of movement. Using the notation introduced in Section 6.2.2, the multistate CTCRW model is defined by

$$\begin{cases} d\boldsymbol{X}_t = \boldsymbol{V}_t dt, \\ d\boldsymbol{V}_t = -\beta_{S_t}\boldsymbol{V}_t dt + \sigma_{S_t} d\boldsymbol{W}_t. \end{cases}$$

This can be viewed as a higher-dimension continuous-time Markov process composed of a continuous component (the location and velocity processes) and a discrete component (the discrete state process), as described e.g. by Berman (1994).

The N-state CTCRW model has $2N$ movement parameters, $(\beta_1, \ldots, \beta_N, \sigma_1, \ldots, \sigma_N)$, and $N(N-1)$ transition rates. Each state $j \in \{1, \ldots, N\}$ is characterised by a particular type of movement, determined by the corresponding parameters β_j and σ_j. In Section 6.2.2.1, we linked the parameters of the CTCRW process to measures of the movement persistence, and of the speed of the animal. Different combinations of movement parameters can therefore capture different types of movement, and a wide range of movement behaviours.

In Sections 6.3 and 6.4, we develop a method of Bayesian inference for the multistate CTCRW model, to estimate all the model parameters, as well as the state process (S_t).

6.2.3.2 Simulation from the multistate CTCRW

We can simulate from the multistate CTCRW using a method similar to the one outlined in the single-state case in Section 6.2.2.5. In addition to the velocity and location processes, we must now also simulate from the state process. The state process is a continuous-time Markov process, with the generator matrix given in Equation 6.13. To simulate from the state process, we use the property that waiting times in state $j \in \{1, \ldots, N\}$ follow an exponential distribution with rate λ_j.

Let $\{t_1, \ldots, t_n\}$ be the time grid on which we wish the simulate a movement track. We first simulate the state process over $[t_1, t_n]$. We initialise the state process to $s_1 = j$ at time $\psi_1 = t_1$. Then, we simulate the state process by iterating the following steps, starting from $i = 2$, and until $\psi_i \geq t_n$.

1. Generate the waiting time $d \sim Exp(\lambda_j)$.

2. Derive the time of the next transition as $\psi_i = \psi_{i-1} + d$.
3. Pick s_{ψ_i} from $\{1, \ldots, N\} \setminus \{j\}$, with probabilities given by the transition probabilities p_{jk} out of state j.
4. Update the current state $j = s_{\psi_i}$, and increment the iterator $i = i + 1$.

From these steps, we obtain a sequence of transition times, and of states, between the times t_1 and t_n. For any i, the state process is constant over the time interval $[\psi_i, \psi_{i+1}]$. We can therefore use the simulation procedure described in Section 6.2.2.5 in the single-state CTCRW, to simulate from ψ_i to ψ_{i+1}, starting from $\psi_1 = t_1$. In the multistate CTCRW model, the movement parameters depend on the current state, so we use the substitute β_j for β and σ_j for σ, where $s_{\psi_i} = j$.

Figure 6.3 shows a trajectory simulated from a 2-state CTCRW model. The two states are characterised by different movement parameters β and σ, and they display different movement patterns. In this simulated example, the movement in state 1 is slow and tortuous, and the movement in state 2 is faster and more directed.

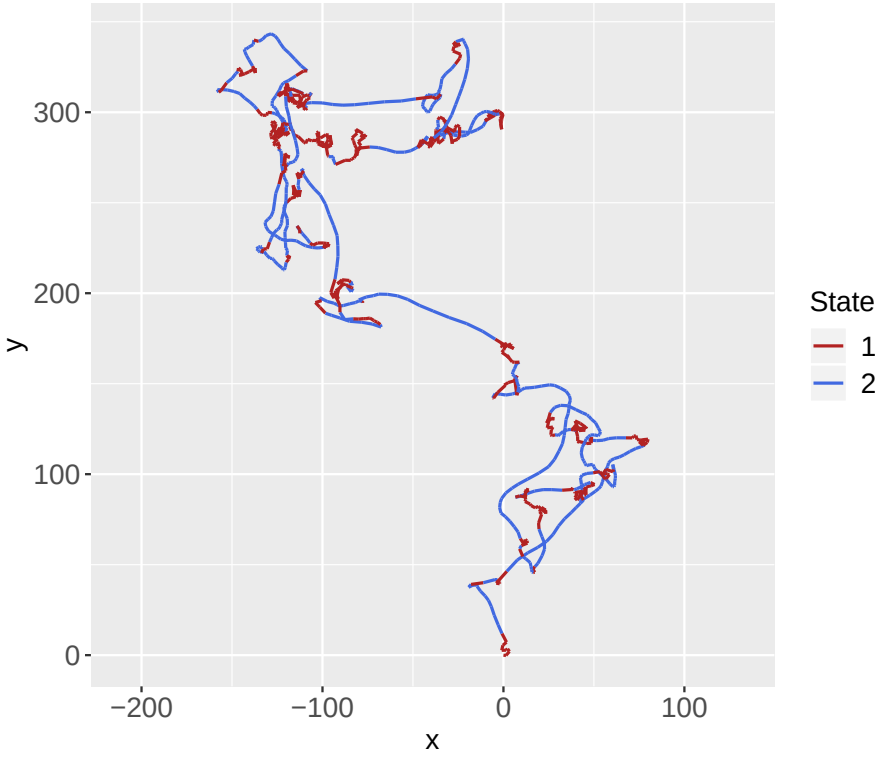


Fig. 6.3 Illustration of the multistate continuous-time correlated random walk.

6.3 The Kalman filter

6.3.1 State-space model formulation

We present a method to evaluate the likelihood of the multistate CTCRW model, given a reconstruction of the underlying state process. Johnson et al. (2008a) formulated the single-state CTCRW as a state-space model, and implemented a Kalman filter to obtain the likelihood of observed movement trajectories. With only minor changes, it can be extended to evaluate the likelihood of the multistate model, conditionally on the behavioural state sequence. The behaviours of the animal are typically not known, and will thus need to be imputed. One method of reconstructing the state sequence is presented in Section 6.4.

State-space models are a very wide class of time series models (Durbin and Koopman, 2012). They are often used in statistics to estimate the dynamics of a phenomenon that cannot be observed directly. They formulate the evolution of a system composed of two processes: an unobserved state process, and an observation process. The state process describes the hidden phenomenon, and the observation process links the state process to an observable output. In movement ecology, state-space models have been used for a long time to analyse telemetry data that include measurement error (Jonsen et al., 2003; Patterson et al., 2008). In that context, the hidden process models the movement of the animal, and the observation process models the measurement error. Note that state-space models generally have the same dependence structure as hidden Markov models, that we described in Section 4.6. Some use the two terms interchangeably, but hidden Markov models usually refer more specifically to state-space models with a discrete state space (Rabiner, 1989; Zucchini et al., 2016). Following Johnson et al. (2008a), the CTCRW can be written as a state-space model, where the observation process is the animal's observed location (possibly including measurement error), and the continuous state process is composed of both the true location and the velocity of the animal.

We consider a data set of observed locations, augmented with the times of the reconstructed behavioural transitions. The locations associated with the transitions are generally not observed, and they are thus treated as missing data. We denote by $\{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n\}$ the augmented sequence of observed locations, possibly obtained with measurement error, and $\{t_1, \ldots, t_n\}$ the associated times. We denote by s_i the (imputed) behavioural state between t_i and t_{i+1}. We denote by $\boldsymbol{\omega}_i = (x_i, v_i^x, y_i, v_i^y)^\top$ the continuous state vector at time t_i, where $\boldsymbol{x}_i = (x_i, y_i)$ is the true location of the animal, and $\boldsymbol{v}_i = (v_i^x, v_i^y)$ is its velocity. We define $\Delta_i = t_{i+1} - t_i$ the time intervals, and $\xi_i = \xi(\Delta_i)$ and $\zeta_i = \zeta(\Delta_i)$ the stochastic terms of the transition

densities for the location and the velocity, respectively (Equations 6.5 and 6.10, substituting β_{s_i} for β and σ_{s_i} for σ).

Note that both s_i and $\boldsymbol{\omega}_i$ are referred to as "state", in state-switching models and state-space models, respectively. Here, we combine both and, to avoid confusion, we will refer to $\boldsymbol{\omega}_i$ as the "continuous" state of the process (as opposed to the "discrete" behavioural state s_i).

The observation equation of the CTCRW is

$$\tilde{\boldsymbol{x}}_i = \boldsymbol{Z}\boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i, \qquad \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{H}_i),$$

where $\boldsymbol{H}_i$ is the 2×2 measurement error covariance matrix, and

$$\boldsymbol{Z} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

That is, the observed location $\tilde{\boldsymbol{x}}_i$ is the sum of the true location $\boldsymbol{x}_i = (x_i, y_i)$ and an error term $\boldsymbol{\varepsilon}_i$. Using block matrix notation, the continuous state equation is

$$\boldsymbol{\omega}_{i+1} = \begin{pmatrix} \boldsymbol{T}_i & \mathbf{0} \\ \mathbf{0} & \boldsymbol{T}_i \end{pmatrix} \boldsymbol{\omega}_i + \boldsymbol{\eta}_i, \qquad \boldsymbol{\eta}_i \sim N\left[\mathbf{0}, \begin{pmatrix} \boldsymbol{Q}_i & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}_i \end{pmatrix}\right]$$

with

$$\boldsymbol{T}_i = \begin{pmatrix} 1 & (1 - e^{-\beta_{s_i}\Delta_i})/\beta_{s_i} \\ 0 & e^{-\beta_{s_i}\Delta_i} \end{pmatrix}, \quad \boldsymbol{Q}_i = \begin{pmatrix} \text{Var}(\xi_i) & \text{Cov}(\xi_i, \zeta_i) \\ \text{Cov}(\xi_i, \zeta_i) & \text{Var}(\zeta_i) \end{pmatrix},$$

The variances of ξ_i and ζ_i are given in Equation 6.5 and 6.10, and their covariance is given in Equation 6.11. Note that the continuous state equation is the same equation that we presented in Section 6.2.2.5 to simulate from the CTCRW.

6.3.2 Kalman equations

The Kalman filter is a set of recursive equations to carry out inference from linear Gaussian state-space models (Kalman, 1960). A state-space model is linear if its state and observation equations are linear; it is Gaussian if the transition density of the state process is normal, and if the observation noise is normal. The multistate CTCRW model, formulated in Section 6.3.1, satisfies both conditions. The Kalman filter is used to learn about the (continuous) state process, i.e. the hidden phenomenon of interest, from a set of observations. The Kalman filter predicts the continuous state of the process, conditional on all past observations.

The prediction is updated with each new observation, based on the state equation and the observation equation. Using the assumptions made on the state and observation processes (i.e. normality), the likelihood of the observations under the state-space model can be calculated as a by-product of the Kalman filter iterations. The Kalman filter efficiently integrates over the continuous state to evaluate the likelihood, and it has largely contributed to the popularity of state-space models. It is conceptually similar to the forward algorithm used in hidden Markov models (described in Section 4.6). The Kalman filter is here used to integrate over both measurement error and unobserved velocities simultaneously.

Although it is not needed to fit the model, we can also implement the so-called Kalman smoother algorithm. The Kalman smoother provides an estimate $\hat{\boldsymbol{\omega}}_i$ of the continuous state at each time step, as well as the covariance matrix $\hat{\boldsymbol{\Sigma}}_i$ of the estimate, conditional on all observations (rather than conditional on past observations only). It consists of a set of backwards recursive equations, starting from the last observation and iterating back until the first observation.

In this section, we give the equations of the Kalman filter and smoother, for the multistate CTCRW model. These equations can be found for the single-state CTCRW in Appendix B of Johnson et al. (2008a), and in the case of a general state-space model in Durbin and Koopman (2012). We use the notation introduced in Section 6.3.1; $\{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n\}$ are the observed locations augmented with the (missing) locations at which state transitions occur, and $\{t_1, \ldots, t_n\}$ the corresponding times. Note that $\tilde{\boldsymbol{x}}_i$ is treated as missing data if t_i corresponds to a behavioural transition (see below for processing of missing data). For all $i = 1, \ldots, n-1$, the behavioural state is constant between t_i and t_{i+1}, and we denote its value by s_i.

Filtering A standard choice to initialise the state estimate is $\hat{\boldsymbol{a}}_1 = (\tilde{x}_1, 0, \tilde{y}_1, 0)'$, i.e. the mean initial location is the first observation $\tilde{\boldsymbol{x}}_1$, and the mean initial velocity is $\mathbf{0}$. The initial covariance matrix $\hat{\boldsymbol{P}}_1$ is typically taken to be diagonal, and its elements measure the uncertainty on the initial state estimate $\hat{\boldsymbol{a}}_1$.

Then, for $i = 1, 2, \ldots, n-1$, the Kalman filter equations are

$$\begin{aligned}
\boldsymbol{u}_i &= \tilde{\boldsymbol{x}}_i - \boldsymbol{Z}\hat{\boldsymbol{a}}_i \\
\boldsymbol{F}_i &= \boldsymbol{Z}\hat{\boldsymbol{P}}_i\boldsymbol{Z}' + \boldsymbol{H}_i \\
\boldsymbol{K}_i &= \begin{pmatrix} \boldsymbol{T}_i & \mathbf{0} \\ \mathbf{0} & \boldsymbol{T}_i \end{pmatrix} \hat{\boldsymbol{P}}_i\boldsymbol{Z}'\boldsymbol{F}_i^{-1}
\end{aligned}$$

$$\hat{\boldsymbol{a}}_{i+1} = \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix} \hat{\boldsymbol{a}}_i + \boldsymbol{K}_i \boldsymbol{u}_i$$

$$\hat{\boldsymbol{P}}_{i+1} = \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix} \hat{\boldsymbol{P}}_i \left[\begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix} - \boldsymbol{K}_i \boldsymbol{Z}' \right]' + \begin{pmatrix} \boldsymbol{Q}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_i \end{pmatrix}$$

with the model matrices $\boldsymbol{Z}$, $\boldsymbol{T}_i$, and $\boldsymbol{Q}_i$, defined in Section 6.3.1, and where $\boldsymbol{Z}'$ is the transpose of $\boldsymbol{Z}$.

The Kalman filter iterations output a state estimate $\hat{\boldsymbol{a}}_i$ and an estimate covariance matrix $\hat{\boldsymbol{P}}_i$ at each time t_i, for $i \in \{1, \ldots, n\}$. These estimates are conditional on all previous observations, i.e. on $\{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_{i-1}\}$.

Log-likelihood Due to the assumptions of normality, the log-likelihood can be calculated as a by-product of the Kalman filter,

$$l(\boldsymbol{\theta}|\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n) = -n \log(2\pi) - \frac{1}{2} \sum_{i=1}^{n} \left(\log |\boldsymbol{F}_i| + \boldsymbol{u}_i' \boldsymbol{F}_i^{-1} \boldsymbol{u}_i \right), \tag{6.14}$$

where $|\boldsymbol{F}_i|$ denotes the determinant of $\boldsymbol{F}_i$, and $\boldsymbol{\theta}$ is the vector of movement parameters β_j and σ_j ($j = 1, \ldots, N$). In the single-state case (i.e. the CTCRW model of Johnson et al., 2008a), or if the behavioural process is known, this expression of the log-likelihood can directly be used to obtain estimates of the movement parameters. For example, Equation 6.14 can be optimised over $\boldsymbol{\theta}$ to obtain maximum likelihood estimates. Here, we are interested in the case where there are several behavioural states and, in most applications, the state process is not known (i.e. the behaviour of the animal is not observed, and it must also be estimated). In Section 6.4, we describe a method of inference for the state-switching CTCRW, based on the likelihood given by the Kalman filter.

Smoothing We initialise $\boldsymbol{r}_n = \boldsymbol{0}$ and $\boldsymbol{N}_n = \boldsymbol{0}$. Then, for $i = n, n-1, \ldots, 1$, the Kalman smoother equations are

$$\boldsymbol{L}_i = \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix} - \boldsymbol{K}_i \boldsymbol{Z}$$

$$\boldsymbol{r}_{i-1} = \boldsymbol{Z}' \boldsymbol{F}_i^{-1} \boldsymbol{u}_i + \boldsymbol{L}_i' \boldsymbol{r}_i$$

$$\boldsymbol{N}_{i-1} = \boldsymbol{Z}' \boldsymbol{F}_i^{-1} \boldsymbol{Z} + \boldsymbol{L}_i' \boldsymbol{N}_i \boldsymbol{L}_i$$

$$\hat{\boldsymbol{\omega}}_i = \hat{\boldsymbol{a}}_i + \hat{\boldsymbol{P}}_i \boldsymbol{r}_{i-1}$$

$$\hat{\boldsymbol{\Sigma}}_i = \hat{\boldsymbol{P}}_i - \hat{\boldsymbol{P}}_i \boldsymbol{N}_{i-1} \hat{\boldsymbol{P}}_i$$

with the model matrices $\boldsymbol{Z}$, $\boldsymbol{T}_i$ and $\boldsymbol{Q}_i$ defined in Section 6.3.1, and where $\boldsymbol{u_i}$, $\boldsymbol{F}_i$, $\boldsymbol{K}_i$, $\hat{\boldsymbol{a}}_i$, and $\hat{\boldsymbol{P}}_i$ are computed in the filtering equations.

Then, $\hat{\boldsymbol{\omega}}_i$ is the smoothed state vector at time t_i, and $\hat{\boldsymbol{\Sigma}}_i$ the smoothed state variance matrix. These estimates are conditional on all observations $\{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n\}$, rather than only on previous observations. The filter and smoother algorithms are a popular method to estimate the true location of the animal on any time grid, e.g. to estimate locations that were observed with measurement error, or to estimate a trajectory at regular time intervals.

Missing data The Kalman filter and smoother are often used to estimate the animal's location (and velocity) on an arbitrary time grid, rather than only at the times of observations. The times that do not coincide with an observation are passed to the algorithm as missing data, for which an estimate of the continuous state (and an associated state covariance matrix) are computed. Moreover, in the case of the state-switching CTCRW model, the locations corresponding to the times of the behavioural transitions are generally not observed, and they must be treated as missing data. Missing observations only require minor changes to the equations given above. If the index i corresponds to a missing value, the filtering equations become

$$\begin{aligned}
\hat{\boldsymbol{a}}_{i+1} &= \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix} \hat{\boldsymbol{a}}_i \\
\hat{\boldsymbol{P}}_{i+1} &= \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix} \hat{\boldsymbol{P}}_i \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix}' + \begin{pmatrix} \boldsymbol{Q}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_i \end{pmatrix},
\end{aligned}$$

and the smoothing equations become

$$\begin{aligned}
\boldsymbol{r}_{i-1} &= \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix}' \boldsymbol{r}_i \\
\boldsymbol{N}_{i-1} &= \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix}' \boldsymbol{N}_i \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix}.
\end{aligned}$$

The other equations are unchanged. Note also that missing observations do not contribute to the log-likelihood.

6.4 Bayesian inference

We describe a method to infer the parameters of the multistate CTCRW model introduced in Section 6.2.3. In Section 6.3, we explained how the Kalman filter can be used to compute the likelihood of the model, conditionally on the behavioural state sequence. However, in most applications, the behaviours of the animal are not observed. We propose to estimate the unobserved states by sampling over all possible sequences in a Markov chain Monte Carlo (MCMC) algorithm.

6.4.1 MCMC algorithm

We follow the Metropolis-within-Gibbs approach introduced by Blackwell (2003). It relies on the successive updates of the three components of the model:

1. the reconstructed behavioural state sequence,
2. the movement parameters, i.e. the parameters of the CTCRW for each state,
3. the transition rates of the behavioural state process.

This is a "Gibbs" algorithm because we update each component in turn, conditional on the other two. However, we cannot sample directly from the posterior distributions of the behavioural state sequence, or from the posterior distribution of the movement parameters. In both cases, we use a Metropolis step to update the process.

We denote by $\mathcal{S}$ the reconstructed state sequence, composed of the times of the state transitions and the values of the state process between the transitions. We denote by $\boldsymbol{\theta}$ the vector of parameters of the movement process, i.e. $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_N, \sigma_1, \ldots, \sigma_N)$ for a N-state model. We write $p(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}, \mathcal{S})$ as the likelihood of a sequence of observed locations $\tilde{\boldsymbol{x}}$, given $\boldsymbol{\theta}$ and $\mathcal{S}$. This conditional likelihood is obtained from the Kalman filter presented in Section 6.3. We denote by $\boldsymbol{\Lambda}$ the generator matrix of the behavioural state process, composed of the transition rates and the transition probabilities, as defined in Equation 6.13.

We initialise the state sequence to $\mathcal{S}^{(0)}$, the movement parameters to $\boldsymbol{\theta}^{(0)}$, and the generator matrix to $\boldsymbol{\Lambda}^{(0)}$. Then, for K iterations ($k = 1, \ldots, K$), we alternate between the three steps described in Sections 6.4.1.1-6.4.1.3, to sample $\mathcal{S}^{(k)}$, $\boldsymbol{\theta}^{(k)}$, and $\boldsymbol{\Lambda}^{(k)}$.

6.4.1.1 Update of the behavioural state process

The evaluation of the likelihood of the model, described in Section 6.3.2, is conditional on the sequence of underlying states. In practice, the states are generally not observed, such

that we need to impute them. In this MCMC algorithm, we sample over the state sequences, using a Metropolis update.

The sequence of states $\mathcal{S}$ is composed of the times of the state transitions, and the values of the states. At each iteration, an updated state sequence $\mathcal{S}^*$ is proposed as follows. We choose an interval $[t_a, t_b]$, where a and b are two integers such that $t_1 \leq t_a < t_b \leq t_n$. We simulate the state process (S_t) from t_a to t_b, conditional on s_{t_a} and s_{t_b}, using the generator matrix $\mathbf{\Lambda}^{(k-1)}$. This can for example be done using the endpoint-conditioned continuous-time Markov process simulation methods of Hobolth and Stone (2009). The proposed sequence of states $\mathcal{S}^*$ remains identical to $\mathcal{S}^{(k-1)}$ outside $[t_a, t_b]$. The acceptance ratio for $\mathcal{S}^*$ is

$$r = \frac{p(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}^{(k-1)}, \mathcal{S}^*)}{p(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}^{(k-1)}, \mathcal{S}^{(k-1)})}.$$

The proposed state process reconstruction is accepted with probability $\min(1, r)$. If the proposed sequence is accepted, then $\mathcal{S}^{(k)} = \mathcal{S}^*$, else $\mathcal{S}^{(k)} = \mathcal{S}^{(k-1)}$.

The length of the interval $[t_a, t_b]$ over which the state sequence is updated is a tuning parameter of the sampler, and updates over longer intervals are generally less likely to be accepted. Note that the state sequence is generated from a continuous-time Markov process, and transitions can therefore occur at any point in (continuous) time. In particular, they are not constrained to happen at the times of the observations.

6.4.1.2 Update of the movement parameters

We use a Metropolis-Hastings step to update the movement parameters. At iteration k, we propose new movement parameters $\boldsymbol{\theta}^*$, from a proposal density $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(k-1)})$, and the acceptance ratio is

$$r = \frac{p(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}^*, \mathcal{S}^{(k)})p(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^{(k-1)}|\boldsymbol{\theta}^*)}{p(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}^{(k-1)}, \mathcal{S}^{(k)})p(\boldsymbol{\theta}^{(k-1)})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(k-1)})},$$

where $p(\boldsymbol{\theta})$ denotes the prior distribution on the movement parameters $\boldsymbol{\theta}$.

The parameters are updated to $\boldsymbol{\theta}^*$ with probability $\min(1, r)$. Note that, if the proposal distribution is symmetric such that $\forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, q(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) = q(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1)$, then r simplifies to

$$r = \frac{p(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}^*, \mathcal{S}^{(k)})p(\boldsymbol{\theta}^*)}{p(\tilde{\boldsymbol{x}}|\boldsymbol{\theta}^{(k-1)}, \mathcal{S}^{(k)})p(\boldsymbol{\theta}^{(k-1)})}.$$

In practice, a standard choice is to use a multivariate normal proposal distribution on the working scale of the parameters (in this case, on the log scale). Its variance can be tuned to obtain different acceptance rates, and covariance structure can be added to explore the parameter space more efficiently.

6.4.1.3 Update of the transition rates

Following Blackwell (2003), using conjugate priors, the transition rates can be directly sampled from their posterior distribution, which is known conditionally on the reconstructed state sequence $\mathcal{S}^{(k)}$. We consider the formulation of the generator matrix given in Equation 6.13, in terms of transition rates and transition probabilities. The transition rates and the transition probabilities can be sampled separately, as follows.

For each $i \in \{1, \ldots, N\}$, we denote n_i the number of time intervals spent in state i, and $(d_i^{(1)}, d_i^{(2)}, \ldots, d_i^{(n_i)})$ their lengths, derived from the state sequence $\mathcal{S}^{(k)}$. These dwell times are exponentially distributed with rate λ_i. The conjugate prior of the exponential distribution is the gamma distribution such that, with the prior

$$\lambda_i \sim \text{gamma}(\alpha_1, \alpha_2),$$

the transition rates are sampled from the posterior distribution

$$\lambda_i^{(k)} | \mathcal{S}^{(k)} \sim \text{gamma}\left(\alpha_1 + n_i,\ \alpha_2 + \sum_{j=1}^{n_i} d_i^{(j)}\right).$$

For $i \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, N\}$ such that $i \neq j$, we denote by n_{ij} the number of transitions from state i to state j, and $n_i = \sum_j n_{ij}$ the number of transitions out of state i, derived from the state sequence $\mathcal{S}^{(k)}$. Then, we have

$$(n_{i1}, n_{i2}, \ldots, n_{i_N}) \sim \text{multinom}(n_i, p_{i1}, p_{i2}, \ldots, p_{iN}).$$

The conjugate prior of the multinomial distribution is the Dirichlet distribution such that, with the prior

$$\boldsymbol{p}_i \sim \text{Dir}(\kappa_{i1}, \kappa_{i2}, \ldots, \kappa_{iN}),$$

the posterior distribution of the transition probabilities is

$$\boldsymbol{p}_i^{(k)} | \mathcal{S}^{(k)} \sim \text{Dir}(\kappa_{i1} + n_{i1}, \kappa_{i2} + n_{i2}, \ldots, \kappa_{iN} + n_{iN}).$$

where $\boldsymbol{p}_i = (p_{i1}, p_{i2}, \ldots, p_{iN})$ is the vector of transition probabilities out of state i.

6.4.2 Implementation

The MCMC algorithm described in Section 6.4.1 can be implemented to obtain samples from the posterior distributions of the movement parameters, of the transition rates, and of the hidden state sequence. In practice, the algorithm must be tuned appropriately, to avoid numerical issues and ensure fast mixing. It is difficult to give general guidance for the choice of initial and tuning parameters, but we provide some suggestions below.

Initial parameters The sampler requires initial values for all parameters, and for the reconstructed state sequence, from which to start the exploration of the parameter space. Initial values for the movement parameters, and for the transition rates of the state process, must be provided. If the starting values are poorly chosen (i.e. not in a plausible range), the chain may need many iterations to overcome the initialisation bias. In practice, it may often be easier to provide initial values for the alternative parameters τ and ν defined in Section 6.2.2.1 (rather than for β and σ), because they can be interpreted as the time scale of movement persistence, and the mean speed of movement of the animal.

The state sequence, defined by the times of transitions and the values of the state process between transitions, must also be initialised. One possibility would be to use another method of classification of telemetry data, such as hidden Markov models (Michelot et al., 2016), to obtain a rough partitioning of the track. However, such methods may not be appropriate if the data are collected at irregular time intervals, for example. Another option is to initialise the state sequence at random, by assigning the value 1 or 2 to each observation with probability 0.5 each (in a 2-state model). Then, once the observations have been initially classified, the exact timing of each state transition must be chosen, for example at random in the time interval where it occurs.

Kalman parameters The Kalman filter needs an initial value for the continuous state process, i.e. for the true location and the velocity of the animal at the first time point. As mentioned in Section 6.3, the initial location is the first observation, and a standard choice for the initial velocity is $\mathbf{0}$. This is reasonable if the mean velocity is $\mathbf{0}$, as assumed throughout this chapter. An initial covariance matrix is also passed to the Kalman filter, and it reflects the uncertainty on the initial state vector. The uncertainty associated with the initial location will typically be the measurement error variance. It is more difficult to define the uncertainty

associated with the initial velocity. It may be chosen as a plausible value for the long-term variance of the distribution of the velocity, for example.

Here, we assume that the variance of the measurement error is known, and that it is provided as input for the Kalman filter. This is usually a reasonable assumption, because the error of standard telemetry devices (e.g. Argos) is known (see e.g. Lopez et al., 2014). Alternatively, we could estimate the error variance parameters, in addition to the other parameters of the model. In the framework described here, we would need to sample over them, in a Metropolis step similar to the one used to sample over the movement parameters (Section 6.4.1.2).

Proposal distributions To propose updates of the state sequence, we consider the endpoint-conditioned simulation method suggested by Hobolth and Stone (2009). It is implemented in the R package ECctmc, with the function `sample_path` (Fintzi, 2018). The length of the interval over which the state sequence is updated determines its "proposal distribution": longer updates are less likely to be accepted. If the length of the update interval is sampled from a heavy-tailed distribution, it allows for occasional long updates.

Proposal distributions must also be defined for the movement parameters. We consider normal proposals on the working scale, and the choice of the proposal variance determines the acceptance rate. It can usually be tuned to obtain an optimal acceptance rate, i.e. around 23% in a standard Metropolis algorithm (Roberts et al., 1996).

Prior distributions Prior distributions must be specified for the movement parameters, and for the transition rates. Like for the choice of initial parameters, it may be easier to define prior distributions on the parameters τ and ν than on β and σ, because of their biological interpretation. Prior distributions on the transition rates define a plausible range for the time scale over which the behavioural states occur.

6.5 Simulation study

We used simulations to investigate the performance of the MCMC algorithm described in Section 6.4 to infer the hidden state sequence and the movement parameters from irregular movement data, and to compare this method to analogous discrete-time approaches.

6.5.1 Estimation from irregular intervals

We simulated 10,000 locations from a 2-state model at a fine time scale (every 0.1 time unit), and thinned them by keeping 10% of the points at random (i.e. 1000 irregularly-spaced

locations), to emulate real movement data. The time intervals in the resulting data set ranged between 0.1 and 8 time units.

The movement parameters and switching rates of the simulated process were chosen as

$$(\beta_1, \beta_2) = (1, 0.3), \quad (\sigma_1, \sigma_2) = (1, 3), \quad \mathbf{\Lambda} = \begin{pmatrix} -0.03 & 0.03 \\ 0.03 & -0.03 \end{pmatrix}.$$

In state 1, the variance was smaller and the reversion to the mean larger, which resulted in slower and more sinuous movement (perhaps analogous to "area-restricted search" behaviour). State 2 corresponded to faster and more directed movement (analogous to "transit"). This can be seen from the time scale of autocorrelation τ and mean speed ν for each state, as defined in Section 6.2.2.1. In this simulation, we have $(\tau_1, \tau_2) = (1, 3.33)$ and $(\nu_1, \nu_2) = (0.89, 4.85)$, i.e. more persistent and faster movement in state 2. The transition rates were chosen such that the process would on average stay about 30 time units in a state before switching to the other state. The simulated track (after thinning) is shown in Figure 6.4.

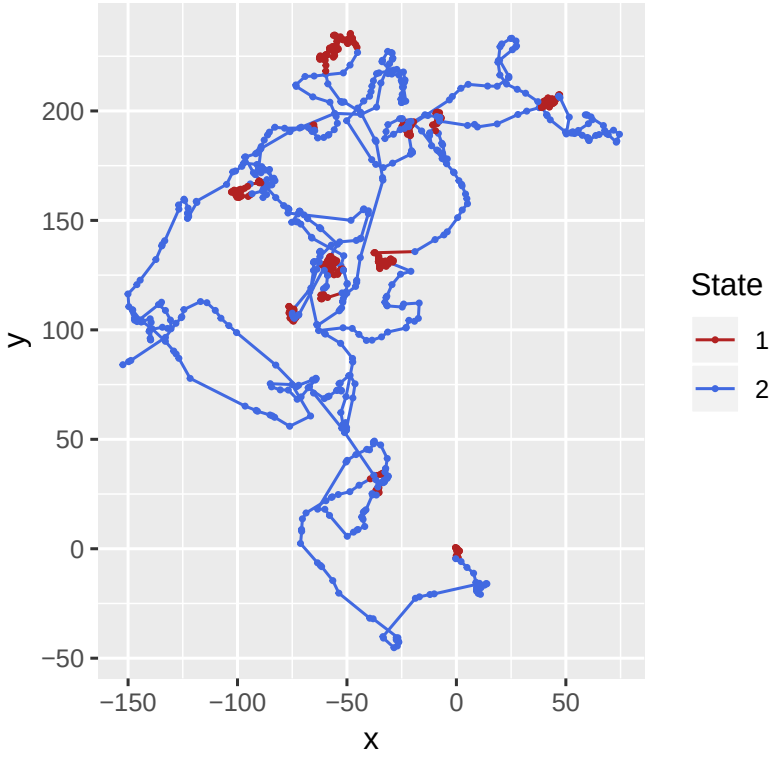


Fig. 6.4 Track simulated from a 2-state CTCRW model.

We initialised the reconstructed state sequence by classifying each observation randomly as being in state 1 or state 2, with probability 0.5 each. At each iteration, the state process was updated over a randomly-selected interval of random length (uniform between 3 and 70 time steps). We used independent normal proposal distributions (on the working log scale) to update the movement parameters. The proposal variances were tuned based on initial test runs, to obtain near-optimal acceptance rates. We chose normal prior distributions on the working scale for the movement parameters, centred on the true values of the parameters, and with large variances.

We ran 10^5 MCMC iterations, which took around 20 min on a 2GHz i5 CPU, and discarded the first 5×10^4 as burn-in. Figure 6.5 shows the posterior probabilities of being in state 2 at the times of the observations, to compare with the "true" simulated state sequence. For each $i = 1, \ldots, n$, we calculated the posterior probability as the proportion of reconstructions of the state process in which the location $\boldsymbol{x}_i$ was classified in state 2. (It would have been equivalent to consider the posterior probability of being in state 1; the two probabilities sum to 1.) We considered that an observation was misclassified if the posterior probability of being in the true state was less than 0.5. The states were correctly estimated at the vast majority of observation times, with only 2% of missclassified observations. The posterior probability of being in the true state was less than 0.9 for about 4.5% of the observed locations.

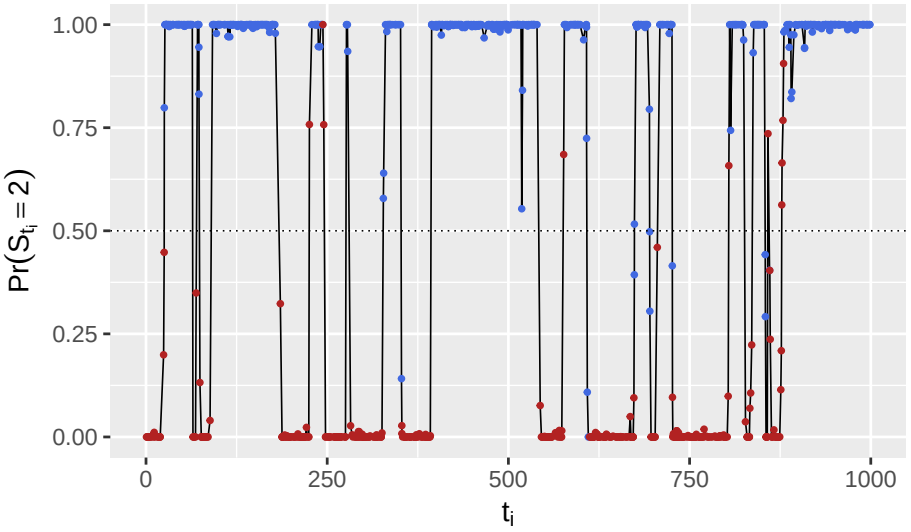


Fig. 6.5 Posterior probabilities of being in state 2 at the times of the observations. The true (simulated) states are shown by the colours (red: state 1, blue: state 2).

Figure 6.6 displays posterior samples for the state-dependent movement parameters, β_1, β_2, σ_1, and σ_2, as well as the true parameter values used in the simulation. The posterior

distributions seem to appropriately estimate all movement parameters. Although there appears to be some possible bias, replications (not shown here) reveal that it is only due to randomness, and not consistent across simulations.

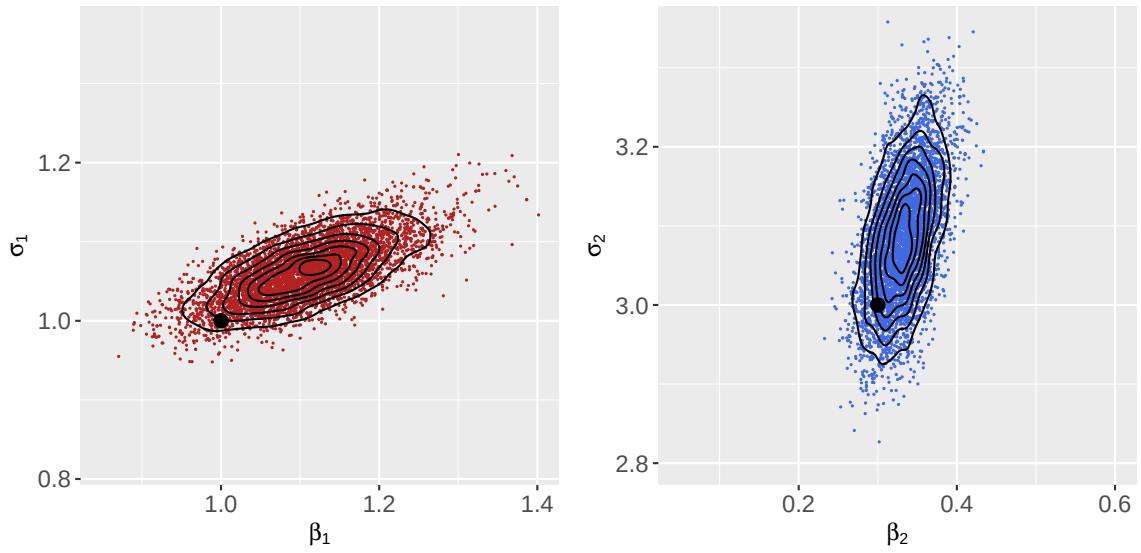


Fig. 6.6 Posterior samples of the movement parameters in state 1 (left) and in state 2 (right), in the simulation study. The black dots are the true values of the parameters, used in the simulation. The black lines show contours of a kernel density estimate of the posterior samples. The samples are thinned to every tenth value, for visualisation purposes.

	True value	Posterior mean	95% CI	ESS
β_1	1	1.11	(0.97,1.26)	1054
β_2	0.3	0.33	(0.28,0.39)	801
σ_1	1	1.06	(0.99,1.14)	1511
σ_2	3	3.09	(2.94,3.24)	2241

Table 6.1 Results of the simulation study, based on 5×10^4 iterations (excluding burn-in). Posterior means and 95% credible intervals of the four movement parameters, and effective sample sizes.

Posterior means, 95% credible intervals, and effective sample sizes for the four movement parameters are given in Table 6.1. The effective sample sizes were calculated with the R package coda (Plummer et al., 2006). The credible intervals all included the true values of the parameters. We were able to recover the values of the state process and of the state-dependent movement parameters from a simulated track thinned to irregular intervals. This demonstrates the ability of the method to work across temporal scales, and to cope with irregular sampling.

6.5.2 Measurement error and path reconstruction

The state-space model formulation of the multistate CTCRW, presented in Section 6.3.1, makes it possible to account for measurement error in the observed locations. This is particularly useful for data collected via the Argos system, which is common for marine mammals. In Section 6.3, we described the Kalman smoother, as a method to obtain estimates of the continuous state of the process, i.e. estimates of the true location and velocity of the animal. Here, we investigate the effect of measurement error on the accuracy of the parameter estimation, and we use the Kalman smoother to reconstruct the location process from thinned and noisy observations.

We started from the simulated path described in Section 6.5.1, and added isotropic Gaussian noise to the thinned locations, with variance $\sigma^2_{\text{obs}} = 4$. Mixing was slower than in the absence of measurement noise, so we ran 4×10^5 iterations of the MCMC algorithm, and discarded the first 2×10^5 as burn-in, to obtain estimates of the movement parameters and of the state sequence. The estimates, credible intervals, and effective sample sizes for the four movement parameters are given in Table 6.2.

	True value	Posterior mean	95% CI	ESS
β_1	1	1.37	(0.52,4.03)	129
β_2	0.3	0.35	(0.26,0.45)	4679
σ_1	1	0.97	(0.36,2.73)	147
σ_2	3	3.13	(2.73,3.62)	3950

Table 6.2 Results of the simulation study with measurement error, based on 2×10^5 iterations (excluding burn-in). Posterior means and 95% credible intervals of the four movement parameters, and effective sample sizes.

The posterior distributions generally captured the true parameter values. However, the effective sample sizes for β_1 and σ_1 were very small, indicating that mixing was very slow for the movement parameters in state 1. The credible intervals for the parameters in state 1 were also much wider than those in state 2. This is because the simulated movement was much slower in state 1, such that the introduced noise was large compared with the scale of movement. Despite this difficulty, the posterior state probabilities gave a good estimate of the true state sequence. Indeed, about 94% of the locations were correctly classified.

We then ran the Kalman smoother iterations, based on the mean posterior estimates of the movement parameters and the estimated state sequence, to estimate the true location process on a fine time grid. Figure 6.7 shows the simulated track, the thinned and noisy observations,

and the path reconstruction. The estimates obtained from the Kalman smoother successfully approximated the true underlying location process.

6.5.3 Comparison to discrete-time model

We simulated two tracks from the state-switching CTCRW model, at a time resolution of 0.1, with the same parameters used in the simulations of Section 6.5.1. We then thinned them to obtain two data sets:

(a) A track of 1000 locations, at a regular time resolution of $\Delta = 0.5$;

(b) A track of 1000 locations, at a regular time resolution of $\Delta = 5$.

Both tracks arise from the same underlying process (the true state-switching CTCRW), and they emulate data sets observed at different time resolutions. We fitted a discrete-time state-switching model, and the state-switching CTCRW model, to the two data sets, to illustrate the differences in their scaling properties. In this experiment, we thinned the tracks at regular time intervals (rather than irregularly, at random) because discrete-time models cannot be applied to irregular location data. The two thinned tracks are shown in Figure 6.8.

We fitted a 2-state discrete-time hidden Markov model to each data set, using the R package moveHMM (Michelot et al., 2016). The standard hidden Markov model, described by Patterson et al. (2009) and Langrock et al. (2012), is formulated in terms of step lengths (distances between successive locations) and turning angles (angles between successive locations). We modelled the step lengths with gamma distributions, and the turning angles with von Mises distributions. There were six estimated movement parameters: the step length means (μ_1, μ_2), the step length standard deviations (s_1, s_2), and the turning angle concentrations (κ_1, κ_2). The mean of the distribution of turning angles was fixed to zero, which is a standard choice to model movement persistence. See Michelot et al. (2016) for more detail about the hidden Markov model formulation implemented in moveHMM. The parameters of the step length distribution are related to the speed of movement, and the parameters of the turning angle distribution are related to the directional persistence. There were two estimated parameters of the state process: the transition probabilities γ_{12} and γ_{21}, where γ_{ij} is the probability of a transition from state i to state j over one time interval.

We also fitted the state-switching CTCRW model to each thinned data set separately. Like in the simulation study of Section 6.5, we ran 10^5 MCMC iterations, and discarded the first 5×10^4 as burn-in. We estimated four movement parameters: $(\beta_1, \beta_2, \sigma_1, \sigma_2)$. There were two estimated parameters of the state process: the transition rates λ_1 and λ_2, as defined

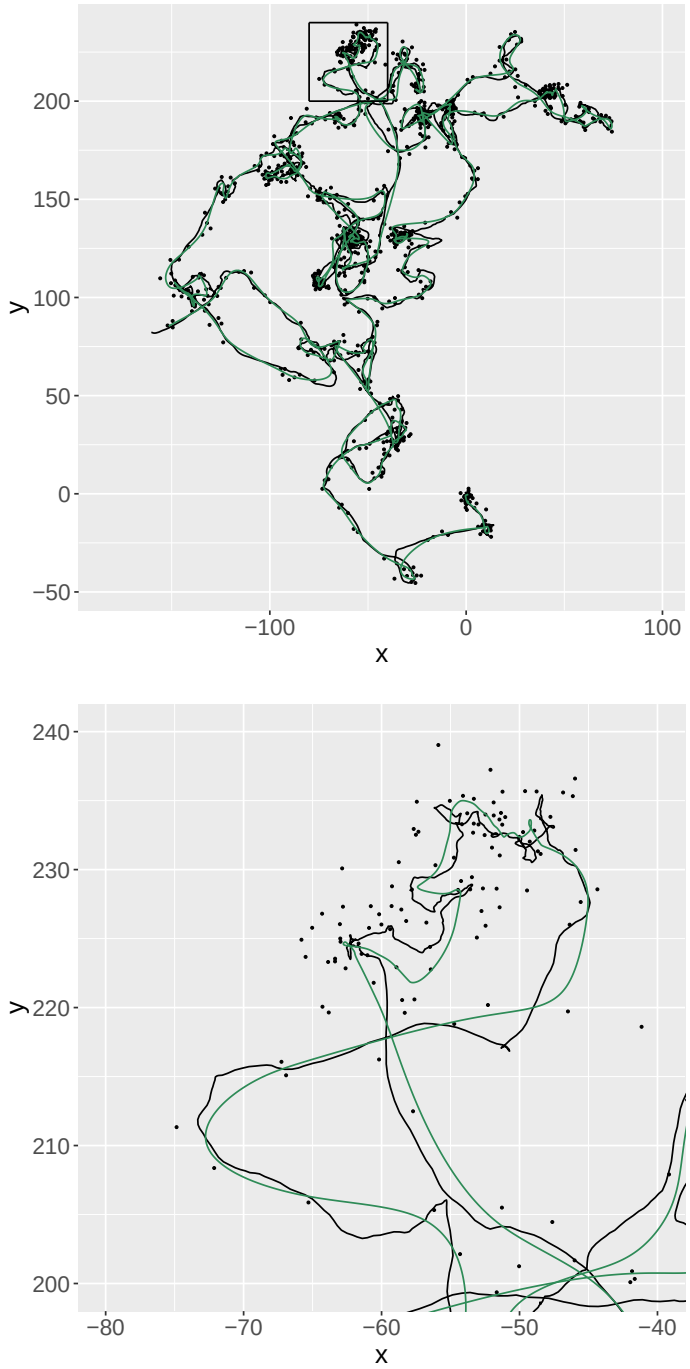


Fig. 6.7 Reconstruction of a simulated track. The black line is the true simulated process, the black dots are the thinned and noisy locations used to fit the model, and the green line is the estimated location process obtained with the Kalman filter and smoother. The bottom plot magnifies the part of the track delimited by a rectangle in the top plot.

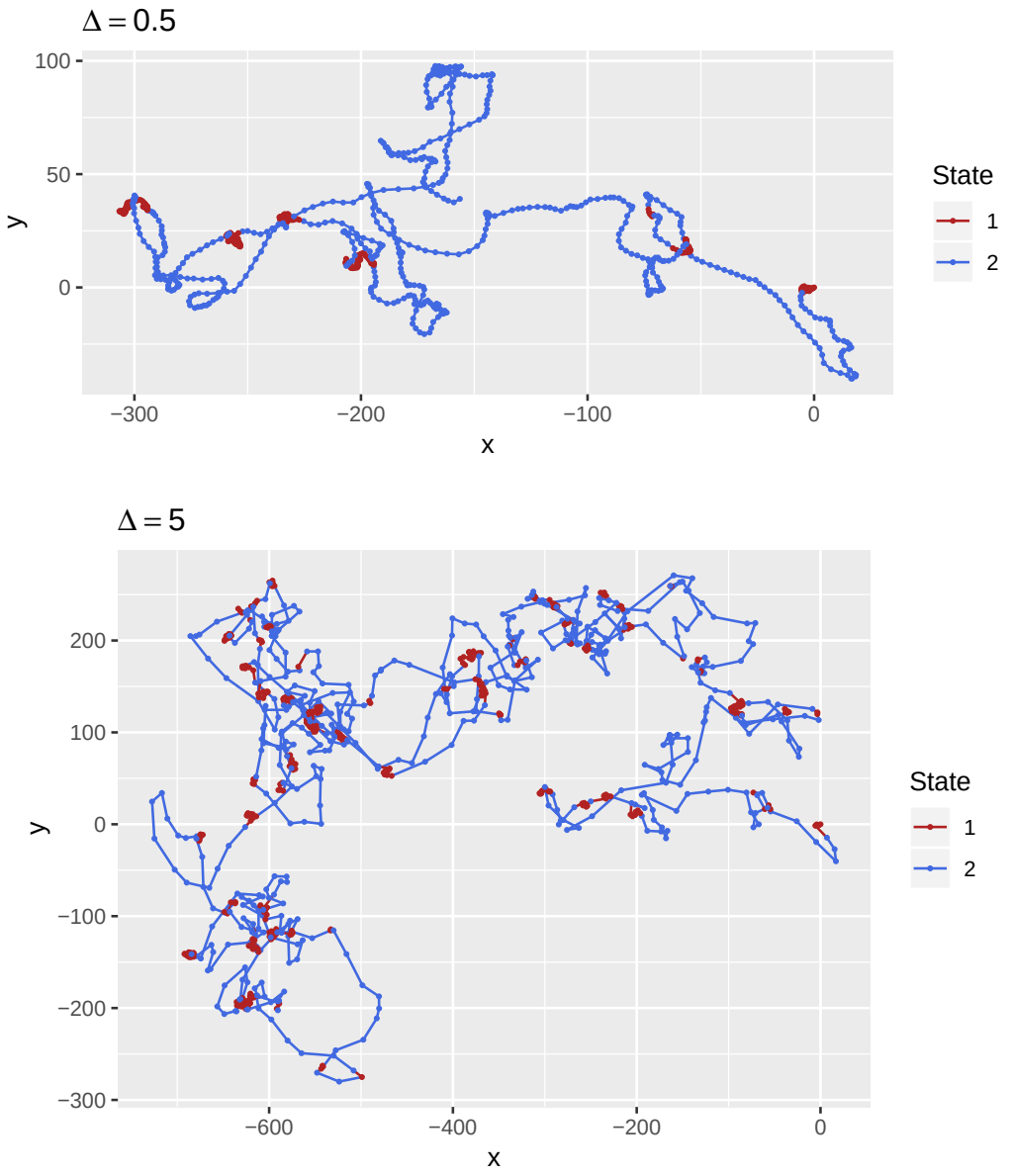


Fig. 6.8 Two tracks simulated from the same state-switching CTCRW model, and thinned to different time resolutions (top: $\Delta = 0.5$; bottom: $\Delta = 5$).

in Equation 6.13. The parameter estimates for the hidden Markov models and for the state-switching CTCRW models are given in Table 6.3.

The parameters of the continuous-time model (state-switching CTCRW) were very similar in both analyses. Although the two thinned data sets have very different sampling intervals ($\Delta = 0.5$ and $\Delta = 5$), the parameters of the CTCRW are independent of the time intervals of observations.

However, the parameters of the discrete-time model (hidden Markov model) were very different in the two analyses. This should be expected, because the model is formulated in terms of scale-dependent metrics: the step lengths and turning angles. The step lengths increase with the time interval of observation, and the estimates reflect this. Indeed, the mean and standard deviation of the distribution of step lengths were smaller in the analysis with $\Delta = 0.5$ than when $\Delta = 5$. It should be noted that there is no clear scaling rule for those parameters. In particular, the mean step length is generally *not* 10 times longer over $\Delta = 5$ than over $\Delta = 0.5$. This highlights a problem that often arises in analyses of irregular tracking data with discrete-time models. A method that has been proposed to deal with irregular data is to derive a "movement rate" for each time step, by dividing the step length by the time interval. Then, it is assumed that the movement rates do not depend on the time interval, and represent a measure of the animal's speed of movement. However, we can see from this simulated example that the movement rates do, in fact, depend on the time intervals of observation. Indeed, the mean movement rate can be obtained as the mean step length divided by the time interval. In the first data set, the mean movement rates in the two states are $0.52/0.5 = 1.04$ and $2.60/0.5 = 5.2$. In the second data set, the mean movement rates are $2.83/5 = 0.57$ and $18.95/5 = 3.79$. In general, the sparser the data, the more the speed of movement is underestimated by this method. A similar problem arises when linear interpolation is used to obtain locations at regular intervals in time. Linear interpolation draws straight lines between observed locations, and will therefore tend to underestimate the speed of movement, and overestimate the persistence in direction.

The turning angles tend to be less concentrated around 0 as the time interval of observation increases, because movement persistence is less clearly visible at a coarse time resolution. This can be seen in the parameter estimates, as κ_1 and κ_2 are much larger in the analysis with $\Delta = 0.5$ than in the analysis with $\Delta = 5$. However, similarly to the mean step length, it is not clear how the concentration parameter scales with the time interval of observations.

This dependence on the time scale of observations must be taken into account in the interpretation of the parameters of discrete-time movement models. For example, it is difficult to interpret the mean step length as a measure of the animal's speed, because its definition

			$\Delta = 0.5$		$\Delta = 5$	
	Parameter	True value	Estimate	95% CI	Estimate	95% CI
Discrete time						
Movement parameters	μ_1	—	0.52	(0.48,0.56)	2.83	(2.66,3.02)
	μ_2	—	2.60	(2.50,2.71)	18.95	(17.98,19.98)
	s_1	—	0.33	(0.30,0.37)	1.69	(1.53,1.87)
	s_2	—	1.01	(0.94,1.09)	10.07	(9.27,10.94)
	κ_1	—	1.38	(1.22,1.55)	0.21	(0.12,0.39)
	κ_2	—	7.03	(5.94,8.32)	0.75	(0.62,0.90)
Transition probabilities	γ_{12}	—	0.063	(0.045,0.089)	0.132	(0.104,0.167)
	γ_{21}	—	0.068	(0.048,0.097)	0.134	(0.104,0.172)
Continuous time						
Movement parameters	β_1	1	1.12	(0.95,1.30)	0.93	(0.69,1.23)
	β_2	0.3	0.32	(0.25,0.39)	0.27	(0.23,0.32)
	σ_1	1	1.04	(0.98,1.11)	0.95	(0.74,1.22)
	σ_2	3	3.03	(2.90,3.17)	2.84	(2.58,3.14)
Transition rates	λ_1	0.03	0.026	(0.013,0.045)	0.035	(0.026,0.044)
	λ_2	0.03	0.034	(0.016,0.058)	0.033	(0.025,0.042)

Table 6.3 Parameter estimates in the comparison of hidden Markov models ("discrete time") and state-switching CTCRW models ("continuous time") over two different temporal scales. For the hidden Markov model parameters, maximum likelihood estimates and 95% confidence intervals are shown (obtained with the package moveHMM). For the CTCRW parameters, mean posterior estimates and 95% credible intervals are shown.

is tied to the time interval of observation. It will tend to increase with the time interval, but generally not linearly (because animals do not move in straight lines between observed locations). Crucially, the results of different discrete-time analyses cannot be compared if the time interval of observation is different, as illustrated in this simulation experiment.

The same scale dependence arises for the parameters of the state process. The transition probabilities of the hidden Markov models are defined over a given (fixed) time interval, and can only be interpreted over that time interval. For example, in the first data set, we estimated that the probability of a transition from state 1 to state 2 was $\hat{\gamma}_{12} = 0.06$. This means that, if the animal is in state 1, there is a probability 0.06 of switching to state 2 over each time step of length $\Delta = 0.5$. On the other hand, the transition rates of the continuous-time model are defined independently of any particular time interval. Regardless of the time interval of observation, the estimated transition rate can be interpreted as the "mean number of transitions per hour".

The objective of this simulation study is not to compare how well the discrete-time and continuous-time models captured the true underlying process, or how well they recovered the true parameter values. This would be unfair, because the true process used to simulate the tracks is the state-switching CTCRW. Nevertheless, we believe that the results presented here are useful to illustrate the fundamental difference in the formulations of discrete-time and continuous-time models. In particular, a considerable practical advantage of continuous-time approaches is that the estimated parameters, and therefore the interpretation and inference, do not depend on the time interval of observation.

6.6 Grey seal case study

We illustrate the use of the state-switching CTCRW model for the analysis of a grey seal (*Halichoerus grypus*) movement track. We considered a trajectory of 1875 observations, collected in the North Sea between April and August 2008, available on Movebank (McConnell, 2019) and previously described by Russell et al. (2015). The base sampling frequency was of one location every 30 minutes, but many fixes were missed, and the resulting time grid was highly irregular ($P_{0.025}$ = 27 min, $P_{0.975}$ = 8 hours). Note that the CTCRW model describes movement on a plane, and thus requires that the longitude-latitude locations be projected to UTM coordinates. The animal was equipped with a GPS tag, with low location uncertainty (less than 50m), so we did not model measurement error in the analysis.

We considered a 2-state CTCRW model, with four movement parameters to estimate: β_1, β_2, σ_1, σ_2. Similarly to the simulation study, we initialised the state reconstruction to a random

sequence of 1s and 2s (with probability 0.5 each). We used independent normal proposal and prior distributions on the working scale of the movement parameters. We selected the proposal variances based on test runs, and used weakly informative prior distributions. We ran 2 million MCMC iterations, discarding the first half as burn-in, which took about 14 hours on a 2GHz i5 CPU. We only saved every 100th reconstructed state sequence, because of memory limitations.

Figure 6.9 shows a map of the track, coloured by posterior state probabilities, and Figure 6.10 shows posterior samples for the four movement parameters ($\beta_1, \beta_2, \sigma_1, \sigma_2$). Posterior means, 95% credible intervals, and effective sample sizes for the four movement parameters are given in Table 6.4.

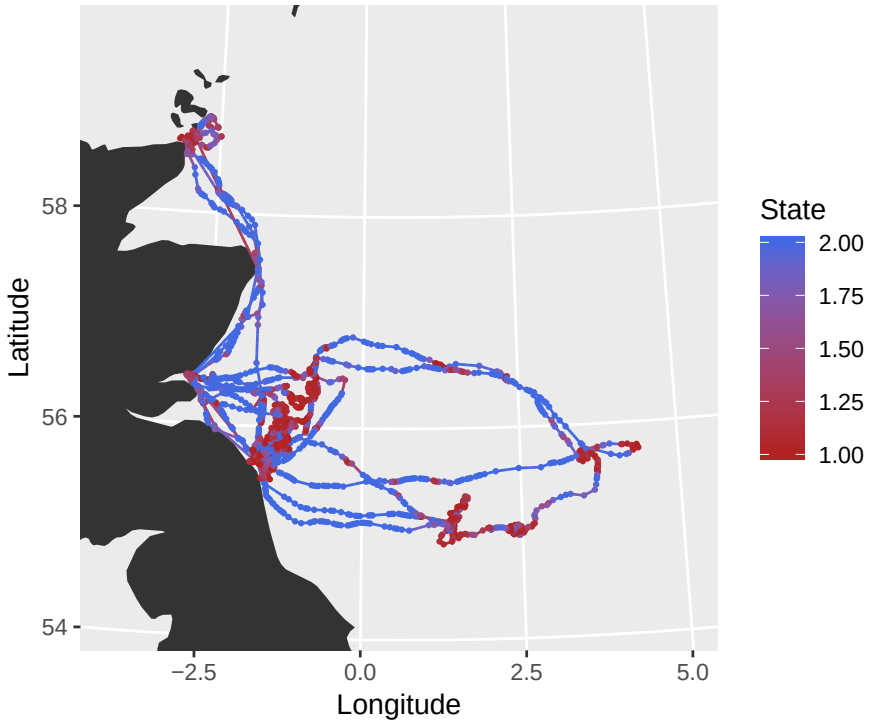


Fig. 6.9 Grey seal track, off the East coast of Great Britain, coloured by posterior state probabilities.

State 2 captured very directed movements, corresponding to periods of transit between areas of interest, and state 1 captured more tortuous phases of the track. This can be seen in Figure 6.10: the posterior distribution of β_1 covers much larger values than that of β_2 (posterior means of 1.74 and 0.06, respectively), indicating stronger reversion to the mean in state 1, and thus less movement persistence. From Figure 6.10, we can also see that the posterior distributions of the movement parameters in the two states do not overlap. This is a good

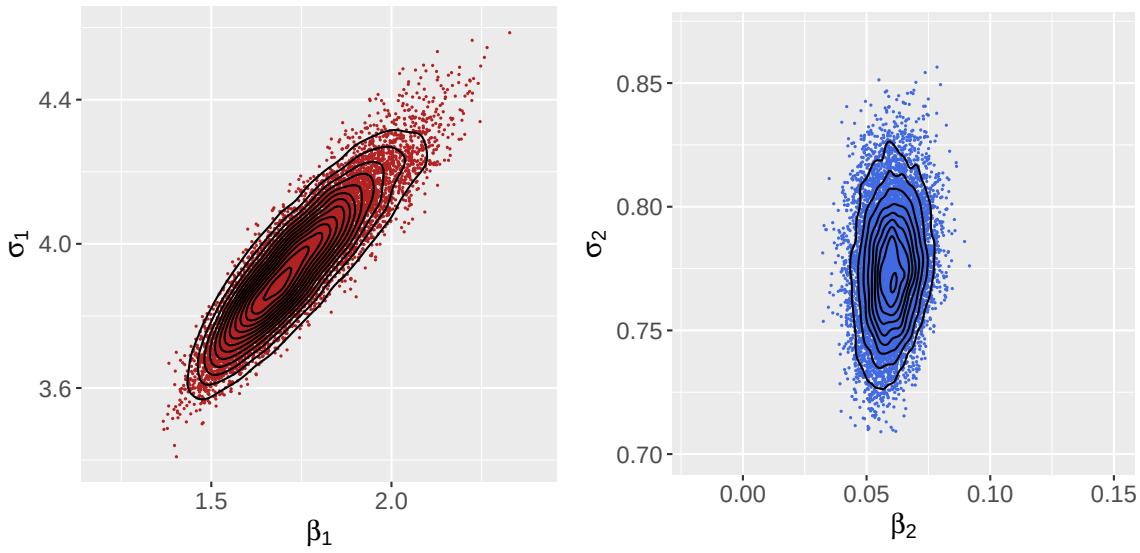


Fig. 6.10 Posterior samples of the movement parameters in state 1 (left) and in state 2 (right), in the grey seal case study. The black lines show contours of a kernel density estimate of the posterior samples. The samples are thinned to every 100th value, for visualisation purposes.

	Posterior mean	95% CI	ESS
β_1	1.74	(1.49,2.04)	1493
β_2	0.06	(0.04,0.08)	2413
σ_1	3.94	(3.65,4.28)	3150
σ_2	0.78	(0.73,0.82)	1234

Table 6.4 Results of the grey seal case study, based on 10^6 iterations (excluding burn-in). Posterior means and 95% credible intervals of the four movement parameters, and effective sample sizes.

indication that the two states of the model captured two distinct types of movement, which can be interpreted as distinct types of behaviour. There were no signs of label switching in the posterior samples; if there were, a straightforward solution would be to constrain (β_1, β_2) and (σ_1, σ_2) to be ordered (McClintock et al., 2014).

We transformed the posterior samples of movement parameters to obtain estimates of τ and ν in both states (as defined in Section 6.2.2.1). Histograms of the posterior samples for the transformed parameters are shown in Figure 6.11. The posterior means were $\hat{\tau}_1 = 0.55$h and $\hat{\tau}_2 = 17.21$h for the time scales of autocorrelation, and $\hat{\nu}_1 = 2.63$km/h and $\hat{\nu}_2 = 2.90$km/h for the mean speeds. This indicates that the two states are very similar in terms of the speed of movement, but that the autocorrelation function of the velocity drops much faster in state 1 than in state 2. As explained in Section 6.2.2.1, the velocity autocorrelation decreases to 0.05 over a time interval of 3τ. In the fitted model, the velocity is therefore approximately independent after about 1 hour 40 minutes in state 1, and after about two days in state 2.

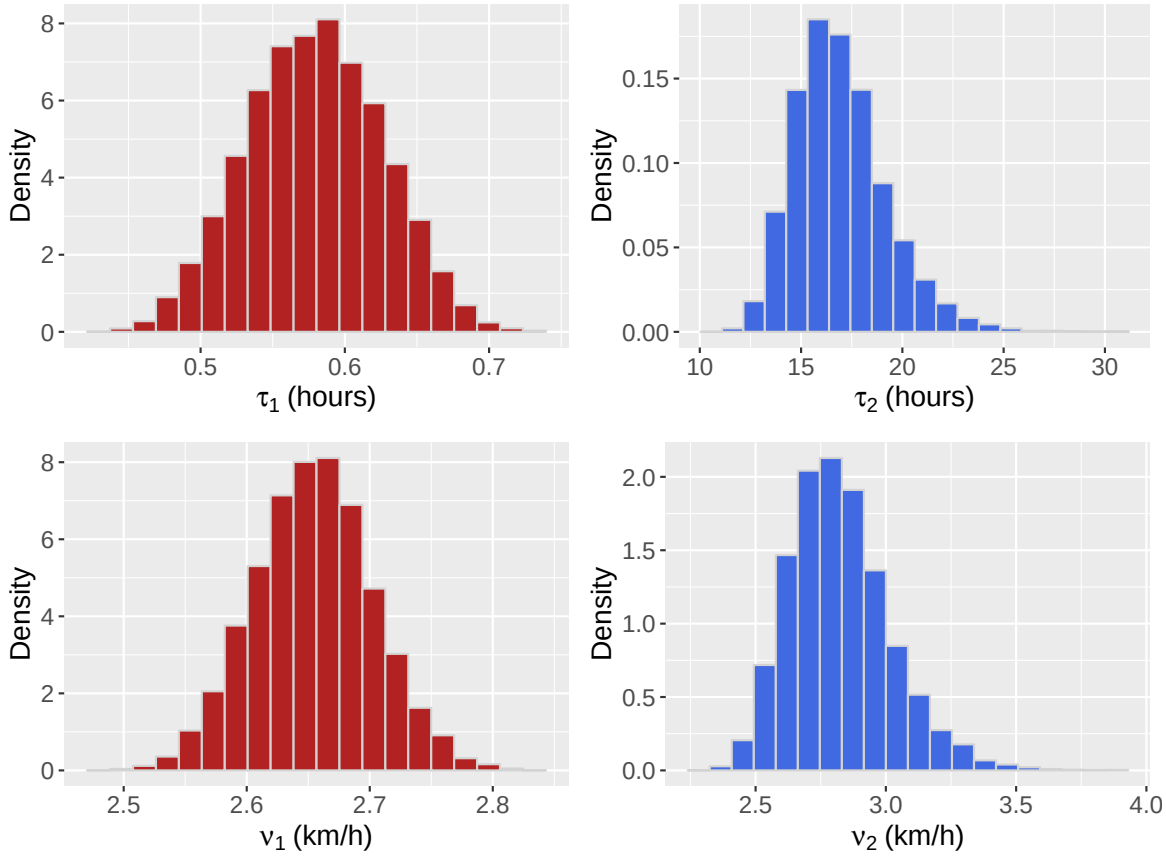


Fig. 6.11 Histograms of posterior samples for alternative movement parameters in the grey seal case study.

The posterior samples of the transition rates can be used to derive mean dwell times in each state, and long-term activity budgets. The dwell times in state i follow an exponential distribution with rate λ_i (the rate of transition out of state i). The mean dwell time can thus be derived as $d_i = 1/\lambda_i$. In the grey seal analysis, the posterior means for the mean dwell time were $\hat{d}_1 = 7.5$h in state 1, and $\hat{d}_2 = 7.8$h in state 2, indicating similar dwell times in both states. Activity budgets refer to the proportion of time spent by an animal in each of its behavioural states (Russell et al., 2015). In a time-homogeneous state-switching model, an estimate of the long-term activity budget can be calculated as the stationary distribution of the underlying Markov process. The stationary distribution of a N-state continuous-time Markov process is the vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)$ which satisfies $\boldsymbol{\pi}\boldsymbol{\Lambda} = \mathbf{0}$, subject to the constraint $\sum_{i=1}^{N} \pi_i = 1$, where $\boldsymbol{\Lambda}$ is the generator matrix defined in Equation 6.13. In the 2-state case, solving the equation yields $\pi_1 = \lambda_2/(\lambda_1 + \lambda_2)$ and $\pi_2 = \lambda_1/(\lambda_1 + \lambda_2)$. The posterior mean estimate for the stationary distribution was $(\hat{\pi}_1, \hat{\pi}_2) = (0.48, 0.52)$, i.e. the seal will tend to spend roughly the same proportion of time in both states, in the long term. Histograms of posterior draws for the dwell times and stationary distribution are displayed in Appendix E.

The Kalman filter and smoother recursions given in Section 6.3 can be used to compute estimated velocities at the times of the observations. The velocities obtained with the mean posterior movement parameters are displayed in Figure 6.12, and split by posterior state estimates. The strong movement autocorrelation in state 2 can best be seen in the outer rim of the plot, where the velocity sometimes persists with little variation over many time steps. Interestingly, a cluster of very small velocities (close to zero) was also classified in state 2. This is because, as seen in the estimates of the movement parameters, the main difference between the two states is not in the speed, but in the velocity persistence. State 2 therefore captures both fast persistent and slow persistent movements of the seal. On the other hand, state 1 captures less persistent movement, characterised by a weaker autocorrelation in the velocity process. This suggests that a 3-state model may be more appropriate to model these data. Indeed, a 3-state model could be used to differentiate between non-persistent movement, slow persistent movement, and fast persistent movement.

As illustrated in the simulation of Section 6.5.2, the Kalman algorithm can also provide estimates of the locations (and possibly velocities) of the animal – and associated standard errors – on any time grid, e.g. on a finer time grid than that of the observations.

To assess convergence, we ran the MCMC algorithm ten times, from ten different sets of starting values, and we calculated the potential scale reduction factor from the ten simulated chains (PSRF, Gelman et al., 2013). The PSRF compares the variance between the chains and within the chains, to determine whether the algorithm seems to have converged. A

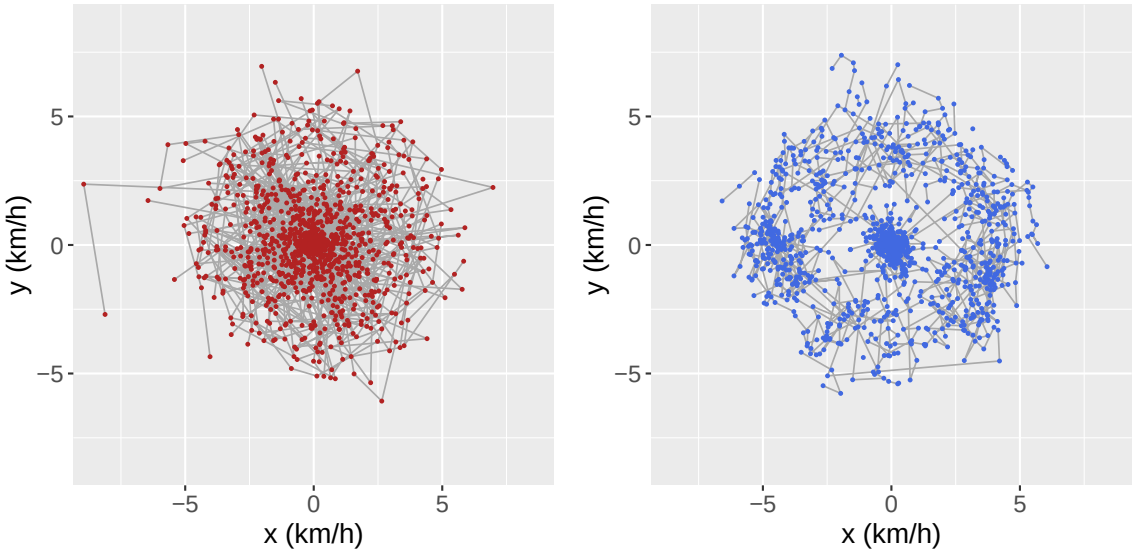


Fig. 6.12 Predicted velocities from the grey seal example, obtained with the mean posterior movement parameter estimates, for time steps classified in state 1 (left) and in state 2 (right). The grey segments link consecutive velocities that were classified in the same state.

score close to 1 indicates convergence. In this analysis, we evaluated the PSRF using the function 'gelman.diag' of the package coda. The scores for the four movement parameters were PSRF(β_1) = 1.11, PSRF(β_2) = 1.02, PSRF(σ_1) = 1.07, PSRF(σ_2) = 1.08, and the factors for the two transition rates were PSRF(λ_1) = 1.04 and PSRF(λ_2) = 1.06. There was therefore no indication of convergence issues in the ten simulated chains.

6.7 Discussion

We presented a Bayesian framework to infer discrete behavioural states and movement parameters from a continuous-time model of animal movement. The continuous-time formulation is consistent across temporal scales. The conditional likelihood of the model, used in the MCMC algorithm, is implemented using the Kalman filter, making it relatively fast. The state-switching CTCRW model described in this chapter is widely applicable to many types of telemetry data. Its flexible formulation can accommodate data with the following features: (1) movement persistence, (2) behavioural heterogeneity, (3) irregular time intervals, and (4) measurement error.

Different combinations of these four features can often be found in telemetry data sets, and failing to deal with them can lead to wrong inferences about the movement and behaviour of animals. To our knowledge, the only other existing approach which can account for this

range of features is that developed by Parton and Blackwell (2017). They described a state-switching continuous-time movement model, in which the movement is described by two stochastic processes: the bearing (i.e. compass direction), and the speed of movement. Parton and Blackwell (2017) modelled the bearing with a Brownian motion, and the speed with a one-dimensional Ornstein-Uhlenbeck process. Three biologically-relevant parameters of movement can be obtained from that model: the mean and variance of the speed process, and the variance of the bearing process. Their approach is very appealing due to its similarities with the widely-used discrete-time models based on step lengths and turning angles (Morales et al., 2004; Michelot et al., 2016). Indeed, the bearing process is analogous to the turning angle, and the speed process is analogous to the step length. The 'step-and-turn' model of Parton and Blackwell (2017) is therefore convenient to formulate the characteristics of movements in terms of familiar metrics. The main limitation of their approach, compared with the method presented in this chapter, is its computational cost. To help with the biological interpretation of the CTCRW, it may be possible to link its parameters (β and σ) to those of the step-and-turn model. In Section 6.2.2.1, we explained how the mean speed of movement ν can be calculated and, following Appendix D, we could also derive the speed variance of the CTCRW. Similar calculations may be used to obtain distribution of bearings for the velocity Ornstein-Uhlenbeck process.

The MCMC algorithm of Section 6.4 closely resembles that developed by Blackwell (1997, 2003). However, they modelled the location of an animal (rather than its velocity) with an Ornstein-Uhlenbeck process, which could not capture strong movement persistence. Here, we adapted the approach for the CTCRW process, to model autocorrelation in the velocity of an animal, i.e. persistence in the speed and direction of movement (which is often present in high-frequency telemetry data). However, the CTCRW may not be adequate to analyse movement data that do not display this type of persistence. In particular, data collected on a coarse time grid may not exhibit autocorrelated velocities, and could instead be analysed with the state-switching Ornstein-Uhlenbeck model of Blackwell (1997, 2003).

The inferential approach introduced in this paper could in principle be used to implement a state-switching version of the OUF model described by Fleming et al. (2014). The OUF process is a generalisation of the CTCRW used in this paper, and of the Ornstein-Uhlenbeck location process used e.g. by Blackwell (1997, 2003). It features both persistence in velocity and long-term attraction towards a point in space, making it a very flexible model of animal movement. Like the CTCRW, it can be written as a state-space model, and the Kalman filter can be used to derive the likelihood of the model (Fleming et al., 2017). The MCMC algorithm described in Section 6.4 could then be used to fit a multistate OUF model to animal

movement. However, the OUF process has five parameters (against 2 only for the CTCRW), which could make estimation more challenging. More generally, this methodology could be applied to a model switching between processes with different formulations (e.g. a 2-state model switching between a CTCRW and OUF process). These complex multistate models could for example capture the structured trips of central place foragers (similarly to the discrete-time models of Michelot et al., 2017).

Analyses of animal movement and behaviour often combine telemetry and environmental data. In state-switching models, the effect of environmental (or other) covariates on the transition probabilities is of particular interest, and is used to uncover the drivers of animal behavioural and movement decisions (Patterson et al., 2009; Bestley et al., 2012; Blackwell et al., 2016). Blackwell et al. (2016) described a method of inference for the state-switching Ornstein-Uhlenbeck movement model. They allow the transition rates to be functions of spatial covariates (i.e. that can be evaluated at any point of the study region), or to be functions of the time of day (to analyse circadian cycles in the behaviour of an animal). The MCMC algorithm of Section 6.4 could be extended, following Blackwell et al. (2016), to allow for the inclusion of covariates in the state-switching dynamics.

Although we refer to the states of the Markov process as "behavioural states", it is important to note that they really are *statistical* states, that capture the temporal autocorrelation in the velocity process. They should be interpreted with caution, and may not exactly correspond to separate behaviours (Patterson et al., 2017). In particular, there can be greater uncertainty in the partitioning of a track if the states are not very distinct, i.e. if they do not clearly differ in terms of the animal's velocity process (Beyer et al., 2013). In the simulation study and grey seal case study, we focused on the 2-state model formulation, because the interpretation becomes more difficult in models with more states. Pohle et al. (2017) discussed this problem in hidden Markov models, which are the discrete-time analogue of the model presented in this work. A possible solution is to use auxiliary data, if available, to identify behavioural states. For example, if the behaviour of the animal is known for some of the observations, the corresponding states can be fixed throughout the algorithm, in a semi-supervised framework (Leos-Barajas et al., 2017). We could also include observation variables to the state-space model, in addition to the locations, to inform the behavioural states. For example, information about vertical movement has been used to identify behaviours in marine mammals (DeRuiter et al., 2017; McClintock et al., 2017). In our framework, the additional observation variables would have to be modelled with a normal transition density, to be integrated into the standard Kalman filter likelihood computations.

The computational cost of the method presented in this paper greatly depends on the number of MCMC iterations needed to obtain reliable estimates. Like for any MCMC algorithm, convergence should be checked, for example using trace plots of the posterior samples, or convergence diagnostics such as the potential scale reduction factor (Gelman et al., 2013). We can improve the mixing speed of the algorithm with the choice of the tuning parameters, i.e. the variances of the proposal distributions for the movement parameters, and the length of the time interval over which the state sequence should be updated at each iteration. In the applications of Sections 6.5 and 6.6, we chose the tuning parameters to obtain acceptance rates that are close to the optimal value (i.e. around 23%; see Roberts et al., 1997).

In this paper, we focused on the estimation of the hidden state sequence and of movement parameters, for the state-switching CTCRW model. However, this modelling framework offers other possibilities. In particular, it can be used to predict the location of the animal between observations (with associated uncertainty). This has been one of the main applications of the single-state CTCRW, as implemented in the R package crawl (Johnson and London, 2018), to obtain smooth estimates of a trajectory from noisy and irregular telemetry data (e.g. Robinson et al., 2012; Baylis et al., 2015; Rode et al., 2015). Predictions from the state-switching CTCRW model account for behavioural heterogeneity in the movement patterns, and are therefore more susceptible to provide realistic location and uncertainty estimates. More generally, all functionalities of the single-state CTCRW can be implemented in the state-switching case, based on the Kalman filter and smoother algorithms presented in Section 6.3.2.

Chapter 7

Discussion and future work

7.1 Implications of scale-free habitat selection models

The MCMC step selection model, introduced in Chapter 3, has important implications for the future development of habitat selection analyses. Our approach reconciles the two main existing approaches to estimating habitat preferences in animals: resource selection functions (RSFs), and step selection functions (SSFs). The two methods have been shown to yield different results, because they model the selection at two different scales (Moorcroft and Barnett, 2008). RSFs capture the selection at a global spatio-temporal scale, whereas SSFs describe the selection at the local scale of the movement step. The MCMC step selection model bridges the gap between the two scales, and it describes both local movement decisions and global space use as the response to the same underlying habitat selection mechanism. It therefore has the potential to improve our understanding of both individual animal movement and species distributions.

Models of individual movement provide a detailed description of the movement patterns, but they often fail to properly account for the effects of the environment. In recent years, there has been some effort to integrate environmental variables into analyses of telemetry data. In particular, SSFs were developed to model movement in response to habitat features. In state-switching models, such as hidden Markov models, it is possible to formulate the transition probabilities as functions of covariates, and this has been used to investigate the environmental drivers of animal behaviour (Patterson et al., 2009; Michelot et al., 2016; Towner et al., 2016). The parameters of movement can also be functions of environmental covariates, such that the speed or directionality of movement may vary based on the conditions (McClintock and Michelot, 2018). In fact, SSFs can describe increasingly complex movement patterns, and

hidden Markov models can incorporate more and more environmental information, so we can conjecture that the two approaches will converge in the future. Although these refinements are certainly a step forward to understand the environmental drivers of animal behaviour, existing individual-based movement models may only describe habitat selection at a local scale. A consequence is that they are typically not stationary models, and they cannot readily be used to estimate space use by animals (Michelot et al., 2017). The MCMC step selection model directly links movement and space use, which are both defined by the same habitat selection parameters.

The MCMC step selection approach introduced in this thesis does not offer the same flexibility as standard SSF models to capture short-term movement and habitat selection. Indeed, one advantage for SSF models is the possibility to express the movement process and the selection process separately. The movement model can then be tailored to capture detailed features of the observed movement. In the MCMC movement model, however, the flexibility of the movement model is restricted, because it needs to satisfy certain conditions to ensure stationarity. For example, in the models described in Chapters 3 and 4, we required that the transition density satisfies the detailed balance condition. In SSF analyses, habitat selection can also be defined with more flexibility. In particular, as mentioned in Section 3.8, the environmental covariates of SSFs can be measured over the whole movement step, rather than only at points in space. Thus, SSFs provide a more attractive framework for the analysis of short-term movement and local habitat selection.

Although the parameters of SSFs do not directly describe long-term space use, simulations or other numerical methods can be used to estimate the stationary distribution of a SSF (Potts et al., 2014a; Signer et al., 2017). In that context, it can be useful to compare the long-term predictions of the SSF with the observed long-term distribution of an animal, to illuminate aspects of habitat selection that are not captured by the SSF. We could imagine a similar diagnostic for the MCMC step selection model. For each observed step, we could simulate from the estimated movement model, and compare the one-step-forward simulations with the observed steps. This would uncover aspects of short-term movement and selection that are not captured by the MCMC step selection model.

Models of species distributions, and models of space use, focus on the global scale. They describe a long-term distribution over the whole study region. RSFs model space use as the consequence of habitat selection on that global scale. They do not provide a mechanistic description of this selection, however, because they ignore the dependence of the movement data. Because they do not estimate local habitat selection, RSFs suffer a certain degree of arbitrariness in the definition of habitat availability. In RSF analyses, the availability usually

covers the whole study region. The choice of the extent of the study region is arbitrary, and it can affect the estimated habitat selection parameters (Beyer et al., 2010). On the other hand, in step selection models (such as SSFs and the MCMC step selection model), availability is constrained by the movement model, and it can be determined from telemetry data. For example, in the local Gibbs model, habitat availability is captured by the transition density, which is estimated jointly with the parameters of selection.

Results from the MCMC step selection model are easier to interpret, and to compare across studies. They can consistently be interpreted on the small scale of movement and on the large scale of space use. This scalability means that inferences about habitat preferences do not depend on the sampling frequency of the data, or on an ad hoc definition of availability. The Langevin movement model has the same advantage over existing approaches. In addition, it describes the movement in continuous time, and can therefore be applied to tracking data with irregular time intervals.

7.2 Alternative MCMC step selection models

We described the local Gibbs sampler as the basis for a wide family of movement models. Properties of hybrid algorithms further increase the array of movement patterns that can be captured by the model. In particular, they can be used to model behavioural switching, as suggested in Section 3.6.2. However, the MCMC algorithms that we have used in Chapters 3 and 4 fail to capture persistence in the direction of movement. In Chapter 6, in a different context, we argued that movement persistence is ubiquitous in high-resolution telemetry data, and that it is an important feature for realistic models of movement and behaviour. It is difficult to introduce this type of autocorrelation in MCMC step selection models, because most MCMC algorithms are based on reversible Markov chains (Section 2.3). Reversible Markov chains satisfy the detailed balance condition, which guarantees a form of symmetry in the transition kernel and precludes movement persistence. As we suggested in Chapter 3, the possibility to model animal movement using non-reversible MCMC algorithms should be investigated.

Adaptive MCMC algorithms are another class of samplers that could be used to define more realistic movement models. The transition kernel of an adaptive algorithm evolves as time goes by, based on the current state and previous states of the process (Haario et al., 1999, 2001). This is a useful feature in MCMC sampling problems, to explore the target distribution more efficiently than with fixed transition rules. It may also be useful in the context of animal movement, because the mechanisms of habitat selection are susceptible to vary in time, as

the animal becomes familiar with its environment. Effectively, adaptive MCMC algorithms could be used to account for memory in MCMC step selection models.

The normal kernel model is a special case of the local Gibbs model, with a normal transition density. Although it is a simplistic model of movement, as illustrated in the case study of Chapter 4, its formulation makes it a good candidate to analyse irregular tracking data. Indeed, in Section 3.6.3, we proposed a slightly different parameterisation of the model, to account for irregular time intervals. We suggested that the variance parameter could be scaled by the time interval, to mimic the properties of the Brownian motion. In the simulations of Section 4.7.3, we found that the habitat selection and variance parameters of the model could be recovered from irregular data, using this formulation. It would be interesting to repeat a similar experiment, using real telemetry data. One possibility would be to thin the zebra track of Chapter 4 at random, to obtain an irregular data set. Then, we could fit the "irregular" normal kernel model to the thinned track, and check whether the estimated parameters are consistent with those found from the complete data set. From a more theoretical perspective, it would also be valuable to investigate the relationship between the irregular normal kernel model and continuous-time models, such as the Langevin movement model of Chapter 5.

7.3 Better inference for the local Gibbs model

We presented a method of inference for the local Gibbs model, based on maximum likelihood estimation. The likelihood function is generally not analytically tractable, and we used Monte Carlo approximation methods to evaluate it. The accuracy of the estimates therefore depends on the accuracy of the approximation, and will tend to increase as the size of the Monte Carlo samples increases (to the detriment of computational speed). In Chapter 4, we used simulations to find the Monte Carlo sample sizes that were needed to obtain accurate parameter estimates. The approximation of the likelihood also affects uncertainty quantification. Indeed, we derived standard errors from the Hessian of the likelihood function, evaluated at the estimated maximum. In practice, we evaluated the Hessian of the *approximate* likelihood, for different sizes of the Monte Carlo samples, to ensure that it converged for large enough samples. Although it would require very computationally intensive simulations, it would be interesting to check the coverage of the confidence intervals obtained with this method.

Other techniques of uncertainty estimation could also be considered, such as the bootstrap (Efron and Tibshirani, 1993). The bootstrap is a general method based on random resampling with replacement. It could be used as follows to obtain confidence intervals on the parameters

of the local Gibbs model. We consider a movement track of T locations $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$. We cannot directly resample the locations, because of the temporal structure of the data. However, we can consider the set of $T-1$ steps $(\boldsymbol{d}_1, \ldots, \boldsymbol{d}_{T-1})$, where $\boldsymbol{d}_t = \boldsymbol{x}_{t+1} - \boldsymbol{x}_t$. Based on the Markov assumption made in the local Gibbs model, the steps are independent. We generate a new set of $T-1$ steps, resampled with replacement from the set of observed steps. We fit the local Gibbs model to the new sample, and obtain estimates for the habitat selection and movement parameters. We repeat this a large number of times, and we use the empirical distribution of estimates to derive confidence intervals. This would be a very computational approach, because many replications would be needed to obtain reliable error estimates. It may be feasible for small data sets, and simple local Gibbs formulations.

Additional simulation studies could be implemented, to further investigate the performance of the local Gibbs model for different scenarios. In Chapter 4, we mentioned that the MCMC movement model may be used to obtain combined inference from (dependent) telemetry data and (independent) survey data. This is a great practical advantage, which follows from the scale-invariant properties of the habitat selection parameters of the model. The estimates of habitat selection would combine two different scales of selection: the short-term selection observed in the telemetry data, and the long-term selection observed in the survey data. It would be easy to set up a simulation experiment, to verify that this method works. We could simulate two data sets: a movement trajectory from the local Gibbs model, and a set of independent random samples from the (known) utilisation distribution. Then, we could compute the likelihood of the first data set using the local Gibbs model, and the likelihood of the second data set using standard RSF methods. We could then optimise the combined likelihood with respect to the habitat selection parameters, and obtain maximum likelihood estimates. This would be a valuable experiment to assess the gain in accuracy obtained from the combination of the two types of data.

7.4 Langevin movement model

7.4.1 Implementation

The Langevin movement model is a continuous-time model of movement and habitat selection. As such, it can directly be applied to tracking data collected at irregular time intervals, and parameter estimates can be compared across studies. However, the transition density of the underlying process is generally intractable, and methods of approximate inference must be used. We presented one such method, based on the Euler-Maruyama discretization scheme. The results of the simulation study of Section 5.5.3 suggest that, for a constant

number of observed locations, there is a trade-off between the precision and the accuracy of the estimation. If the time intervals of observation are short, the error due to the discretization is small, and the estimates are accurate. However, because the time intervals are short, the track may only cover a small geographical area, which leads to low precision. In contrast, observations obtained at long intervals lead to precise but biased estimates. In many cases, if the telemetry data have already been collected, we cannot choose the time interval of discretization, because it is the sampling interval. However, the observed trade-off can provide guidelines for the design of future studies. For practical reasons, it is often the case that the total number of locations retrieved from a telemetry tag is determined by the battery time. For a given study duration, the sampling rate is fixed, and the converse is also true (Brown et al., 2012). The choice of this trade-off could be based on the results of the simulations. The mean squared error of the estimates could for example be used to determine which time interval is preferable.

The implementation of the Langevin movement model requires the evaluation of the gradient of the covariate fields at the observed locations. We suggested that bilinear interpolation could be used to obtain continuous functions from the piecewise-constant covariates. It is a very convenient method, because the gradient of the interpolated function can easily be derived from the grid of values of the function. However, bilinear interpolation is a simple method, and it may not always be appropriate, e.g. for a very fragmented habitat covariate. The evaluation of the gradient of a two-dimensional surface, based on a discrete grid of observations, is a common problem in geostatistics (see e.g. Skidmore, 1989; Meyer et al., 2001, and references therein). It is possible that other approaches, such as finite differences, could work better in some situations. Different interpolation and gradient estimation methods should be compared in simulations to determine which one may be preferred in practice.

7.5 Further work on the state-switching CTCRW

In Chapter 6, we described a method of Bayesian inference for the state-switching CTCRW model, to estimate the movement parameters and behavioural states from telemetry data. We focused on the case where the data set only comprises one movement track. The method can be extended to the case of multiple tracks, e.g. if several individuals were observed. Here, we consider several independent tracks, and we assume that the movement parameters and the transition rates are shared by all individuals. The conditional likelihood derived from the Kalman filter can be computed separately for each track, and the full likelihood is the product of the individual likelihoods. The MCMC algorithm described in Section 6.4 requires only minor alterations. The update of the state process (Section 6.4.1.1) must be done separately

for each track. In practice, we may first choose a track, and then apply the update step described in the single-track case. To improve mixing, we could also try to update the state sequences of several tracks at once, although this would generally reduce the acceptance rate. The update of the movement parameters (Section 6.4.1.2) is unchanged, because the parameters are assumed to be shared by all individuals. The update of the transition rates (Section 6.4.1.3) is also almost unchanged. The derivation of the posterior distribution of the transition rates is based on the number of state transitions and the time spent in each state, which can be calculated separately for each track, and then summed.

We presented the state-switching CTCRW model as a good alternative to the widely-used hidden Markov models, because it is formulated in continuous time. In Section 6.5.3, we used simulations to highlight the difference in the scaling properties of the parameters of discrete-time and continuous-time models. We found that the parameters of discrete-time models depend on the time scale of the observations This is not surprising, because they describe metrics of movement that are only defined for a particular time interval of observation (e.g. the step lengths and turning angles). On the contrary, estimates of the parameters of the state-switching CTCRW did not depend on the sampling frequency. These results illustrate the limitations of discrete-time models, in particular to analyse data collected irregularly in time. A more extensive simulation study could reveal the practical consequences of this limitation of discrete-time analyses. In particular, it would be interesting to investigate the effect of the time scale of observation on the estimated behavioural states. The classification of the movement tracks into "behavioural phases" may be sensitive to the sampling frequency, which would suggest that great caution is needed in their interpretation.

Model checking could be used to verify that the state-switching CTCRW appropriately captures features of the real data. Conn et al. (2018) presented a review of the Bayesian model checking techniques that have been used in ecological studies. The most common method is posterior predictive checks (e.g. Morales et al., 2004). In the context of the state-switching CTCRW model, we could simulate a movement track from each posterior sample, and compare the simulated movement to the real (observed) movement of the animal. To assess the movement component of the model, we could for example compare the step lengths in the observed data with the simulated step lengths. To assess the behavioural component, we could compare the proportion of time spent in each state, or the mean dwell times. Following Morales et al. (2004), we could also compare the autocorrelation structure of the observed step lengths to that of the simulated step lengths. Discrepancies between features of the real and simulated data point to model misspecifications.

References

Aarts, G., Fieberg, J. and Matthiopoulos, J. (2012). Comparative interpretation of count, presence–absence and point methods for species distribution models, *Methods in Ecology and Evolution*, **3 (1)**: 177–187.

Anderson-Sprecher, R. and Ledolter, J. (1991). State-space analysis of wildlife telemetry data, *Journal of the American Statistical Association*, **86 (415)**: 596–602.

Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC, *Statistics and Computing*, **18 (4)**: 343–373.

Arthur, S. M., Manly, B. F., McDonald, L. L. and Garner, G. W. (1996). Assessing habitat selection when availability changes, *Ecology*, **77 (1)**: 215–227.

Avgar, T., Baker, J. A., Brown, G. S., Hagens, J. S., Kittle, A. M., Mallon, E. E., McGreer, M. T., Mosser, A., Newmaster, S. G., Patterson, B. R. et al. (2015). Space-use behaviour of woodland caribou based on a cognitive movement model, *Journal of Animal Ecology*, **84 (4)**: 1059–1070.

Avgar, T., Potts, J. R., Lewis, M. A. and Boyce, M. S. (2016). Integrated step selection analysis: bridging the gap between resource selection and animal movement, *Methods in Ecology and Evolution*, **7 (5)**: 619–630.

Barnett, A. H. and Moorcroft, P. R. (2008). Analytic steady-state space use patterns and rapid computations in mechanistic home range analysis, *Journal of Mathematical Biology*, **57 (1)**: 139–159.

Bastille-Rousseau, G., Murray, D. L., Schaefer, J. A., Lewis, M. A., Mahoney, S. P. and Potts, J. R. (2018). Spatial scales of habitat selection decisions: implications for telemetry-based movement modelling, *Ecography*, **41 (3)**: 437–443.

Baylis, A., Orben, R., Arnould, J., Peters, K., Knox, T., Costa, D. and Staniland, I. (2015). Diving deeper into individual foraging specializations of a large marine predator, the Southern sea lion, *Oecologia*, **179 (4)**: 1053–1065.

Berman, S. M. (1994). A bivariate Markov process with diffusion and discrete components, *Stochastic Models*, **10 (2)**: 271–308.

Bestley, S., Jonsen, I. D., Hindell, M. A., Guinet, C. and Charrassin, J.-B. (2012). Integrative modelling of animal movement: incorporating in situ habitat and behavioural information for a migratory marine predator, *Proceedings of the Royal Society of London B: Biological Sciences*, p. rspb20122262.

Beyer, H. L., Haydon, D. T., Morales, J. M., Frair, J. L., Hebblewhite, M., Mitchell, M. and Matthiopoulos, J. (2010). The interpretation of habitat preference metrics under use–availability designs, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **365 (1550)**: 2245–2254.

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

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

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

Blackwell, P. G., Niu, M., Lambert, M. S. and LaPoint, S. D. (2016). Exact Bayesian inference for animal movement in continuous time, *Methods in Ecology and Evolution*, **7 (2)**: 184–195.

Block, B. A., Jonsen, I. D., Jorgensen, S. J., Winship, A. J., Shaffer, S. A., Bograd, S. J., Hazen, E. L., Foley, D. G., Breed, G., Harrison, A.-L. et al. (2011). Tracking apex marine predator movements in a dynamic ocean, *Nature*, **475 (7354)**: 86.

Bovet, P. and Benhamou, S. (1988). Spatial analysis of animals' movements using a correlated random walk model, *Journal of Theoretical Biology*, **131 (4)**: 419–433.

Boyce, M. S. and McDonald, L. L. (1999). Relating populations to habitats using resource selection functions, *Trends in Ecology & Evolution*, **14 (7)**: 268–272.

Brillinger, D. R., Preisler, H. K., Ager, A. A. and Kie, J. (2001). The use of potential functions in modelling animal movement, in *Data Analysis from Statistical Foundations*, edited by Saleh, A. E., pp. 369–386, Nova Science Publishers.

Brown, D. D., Kays, R., Wikelski, M., Wilson, R. and Klimley, A. P. (2013). Observing the unwatchable through acceleration logging of animal behavior, *Animal Biotelemetry*, **1 (1)**: 20.

Brown, D. D., LaPoint, S., Kays, R., Heidrich, W., Kümmeth, F. and Wikelski, M. (2012). Accelerometer-informed GPS telemetry: reducing the trade-off between resolution and longevity, *Wildlife Society Bulletin*, **36 (1)**: 139–146.

Brown, R. (1828). A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies, *The Philosophical Magazine*, **4 (21)**: 161–173.

Cagnacci, F., Boitani, L., Powell, R. A. and Boyce, M. S. (2010). Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **365 (1550)**: 2157–2162.

Chang, K.-T. (2018). *Introduction to geographic information systems, Ninth Edition*, McGraw-Hill Education, New York.

Codling, E. and Hill, N. (2005). Sampling rate effects on measurements of correlated and biased random walks, *Journal of Theoretical Biology*, **233 (4)**: 573–588.

Codling, E. A., Plank, M. J. and Benhamou, S. (2008). Random walk models in biology, *Journal of the Royal Society Interface*, **5 (25)**: 813–834.

Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R. and Hooten, M. B. (2018). A guide to Bayesian model checking for ecologists, *Ecological Monographs*, **88 (4)**: 526–542.

Coulon, A., Morellet, N., Goulard, M., Cargnelutti, B., Angibault, J.-M. and Hewison, A. M. (2008). Inferring the effects of landscape structure on roe deer (Capreolus capreolus) movements using a step selection function, *Landscape Ecology*, **23 (5)**: 603–614.

Courbin, N., Loveridge, A. J., Macdonald, D. W., Fritz, H., Valeix, M., Makuwe, E. T. and Chamaillé-Jammes, S. (2016). Reactive responses of zebras to lion encounters shape their predator–prey space game at large scale, *Oikos*, **125 (6)**: 829–838.

DeRuiter, S. L., Langrock, R., Skirbutas, T., Goldbogen, J. A., Calambokidis, J., Friedlaender, A. S., Southall, B. L. et al. (2017). A multivariate mixed hidden Markov model for blue whale behaviour and responses to sound exposure, *The Annals of Applied Statistics*, **11 (1)**: 362–392.

Duchesne, T., Fortin, D. and Rivest, L.-P. (2015). Equivalence between step selection functions and biased correlated random walks for statistical inference on animal movement, *PloS One*, **10 (4)**: e0122947.

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

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*, vol. 38, Oxford University Press, Oxford.

Edelhoff, H., Signer, J. and Balkenhol, N. (2016). Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns, *Movement Ecology*, **4 (1)**: 21.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap*, vol. 57 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, New York.

Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, *Annalen der Physik*, **322 (8)**: 549–560.

Fintzi, J. (2018). *ECctmc: Simulation from Endpoint-Conditioned Continuous Time Markov Chains*, R package version 0.2.5.

Fleming, C. H., Calabrese, J. M., Mueller, T., Olson, K. A., Leimgruber, P. and Fagan, W. F. (2014). From fine-scale foraging to home ranges: a semivariance approach to identifying movement modes across spatiotemporal scales, *The American Naturalist*, **183 (5)**: E154–E167.

Fleming, C. H., Sheldon, D., Gurarie, E., Fagan, W. F., LaPoint, S. and Calabrese, J. M. (2017). Kálmán filters for continuous-time movement models, *Ecological Informatics*, **40**: 8–21.

Forester, J. D., Im, H. K. and Rathouz, P. J. (2009). Accounting for animal movement in estimation of resource selection functions: sampling and data analysis, *Ecology*, **90 (12)**: 3554–3565.

Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T. and Mao, J. S. (2005). Wolves influence elk movements: behavior shapes a trophic cascade in Yellowstone National Park, *Ecology*, **86 (5)**: 1320–1330.

Frair, J. L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N. J. and Pedrotti, L. (2010). Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **365 (1550)**: 2187–2200.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85 (410)**: 398–409.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian data analysis, Third Edition*, Chapman and Hall/CRC, Boca Raton.

Gilbert, P. and Varadhan, R. (2016). *numDeriv: Accurate Numerical Derivatives*, R package version 2016.8-1.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. (1996). *Markov chain Monte Carlo in practice*, CRC press, London.

Gilks, W. R., Roberts, G. O. and Sahu, S. K. (1998). Adaptive Markov chain Monte Carlo through regeneration, *Journal of the American Statistical Association*, **93 (443)**: 1045–1054.

Gillespie, D. T. (1996). Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral, *Physical review E*, **54 (2)**: 2084.

Gloaguen, P., Etienne, M.-P. and Le Corff, S. (2018). Stochastic differential equation based on a multimodal potential to model movement data in ecology, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **67 (3)**: 599–619.

Grecian, W. J., Lane, J. V., Michelot, T., Wade, H. M. and Hamer, K. C. (2018). Understanding the ontogeny of foraging behaviour: insights from combining marine predator bio-logging with satellite-derived oceanography in hidden Markov models, *Journal of The Royal Society Interface*, **15 (143)**: 20180084.

Gurarie, E., Andrews, R. D. and Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data, *Ecology Letters*, **12 (5)**: 395–408.

Gurarie, E., Fleming, C. H., Fagan, W. F., Laidre, K. L., Hernández-Pliego, J. and Ovaskainen, O. (2017). Correlated velocity models as a fundamental unit of animal movement: synthesis and applications, *Movement Ecology*, **5 (1)**: 13.

Gurarie, E. and Ovaskainen, O. (2011). Characteristic spatial and temporal scales unify models of animal movement, *The American Naturalist*, **178 (1)**: 113–123.

Guttorp, P. (1995). *Stochastic Modeling of Scientific Data*, Chapman and Hall/CRC.

Haario, H., Saksman, E. and Tamminen, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm, *Computational Statistics*, **14 (3)**: 375–396.

Haario, H., Saksman, E., Tamminen, J. et al. (2001). An adaptive Metropolis algorithm, *Bernoulli*, **7 (2)**: 223–242.

Hanks, E. M., Hooten, M. B., Alldredge, M. W. et al. (2015). Continuous-time discrete-space models for animal movement, *The Annals of Applied Statistics*, **9 (1)**: 145–165.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57 (1)**: 97–109.

Hays, G., Akesson, S., Godley, B., Luschi, P. and Santidrian, P. (2001). The implications of location accuracy for the interpretation of satellite-tracking data, *Animal Behaviour*, **61 (5)**: 1035–1040.

Hebblewhite, M. and Merrill, E. (2008). Modelling wildlife–human relationships for social species with mixed-effects resource selection models, *Journal of applied ecology*, **45 (3)**: 834–844.

Hebblewhite, M., Merrill, E. and McDonald, T. (2005). Spatial decomposition of predation risk using resource selection functions: an example in a wolf–elk predator–prey system, *Oikos*, **111 (1)**: 101–111.

Hobolth, A. and 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 (3)**: 1204.

Hooten, M. B., Hanks, E. M., Johnson, D. S. and Alldredge, M. W. (2013). Reconciling resource utilization and resource selection functions, *Journal of Animal Ecology*, **82 (6)**: 1146–1154.

Hooten, M. B., Johnson, D. S., McClintock, B. T. and Morales, J. M. (2017). *Animal movement: statistical models for telemetry data*, CRC Press, Boca Raton.

Hooten, M. B., Scharf, H. R. and Morales, J. M. (2018). Running on empty: Recharge dynamics from animal movement data, *Ecology Letters*, DOI:10.1111/ele.13198.

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

Iacus, S. M. (2009). *Simulation and inference for stochastic differential equations: with R examples*, Springer Science & Business Media, New York.

Jacobson, M. T. and Matthews, P. (1996). Generating uniformly distributed random Latin squares, *Journal of Combinatorial Designs*, **4 (6)**: 405–437.

Johnson, C. J., Nielsen, S. E., Merrill, E. H., McDonald, T. L. and Boyce, M. S. (2006). Resource selection functions based on use–availability data: theoretical motivation and evaluation methods, *Journal of Wildlife Management*, **70 (2)**: 347–357.

Johnson, C. J., Seip, D. R. and Boyce, M. S. (2004). A quantitative approach to conservation planning: using resource selection functions to map the distribution of mountain caribou at multiple spatial scales, *Journal of Applied Ecology*, **41 (2)**: 238–251.

Johnson, D. H. (1980). The comparison of usage and availability measurements for evaluating resource preference, *Ecology*, **61 (1)**: 65–71.

Johnson, D. S. and London, J. M. (2018). crawl: an R package for fitting continuous-time correlated random walk models to animal movement data.

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

Johnson, D. S., Thomas, D. L., Ver Hoef, J. M. and Christ, A. (2008b). A general framework for the analysis of animal resource selection from telemetry data, *Biometrics*, **64 (3)**: 968–976.

Jonsen, I. D., Flemming, J. M. and Myers, R. A. (2005). Robust state–space modeling of animal movement data, *Ecology*, **86 (11)**: 2874–2880.

Jonsen, I. D., Myers, R. A. and Flemming, J. M. (2003). Meta-analysis of animal movement using state-space models, *Ecology*, **84 (11)**: 3055–3063.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, *Journal of basic Engineering*, **82 (1)**: 35–45.

Kareiva, P. and Shigesada, N. (1983). Analyzing insect movement as a correlated random walk, *Oecologia*, **56 (2-3)**: 234–238.

Kays, R., Crofoot, M. C., Jetz, W. and Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet, *Science*, **348 (6240)**: aaa2478.

Kessler, M., Lindner, A. and Sorensen, M. (2012). *Statistical methods for stochastic differential equations*, Boca Raton, FL: Chapman and Hall/CRC.

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

Lele, S. R. and Keim, J. L. (2006). Weighted distributions and estimation of resource selection probability functions, *Ecology*, **87 (12)**: 3021–3028.

Leos-Barajas, V., Photopoulou, T., Langrock, R., Patterson, T. A., Watanabe, Y. Y., Murgatroyd, M. and Papastamatiou, Y. P. (2017). Analysis of animal accelerometer data using hidden Markov models, *Methods in Ecology and Evolution*, **8 (2)**: 161–173.

Lopez, R., Malarde, J.-P., Royer, F. and Gaspar, P. (2014). Improving Argos doppler location using multiple-model Kalman filtering, *IEEE Transactions on Geoscience and Remote Sensing*, **52 (8)**: 4744–4755.

Manly, B., McDonald, L., Thomas, D., McDonald, T. L. and Erickson, W. P. (2002). *Resource selection by animals: statistical design and analysis for field studies, Second Edition*, Kluwer Academic Publishers, Dordrecht.

Marzluff, J. M., Millspaugh, J. J., Hurvitz, P. and Handcock, M. S. (2004). Relating resources to a probabilistic measure of space use: forest fragments and steller's jays, *Ecology*, **85 (5)**: 1411–1427.

Mathai, A. M. (1999). *An introduction to geometrical probability: distributional aspects with applications*, vol. 1, CRC Press.

McClintock, B. T. (2017). Incorporating telemetry error into hidden Markov models of animal movement using multiple imputation, *Journal of Agricultural, Biological and Environmental Statistics*, **22 (3)**: 249–269.

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

McClintock, B. T., London, J. M., Cameron, M. F. and Boveng, P. L. (2017). Bridging the gaps in animal movement: hidden behaviors and ecological relationships revealed by integrated data streams, *Ecosphere*, **8 (3)**: e01751.

McClintock, B. T. and Michelot, T. (2018). momentuHMM: R package for generalized hidden Markov models of animal movement, *Methods in Ecology and Evolution*, **9 (6)**: 1518–1530.

McConnell, B. J. (2019). Data from: State-switching continuous-time correlated random walks, Movebank data repository, doi:10.5441/001/1.m7j2263r.

McKay, M. D., Beckman, R. J. and Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, **21 (2)**: 239–245.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines, *The Journal of Chemical Physics*, **21 (6)**: 1087–1092.

Meyer, T. H., Eriksson, M. and Maggio, R. C. (2001). Gradient estimation from irregularly spaced data sets, *Mathematical Geology*, **33 (6)**: 693–717.

Michel, M. and Sénécal, S. (2017). Forward event-chain Monte Carlo: a general rejection-free and irreversible Markov chain simulation method, *arXiv preprint arXiv:1702.08397*.

Michelot, T. and Blackwell, P. G. (2019). State-switching continuous-time correlated random walks, *Methods in Ecology and Evolution*, DOI: 10.1111/2041-210X.13154.

Michelot, T., Blackwell, P. G., Chamaillé-Jammes, S. and Matthiopoulos, J. (2018a). Inference in MCMC step selection models, *arXiv preprint arXiv:1810.10930*.

Michelot, T., Blackwell, P. G. and Matthiopoulos, J. (2019). Linking resource selection and step selection models for habitat preferences in animals, *Ecology*, **100 (1)**: e02452.

Michelot, T., Gloaguen, P. and Étienne, M.-P. (2018b). The Langevin diffusion as a continuous-time model of animal movement and habitat selection, *arXiv preprint arXiv:1810.10213*.

Michelot, T., Langrock, R., Bestley, S., Jonsen, I. D., Photopoulou, T. and Patterson, T. A. (2017). Estimation and simulation of foraging trips in land-based marine predators, *Ecology*, **98 (7)**: 1932–1944.

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

Millspaugh, J. J., Nielson, R. M., McDonald, L., Marzluff, J. M., Gitzen, R. A., Rittenhouse, C. D., Hubbard, M. W. and Sheriff, S. L. (2006). Analysis of resource selection using utilization distributions, *The Journal of Wildlife Management*, **70 (2)**: 384–395.

Moorcroft, P. R. and Barnett, A. (2008). Mechanistic home range models and resource selection analysis: a reconciliation and unification, *Ecology*, **89 (4)**: 1112–1119.

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

Nadarajah, S., Kotz, S. et al. (2006). R programs for truncated distributions, *Journal of Statistical Software*, **16 (c02)**.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization, *The computer journal*, **7 (4)**: 308–313.

Neu, C. W., Byers, C. R. and Peek, J. M. (1974). A technique for analysis of utilization-availability data, *The Journal of Wildlife Management*, pp. 541–545.

Nicosia, A., Duchesne, T., Rivest, L.-P., Fortin, D. et al. (2017). A multi-state conditional logistic regression model for the analysis of animal movement, *The Annals of Applied Statistics*, **11 (3)**: 1537–1560.

Nielsen, S. E., Boyce, M. S., Stenhouse, G. B. and Munro, R. H. (2003). Development and testing of phenologically driven grizzly bear habitat models, *Ecoscience*, **10 (1)**: 1–10.

Northrup, J. M., Hooten, M. B., Anderson, C. R. and Wittemyer, G. (2013). Practical guidance on characterizing availability in resource selection functions under a use–availability design, *Ecology*, **94 (7)**: 1456–1463.

Øksendal, B. (2003). *Stochastic differential equations*, Springer.

Ovaskainen, O. (2004). Habitat-specific movement parameters estimated using mark–recapture data and a diffusion model, *Ecology*, **85 (1)**: 242–257.

Ovaskainen, O. (2008). Analytical and numerical tools for diffusion-based movement models, *Theoretical Population Biology*, **73 (2)**: 198–211.

Ovaskainen, O. and Cornell, S. J. (2003). Biased movement at a boundary and conditional occupancy times for diffusion processes, *Journal of Applied Probability*, **40 (3)**: 557–580.

Parton, A. and Blackwell, P. G. (2017). Bayesian inference for multistate 'step and turn' animal movement in continuous time, *Journal of Agricultural, Biological and Environmental Statistics*, **22 (3)**: 373–392.

Patterson, T. A., Basson, M., Bravington, M. V. and Gunn, J. S. (2009). Classifying movement behaviour in relation to environmental conditions using hidden Markov models, *Journal of Animal Ecology*, **78 (6)**: 1113–1123.

Patterson, T. A., McConnell, B. J., Fedak, M. A., Bravington, M. V. and Hindell, M. A. (2010). Using GPS data to evaluate the accuracy of state–space methods for correction of Argos satellite telemetry error, *Ecology*, **91 (1)**: 273–285.

Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L. and King, R. (2017). Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges, *AStA Advances in Statistical Analysis*, DOI: 10.1007/s10182-017-0302-7.

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

Pearce, J. L. and Boyce, M. S. (2006). Modelling distribution and abundance with presence-only data, *Journal of Applied Ecology*, **43 (3)**: 405–412.

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

Pohle, J., Langrock, R., van Beest, F. M. and Schmidt, N. M. (2017). Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement, *Journal of Agricultural, Biological and Environmental Statistics*, **22 (3)**: 270–293.

Porter, W. F. and Church, K. E. (1987). Effects of environmental pattern on habitat preference analysis, *The Journal of Wildlife Management*, pp. 681–685.

Potts, J. R., Bastille-Rousseau, G., Murray, D. L., Schaefer, J. A. and Lewis, M. A. (2014a). Predicting local and non-local effects of resources on animal space use using a mechanistic step selection model, *Methods in Ecology and Evolution*, **5 (3)**: 253–262.

Potts, J. R., Mokross, K. and Lewis, M. A. (2014b). A unifying framework for quantifying the nature of animal interactions, *Journal of The Royal Society Interface*, **11 (96)**: 20140333.

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

Preisler, H. K., Ager, A. A. and Wisdom, M. J. (2013). Analyzing animal movement patterns using potential functions, *Ecosphere*, **4 (3)**: 1–13.

Prokopenko, C. M., Boyce, M. S. and Avgar, T. (2017). Characterizing wildlife behavioural responses to roads using integrated step selection analysis, *Journal of Applied Ecology*, **54 (2)**: 470–479.

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

Rhodes, J. R., McAlpine, C. A., Lunney, D. and Possingham, H. P. (2005). A spatially explicit habitat selection model incorporating home range behavior, *Ecology*, **86 (5)**: 1199–1205.

Robert, C. and Casella, G. (2013). *Monte Carlo statistical methods*, Springer Science & Business Media.

Roberts, G. O., Gelman, A., Gilks, W. R. et al. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms, *The Annals of Applied Probability*, **7 (1)**: 110–120.

Roberts, G. O. and Rosenthal, J. S. (1998). Optimal scaling of discrete approximations to langevin diffusions, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **60 (1)**: 255–268.

Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC, *Journal of Computational and Graphical Statistics*, **18 (2)**: 349–367.

Roberts, G. O., Tweedie, R. L. et al. (1996). Exponential convergence of Langevin distributions and their discrete approximations, *Bernoulli*, **2 (4)**: 341–363.

Robinson, P. W., Costa, D. P., Crocker, D. E., Gallo-Reynoso, J. P., Champagne, C. D., Fowler, M. A., Goetsch, C., Goetz, K. T., Hassrick, J. L., Hückstädt, L. A. et al. (2012). Foraging behavior and success of a mesopelagic predator in the Northeast Pacific Ocean: insights from a data-rich species, the Northern elephant seal, *PloS One*, **7 (5)**: e36728.

Rode, K. D., Wilson, R. R., Regehr, E. V., Martin, M. S., Douglas, D. C. and Olson, J. (2015). Increased land use by Chukchi sea polar bears in relation to changing sea ice conditions, *PloS One*, **10 (11)**: e0142213.

Roever, C. L., Beyer, H., Chase, M. and Van Aarde, R. J. (2014). The pitfalls of ignoring behaviour when quantifying habitat selection, *Diversity and Distributions*, **20 (3)**: 322–333.

Roever, C. L., Boyce, M. S. and Stenhouse, G. B. (2010). Grizzly bear movements relative to roads: application of step selection functions, *Ecography*, **33 (6)**: 1113–1122.

Russell, D. J., McClintock, B. T., Matthiopoulos, J., Thompson, P. M., Thompson, D., Hammond, P. S., Jones, E. L., MacKenzie, M. L., Moss, S. and McConnell, B. J. (2015). Intrinsic and extrinsic drivers of activity budgets in sympatric grey and harbour seals, *Oikos*, **124 (11)**: 1462–1472.

Russell, J. C., Hanks, E. M., Haran, M., Hughes, D. et al. (2018). A spatially varying stochastic differential equation model for animal movement, *The Annals of Applied Statistics*, **12 (2)**: 1312–1331.

Scharf, H., Hooten, M. B. and Johnson, D. S. (2017). Imputation approaches for animal movement modeling, *Journal of Agricultural, Biological and Environmental Statistics*, **22 (3)**: 335–352.

Schlägel, U. E. and Lewis, M. A. (2016). Robustness of movement models: can models bridge the gap between temporal scales of data sets and behavioural processes?, *Journal of Mathematical Biology*, **73 (6-7)**: 1691–1726.

Schlather, M., Malinowski, A., Menck, P. J., Oesting, M. and Strokorb, K. (2015). Analysis, simulation and prediction of multivariate random fields with package RandomFields, *Journal of Statistical Software*, **63 (8)**: 1–25.

Scrafford, M. A., Avgar, T., Heeres, R. and Boyce, M. S. (2018). Roads elicit negative movement and habitat-selection responses by wolverines (Gulo gulo luscus), *Behavioral Ecology*, p. arx182.

Signer, J., Fieberg, J. and Avgar, T. (2017). Estimating utilization distributions from fitted step-selection functions, *Ecosphere*, **8 (4)**.

Siniff, D. B. and Jessen, C. (1969). A simulation model of animal movement patterns, in *Advances in ecological research*, vol. 6, pp. 185–219, Elsevier.

Skidmore, A. K. (1989). A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model, *International Journal of Geographical Information System*, **3 (4)**: 323–334.

Tang, C. Y. and Chen, S. X. (2009). Parameter estimation and bias correction for diffusion processes, *Journal of Econometrics*, **149 (1)**: 65–81.

Thurfjell, H., Ciuti, S. and Boyce, M. S. (2014). Applications of step-selection functions in ecology and conservation, *Movement ecology*, **2 (1)**: 4.

Tierney, L. (1994). Markov chains for exploring posterior distributions, *Annals of Statistics*, **22**: 1701–1728.

Towner, A. V., Leos-Barajas, V., Langrock, R., Schick, R. S., Smale, M. J., Kaschke, T., Jewell, O. J. and Papastamatiou, Y. P. (2016). Sex-specific and individual preferences for hunting strategies in white sharks, *Functional Ecology*, **30 (8)**: 1397–1407.

Turchin, P. (1998). *Quantitative analysis of movement: Measuring and modeling population redistribution in animals and plants*, Beresta Books.

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the Brownian motion, *Physical Review*, **36 (5)**: 823.

White, G. C. and Garrott, R. A. (2012). *Analysis of wildlife radio-tracking data*, Elsevier.

Whitehead, H. and Jonsen, I. D. (2013). Inferring animal densities from tracking data using Markov chains, *PloS One*, **8 (4)**: e60901.

Wilson, K., Hanks, E. and Johnson, D. (2018). Estimating animal utilization densities using continuous-time Markov chain models, *Methods in Ecology and Evolution*, **9 (5)**: 1232–1240.

Xifara, T., Sherlock, C., Livingstone, S., Byrne, S. and Girolami, M. (2014). Langevin diffusions and the Metropolis-adjusted Langevin algorithm, *Statistics & Probability Letters*, **91**: 14–19.

Zucchini, W., MacDonald, I. L. and Langrock, R. (2016). *Hidden Markov models for time series: an introduction using R, Second Edition*, Chapman and Hall/CRC.

Appendix A

Convolution of normal densities

Here, we verify that the resource-independent step density of the normal kernel model is a normal density with mean $\boldsymbol{x}_t$ and variance $2\sigma^2$. Because we consider the isotropic case, we only need to prove it in either dimension. For a flat target distribution, the likelihood of a (one-dimensional) transition from x to y is

$$
\begin{aligned}
p_0(y|x) &= \int p(y|\mu)p(\mu|x)d\mu \\
&= \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\mu-x)^2}{2\sigma^2}\right] d\mu \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{y^2+\mu^2-2\mu y}{2\sigma^2} - \frac{\mu^2+x^2-2\mu x}{2\sigma^2}\right] d\mu \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{2\mu^2-2\mu y-2\mu x}{2\sigma^2} - \frac{x^2+y^2}{2\sigma^2}\right] d\mu \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\mu^2-\mu(x+y)}{\sigma^2} - \frac{x^2+y^2}{2\sigma^2}\right] d\mu \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\mu-(x+y)/2)^2-(x+y)^2/4}{\sigma^2} - \frac{x^2+y^2}{2\sigma^2}\right] d\mu \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{x^2+y^2}{2\sigma^2} + \frac{(x+y)^2}{4\sigma^2}\right] \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\mu-(x+y)/2)^2}{\sigma^2}\right] d\mu \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y-x)^2}{4\sigma^2}\right] \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\mu-(x+y)/2)^2}{\sigma^2}\right] d\mu.
\end{aligned}
$$

We define $\omega = \sigma/\sqrt{2}$, and substitute $\sqrt{2}\omega$ for σ in the integral,

$$\begin{aligned} p_0(y|x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y-x)^2}{4\sigma^2}\right] \int \frac{1}{\sqrt{2\pi}\sqrt{2}\omega} \exp\left[-\frac{(\mu-(x+y)/2)^2}{2\omega^2}\right] d\mu \\ &= \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma} \exp\left[-\frac{(y-x)^2}{4\sigma^2}\right] \int \frac{1}{\sqrt{2\pi}\omega} \exp\left[-\frac{(\mu-(x+y)/2)^2}{2\omega^2}\right] d\mu. \end{aligned}$$

The function in the integral is a normal density with mean $(x+y)/2$ and variance ω^2, so its integral is 1. The expression becomes

$$p_0(y|x) = \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma} \exp\left[-\frac{(y-x)^2}{4\sigma^2}\right],$$

which is the probability density function of the normal distribution with mean x and variance $2\sigma^2$.

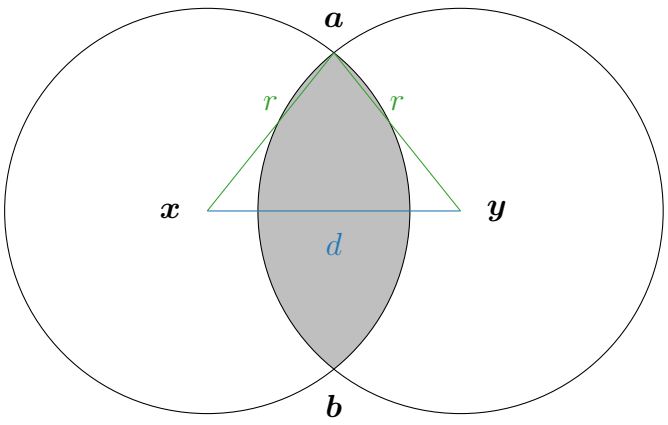
We have shown that the resource-independent step density of the normal kernel model is a normal distribution, centred on the current location, and with variance $2\sigma^2$.

Appendix B

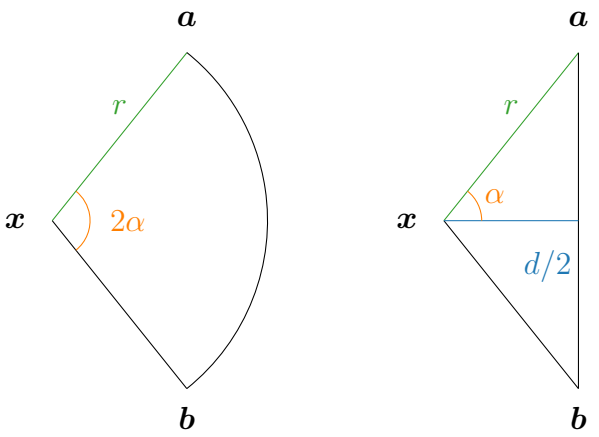
Intersection of two discs

B.1 Area of intersection

In the following, we show that the area $A(D_r(\boldsymbol{x}) \cap D_r(\boldsymbol{y}))$ can be written in terms of the radius r and of $d = \|\boldsymbol{y} - \boldsymbol{x}\|$, the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$. We focus on the case where $d \leq 2r$ (because the area of intersection is zero otherwise). The graph below shows the configuration.



The area of intersection of the two discs is shown in grey, and can be calculated as follows. Consider the two subregions of the disc shown below,



The area of intersection of the two discs is twice the difference between the area of the slice (left) and the area of the triangle (right). The area of the slice is $A_S = \alpha r^2$, and we can write α in terms of r and d, using trigonometry in the rectangle:

$$\alpha = \cos^{-1}\left(\frac{d}{2r}\right),$$

such that

$$A_S = r^2 \cos^{-1}\left(\frac{d}{2r}\right).$$

The area of the triangle is

$$\begin{aligned} A_T &= \frac{d}{2} r \sin(\alpha) \\ &= \frac{rd}{2} \sin\left(\cos^{-1}\left(\frac{d}{2r}\right)\right). \end{aligned}$$

Using the property that $\sin(\cos^{-1}(a)) = \sqrt{1-a^2}$, this becomes

$$A_T = \frac{rd}{2}\sqrt{1 - \frac{d^2}{4r^2}}.$$

Eventually, the area of intersection of $D_r(\boldsymbol{x})$ and $D_r(\boldsymbol{y})$ is

$$A(D_r(\boldsymbol{x}) \cap D_r(\boldsymbol{y})) = 2(A_S - A_T) = 2r^2 \cos^{-1}\left(\frac{d}{2r}\right) - rd\sqrt{1 - \frac{d^2}{4r^2}}$$

if $d < 2r$, and 0 otherwise. For a given r, this area only depends on the distance d between the centres $\boldsymbol{x}$ and $\boldsymbol{y}$ of the two discs.

B.2 Acceptance rate of sampling procedure

In Section 4.4.2, we proposed a method to generate samples from the intersection of two discs, based on the smallest rectangle that includes the intersection (see Figure 4.2). As before, we denote by r the radius of the discs, and d the distance between their centres. In this appendix, we calculate the expected acceptance rate of the proposed sampling procedure when $d = 0$, and when it tends to $2r$.

When $d = 0$, the rectangle is a square with sides of length $2r$, with area $4r^2$. In that case, the expected acceptance rate Q is the ratio of the area of a disc of radius r and the area of that square, i.e. $Q = (\pi r^2)/(4r^2) = \pi/4$.

The expected acceptance rate is undefined for $d = 2r$, so we derive its limit when $d \to 2r$. In general, for $0 \leq d < 2r$, the expected acceptance rate is given by the ratio

$$Q = \frac{\mathcal{A}_r(d)}{\mathcal{A}_{\text{rect}}},$$

where $\mathcal{A}_r(d)$ is the area of intersection of the two discs, and $\mathcal{A}_{\text{rect}}$ is the area of the rectangle. We previously found

$$\mathcal{A}_r(d) = 2r\cos^{-1}\left(\frac{d}{2r}\right) - rd\sqrt{1 - \frac{d^2}{4r^2}},$$

and it is easy to see that

$$\mathcal{A}_{\text{rect}} = (2r - d)\sqrt{4r^2 - d^2}.$$

To simplify the expression, we define $c = d/(2r)$, and we can write

$$\begin{aligned} Q(c) &= \frac{2r^2\cos^{-1}(c) - 2r^2c\sqrt{1-c^2}}{2r(1-c)\sqrt{2r(1-c)}\sqrt{2r(1+c)}} \\ &= \frac{1}{2}\left[\frac{\cos^{-1}(c) - c\sqrt{1-c^2}}{(1-c)\sqrt{1-c^2}}\right]. \end{aligned}$$

We are looking for the limit of Q when $d \to 2r$, i.e. when $c \to 1$. Q is undefined in 1, so we consider the Taylor expansions of the numerator and denominator to derive the limit. From the extensions of the functions to the complex plane, we find the following expansions at

$c = 1$,

$$\begin{aligned}
\cos^{-1}(c) &= \sqrt{2}i\sqrt{c-1} - \frac{\sqrt{2}}{12}i(c-1)^{3/2} + o((c-1)^{3/2}),\\
c\sqrt{1-c^2} &= \sqrt{2}i\sqrt{c-1} + \frac{5}{4}\sqrt{2}i(c-1)^{3/2} + o((c-1)^{3/2}),\\
(1-c)\sqrt{1-c^2} &= -\sqrt{2}i(c-1)^{3/2} + o((c-1)^{3/2}).
\end{aligned}$$

Substituting these expressions, the numerator of Q becomes

$$\begin{aligned}
\cos^{-1}(c) - c\sqrt{1-c^2} &= -\left(\frac{\sqrt{2}}{12} + \frac{5\sqrt{2}}{4}\right) i(c-1)^{3/2} + o((c-1)^{3/2})\\
&= -\sqrt{2}\left(\frac{1}{12} + \frac{15}{12}\right) i(c-1)^{3/2} + o((c-1)^{3/2})\\
&= -\frac{4}{3}\sqrt{2}i(c-1)^{3/2} + o((c-1)^{3/2}).
\end{aligned}$$

Taking the limit, we find

$$\begin{aligned}
\lim_{c\to 1} Q(c) &= \lim_{c\to 1} \frac{1}{2}\left[\frac{-(4/3)\sqrt{2}i(c-1)^{3/2}}{-\sqrt{2}i(c-1)^{3/2}}\right]\\
&= \frac{1}{2} \times \frac{4}{3}\\
&= \frac{2}{3}.
\end{aligned}$$

We can calculate the derivative of Q with respect to c, over $c \in (0, 1)$, to verify that the function is monotonic decreasing over the interval. This proves that Q is bounded above by $\pi/4$ and bounded below by $2/3$.

Appendix C

Bilinear interpolation gradient

In this appendix, we derive the analytical expression of the gradient for the bilinear interpolation of a surface. We used the bilinear interpolation to interpolate the piecewise-constant covariate functions, in the Langevin movement model.

Let f be the two-dimensional function of interest. We want to approximate f and its gradient at a point (x, y). The point is in a grid cell delimited by x_1 (lower) and x_2 (upper) along the x axis, and by y_1 (lower) and y_2 (upper) along the y axis. The function f is known at the four corner points of the grid cell, and we write

$$\begin{aligned} f(x_1, y_1) &= f_{11} \\ f(x_1, y_2) &= f_{12} \\ f(x_2, y_1) &= f_{21} \\ f(x_2, y_2) &= f_{22}. \end{aligned}$$

The bilinear interpolation of f at the point (x, y) is

$$\hat{f}(x, y) = \frac{y_2 - y}{y_2 - y_1}\left(\frac{x_2 - x}{x_2 - x_1} f_{11} + \frac{x - x_1}{x_2 - x_1} f_{21}\right) + \frac{y - y_1}{y_2 - y_1}\left(\frac{x_2 - x}{x_2 - x_1} f_{12} + \frac{x - x_1}{x_2 - x_1} f_{22}\right).$$

For convenience, we rewrite it as

$$\begin{aligned} \hat{f}(x, y) = \frac{1}{(y_2 - y_1)(x_2 - x_1)}\Big\{ &(y_2 - y)\big[(x_2 - x) f_{11} + (x - x_1) f_{21}\big] + \\ &(y - y_1)\big[(x_2 - x) f_{12} + (x - x_1) f_{22}\big]\Big\}. \end{aligned}$$

We can then derive the partial derivatives of the interpolated function,

$$\frac{\partial \hat{f}}{\partial x}(x, y) = \frac{(y_2 - y)(f_{21} - f_{11}) + (y - y_1)(f_{22} - f_{12})}{(y_2 - y_1)(x_2 - x_1)}$$
$$\frac{\partial \hat{f}}{\partial y}(x, y) = \frac{(x_2 - x)(f_{12} - f_{11}) + (x - x_1)(f_{22} - f_{21})}{(y_2 - y_1)(x_2 - x_1)}.$$

Appendix D

Alternative parametrisation of the CTCRW

Gurarie et al. (2017) suggest an alternative parametrisation of the CTCRW process, with parameters

$$\tau = \frac{1}{\beta}, \quad \nu = \frac{\sqrt{\pi}\sigma}{2\sqrt{\beta}}.$$

In this formulation, τ is the time interval over which the autocorrelation function of the velocity process decreases by a factor e. A larger value of τ indicates stronger movement persistence. The parameter ν is the mean speed, and a larger value of ν corresponds to faster movement. In this appendix, we show how the formulas above can be obtained.

Time scale of autocorrelation

We consider the one-dimensional velocity process $(V_t)_{t\geq 0}$, modelled with an Ornstein-Uhlenbeck process with mean parameter $\gamma = 0$. The results also apply to the isotropic two-dimensional case of the CTCRW process. The velocity satisfies

$$V_t|\{V_s = v_s\} \sim N\left[e^{-\beta(t-s)}v_s,\ \frac{\sigma^2}{2\beta}(1 - e^{-2\beta(t-s)})\right].$$

The Orstein-Uhlenbeck process is stationary, with long-term mean $\gamma = 0$ and long-term variance $w = \sigma^2/(2\beta)$. Its autocorrelation between times s and t is therefore

$$R(s,t) = \frac{\text{Cov}(V_s, V_t)}{w}$$

We first find the autocovariance of the velocity process between times s and t,

$$\text{Cov}(V_s, V_t) = E(V_s V_t) - E(V_s)E(V_t).$$

By property of the Ornstein-Uhlenbeck process, $E(V_s) = E(V_t) = \gamma = 0$. So, we have

$$\begin{aligned}
\text{Cov}(V_s, V_t) &= E(V_s V_t) \\
&= \int E(V_t|V_s = v_s) v_s p(v_s) dv_s \\
&= \int v_s^2 e^{-\beta(t-s)} p(v_s) dv_s \\
&= e^{-\beta(t-s)} \int v_s^2 p(v_s) dv_s \\
&= e^{-\beta(t-s)} E(V_s^2).
\end{aligned}$$

We also have

$$\text{Var}(V_s) = E(V_s^2) - E(V_s)^2 = E(V_s^2)$$

and so

$$\text{Cov}(V_s, V_t) = e^{-\beta(t-s)}\text{Var}(V_s) = we^{-\beta(t-s)}.$$

The autocorrelation function of the process between times s and t becomes

$$R(s,t) = \frac{we^{-\beta(t-s)}}{w} = e^{-\beta(t-s)}.$$

We can then verify that the autocorrelation between times t and $t+\tau$ (with $\tau = 1/\beta$) is

$$R(t, t+\tau) = e^{-\beta\tau} = e^{-1}.$$

That is, the autocorrelation function of the velocity process decreases by a factor e over a time interval of length τ.

Mean speed

The long-term distribution of the velocity Ornstein-Uhlenbeck process with mean $\boldsymbol{\gamma} = \mathbf{0}$ is

$$\boldsymbol{V} \sim N\left(\mathbf{0},\ \frac{\sigma^2}{2\beta}\boldsymbol{I}_2\right).$$

The mean speed of the animal is obtained as the mean norm of the two-dimensional vector $\boldsymbol{v}$. Denote $\boldsymbol{V} = (V_x, V_y)'$, with

$$V_x \sim N\left(0,\ \frac{\sigma^2}{2\beta}\right), \quad V_y \sim N\left(0,\ \frac{\sigma^2}{2\beta}\right),$$

the one-dimensional components of the velocity. We can define two standard normal variables $\tilde{V}_x$ an $\tilde{V}_y$ by

$$\tilde{V}_x = \frac{V_x}{\sigma/\sqrt{2\beta}}, \quad \tilde{V}_y = \frac{V_y}{\sigma/\sqrt{2\beta}}.$$

The mean norm of the velocity is

$$\begin{aligned}
\nu &= E\left[\sqrt{V_x^2 + V_y^2}\right] \\
&= E\left[\sqrt{\frac{\sigma^2}{2\beta}\tilde{V}_x^2 + \frac{\sigma^2}{2\beta}\tilde{V}_y^2}\right] \\
&= \frac{\sigma}{\sqrt{2\beta}} E\left[\sqrt{\tilde{V}_x^2 + \tilde{V}_y^2}\right].
\end{aligned}$$

We know that $\sqrt{\tilde{V}_x^2 + \tilde{V}_y^2}$ follows a χ distribution with two degrees of freedom. Its mean is

$$E\left[\sqrt{\tilde{V}_x^2 + \tilde{V}_y^2}\right] = \sqrt{2}\frac{\Gamma(1.5)}{\Gamma(1)} = \sqrt{2}\frac{\sqrt{\pi}}{2} = \sqrt{\frac{\pi}{2}}.$$

Eventually, we find the mean speed,

$$\nu = \frac{\sigma}{\sqrt{2\beta}} \times \sqrt{\frac{\pi}{2}} = \frac{\sqrt{\pi}\sigma}{2\sqrt{\beta}}.$$

Appendix E

Additional figures for the grey seal case study

In Chapter 6, we described the link between the transition rates of the continuous-time behaviour process and measures of behaviour persistence. In particular, the transition rates can be used to derive the mean dwell time in each state, i.e. the mean duration spent before switching to another state. They can also be linked to the stationary distribution of the behavioural process, which describes the proportion of time spent in each state, in the long run. Histograms of the posterior samples of the mean dwell times are shown in Figure E.1, and histograms of the stationary state probabilities are shown in Figure E.2.

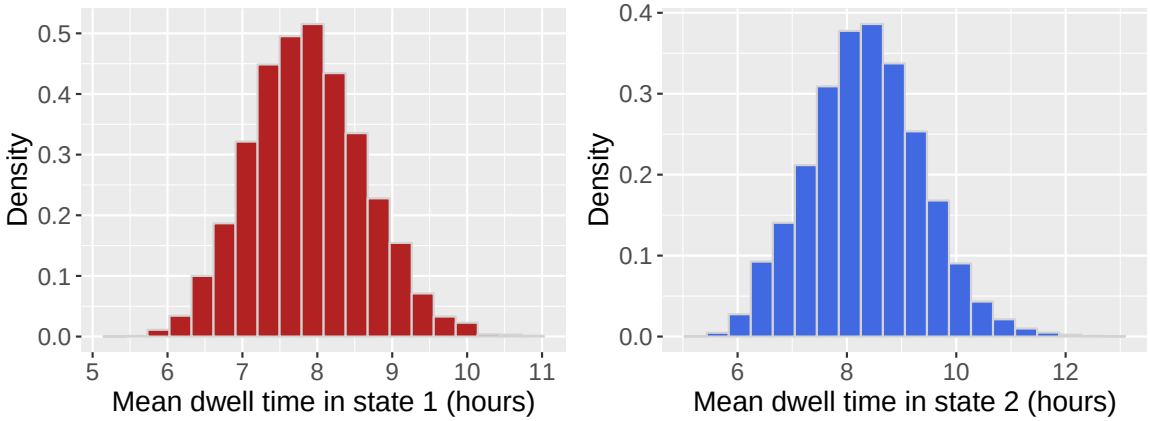


Fig. E.1 Histogram of posterior samples of dwell times in the grey seal case study.

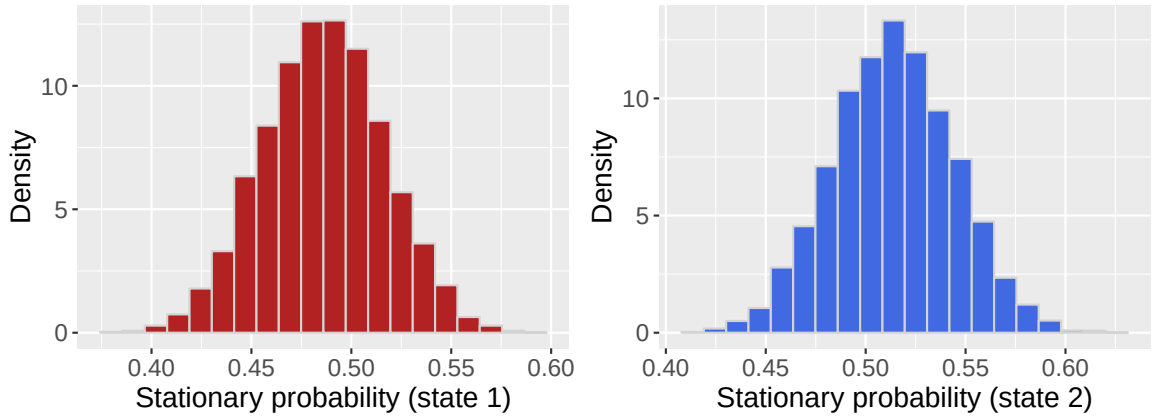


Fig. E.2 Histogram of posterior samples of stationary state probabilities in the grey seal case study.