# Models on the Move: Memory and Temporal Discretization in Animal Movement

by

Ulrike E. Schlägel

A thesis submitted in partial fulfillment of the requirements for the degree of

**Doctor of Philosophy**

in

**Applied Mathematics**

Department of Mathematical and Statistical Sciences

**University of Alberta**

# Abstract

Movement ecology thrives from a successful synergy of data and models. In a field where experiments are difficult or impossible, linking field data with mathematical and statistical models allows us to test hypotheses and increase our quantitative understanding of movement processes. Owing to technological progress, data availability and quality are growing rapidly, inspiring new questions and challenging methodology. In my thesis, I address two modelling challenges, one at the forefront of current research on memory-based movement and the other long-standing, yet prevailing, in movement data analysis.

Movement serves needs, such as foraging, but also requires time and energy. Therefore, we expect animals to have evolved strategies for efficient movement, likely drawing on cognitive abilities. Indeed, one of the current challenges in movement ecology is to understand the role of cognition, including memory, for movement. To date, very few models that include memory mechanisms have been confronted with data. In my thesis, I present a new cognitive-based model, in which an individual's travel history feeds back to future movement decisions. I focused on the pure spatio-temporal aspect of the travel history, assuming that an individual keeps track of elapsed times since last visits to locations and uses this information during the movement process. I showed that, despite the dynamic interplay of information gain and use, statistical inference can successfully identify this mechanism. I further applied the new modelling framework to wolf (*Canis lupus*) movement data to test whether wolves adopt a prey management strategy, based on memory, that is directed at reducing impacts of behavioural depres-

sion of prey through optimal timing of returns to hunting sites. I found support for the hypothesis but also point out the need to analyze a larger number of individuals to reach stronger conclusions.

Data collection methods, as well as standard modelling approaches, discretize the temporal dimension of movement processes. This discretization is a challenge for data analysis, because results may be affected by data sampling rate. In my thesis, I develop the formal concept of movement models' robustness against varying temporal resolution. I provide a series of definitions for movement model robustness. These definitions vary in their strength of conditions but all rest on the same requirement that a model can validly be applied to data with varying resolutions, while parameters change in a systematic way that can be predicted. In an analysis of random walks and spatially-explicit extensions thereof, I found that while true robustness is rare, approximate robustness is more widely present in models. I further demonstrate how robustness can be used to mitigate the influence of temporal resolution on statistical inference.

# Preface

This thesis is an original work by Ulrike E. Schlägel.

Chapter 2 of this thesis has been published as: Schlägel, U.E. & Lewis, M.A. (2014). Detecting effects of spatial memory and dynamic information on animal movement decisions. Methods in Ecology and Evolution, 5(11), 1236–1246. I developed and performed the analysis and wrote the manuscript. Dr. M.A. Lewis was supervisory author and provided feedback on concept formation and the manuscript.

The data used for Chapter 3 was provided by Dr. Evelyn Merrill at the University of Alberta (University of Alberta Animal Care Protocol Nos. 391305, 353112, and 411601).

To my grandparents' memory

and Harald Lammeck

# Acknowledgements

First and foremost, I would like to thank Dr. Mark Lewis for supervising me during my PhD. By accepting me as a student, he made a dream come true. Working at the intersection of mathematics, statistics and ecology – Mark made it not only possible but also an invaluable experience. There is so much to learn from Mark that I should be happy to have internalized a fraction of it. Mark was also very supportive in difficult times, and I am extremely grateful to him for all support, understanding and autonomy that he granted me.

I am also very grateful to Dr. Evelyn Merrill and Dr. Nathan Webb for adding a substantial amount of excitement to my PhD by sharing their wolf data with me. To work on real data would have been exciting, but to work on real wolf data was a true highlight. I also thank my supervisory committee, which next to Mark and Evie included Drs. Subhash Lele and Michael Kouritzin. I am grateful that they took the time to be on my committee and to provide valuable feedback during the course of my program.

I thank iCORE, now part of Alberta Innovates - Technology Futures, the PIMS International Graduate Training Centre in Mathematical Biology, and the University of Alberta for funding during my PhD. I further thank the IGTC for enriching experiences during the yearly summits.

I am grateful to the entire Lewis Research Group for all the feedback I received during lab talks, for many interesting and lively discussions, and for their social support. I would like to express special thanks to Cecilia Hutchinson, who always had a friendly word, smile, or comforting hug, ready to give. I am also very grateful to Kimberly Wilke-Budinsky for all her cheerfulness and her eager help with any issue of administration.

I would like to thank a few fellow students and postdocs individually for making a special contribution to my learning experience and my growth. I am incredibly grateful to Marie Auger-Méthé. I learned a lot from her. In particular, she was the first to introduce me to the magic of data analysis during a summer course at the

Bamfield Marine Science Centre ("stationary bears on moving ice"). Later, she and Craig DeMars invited me to work on a project with them, analyzing caribou movement data, which was a fun and very illuminating experience for me. For this, I would also like to thank Craig. Furthermore, I learned a lot during meetings of the Ecological Modelling group. I would like to specifically thank Devin Goodsman for initiating and leading the group. I am grateful to Jonathan Potts and Andrew Bateman for supporting me during my first paper submission. They helped me by both pushing and encouraging me and by providing useful feedback. Finally, I would like to thank Greg Breed for lending an ear to my worries and doubts, for encouraging me, and for relaxing hours alongside a glass of Whisky.

I am indebted to my mother and my sister Sonja for their support during all the last years. Only my Mum's unconditional love and patience allowed her to be happy about my phone calls no matter the time, joining me in cheerful moments and also helping me to overcome moments of struggle. I thank my sister for staying true to her profession by being there for me in any case of emergency, and also for her love, help and encouragement during my first year.

I dedicate my thesis to my grandparents' memory, whose support was crucial in my decision to pursue a University career. I am immeasurably grateful to them for all their help. I also dedicate my thesis to Harald Lammeck, my secondary school math teacher. As part of a game, he conjectured that I would become a mathematician – and his guess, whether insightful or just luck, became a self-fulfilling prophecy. I am thankful to him for installing the idea in my mind of becoming a mathematician.

Finally, I would like to express my thanks to Dr. Thomas Mueller at BiK-F, Frankfurt, for providing motivation during the last stretch of my PhD by offering me exciting future opportunities.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Advances in movement ecology spur new methodological challenges

Animals show fascinating movement capacities. Humpback whales travel more than 8,000 km between wintering and feeding areas (Rasmussen *et al.*, 2007), and arctic terns even reach one-way migration distances of more than 25,000 km (Egevang *et al.*, 2010). Many fish aggregate in schools and coordinate their movements in large groups of thousands, even millions, of individuals spaced less than a body length apart (Misund, 1993). The well-known cheetah, the fastest mammal, can sprint at 100 km/h (Sharp, 1997), while diving falcons such as the peregrine can even reach speeds up to 360 km/h (Tucker, 1998).

It is not only these spectacular and eye-catching phenomena that attract our attention to movement but also the importance of movement for many ecological processes. Many animals must move to meet their daily and lifetime needs related to maintenance (e.g. foraging), survival (e.g. escaping predation) and reproduction (e.g. finding mates, travelling to breeding sites). Movement of individuals scales up to population patterns, affecting populations' abundances and distributions as well as community structures (Mueller *et al.*, 2008; Morales *et al.*, 2010; Buchmann *et al.*, 2011). Additionally, movement processes have far-reaching consequences for disease dynamics in host-pathogen systems, being able to both promote and impede transmissions (White *et al.*, 2000; Altizer *et al.*, 2011). By facilitating seed dispersal, motile frugivorous animals provide important ecosystem service to their mutualistic plant partners (Mueller *et al.*, 2014).

During the last decades, studies of animal movement have benefitted from major advances in tracking technology that opened new dimensions for data collection. GPS- and Argos-based tracking devices have become widespread and allow automatic data collection over large spatial and temporal scales (Rutz & Hays, 2009; Cagnacci *et al.*,

2010). Owing to increasing miniaturization of these tagging devices, they can be fitted to a wide range of species, which continues to grow (Bridge *et al.*, 2011; Recio *et al.*, 2011). Not only the extent to which data can be collected has increased but also the precision and frequency (Frair *et al.*, 2007; Tomkiewicz *et al.*, 2010; Bridge *et al.*, 2011). For example, by now GPS devices often reach an average precision of 10-28 m (Frair *et al.*, 2007). These technological advances have led to an unprecedented availability of movement data across the globe, as for example demonstrated by the Movebank project (Kranstauber *et al.*, 2011).

At the same time, mathematics, statistics and computing have contributed new methods and tools to study movement theoretically and to analyze the newly available movement data. Within both the individual-based Lagrangian and the population-based Eulerian framework, descriptions of movement have progressed from simple random walks and diffusion models towards models that explicitly acknowledge spatial heterogeneity of environments and selective behaviour of animals. Contemporary models include, for example, resource selection mechanisms (Rhodes *et al.*, 2005; Hanks *et al.*, 2011; Mckenzie *et al.*, 2012; Potts *et al.*, 2014), interactions of predators with their prey and conspecifics (Lewis & Murray, 1993; Moorcroft *et al.*, 2006), and switches in behavioural modes (Morales *et al.*, 2004; McClintock *et al.*, 2012; Langrock *et al.*, 2013). Studies of movement have even started to consider the role of cognitive processes within contexts such as optimal foraging (Barraquand & Benhamou, 2008; Boyer & Walsh, 2010; Grove, 2013) and home range formation (Börger *et al.*, 2008; Van Moorter *et al.*, 2009).

It is in the nature of scientific progress that advances lead not only to increased knowledge and understanding but also to new questions. Thus, advances in technology, methodology and theory have resulted in a movement ecology paradigm and continue to inspire new goals and ambitions (Nathan *et al.*, 2008; Holyoak *et al.*, 2008; Börger *et al.*, 2008; Tomkiewicz *et al.*, 2010; Fagan *et al.*, 2013). In my PhD thesis, I address methodological challenges in movement ecology, contributing to two different aspects of modelling movement.

## 1.1 Memory matters: modelling informed animal movement

Animals demonstrate a variety of cognitive skills. For example, food-caching birds find previously buried seeds up to nine months later (Balda & Kamil, 1992); bees can-

not only navigate efficiently but also communicate spatial information through their famous waggle dances (Menzel *et al.*, 2006); birds parasitized by cuckoos can adjust their probability of rejecting eggs depending on whether they have seen a cuckoo on their nest (Davies *et al.*, 1996). In general, animal cognition refers to all mechanisms by which animals acquire, process and use information (Shettleworth, 2010). This includes mechanisms such as conditioning and associative learning, in which animals learn to respond to cues and stimuli, as well as mechanisms by which animals act on information retrieved from memory. We have learned much about animal cognition through carefully designed experiments. Manipulative experiments allow researchers to construct test situations that are tailored to induce specific, and often predicted, behaviours in animals, while minimizing and controlling for potential confounding factors (for a variety of examples, see e.g. Shettleworth, 2010). At the same time, experiments are often limited to small scales and may require habituation of animals (but see Thorup *et al.*, 2007; Tsoar *et al.*, 2011).

One of the present challenges in movement ecology is to understand the role of cognition, including memory, for movement processes (Börger *et al.*, 2008; Smouse *et al.*, 2010; Fagan *et al.*, 2013). Movement usually serves a goal ("why move?"), triggered by an individual's internal state. Motion capacities ("how to move?") work together with navigation capacities ("where to move?"), under the influence of external factors, to achieve the goal (Nathan *et al.*, 2008). Within this paradigm, cognition serves the navigation process, which includes orientation in space and time and selection of target locations. Orientation and navigation skills allow animals to reach target locations and thus assist processes such as homing and migration, but also foraging. Foraging animals benefit from information about the location of resources, temporal availability of resources and resource quality. At any time, information may be perceived or obtained from memory. If animals remember profitable places, this involves both spatial memory of the location and attribute memory about the locations' features (Fagan *et al.*, 2013).

Much work has concerned the mechanisms by which animals orient themselves and navigate in space. A relatively simple mechanism is path integration (also termed dead reckoning), in which an individual internally keeps track of all distances and directional changes of its path and at any point can orient back straight towards its starting location (Wittlinger *et al.*, 2006; Collett *et al.*, 2006). Path integration is an egocentric mechanism, that is spatial information is encoded with respect to the individual itself. In contrast, exocentric mechanisms establish spatial orientation based on external references, such as beacons and landmarks (Shettleworth, 2010). It is

generally thought that landmarks can be integrated into cognitive maps for navigation, but it is still debated how landmarks are used, e.g. whether interlandmark relations or landscape contours are more important, and how flexible maps needs to be, e.g. whether they allow a euclidian view that preserves distances and angles (O'Keefe & Nadel, 1978; Bennett, 1996; Lew, 2011). To date, navigation strategies have mainly been tested in insects such as ants and bees, small rodents and in particular rats, and to some degree birds, because of these species' suitability for experiments (Griffin, 1952; Collett *et al.*, 2006; Shettleworth, 2010).

The relevance of cognition for foraging and search for resources has started to be investigated theoretically, often using individual-based simulation models. These studies evaluate the benefits that information-based and especially memory-based movement strategies can provide, and under which conditions. Results suggest that memory-based foraging strategies are particularly successful in environments with patchy, or clumped, resources (Barraquand *et al.*, 2009; Fronhofer *et al.*, 2013) and in temporally predictable landscapes (Mueller *et al.*, 2010). When temporal unavailability of resources reduces predictability, a mixed strategy of both random and informed steps can be optimal (Boyer & Walsh, 2010). Barraquand *et al.* (2009) point towards the importance of considering costs, e.g. higher energy requirements, of cognitive abilities that reduce their benefits, which still needs to be further investigated (Fagan *et al.*, 2013). Memory has also sparked interest with respect to the mechanisms underlying home range formation. Simulations of memory-based movement have been able to lead to restricted space-use patterns when memorized information increased the expected value of familiar resources patches compared to unfamiliar patches despite temporary resource depletion (Van Moorter *et al.*, 2009; Spencer, 2012).

To test predictions from theoretical findings, we need to confront them with data. However, studying the link between cognition and movement in free-ranging animals in their natural habitats is challenging. First, due to less controllability, alternative explanations than the hypothesized ones can be difficult to rule out. Nonetheless, there have been attempts to infer spatial cognition and memory in foraging animals, primarily primates (Asensio *et al.*, 2011; Janmaat *et al.*, 2006; Janmaat & Chancellor, 2010). These studies benefited from the possibility to follow the animals and record detailed behavioural and environmental data. Second, when relying on models, movement is a complex process because of its both spatial and temporal dimension. This process becomes even more intricate when detailed behavioural mechanism such as memory-based resource selection are included. Applying memory-based models to data demands suitable, including identifiable, models, appropriate statistical methods,

and also computational tools for model fitting. Therefore, many approaches at the memory-movement interface have remained theoretical, and very few inferential studies exist (Dalziel *et al.*, 2008; Avgar *et al.*, 2015). In my thesis, I propose a new model that incorporates a memory-based movement strategy, yet is tractable enough to allow likelihood-based statistical inference.

## 1.2 Fitting models to data: mitigating impacts of temporal discretization

To test hypotheses about movement behaviour and to estimate parameters of movement processes, we need both data and quantitative methods. Pure data, e.g. in form of field observations, and descriptive analyses give important impulses to research, generating questions and hypothesis. On the other hand, mathematical models provide abstraction and simplification for hypotheses and allow us to test "what if" situations. Theoretical model analyses can help us to identify key mechanisms of a process, reveal threshold phenomena and make quantitative predictions (e.g. Lewis & Kareiva, 1993; Neubert *et al.*, 1995; Lewis *et al.*, 1997). Computer-based simulations of models (also termed individual-based or agent-based models) are a useful tool to hypothetically explore more complex processes that are hard to tract analytically (e.g. Barraquand *et al.*, 2009; Berbert & Fagan, 2012). However, only a synergy of movement data and models, mediated by rigorous statistical methods, enables us to explain and quantify animal movement processes and their implications for ecology (Hilborn & Mangel, 1997; Turchin, 1998). Additionally, simulation models that are qualitatively and quantitatively informed by previous data analyses, are a valuable tool for wildlife management and conservation (e.g. Colchero *et al.*, 2010; DeCesare *et al.*, 2012; Webb & Merrill, 2012).

Both movement data and models are approximations of reality. Animal movement often spans large spatial and temporal scales and may occur in inaccessible habitat. Thus, our ability to observe movement is limited, and we usually rely on techniques that deliver series of snapshots of individuals' locations and behaviours (Turchin, 1998). By contrast, it is the very nature of models to approximate and simplify. Approximations can ease analysis and application, as expressed in the famous statement "all models are wrong, but some are useful" (Box & Draper, 1987). Additionally, through simplifying reductions we can focus on a few important mechanisms of a process without being distracted by too much detail. This is, in fact, one of the merits of modelling, allowing

us to identify and understand key mechanisms that generate observed patterns and phenomena (Levin, 1992). Whether approximations are necessary or desirable, we should, if possible, conduct them with care and be mindful about their consequences.

Movement processes occur in space and time, and approximations can pertain to both of these dimensions. Much progress has been made with respect to improving the accuracy of spatial approximations of movement data. Often, data is collected by automated tracking devices, which are attached to an individual (e.g. as collars or backpacks). Common technologies include GPS, mainly for terrestrial species, and Argos, mainly for marine species (Rutz & Hays, 2009; Frair *et al.*, 2010). GPS data has reached high spatial accuracy, many devices reaching an average precision of 10-28 m (Frair *et al.*, 2007). Argos devices are less precise, with measurement errors ranging from a few hundred meters to several kilometres (Rutz & Hays, 2009; Patterson *et al.*, 2010). However, Argos technology offers other advantages (e.g. tags require less battery power and do not need to be retrieved for data recovery), and a body of research has successfully attended to state-space models and associated methods as a means to correct for measurement errors (Jonsen *et al.*, 2005; Patterson *et al.*, 2008; Patterson & Hartmann, 2011).

Temporal resolution of movement data still poses challenges. In view of data collection, one of the major limiting factors of data sampling rate is battery life of tracking devices. Signal reception, transmission and on-board processing require battery power, constraining the number of possible measurements (Breed *et al.*, 2011; Patterson & Hartmann, 2011). Ideally, we would like to measure movement paths both over long time spans and at high frequency, however, usually we must make compromises. For example, to collect data over a year or longer, GPS collars are often programmed to attempt location fixes every 2-4 hours or similar (Webb & Merrill, 2012; DeMars *et al.*, 2013; Avgar *et al.*, 2015). An additional factor to consider is signal to noise ratio. Spatial measurement error should be small compared to distances between successive location fixes, which may require larger time intervals, especially when movement is slow (Ryan *et al.*, 2004; Jerde & Visscher, 2005; Bradshaw *et al.*, 2007). Therefore, most current data sets represent temporally discretized movement paths, where the resolution is at least partly dictated by technological constraints.

From a modelling perspective, temporal discretization of movement processes arises mainly in the Lagrangian view of movement. When analyzing movement data with a place-based Eulerian approach, for example to estimate population space-use patterns, models are often of diffusion type (Turchin, 1998; Smouse *et al.*, 2010). In this case, high temporal resolution of data is less important, on the contrary, some models rely on

location measurements to be spaced enough in time to be void of directional correlations (Moorcroft & Lewis, 2006). In the Lagrangian view, the focus lies on understanding the movement behaviour of individuals in relation to internal and external cues (Turchin, 1998; Nathan *et al.*, 2008; Smouse *et al.*, 2010). Here, the movement process is often modelled as some type of random walk or extension thereof. A few studies have linked discrete-time movement data with continuous-time random walks, in particular the Ornstein-Uhlenbeck process (Johnson *et al.*, 2008; Hanks *et al.*, 2011). However, most approaches use discrete-time random walks (for a review, see McClintock *et al.*, 2014). Discrete-time descriptions are particularly intuitive at scales at which movement occurs between natural stopping sites or resource patches, for example, when insects fly between ovipositing sites (Kareiva & Shigesada, 1983). But even in more complex situations, discrete-time random walks remain useful to disentangle scales (at which movement processes are stationary) and searching modes (intensive versus extensive searching behaviour) and to understand the movement behaviour of individuals with respect to their environment (Morales *et al.*, 2004; Langrock *et al.*, 2013; Benhamou, 2013; McClintock *et al.*, 2014; Avgar *et al.*, 2015).

We are therefore often faced with a situation in which both data and model approximate the temporal dimension of a movement process. Several problems arise from this. First, because the discretization of the data is constrained by collection methods, the available snapshots of an animal's path may not necessarily correspond to the biologically most relevant events, e.g. behavioural change points. The mismatch between behavioural "moves" and modelled "steps" has already been pointed out by Turchin (1998). However, this problem may be better understood in the context of multiple-mode multiple-scale movement, and a random-walk model can still be useful to describe movement between behavioural change points (Benhamou, 2013). Second, the temporal resolutions of data and model are usually treated in unison. An exception are models that need to handle temporally irregular data, for example as they arise when tracking marine mammals. For this situation, state-space models have become a tool to connect irregular data with a discrete-time movement model (Breed *et al.*, 2012; McClintock *et al.*, 2012). In contrast, time series of regular observations are usually matched directly with the model. However, before doing so, ideally one should evaluate whether the temporal resolution of the data is a suitable approximation of the movement process of interest.

Third, given the implied link between the data and model's temporal resolution, a major problem is that the data's resolution can influence statistical inference. Movement ecologists have recognized this issue and devoted a number of studies to demon-

strating especially the negative effects of low resolution, which lead to inaccurate biological results as well as difficulties in generalizing and comparing results (Turchin, 1998; Breed *et al.*, 2011; Rowcliffe *et al.*, 2012; Postlethwaite & Dennis, 2013; Yackulic *et al.*, 2011). A few attempts have been made to design methods that can compensate for varying temporal resolution of data, especially with respect to measuring travel distance and path tortuosity (Pépin *et al.*, 2004; Benhamou, 2004; Codling & Hill, 2005; Avgar *et al.*, 2013). However, approaches to the problem have remained sporadic. Given this situation, we may expect that many current movement analyses are in danger of being influenced by assumptions about an implicitly adopted temporal discretization of data and model, without knowing how large the influence is. In my thesis, I draw on the rigour of mathematics and statistics to address the problem with a new perspective as well as generality.

## 1.3 Dissertation outline

The following chapters of my thesis address the previously introduced methodological challenges in equal parts. Chapters 2 and 3 answer the call for new inferential models at the memory-movement interface. Chapters 4 and 5 establish a new theoretical framework for understanding and mitigating the impact of data and models' temporal discretization on statistical inference.

In Chapter 2, I propose a new model that includes a dynamic interplay of movement decisions and information gain. This model builds on previous formulations of movement as result of general movement capacities and available resources. In my model, I account for movement not only being influenced by information, e.g. resource information, but also affecting information itself. This takes up observations that animals use their travels not only to forage, for which they may use information, but also to acquire information, e.g. about temporary resource statuses. Also, animals may change information as they move, e.g. by depleting resources. As an example of such dynamic interplay, I model how an individual's movement decisions interact with information about its own travel history. I focus on the pure spatial and temporal aspect of the travel history in form of *time since last visit* to locations. The model yields a likelihood function and can thus be fitted to empirical movement data. Because the variable time since last visit is a dynamic covariate, not only influencing movement decisions but being affected by them, I test the functionality of the framework with simulated data. With this, I show that a classic model selection approach can identify the cognitive-based movement strategy and that parameter estimation can recover the

quantities that are shaping the process.

In Chapter 3, I apply the new model developed in Chapter 2 to wolf (*Canis lupus*) movement data. I use the model's dynamic interplay between movement decisions and travel history to formulate a movement strategy, in which wolves move so as to minimize their impact on prey. This impact includes in particular behavioural depression of prey caused by the wolves. Wolves, however, can reduce prey's behavioural changes by leaving and returning to hunting sites in a timely manner. By confronting the model with data, I show that observed movement patterns of wolves support this hypothesized prey-management strategy, although quantitative results do not fully align with expectations. In addition to the ecological relevance of the analysis, it also serves to demonstrate how to infer a sophisticated movement strategy that likely involves spatial and temporal memory from empirical data.

With Chapter 4, I turn towards the problem that statistical inference is often affected by the temporal discretization of movement data and models. I approach the problem by dissociating a model's resolution from the data's resolution, thereby allowing the view that data may in fact only represent a subprocess of the behavioural process of interest (which is formulated in the model). In this view, the problem becomes related to the formal concept of robustness in statistics. In Chapters 4 and 5, I therefore develop a series of definitions for movement models' robustness against changes in temporal resolution, and I examine if, and which, existing models have this property. In Chapter 4, I start by defining robustness and a weaker version, semi-robustness, for classic random walks with independently and identically distributed steps. In this case, robustness implies that a model can be validly applied to both finer and coarser data, while keeping information about parameters intact via an appropriate parameter transformation. Semi-robustness only allows the model to be scaled up to coarser resolutions, which, however is a useful property already. I investigate which step distributions lead to (semi-) robust random walks, and how we can make models robust by extending step distributions within larger families. I also show how robustness and semi-robustness relate to the probabilistic concept of infinite divisibility.

In Chapter 5, I extend the idea of robustness to more general first-order Markovian models. Due to the more complex situation, robustness is defined slightly different than in Chapter 4. In fact, the robustness definition in Chapter 5 is closer related to the semi-robustness definition in Chapter 4 in the sense that it addresses the issue of linking a model with suboptimal, i.e. coarser, data. Because exact robustness is a very strong condition, I also propose two alternative definitions, asymptotic and approximate robustness. Especially the new definition of approximate robustness is designed

to become a practical tool. Again, I investigate which models are robust. I focus on spatially-explicit resource selection models, which constitute the class of models I use in Chapters 2 and 3, without the dynamic information component. In addition to presenting theoretical results based on analytical calculations, I also show how the new definitions can be used more practically via numerical calculations. Finally, I contrast robustness properties of two spatially-explicit resource selection models that incorporate an exponential and logistic resource selection function, respectively.

In Chapter 6, I conclude my thesis with a discussion of the newly developed models and methods in light of their application for movement ecology. I highlight specific features of the new movement model and how these can be used to test hypotheses about cognitive-based movement strategies. I also discuss the relevance of spatial and temporal memory for the modelling approach. Furthermore, I discuss my theoretical work on movement models' robustness in the context of the well-established concept of robustness in statistics. I highlight the key findings of my analysis, discuss how the new concept can be applied to analyses of movement data, and suggest directions for further research.

# Chapter 2

# Detecting effects of spatial memory and dynamic information on animal movement decisions[1]

## 2.1 Introduction

Animal movement serves important needs such as food acquisition, escape from predators, and travel to reproduction sites. Consequently, many species have evolved capacities to move efficiently and purposefully by considering varying sources of information for their movement decisions (Janson & Byrne, 2007; Sulikowski & Burke, 2011). Explaining the mechanisms that underly such informed movement behaviour will allow us to better understand animal space-use patterns and their responses to environmental changes (Dalziel *et al.*, 2008; Nathan *et al.*, 2008; Sutherland *et al.*, 2013).

Most animals live in heterogenous environments, and the link between movement and environment has received much attention. Using classical resource-selection analyses (Manly *et al.*, 2002), a wide range of studies have demonstrated that animals selectively use the biotic and abiotic features that are available to them (Fortin *et al.*, 2005; Gillies *et al.*, 2011; Squires *et al.*, 2013). Analyses of movement characteristics have shown that animals express different movement behaviours, e.g. encampment or travel, in different habitats (Morales *et al.*, 2004; Forester *et al.*, 2007).

Most mechanistic models have concentrated on incorporating relationships between environmental factors and movement behaviour within a static environment (but see

[1] A version of this chapter has been published as: Schlägel, U.E. & Lewis, M.A. (2014). Detecting effects of spatial memory and dynamic information on animal movement decisions. Methods in Ecology and Evolution, 5(11), 1236–1246.

Avgar *et al.*, 2013); however, observations show that animals also take into account dynamically changing information and respond with their movements to temporal availability or unavailability of resources (Martin-Ordas *et al.*, 2009). For instance, fruit-eating primates express goal-oriented travel towards those trees in their home range that carry ripe fruit (Asensio *et al.*, 2011), and it has been suggested that monkeys use their daily travels to monitor fruiting histories of trees (Janmaat *et al.*, 2013; Janson & Byrne, 2007). On the other hand, many resources, once depleted, need some time before they become available again, providing reason for animals to avoid depleted food patches (Davies & Houston, 1981; Owen-Smith *et al.*, 2010; Bar-Shai *et al.*, 2011). Avoidance behaviour may be a response not only to depletion of resources, such as plant biomass or prey, but also to behavioural depression. Behavioural depression refers to a reduction in prey availability that is caused by behavioural changes of the prey in response to predation (Charnov *et al.*, 1976). For example, prey may show greater alertness or seek shelter. This reduces capture rates, to which predators may respond in turn by changing their hunting areas (Jedrzejewski *et al.*, 2001; Amano & Katayama, 2009). Temporal considerations also become important for movement decisions if territorial defence mechanisms require animals to visit certain locations regularly, e.g. to scent-mark territory boundaries (Moorcroft & Barnett, 2008; Giuggioli *et al.*, 2011).

As the above examples highlight, spatio-temporal information drives movement decisions and at the same time movement allows animals to update this information. Experimental findings additionally support that animals make decisions based on information that they have obtained through previous experiences. Memory of information about the "what, where and when" of events, obtained through subjective experience, is termed "www-memory" (Martin-Ordas *et al.*, 2009) or "episodic-like memory" (Griffiths *et al.*, 1999). It is possible that animals acquire information about current environmental conditions through perceptual cues, even over large distances (Tsoar *et al.*, 2011), and that information about the recent travel history is stored in externalized "memory", such as pheromone trails or slime (Deneubourg *et al.*, 1989; Reid *et al.*, 2012). However, it is likely that many animals draw upon internal memory, especially for behaviours that require information about temporal distances ("how long ago?") (Griffiths *et al.*, 1999; Martin-Ordas *et al.*, 2009; Janmaat *et al.*, 2013). During recent years, movement models have started to incorporate influences of memorized information on movement decisions (for a review see Fagan *et al.*, 2013). Most of these are simulation models that are used for theoretical considerations only (but see Avgar *et al.*, 2013); however to test our understanding of the feedbacks between movement and information acquisition, we must also interface memory-based models with data

(Smouse *et al.*, 2010).

Here, I present a new model for animal movement that is amenable to likelihood-based inference, and in which I mechanistically incorporate the interplay of movement decisions, environmental information and dynamically changing temporal information. The model is similar in its form to recent spatially explicit resource-selection models (e.g. Rhodes *et al.*, 2005; Forester *et al.*, 2009), in which movement steps are assigned probabilities based on general movement tendencies and resource preferences. In previous models, resource information enters as a static covariate, providing knowledge about features of the landscape, such as land cover type or topographical features. In my model, I add dynamic information obtained through experiences made during movement. To realize the interplay of movement and information acquisition, I draw on the concept of a cognitive map (Tolman, 1948; Asensio *et al.*, 2011). I use this concept here as a helpful mathematical construct that provides a map-like representation of the animal's environment containing all relevant information. For an example of a dynamic information-gain process I introduce information about the time since last visit to locations. Time since last visit is useful information that can play a role, for example, in the process of patrolling in canids or food acquisition across species if food availability varies (Davies & Houston, 1981). With the inclusion of this information acquisition process, I present a practical model that incorporates both dynamic information and spatial memory.

I place the model into a model selection framework that allows to identify which types of information most likely shape the movement decision process. I first outline the general formulation of my model and how memory effects can be integrated. Subsequently, I present the details of several candidate models that correspond to different underlying mechanisms of animal movement behaviour. Next, I show how the models can be fitted to empirical movement trajectories to perform statistical inference. Finally, using simulated data, I test the functionality of my framework and assess whether the method can correctly detect effects of static resource information and dynamically changing temporal information and whether model parameters can be estimated reliably.

## 2.2 Methods

For several decades, the basis of many animal movement models have been random walks. In a classical random walk, movement is described as a series of discrete steps that have independent and identical probability distributions. This has been extended

to include correlations between steps, biases towards specific locations, and step probabilities that depend on the behavioural state of the individual (Morales *et al.*, 2004; McClintock *et al.*, 2012; Breed *et al.*, 2012; Langrock *et al.*, 2013). Random walks and their extensions have been used both to analyze movement behaviour at an individual level (Lagrangian approach; e.g. Smouse *et al.*, 2010) and to derive partial-differential equation models that describe spatio-temporal patterns at a population level or expected space-use of individuals (Eulerian approach; e.g. Codling *et al.*, 2008).

I am interested in understanding decision processes that underly movement behaviour on the scale of individuals. I draw upon a modelling framework that bridges the gap between statistical resource-selection analysis and spatially explicit movement models (Rhodes *et al.*, 2005; Moorcroft & Barnett, 2008; Forester *et al.*, 2009). The framework builds on a random walk and defines movement via step probabilities, which have two components. A resource-independent movement kernel assigns probabilities to steps based on the animal's general movement tendencies. Given this, a weighting function evaluates the attractiveness of steps according to resource availability and resource preferences. I extend this framework by generalizing the weighting function. In this generalization, the weighting function does not only describe the influence of resources but allows for the inclusion of any information relevant to the animal. Information can pertain to landscape features and resources, as in previous models, but also to memories of past events and timing aspects, which cannot be obtained externally but only through the movement process and the animal's behaviour itself. I assume that information at a given time is either obtained through direct perception or retrieved from the animal's cognitive map (i.e. memory) which itself is updated through experience. In my model, the cognitive map is a function that assigns values to locations according to their information content at a given time. Thus, it serves as a mathematical tool without the claim that it truly represents the underlying cognitive mechanism. With the framework of the cognitive map I provide a general method for including an explicit information-acquisition process. The cognitive map itself can take many forms, depending on the species and behaviour of interest. In my candidate models, I demonstrate examples of types of information the cognitive map may contain.

### 2.2.1 The modelling framework

I consider movement paths of individual animals, and I assume that an individual's trajectory consists of a series of locations $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ at regular times $T = \{1, \ldots, N\}$. Each location has an Easting and a Northing in two-dimensional space, which is dis-

cretized into a regular grid of square cells. The resolution of the spatial discretization depends on the available environmental data and should be fine enough compared to the animal's movement such that steps generally range over multiple cells. I model movement as a stochastic process, where the probability of making a step to location $\boldsymbol{x}_t$ depends on the location at time $t-1$ and, if movement is persistent, on the previous step from $\boldsymbol{x}_{t-2}$ to $\boldsymbol{x}_{t-1}$. I define this step probability as

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \boldsymbol{\theta}) = \frac{k(\boldsymbol{x}_t; \boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \boldsymbol{\theta}_1)\, w_t(\boldsymbol{x}_t; \boldsymbol{\theta}_2)}{\sum_{\boldsymbol{y}\in\Omega} k(\boldsymbol{y}; \boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \boldsymbol{\theta}_1)\, w_t(\boldsymbol{y}; \boldsymbol{\theta}_2)}, \tag{2.1}$$

where $k$ is an information-independent movement kernel, $w_t$ is an information-based weighting function, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is a collection of model parameters. The sum in the denominator ensures that $p$ is an appropriately normalized probability mass function over space. The spatial domain $\Omega$ is the area within which the animal can choose to travel during the time relevant to the study.

Using the conceptual framework of Nathan *et al.* (2008), the kernel $k$ can be interpreted as describing the animal's motion capacity and $w_t$ as formulating the influence of external factors, to which I add memorized information. Both $k$ and $w_t$ can be affected by the animal's internal goal. For instance, if a herbivore is foraging it is likely that it moves slowly, changes its movement direction frequently and generally stays in an environment with suitable foraging material. It may additionally prefer to forage in an area with low predation risk. Such behaviour could be implemented by a kernel that assigns higher probabilities to locations in the animal's close vicinity with the same values in all directions and a weighting function that has highest values in preferred foraging habitat. The weighting function could also include information about previously experienced presence of predators (Latombe *et al.*, 2014).

In general, the movement kernel $k$ can be very simple, e.g. constant within the animal's maximum movement radius (Rhodes *et al.*, 2005); however, we can also use a more complex kernel that accounts for persistence in movement direction or biases towards specific locations (Moorcroft & Lewis, 2006). Directions can be measured by either turning angles (the angles between successive steps) or bearings (the angles of steps with respect to a fixed direction, e.g. North).

I model the weighting function $w_t$ as a resource selection function (Manly *et al.*, 2002; Lele & Keim, 2006). There are several forms available for resource selection functions, and here I present the logistic form,

$$w_t(\boldsymbol{x}; \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \left[1 + \exp\left(-\alpha - \boldsymbol{\mathcal{I}}_t(\boldsymbol{x}) \cdot \boldsymbol{\beta} - f(\boldsymbol{\mathcal{I}}_t(\boldsymbol{x}), \boldsymbol{\gamma})\right)\right]^{-1}, \tag{2.2}$$

where $\cdot$ denotes the dot product of two vectors. The vector $\boldsymbol{\mathcal{I}}_t(\boldsymbol{x}) \in \mathbb{R}^n$ is the cognitive map content at location $\boldsymbol{x}$ at time $t$ containing the values of all information variables of location $\boldsymbol{x}$ at time $t$, and $\boldsymbol{\beta} \in \mathbb{R}^n$ is a parameter vector describing the animal's preference for a location of type $\boldsymbol{\mathcal{I}}_t(\boldsymbol{x})$. The intercept $\alpha \in \mathbb{R}$ determines the baseline weight of a location when all information variables are zero. The function $f$ and parameter vector $\boldsymbol{\gamma}$ account for possible interactions between different information variables. Locations with preferred features have high weights, thereby increasing the chance that an animal will visit those. The logistic form of the weighting function restricts weights to be between zero and one, and therefore the weighting function can in fact be viewed as a resource selection probability function (Lele & Keim, 2006).

Because of the dependence structure of the step probabilities in equation (3.1), they are only valid for times $t \geq 3$. Here, I chose to define an initial probability for the first two locations, $p(\boldsymbol{x}_1, \boldsymbol{x}_2|\boldsymbol{\theta}) = p(\boldsymbol{x}_2|\boldsymbol{x}_1, \boldsymbol{\theta})\, p(\boldsymbol{x}_1|\boldsymbol{\theta})$. A simple option is to assume that every location in the spatial domain has the same probability to be the first location, $p(\boldsymbol{x}_1|\boldsymbol{\theta}) = \frac{1}{|\Omega|}$, and to let

$$p(\boldsymbol{x}_1|\boldsymbol{x}_2, \boldsymbol{\theta}) = \frac{\tilde{k}(\boldsymbol{x}_2; \boldsymbol{x}_1, \kappa, \lambda)\, w_t(\boldsymbol{x}_2; \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})}{\sum_{\boldsymbol{y} \in \Omega} \tilde{k}(\boldsymbol{y}; \boldsymbol{x}_1, \kappa, \lambda)\, w_t(\boldsymbol{y}; \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})}, \tag{2.3}$$

where $\tilde{k}$ is possibly a simplified form of $k$ in case that $k$ describes persistent movement.

### 2.2.2 Candidate models

I consider four different models that represent biological hypotheses about the types of information that an individual may consider for making movement decisions. In the simplest case, the *null model*, I assume that the animal considers no specific information. In the *resource model*, an individual considers static information about the environment, where 'static' means that the information content remains constant over the time span of the analysis. Information can be given about any resources pertaining to the animal, e.g. any variables as they are typical in resource-selection analyses. To include dynamically changing information, I allow information, and thereby the weighting function, to change through time. If information were only given externally, this would constitute a dynamic version of the resource model. However, my aim is to model a dynamic interplay of movement decisions and information content. In the *memory model*, I therefore introduce *time since last visit* as new type of information. To account for the possibility that both resources and the dynamic variable time since last visit influence movement decisions simultaneously, I consider a *combination model*

as the most complex model.

I implement the different models by varying the information variable $\boldsymbol{\mathcal{I}}_t$ in the weighting function (3.2) while using the same movement kernel for all models.

### Null model

In the null model, I assume that the information content of all locations is zero. Therefore, the weighting function is homogeneous across the landscape and constant over time, $w_t(\boldsymbol{x}) = 1$ for all $x \in \Omega, t \in T$. This means that the animal moves only according to the kernel $k$.

### Resource model

In the resource model, information is static and includes all resource variables of interest, $\boldsymbol{\mathcal{I}}_t(\boldsymbol{x}) = \boldsymbol{\mathcal{I}}(\boldsymbol{x}) = (r_1(\boldsymbol{x}), \ldots, r_n(\boldsymbol{x}))$ for every location $x \in \Omega$. It is straightforward to extend this to dynamic resource information to include, e.g. seasonal changes in the landscape or disturbance events.

### Memory model

In the memory model, I assume that while the animal moves through the environment, it monitors the time since last visit from locations and uses this information for movement decisions. For instance, recently visited areas may be avoided for a period of time, whereas locations with long absence may be attractive. In my model, I include this feature by defining the cognitive map as $m_t : \Omega \to \mathbb{N}$, which at any time assigns values to all locations in the spatial domain based on the map values at the previous time and the most recent movement step. If the animal moves from location $\boldsymbol{x}_{t-1}$ to $\boldsymbol{x}_t$ between times $t-1$ and $t$, I define for any location $y$ in the spatial domain

$$m_t(\boldsymbol{y}) = \begin{cases} 0, & \text{if } d(\boldsymbol{y}, \boldsymbol{z}) \leq \delta \text{ for any } \boldsymbol{z} \in \text{path}(\boldsymbol{x}_{t-1} \to \boldsymbol{x}_t) \\ m_{t-1}(\boldsymbol{y}) + 1, & \text{otherwise.} \end{cases} \tag{2.4}$$

Because of the spatial discretization, I use $d(\boldsymbol{y}, \boldsymbol{z}) = |y_E - z_E| + |y_N - z_N|$ as the distance between two locations $\boldsymbol{y}$, $\boldsymbol{z}$ with Easting and Northing $\boldsymbol{y} = (y_E, y_N)$ and $\boldsymbol{z} = (z_E, z_N)$, such that all locations within a distance $\delta$ of a fixed location $\boldsymbol{z}$ form a diamond-shaped area around $\boldsymbol{z}$. I assume that $\text{path}(\boldsymbol{x}_{t-1} \to \boldsymbol{x}_t)$ is the straight line between $\boldsymbol{x}_{t-1}$ and $\boldsymbol{x}_t$. Via equation (2.4), an individual counts the number of steps it remains absent from locations, and therefore $m_t(\boldsymbol{x})$ is the time since last visit to

location $\boldsymbol{x}$ at time $t$. A location is considered visited when the animal comes within a distance $\delta > 0$. Because $m_t$ is obtained recursively, I have to define appropriate starting values. Here, I use movement data prior to the trajectory $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ for initialization. If a location $\boldsymbol{x}$ was visited during the initialization phase, I calculate the time between the last visit to this location and the beginning of the actual trajectory and thus reconstruct time since last visit at time $t = 1$, $m_1(\boldsymbol{x})$. For all locations not visited during initialization, I set time since last visit as the length of the initialization phase. The dynamic variable time since last visit is used in the memory model to inform movement decisions via $\boldsymbol{\mathcal{I}}_t(\boldsymbol{x}) = m_{t-1}(\boldsymbol{x})$. Once $\boldsymbol{x}_t$ is chosen according to the probability mass function in (3.1), $m_t$ is updated via (2.4). Here, I track time since last visit for the entire spatial domain $\Omega$. If the selection coefficient with respect to $m_t(\boldsymbol{x})$ is positive, this leads to any location eventually becoming highly attractive after long enough absence. If this behaviour is not desired, one may adjust the definition of the cognitive map or weighting function appropriately. For example, if prior information about an animal's behaviour is given, it is possible to track time since last visit only for certain locations of specific interest.

### Combination model

In the *combination model*, I allow information types from both the resource and the memory model to influence movement simultaneously by letting the information vector be $\boldsymbol{\mathcal{I}}_t(\boldsymbol{x}) = (r_1(\boldsymbol{x}), \ldots, r_n(\boldsymbol{x}), m_{t-1}(\boldsymbol{x}))$. In particular, this models allows for interactive effects of time since last visit and resource variables, e.g. by incorporating multiplicative terms of the form $\gamma\, r(\boldsymbol{x})\, m_{t-1}(\boldsymbol{x})$ into the interaction term $f(\boldsymbol{\mathcal{I}}_t(\boldsymbol{x}), \boldsymbol{\gamma})$ in the weighting function (3.2). This is important in situations where return times to locations matter depending on the resources at the location, e.g. average return times to preferred foraging areas may differ from those to locations used as shelter.

### Information-independent kernel

I define the movement kernel $k$ based on a step length distribution with density $\mathcal{S}$ and a distribution for movement directions with density $\Phi$. For step length, I use a Weibull distribution with scale and shape parameter $\kappa > 0$ and $\lambda > 0$, respectively, because it has a flexible form and generally shows a good fit with empirical data (Morales *et al.*, 2004). Thus,

$$\mathcal{S}(\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|; \kappa, \lambda) = \frac{\kappa}{\lambda}\left(\frac{\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|}{\lambda}\right)^{\kappa-1} \exp\left(-\left(\frac{\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|}{\lambda}\right)^{\kappa}\right). \tag{2.5}$$

To measure movement directions, I use bearings, and I denote the bearing of the step from $\boldsymbol{x}_{t-1}$ to $\boldsymbol{x}_t$ by $\varphi(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t) \in [-\pi, \pi)$. I include directional persistence by choosing a wrapped Cauchy distribution for bearings with scale parameter $\rho > 0$ and mode at the previous step's bearing $\varphi(\boldsymbol{x}_{t-2}, \boldsymbol{x}_{t-1})$,

$$\Phi\big(\varphi(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t); \varphi(\boldsymbol{x}_{t-2}, \boldsymbol{x}_{t-1}), \rho\big) = \frac{1}{2\pi} \frac{\sinh \rho}{\cosh \rho - \cos(\varphi(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t) - \varphi(\boldsymbol{x}_{t-2}, \boldsymbol{x}_{t-1}))}. \tag{2.6}$$

The wrapped Cauchy distribution is convenient for implementation, and it has been used commonly to model movement directions (Morales *et al.*, 2004; Codling *et al.*, 2008, note that (2.6) is equivalent to their formula with parameter transformation $r = \exp(-\rho)$). One could use alternative distributions, such as the von Mises distribution or wrapped normal distribution (Codling *et al.*, 2008). Assuming that the choices for step length and movement direction are independent, the kernel becomes the product of $\mathcal{S}$ and $\Phi$, describing a correlated random walk,

$$k(\boldsymbol{x}_t; \boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \kappa, \lambda, \rho) = \mathcal{S}(\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|; \kappa, \lambda)\, \Phi\big(\varphi(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t); \varphi(\boldsymbol{x}_{t-2}, \boldsymbol{x}_{t-1}), \rho\big). \tag{2.7}$$

Because the kernel formulates persistent movement and takes into account the bearing of the previous step, I define a simplified kernel for $t = 2$ as

$$\tilde{k}(\boldsymbol{x}_2; \boldsymbol{x}_1, \kappa, \lambda) = \frac{1}{2\pi} \frac{\kappa}{\lambda} \left( \frac{\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|}{\lambda} \right)^{\kappa - 1} \exp\left( - \left( \frac{\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|}{\lambda} \right)^{\kappa} \right). \tag{2.8}$$

This means that I assume a uniform distribution for the first bearing.

Note that this definition of the movement kernel from step length and bearing distributions does not mean that I obtain the kernel from empirical step lengths and bearings in advance and then use this observed kernel to estimate the weighting function parameters in a case-control study, as has been previously suggested for resource-selection analysis (Fortin *et al.*, 2005; Forester *et al.*, 2009). Because movement and resource selection are not independent processes, a decoupled treatment of the processes can lead to biased estimates. I circumvent this problem, and I use the formulation in terms of step length and bearing only to define the functional form of the information-independent movement kernel. During model fitting (see next section) I estimate all model parameters simultaneously from the data.

### 2.2.3 Statistical inference

If information $\boldsymbol{\mathcal{I}}_t$ is known, the likelihood function for the collection of parameters $\boldsymbol{\theta} = (\kappa, \lambda, \rho, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$ for the general model is

$$L(\boldsymbol{\theta}) = p(\boldsymbol{x}_1, \boldsymbol{\theta})\, p(\boldsymbol{x}_2|\boldsymbol{x}_1\boldsymbol{\theta}) \times \prod_{t=3}^{N} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \boldsymbol{\theta}). \tag{2.9}$$

In the memory and combination model, $\boldsymbol{\mathcal{I}}$ includes the variable time since last visit $\boldsymbol{m} = (m_t, t \in T)$, which represents internal information of the animal that in general cannot be observed. However, because of the way I define and initialize $\boldsymbol{m}$, I am able to iteratively calculate the time series $(m_1, \ldots, m_N)$ based on the movement trajectory. Therefore, given the data $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$, time since last visit becomes a known covariate, and the likelihood function in equation (5.11) is valid for all models.

To obtain parameter estimates and their confidence intervals for all models I use data cloning (Lele *et al.*, 2007). Data cloning uses Markov Chain Monte Carlo (MCMC) methods, which are usually employed in Bayesian statistical inference. However, data cloning provides approximations to maximum likelihood estimates (MLE), together with Wald-type confidence intervals, thus facilitating frequentist inference; see Appendix A.1.

I use the approximate MLEs for the model parameters in (5.11) to calculate the corresponding approximate maximum likelihood values. From these, I obtain the Bayesian Information Criterion (BIC) for each of the four models (Burnham & Anderson, 2002). Alternatively, I could have used Akaike information criterion (AIC); however for large datasets, AIC tends to favour overly complex models (Link & Barker, 2006). For each trajectory, I select the model with smallest BIC as the one that explains the decision mechanism of the trajectory best. I use the BIC of this best model as a reference to calculate BIC differences for all alternative models ($\Delta\text{BIC} = \text{BIC}_{\text{alternative}} - \text{BIC}_{\text{best}}$).

A common problem in statistical inference are missed observations. Missed locations in an otherwise regular movement trajectory occur, for instance when GPS devices fail to acquire satellite signal due to closed canopy or otherwise limited available sky. In an autocorrelated trajectory, with each missed location we additionally lose associated information. Calculations of step lengths and bearings require two successive locations. In models with persistent movement, we require not only the current but also the previous bearing for step probabilities. Therefore, in a correlated random walk, one missed location can effectively lead to a gap of two full steps. In MCMC-based data cloning, we can treat missed locations explicitly as unknown variables and

account for this in the likelihood function,

$$L(\boldsymbol{\theta}) = \int p(\boldsymbol{x}_{\text{avail}}, \boldsymbol{x}_{\text{miss}})\, \mathrm{d}\boldsymbol{x}_{\text{miss}}. \tag{2.10}$$

This allows to preserve the entire dependency structure of the trajectory and avoids the need to discard any information. For more information on this, see Appendix A.2.

## 2.3 Simulation study

To verify the functionality of my method I applied the modelling framework and statistical inference method to simulated data. Because (3.1) defines probability mass functions for movement steps, I can sample from them to iteratively generate individual movement trajectories according to the four candidate models. These data have the advantage that I know both a trajectory's underlying mechanism and the parameter values that were used to generate the trajectory. By applying the inference procedure to these data, I investigated whether it was possible to identify the true underlying mechanism of a trajectory and to correctly estimate parameter values.

### 2.3.1 Simulation of landscapes

Because movement decisions in the resource and combination model are based on environmental information, I first simulated landscapes of covariate values. I consider two resources $(r_1, r_2)$, one having a continuous range of values, e.g. a biomass measure or elevation, and the other representing presence or absence of a feature, e.g. a preferred food source, via a binary variable that takes either value 1 or 0. To include biological realism, I accounted for spatial correlations in the covariate values. I simulated five pairs of landscapes with varying spatial structures. For more information see Appendix A.3.

### 2.3.2 Simulation of movement trajectories

I generated movement trajectories using the four candidate models presented above. When I used the null model, I called the resultant trajectory a *null trajectory*, and I named trajectories analogously for the other models.

On each of the five landscape pairs, I simulated a null, resource, memory and combination trajectory, using the same movement parameter values on all landscapes and across all four models, as applicable (Figure 2.3). The kernel parameters $\kappa$, $\lambda$,

$\rho$ appear in all models. The resource model has additional parameters $\alpha_{\text{res}}$, $\beta_1$, $\beta_2$, which are the intercept and the selection parameters with respect to the two resources $(r_1, r_2)$ of the weighting function (3.2). In this model, I assumed there is no interaction between the two resources. The memory model instead has additional parameters $\alpha_{\text{men}}$ and $\beta_{\text{mem}}$, which describe the animal's preferences with respect to time since last visit $\mathbf{m}$. In the combination model, the weighting function includes all effects, such that it has parameters $\alpha_{\text{com}} = \alpha_{\text{res}} + \alpha_{\text{mem}}$, $\beta_1$, $\beta_2$, and $\beta_{\text{mem}}$. In this model, I further allowed for interactions between resources and time since last visit by defining the interaction term in the weighting function as $f\big((r_1, r_2, m_t), \gamma_1, \gamma_2\big) = \gamma_1 r_1 m_t + \gamma_2 r_2 m_t$, where $\gamma_1$ and $\gamma_2$ are the interaction parameters. I chose the main set of parameter values to represent realistic movement behaviour. To account for scenarios for which parameter values were potentially more difficult to estimate from data, e.g. small values of selection parameters, I generated two supplemental data sets, comprising two additional sets of 20 trajectories each generated from alternative sets of parameters; see Appendix A.4.

For all trajectories, I simulated 2600 time steps, of which I discarded the first 1400 steps as initialization. This was particularly important for the memory model, in which I started with a cognitive map having value 0 everywhere. I used the last 400 steps from the initialization phase to calculate $m_1$. Each final trajectory consisted of 1200 time steps, which I considered a length commonly available (e.g. 1200 time steps could represent 50 days of 1-hr data or 100 days of 2-hr data).

For an example of how to handle missed observations, I simulated a combination trajectory with 90% fix rate by removing locations from a trajectory, 5% as single locations and 5% as two successive locations. I chose a trajectory from the main data set, which allowed me to compare results for completely and incompletely observed trajectory; see Appendix A.2 and B.1.2.

### 2.3.3 Analysis of simulated data

To every simulated trajectory, I fitted all four candidate models (Figure 2.3) using data cloning. For details about the data cloning and MCMC procedures, such as number of clones and iterations used, see Appendix A.1. There were two basic types of model fits that I distinguished in my analyses. A model could be fitted to a matching trajectory, i.e. a trajectory that had been simulated using the same model's mechanism (e.g. a resource model fitted to a resource trajectory). Or, a model could be fitted to a non-matching trajectory (e.g. a resource model fitted to a null, memory or combination trajectory). Each model fit generated estimates of the model parameters, together

with Wald-type confidence intervals. Here, I used 95% confidence intervals. Using the approximate maximum-likelihood parameter estimates from data cloning, I estimated the maximum value of the model likelihood, and BIC, for each model fit.

For all parameter estimates, I examined whether their potential scale reduction factors $\hat{R}$ were close to 1 (Gelman & Rubin, 1992). For an MCMC fit, in which parallel Markov chains are used to generate the posterior distribution of a parameter, the potential scale reduction factor of a parameter indicates whether the chains have mixed well and converged. If this is not the case, the estimate that results from such an MCMC is not meaningful. I considered a potential scale reduction factor $\hat{R} \leq 1.1$ to be sufficiently close to 1 (Gelman & Rubin, 1992), and I excluded all parameter estimates that did not meet this condition from my analysis. Whenever such a non-convergent or non-mixing parameter occurred within a model fit, the resultant likelihood and BIC values of the fit were possibly inaccurate. Therefore, if a model fit included one or more parameters with $\hat{R} > 1.1$, I excluded the BIC value from the model-selection analysis.

For each trajectory, I compared whether the best model according to model selection via BIC coincided with the true underlying model of the trajectory. Under the assumption that the framework is functional, I expected the model that matched a trajectory's underlying mechanism to have minimal BIC. Because both the resource model and the memory model are nested within the combination model, I further expected the combination model, when applied to a resource or memory trajectory, to perform better than the simple alternative (e.g. a memory model applied to a resource trajectory).

For matching model fits, I compared true parameter values that were used to generate a trajectory to the parameter estimates obtained from applying the matching model, and I examined whether 95% confidence intervals of parameters included the true values. This should be achieved 95% of the time if parameters are identifiable and the statistical methodology is functional.

In resource-selection analysis, it is usual to use hypothesis testing to determine whether a covariate has an effect or not. I performed an equivalent analysis and examined confidence intervals of the selection parameters $\beta_1$, $\beta_2$, $\beta_{\text{mem}}$, $\gamma_1$, $\gamma_2$ in those model fits, in which the combination model was fitted to a trajectory. The combination model includes all possible covariates, but not all covariates were simulated to have an effect in all trajectories, e.g. a resource trajectory includes effects of the resource variables but not time since last visit. Confidence intervals that corresponded to true underlying effects should exclude zero and vice versa. However, by definition, an $\alpha$-level hypothesis test results in a Type I error of $\alpha$, which I expected to observe approximately

in this analysis. Additionally, I expected a Type II error to occur, where a confidence interval included zero, although the corresponding covariate had an effect. I compared the outcome of this method with the results from model selection via BIC.

I performed all simulations of movement trajectories and statistical analyses in R (R Core team, 2013), using additionally package 'dclone' (Solymos, 2010). To generate MCMC samples, I used JAGS via the R package 'rjags' (Plummer, 2013).

## 2.4 Simulation results

Here, I present results for data generated with the main set of parameters $\boldsymbol{\theta}_1$ (Figure 2.3). Results for supplemental data generated by additional sets of parameters can be found in Appendix B.1.1.

Of all 80 model fits (four models fitted to 20 trajectories each), 80% had potential scale reduction factor $\hat{R} \leq 1.1$ for all model parameters. In the remaining model fits, at least one parameter had $\hat{R} > 1.1$ (Figure 2.4). Convergence or mixing problems never occurred when the null model was fitted to a trajectory, even if the trajectory had a more complex underlying mechanism. Large $\hat{R}$ values only occurred if the fitted model contained parameters that were inapplicable to the model that was used to generate the trajectory. This was the case when any of the more complex models was fitted to a null trajectory, when the combination model was fitted to a resource or memory trajectory, or when the memory model was fitted to a resource trajectory and vice versa. In these model fits, the non-convergent parameters were mainly those that did not correspond to true underlying covariate effects. However, when in a model fit problems occurred for multiple parameters, occasionally even applicable parameters failed to converge. In matching model fits, Markov chains always mixed well and converged. For more details on convergence, see Appendix B.2 and B.3.

Our model selection framework was able to correctly identify the true underlying model for all trajectories (Figure 2.4). When a trajectory had underlying resource or memory mechanism, the next best model was always the combination model with $\Delta$BIC being a magnitude smaller than for the alternatives. This pattern was only disturbed if the combination model experienced convergence problems and was therefore excluded from further analysis.

Parameter estimates in matching model fits agreed well with true underlying parameter values. Parameter estimates generally were both close to and balanced around their true values (Figure 2.5). 95% confidence intervals (n=115) included the true parameter value 91% of the time. Considering also results from the supplemental data,

94% of all confidence intervals (n=345) included the true value.

Our hypothesis test as to whether covariates had an effect agreed with my expectations. The combination model fitted to the 20 trajectories lead to 73 estimates of selection parameters, of which 39 corresponded to true underlying effects. Analyzing their confidence intervals, I obtained a false positive rate (Type I error rate) of 0.09 and a false negative rate of zero, i.e. Type II errors did not occur. However, if I also considered the supplemental data and thereby increased the amount of resultant parameter estimates with confidence intervals to a total of 217, I obtained a Type I error rate of 0.04 and again a Type II error rate of 0.09.

## 2.5 Discussion

In recent years, the link between animal movement and spatial memory has received increasing attention (Smouse *et al.*, 2010; Fagan *et al.*, 2013). Studies of animal behaviour and cognition have given useful insights into animals' capacities to remember past experiences and use spatial memory. Most results have been obtained through experiments in confined and synthetic settings. However, to better understand how important ecological processes such as movement and dispersal are shaped by cognitive processes and memory, we also need to look at animals in their natural environments (Tsoar *et al.*, 2011). Understanding the components of individual movement decisions and their interactions ultimately will help us to predict how population distribution patterns respond to environmental changes, such as landscape fragmentation and changing climate.

I have presented a modelling framework that can be used to detect the influence of memorized information on movement decisions. I recognize that in many situations it is difficult to confirm that animals draw upon memorized information instead of momentarily perceived information; however, there is evidence that animals use information that they have obtained during past experiences (Martin-Ordas *et al.*, 2009; Janmaat *et al.*, 2013). As an example of such information, I use time since last visit to locations. In my model, time since last visit is continuously updated during the movement process and at the same time influences movement decisions. I formulate my models in a way that makes them amenable to likelihood-based statistical inference. This allows to fit the models to data to test whether the timing of events plays a role for movement decisions. Fitting the full model (3.1), encompassing both general movement tendencies and selective behaviour, to data via the likelihood function (5.11) enables simultaneous estimation of parameters of both the general movement kernel

and weighting function. This distinguishes my method from step selection approaches that use an empirical movement kernel to estimate resource selection parameters in a case-control framework (Fortin *et al.*, 2005; Forester *et al.*, 2009).

In my definition of the weighting function (3.1), I followed the classical formulation of resource-selection functions and evaluated a movement step based on the information at the endpoint of the step. In the memory model this means that an animal may cross recently visited locations on its path although these have low weights. Depending on the behaviour of the study species, it may be appropriate to change this so that cognitive map values along the entire path are considered, thus following the idea of step selection functions (Fortin *et al.*, 2005; Potts *et al.*, 2014). In my framework, it is straightforward to define the weighting function as a function not only of $\boldsymbol{x}_t$ but also $\boldsymbol{x}_{t-1}$ and to include any information related to the step from $\boldsymbol{x}_{t-1}$ to $\boldsymbol{x}_t$. Endpoints are observed locations and therefore have certainly been used. To include information about entire steps, we must make an assumption about which locations were visited between observed locations. In the memory model, I assume this is a straight line, however one may use more sophisticated methods similar to Brownian bridges (Horne *et al.*, 2007).

I used simulated landscapes and movement data to verify the functionality of my modelling framework and statistical inference method. Adding the memory process to the modelling framework considerably increased model complexity and the amount of data that had to be processed. I was therefore interested in whether I could correctly detect memory effects in empirical movement patterns and whether parameters that describe the memory process and its interactions with other variables were possible to estimate reliably. To perform inference, I used data cloning, which uses MCMC techniques but facilitates frequentist inference. I used the software package JAGS, which allowed me to define models in an easily understandable language and provides a stable implementation of MCMC sampling. JAGS was able to adapt the sampling process successfully so that parameters of very different magnitude could be reliably estimated. However, this came at the cost of long computation times (ranging 0.5-5 days per single chain for different models) and high memory needs (ranging 1-5 GB RAM). Alternatively, I could have used conventional numerical maximization of the likelihood function, which in this case may have been faster but at the same time more limited. Because data cloning is based on MCMC, it is amenable to extensions of my model to include partially observed and hidden processes. I have demonstrated this with an example on missed observations. Any Bayesian method would provide this option and it may be a matter of belief whether frequentist of Bayesian approaches

are used. However, data cloning additionally provides tools to detect parameter estimability problems (Lele *et al.*, 2010), which was relevant in my analysis; compare Appendix B.2. At this stage, data cloning via JAGS was computationally intense, and it may be worth to explore alternative options, e.g. a 'home made' MCMC sampler in a fast language such as C/C++. Still, with quickly increasing computational capacities and advances in statistical software, I believe that my method has a promising future.

Verification of the method was successful. In matching model fits, almost all MCMC runs mixed well and converged. Convergence and mixing problems occurred in non-matching model fits and especially for parameters that were not meaningful to the trajectory (e.g. a resource selection parameter for a memory trajectory). For further application of the method, I have given recommendations how to proceed in cases on non-convergent model fits (Appendix B.3). Model selection via BIC successfully identified trajectories' true underlying mechanisms, and if parameters in a model fit were applicable to the underlying trajectory, I was able to recover true parameter values. Simulated movement trajectories were samples of stochastic processes, and therefore realized parameter values were subject to stochasticity. Thus, parameter estimates could not be expected to exactly coincide with the true values. Verifying the functionality of my method was particularly important with respect to the newly introduced memory process. I conclude that if time since last visit is a driver of observed movement trajectories, my framework is able to detect this.

When I compared results from model selection to outcomes of hypothesis tests, I found that model selection was better able to distinguish true underlying mechanisms of trajectories. By definition, hypothesis tests allow for a Type I error, the size of which is influenced by the level of the test. However, decreasing the Type I error simultaneously decreases the power to correctly detect effects of covariates and increases the Type II error. The model selection framework is not based on this concept, and it proved to be more accurate in my analysis.

I have built on the framework of spatially explicit resource-selection models and added the influence of a dynamic memory process on movement decisions by introducing a dynamic cognitive map and linking it with the movement and resource-selection process. The existence of cognitive maps in animals is debated, and there is especially controversy about what form such maps may take, e.g. whether animals use topological cognitive maps for landmark-based navigation or whether animals can create and use geometric cognitive maps that preserve angles and distances between locations (Bennett, 1996; Asensio *et al.*, 2011; Collett *et al.*, 2013). This debate also includes the question whether spatial information in the brain is encoded with respect to the posi-

tion of the viewer, i.e. egocentrically, or independently of the position of the viewer, i.e. allocentrically (Yeap, 2014). In my models I do not focus on navigational mechanisms but decision making processes, and I use the cognitive map as a useful mathematical tool to model spatial information. Investigation of different navigational mechanisms within a model-selection framework similar to that presented here could be the goal of future research. With my model formulation in terms of a cognitive map, I have provided a general framework for linking movement with information use and acquisition. I emphasize that within this general formulation, a variety of more specific formulations of cognitive maps can be realized, tailored to the situations and behavioural processes of interest.

In my candidate models, I have used time since last visit to locations as an example of a form of dynamic information that is mediated by the cognitive map. I have demonstrated how the time since last visit to a location can shape the movement process, either on their own or in interaction with environmental variables. Such behaviours can, for instance, occur when animals patrol their home ranges for defence purposes, when predators counteract behavioural depression, or when animals rely on resources that vary in their availability due to depletion. However, my modelling framework and its elements are flexible and can be extended to include other forms of dynamic information and experiences that animals collect during their movement. For instance, while animals travel they may gather information about seasonally available resources. Observations of Mangabeys show evidence that they remember fruiting statuses of fig trees and use this information to predict the fruiting status of those trees at later times (Janmaat *et al.*, 2013). Prey species can use their movement to collect information about the distribution of predators. Such information can enable prey to reduce costly anti-predatory behaviours and therefore outweigh attack risks connected to the information collection. This has been suggested to explain movement behaviour of caribou towards wolf paths (Latombe *et al.*, 2014).

In my models, I have reconstructed time since last visit from the movement path, using an initialization period as basis. Because an observed movement path consists of a discrete series of locations, we must make an assumption about the path between two successive locations. Here, I simply defined a buffer zone around the straight line between two locations and considered all locations within this buffer as visited. Another option would be to formulate time since last visit as a random variable and incorporate it via a hierarchical model structure, e.g. a state-space model, in which time since last visit is a hidden process. Such a formulation may also become useful when integrating a dynamic variable that cannot be reconstructed from the movement path. Including

a high-dimensional hidden process, however, would increase the computational burden mentioned above even further.

Although my models describe movement behaviour of individual animals, the ideas I have presented can also apply to other systems. A specific feature of my models is the interaction between a movement process and an information, or memory, process. A similar dynamic interplay can arise on a larger scale when a species disperses and expands its range. While moving into a new environment, the dispersing species might alter the environment and its species composition, which in turn could affect the dispersing species (Gilman *et al.*, 2010). Such processes could be analyzed with the same mathematical ideas and modelling tools as I have presented here. Thus, I have presented a powerful modelling approach to identify spatial memory and dynamic information as drivers of movement decisions, and my framework and its elements promise a wide range of applications within movement ecology.

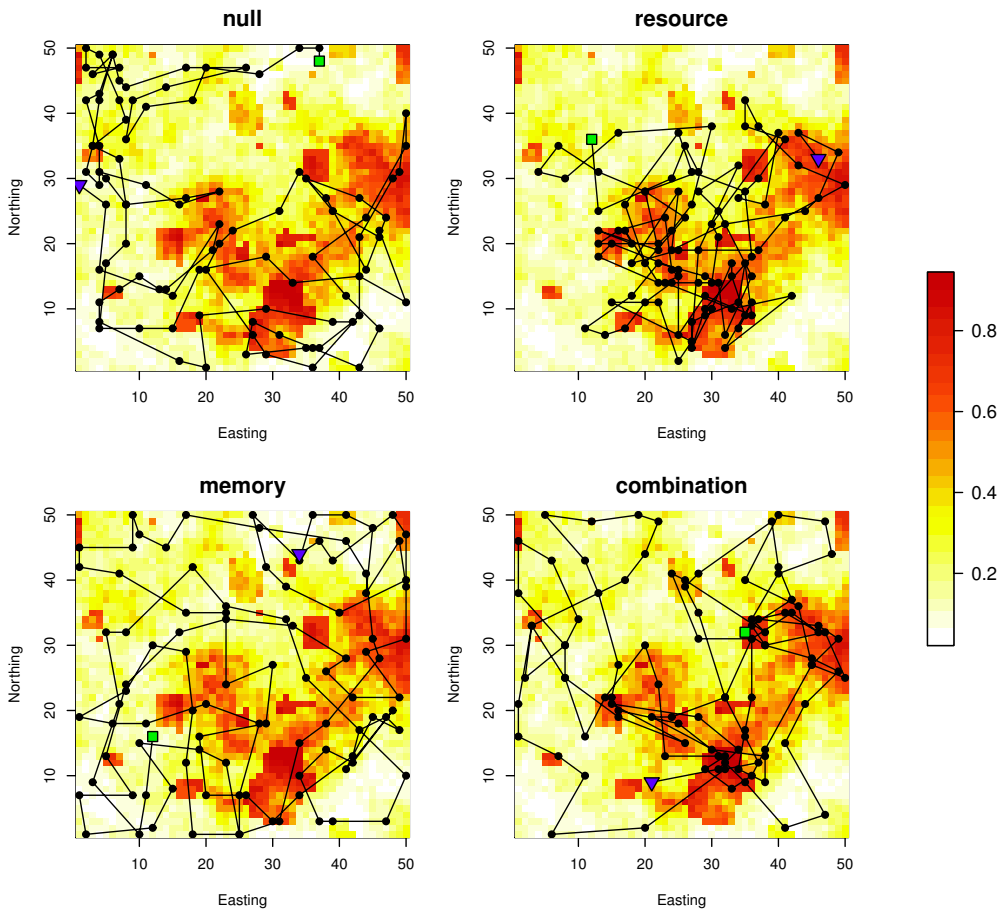



**Figure 2.1.** Example trajectories from the four candidate models, 100 steps long, with starting location marked by a green box and final location marked by a blue triangle. All trajectories are plotted on top of an example resource selection function $w(x;\alpha,\boldsymbol{\beta}) = \big[1+\exp\big(-\alpha-\beta_1\, r_1(\boldsymbol{x})-\beta_2\, r_2(\boldsymbol{x})\big)\big]^{-1}$ generated from two resources $r_1$ and $r_2$. The null model does not consider resource information and therefore the null trajectory visits locations irrespective of the resource selection function. The memory model does not consider resource information either, however, the animal avoids recently visited locations and is attracted to locations with long time since last visit. Therefore, the memory trajectory efficiently explores the spatial domain in a patrolling fashion. In contrast, the resource trajectory mainly remains in areas where the resource selection function has high values. The combination trajectory shows a mixture of behaviours from the resource and the memory model. The trajectories were generated using the first landscape pair and main parameter set from the simulation study; compare Figure 2.3.

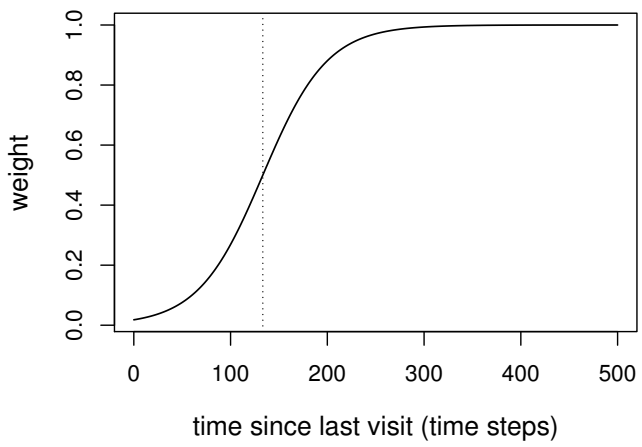



**Figure 2.2.** In the memory model, the weight $w_t(\boldsymbol{x})$ of a location $\boldsymbol{x}$ depends on time since last visit $m_{t-1}(\boldsymbol{x})$ to that location. Locations that have been visited recently have low weights and are thus avoided. A weight of 0.5 is attained when $m_{t-1}(\boldsymbol{x}) = -\frac{\alpha}{\beta}$ (dotted vertical line).

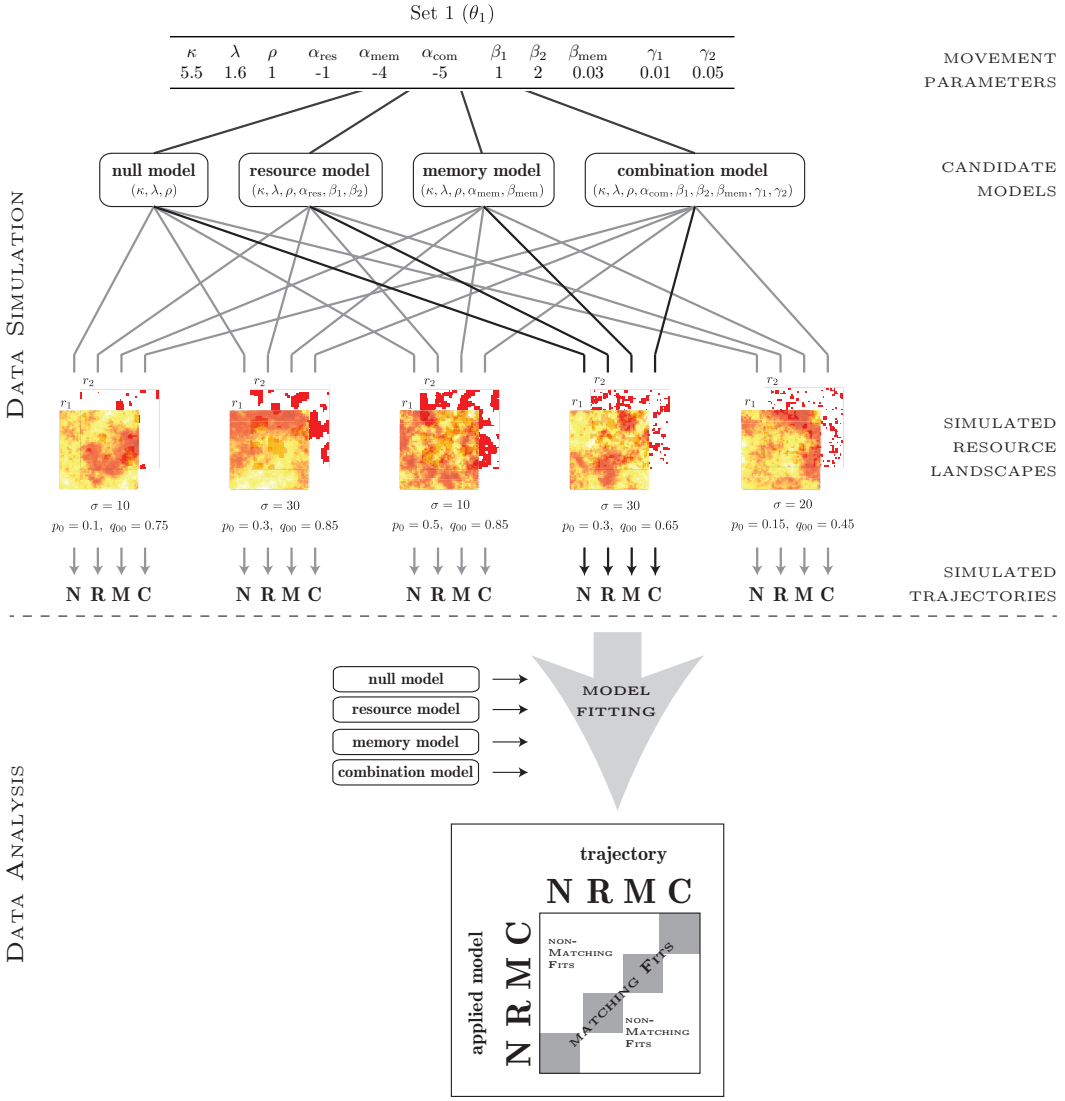



**Figure 2.3.** Overview of data simulation and model fitting. For a set of parameter values, I generated trajectories using all four candidate models. Using each model, I simulated trajectories on five different landscapes resulting in 20 trajectories. Each trajectory was then fitted with all four models, leading to a total of 80 model fits.

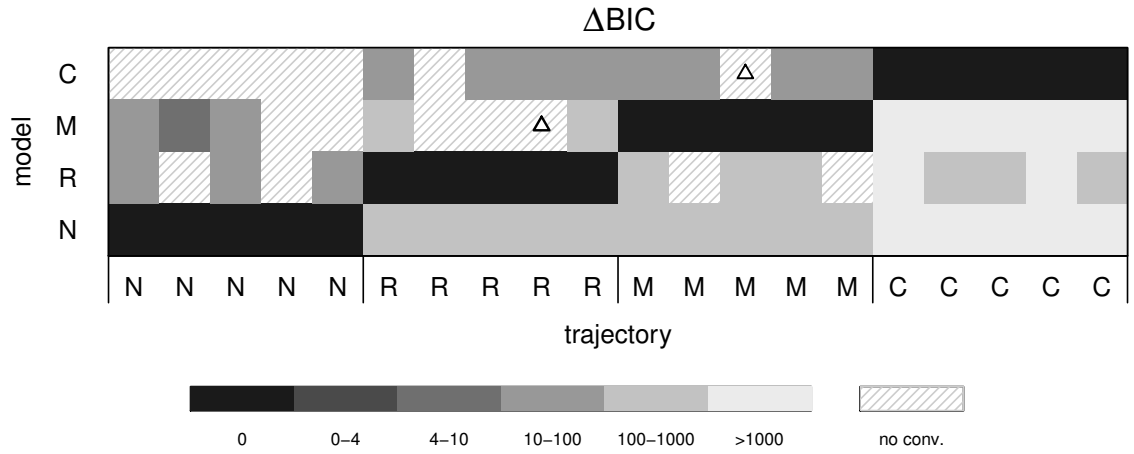



**Figure 2.4.** Each column shows model selection results for one simulated trajectory when fitted with the null (N), resource (R), memory (M) and combination(C) model. For each trajectory, I calculated BIC values for the four fitted models, and the figure shows differences in BIC with respect to the minimal BIC value, i.e. the model with minimal BIC has $\Delta\text{BIC} = 0$. I excluded model fits with non-convergent MCMC. Triangles indicate trajectories for which I calculated estimability diagnostics; compare Appendix B.2.

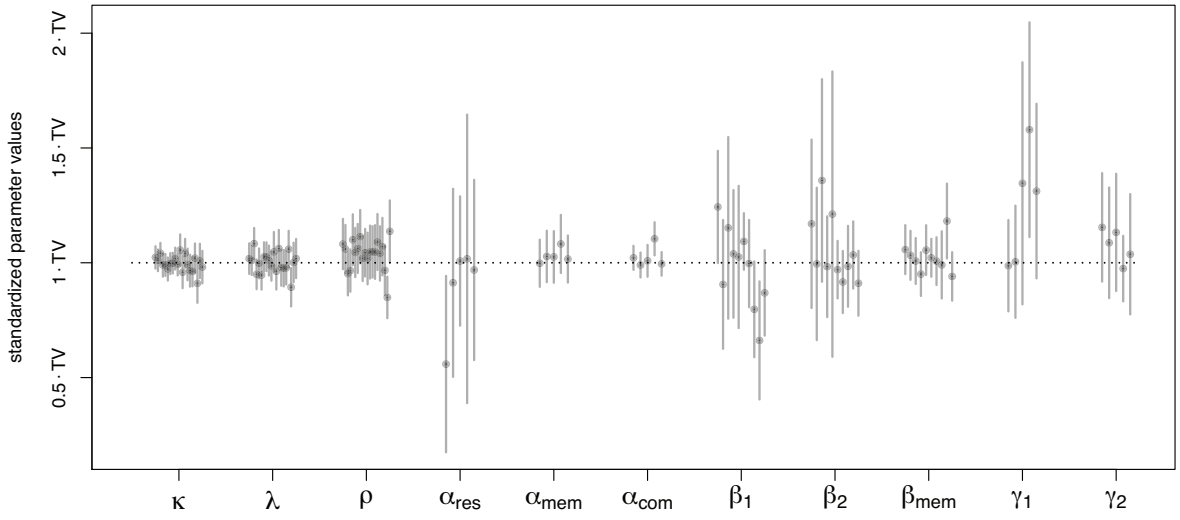



**Figure 2.5.** Parameter estimates and their 95% confidence intervals for matching model fits (each trajectory fitted with the same model that was used to generate the trajectory). Both parameter estimates and Wald-type confidence intervals are scaled by the true parameter values (TV): $\kappa = 5.5$, $\lambda = 1.6$, $\rho = 1$, $\alpha_{\text{res}} = -1$, $\alpha_{\text{mem}} = -4$, $\alpha_{\text{com}} = -5$, $\beta_1 = 1$, $\beta_2 = 2$, $\beta_{\text{mem}} = 0.03$, $\gamma_1 = 0.01$, $\gamma_2$=0.05.

# Chapter 3

# Prudent prey-management: do wolves keep track of space and time?

## 3.1 Introduction

Wolves play on one side of intricate predator-prey interactions. Predators do not only reduce prey numbers through consumption but also have non-lethal effects on prey that can be far-reaching. By posing fatal risks, predators create a "landscape of fear", to which prey respond by altering their foraging behaviour (Brown *et al.*, 1999; Laundré *et al.*, 2001). Although anti-predator behaviours increase survival, they also result in lower consumption rates (Lima & Dill, 1990). For example, in the presence of wolves, ungulates increase their vigilance and spend more time scanning the environment, thus reducing the time available for feeding (Laundré *et al.*, 2001; Liley & Creel, 2007). Additionally, when exposed to predation risk, prey tend to select safer habitats by moving into open areas where predators can be earlier detected or by seeking shelter in forested areas (Fortin *et al.*, 2005; Latombe *et al.*, 2014). Because safe habitats often do not coincide with optimal foraging habitat, this means to forego foraging opportunities. Predator-induced shifts in behaviour have further implications for communities and ecosystems because they release herbivory from plants (Fortin *et al.*, 2005; Ripple & Beschta, 2012) and can also stabilize predator-prey population dynamics (Brown *et al.*, 1999).

Predators' non-lethal effects on prey feed back to predators, requiring them to adapt their behaviour themselves. Anti-predatory behaviours make prey less vulnerable to

predation, thereby reducing capture rates. Prey availability thus reduced is termed behavioural depression of prey (Charnov *et al.*, 1976). While prey should adjust their behaviour to optimize food intake while minimizing predation risk, predators face another optimization problem: they must pursue prey but also temporarily release hunting pressure to allow prey recovery. Charnov's marginal value theorem (MVT) predicts that predators should leave a foraging patch when the rate of food intake drops to the overall expectation over the available space (Charnov *et al.*, 1976; Charnov, 1976). For wolves, food intake rate may not be the only measure to determine patch departure time. Ungulates respond in their behaviour not only to hunting events by wolves but also to mere wolf presence or passage within distances of one kilometre or more (Liley & Creel, 2007; Latombe *et al.*, 2014). With this in mind, we may interpret the MVT in the sense that wolves should give up search in an area when the expectation of making a successful kill drops below the average over their available space.

To optimize hunting success, wolves should not only optimize giving-up times but also adjust return times, which is influenced by the time it takes prey to reduce anti-predator behaviour (Latombe *et al.*, 2014). In a call for a stronger consideration of predators as strategically behaving actors in predator-prey interactions, Lima (2002) proposed that predators may adopt a prey-management strategy that is "prudent" in the sense that it avoids excessive hunting in a given area and instead spreads the risk over all hunting sites. This idea that wolves actively act against behavioural depression has been supported by Jedrzejewski *et al.* (2001), who studied movements of wolves in Białowieża Primeval Forest, Poland. They found that wolves changed their travel routes daily, with little overlap between daily hunting areas, both during spring-summer and fall-winter seasons. Especially in fall and winter, wolves tended to rotate through their territories, returning to the same areas roughly every sixth day. As second explanation for wolves' rotational movement patterns, Jedrzejewski *et al.* (2001) mention territorial mechanisms. Packs strongly defend their territories against intruders and use scent marks, among other cues, to signal their presence to foreign wolves (Peters & Mech, 1975; Mech & Boitani, 2006). To maintain active scent-marks, which are most numerous along territory boundaries, wolves need to revisit locations and traverse their territories regularly (Peters & Mech, 1975; Zub *et al.*, 2003). However, the study by Jedrzejewski *et al.* (2001) remained largely descriptive, not testing the proposed explanations for their observed movement patterns.

I apply the models developed in Chapter 2 to movement data of gray wolves (*Canis lupus*) in Alberta, Canada, to investigate whether wolves move according to the prudent prey-management strategy. The models describe movement as a decision process

that can incorporate effects of the environment as well as cognitive effects. Therefore, I can use them to test wolves' movement strategies and their underlying mechanisms. Unless wolves follow fixed routes very strictly, rotational movement behaviour requires wolves to incorporate information about their own travel history into movement decisions. This information has both a spatial ("where?") and a temporal ("how long ago?") component. Basically, wolves need to follow two rules: 1) avoid recently visited areas, and 2) travel to areas with long absence. I incorporate these behavioural rules into the modelling framework and combine them with effects of prey density to test the hypothesis that wolves counteract behavioural depression of prey. Under the assumption that the hypothesis is true, timing of visits is particularly important for areas with high prey densities and less relevant for areas without prey. Thus, I expect rules 1) and 2) to drive movement decisions in interaction with prey density. I additionally account for possible territorial defense mechanisms. Wolves tend to avoid locations close to the territory boundary, because they pose a greater risk of fatal interactions with foreign wolves (Mech, 1994). However, to maintain scent-marks along the boundary, especially rule 2) becomes important for outer regions of territories, and I expect this effect to reduce an general avoidance of boundary locations, given long absence.

## 3.2 Methods

### 3.2.1 Wolf movement and prey data

The data used for this study were collected in an 25,000 $km^2$ area west of Rocky Mountain House, about 200 km southwest of Edmonton in Alberta, Canada (52°27′N, 115°45′W). The area is part of the east slopes of the Rocky Mountains, and terrain includes gentle foothills in the eastern parts as well as mountains ($< 3100$m) towards the west. Much of the landscape is covered by conifer forest (52.1%), which is interspersed with smaller areas of natural lowlands (10.4%), forestry cut-blocks (5.7%) and stands of deciduous forest (2.7%) (Webb *et al.*, 2008). During the time of data collection, wolf density was on average 22.3 wolves/1000$km^2$ in the eastern part of the study area, which declined to 9.7 wolves/1000$km^2$ towards the more mountainous west (Webb, 2009).

During the years 2004–2006, wolves were captured and fitted with GPS collars

(Lotek 3300Sw store-on-board GPS collars); for details see Webb (2009). The collars were programmed to collect location measurements every two hours. This led to time series of observed movement steps, each step representing an interval of two hours. Unsuccessful fix attempts, for example due to poor satellite signal, led to missed observations. Because I expected rotational movement patterns to be strongest during the winter, which is the season where cubs are old enough to leave den and rendezvous sites and travel with the pack, I focused on movement data during the time period 1 November – 15 April. I analyzed data from three wolves with IDs 220, 284, and 285. To avoid strong effects of elevation, for example movements being concentrated in valleys, I selected individuals with territories in the eastern part of the study area with lower elevations. Additionally, I only chose wolves with long enough time series (more than 1200 steps).

In addition to the wolf movement data, information was collected about winter prey densities in the study area. The four major prey species for wolves were deer (*Odocoileus* spp.), elk (*Cervus elaphus*), moose (*Alces alces*) and feral horses (*Equus caballus*). To obtain spatially-explicit maps of densities, fecal pellet groups deposited over winter were counted across transects after snow melt. Pellet counts from transects were interpolated across the study area using inverse-distance weighting. Counts of pellet groups were converted to numbers of individuals with the help of estimated numbers of individuals within wildlife management units obtained through winter areal surveys. For moose and elk, ratios of number of pellet groups to number of individuals were calculated directly based on the aerial surveys. For deer and feral horses, the ratio obtained for moose was adjusted for deer and horses based on differences in winter defecation rates of the species. For more details, see Webb (2009) and McPhee *et al.* (2012).

From densities of individual prey species (deer, elk, moose and feral horse), I calculated a weighted sum of all prey densities to obtain a combined measure of available prey biomass. For this, I used ungulate weights from the literature, averaging estimates of adult males and females (Knopff *et al.*, 2010). Prey densities (numbers per area) were available at a resolution of 30 m, that is as number per 900 $m^2$. I aggregated these data to a coarser spatial resolution with 300 m $\times$ 300 m cells. I discretized location measurements of the movement trajectories accordingly. That is, I considered trajectories as movement on a discrete spatial grid of land cells, using the coordinates of the cell centres. A continuously measured location was converted to a cell by assigning it to the cell that it fell in.

### 3.2.2 Models

#### Step probabilities

I analyzed the trajectory of each wolf separately, using the individual-based statistical models introduced in Chapter 2 (Schlägel & Lewis, 2014). The models build on random walks and incorporate effects of spatial information on movement decisions. They are based on spatially-explicit resource selection models, which have arisen from traditional resource selection functions (Manly *et al.*, 2002; Boyce *et al.*, 2002) from a need to improve the definition of *availability* of resources (Arthur *et al.*, 1996) and to incorporate spatially-explicit effects of home ranges (Rhodes *et al.*, 2005).

In these models, the probability of a movement step from location $x_{t-1}$ to location $x_t$ between times $t-1$ and $t$ is affected by general movement tendencies as well as preferences for certain locations based on environmental features and past experiences. The step probability is given by

$$p(x_t|x_{t-1}) = \frac{k(x_t; x_{t-1})\, w_t(\boldsymbol{\mathcal{I}}(x_t))}{\sum_{z\in\Omega} k(z; x_{t-1})\, w_t(\boldsymbol{\mathcal{I}}(z))}. \tag{3.1}$$

The function $k$ is an information-independent movement kernel that reflects general movement tendencies, that is how far and and in which direction an individual may travel within a time interval without considering the environment. Thus, the function $k$ takes care of motion capacities, as defined in the framework by Nathan *et al.* (2008). The information-based weighting function $w_t$ assigns weights to locations based on spatial information. Both functions, $k$ and $w_t$, depend on model parameters that I describe in more detail further below. Note that locations $x_{t-1}$, $x_t$, $z$ represent discrete cells in space, as described above. The sum in the denominator of equation (3.1) is a normalization constant over space. To reduce computational burden, I only summed over a large enough area $\Omega$ around the individual's current location $x_{t-1}$. I chose this area large enough so that the probability of stepping outside this area was negligibly small.

The weighting function $w_t$ is a central element of this model. Given the probability that an individual may encounter a location based on its motion capacities, which is given by the kernel $k$, the weighting function gives the probability that an individual may select a location based on certain features of that location. In general, relevant features could be topographical, for example elevation, the land cover type, food availability, or even past experiences about the location. The weighting function is thus a resource-selection probability function (RSPF) (Lele & Keim, 2006; Lele *et al.*, 2013).

An RSPF can take various forms, and here I used one of the most common forms, in which the RSPF is a logistic function. The value of the weighting function at location $x$ is then defined as

$$w_t(\boldsymbol{\mathcal{I}}(x)) = [1 + \exp(-\alpha - \boldsymbol{\beta} \cdot \boldsymbol{\mathcal{I}}_t(x) - f(\boldsymbol{\mathcal{I}}_t(x), \boldsymbol{\gamma}))]^{-1}, \tag{3.2}$$

where the dot $\cdot$ defines the dot product between two vectors. The vector $\boldsymbol{\mathcal{I}}_t(x)$ contains all spatial information about location $x$ at time $t$. The spatial information $\boldsymbol{\mathcal{I}}_t(x)$ about location $x$ may change over time, which also leads to a change in the weighting function. Therefore, I use the subscript $t$ for both the spatial information and the weighting function.

The parameters $\alpha$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ shape the weighting function. The parameter $\alpha$ can be thought of as an intercept that determines the values of the weighting function when the information variables are zero. Generally, parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ influence the direction of selection and selection strength. The weighting function includes an additive term $\boldsymbol{\beta} \cdot \boldsymbol{\mathcal{I}}_t(x)$ and a multiplicative term $f(\boldsymbol{\mathcal{I}}_t(x), \boldsymbol{\gamma})$, which contains terms of the form $\gamma \cdot \mathcal{I}_{t,1}(x) \cdot \mathcal{I}_{t,2}(x)$. Note that because the weighting function is non-linear, both additive and multiplicative terms can lead to interactive effects on the overall probability of selection. This is in contrast to linear models (e.g. linear regression models), in which additive (i.e. independent) and multiplicative (i.e. interactive) terms can be distinguished.

The general movement kernel $k$ is the density function of a random walk in discretized two-dimensional space. For this, I sampled a continuous-space density at discrete points (representing the centre location of each cell in the landscape). The normalization constant in equation (3.1) assures that step probabilities are properly normalized over the discretized space. I used a Weibull distribution for step lengths (Morales *et al.*, 2004) and assumed a uniform distribution for bearings. A major reason for using simply a uniform distribution for bearings was to retain as much steps as possible. In a correlated random walk, bearings are autocorrelated, and therefore three successive location measurements are needed to define the probability for one movement step. With missing measurements in the time series, this would additionally decrease the number of available steps. Thus the kernel is given by

$$k(y; x) = \frac{1}{2\pi\|x - y\|} \frac{\kappa}{\sigma} \left( \frac{\|x - y\|}{\sigma} \right)^{\kappa - 1} \exp\left( - \left( \frac{\|x - y\|}{\sigma} \right)^{\kappa} \right), \tag{3.3}$$

where $\kappa$ and $\sigma$ are the shape and scale parameter of the Weibull distribution, respec-

tively. When the kernel $k$ is included in the step probability (3.1), the factor $\frac{1}{2\pi}$ can be omitted, because it cancels with the denominator. The factor $\frac{1}{\|x-y\|}$ is due to a change from polar coordinates (using step lengths and bearing to define a step) to euclidean spatial coordinates.

In this analysis, I concentrated on movement decisions within a wolf's territory, and I defined probabilities for steps outside the territory to be zero. This means that, given a current location $x_{t-1}$ close to the territory boundary, the step probability (3.1) is zero for $x_t$ outside the territory. For locations inside the territory, the probabilities are adjusted accordingly via the normalization constant in equation (3.1) by only summing over locations within the territory. The main reason for this was to avoid biased results. Especially in areas of high wolf density, territory formation is not only influenced by available resources but also other wolf packs. Therefore, if a wolf did not visit a location outside the territory, this may have been due not only to the spatial information considered in $\boldsymbol{\mathcal{I}}_t$, but also to the presence of other wolf packs or anthropogenic features. I did not consider these effects in the weighting function but concentrated on movement decisions of wolves on a daily basis within the territory. To define the territory, I used an estimate of space use based on Brownian bridges (Horne *et al.*, 2007). Brownian bridges estimate the movement path between two successive locations of a sampled trajectory by assuming Brownian motion, considering both the distance between the locations and the length of the time interval. Brownian bridges can be used to estimate density functions for space use not only from independent observations but based on the full movement path of an animal. To calculate the Brownian bridge home range estimate, I used the R package adehabitatHR, which implements the method proposed by Horne *et al.* (2007) (Calenge, 2006). The calculation of the utilization distribution required two parameters. Because movement trajectories were defined on discretized space with $300 \times 300$ m cells, I chose a small value (0.05) for the parameter representing location measurement error. The second parameter, the Brownian motion variance parameter, is related to the speed of the animal, and was obtained through maximum likelihood estimation based on the movement trajectory (Horne *et al.*, 2007). From the resulting utilization distribution, I defined the territory by all locations within the 99.9% quantile. This high quantile was chosen to ensure that all locations of the trajectory were contained within one connected area (without single points or small islands), and with this choice of the home range this was achieved with very few exceptions. I also chose the high quantile from a conservative perspective to avoid restricting step probabilities too much.

### Spatial information

To test the hypothesized explanations of rotational movement behaviour, I considered three different types of spatial information in the weighting function (3.2). First, I included the combined prey density measure, which was a weighted sum of densities over all prey species according to their body mass. I used this combined prey density to average out species-specific effects. Prey availabilities of individual species varied across territories, and wolves' habitat selection may vary with prey availability but also landscape configurations (Milakovic *et al.*, 2011). Such differences were not the focus here, and the combined prey density allowed to consider prey availability more generally. Second, for territorial defense mechanisms, locations closer to the territory boundary play an important role. Therefore I defined a type of spatial information to give the minimum distance from each location to the territory boundary. The territory boundary is the outer edge of the territory estimate based on Brownian bridges. Note that this is reverse from typically used measures, which calculate the distance from a central location (Rhodes *et al.*, 2005).

The third type of spatial information was based on an individual's own travel history, which changed at every time step. I defined the variable *time since last visit (TSLV)* to specify at each time step $t$, and for each location $x$, the time since the animal had last been to the location, denoted by $m_t(x)$. For example, if between times $t-1$ and $t$ the animal moved from location $x_{t-1}$ to $x_t$, I considered all locations on the path from $x_{t-1}$ to $x_t$ as most recently visited and and set their value of TSLV at time $t$ to be one. That is, I defined $m_t(z) = 1$ for all locations $z$ that lie on the path between $x_{t-1}$ and $x_t$. For the calculation of TSLV, I defined the path to be the straight line between $x$ and $y$. Because it is unlikely that an individual moved in a straight line, I also considered locations within a certain distance of the line as visited (for the purpose of calculating TSLV). For these locations, TSLV was also set to one. Using a buffer around the straight line between two locations is a simple way of accounting for the fact that we do not observe all locations that an animal visits on its path. A more sophisticated approach would be to implement, for example, a Brownian bridge (Horne *et al.*, 2007) for the estimated path between two successive locations. One could even go further and expand a Brownian bridge model to include the more complex movement mechanisms studied here.

For all other locations that were not considered visited, TSLV increased by one at every time step. That is, I set $m_t(z) = m_{t-1}(z) + 1$ for locations $z$ not visited during times $t-1$ and $t$. This led to a map with values of TSLV similar to a map with environmental information, but which changes at every time step. TSLV increased in

areas that the individual stayed away from and was reset to one whenever an individual visited a location, that is when it came sufficiently close to the location. The dynamic map was updated at the end of each movement step, and therefore the weighting function $w_t$ at time $t$ was based on TSLV at time $t-1$, $m_{t-1}(x)$ for all $x$.

Given TSLV for some point in time, it is straightforward to update it for all following time steps based on the animal's movement path. To obtain an initial map of TSLV, I separated movement trajectories into two segments. I used the first 300 or 400 movement steps, depending on data availability, to initialize TSLV and used the rest of the trajectory for statistical inference. The time that corresponded to the beginning of the second part of the trajectory was set to be $t = 1$. I calculated TSLV at $t = 1$ for all locations that were visited during the initialization phase. For locations that were not visited, I set TSLV to the length of the initialization phase, that is to 300 or 400, as appropriate. Note that in Chapter 2, the variable TSLV was described in the section "Memory model" and the spatial map with values of TSLV was referred to as cognitive map. The variable remains the same.

To calculate TSLV, it was necessary to define which locations were considered as visited, additionally to the straight line between the beginning and end of a movement step. Because I aimed to understand the influence of the travel history in relation to prey, I took into account at which distances wolf presence influences prey behaviour. Studies on elk-wolf relationships found that wolf presence can affect elk behaviour, such as group size, vigilance, and movement rates, at distances of 1–5 km (Liley & Creel, 2007; Proffitt *et al.*, 2009). Here, I used discretized space with landscape cells of size 300 × 300 m. I defined a buffer around a cell by using the distance measure $d(x, y) = |x_{\text{east}} - y_{\text{east}}| + |x_{\text{north}} - y_{\text{north}}|$, where the subscripts relate to the eastern and northern component of a two-dimensional location. The coordinates were taken from the centre of each cell. With this distance measure, the buffer becomes a diamond-shaped area around the centre cell. If we define a buffer of size $\delta$ around the location $x$, the corners of the buffer area are those cells that are $\delta$ cells away from $x$ in exact northern, eastern, southern and western direction. For the calculation of TSLV, I used a buffer of size of four cells. I calculated the buffer for each cell that is intersected by the straight line of a step. A distance of four cells in the discretized space corresponds to 1.2 km in continuous space.

All of the trajectories contained missed observations. If at a time step $t$ the corresponding location was missing, I updated TSLV by increasing TSLV for all locations by one. I did not reset any value to one, because there was no current path available. However, I accounted for this later at the next available time step. At that time, I reset

TSLV to one for the entire path since the last available location. Because at least two time intervals had passed since the last location, I increased the buffer size for these longer steps by two.

The different types of spatial information varied greatly in their ranges, and therefore I transformed most of them before inclusion in the models. Because TSLF exhibited a very wide range of values, I log-transformed it. In contrast, prey densities were very low (all below one, often substantially), and I standardized them by subtracting the mean and dividing by the standard deviation (Lele, 2009). To calculate the mean and standard deviation for prey density, I only used values within the minimum convex polygon of the relevant movement trajectory. Measurements of distance to the territory boundary were left untransformed.

I tested the effect of the three different types of spatial information by including them in different combinations in the weighting function. I tested all three variables as additive terms, and additionally considered multiplicative terms between TSLV and prey density and between TSLV and distance to territory boundary.

### 3.2.3 Parameter estimation and model selection

I fitted individual models to data within a maximum-likelihood framework. The likelihood function could be composed from the step probabilities (3.1). If trajectories were available completely, that is if locations were observed for every time step, one could simply multiply the probabilities for all steps because of the Markov structure of the model. However, some observations were missing in all of the trajectories. For steps with missing start or end location, no step length can be determined. Therefore, I omitted these steps and conditioned the likelihood function on the first location of each segment of successively available locations. By using this method, I assumed that steps were missing randomly and not correlated to environmental variables.

I further omitted steps with length below a threshold. This was mainly due to the inclusion of the variable TSLV. Wolves express different behavioural modes, such as handling a kill, resting away from a kill site, or relocating (Franke *et al.*, 2006). When wolves are at a kill site or resting, movement steps tend to be short, and I assumed that TSLV was less likely to be important for these steps. Timing may play a role when wolves revisit a kill site, however, I was interested in understanding the effect of TSLV with respect to territorial defense and behavioural depression of prey. Therefore, I focused on movement steps that were more likely associated with relocating behaviour. Franke *et al.* (2006) used a hidden Markov model to identify the three major modes

"bedding", "localized activity" and "relocating". They found that the relocating mode was characterized by steps with length above 200 m, with the majority of steps between 500–2500 m. These distances were obtained using movement data with hourly location measurements and therefore were not immediately transferrable to my study with 2-hourly movement data. Roughly, steps at a rate of 500 meter per hour may be converted to 1000 meter per two hours, although it is known that measurements of travel distance are influenced by sampling rate, and the longer the time interval between location measurements the larger the risk of underestimating true travel distance (Pépin *et al.*, 2004; Rowcliffe *et al.*, 2012). Still, with these considerations it seemed appropriate to set the threshold for defining relocating steps at about 1000 m. If movement was straight in east-west or north-south direction, 1000m corresponded to about three cells in the discretized space. Another point to consider for the threshold was the use of the buffer for TSLV. If a step was within the buffer size of the last visited location, the step naturally ended at a location with TSLV=1. In contrast, if a step was larger than the buffer size, which was four cells, it could end at a location with TSL=1, especially when the animal backtracked. However, there was also a chance that the step ended in a location outside the buffer of the previous step with TSLV $> 1$. To avoid an artificial bias towards smaller values of TSLV for small steps, I defined the threshold to be five cells, corresponding roughly to 1500 m in continuous space. The likelihood function was then

$$L(\kappa, \sigma, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i}^{N} p(x_{t_i} | x_{t_i-1}, \kappa, \sigma, \boldsymbol{\beta}, \boldsymbol{\gamma}) \tag{3.4}$$

for all available steps from $x_{t_i-1}$ to $x_{t_i}$ with $\|x_{t_i-1} - x_{t_i}\| > 5$. Note that I omitted small steps after calculating TSLV for the entire time series. Therefore, steps used for the final analysis have appropriate values of TSLV, representing correct times based on the full path.

I obtained parameter estimates by optimizing the likelihood function using a Nelder-Mead algorithm implemented in R (R Core team, 2013). To find the global maximum, I optimized the likelihood function starting at various points in parameter space. From these results, I chose the parameters with the highest likelihood value and used them as starting point for the final optimization. I used an estimate of the Hessian matrix of the log-likelihood at the optimal parameter values to obtain standard errors of the maximum likelihood estimates. To find the best fitting model, I performed model selection via AIC.

I analyzed parameter estimates for their effects on movement decisions. Parameter estimates of the weighting function $w_t$ are best understood when considering their

implications for the weighting function itself. In the step probabilities (3.1), the weighting function $w_t(\boldsymbol{\mathcal{I}}_t(x))$ is a function of geographical space (sensu Aarts *et al.* (2011)), but via the spatial information $\boldsymbol{\mathcal{I}}_t(x)$ at each location. The weighting function can alternatively be viewed as a function $w_t(\boldsymbol{\mathcal{I}})$ over the different ranges of the three information variables TSLV, prey density, and distance to territory boundary. This corresponds to viewing the weighting function in environmental space sensu Aarts *et al.* (2011). This perspective revealed the effects of the information variables on the probability of selecting a location.

To test whether wolves showed rotational movement patterns per se, I inspected the resulting weighting function from a fit with the pure model that only included TSLV as information variable. Because the weighting function is modelled after a logistic function, I expected to see a switching behaviour with a generally positive relationship between TSLV and probability of selection: low probability of selection for small values of TSLV and high probability of selection for larger values of TSLV. Under the hypothesis that wolves counteract behavioural depression of prey, I expected models that included prey density to perform better than the null model. The effect of TSLV on the probability of selection should vary between locations with low and locations with high prey density. This could be either in form of shifting the switching curve, by strengthening the switching behaviour (steeper sigmoidal curve), or both. With respect to the distance variable, I generally expected to see a positive effect on the probability of selection, which would reflect a tendency to prefer locations further away from the boundary. If TSLV played a role for territorial defense mechanisms, I expected to see such an overall avoidance of boundary locations become weaker as TSLV increased.

## 3.3 Results

Wolf movement time series spanned approximately four months (December to April; wolves 284, 285) or five months (November to April; wolf 220), with successful fix rates ranging 82–91%. After selecting relocating steps, the number of available steps for analysis were 302 (wolf 220), 243 (wolf 284) and 264 (wolf 285). Prey distributions within territories varied, however, generally many of the high prey density areas were located in the outer regions of territories (Figure 3.1).

When considering only pure models that tested the three information variables independently, the pure time model (with TSLV) was always significantly better than the null model, with large differences in AIC ranging 33.8-63.4 for the three wolves

(Table 3.1). However, when inspecting the effect of TSLV on the weighting function $w_t(\boldsymbol{\mathcal{I}})$, that is the probability of selection in environmental space, I found it to be weak (Figure 3.2a-c). Although locations with very small value of TSLV had low probability of selection, the weighting function quickly reached one, indicating a nearly indifferent behaviour with respect to TSLV. The pure prey model was only in one case better than the null model, and the effective influence of prey density on the probability of selection was negligible. The pure prey model always had considerably higher AIC than the pure time model. The pure distance model was better than the null model for wolves 284 and 285. For wolf 284, locations closer to the territory boundary had a low probability of selection (Figure 3.2d). For wolf 285, this effect was only present for very small distances from the boundary (Figure 3.2e). Note that territory estimates were obtained in the same way for all wolves, but their geometry and the configuration of visited locations within the territory varied (Figure 3.1). Again, the pure distance model had higher AIC than the pure time model.

Model selection via AIC revealed two models that generally described the data best. These were the model that included both additive and multiplicative effects of TSLV and distance to territory boundary and the model that additionally had additive and multiplicative terms of prey density (Table 3.1). As mentioned above, all models with more than one variable must be considered as interactive models because of the non-linearity of the weighting function. However, additive and multiplicative terms in the weighting function act on different aspects of the resulting sigmoidal curve in environmental space, $w_t(\boldsymbol{\mathcal{I}})$. When considering the weighting function as a function of one of the variables, e.g. TSLV, additive terms of the other variables cause the sigmoidal curve to shift horizontally, whereas multiplicative terms influence the steepness of the curve. Thus, additive terms shift an individual's preferences, whereas multiplicative terms may strengthen, weaken, or even reverse, preferential behaviour.

Within the best models, the overall positive relationship between TSLV and probability of selection observed in the pure time model remained (Table 3.2). Although the additive selection coefficient $\beta_{\text{time}}$ for TSLV was negative for wolf 284, the overall coefficient $\beta_{\text{time}} + \gamma_{\text{dist}} \cdot \text{dist} + \gamma_{\text{prey}} \cdot \text{prey}$, considering all interactions, was only negative for very small values of distance. When this was the case, the entire combination of parameters resulted in a weighting function (as function of TSLV) close to zero, such that the negative slope had only a very weak effect. For most parameter combinations, the overall selection coefficient for TSLV was positive. The overall relationship between distance to territory boundary and probability of selection was positive as well, that is wolves generally preferred locations away from the boundary. When prey density

was included in the best model, the relationship between prey density and probability of selection was positive, that is locations with higher prey densities were preferred. For wolf 285, the multiplicative prey parameter $\gamma_{\text{prey}}$ was negative, however, because the additive prey parameter $\beta_{\text{prey}}$ was larger, the overall selection coefficient for prey, $\beta_{\text{prey}} + \gamma_{\text{prey}} \cdot \log(\text{TSLV})$, was always positive.

All best models included an interactive effect of TSLV and distance to territory boundary, and the effect on the weighting function $w_t(\boldsymbol{\mathcal{I}})$, given the probability of selection, was similar for all three wolves. Within central areas of the territory (corresponding to larger values of distance to boundary), the weighting function was nearly one for the entire range of TSLV except for very low values (Figure 3.3a, 3.4a, 3.5a). This reflected an almost constant probability of selection over most of the range of TSLV, and probability of selection was only smaller for very recently visited locations. In areas closer to the territory boundary (corresponding to lower values of distance), the effect of TSLV became more pronounced. The range of TSLV with low probability of selection extended, followed by a switch towards high probability of selection for locations with higher TSLV. When considering the weighting function as a function of distance to boundary, I observed a similar general pattern as presented in the pure distance model, boundary locations having a lower probability of selection than locations further away from the boundary (Figure 3.3b). However, in combination with TSLV, the avoidance of boundary locations became weaker as TSLV increased. After several days of absence, the probability of selection for locations close to the boundary increased to values considerable above zero (Figure 3.3b).

For two of the wolves, the best model was the most complex model with interaction terms for both distance to boundary and prey density. However, for wolf 284, the difference between this model and the model that only included TSLV and distance to boundary was small ($\Delta\text{AIC} = 1$). This was also reflected in an only weak effect of prey density on the weighting function. Close to the territory boundary, the weighting function (as function of TSLV) was slightly steeper for high prey densities compared to lower prey densities (Figure 3.4b,c). Thus, selection with respect to TSLV is stronger in areas with high prey density. This effect vanished as distance to boundary became larger (Figure 3.4d).

In contrast, for wolf 285, the model with prey density was significantly better than the model with distance only ($\Delta\text{AIC} = 5.1$). Again, the effect of prey density on the weighting function occurred mainly in the outer regions of the territory (Figure 3.5b,c,d). Here, prey density acted mainly through the additive term by shifting the sigmoidal curve of TSLV (Figure 3.5b,c) to the right. This resulted in a larger

range of TSLV values being avoided (low probability of selection). Or reworded, longer times of absence were required before the individual became likely to return. For example, fixing distance at the 0.1 quantile of attained values and prey density at the 0.95 quantile of attained values, the probability of selecting such a location started to exceed 0.5 when TSLV corresponded to approximately three days (Figure 3.5b). For areas further away from the boundary, the effect of TSLV vanished almost entirely (Figure 3.5d).

Parameter estimates of the movement kernel $k$ showed a systematic trend when comparing between estimates obtained from the null model and from the best model according to AIC. Both $\hat{\kappa}$ and $\hat{\sigma}$, the shape and scale of the Weibull distribution for step length, respectively, were smaller for the better fitting model with inclusion of spatial information (Table 3.3). This resulted in a lower mean of the resulting Weibull distribution in all cases (Table 3.3).

## 3.4 Discussion

I used a novel method to investigate how the spatio-temporal component of an individual's own travel history influences wolf movement decisions in relation to prey densities and territory effects. With this approach, I tested whether observed wolf movement patterns support the hypothesis that wolves follow a prudent prey-management strategy. This strategy aims at counteracting anti-predator behaviour in prey by avoiding too frequent and prolonged exposure of prey to predation risk. Instead, the predator spreads the risk temporally and spatially. Managing prey in this way should be most successful when a predator has exclusive access to prey, which could be one of the benefits of a territory (Charnov *et al.*, 1976; Davies & Houston, 1981). Although wolves inhabit territories in packs and not alone, pack member usually do not act as competitors in hunting (Mech & Boitani, 2006). Therefore, a prey-management strategy could be advantageous for wolves.

Additionally to conventional environmental variables, I included the variable time since last visit into a movement-based resource selection model. The variable TSLV encodes both spatial and temporal information about previous visits to locations. Therefore, keeping track of TSLV for movement decisions suggests spatial memory and a sense of time. Wolves may also use external cues to guide them, for example their own scent marks can convey information about their travel history (Peters & Mech, 1975). However, scent marks can only be utilized when encountered. TSLV changes permanently as result of movement, while I assumed that it influences movement decisions at

the same time. Resource selection must be viewed in relation to what is available to an individual, and a suitable definition of what constitutes available becomes crucial where availability of the variable TSLV changes permanently and interacts with movement. The movement-based model is suitable for this situations, as it assesses availability for every step based on the current position of the individual via the movement kernel $k$.

An additional advantage of the movement-based approach is that it allows simultaneous estimation of movement kernel and RSPF parameters. This is in contrast to the more traditional step selection approach that first estimates an empirical kernel from the data, and in a second step uses this to sample available control points (Fortin *et al.*, 2005; Forester *et al.*, 2009). The problem with this approach is that the empirical kernel is already likely influenced by selective behaviour, which can bias the following actual resource selection estimation (Forester *et al.*, 2009). Within the movement-based framework here, the step probability given in equation (3.1) is the joint probability of a step being available by movement capacities and being selected based on spatial information. The joint probability is equivalently formulated as probability of selecting a location based on its characteristics, conditional on the location being available by movement capacities. During statistical inference, all parameters are estimated simultaneously. The difference between the two approaches becomes also apparent when comparing parameter estimates of the kernel $k$ between the null and the best model. The estimates from the null model correspond to those that would be obtained for the empirical kernel in the traditional step-selection approach. I observed that these estimates were higher and resulted in a higher mean of the Weibull distribution for step length, compared to the estimates from the integrated approach (estimates from the best model). Thus, the null model overestimates general movement tendencies.

During model selection, models that included TSLV were always considerably better than models that did not include this variable. Therefore, I did not consider an interactive model of distance and prey only. The dominance of TSLV suggests that this variable played a role for movement decisions. However, TSLV only effectively influenced the weighting function, which is the RSPF, in areas close to the territory boundary, whereas further away from the boundary, the RSPF remained nearly constant. The fact that TSLV did not influence movement more generally was contrary to my expectations. If wolves had a general tendency to avoid recently visited areas for several days before returning, as suggested by the work of Jedrzejewski *et al.* (2001), I would have expected a sigmoidal shape of the weighting function, as function of log(TSLV), also within inner areas of the territory. It is possible that I did not observe a stronger effect of TSLV within the territory due to methodological choices. First, the

time interval between successive locations was two hours, and TSLV was measured at the same scale. It is possible that this time scale is unsuitable (possibly too small) to observe an effect of TSLV for rotational movement patterns. Second, an underlying assumption of the movement model was that TSLV influenced each relocating step. This reflects the premise that a wolf completely avoids an area recently visited. However, it may be necessary to release this strict assumption. For example, steps may cross recent paths to reach other important areas of the territory.

My observation that the effect of TSLV on the weighting function changed towards the territory boundary supports the hypothesis that TSLV played a role for territorial defense mechanisms. The results suggest that generally wolves preferred locations within some safety distance from the boundary. This is in line with observations that the risk for a wolf to be killed in an encounter with neighbouring wolves increases towards the boundary (Mech, 1994). The probability of selecting a location closer to the boundary became larger when TSLV increased. This suggests that when durations of absence from boundary locations were long enough, these locations switched from being unattractive to being attractive. This agrees with the need to renew scent marks or similar signs of presence (Peters & Mech, 1975).

Because of the territories' geometry, the amount of locations with certain distance values varies within a territory, and generally there are more boundary than central locations. One may suspect that this could lead to a false positive interaction effect of TSLV and distance. However, under the Null model movement is a simple random walk according to the kernel $k$, which is radially-symmetric and assumes a Weibull distribution for step lengths. For diffusive movement, we would expect a spatially uniform distribution at the steady state (Turchin, 1998). For my Null model, I expect a similar behaviour, which means that under the Null model, an animal should have the same probability of being in any location, whether close to the boundary or not. Therefore, and because my model is spatially-explicit, I consider the observed effect of distance to be a true effect. Additionally, my results from Chapter 2 demonstrate that misclassification of models is rare in this modelling framework. The models in Chapter 2 did not include the geometrical distance measure, however, future work could rule out the possibility of false positive results for effects of distance, using a simulation approach as that presented in Chapter 2.

For one of the wolves, the full model with prey density was significantly better than the less complex model without prey density. For this wolf with ID 285, prey density influenced the weighting function in a way that agreed with the hypothesis that wolves counteract behavioural depression of prey. For locations with low prey density, TSLV

had only a weak effect on the RSPF, suggesting an indifferent behaviour with respect to TSLV in areas of low prey density. With increasing prey density, the effect of TSLV became more pronounced and differentiated, with higher probability of selection for locations with high TSLV. This may be interpreted as a tendency of the wolf to avoid areas with high prey density for some time before returning. Effects of prey occurred primarily towards the territory boundary and less in more central areas of the territory. This could be due to the territories' geometry and prey distribution. Especially in the territory of wolf 285, higher prey densities occurred rather in regions with smaller distance to boundary. In general, in all three territories high prey densities occurred often towards the territory boundaries. This may reflect a strategy of prey to inhabit buffer zones between different wolf pack's territories (Lewis & Murray, 1993). To discern the effects of prey and territory boundary, one would have to investigate possible interactions between prey density and distance to territory boundary. Based on my results, this would be an important step for future analysis.

However, overall effects of prey density were less pronounced than expected. When considering only pure models, the model with prey density received the weakest support from the data. A reason for finding only few effects of prey density may lie in the data. First, is is possible that the measure for prey density used was not the best to reflect prey availability for wolves. Prey availability may not only be driven by prey abundance and prey habitat selection but also by landscape features that enhance encounter rates and prey vulnerability (Bergman *et al.*, 2006; Milakovic *et al.*, 2011). Second, several approximating steps, including interpolation via habitat models and conversion from fecal pellet counts to numbers of individuals, were required to obtain estimates of prey densities across the study area. All these steps likely introduced inaccuracies. Also, fecal pellet counts were based on pellets deposited over an entire winter. This may have identified areas with high prey accumulation at some point during the winter, but accumulation may have been temporary only. Prey densities thus obtained rather gave estimates of expected prey densities based on habitat features, and actual presence of prey may have deviated considerably. All these reasons likely decreased the ability to detect effects of prey densities on wolf movement decisions. Further analyses, considering more individual wolves, will help to reach stronger conclusions.

**Table 3.1.** Model selection results. Values of $\Delta\text{AIC} = \text{AIC} - \text{AIC}_{\text{min}}$ refer to the difference between each model and the best model with $\Delta\text{AIC} = 0$. Highlighted in grey are best models and in addition more parsimonious models with $\Delta\text{AIC} < 2$.

| | wolf 220 | | wolf 284 | | wolf 285 | |
|---|---|---|---|---|---|---|
| Model | AIC | ΔAIC | AIC | ΔAIC | AIC | ΔAIC |
| null | 4189.2 | 44.0 | 3621.4 | 105.2 | 3643.6 | 74.9 |
| TSLV | 4155.4 | 10.2 | 3558.0 | 41.8 | 3602.5 | 33.8 |
| dist | 4189.4 | 44.1 | 3577.3 | 61.1 | 3623.8 | 55.1 |
| prey | 4184.7 | 39.5 | 3625.4 | 109.2 | 3645.6 | 76.9 |
| TSLV+dist | 4153.6 | 8.4 | 3559.0 | 42.7 | 3599.9 | 31.3 |
| TSLV+prey | 4156.5 | 11.3 | 3554.9 | 38.7 | 3600.8 | 32.2 |
| dist+prey | 4190.2 | 44.9 | 3574.8 | 58.5 | 3618.4 | 49.7 |
| TSLV+dist+prey | 4155.6 | 10.4 | 3555.3 | 39.0 | 3599.5 | 30.9 |
| TSLV+dist+TSLV*dist | 4145.2 | 0 | 3517.2 | 1.0 | 3573.8 | 5.1 |
| TSLV+prey+TSLV*prey | 4156.5 | 11.3 | 3556.5 | 40.2 | 3600.1 | 31.4 |
| TSLV+dist+prey +TSLV*dist+TSLV*prey | 4148.4 | 3.2 | 3516.2 | 0 | 3568.7 | 0 |

**Table 3.2.** Parameter estimates with their standard errors for the weighting function $w_t$. Given are estimates from the best model fit; compare Table 3.1. For wolf 284, estimate are given for the two best models with $\Delta\text{AIC} = 0$ (full model) and $\Delta\text{AIC} = 1$ (distance interaction model).

| | | $\alpha$ | $\beta_{\text{time}}$ | $\beta_{\text{dist}}$ | $\beta_{\text{prey}}$ | $\gamma_{\text{dist}}$ | $\gamma_{\text{prey}}$ |
|---|---|---|---|---|---|---|---|
| wolf 220 | Est. | −2.33 | 0.441 | 0.072 | – | 0.206 | – |
| | SE | 0.59 | 0.151 | 0.050 | – | 0.083 | – |
| wolf 284 | Est. | −3.16 | −0.548 | 0.018 | 0.284 | 0.335 | 0.237 |
| $\Delta\text{AIC} = 0$ | SE | 0.85 | 0.292 | 0.055 | 0.328 | 0.108 | 0.174 |
| wolf 284 | Est. | −3.27 | −0.270 | 0.028 | – | 0.255 | – |
| $\Delta\text{AIC} = 1$ | SE | 0.75 | 0.345 | 0.057 | – | 0.029 | – |
| wolf 285 | Est. | −4.25 | 0.251 | 0.204 | 1.465 | 0.863 | −0.363 |
| | SE | 1.05 | 0.228 | 0.097 | 0.657 | 0.308 | 0.157 |

**Table 3.3.** Parameter estimates with their standard errors for the movement kernel $k$. The last column gives the mean of the resulting Weibull distribution for step length. The best model is the one with $\Delta\text{AIC} = 0$. For wolf 284, this is the model with both interactions; compare Table 3.1.

| | | $\hat{\kappa}$ | SE | $\hat{\sigma}$ | SE | Mean $\hat{\sigma}\,\Gamma(1+\frac{1}{\hat{\kappa}})$ |
|---|---|---|---|---|---|---|
| wolf 220 | null | 2.36 | 0.10 | 13.37 | 0.37 | 11.84 |
| | best | 2.01 | 0.12 | 12.22 | 0.12 | 10.83 |
| wolf 284 | null | 1.94 | 0.08 | 15.95 | 0.60 | 14.15 |
| | best | 1.55 | 0.10 | 13.96 | 0.10 | 12.55 |
| wolf 285 | null | 2.26 | 0.10 | 12.83 | 0.39 | 11.37 |
| | best | 1.87 | 0.12 | 11.45 | 0.12 | 10.17 |

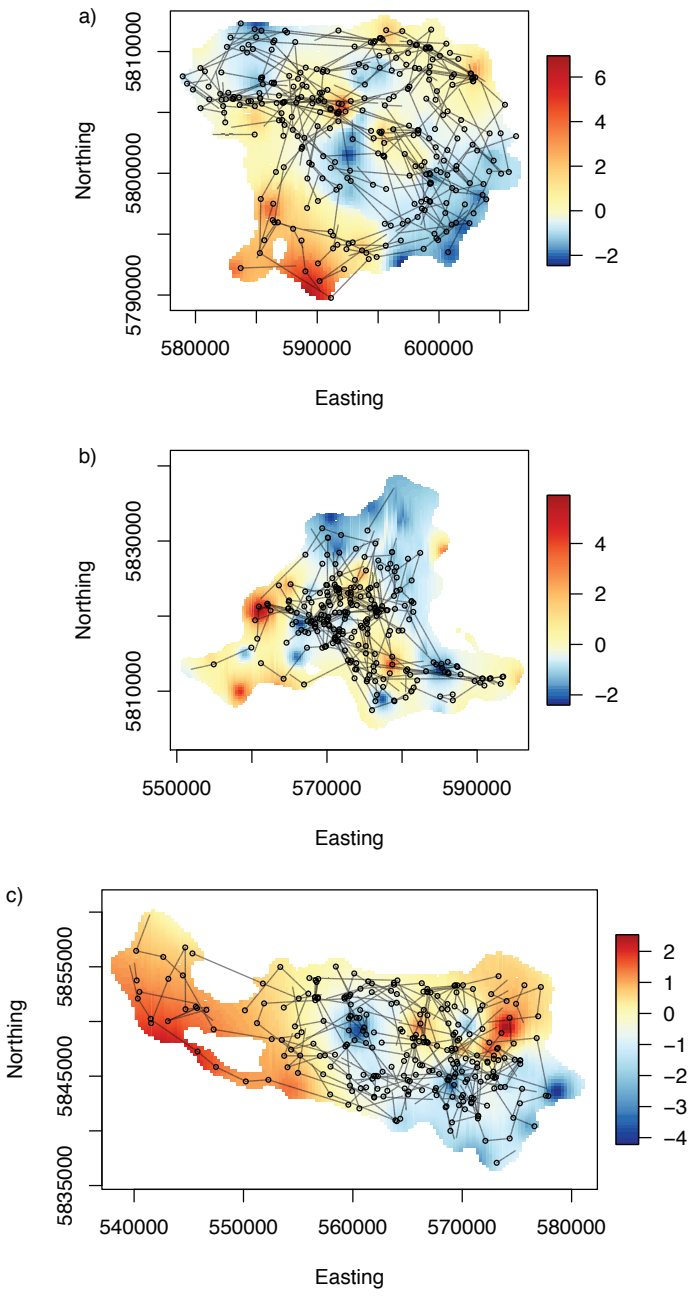



**Figure 3.1.** Wolf observations on prey density maps. Panels correspond to wolves with IDs **a)** 220, **b)** 284, and **c)** 285. Depicted are only relocating steps (lines), with end points of steps marked by small circles. Colours reflect the combined prey density measure, which was standardized for each wolf separately. The maps show the territories obtained from a Brownian bridge home range estimate, which was based on the full trajectory (consisting of relocating and all other steps).

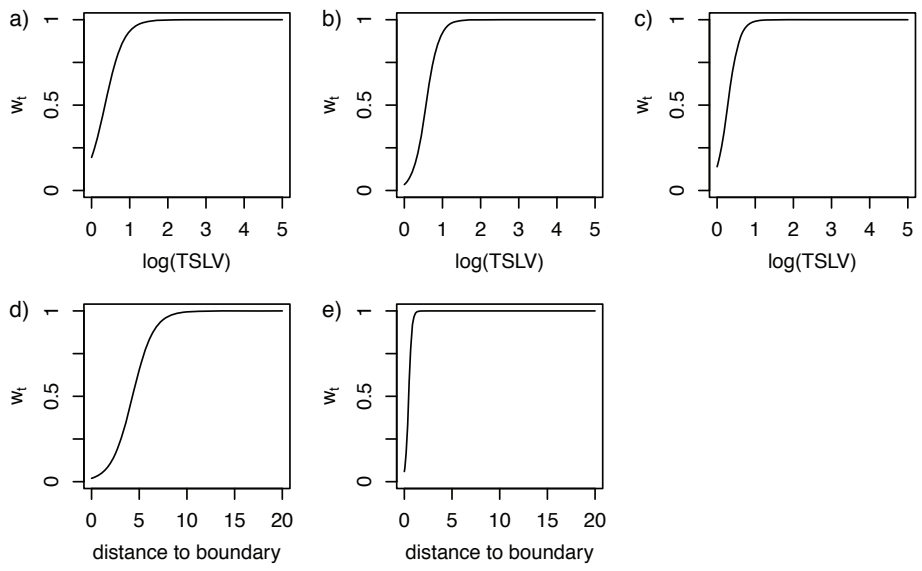



**Figure 3.2. Panels a), b), c)**: Weighting function $w_t(\boldsymbol{\mathcal{I}})$ as a function of TSLV (on a logarithmic scale as used in the model) using parameter estimates from a fit with the pure model that only included TSLV for **a)** wolf 220, **b)** wolf 284, and **c)** wolf 285. The probability of selection reaches values close to one quickly, for values of log(TSLV) that correspond to only very few time steps. **Panels d), e)**: Weighting function $w_t(\boldsymbol{\mathcal{I}})$ as a function of distance using parameter estimates from a fit with the pure distance model for the two wolves for which this fit was significantly better than the null model, which were **d)** wolf 284, **e)** wolf 285.

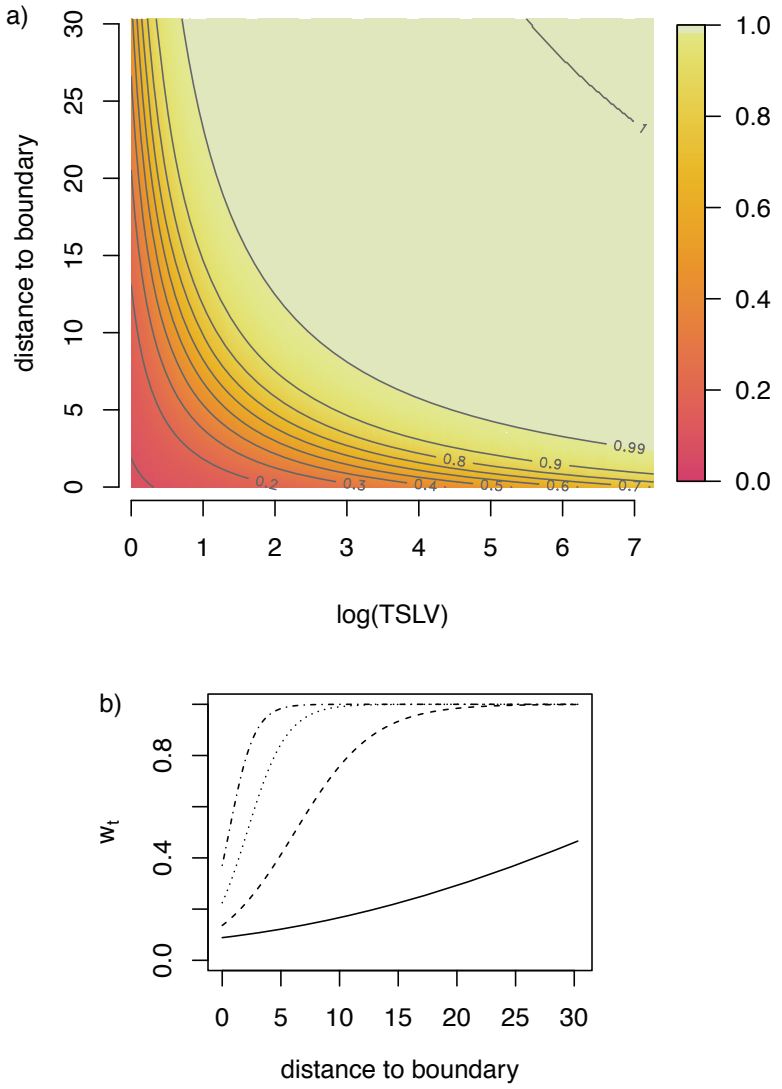



**Figure 3.3.** Wolf 220. **Panel a)**: Weighting function $w_t(\boldsymbol{\mathcal{I}})$ as a function of TSLV (on a logarithmic scale as used in the model) and distance to territory boundary, using parameter estimates from the best fitting model; compare Table 3.2. Note that here the weighting function is displayed over the ranges of the information variables instead of geographical space. **Panel b)**: Slice through the weighting function for four fixed values of TSLV: 1 (visited during last step, solid line), 3 (last visit few hours ago, dashed line), 12 (last visit one day ago, dotted line), 60 (last visit five days ago, dot-dashed line).

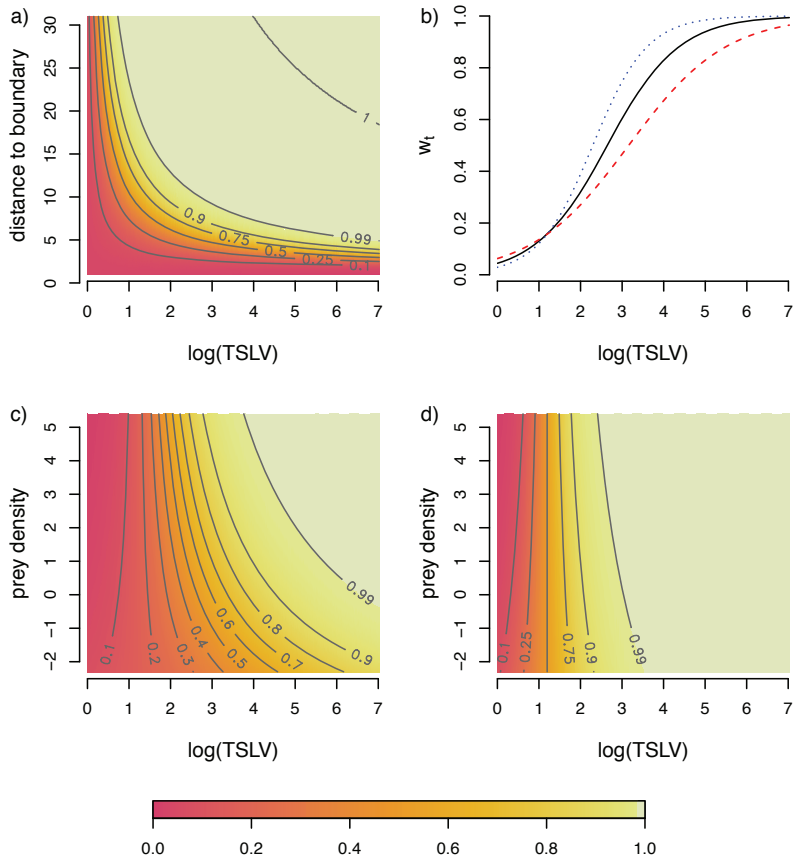



**Figure 3.4.** Wolf 284. Weighting function $w_t(\boldsymbol{\mathcal{I}})$ using parameter estimates from the best fitting model ($\Delta\text{AIC} = 0$) with both interaction terms; compare Table 3.2). Note that here the weighting function is displayed over the ranges of the information variables instead of geographical space. **Panel a)**: Prey density is fixed at zero, corresponding to the mean value over the territory. **Panel b)**: Slice through the previous plot at low distance 5.1 (0.05 quantile of attained values among relocating steps; black solid line). The same slice is depicted when prey density is fixed at -1.3 (0.05 quantile of attained values; red dashed line) and at 1.6 (0.95 quantile of attained values; blue dotted line). **Panel c)**: Distance to boundary fixed at 5.1 (0.05 quantile of attained values). **Panel d)**: Distance to boundary fixed at 9.1 (mean distance value over the territory).

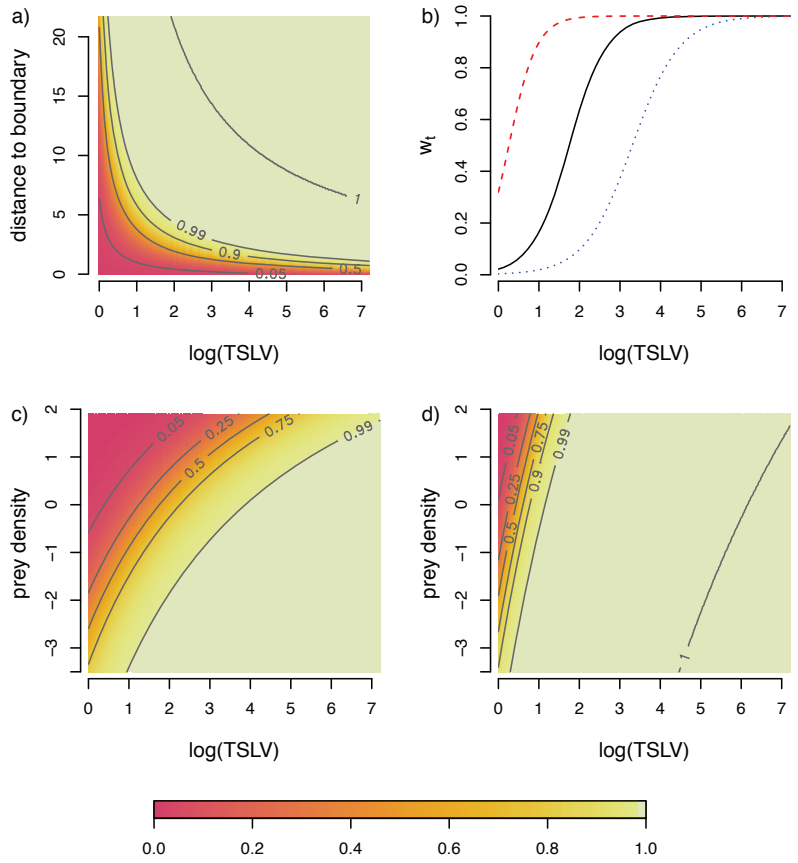



**Figure 3.5.** Wolf 285. Weighting function $w_t(\boldsymbol{\mathcal{I}})$, using parameter estimates from the best fitting model with both interaction terms; compare Table 3.2). Note that here the weighting function is displayed over the ranges of the information variables instead of geographical space. **Panel a)**: Prey density is fixed at zero, corresponding to the mean value over the territory. **Panel b)**: Slice through the previous plot at low distance 2.2 (0.1 quantile of attained values among relocating steps; black solid line). The same slice is depicted when prey density is fixed at -2.1 (0.05 quantile of attained values; red dashed line) and at 1.27 (0.95 quantile of attained values; blue dotted line). **Panel c)**: Distance to boundary fixed at 2.2 (0.1 quantile of attained values). **Panel d)**: Distance to boundary fixed at 7.14 (mean distance value over the territory).

# Chapter 4

# A framework for analyzing movement models' robustness against varying temporal discretization[1]

## 4.1 Introduction

To learn about animal movement behaviour, researchers across the world collect increasing amounts of data for many different species. Often, these data consist of time series of individual animals' locations, together with information about environmental aspects such as habitat and resource availability (Rhodes *et al.*, 2005; Mueller *et al.*, 2008), topography (Moorcroft & Lewis, 2006; Squires *et al.*, 2013), presence of predators (Latombe *et al.*, 2014), and anthropogenic features (Tracey *et al.*, 2010; Colchero *et al.*, 2010; Latham *et al.*, 2011). Additional biotelemetry data even can give us cues about activity patterns, for example via acceleration sensors (Brown *et al.*, 2012; Fröhlich *et al.*, 2012), or about the animal's internal state, for example via internal sensors that measure stomach temperature and indicate food ingestion (Austin *et al.*, 2006; Bestley *et al.*, 2008). Our ability to collect more extensive and detailed data provides an unprecedented opportunity to understand the mechanisms of movement behaviour in greater detail (Hebblewhite & Haydon, 2010; Bridge *et al.*, 2011).

A lot of work has been dedicated towards increasing the efficacy and accuracy of

[1] A version of the chapter has been submitted to Journal of Mathematical Biology as: Schlägel, U.E. & Lewis, M.A. A framework for analyzing movement models' robustness against varying temporal discretization.

tagging devices, and much improvement has been achieved. For example, modern GPS devices can measure locations with an accuracy of 10-20 m and less than 24 per cent data loss (Frair *et al.*, 2010; Rowcliffe *et al.*, 2012). At the same time, new methods for analyzing movement data are emerging, allowing us to ask more complex and in-depth questions about movement behaviour. Hierarchical modelling approaches including state-space models help us to split movement paths into behaviourally meaningful segments and to estimate activity budgets for behaviours such as resting, foraging and travelling (Morales *et al.*, 2004; Breed *et al.*, 2012; McClintock *et al.*, 2013). New models combine resource selection with cognitive processes in creative ways to study animals' navigational mechanisms and the role of spatial memory (Dalziel *et al.*, 2008; Mueller *et al.*, 2010; Avgar *et al.*, 2013). While much progress has been made, challenges remain (Börger *et al.*, 2008; Smouse *et al.*, 2010; Fagan *et al.*, 2013).

When tracking an animal's movement path, e.g. via GPS-based telemetry, locations are measured at discrete times, and the rate and regularity of measurements are critical features. From raw location data we can estimate classic movement characteristics such as mean square displacement, measures of directional persistence or tortuosity, and travel distance (Turchin, 1998; Codling *et al.*, 2008; Rowcliffe *et al.*, 2012). These quantities can vary largely when derived from movement data with different temporal resolutions (Ryan *et al.*, 2004; Codling & Hill, 2005; Nouvellet *et al.*, 2009; Rowcliffe *et al.*, 2012). A few studies used fine-scale movement data to empirically estimate correction factors to adjust measured travel distances according to the sampling interval (Pépin *et al.*, 2004; Ryan *et al.*, 2004). While this is a first approach to understand the influence of sampling interval on measured travel distance, it is unclear whether results can be generalized from these studies to other species and systems. Using simulations of unbiased and biased velocity jump processes, Codling & Hill (2005) estimated linear equations that relate the sampling rate of a discretized path to the angular deviation (or sinuosity) and apparent speed of the movement. They found that these relationships break down when the observed sinuosity becomes large, either due to large sinuosity in the underlying movement or a relatively large sampling time step. Generally, sampling a continuous path of an animal at discrete intervals can lead to various degrees of information loss (Turchin, 1998). When we fit a movement model to data to perform statistical inference, the temporal resolution of our sampling can affect parameter estimates and result in erroneous inference such as misclassified behavioural states (Breed *et al.*, 2011; Postlethwaite & Dennis, 2013).

One may think that the best solution to avoid undersampling and information loss is to take measurements at high rates to approximate a continuous path as best as

possible. However, this is often not feasible, because limited battery life of tagging devices gives rise to a tradeoff between sampling frequency and total sampling time span (Mills *et al.*, 2006; Breed *et al.*, 2011). In addition, oversampled movement paths can be problematic in data analysis, because they lead to strong and long-lasting autocorrelations and require the processing of very long time series. Also, very frequent fix attempts can reduce GPS transmitter efficiency (measured as total number of successful locations obtained during the deployment time) (Mills *et al.*, 2006), and noise can become very large compared to the actual signal, especially if animals are resting or moving slowly (Ryan *et al.*, 2004). It is therefore important to choose measurement rates appropriately to the behavioural scale of interest. Even if we decide about sampling rates with care, it remains a problem that results are often tied to the scale of particular studies. Generalizing or transferring results as well as comparison between different studies is limited (Tanferna *et al.*, 2012; Postlethwaite & Dennis, 2013).

Here, I introduce a new theoretical framework for analyzing movement models' robustness against varying resolutions of temporal discretization. I formally define robustness as a specific property of a model. Generally speaking, I consider a model to be robust if it can be applied validly to movement data with different temporal resolutions, thus allowing consistent statistical inference. While I do not require important movement characteristics expressed in model parameters to be the same across sampling rates, I ask for them to vary systematically in a way that allows translation of results between scales. Here, I present the new framework in terms of random walk models with independently and identically distributed steps. Although many contemporary movement models have surpassed these classical random walk models in complexity, I believe that my analyses here are important to understand the new concept of robustness and to put it in context with other established ideas in probability theory and movement ecology.

This chapter is organized as follows. In section 4.2, I describe the set-up of my study, after which I follow with two introductory example models that illustrate my framework. I then give formal definitions of two types of robustness that vary in their strength of requirements but also benefits. In section 4.3, I analyze robustness properties of one-dimensional models. I present models that are robust, suggest a way to construct robust models from non-robust models and relate robustness to the probabilistic concept of infinite divisibility. In section 4.4, I extend results about robustness to two-dimensional models, in particular models with radially symmetric step densities. My framework provides a new systematic, mathematically founded approach to analyze if, and how, sampling rate of movement paths influences inference.

## 4.2 The robustness framework

### 4.2.1 Temporal resolution of random walks

Random walks have a long history as animal movement models. As a basis for deriving partial-differential equation models for population distributions (Patlak, 1953; Skellam, 1951), for building simulation models for moving individuals (Kaiser, 1976; Jones, 1977), and for developing metrics that summarize movement characteristics (Kareiva & Shigesada, 1983), random walks have proved useful early on in movement ecology and have remained so. Although models have become more complex to include behavioural mechanisms such as territorial defense (Moorcroft & Lewis, 2006; Potts *et al.*, 2013) or resource selection (Mckenzie *et al.*, 2012; Potts *et al.*, 2014), to describe temporally switching behaviour (Morales *et al.*, 2004; McClintock *et al.*, 2013), and to account for stochasticity of the measurement process (Patterson *et al.*, 2008; Breed *et al.*, 2012), random walks remain at the root of many movement models (Börger *et al.*, 2008; Smouse *et al.*, 2010).

The classic random walk model for movement is a stochastic process $\{\boldsymbol{X}_t, t \in \mathbb{N}\}$, where the location $\boldsymbol{X}_t \in \mathbb{R}^2$ of an organism for each time index $t \in \mathbb{N}$ is given as a sum of independently identically distributed (i.i.d.) steps (Klenke, 2008). That is,

$$\boldsymbol{X}_t = \boldsymbol{x}_0 + \sum_{i=1}^{t} \boldsymbol{S}_i, \tag{4.1}$$

where $\boldsymbol{x}_0$ is the (fixed) start location of the movement path, and $\boldsymbol{S}_i$ is the vector, that is the step, between location $\boldsymbol{X}_{i-1}$ and $\boldsymbol{X}_i$. Note that here I use $\boldsymbol{S}$ to denote steps and $\boldsymbol{X}$ to denote locations, which are sums of steps. In the statistical literature, often $\boldsymbol{S}$ is used for sums of random variables. However, I have chosen notation according to the movement context. For a graphical clarification of notations, compare Figure 4.1. The random walk models an observed movement path, that is a series of locations $\underline{\boldsymbol{x}} = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, where $\boldsymbol{x}_t \in \mathbb{R}^2$, measured at regular time intervals.

As a convenient way for systematically studying varying temporal discretization of movement data, we can mimic different sampling rates of movement paths via subsampling. The $n$th subsample of $\boldsymbol{x}$ consists of every $n$th location, that is $\underline{\boldsymbol{x}}_n = \{\boldsymbol{x}_0, \boldsymbol{x}_n, \boldsymbol{x}_{2n}, \ldots\}$. As $n$ increases, the temporal resolution of the data becomes coarser. Note that $\underline{\boldsymbol{x}}_1 = \underline{\boldsymbol{x}}$ is the original time series. If $\underline{\boldsymbol{x}}$ is modelled by the process $\{\boldsymbol{X}_t, t \in \mathbb{N}\}$, then the subsample $\underline{\boldsymbol{x}}_2$, which consists of every second location of the original time series, is correctly described by the subprocess $\{\boldsymbol{X}_{2t}, t \in \mathbb{N}\}$. A priori, the subprocess

has different probability distributions than the original process, however, there is a simple relationship between the two processes. For the subprocess we have $\boldsymbol{X}_{2t} = \boldsymbol{x}_0 + \sum_{i=1}^{2t} \boldsymbol{S}_i = \boldsymbol{x}_0 + \sum_{i=1}^{t} \tilde{\boldsymbol{S}}_{i,2}$, for steps $\tilde{\boldsymbol{S}}_{i,2} = \boldsymbol{S}_{2i-1} + \boldsymbol{S}_{2i}$. Compare also Figure 4.1. More generally, for an arbitrary subprocess, we have

$$\boldsymbol{X}_{nt} = \boldsymbol{x}_0 + \sum_{i=1}^{nt} \boldsymbol{S}_i = \boldsymbol{x}_0 + \sum_{i=1}^{t} \tilde{\boldsymbol{S}}_{i,n}, \tag{4.2}$$

for the larger steps $\tilde{\boldsymbol{S}}_{i,n} = \sum_{j=0}^{n-1} \boldsymbol{S}_{ni-j}$. Therefore, the distribution of $\boldsymbol{X}_{nt}$ is based on steps that are themselves sums of steps of the original process. Recall that for a random walk with i.i.d. steps, all $\boldsymbol{S}_i$ have the same distribution, however, their sum may generally have a different distribution.

If a movement model were robust against changes in temporal resolution, the same model should be able to describe validly both a path $\underline{\boldsymbol{x}}$ and its subsample $\underline{\boldsymbol{x}}_n$. As I have described above, in a random walk model the distributions of the steps define the process. If the steps $\{\boldsymbol{S}_i, i \in \mathbb{N}\}$ and $\{\tilde{\boldsymbol{S}}_{i,n}, i \in \mathbb{N}\}$ for a range of subsampling indices $n \in \mathbb{N}$ can be described by the same probability model, with appropriate adjustment of model parameters, then I consider the model to be robust against varying temporal discretization within that range.

### 4.2.2 Two illustrative examples

I illustrate the concept of robustness with two simple examples. For simplicity, I consider one-dimensional models. First, for an example of a robust model, I assume that all steps $S_i$ have identical normal distribution, with zero mean and some positive variance $\sigma^2$, which I denote by $S_i \sim \text{Normal}(0, \sigma^2)$. A step density centred at zero means that steps to the right and left have the same probability. Because of the linearity of the normal distribution, it follows that the location $X_t$ is normally distributed as well, $X_t \sim \text{Normal}(x_0, t\sigma^2)$. The steps $\tilde{S}_{i,2}$ of the subsampled process $\{X_{2t}, t \in \mathbb{N}\}$ are sums of two normally distributed random variables, and therefore we have $\tilde{S}_{i,2} \sim \text{Normal}(0, 2\sigma^2)$ and $X_{2t} \sim \text{Normal}(x_0, 2t\sigma^2)$. Thus, the probability distributions that describe the original and the subsampled process are both normal with the same mean but different variances. However, the variances are related through a simple linear function. Therefore, we can make inference using the subsampled data and process and simply divide the estimated variance by 2 to obtain an estimate of the variance of the original process. Conversely, we can multiply the variance obtained using the original process by 2 to obtain the valid variance for the coarser process. This also

works analogously for $n > 2$. Because of this property, the random walk model with normally distributed steps is robust.

For a counterexample of robustness, I consider steps that have Laplace distribution, which is also termed double-exponential distribution. The Laplace distribution, similar as the normal distribution, is symmetric, however it is more peaked and has slightly heavier tails than the normal distribution. It commonly serves as a one-dimensional (or marginal, in two-dimensional models) redistribution kernel in models for dispersing organisms (Neubert *et al.*, 1995). I assume that steps $S_i$ are i.i.d. Laplace distributed with location parameter zero, i.e. the density is centred at zero, and scale parameter $\sigma$, that is $S_i \sim \text{Laplace}(0, \sigma)$. Consequently, the location $X_t$ is distributed as a sum of Laplace distributions. Sums of Laplace distributed random variables are not as simple or well-known as the previous Normal example. Still, we can employ characteristic functions to look into this case further. The characteristic function (ch.f.) of a random variable $X$ is defined by the expectation $\phi_X(u) = E(\mathrm{e}^{iuX})$. Characteristic functions uniquely define distributions, and they have the convenient property that summing independent random variables corresponds to multiplying their characteristic functions (Klenke, 2008). The ch.f. for the above step distribution is given by

$$\phi_{S_i}(u) = \frac{1}{1 + \sigma^2 u^2}. \tag{4.3}$$

The steps of the subsampled process, $\tilde{S}_{i,2} = S_{2i-1} + S_{2i}$, consequently have ch.f.

$$\phi_{\tilde{S}_{i,2}}(u) = \phi_{S_i}(u)^2 = \frac{1}{1 + (\sqrt{2}\sigma)^2 u^2 + \sigma^4 u^4}. \tag{4.4}$$

This function cannot be expressed as the characteristic function of any Laplace distribution, which would have to be of the form $\mathrm{e}^{i\mu t}(1+\sigma^2 u^2)^{-1}$ for some location parameter $\mu \in \mathbb{R}$ and scale $\sigma > 0$. With a bit more work, one can also compare probability density functions. While the step $S_i$ has the Laplace density

$$f_{S_i}(s) = \frac{1}{2\sigma} \mathrm{e}^{-\frac{|s|}{\sigma}}, \tag{4.5}$$

the density of the sum of two such random variables is given by

$$f_{\tilde{S}_{i,2}}(s) = \frac{1}{4\sigma^2} \mathrm{e}^{-\frac{|s|}{\sigma}} (\sigma + |s|), \tag{4.6}$$

which we cannot write in form of $f_{S_i}(s)$ by transforming the parameters. It follows that the step distribution for the subsampled process does not belong to the same family of

distributions as the original process, namely the Laplace family. This means that if we fit the original model with Laplace distributed steps to both $\boldsymbol{x}$ and $\boldsymbol{x}_2$, the resulting parameter estimates are not truly comparable. If, however, instead we fit a different model to $\boldsymbol{x}_2$ that uses densities (4.6), the parameter $\sigma$ describes the same quantity as in the original model. Therefore, the model that has Laplace distributed steps is not robust against varying temporal resolution; but see section 4.3.2.

### 4.2.3 Formal definition of robustness

I now define robustness formally. As demonstrated above, the step distribution plays an essential role for the robustness of random walk models. In the Laplace example, the characteristic function has been a convenient tool to analyze step distributions of random walk models. Therefore, I use them in my definitions of robustness. For a two-dimensional model, the ch.f. of a step $\boldsymbol{S}_i \in \mathbb{R}^2$ is $\phi(\boldsymbol{u}) = E(\mathrm{e}^{i\boldsymbol{u}\cdot\boldsymbol{S}_i})$ for $\boldsymbol{u} \in \mathbb{R}^2$, where $\cdot$ denotes the scalar product of vectors.

**Definition 4.1** (Semi-robustness)**.** Let $\phi(\boldsymbol{u};\boldsymbol{\theta})$ be the characteristic function of the i.i.d. steps in a random walk movement model, where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is the vector of model parameters. The movement model is *semi-robust (with respect to distribution class)* if for every $n \in \mathbb{N}$ there exists a function $g_n : \boldsymbol{\Theta} \to \boldsymbol{\Theta}$ such that

$$\phi(\boldsymbol{u};\boldsymbol{\theta})^n = \phi(\boldsymbol{u}; g_n(\boldsymbol{\theta})). \tag{4.7}$$

As mentioned before, summing independent random variables (here, steps in a random walk) corresponds to multiplying their respective characteristic functions. In the random walk models, steps are identically distributed. Therefore, the LHS of equation (4.7) is the ch.f. of the sum of $n$ steps and therefore defines the distribution for the steps $\tilde{\boldsymbol{S}}_{i,n}$ of the model for the $n$th subsample. The RHS of the equation is the ch.f. of the steps $S_i$, however with transformed parameters. Therefore, semi-robustness requires that subsamples of the random walk are defined by the same step distribution up to a known parameter transformation. The parameter transformation $g_n$ is an important part of the definition, because it allows to scale up model parameters to a coarser discretization. Say, a model represents a temporal discretization $\tau$, that is $\tau$ is the time interval between two locations. If the model is semi-robust, it is also valid for any discretization $n\tau$, $n \in \mathbb{N}$, with parameter $g_n(\boldsymbol{\theta})$.

If we want to be able to compare results of studies that use different temporal resolutions for their models more generally, we also need be able to translate parameters

downwards, that is to a finer discretization. The following definition characterizes models that can be scaled both upwards and downwards.

**Definition 4.2** (Robustness)**.** A semi-robust movement model is *robust (with respect to distribution class)* if the function $g_n$ in Definition 4.1 is bijective, that is both one-to-one and onto.

This definition allows scaling upwards just as before. Additionally, we can translate the parameter $\boldsymbol{\theta}$ to a finer scale $\frac{1}{n}\tau$. The surjectivity of $g_n$ guarantees that there exits an inverse image $\boldsymbol{\psi} = g_n^{-1}(\boldsymbol{\theta}) \in \boldsymbol{\Theta}$, which is unique by injectivity. Therefore, $\phi(\boldsymbol{u}; \boldsymbol{\psi})$ defines a valid characteristic function, and by property (4.7) we have

$$\phi(\boldsymbol{u}; \boldsymbol{\psi})^n = \phi(\boldsymbol{u}; g_n(\boldsymbol{\psi})) = \phi(\boldsymbol{u}; \boldsymbol{\theta}). \tag{4.8}$$

This means that there is a valid sub-model for the discretization $\frac{1}{n}\tau$ with parameter vector $\boldsymbol{\psi}$.

The introductory example model with Normally distributed steps is robust. The transformation for the only model parameter, the standard deviation $\sigma$, is $g_n(\sigma) = \sqrt{n}\sigma$. The second example with Laplace distributed steps is neither robust nor semi-robust since property (4.7) is not met. In section 4.3.2, I will show that it is possible to embed the Laplace model within an extension so as to make it robust.

## 4.3 One-dimensional models

In the following, I look further into the question which random walk models are robust. First, I focus on one-dimensional models, that is random walks on the real line. These models can play a role in situations where movement is naturally limited, e.g. movement within a stream or along a river bank. Also, univariate step distributions arise as marginals of two-dimensional movement or dispersal kernels; see section 4.4.2. After presenting classes of robust models, I describe the relationship of robustness with the probabilistic concept of infinite divisibility. With this, I hope to deepen the reader's understanding of robustness and to set robustness apart from other concepts.

### 4.3.1 Robust random walk models

To find robust models, I look for steps with probability distributions that are closed under summation. Such a property ensures semi-robustness, which is a necessary condition for robustness. Whether a semi-robust model is also robust depends largely on

the parameter space for which the step distribution is well-defined. A straightforward example is given by distributions, whose ch.f. is a power of some function and the power is a model parameter.

**Theorem 4.1.** *Consider a one-dimensional random walk movement model with i.i.d. steps that have characteristic function of the form* $\phi(u;\boldsymbol{\theta}) = h(u;\boldsymbol{\theta}_1)^{\theta_2}$ *for some function*
$h : \mathbb{R} \times \boldsymbol{\Theta}_1 \to \mathbb{C}$ *and model parameters* $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \theta_2) \in \boldsymbol{\Theta}_1 \times \Theta_2$. *If the parameter space is such that* $n\Theta_2 = \{n\theta_2; \theta_2 \in \Theta_2\} \subset \Theta_2$ *for all* $n \in \mathbb{N}$, *the model is semi-robust. If additionally* $\frac{1}{n}\Theta_2 \subset \Theta_2$ *for all* $n \in \mathbb{N}$, *then the model is robust.*

*Proof.* Define the parameter transformation as $g_n(\boldsymbol{\theta}) = g_n(\boldsymbol{\theta}_1, \theta_2) = (\boldsymbol{\theta}_1, n\theta_2) \in \boldsymbol{\Theta}_2 \times \Theta_2$. Then, trivially, we have $\phi(u;\boldsymbol{\theta})^n = h(u;\boldsymbol{\theta}_1)^{n\theta_2} = \phi(u; g_n(\boldsymbol{\theta}))$, and semi-robustness follows. Let $\frac{\theta_2}{n} \in \Theta_2$ for all $n \in \mathbb{N}$ and all $\theta_2 \in \Theta_2$. Then for each $\boldsymbol{\theta}$ there is a unique inverse image $g_n^{-1}(\boldsymbol{\theta}) = (\boldsymbol{\theta}_1, \frac{\theta_2}{n})$, which lies within the valid parameter range. Therefore, the model is robust. □

For such models, the parameter transformation only affects the parameter that constitutes the power in the ch.f. For example, consider i.i.d. steps $S_i$ that have Gamma distribution with shape $\kappa > 0$ and scale $\sigma > 0$. Note that the support of the Gamma density is only the positive real line, so movement steps are always into the same direction (to the right). The Gamma distribution has the well-known property that a sum of independent Gamma random variables, all having the same scale parameter, again has a Gamma distribution (Casella & Berger, 2002). The ch.f. of the Gamma distribution is $\phi(u;\kappa,\sigma) = (1 - \sigma i u)^{-\kappa}$. Therefore, we directly obtain $\phi(u;\kappa,\sigma)^n = (1 - \sigma i u)^{-n\kappa} = \phi(u; n\kappa, \sigma)$. Hence, the summation affects the shape parameter, and we have $g_n(\kappa,\sigma) = (n\kappa,\sigma)$. Because the Gamma distribution is defined for all positive shapes $\kappa \in \mathbb{R}^+$, the transformation $g_n$ is invertible, and it follows that steps with Gamma distribution lead to robust models.

The chi-squared distribution is a special case of the Gamma distribution for a scale $\sigma = 2$ and shape $\kappa = \frac{k}{2}$ for degrees of freedom $k \in \mathbb{N}$. The ch.f. is

$$\phi(u;k) = (1 - 2iu)^{\frac{k}{2}}. \tag{4.9}$$

The $n$th power of $\phi$ is still a ch.f. of a chi-squared distribution with degrees of freedom $nk \in \mathbb{N}$, and therefore a model with chi-squared steps is semi-robust. However, for an arbitrary $k \in \mathbb{N}$, the fraction $\frac{k}{n}$ is a rational but not necessarily a natural number. Thus, the second condition of Theorem 4.1 is not satisfied. For more examples of

distributions that meet the conditions of Theorem 4.1, see Table 4.1. Note that there are also discrete distributions that belong to the group of distributions described in the theorem (e.g. the binomial, Poisson and negative-bionomial).

Another class of distributions that are suitable as step distributions for robust models is given by the family of stable distributions (Samorodnitsky, 1994; Nolan, 1997; Klenke, 2008). The stable distributions comprise a four-parameter family of distributions, which I denote by $\mathcal{S}(\alpha, \beta, \sigma, \mu)$, with index of stability $0 < \alpha \leq 2$, skewness $-1 \leq \beta \leq 1$, scale $\sigma > 0$ and location $\mu \in \mathbb{R}$. Note that the scale parameter does not necessarily correspond to the variance of the distribution, which is in fact infinite for most stable distributions. Only for certain values of $\alpha$ and $\beta$, stable distributions have closed-form density functions. However, for any parameter values, we can define a stable distribution uniquely by its characteristic function. There are multiple ways to parameterize stable distributions, which differ slightly in the interpretation of the parameters $\sigma$ and $\mu$. Here I use the form of the ch.f. provided in Nolan (1997),

$$\phi(u; \alpha, \beta, \sigma, \mu) = \begin{cases} \exp\big[i\mu u - \sigma^\alpha |u|^\alpha \big(1 - i\beta \tan(\tfrac{\pi\alpha}{2}) \operatorname{sign}(u)\big)\big], & \alpha \neq 1 \\ \exp\big[i\mu u - \sigma|u|\big(1 + i\beta \operatorname{sign}(u) \ln|u|\big)\big], & \alpha = 1. \end{cases} \tag{4.10}$$

The most famous example of a stable distribution is the normal distribution for $\alpha = 2$. Using the above parameterization of the stable distribution, the mean and variance of the normal distribution are $\mu$ and $2\sigma^2$, respectively. For $\alpha = 2$, the term including the parameter $\beta$ vanishes. For $\alpha = 1$ and $\beta = 0$, the Cauchy distribution is another well-known case, for which a closed-form density is known. While the normal and Cauchy distribution are symmetric, the Lévy distribution for $\alpha = \frac{1}{2}$ and $\beta = 1$ is an example of a stable distribution with skewed density function (Samorodnitsky, 1994).

**Theorem 4.2.** *A one-dimensional random walk movement model with i.i.d. steps is robust if steps are distributed according to the stable law $\mathcal{S}(\alpha, \beta, \sigma, \mu)$, i.e. have characteristic function* (4.10)*.*

*Proof.* We can easily verify that the ch.f. of the stable distribution satisfies property (4.7). We have

$$\phi(u; \alpha, \beta, \sigma, \mu)^n = \begin{cases} \exp\big[i(n\mu)u - (n^{\frac{1}{\alpha}}\sigma)^\alpha |u|^\alpha \big(1 - i\beta \tan(\tfrac{\pi\alpha}{2}) \operatorname{sign}(u)\big)\big], & \alpha \neq 1 \\ \exp\big[i(n\mu)u - (n\sigma)|u|\big(1 + i\beta \operatorname{sign}(u) \ln|u|\big)\big], & \alpha = 1. \end{cases} \tag{4.11}$$

Therefore, choose $g_n(\alpha, \beta, \sigma, \mu) = (\alpha, \beta, n^{\frac{1}{\alpha}}\sigma, n\mu)$. It is easy to see that $g_n$ is a bijection of the parameter space, leaving $\alpha$ and $\beta$ unchanged and being monotone on $\mathbb{R}^+ \times \mathbb{R}$ in

the last two arguments. Therefore, stable steps distributions lead to robust models. □

I have just demonstrated that if we sum $n$ steps, each having stable distribution $S_i \sim \mathcal{S}(\alpha, \beta, \sigma, \mu)$, the sum is again stable according to

$$\tilde{S}_{i,n} \sim \mathcal{S}(\alpha, \beta, n^{\frac{1}{\alpha}}\sigma, n\mu). \tag{4.12}$$

In fact, stable distributions are a family of distributions that have been constructed to have this special summation property. Equivalently to defining a stable distribution by its characteristic function, we can also say a random variable $S$ has stable distribution if the sum of independent copies of $S$ is a scaled and shifted version of $S$, that is we have

$$\sum_{i=1}^{n} S \stackrel{d}{=} a_n S + b_n \tag{4.13}$$

for some $a_n > 0, b_n \in \mathbb{R}$, where $\stackrel{d}{=}$ stands for equality in distribution (Samorodnitsky, 1994; Kotz *et al.*, 2001). In fact, the only choice for $a_n$ is $a_n = n^{\frac{1}{\alpha}}$ (Samorodnitsky, 1994). Because the location $X_t$ is a sum of steps, $X_t = x_0 + \sum_{i=1}^{t} S_i$, the distribution of the location $X_t$ is also stable,

$$X_t \sim \mathcal{S}(\alpha, \beta, t^{\frac{1}{\alpha}}\sigma, x_0 + t\mu), \tag{4.14}$$

for any $t \in \mathbb{N}$. The analogue holds for the locations of the subsampled process $\{X_{nt}, t \in \mathbb{N}\}$,

$$X_{nt} \sim \mathcal{S}(\alpha, \beta, n^{\frac{1}{\alpha}} t^{\frac{1}{\alpha}}\sigma, x_0 + nt\mu). \tag{4.15}$$

The parameters $\alpha$ and $\beta$ remain unchanged under summation. The parameter $\beta$ determines skewness, with $\beta = 0$ corresponding to a symmetric density, and therefore a stable distribution $\mathcal{S}(\alpha, 0, \cdot, \cdot)$ is also termed $\alpha$-symmetric stable distribution.

A special case is given by models that have starting location $x_0 = 0$ and step distribution $S \sim \mathcal{S}(\alpha, 0, \sigma, 0)$. These specific stable distributions are symmetric with centre at zero, and they lead to

$$X_t \sim \mathcal{S}(\alpha, 0, t^{\frac{1}{\alpha}}\sigma, 0). \tag{4.16}$$

Such a random walk is self-similar because

$$X_{nt} \stackrel{d}{=} n^{\frac{1}{\alpha}} X_t. \tag{4.17}$$

Also, the probability density function of the step distribution, $p_S(s)$, is related to the density of the summed steps $\tilde{S}_{i,n}$ via a scaling property (Klafter *et al.*, 1995),

$$p_{\tilde{S}_{i,n}}(s) = \frac{1}{n^{\frac{1}{\alpha}}} p_{S_i}\left(\frac{s}{n^{\frac{1}{\alpha}}}\right). \tag{4.18}$$

This specific random walk is called a Lévy flight (Klafter *et al.*, 1995). Note that this (original) definition of a Lévy flight is different from a Lévy walk. A Lévy walk is based on a continuous-time random walk, describing the movement of an organism at constant speed between reorientation events (Klafter *et al.*, 1995). In the movement literature, the two terms are often used interchangeably (Reynolds & Rhodes, 2009; James *et al.*, 2011). Note that because of the different assumptions data are processed slightly different in a Lévy walk analysis, where usually steps (as I have defined them here) are combined as long as directional changes between them remain under a certain threshold (Plank *et al.*, 2013).

Although stable step distributions are predestined to lead to robust models, robustness is a more general concept. In terms of the characteristic function $\phi$ of $S$, the summation property (4.13) is $\phi(u)^n = \mathrm{e}^{iub_n}\phi(a_n u)$, or simply $\phi(u)^n \propto \phi(a_n u)$. In comparison, the robustness property (4.7) is a weaker condition. It means that the sum of $n$ i.i.d. steps has the same distribution as a single step up to adjusted parameter values according to a known function $g_n$. In the case where steps have stable distribution, the function $g_n$ affects the scale and location parameter of a distribution. However, distributions may have other types of parameters that can be affected. For example, in the above case of Gamma distributed steps, summation of steps results in a modified shape parameter. In contrast, scaling a Gamma distributed random variable by a constant $c$ leads to a Gamma distribution with same shape $\kappa$ but adjusted scale $c\sigma$. Therefore, the Gamma distribution is not stable, and the resulting random walk does not exhibit self-similarity. However, the random walk model with Gamma distributed steps in robust.

### 4.3.2 Robust model extensions

As I have shown in Theorem 4.1, a step distribution having ch.f. that is the power of some function leads to a semi-robust or robust model, depending on the definition of the parameter space. This leads to the idea that we can obtain robustness by embedding a distribution into a larger family of distributions by adding an additional power parameter to the ch.f. Starting with a ch.f. $\phi(u;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we can augment the

model parameters by $k \in \mathbb{N}$, that is we define a new parameter vector $\bar{\boldsymbol{\theta}} = (\boldsymbol{\theta}, k) \in \boldsymbol{\Theta} \times \mathbb{N}$. We can then define a new distribution via the ch.f. $\psi(u; \bar{\boldsymbol{\theta}}) = \phi(u; \boldsymbol{\theta})^k$. For $k \in \mathbb{N}$ we know that $\psi$ is again a ch.f., because by construction it is the ch.f. of a distribution of a sum of $k$ independent random variables. Because $nk \in \mathbb{N}$ for all $n, k \in \mathbb{N}$, and according to Theorem 4.1, a step distribution with ch.f. $\psi(u; \bar{\boldsymbol{\theta}})$, where $k$ is simply one of the model parameters, leads to a semi-robust random walk model with $g_n(\boldsymbol{\theta}, k) = (\boldsymbol{\theta}, nk)$. To go a step further and construct a robust model, the range of the parameter $k$ would need to include positive rational numbers. However, for $k \neq \mathbb{N}$, we have in general no guarantee that $\psi$ is again the ch.f. of a distribution

As an illustration of these ideas, consider the Laplace distribution. The Laplace distribution with mean zero and scale parameter $\sigma > 0$ has ch.f.

$$\phi(u; \sigma) = \frac{1}{1 + \sigma^2 u^2}. \tag{4.19}$$

I have shown above that a model with Laplace distributed steps is not robust. However, we can define a new family of distributions via the ch.f.

$$\psi(u; \sigma, k) = \frac{1}{(1 + \sigma^2 u^2)^k}, \tag{4.20}$$

where $k \in \mathbb{N}$. This is the ch.f. of the sum of $k$ independent Laplace random variables and therefore a valid ch.f. Using this distribution for steps and treating $k$ as a regular model parameter leads to a semi-robust model. In this particular case of the extended Laplace distribution, the function $\psi$ in equation (4.20) is also a valid ch.f. for any non-negative, real $k \in \mathbb{R}_{\geq 0}$ (Kotz *et al.*, 2001). It corresponds to a generalized asymmetric Laplace distribution with location parameter zero and symmetry parameter being zero (and hence being symmetric); compare also Table 4.1. This generalized Laplace distribution is not widely known, however, it has found several applications. In particular, it has been used in financial modelling, where it is also known as variance gamma model (Madan & Seneta, 1990; Seneta, 2004). A movement model with step distribution determined by the ch.f. (4.20) for $k \in \mathbb{R}_{\geq 0}$ is robust.

For applications in which likelihood functions play an important role, e.g. for statistical inference, a remaining question is whether we can find the corresponding probability density function for the ch.f. $\psi$. In principle, the probability density function of a distribution can be calculated as inverse Fourier transform of the characteristic function (Klenke, 2008). Alternatively, for $k \in \mathbb{N}$, the density of $\psi$ can be obtained as the convolution of the $k$ single step densities. Both methods can be difficult or may

not result in a closed-form density. However, for the above example of the generalized asymmetric Laplace distribution, a density function is available in terms of a Bessel function (Kotz *et al.*, 2001). In the symmetric case with location parameter zero, the density that corresponds to the ch.f. $\phi$ in equation (4.20) is

$$f(x) = \frac{1}{\sqrt{\pi}(k-1)!} 2^{-k+\frac{1}{2}} \sigma^{-k-\frac{1}{2}} |x|^{k-\frac{1}{2}} K_{k-\frac{1}{2}}\Big(\frac{|x|}{\sigma}\Big), \tag{4.21}$$

where $K_{k-\frac{1}{2}}(x)$ is a modified Bessel function of the third kind. This formula is valid for any $k \geq 0$. For the case where $k$ is restricted to the non-negative integers, $k \in \mathbb{N}_0$, the Bessel function $K_{k-\frac{1}{2}}(x)$ has a closed form (Kotz *et al.*, 2001, Appendix C), and we can alternatively write

$$f(x) = \frac{\mathrm{e}^{-\frac{|x|}{\sigma}}}{\sigma\,(k-1)!\,2^k} \sum_{j=0}^{k-1} \frac{(k-1+j)!}{(k-1-j)!\,j!} \cdot \frac{\big(\frac{|x|}{\sigma}\big)^{k-1-j}}{2^j}. \tag{4.22}$$

This density function can be used for likelihood-based inference, and both $\sigma$ and $k$ can be estimated simultaneously. While the new parameter $k$ may take the role of a nuisance parameter, it allows the distribution to be more flexible. Most importantly, estimates of $\sigma$ become comparable across different temporal resolutions; see Figure 4.2.

Robust model extensions highlight that (semi-) robustness is defined with respect to a distribution class. If we want to use a random walk model validly across a range of resolutions, we need to preserve the statistical distribution that defines movement steps. If a model is not robust *a priori*, we can embed its step distribution in a larger class of distributions, within which the model become robust with respect to this larger class.

### 4.3.3 Robustness and infinite divisibility

Robustness is related to the probabilistic concept of infinite divisibility. Roughly speaking, a distribution is infinitely divisible if it can be expressed as the distribution of a sum of independent random variables. More precisely, in terms of the characteristic function $\phi$ of a distribution, $\phi$ is infinitely divisible if for every $n \in \mathbb{N}$, there exists another ch.f. $\phi_n$ such that $\phi(u) = \phi_n(u)^n$ (Steutel & Van Harn, 2004; Klenke, 2008). It is important that $\phi_n$ is not just any function but the ch.f. of a random variable. An example of an infinitely divisible distribution is the normal distribution with mean

$\mu \in \mathbb{R}$ and standard deviation $\sigma \in \mathbb{R}^+$. Its ch.f. is

$$\phi(u; \mu, \sigma) = \mathrm{e}^{i\mu u - \frac{1}{2}\sigma^2 u^2}. \tag{4.23}$$

We can choose $\phi_n(u) = \phi(u; \frac{\mu}{n}, \frac{\sigma}{\sqrt{n}})$, which is the ch.f. of another normal distribution with mean $\frac{\mu}{n} \in \mathbb{R}$ and standard deviation $\frac{\sigma}{\sqrt{n}} \in \mathbb{R}^+$. In general, many of the commonly known distributions are infinitely invisible.

Both concepts, robustness and infinite divisibility, are linked to sums of random variables. However, the two concepts are not the same. The Laplace distribution is infinitely divisible, however, the factors of the ch.f. do not again correspond to Laplace distributions. Instead, the ch.f. of a zero-mean Laplace distribution can be factored as follows (Kotz *et al.*, 2001),

$$\phi(u) = \frac{1}{1+\sigma^2 u^2} = \left[ \left( \frac{1}{1 - i\sigma u} \right)^{\frac{1}{n}} \left( \frac{1}{1 + i\sigma u} \right)^{\frac{1}{n}} \right]^n = \phi_n(u)^n. \tag{4.24}$$

Each factor $\phi_n$ is the ch.f. of a random variable that is a difference between two i.i.d. Gamma random variables (Kotz *et al.*, 2001). This second example highlights that a distribution can be infinitely divisible but, as a step distribution, does not lead to a robust model. This is due to the fact that infinite divisibility only requires the existence of random variables that sum up to the variable in question. Robustness additionally requires that the summands belong to the same distribution as the original, only with modified parameter values. On the other hand, the converse is true and every robust random walk model of the form considered here must have infinitely divisible step distribution.

**Theorem 4.3.** *Let $S_i$, $i \in \mathbb{N}$, denote the i.i.d. steps of a random walk movement model. If the step distribution leads to a robust model, then $S_i$ is infinitely divisible. The converse is not true, that is not every infinitely divisible step distribution leads to a robust model.*

*Proof.* Let $\phi(u; \boldsymbol{\theta})$, with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, be the ch.f. of a single step $S_i$. Let $n \in \mathbb{N}$, and let $g_n$ be the parameter transformation given by robustness. Because $g_n$ is bijective, we can define a unique $\boldsymbol{\psi} := g_n^{-1}(\boldsymbol{\theta}) \in \boldsymbol{\Theta}$ and choose $\phi_n(u) := \phi(u; \boldsymbol{\psi})$. It follows that

$$\phi_n(u)^n = \phi(u; \boldsymbol{\psi})^n = \phi(u; g_n(\boldsymbol{\psi})) = \phi(u; g_n(g_n^{-1}(\boldsymbol{\theta}))) = \phi(u; \boldsymbol{\theta}), \tag{4.25}$$

which shows infinite divisibility. As a counterexample for the converse, I have demonstrated above that the Laplace distribution is infinitely divisible, but a model with

Laplace distributed steps is not robust. □

In the preceding proof, the bijectivity, and in particular the surjectivity, of the transform $g_n$ is crucial for the existence of $\phi_n$. Therefore, semi-robustness is not a sufficient criterion for infinite divisibility. Consider the Binomial distribution, which is discrete and not typically used as distribution for movement steps. Still, it serves as a counterexample for a distribution that is not infinitely divisible, yet as step distribution leads to a semi-robust model. For its ch.f. is $\phi(u; p, n) = (1 - p + p\mathrm{e}^{iu})^n$ for $p \in [0, 1]$ and $n \in \mathbb{N}$, and therefore meets the first, but not the second, condition of Theorem 4.1. On the other hand, as a distribution with bounded support, namely $\{k \in \mathbb{N}, k \leq n\}$, it is not infinitely divisible (Steutel & Van Harn, 2004).

Even if a model both is semi-robust and has infinitely divisible step distribution, it does not follow that it is robust. Consider the model with chi-squared distributed steps. As I have illustrated in section 4.3.1, this model is semi-robust but not robust. Still, the chi-squared distribution is a special case of the Gamma distribution and thus infinitely divisible; compare Table 4.1. The reason for the model not being robust is that the summands, which a chi-squared random variable can be decomposed into, are generally Gamma and not again chi-squared random variables. This examples highlights that the definition of the model parameter space is an important consideration for robustness. If instead of the chi-squared distribution, which is embedded in the Gamma distribution, we directly use the Gamma distribution as probability model for steps, we immediately obtain a robust model.

I have used the same idea in section 4.3.2 to embed the Laplace distribution within the more comprehensive generalized Laplace distribution. Although the Laplace distribution is infinitely divisible, Laplace distributed steps lead to neither a robust nor a semi-robust model. If we define the extension described by the ch.f. (4.20) for $k \in \mathbb{N}$, we obtain a random walk model that is semi-robust. If we go even further and define the extension for $k \in \mathbb{R}_{\geq 0}$, the resulting model is robust.

From these considerations I conclude that robust random walk models lie within the intersection of semi-robust models and models with infinitely divisible steps, however, they do not constitute the entire intersection; see Figure 4.3.

## 4.4 Two-dimensional models

### 4.4.1 Radially-symmetric step densities

Many applications of movement modelling, especially those that consider movement of terrestrial animals, require the use of two-dimensional models. We then often describe steps by their length and bearing, which corresponds to describing a vector in polar coordinates. Accordingly, instead of assigning a distribution to steps directly, we compose step distributions from a step length distribution and a distribution for the bearing. From these, we can obtain a step distribution (i.e. a distribution for the two-dimensional vector) by taking into account the transformation from polar coordinate formulation to euclidean space. Let $\boldsymbol{S} = \binom{S_1}{S_2} \in \mathbb{R}^2$ be the two-dimensional step. Then let

$$R = \sqrt{S_1^2 + S_2^2} \tag{4.26}$$

denote the step length, which is the radius of the vector in polar coordinates, and let $p_R(r)$ be the step length distribution. Let $p_B(\beta)$ denote the distribution of the bearing. Note that, in accordance with common usage, I use capital letters for random variables and small letters for their realizations. The transformation between the two coordinate systems is given by $S_1 = R\cos B$ and $S_2 = R\sin B$. Assuming that step length and bearing distributions are independent, we obtain as step density

$$p_{S_1,S_2}(s_1, s_2) = \frac{1}{\sqrt{s_1^2 + s_2^2}} \cdot p_R\big(\sqrt{s_1^2 + s_2^2}\big) \cdot p_B(\mathrm{Arg}(s_1 + is_2)), \tag{4.27}$$

where $\mathrm{Arg}(\cdot)$ denotes the principle argument of a complex number. The factor $(\sqrt{s_1^2 + s_2^2})^{-1}$ is due to the coordinate system transformation.

A classic assumption for simple random walk models is that bearings have uniform distribution on the interval $(-\pi, \pi]$, which leads to a bearing density $p_B(\beta) = \frac{1}{2\pi}$ (Bartumeus *et al.*, 2005; Smouse *et al.*, 2010; James *et al.*, 2011). If movement is assumed to be persistent in its direction, we may release this assumption and use a von Mises or wrapped Cauchy distribution instead (Morales *et al.*, 2004; Codling *et al.*, 2008; McClintock *et al.*, 2013). Here, I only consider models with uniform bearing distribution and therefore step densities of the form

$$p_{S_1,S_2}(s_1, s_2) = \frac{1}{2\pi\sqrt{s_1^2 + s_2^2}} \cdot p_R\Big(\sqrt{s_1^2 + s_2^2}\Big). \tag{4.28}$$

This density function is radially symmetric, and we can simply write

$$p_{S_1,S_2}(r) = \frac{1}{2\pi r}\, p_R(r) \tag{4.29}$$

for $r = \sqrt{s_1^2 + s_2^2}$. Note that I distinguish the radius density $p_R$ and radially-symmetric step density $p_{S_1,S_2}$ via the subscript.

The radial symmetry of the density (4.28) enables us to compute its ch.f. via a Hankel transform. The Hankel transform of order $\nu$ of a function $f(r)$ for $r \geq 0$ is given by the integral

$$\mathcal{H}_\nu\{f\}(u) = \int_0^\infty r f(r) J_\nu(ru)\, \mathrm{d}r, \tag{4.30}$$

where $J_\nu$ denotes the Bessel function of the first kind of order $\nu$ (Piessens, 2000). The ch.f. of a two-dimensional random vector with joint density (4.29) can be calculated as

$$\phi(\boldsymbol{u}) = 2\pi\, \mathcal{H}_0\{p_{S_1,S_2}\}(\|\boldsymbol{u}\|). \tag{4.31}$$

For details about the calculation, see Appendix C. Because $\phi$ only depends on the norm of $\boldsymbol{u}$ and hence is radially symmetric as well, I also use the notation $\phi(\|\boldsymbol{u}\|)$. Hankel transforms have been computed for a variety of functions, which in the following simplifies my analysis of characteristic functions for two-dimensional step distributions.

### 4.4.2 Robust two-dimensional models

In the following, I look for robustness among two-dimensional models. A direct way of verifying robustness is via the two-dimensional ch.f. according to Definition 4.1 or 4.2. In the case where the step distribution has a radially symmetric density function, it depends on the step distribution $p_{S_1,S_2}(r)$ whether or not the Hankel transform in formula (4.31) can be readily obtained. Alternatively, we can draw on previous results for one-dimensional models.

**Theorem 4.4.** *Consider a random walk model with two-dimensional steps that have radially symmetric density of the form* (4.28)*. If the marginal step distribution, given by the density* $p_{S_1}(s_1) = \int_{-\infty}^{\infty} p_{S_1,S_2}(s_1, s_2)\, ds_2$*, leads to a (semi-) robust model in one dimension, then the two-dimensional model is (semi-) robust as well.*

*Proof.* Let $\phi(\|\boldsymbol{u}\|; \boldsymbol{\theta})$ denote the radially symmetric ch.f. of the two-dimensional steps,

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ are the model parameters. The ch.f. of the marginal density is

$$\int_{-\infty}^{\infty} \mathrm{e}^{iu_1 s_1} p_{S_1}(s_1)\,\mathrm{d}s_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{iu_1 s_1} p_{S_1,S_2}(s_1,s_2)\,\mathrm{d}s_1\,\mathrm{d}s_2$$
$$= \phi(\|\boldsymbol{u}\|;\boldsymbol{\theta})\big|_{u_2=0} = \phi(|u_1|;\boldsymbol{\theta}) =: \phi_{S_i}(u_1;\boldsymbol{\theta}) \quad (4.32)$$

Let $n \in \mathbb{N}$. By assumption, there exists a function $g_n$ such that

$$\phi_{S_i}(u_1;\boldsymbol{\theta})^n = \phi_{S_i}(u_1; g_n(\boldsymbol{\theta})). \quad (4.33)$$

Because of the previous calculations, we also have $\phi(|u_1|;\boldsymbol{\theta})^n = \phi(|u_1|; g_n(\boldsymbol{\theta}))$. Replacing $u_1$ by $\|\boldsymbol{u}\|$ yields semi-robustness for the two-dimensional model. The parameter transformation is the same for the two-dimensional and the marginal one-dimensional model, therefore if the one-dimensional model is robust, the same holds for the two-dimensional one. □

With this result, I have established a link between one- and two-dimensional models. The correspondence of the characteristic functions given in equation (4.32) allows to compute the ch.f. of the radially symmetric two-dimensional model directly from the ch.f. of the one-dimensional model, and vice versa. Whether it is easier to obtain the two-dimensional ch.f. via the Hankel transform of the two-dimensional density or via the ch.f. of the one-dimensional marginal depends on which of the two densities is available. Conversely, from the two-dimensional ch.f. we can calculate the two-dimensional, radially symmetric step density via an inverse Hankel transform, which is self-reciprocal.

To demonstrate these relationships, I now present three example models and their robustness properties.

*Example* 4.1 (*Exponential step length).* A common step length distribution used for movement analyses is the exponential distribution (Smouse *et al.*, 2010; DeMars *et al.*, 2013), which has density $p_R(r) = \frac{1}{\lambda}\mathrm{e}^{-\frac{r}{\lambda}}$. Using this in the step density (4.29), we obtain

$$p_{S_1,S_2}(r) = \frac{1}{2\pi\lambda r}\mathrm{e}^{-\frac{r}{\lambda}}. \quad (4.34)$$

The Hankel transform of order zero is given by $\mathcal{H}_0\{p_{S_1,S_2}\}(u) = \frac{1}{2\pi}(1+\lambda^2u^2)^{-\frac{1}{2}}$ (Piessens, 2000), and thus the ch.f. is

$$\phi(\|\boldsymbol{u}\|;\lambda) = \frac{1}{\sqrt{1+\lambda^2\|u\|^2}} \quad (4.35)$$

From this, we can already see that the exponential step length model, where $\lambda > 0$ is the only parameter, is neither robust nor semi-robust. The marginal of the density $p_{S_1,S_2}$ is

$$p_{S_1}(s_1) = \frac{1}{\lambda\pi} K_0\left(\frac{|s_1|}{\lambda}\right), \tag{4.36}$$

where $K_0$ denotes the Bessel function of the second kind of order zero. The ch.f. of the marginal is $\phi(u;\lambda) = (1+\lambda^2 u^2)^{-\frac{1}{2}}$. This is in fact the ch.f. of a generalized (asymmetric) Laplace distribution with location and asymmetry parameters being zero, and with scale $\lambda$ and power $k = \frac{1}{2}$, which was shown before to be robust; compare section 4.3.2 and Table 4.1. Therefore, if we embed the exponential step length model in an extended model with step characteristic function

$$\phi(\|\boldsymbol{u}\|;\lambda,k) = \frac{1}{(1+\lambda^2\|\boldsymbol{u}\|^2)^k}, \tag{4.37}$$

for $k \in \mathbb{R}_{\geq 0}$, we obtain a robust model with the two parameters $\lambda > 0$ and $k \in \mathbb{R}_{\geq 0}$. In the one-dimensional case, we could obtain the density from the ch.f. (4.20) via an inverse Fourier transform. However, the two-dimensional step density needs to be computed from (4.37) as an inverse Hankel transform. Unfortunately, the inverse Hankel transform of order zero of the function (4.37) is not readily available.

*Example* 4.2 (*Heavy-tailed step length distribution).* In one dimension, I have shown that stable step distributions lead to robust models. An example of a stable distribution with closed-form density function is the Cauchy distribution. According to Theorem 4.4, we can therefore construct a robust two-dimensional model by finding the two-dimensional density (4.29) that has the Cauchy density as marginal. We can achieve this via the identity of characteristic functions established in (4.32). From the ch.f. of the Cauchy distribution, we obtain a corresponding two-dimensional ch.f. $\phi(\|\boldsymbol{u}\|;\sigma) = \mathrm{e}^{-\sigma\|\boldsymbol{u}\|}$. Applying an inverse Hankel transform according to the identity (4.31), we obtain (Piessens, 2000)

$$p_{S_1,S_2}(r) = \frac{\sigma}{2\pi(\sigma^2+r^2)^{\frac{3}{2}}}. \tag{4.38}$$

According to (4.29), this results in a step length density for the two-dimensional models as follows

$$p_R(r) = \frac{\sigma\, r}{(\sigma^2+r^2)^{\frac{3}{2}}}. \tag{4.39}$$

The variance does not exist for this density, and the density is heavy-tailed. More

precisely, the tail is of order $\frac{1}{r^2}$, that is we have

$$\frac{\sigma\, r}{(\sigma^2 + r^2)^{\frac{3}{2}}} = \mathcal{O}\Big(\frac{1}{r^2}\Big), \tag{4.40}$$

as $r \to \infty$. I will later show that the step distribution in this example is a special case of a bivariate stable distribution. Because of its relation with the univariate Cauchy, it is also known as bivariate (isotropic) Cauchy (Achim & Kuruoglu, 2005; Nadarajah & Kotz, 2007).

*Example* 4.3 (*Normally distributed steps, or Rayleigh step length distribution).* The normal distribution is another special case of a stable distribution. Its radially symmetric two-dimensional version with mean zero is the bivariate normal distribution with covariance matrix $\left(\begin{smallmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{smallmatrix}\right)$, having density

$$p_{S_1,S_2}(r) = \frac{1}{2\pi\sigma^2} \mathrm{e}^{-\frac{r^2}{2\sigma^2}}, \tag{4.41}$$

and ch.f. $\phi(\|\boldsymbol{u}\|; \sigma) = \mathrm{e}^{-\frac{1}{2}\sigma\|\boldsymbol{u}\|^2}$. The corresponding step length distribution with density

$$p_R(r) = \frac{r}{\sigma^2} \mathrm{e}^{-\frac{r^2}{2\sigma^2}} \tag{4.42}$$

is a Rayleigh distribution with scale parameter $\sigma > 0$. As can be easily seen from the ch.f. and also via Theorem 4.4, this model with normally distributed steps is robust.

In the latter two examples, the step distributions are special cases of bivariate stable distributions. Analogously to one-dimension, an $\alpha$-stable random vector $\boldsymbol{S} \in \mathbb{R}^2$, $0 < \alpha \leq 0$, by construction has the property

$$\sum_{i=1}^{n} \boldsymbol{S} \overset{d}{=} n^{\frac{1}{\alpha}} \boldsymbol{S} + \boldsymbol{b}_n \tag{4.43}$$

for some $\boldsymbol{b}_n \in \mathbb{R}^2$ (Samorodnitsky, 1994). If $\boldsymbol{S}$ is elliptically contoured, its ch.f. is

$$E\big(\mathrm{e}^{i\boldsymbol{u}\cdot\boldsymbol{S}}\big) = \exp\Big(i\boldsymbol{u}\cdot\boldsymbol{\mu} - (\boldsymbol{u}^T\Sigma\boldsymbol{u})^{\frac{\alpha}{2}}\Big) \tag{4.44}$$

for location vector $\boldsymbol{\mu} \in \mathbb{R}^2$ and positive definite shape matrix $\Sigma$ (Nolan, 2013). From this form of the ch.f., it can easily be seen that the $n$th power is again the ch.f. of an $\alpha$-stable random vector, with location vector $n\boldsymbol{\mu}$ and shape matrix $n^{\frac{2}{\alpha}}\Sigma$. Therefore, we immediately obtain the following theorem.

**Theorem 4.5.** *A two-dimensional random walk model with elliptically contoured steps* $\boldsymbol{S}$ *that have bivariate stable distribution, i.e. have ch.f.* (4.44)*, is robust.* □

The bivariate normal distribution with mean $\boldsymbol{\mu}$ and a general covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \tag{4.45}$$

where $\rho$ is the correlation, is an example of such a bivariate stable distribution for $\alpha = 2$. If $\boldsymbol{S}$ is not only elliptically contoured but even radially symmetric with location $\boldsymbol{\mu} = 0$, the ch.f. (4.44) simplifies to

$$\phi(\|\boldsymbol{u}\|; \alpha, \sigma) = \mathrm{e}^{-\sigma^\alpha \|\boldsymbol{u}\|^\alpha}, \tag{4.46}$$

for $\sigma > 0$. Example 4.2 and 4.3 were special cases for $\alpha = 1$ and $\alpha = 2$, respectively.

As in the univariate case, closed-form expressions for the density of bivariate stable distributions are available only for some special cases, e.g. the cases presented in Examples 4.2 and 4.3. However, there are results that allow simulation of random variables with stable distributions. For an $\alpha$-stable, radially symmetric stable random vector $\boldsymbol{S}$, we have

$$\boldsymbol{S} \overset{d}{=} \sqrt{AT}\,\boldsymbol{U}, \tag{4.47}$$

where $\boldsymbol{U}$ is a random vector with uniform distribution on the unit circle, $T$ is a chi-squared random variable with degrees of freedom 2, and $A$ is a univariate stable random variable, $A \sim \mathcal{S}(\frac{\alpha}{2}, 1, 2\sigma^2(\cos\frac{\pi\alpha}{4})^{\frac{2}{\alpha}}, 0)$ (Nolan, 2013). Thus, to obtain a bivariate stable random vector, it is enough to generate a univariate stable random variable. For this, an algorithm is available (Weron, 1996), which has been implemented in the R package stabledist (Wuertz *et al.*, 2013). This package also provides numerical calculations of density and cumulative distribution functions.

## 4.5 Discussion

I presented a new way of classifying movement models according to their robustness against changes in temporal discretization. After providing a formal definition for movement model's robustness, I explored which models have this property. My definition emphasizes a systematic transformation of model parameters between temporal resolutions. This ensures that, if a model is robust, we can fit it to movement data with varying time intervals between locations, and we know how to translate model

parameters between resolutions. Conversely, if a model is not robust, any results we derive from it are tied to its particular temporal scale, and therefore comparison of studies is difficult if they use data obtained at different sampling rates.

The question of robustness may already arise at a fundamental level when interfacing models with data. If a model is not robust, then it cannot use data measured at a particular temporal scale of sampling to make inferences about movement behaviours at shorter and longer time scales. This is of particular concern in movement ecology, because time scales for sampling animal movement data are often subject to logistical constraints. For example, limited battery life of GPS devices often leads to lower sampling rates in favour of longer total time spans. The time scales thus imposed on data may be very different than those for behavioural or ecological questions about movement. If a model is not robust, then it may still be semi-robust, which means that inference can be made at longer but not at shorter time scales. Because the conditions for robustness and semi-robustness are rather stringent, it appears that many existing movement models may fail in this regard.

Previous approaches to the problem have been empirical or based on simulations. Several studies used fine-scale movement data with sampling intervals of a few minutes (Pépin *et al.*, 2004; Postlethwaite & Dennis, 2013) or even a few seconds (Ryan *et al.*, 2004). These data were subsampled at various scales to obtain empirical relationships between the sampling interval and measured or inferred movement parameters. Such investigations have demonstrated that the sampling interval can have a strong effect on results from movement analyses. However, each of these studies is based on a specific species within a particular environment, and it is unclear whether the obtained relationships and possible correction factors can be transferred to other species and systems. Also, fine-scale movement data is rarely available, and therefore we need a more general method that relates sampling rates to movement parameters. As an alternative to using very fine movement data, Codling & Hill (2005) and Rowcliffe *et al.* (2012) simulated synthetic data from movement models and subsampled these. All of these approaches remain empirical and constitute case studies for particular data sets, parameters, and models. In contrast, my newly introduced framework is mathematically rigorous and provides analytical methods and results to determine whether a model is robust against changes in sampling intervals.

In my investigations, I found that robustness is a rather strong condition for a model. This is in line with previous empirical results that highlight the sensitivity of movement characteristics to the sampling interval. For one-dimensional models, I encountered two groups of step distributions that lead to robustness. First, Theorem 4.1

established robustness for distributions whose characteristic function is a simple power function. Among the common distributions, those that meet this condition have support $\mathbb{R}_{\geq 0}$ and therefore only allow steps into positive direction. Such models can be applicable in situations where movement experiences external forces, such as movement within strong water currents (Luschi *et al.*, 2003) or wind-driven dispersal (de la Giroday *et al.*, 2011). The second class of step distributions that lead to robust models are the stable distributions. If steps have $\alpha$-symmetric stable distribution $\mathcal{S}(\alpha, 0, \sigma, 0)$, the resulting random walk is a Lévy flight (Klafter *et al.*, 1995). In my analysis of two-dimensional models, I found few robust models. It is, again, mainly the stable distributions that constitute examples of robust models. Stable distributions are fat-tailed and do not have second (and higher) moments, the normal distribution for $\alpha = 2$ being the only exception. To circumvent this problem, the related Lévy walk was introduced (Klafter *et al.*, 1995).

On the one hand, Lévy walks may be attractive models because of their scale-invariance and optimality in certain foraging situations (Viswanathan *et al.*, 1999). On the other hand, it is highly debated whether Lévy walks are suitable models for movement and fit empirical data (Benhamou, 2007; James *et al.*, 2011; Edwards, 2011; Pyke, 2015). A major point of controversy arises from the difficulty of inferring processes from patterns. Although movement patterns may fit Lévy walks, the underlying process does not necessarily need to be a Lévy walk but may be due to more complex behaviour (Benhamou, 2007; Plank *et al.*, 2013). The debate further concerns statistical methods that are used to detect Lévy walk behaviour in data (White *et al.*, 2008; Edwards, 2011). Even the application of Lévy walks within optimal foraging theory as Lévy foraging hypothesis has been met with scepticism (Pyke, 2015).

In this chapter, I was merely interested in the question if there are models that are robust against changes in sampling rates, and which models these are. Because of the complexity of the issue, I concentrated on investigating this question for basic random walks. Even among these rather simple models, I found few that are robust. This foreshadows that robustness may be rare, if existent at all, among more complex models. But many contemporary models include forms of behavioural mechanisms beyond the mere random walk and will likely continue to become more sophisticated (Holyoak *et al.*, 2008; Smouse *et al.*, 2010; Fagan *et al.*, 2013). This means that most analyses of movement data to date are restricted to the measurement time scale of each study, limiting extrapolation of results and comparison between studies. With the analysis here, I have proposed a new fundament for analyzing movement models' robustness against varying sampling rates. An important next step will be to extend the

framework to more complex models that include temporal or spatial heterogeneities.

I suggest that a path for further investigation lies in continuing to look for robust extensions of models. The results I have presented about robust random walk models need not be exhaustive. In Example 4.1, I have shown that the two-dimensional model with exponential step length is not robust but can be extended to a robust model with an additional parameter (the power of the ch.f.). This would be similar to the one-dimensional example in section 4.3.2, where I demonstrated a robust extension to the Laplace model. If we would use this extended model and during statistical inference estimate the power parameter together with all other parameters, we would be using a robust model. Such an extension is, in theory, also possible for other models. Unfortunately, although we may be able to straightforwardly construct the characteristic function of such a robust extension, it can be more difficult to derive the bivariate step density. To overcome this problem, one could fall back on numerical solutions. For example, one could solve the inverse Hankel transform of equation (4.37) numerically and embed this process into an inferential optimization routine such as likelihood maximization or an MCMC algorithm.

Another possibility is to somewhat release the strict conditions of robustness. In my definition presented here, the parameter transformation $g_n$ is a key element. It assures that we can systematically translate results about model parameters between analyses using different sampling rates. The works by Pépin *et al.* (2004) and Codling & Hill (2005) tried to establish such a transformation empirically for some specific movement quantities. The relationships they found were able to correct for different sampling rates to some extent. This suggests that although my robustness is a strong condition on a model, there may be models that are approximately robust within certain ranges of sampling intervals. Often, we do not wish to compare data analyses with widely varying sampling intervals. When we analyze movement, we always have to be aware of the behavioural scale of interest, as the behavioural processes may vary across scales (Yackulic *et al.*, 2011; Fleming *et al.*, 2014). However, it may be a reasonable goal to compare movement analyses with sampling intervals within a range of a couple hours or so. Within this range, an approximate type of robustness may be sufficient.

**Table 4.1.** List of univariate distributions, which as random walk step distributions lead to semi-robust or robust models. The table indicates which of these distributions are also robust or infinitely divisible.

| Distribution | Ch.f. $\phi(u)$ | Parameter space | Semi-rob. | Rob. | Inf. div. |
|---|---|---|---|---|---|
| Continuous distributions with support $\mathbb{R}$ | | | | | |
| Normal | $\mathrm{e}^{iu\mu-\frac{1}{2}\sigma^2u^2}$ | $\mu\in\mathbb{R}$, $\sigma\in\mathbb{R}^+$ | ✓ | ✓ | ✓ |
| Cauchy | $\mathrm{e}^{iu\mu-\sigma\lvert u\rvert}$ | $\mu\in\mathbb{R}$, $\sigma\in\mathbb{R}^+$ | ✓ | ✓ | ✓ |
| Lévy | $\mathrm{e}^{iu\mu-\lvert\sigma u\rvert^{\frac{1}{2}}(1-i\cdot\mathrm{sign}(t))}$ | $\mu\in\mathbb{R}$, $\sigma\in\mathbb{R}^+$ | ✓ | ✓ | ✓ |
| Laplace extension | $\left(\frac{1}{1+\sigma^2u^2}\right)^k$ | $\mu\in\mathbb{R}$, $\sigma\in\mathbb{R}^+$, $k\in\mathbb{N}$ | ✓ | ✗ | ✓ |
| Generalized asymmetric Laplace | $\frac{\mathrm{e}^{iu\mu}}{(1+\sigma^2u^2+i\nu u)^k}$ | $\mu,\nu\in\mathbb{R}$, $\sigma,k\in\mathbb{R}_{\geq 0}$ | ✓ | ✓ | ✓ |
| Continuous distributions with support $\mathbb{R}_{\geq 0}$ | | | | | |
| Gamma | $\frac{1}{(1-\sigma u)^\kappa}$ | $\sigma\in\mathbb{R}^+$, $\kappa\in\mathbb{R}^+$ | ✓ | ✓ | ✓ |
| Chi-squared | $\frac{1}{(1-2u)^{\frac{k}{2}}}$ | $k\in\mathbb{N}$ | ✓ | ✗ | ✓ |
| Discrete distributions | | | | | |
| Poisson | $\mathrm{e}^{\lambda(\mathrm{e}^{iu}-1)}$ | $\lambda\in\mathbb{R}_{\geq 0}$ | ✓ | ✓ | ✓ |
| Bionomial | $(p\mathrm{e}^{iu}+(1-p))^n$ | $p\in[0,1]$, $n\in\mathbb{N}_0$ | ✓ | ✗ | ✗ |

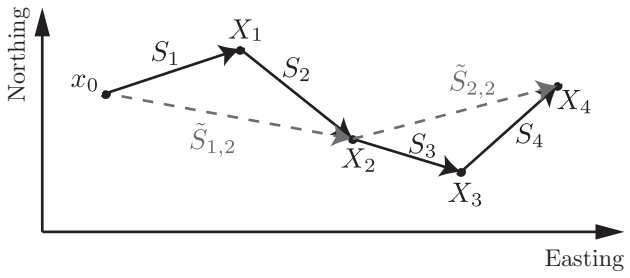



**Figure 4.1.** Schematic of locations $\boldsymbol{X}_i$ and steps $\boldsymbol{S}_i$ between locations. Solid lines indicate the original process, grey dashed lines represent the subprocess for n=2. The subprocess consists of steps $\tilde{\boldsymbol{S}}_{i,2}$.

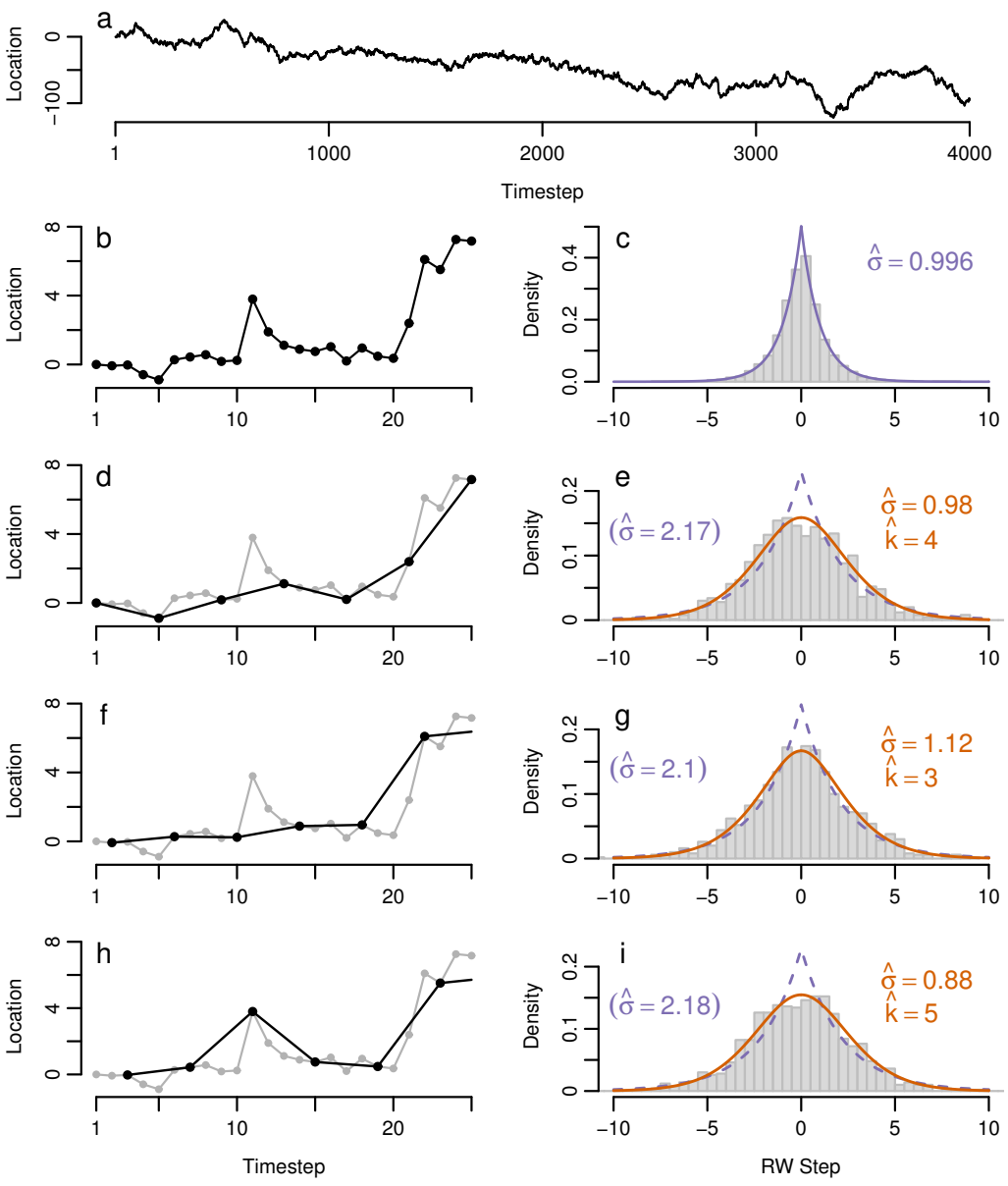



**Figure 4.2.** Inference results when using the Laplace model versus the generalized Laplace model. **Panel a:** Simulated 1D-random walk with Laplace distributed steps with mean zero and scale $\sigma = 1$. **Panel b:** Excerpt of panel a for time steps 1 to 25. **Panel c:** Histogram of realized steps of the random walk, fitted with a Laplace distribution with mean zero. The estimate of the scale $\sigma$ is denoted by $\hat{\sigma}$. **Panels d, f, h:** I subsampled the random walk, taking every 4th location. The panels show the original random walk (in grey) and the subsample (in black). Different subsamples arose depending on the starting location of the subsampling procedure. The three panels start the subsampling at $x_1$, $x_2$, and $x_3$, respectively. Each subsampled path is 1000 time steps long.

**Figure 4.2. Panels e, g, i**: Histograms of realized steps of the subsampled paths. Each histogram corresponds to the subsample to its left. Steps were fitted with a Laplace distribution (dashed purple line) and with a generalized Laplace distribution as given in equation (4.22) (solid red line). The generalized Laplace model accounts for the subsampling with its additional parameter $k$ (here $k = 4$) and is thus the correct model. When fitted to the subsampled random walks, $k$ was estimated simultaneously with $\sigma$. The estimate of $k$ varies for the different subsamples, reflecting the stochasticity of the data, but it is always close to 4. When using the generalized Laplace model, estimates of the scale $\sigma$ are valid estimates for the scale of the original random walk as well. In contrast, the scale estimate from the simple Laplace model (given in parenthesis) cannot validly represent the original scale and naturally overestimates $\sigma$.

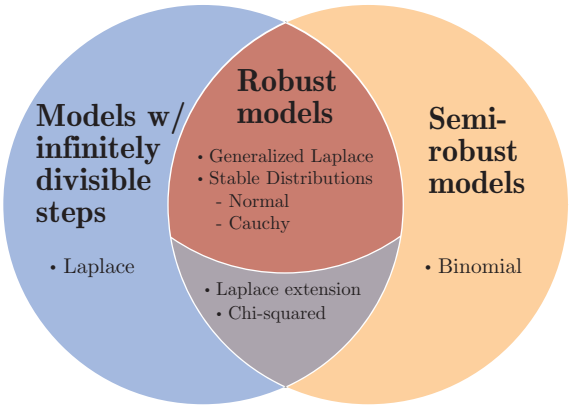



**Figure 4.3.** Graphic depiction of the relationships between semi-robust and robust models and models with infinitely divisible step distributions. Each section contains examples from the text for step distributions that lead to the type of model.

# Chapter 5

# Robustness of movement models: can models bridge the gap between temporal scales of data sets and behavioural processes?[1]

## 5.1 Introduction

Major advances in tracking technology during the last decades have made large datasets of animal movement available to ecologists, and analyses of data have become widespread in ecology. These analyses have shed light on mechanisms that underly fundamental processes such as migration (Robinson *et al.*, 2009; Costa *et al.*, 2012), navigation (Tsoar *et al.*, 2011; Benhamou *et al.*, 2011), or home range behaviour and territoriality (Börger *et al.*, 2008; Potts & Lewis, 2014; Giuggioli & Kenkre, 2014). They have helped to identify conservation goals by revealing habitat preferences and critical environmental features for populations (Sawyer *et al.*, 2009; Colchero *et al.*, 2010; Ito *et al.*, 2013; Masden *et al.*, 2012), as well as the role of important mutualistic interactions between mobile animals and immobile plants (Côrtes & Uriarte, 2013; Mueller *et al.*, 2014). These are only few of the many facets of movement ecology.

Mathematical and statistical models provide a framework for studying movement (Schick *et al.*, 2008; Smouse *et al.*, 2010; Langrock *et al.*, 2013). When linking models to data, we can estimate model parameters and identify best-fitting models, thus infer-

[1] A version of the chapter has been submitted to Journal of Mathematical Biology as: Schlägel, U.E. & Lewis, M.A. Robustness of movement models: can models bridge the gap between temporal scales of data sets and behavioural processes?

ring unknown quantities or mechanisms in movement behaviour. Although movement itself is a continuous process, many individual-based movement models treat time as a discrete variable, viewing movement as a series of locations in space, or equivalently as a series of steps (Turchin, 1998; McClintock *et al.*, 2014). This may largely be ascribed to data being available in this format. Discrete-time models may thus be an intuitive first choice to describe a sampled movement path. However, there may be more reasons to use discrete-time models. The continuous movement path of an animal may consist of various behaviours at different scales (Johnson *et al.*, 2002; Benhamou, 2013). Using a discrete-time model at the scale of interest allows us to focus on the behavioural mechanisms at that scale, while, for example, combining other unknown processes as stochastic effects. Also, the choice of time formulation in a movement model can have side effects that impact inference results. For example, McClintock *et al.* (2014) demonstrated that using a continuous-time Ornstein-Uhlenbeck process in a hierarchical model for identifying behavioural states led to difficulties discriminating between states, due to an inherent correlation between the variables step length and bearing in the Ornstein-Uhlenbeck process.

Many movement models are based on random walks (Turchin, 1998; Codling *et al.*, 2008; McClintock *et al.*, 2014). From simple random walks that assume independently and identically distributed steps, we have moved to correlated random walks, which include directional persistence (Kareiva & Shigesada, 1983), and biased random walks, which can, for example, be used to model centralizing tendencies or long-term directional goals (Benhamou, 2006; Börger *et al.*, 2008; McClintock *et al.*, 2012). Many animals live in heterogeneous environments, and the composition of the environment and availability of resources influence movement decisions (Fortin *et al.*, 2005; McPhee *et al.*, 2012). Therefore, another trend of random-walk extensions has left behind assumptions about environmental homogeneity in favour of spatially-explicit models that incorporate habitat effects on movement decisions (Rhodes *et al.*, 2005; Avgar *et al.*, 2013; Potts *et al.*, 2014). These models have an advantage over statistical resource-selection and step-selection functions (Manly *et al.*, 2002; Fortin *et al.*, 2005; Forester *et al.*, 2009) by allowing simultaneous estimation of movement parameters and environmental effects.

When linking discrete-time models to data, the temporal resolution of the discretization is a critical feature that must be chosen with care. Different time scales may come into play and need to be consolidated. On the one hand, a time scale is given by the biological process of interest. For example, we may be interested in inferring behavioural mechanisms of a movement process and thus need to consider the time

scale at which these mechanisms are relevant. The discretization of a model should represent this scale appropriately. On the other hand, a different time scale may be given by the data collection rate. In practice, the sampling rate of data is subject to technological constraints. One of the major limitations of electronic tagging devices such as Argos or GPS tags is battery life, imposing a tradeoff between measurement rate and total deployment time (Ryan *et al.*, 2004; Breed *et al.*, 2011). Also, to avoid a large noise to signal ratio, the time interval should be chosen so that measurement error relative to distance travelled during a time interval is small (Ryan *et al.*, 2004). For slow moving animals and depending on the accuracy of the tagging device, a minimum time interval of an hour may be necessary (Jerde & Visscher, 2005). Therefore, the resolution of the data may not always match the time scale of the behavioural process of interest. In this case, it becomes a challenge for a model to overcome the conflict.

A related problem is that sampling rate can affect data analysis results (Codling & Hill, 2005; Rowcliffe *et al.*, 2012; Postlethwaite & Dennis, 2013). A common measure calculated from raw movement data is the total distance travelled, which can provide useful information about an animal's energetic expenditures. It is well documented that this quantity is highly influenced by the sampling rate of the data (Ryan *et al.*, 2004; Mills *et al.*, 2006; Tanferna *et al.*, 2012; Rowcliffe *et al.*, 2012). A range of studies demonstrated that other fundamental movement characteristics vary with data sampling frequency as well, for example path sinuosity and apparent speed (Codling & Hill, 2005), movement rate and turning angle (Postlethwaite & Dennis, 2013), and estimates of territory size (Mills *et al.*, 2006). One of the main problems underlying these effects is information loss when subsampling a movement path. This also impairs our capacity to correctly estimate behavioural states through hierarchical modelling approaches that have become widespread in movement analyses (Breed *et al.*, 2011; Rowcliffe *et al.*, 2012). These findings demonstrate that great care is needed when extrapolating movement analysis results beyond the temporal scale of a study. Comparisons of results may not be appropriate if the temporal resolution of the data varies too much, but it is unclear what constitutes "too much".

Despite the evidence of the extent of the problem, little is known about how to solve it. Previous approaches have been mainly empirical, using very fine scale data or synthetic data from simulations, which are subsampled at various resolutions. Movement characteristics calculated at these varying sampling rates are then compared to the values based on the full data, which represent the true values. Some studies have fitted functions to the relationships of movement characteristics and sampling rate (Pépin *et al.*, 2004; Codling & Hill, 2005; Mills *et al.*, 2006). These empirically ob-

tained functions may be used to correct movement characteristics for sampling rate. While correction factors derived from movement data remain situation-specific and cannot easily be applied across species (Ryan *et al.*, 2004; Rowcliffe *et al.*, 2012), we can obtain more general results by analyzing the effects of sampling rate at the level of the model (Codling & Hill, 2005; Rosser *et al.*, 2013). Often, important characteristics of movement can be well captured by models, and therefore analyzing the properties of models can provide more general insights. However, only few such studies exist. An approach to circumvent the problem of scale-dependent statistical inference has been taken by Fleming *et al.* (2014), who use the semivariance function of a stochastic movement process to identify multiple movement modes acting at different temporal scales. The method takes into account all possible time lags between observations. However, there are limitations as to the movement processes that can be included in this analysis (Fleming *et al.*, 2014).

Here, I present a rigorous framework for studying how movement models react to changes in sampling rate, and I use this framework to analyze a class of models based on random walks. With my analysis, I seek to understand whether, and how, models can help to compensate mismatching temporal scales between different data sets or between data and behavioural process of interest. Focusing on spatially-explicit random walks, I investigate whether there are models that can validly be applied to data with different temporal resolutions and how we can account for the differences in resolutions in our interpretation of statistical inference results. In particular, I am interested in how model parameters, and their estimates, change with decreasing temporal resolution. While estimates may change due to a shift in behavioural scale, which we always need to be aware of, I am interested in the changes that arise from the method, that is the model. My framework is related to the statistical concept of robustness, which aims at safeguarding statistical procedures against violations of model assumptions (Hampel, 1986; Huber & Ronchetti, 2009). Often, such violations refer to deviations from assumed probability distributions (e.g. Normal errors), which may result in outliers, misspecified relationships between response and explanatory variables in regression analyses, or violations of the common independence assumption. Here, I define robustness of movement models against changes in temporal discretization. In my framework, I treat robustness as a formal property of a model, namely the movement model. If a model has this property, it can be applied to data with varying resolutions. Additionally, while model parameters do not stay the same, they change systematically and can be translated between resolutions.

Our paper is outlined as follows. In section 5.2, I define what I mean by a movement

model to be robust against changes in temporal resolution. I provide three different definitions, varying in their strength of conditions. In section 5.3, I present different approaches how the definitions can be used to analyze robustness of movement models. Depending on models' complexity and preexisting information, we can use formal analytical methods, numerical calculations, as well as Monte Carlo and simulation approaches. I use these approaches to examine robustness of spatially-explicit random walks and resource-selection models, and I summarize my findings in section 5.4. In section 6, I discuss the relevance of my robustness framework for statistical inference and also draw specific conclusions for spatially-explicit resource-selection models.

## 5.2 Robustness of Markovian movement models

I consider movement models that are discrete-time Markov processes of the form $(\boldsymbol{X}_t,\, t \in T)$, where $\boldsymbol{X}_t \in \mathbb{R}^2$ is an individual's location and $T = \{0, \tau, 2\tau, \dots\}$ is a set of regularly spaced times. This means that I assume that the time interval $\tau > 0$ between two successive location measurements is fixed. Such data often arise from terrestrial animals fitted with GPS devices (Frair *et al.*, 2010). The time interval $\tau$ of the model is usually specified by the resolution of the data. I denote the 1-step transition density for the probability of moving from location $\boldsymbol{y}$ to $\boldsymbol{x}$ between times $t-\tau$ and $t$ by $p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is a vector of model parameters. This notation highlights that the transition density can be time-heterogeneous.

I consider sub-models that consist of every $n$th location of the original model for $n \in \mathbb{N}$. The transition density of the $n$th sub-model for the probability of moving from location y to x between times $t - n\tau$ and $t$ is denoted by $p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$; compare Figure 5.1. A priori, the function $p_{t-n\tau,t}$ can have an entirely different form than $p_{t-\tau,t}$ and may correspond to a different probability distribution. However, via the Chapman-Kolmogorov equation, the $n$-step transition density can be written as a marginal density,

$$\begin{aligned} &p_{t-n\tau,t}(\boldsymbol{x}_t|\boldsymbol{x}_{t-n\tau},\boldsymbol{\theta}) \\ &\qquad = \int_{\mathbb{R}^2\times\cdots\times\mathbb{R}^2} p_{\text{joint}}(\boldsymbol{x}_t, \boldsymbol{x}_{t-\tau}, \dots, \boldsymbol{x}_{t-(n-1)\tau}|\boldsymbol{x}_{t-n\tau},\boldsymbol{\theta})\, d\boldsymbol{x}_{t-\tau}\dots \mathrm{d}\boldsymbol{x}_{t-(n-1)\tau}, \quad (5.1) \end{aligned}$$

where the marginalization is over all intermediate locations visited between times $t-n\tau$ and $t$. For simplicity, I use the general subscript "joint" to denote any joint density of multiple locations. From the notation of the locations it is clear which joint density

is meant. The Markov property of the model allows us to stepwise split up the joint density as follows

$$p_{\text{joint}}(\boldsymbol{x}_t, \boldsymbol{x}_{t-\tau}, \ldots, \boldsymbol{x}_{t-(n-1)\tau} | \boldsymbol{x}_{t-n\tau}, \boldsymbol{\theta}) \\ = p_{t-\tau,t}(\boldsymbol{x}_t | \boldsymbol{x}_{t-\tau}, \boldsymbol{\theta})\, p_{\text{joint}}(\boldsymbol{x}_{t-\tau}, \ldots, \boldsymbol{x}_{t-(n-1)\tau} | \boldsymbol{x}_{t-n\tau}, \boldsymbol{\theta}). \quad (5.2)$$

We can continue this until we obtain

$$p_{t-n\tau,t}(\boldsymbol{x}_t | \boldsymbol{x}_{t-n\tau}, \boldsymbol{\theta}) \\ = \int_{\mathbb{R}^2 \times \cdots \times \mathbb{R}^2} \prod_{k=1}^{n-1} p_{t-k\tau,t-(k-1)\tau}(\boldsymbol{x}_{t-(k-1)\tau} | \boldsymbol{x}_{t-k\tau}, \boldsymbol{\theta})\, d\boldsymbol{x}_{t-\tau} \ldots \mathrm{d}\boldsymbol{x}_{t-(n-1)\tau}. \quad (5.3)$$

Therefore, we can use the 1-step densities to calculate the $n$-step density; compare Figure 5.1. For statistical inference, and thus for my robustness concept, the model parameter vector $\boldsymbol{\theta}$ plays a crucial role. Although the $n$-step density may belong to a different distribution than the 1-step density, equation (5.3) justifies to use the same parameter $\boldsymbol{\theta}$ in the notation of the $n$-step density as in the 1-step density.

I define robustness in terms of the 1-step and $n$-step densities of a model.

**Definition 5.1** (Robustness of degree $n$)**.** Let $n \in \mathbb{N}$ be finite. A movement model of the above type is *robust of degree* $n$ if there exists an injective function $g_n : \boldsymbol{\Theta} \to \boldsymbol{\Theta}$ such that

$$p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}) = p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y}, g_n(\boldsymbol{\theta})) \text{ for all } t \in T \text{ and } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2. \quad (5.4)$$

This definition requires that the $n$-step densities are of the same functional form as the 1-step transitions, where parameters of the model are appropriately transformed via the function $g_n$. This means that if a model is robust, the $n$th sub-model is in fact the same as the original model but with systematically adjusted parameters. The parameter transformation $g_n$ allows us to extrapolate the original parameter $\boldsymbol{\theta}$ to the coarser temporal discretization of the $n$th sub-model. Additionally, we can use the $n$th sub-model to infer the parameter $\boldsymbol{\theta}$ of the original model, because we can invert $g_n(\boldsymbol{\theta})$. Note, however, that this rests on the assumption that the original model defines the process of interest. If, instead we start at the coarser resolution, we would also need surjectivity of the function $g_n$ to conclude the existence of the finer model.

Robustness of degree $n$ has important implications. Given a behavioural process of interest, described by a robust model with parameter $\boldsymbol{\theta}$, we can apply the model

not only to data with matching temporal resolution $\tau$ but also to coarser data with resolution $n\tau$ (e.g. double time interval for $n = 2$). The parameter estimate $\boldsymbol{\psi}$ that we obtain from the coarser data is in fact an estimate of $g_n(\boldsymbol{\theta})$. From this, we can infer the value of $\boldsymbol{\theta}$ via $\boldsymbol{\theta} = g_n^{-1}(\boldsymbol{\psi})$. Additionally, robustness allows us to compare studies pertaining to the same behavioural process but using data sets with different resolutions. If $\boldsymbol{\theta}$ is the estimate based on the finer data, it can be extrapolated to the coarser scale via the parameter transformation $g_n(\boldsymbol{\theta})$, for all degrees $n$ for which the model is robust.

Robustness as in Definition 5.1 is a strong condition that I do not expect to hold but in few special cases of the density $p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$. However, equation (5.4) may hold up to a function $v(\boldsymbol{x},\boldsymbol{y})$, where $v$ is a bounded function that could also depend on $n$ or $\tau$. For practical applications, such *approximate* or *asymptotic robustness* may be sufficient. Therefore, I provide two additional definitions.

**Definition 5.2** (Asymptotic robustness of degree $n$)**.** Let $n \in \mathbb{N}$ be finite. A movement model of the above type is said to be *asymptotically robust of degree n* if there exists an injective function $g_n : \boldsymbol{\Theta} \to \boldsymbol{\Theta}$ and a function $v : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{R}^+$ with the property $v(\boldsymbol{x},\boldsymbol{y};\tau) - 1 = \mathcal{O}(\tau)$ on $\mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^+$, such that

$$p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) = p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},g_n(\boldsymbol{\theta}))\, v(\boldsymbol{x},\boldsymbol{y};\tau) \text{ for all } t \in T \text{ and } \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^2. \tag{5.5}$$

Here, $\mathcal{O}$ denotes the Landau symbol for the order of a function. If a model is asymptotically robust, the $n$-step densities are not exactly the same as the 1-step densities, as was required in Definition 5.1. However, the discrepancy between the densities is bounded by a function that is proportional to $\tau$. More precisely, for an asymptotically robust model we have

$$1 - C\tau \leq \frac{p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})}{p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},g_n(\boldsymbol{\theta}))} \leq 1 + C\tau \tag{5.6}$$

for all $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{\theta}$, for some constant $C > 0$. Therefore, if the time interval $\tau$ of the model is sufficiently small, the $n$-step density will closely resemble the 1-step density with appropriately adjusted parameters. Asymptotic robustness of degree $n$ implies that robustness of degree $n$ is achieved as $\tau \to 0$, that is when the time interval $\tau$ approaches zero.

In applications, the time interval $\tau$ may not be chosen sufficiently small for Definition 5.2 to be useful. Therefore, I give a variation of Definition 5.2, in which the function $v$ does not depend on $\tau$.

**Definition 5.3** (Approximate robustness of magnitude $\delta$ and degree $n$)**.** Let $n \in \mathbb{N}$ be finite. A movement model of the above type is said to be *approximately robust of magnitude $\delta$ and degree $n$* if there exists an injective function $g_n : \boldsymbol{\Theta} \to \boldsymbol{\Theta}$ and a function $v : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^+$ with the property $0 < 1 - \delta \leq v(\boldsymbol{x}, \boldsymbol{y}) \leq 1 + \delta$ for all $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{R}^2$, for a $\delta > 0$, such that

$$p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta}) = p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y}, g_n(\boldsymbol{\theta}))\, v(\boldsymbol{x}, \boldsymbol{y}) \text{ for all } t \in T \text{ and } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2. \tag{5.7}$$

Analogously to equation (5.6), condition (5.7) can be written as

$$1 - \delta \leq \frac{p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})}{p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y}, g_n(\boldsymbol{\theta}))} \leq 1 + \delta. \tag{5.8}$$

In fact, we may consider two different types of magnitudes. Setting

$$v(\boldsymbol{x}, \boldsymbol{y}) := \frac{p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})}{p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y}, g_n(\boldsymbol{\theta}))}, \tag{5.9}$$

this function depends a priori on the parameters, that is we have $v(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta})$, and the magnitude is $\delta_{\boldsymbol{\theta}}$. If $\max_{\boldsymbol{\theta}} \delta_{\boldsymbol{\theta}}$ exists, then this is the overall magnitude for the model with all possible parameter values. The magnitude determines how close $n$-step densities are to the parameter-adjusted 1-step densities. If $\delta$ is small, then the correction function $v$ is close to one everywhere, and thus the $n$-step density has similar values as the 1-step density over its entire domain.

Asymptotic and approximate robustness have similar implications for inference as robustness, but only approximately. The quality of the approximation depends on $\tau$ or the magnitude $\delta$. Suppose we wish to estimate parameters of a behavioural process that we formulate in a model. Suppose we consider the time interval $\tau$ as suitable for the process. If the model is robust of degree $n$, we can use data not only at the matching scale but also at a coarser scale. For example, if the model is robust of degree 2, we can use data obtained at time interval $2\tau$. Because the model is also valid for the coarser scale, we can translate parameter estimates between the scales via the function $g_n$. If a model is asymptotically or approximately robust, the model is not exactly but still approximately valid for the coarser scale. To see this, consider the likelihood function

$$L_1(\boldsymbol{\theta}|\{\boldsymbol{x}_0, \boldsymbol{x}_\tau, \boldsymbol{x}_{2\tau}, \ldots, \}) = \prod_{t \in \{\tau, 2\tau, \ldots\}} p_{t-\tau,t}(\boldsymbol{x}_t|\boldsymbol{x}_{t-\tau}, \boldsymbol{\theta}). \tag{5.10}$$

If a model is robust of degree $n$, the likelihood for data at time interval $n\tau$ is

$$\begin{aligned} L_n(\boldsymbol{\theta}|\{\boldsymbol{x}_0, \boldsymbol{x}_{n\tau}, \boldsymbol{x}_{(n+1)\tau}, \dots, \}) &= \prod_{t\in\{n\tau,(n+1)\tau,\dots\}} p_{t-n\tau,t}(\boldsymbol{x}_t|\boldsymbol{x}_{t-n\tau}, \boldsymbol{\theta}) \\ &= L_1(g_n(\boldsymbol{\theta})|\{\boldsymbol{x}_0, \boldsymbol{x}_{n\tau}, \boldsymbol{x}_{(n+1)\tau}, \dots, \}). \end{aligned} \tag{5.11}$$

If a model is asymptotically robust, we have instead

$$L_1(g_n(\boldsymbol{\theta})) \cdot (1 - C\tau + \mathcal{O}(\tau^2)) \leq L_n(\boldsymbol{\theta}) \leq L_1(g_n(\boldsymbol{\theta})) \cdot (1 + C\tau + \mathcal{O}(\tau^2)), \tag{5.12}$$

omitting the notation of the data, which is the same as in equation (5.11). Analogously, for approximate robustness we have

$$L_1(g_n(\boldsymbol{\theta})) \cdot (1 - \delta + \mathcal{O}(\delta^2)) \leq L_n(\boldsymbol{\theta}) \leq L_1(g_n(\boldsymbol{\theta})) \cdot (1 + C\delta + \mathcal{O}(\delta^2)). \tag{5.13}$$

Therefore, if a model is asymptotically or approximately robust of degree $n$, we may loosely write $L_n(\boldsymbol{\theta}) \approx L_1(g_n(\boldsymbol{\theta}))$, that is the likelihood functions based on data at time interval $\tau$ and on data at interval $n\tau$ are approximately the same. Thus, if data at time interval $\tau$ is not available, we can analyze data at time interval $n\tau$ instead, using the likelihood $L_1$ of the original model. Parameter estimates obtained in this way can be translated to the scale $\tau$ by using the inverse parameter transformation $g_n^{-1}$. Such results from statistical inference based on $L_1$ may be close to results based on the correct $L_n$, which may be difficult to compute. How close results are depends on the quality of the approximations in Definitions 5.2 and 5.3 via $\tau$ or $\delta$. For example, if a model is approximately robust with a very small magnitude $\delta$, the likelihood $L_1$ will describe data at time interval $n\tau$ almost as well as $L_n$.

## 5.3 Analyzing spatially-explicit random walks

I used the robustness definitions to analyze spatially-explicit random walk models. These models merge general movement tendencies of an individual with decisions based on specific characteristics of locations, such as environmental features and available resources. I investigated how the models react when applied to data with increasingly coarser temporal resolution.

My robustness definitions have two key features. First, the 1-step transition densities of the model and the $n$-step densities of the sub-models need to have the same form. Second, model parameters, which are parameters of the densities, need to be

transformed by a known function $g_n$. We can assume different approaches to investigate robustness properties of a model, depending on whether we have a candidate for the parameter transformation $g_n$ or not. If prior knowledge allows us to investigate robustness for a given or hypothesized parameter transformation, we can calculate and compare the $n$-step density $p_{t-n\tau,t}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$ and the parameter-adjusted 1-step density $p_{t-\tau,t}(\boldsymbol{x}|\boldsymbol{y},g_n(\boldsymbol{\theta}))$. By showing equality of the two densities, we can verify robustness. For complex models, analytical calculations may be difficult, or even impossible. In these cases, we may resort to numerical calculations, especially when approximate robustness is sufficient.

In many situations, we may not know $g_n$ a priori, nor have any anticipation. Or, we may have tested robustness for a hypothesized parameter transformation but got poor results. In these cases, we need to establish some information on possible forms of the parameter transformation. Additionally, for complex models, numerical calculation of the high-dimensional integral required for the $n$-step density (compare equation (5.3)) may become inaccurate. A solution is then to draw on the ideas of Monte Carlo sampling. Monte Carlo methods and simulations are useful when probability densities are difficult to compute in closed-form but can conveniently be sampled from (e.g. Robert & Casella, 2000). In the following, I demonstrate both approaches for analyzing movement models' robustness.

### 5.3.1 Analytical and numerical approach

Spatially-explicit random walks can be created by merging two elements in the transition density of the model. One component is the general movement kernel $k_{\boldsymbol{\theta}_1}(x;y)$, which can be the transition density of any standard random walk, describing the probability that an individual takes a step from $y$ to $x$ if there were no environmental information available. A second part of the model, given by the weighting function $w_{\boldsymbol{\theta}_2}(x)$, rates each possible step based on the location $x$. The transition densities of the full model takes the form

$$p_{t-\tau,t}(x|y,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \frac{k_{\boldsymbol{\theta}_1}(x;y)\, w_{\boldsymbol{\theta}_2}(x)}{\int_{\mathbb{R}} k_{\boldsymbol{\theta}_1}(z;y) w_{\boldsymbol{\theta}_2}(z)\,\mathrm{d}z}. \tag{5.14}$$

The integral in the denominator serves as a normalization constant.

For simplicity, I restricted my analysis to the 1-dimensional case, that is I assumed that $X_t \in \mathbb{R}$. I further focused on Gaussian kernels $k_{\boldsymbol{\theta}_1}(x;y) = k_\sigma(x;y)$, where $k_\sigma(x;y)$ is a Gaussian density with mean $y$ and standard deviation $\sigma$. The weighting function $w_{\boldsymbol{\theta_2}}(x)$ was assumed to be positive everywhere to ensure that equation (5.14) defines

a density. In the following I simply use $\boldsymbol{\theta}$ for the parameter vector of the weighting function, or, when it is clear which parameters refer to the weighting function, I drop the subscript for the parameter in the notation of the weighting function entirely.

Note that the transition density (5.14) does not depend on time explicitly. Still, as the individual moves through space over time, the centre location $y$ of the kernel shifts. Although the kernel is a function of the distance $\|x - y\|$ only, the weighting function adds a spatially explicit component. Therefore, unless the individual remains at the same location, the transition kernel effectively changes at every time step. In the following, I omit the time-related subscript in the notation of the density and simply write $p_1$ for the transition density (5.14) and $p_n$ for the $n$-step density. The time interval of the original process is always assumed to be $\tau$. The distinction between 1-step and $n$-step density is still important, because the $n$-step density is in fact an integral over multiple 1-step densities; compare equation (5.3).

I investigated whether I could find weighting functions $w_{\boldsymbol{\theta}}(x)$ such that the model with transition density (5.14) is robust, asymptotically robust or approximately robust. I started by verifying Definition 5.1 for simple cases of the weighting function for a fixed parameter transformation $g_n$. As highlighted above, the parameter transformation is a key element, translating parameters between different temporal resolutions. For the parameter of the Gaussian movement kernel $k_\sigma$, I obtained a candidate for the transformation based on the linearity of the Gaussian distribution. If we only consider the kernel $k_\sigma$, we have a simple random walk with normally distributed steps between locations. The $n$-step density (5.3) is then the density of a sum of $n$ normally distributed random variables, which is again normal with standard deviation $\sqrt{n}\sigma$. Therefore, I assumed that the transformation of the kernel's standard deviation was given by $g_n(\sigma) = \sqrt{n}\sigma$. For the parameters of the weighting function I assumed that they remain unaffected, that is $g_n(\boldsymbol{\theta}) = \boldsymbol{\theta}$. This is an ideal property for a weighting function, as it guarantees validity of inference results across different sampling rates without further translation.

In a next step, I used the same parameter transformation $g_n(\sigma, \boldsymbol{\theta}) = (\sqrt{n}\sigma, \boldsymbol{\theta})$ to establish conditions on the weighting function such that the model is asymptotically robust. For this, I assumed that the parameter of the kernel, the standard deviation, was influenced by the time interval $\tau$, that is $\sigma = \sigma(\tau)$. This reflects that an individual may travel larger distances during longer time intervals. Because of the linearity of the Gaussian distribution, I assumed the relationship $\sigma(\tau) = \sqrt{\tau}\omega$, for some $\omega > 0$. For certain conditions on the weighting function, I verified Definition 5.2 analytically for the robustness degree $n = 2$ by calculating the function $v(x, y; \tau)$ and placing bounds

on it.

As alternative to an analytical approach, we can calculate the ratio of 2-step and 1-step density numerically to see whether we can find a function $v(x, y; \tau)$ according to Definition 5.2 for the degree $n = 2$. Define $\delta(\tau) := \max_{x,y} |v(x, y; \tau) - 1|$. Note that since step densities depend on $\tau$ through $\sigma(\tau)$, we may equivalently consider $\delta(\sigma)$. If this is independent of the other parameters $\boldsymbol{\theta}$, we can obtain the bound on $v$ as $\delta := \max_\sigma \delta(\sigma)$, if this maximum exists. More generally, we can consider $v(x, y, \sigma, \boldsymbol{\theta})$ and calculate $\delta_{\boldsymbol{\theta}}(\sigma) := \max_{x,y} |v(x, y; \sigma, \boldsymbol{\theta}) - 1|$. This $\delta_{\boldsymbol{\theta}}(\sigma)$ is the magnitude of approximate robustness (degree 2) for a model with a fixed weighting function, including parameter values. An overall magnitude for the family of models consisting of the model for all parameter values can be obtained as $\delta := \max_{\sigma,\boldsymbol{\theta}} \delta_{\boldsymbol{\theta}}(\sigma)$. I demonstrate these two numerical approaches with an example weighting function.

### 5.3.2 Simulation approach

#### Resource selection models

Resource selection analyses link animal location data and environmental variables to understand animals' space-use patterns in relation to their habitat. These studies provide insight into species' preferences or avoidance of habitat characteristics, which is important information for wildlife management and conservation purposes (Hebblewhite & Merrill, 2008; Latham *et al.*, 2011; Squires *et al.*, 2013). Central methodological elements are resource selection functions (RSF) and resource selection probability functions (RSPF), describing the probability of selection of certain units (e.g. pixels of land) by an individual based on environmental covariates (Manly *et al.*, 2002; Boyce *et al.*, 2002; Lele & Keim, 2006). RSF and RSPF have been used on their own in a mere statistical framework (Boyce *et al.*, 2002; Courbin *et al.*, 2013), incorporated into spatially-explicit models (Rhodes *et al.*, 2005; Aarts *et al.*, 2011), and become part of mechanistic movement models (Moorcroft & Barnett, 2008; Potts *et al.*, 2014)).

I included resource selection in the spatially-explicit random walk with transition density (5.14) by letting the weighting function take the form $w_{\boldsymbol{\theta}}(x) = w_{\boldsymbol{\theta}}(\boldsymbol{r}(x))$, where $\boldsymbol{r}(x) = (r_1(x), \ldots, r_n(x))$ is a vector of resource covariates at location $x$. Each $r_j$ is a function over space, representing resource covariates such as elevation, biomass measures, land cover type, and much more. The transition density becomes

$$p_1(x|y, \sigma, \boldsymbol{\theta}) = \frac{k_\sigma(x; y)\, w_{\boldsymbol{\theta}}(\boldsymbol{r}(x))}{\int_{\mathbb{R}} k_\sigma(z; y) w_{\boldsymbol{\theta}}(\boldsymbol{r}(z))\, \mathrm{d}z}. \tag{5.15}$$

In practice, geographical information is spatially discrete, and therefore the normalizing integral in equation (5.15) becomes a sum over pixels, or cells, of land. Note that I still restrict attention to one-dimensional models.

The weighting function can take various forms, and here I consider two forms modelled after commonly used resource selection functions (Manly *et al.*, 2002; Lele & Keim, 2006), an exponential function,

$$w_{\exp}(\boldsymbol{r}(x)) = \exp\left(\boldsymbol{\beta} \cdot \boldsymbol{r}(x)\right) \tag{5.16}$$

and a logistic function,

$$w_{\log}(\boldsymbol{r}(x)) = \frac{\exp\left(\alpha + \boldsymbol{\beta} \cdot \boldsymbol{r}(x)\right)}{1 + \exp\left(\alpha + \boldsymbol{\beta} \cdot \boldsymbol{r}(x)\right)}. \tag{5.17}$$

The vector $\boldsymbol{\beta}$ comprises all selection parameters with respect to resource covariates $\boldsymbol{r}$. A higher selection parameter means stronger selection with respect to the corresponding resource. In the logistic form, $\alpha$ is an intercept parameter, which can shift the inflection point of the logistic function away from zero. The inflection point is the point where the logistic function attains a value of 0.5, that is where the probability of selecting a resource is 50%. If the exponential form (5.16) is used, an intercept similar to the one used in equation (5.17) is not identifiable, because it cancels in the definition of the transition density (5.15). Therefore I have omitted it in equation (5.16). The function $w_{\log}$ has range $(0, 1)$ and can therefore be used to describe probabilities. This means that this form can be interpreted as an RSPF, which for a given location $y$ specifies the probability that an animal selects this location, given the covariate values of the location. In contrast, the exponential weighting function can only specify values proportional to this probability, with unknown proportionality constant (Lele *et al.*, 2013).

### Sampling models and sub-models

I examined the two models with weighting functions $w_{\exp}$ and $w_{\log}$ for their robustness. Because the weighting functions depend on space through environmental information $\boldsymbol{r}$ they are highly non-linear, and therefore the transition densities are difficult to examine analytically. Sampling probability distributions is a convenient work around and has the additional advantage that we can control parameters and isolate processes of interest. I thus simulated sample trajectories from the model with transition densities (5.15). The joint density of a movement trajectory $(x_1, \ldots, x_N) \in \mathbb{R}^N$ of length

$N \in \mathbb{N}$ is given by

$$p_{\text{joint}}(x_1, \dots, x_N, \boldsymbol{\theta}) = p_1(x_1, \boldsymbol{\theta}) \prod_{t=2}^{N} p_1(x_t | x_{t-1}, \boldsymbol{\theta}). \tag{5.18}$$

Thus, I sampled successively from the transition densities to obtain a full movement trajectory. I obtained samples from the subprocess $\boldsymbol{x}_n = (x_1, x_{n+1}, \dots)$ consisting of every $n$th location by subsampling the full trajectories. These subsamples represent samples from the model with transition densities being the $n$-step densities $p_n(\cdot|\cdot, \boldsymbol{\theta})$.

Because the models rely on environmental data, I simulated resource landscapes as realizations of Gaussian random fields with exponential covariance model (Haran, 2011; Schlather *et al.*, 2013). This resulted in spatially correlated resource landscapes, thus ensuring realism; compare Figure D.1. The sampled movement trajectories were based on these simulated landscapes. To avoid confounding effects and to keep results as clear as possible, I assumed that the weighting function was based on only one resource $r$, thus we have $w_{\boldsymbol{\theta}}(r(x))$. With the exponential covariance model, I assumed that the covariance of resource values at two different locations is given by

$$\text{Cov}(r(x), r(y)) = \exp\Big(\frac{|x - y|}{s}\Big), \tag{5.19}$$

where $s$ affects the decrease of the spatial autocorrelation with increasing distance.

I sampled trajectories for varying parameter values. I used $\sigma \in \{5, 6, 7\}$ and $\beta \in \{0.5, 1, 1.5, 2\}$ in all combinations. In the model with logistic weighting function $w_{\log}$, I further combined the values $\alpha \in \{-1, -0.5, 0, 0.5, 1\}$ with all other parameters. For each parameter combination, I sampled 16 trajectories for 15,000 time steps each; compare Figures D.2 and D.3. For each of the 16 trajectories, I used a different resource landscape, repeating the same set of resource landscapes across different parameter combinations. The 16 landscapes were generated with varying spatial autocorrelation, $s$ ranging between 200–500. This led to a total of 192 sampled trajectories for the model with exponential weighting function and 960 trajectories for the model with logistic weighting function. I subsampled every trajectory at levels $n = 1, \dots, 15$, leaving 1000 steps for the coarsest time series. The subsample for $n = 1$ is the original trajectory.

### Analyzing parameters

While the simulated trajectories represent samples from the original model with transition densities $p_1(\cdot|\cdot, \boldsymbol{\theta})$, the subsamples of the full trajectories provide us with samples

from the sub-models with $n$-step densities $p_n(\cdot|\cdot, \boldsymbol{\theta})$. To learn about the models' robustness properties, we need to test whether the subsamples reconcile with the parameter-adjusted 1-step densities $p_1(\cdot|\cdot, g_n(\boldsymbol{\theta}))$ for some parameter transformation $g_n$. For a given parameter transformation, we can achieve this by analyzing the fit of the model with transitions $p_1(\cdot|\cdot, g_n(\boldsymbol{\theta}))$ with the subsamples. When $g_n$ is unknown, or when the fit for a hypothesized $g_n$ is poor, we first need to investigate the behaviour of the parameters under subsampling to see whether we can find a function $g_n$ as required by the robustness definitions.

Here, I both tested a priori expectations on the parameter transformation and searched for better alternatives. I estimated parameters for all trajectories and their subsamples using maximum likelihood optimization. The likelihood function for the full trajectories is given in equation (5.18). For subsamples, I applied the same model, although I did not know whether subsamples of trajectories followed the same (parameter-adjusted) process as full trajectories. I expected parameter estimates for the full trajectories to be close to the values that I used during the simulations. I call these the "true values", although deviations in the simulations are possible, because simulated trajectories are realizations of stochastic processes. My main interest are parameter estimates for the subsamples. To distinguish estimates from underlying true parameters, I denote the estimate with a hat, e.g. $\hat{\sigma}$. Ideally, the parameters of the subsamples should follow some function $g_n(\sigma, \alpha, \beta)$, and so should the estimates. To see whether such a function exists, I fitted non-linear regression models to the relationship of parameter estimates of subsamples and the subsampling amount $n$. For each parameter, I fitted two models. One model was more restrictive and represented a priori expectations, whereas the other model had an additional free parameter that allowed more flexibility for the parameter transformation.

The general movement kernel $k$ has one parameter, the standard deviation $\sigma$ of the Gaussian distribution. This kernel describes the general movement tendencies of the animal, and $\sigma$ influences the distance covered in each step. With increasing subsampling, the temporal resolution of the movement path becomes coarser, and I thus expected the standard deviation of the kernel to increase. Each step in a subsample is in fact the accumulated result of one or several steps in the full trajectory. If the kernel is the only force driving the movement, the linearity of the Gaussian distribution caused me to expect the standard deviation of the kernel to increase as $\sqrt{n}\sigma$; compare section 5.3.1. With additional resource selection, however, there may be deviations from this behaviour.

For the resource selection parameters $\alpha$ and $\beta$, an ideal behaviour would be that

they remain unaffected by the subsampling, analogously to my assumptions in section 5.3.1. In my model, I assume that each step is influenced by the weighting function. One of the underlying assumptions of a traditional RSF or RSPF is that it gives weights to locations independently of the values of other locations, which means each location is weighted by its present resource only, without consideration of alternative locations. Therefore, resource selection parameters should be independent of the temporal resolution of the data. However, within the spatially-explicit movement framework, resource selection always occurs in the context of the current location and the available surrounding area as defined by the general movement kernel. Therefore, a change in the movement kernel due to increased subsampling may be accompanied by a change in resource selection parameters.

I fitted the non-linear regression models to the parameter estimates separately for each parameter combination. This means that in each regression, I fitted estimates of 16 trajectories and their subsamples. Because of my previous considerations about the kernel parameter $\sigma$, I assumed a power relationship between the estimate $\hat{\sigma}$ and the subsampling amount $n$, stratified by trajectories. I chose the stratification because trajectories were simulated on different landscapes. Also, for the resource selection parameters, especially when their true values were close to zero, estimates could vary between being positive and negative. In these cases, the stratification allowed for flexibility. The model for the estimate of the $n$th subsample of trajectory $i$ is

$$\hat{\sigma}_{i,n} = \hat{\sigma}_{i,1} \cdot n^b + \varepsilon, \qquad 1 \leq n \leq 15, \quad 1 \leq i \leq 16, \tag{5.20}$$

where the error term $\varepsilon$ is normally distributed with mean zero and positive standard deviation $\zeta$. The maximum likelihood estimate of $b$ should ideally be close to 0.5, however as noted above, it may deviate from this value because of resource-selection mechanisms. To test whether $b$ differs from 0.5, I used model selection via AIC between the model in equation (5.20) and the model in which I fixed $b = 0.5$.

Model choice for the resource selection parameters was less clear. Visual inspection of the estimates, preliminary fits with varying models and inspection of residuals suggested a power law for the parameter $\beta$ as well. I thus fitted the following model,

$$\hat{\beta}_{i,n} = \hat{\beta}_{i,1} \cdot n^b + \varepsilon, \qquad 1 \leq n \leq 15, \quad 1 \leq i \leq 16. \tag{5.21}$$

I compared the fit of this model with the model in which I assumed that subsampling does not change the estimate by setting $b = 0$.

For the intercept parameter $\alpha$ in the logistic form of the resource selection function,

I chose a linear model,

$$\hat{\alpha}_{i,n} = \hat{\alpha}_{i,1} + b\,(n-1) + \varepsilon, \qquad 1 \leq n \leq 15, \quad 1 \leq i \leq 16. \tag{5.22}$$

Inspection of residuals suggested that in some cases the relationship between $\hat{\alpha}$ and $n$ was non-linear. However, a power-law model or other non-linear relationships were not consistently more suitable either. Therefore I remained with the simpler, the linear, model, noting that this is a mainly illustrative analysis.

**Calculating approximate robustness**

To accompany the simulation analysis, I examined approximate robustness properties of the two models with exponential and logistic weighting functions. I focused on approximate robustness of degree 2, and I tested the ideal parameter transformations $g_2(\sigma, \beta) = (\sqrt{2}\sigma, \beta)$ and $g_2(\sigma, \alpha, \beta) = (\sqrt{2}\sigma, \alpha, \beta)$ for $w_{\text{exp}}$ and $w_{\log}$, respectively. I numerically calculated a magnitude $\delta = \max_{x,y}(|v(x,y)-1|)$ for every possible scenario that I used in the previous section. This means that I calculated a magnitude for each combination of the parameters $\sigma$, $\beta$, and $\alpha$ (in case of the logistic weighting function) and for each of the 16 simulated resource landscapes. We may therefore think of $\delta$ as $\delta(\sigma, \alpha, \beta, i)$, for $1 \leq i \leq 16$; compare Figure 5.2. I examined whether magnitudes were influenced by parameter values and specific characteristics of the landscapes, such as their spatial autocorrelation and their overall variation $\text{Var}(r(x))$ over the spatial domain. I further calculated an overall maximum $\max_{\sigma,\alpha,\beta,i} \delta(\sigma, \alpha, \beta, i)$. I compared results between the models with exponential and logistic weighting functions, $w_{\text{exp}}$ and $w_{\log}$.

## 5.4 Results

### 5.4.1 Analytical and numerical results

I found few special cases of weighting functions $w_{\boldsymbol{\theta}}$ that, together with the Gaussian kernel $k_\sigma$, resulted in a robust movement model according to Definition 5.1.

The simplest case was a constant weighting function. Such a weighting function reduces equation (5.14) to the case of a homogeneous environment, where only general movement tendencies play a role, but no environmental information. The model is then a simple random walk with normally distributed steps between locations. Because of the linearity of the normal distribution, the model is robust of degree $n$ for all $n \in \mathbb{N}$

for the assumed parameter transformation $g_n(\sigma) = \sqrt{n}\sigma$; compare also Theorem 5.2 for parameters $a = b = 0$.

A natural next step was to consider a linear weighting function. However, a linear weighting function violates the assumption of being strictly positive everywhere. If in equation (5.14) the current location $y$ is the point at which $w$ becomes zero, the normalization integral vanishes. Also, equation (5.14) can become negative and cease to be a valid density function. Still, I could draw on the linearity of the expectation of a random variable to look into this further. The normalization constant in the transition density (5.14) can be viewed as an expectation of the form $E(w(Z))$ for a normally distributed random variable $Z$ with mean $y$. Therefore, if the function $w$ is linear, the normalization constant reduces to $w(y)$. Equation (5.14) then becomes

$$p_1(x|y, \sigma, \boldsymbol{\theta}) = k_\sigma(x; y)\frac{w_{\boldsymbol{\theta}}(x)}{w_{\boldsymbol{\theta}}(y)}. \tag{5.23}$$

The right-hand side of the equation is positive whenever $x$ and $y$ are either both negative or both positive. If movement only occurs in the domain where the weighting function is positive the model is robustness within this domain. The details of the proof can be found in Appendix E.1.

**Theorem 5.1** (Linear weighting function)**.** *Let $w$ be a linear function $w(x) = ax + b$, for $a, b \in \mathbb{R}$. Let $\mathcal{I} \subset \mathbb{R}$ be the interval where $w > 0$. For the restricted domain $\mathcal{I}$, the movement model with transition densities* (5.14) *is robust of degree $n$ for all $n \in \mathbb{N}$. The parameter transformation is given by $g_n(\sigma, a, b) = (\sqrt{n}\sigma, a, b)$.*

I found another special case to be given by an exponential weighting function. Here, no restriction on the domain is necessary. Again, see Appendix E.1 for details of the proof.

**Theorem 5.2** (Exponential weighting function)**.** *Let $w$ be an exponential function of the form $w(x) = Ce^{ax+b}$ for $C, a, b \in \mathbb{R}$. Then the movement model with transition densities* (5.14) *is robust of degree $n$ for all $n \in \mathbb{N}$ with parameter transformation $g_n(\sigma, C, a, b) = (\sqrt{n}\sigma, C, a, b)$.*

The above two theorems show that it is possible to verify exact robustness with the ideal parameter transformation $g_n(\sigma, \boldsymbol{\theta}) = (\sqrt{n}\sigma, \boldsymbol{\theta})$ for certain weighting functions. However, the cases are very restrictive, and robustness will fail for many other, and especially more complex, weighting functions.

I could additionally establish asymptotic robustness for more general conditions on the weighting function. The main result is summarized in the following theorem. For a detailed proof of the theorem, see Appendix E.2.

**Theorem 5.3** (Asymptotic robustness of degree 2)**.** *Let $w_{\boldsymbol{\theta}}$ be continuous and bounded away from zero. Let $w_{\boldsymbol{\theta}}$ further be twice differentiable with bounded second derivative. Then the model with transition densities* (5.14) *is asymptotically robust of degree 2 with parameter transformation $g_2(\sigma, \boldsymbol{\theta}) = (\sqrt{2}\sigma, \boldsymbol{\theta})$.*

Thus, if the weighting function is well-behaved according to the theorem, we can place a bound on the factor by which the 1- and 2-step density vary; compare equation (5.6). This bound is of order $\tau$, such that the discrepancy between 1- and 2-step density decreases with the time interval.

*Example* 5.1 (*Asymptotic robustness of degree 2).* As a simple example, consider the weighting function $w(x) = \sin(\alpha x) + \beta$ for $\alpha > 0$ and $\beta > 1$. The choice of $\beta$ guarantees that the weighting function is positive everywhere. The function $w$ is bounded between $0 < \beta - 1 \leq w(x) \leq \beta + 1$ for all $x \in \mathbb{R}$, and its second derivative is bounded by $|w''(x)| = \alpha^2$. Therefore, Theorem 5.3 holds.

The proof of Theorem 5.3 is constructive in the sense that it provides us with a constant $C$ for equation (5.6) in terms of the bounds on $w$ and $w''$. However, this constant may be rather large and does not necessarily provide the closest bound on the function $v$. Therefore, it can be informative to calculate approximate robustness numerically.

*Example* 5.2 (*Approximate robustness of degree 2).* I continue the above example with weighting function $w(x) = \sin(\alpha x) + \beta$ for $\alpha > 0$ and $\beta > 1$. I calculated the function $v(x, y; \sigma, \alpha, \beta)$ from Definition 5.3 numerically, using different values of $\alpha$ and $\beta$ (Figure 5.3a). From this, I obtained $\delta_{\alpha,\beta}(\sigma)$ (Figure 5.3b), which is the magnitude of approximate robustness (degree 2) for the model with specific weighting function, that is with specific parameters; compare Figure 5.2. In each case, after reaching a maximum the function vanishes for increasing $\sigma$. Therefore it appears that it is possible to find $\delta_{\alpha,\beta} := \max_\sigma \delta_{\alpha,\beta}(\sigma)$. The wavelength of the sine curve, determined by $\alpha$, and the intercept $\beta$ have different effects on the function $\delta_{\alpha,\beta}(\sigma)$. While $\alpha$ shifts the curve, $\beta$ changes the height of the peak (Figure 5.3b). Therefore, it appears that $\delta_{\alpha,\beta}$ is independent of $\alpha$ and decreases for larger $\beta$. For the weighting function to be positive, $\beta$ needs to be larger than one. For $\beta = 1$, the function $\delta_{\alpha,\beta}$ has a maximum at one. From these considerations, I conclude that $\max_{\alpha,\beta} \delta_{\alpha,\beta} = 1$. This is the overall magnitude of approximate robustness (degree 2) for the family of weighting functions

$w(x) = \sin(\alpha x) + \beta,\ \alpha > 0,\ \beta > 1$; compare Figure 5.2. As a word of caution, I emphasize that I only calculated $\delta_{\alpha,\beta}$ for a fixed number of parameter values and only within finite intervals for $x$ and $y$, and therefore results may be limited to these ranges.

In the region where $\delta(\sigma)$ peaks, the approximation of the parameter-adjusted 1-step density $p_1(x|y, \sqrt{2}\sigma, \alpha, \beta)$ to the actual 2-step density $p_2(x|y, \sigma, \alpha, \beta)$ is only rough. However, for larger values of $\sigma$, and independent of $\alpha$ and $\beta$, the function $\delta_{\alpha,\beta}(\sigma)$ seems to vanish, which means that the approximation is good and the discrepancy between 2- and 1-step densities may be neglected. Theorem 5.3 allows to conclude that $\delta_{\alpha,\beta}(\sigma(\tau))$ is bounded by $C\tau$, for a constant $C > 0$, for all $\alpha > 0$ and $\beta > 1$. As can be seen from the steep initial slope of $\delta_{\alpha,\beta}(\sigma)$, especially for higher values of $\alpha$, the constant $C$ would need to be rather large (Figure 5.3b). The calculations of approximate robustness could additionally show that the bound on $v(x, y)$ is in fact much smaller.

### 5.4.2 Simulation results

#### Results for parameter estimates

When analyzing parameter estimates from the simulated trajectories and their subsamples, I found a difference in the behaviour of parameters between the exponential and the logistic weighting function. Generally, subsampling had less effect on the value of parameter estimates using the logistic form, and the behaviour of estimates agreed closer with my expectations.

For both weighting functions, estimates $\hat{\sigma}$ showed a good fit with the power-law model. When I used the exponential form, the estimated power $b$ ranged from 0.45 to 0.5 for varying parameter combinations, thus deviating from expected behaviour for some parameter combinations (Figure 5.4a). For small selection parameter $\beta$, the estimate $\hat{\sigma}$ showed the expected increase as $\hat{\sigma}\sqrt{n}$. With increasingly strong selection, i.e. higher value of $\beta$, estimates $\hat{\sigma}$ became smaller with increased subsampling relative to the ideal relationship. An increase in $\sigma$ did not influence the fit other than leading to a larger residual standard error $\hat{\zeta}$, which is to be expected because of the overall larger values of the dependant variable. In contrast, when using the logistic form, the estimated power $b$ differed only very slightly from 0.5 and in some cases, the simpler model with fixed $b$ was preferred by model selection right away (Figure 5.4b).

The behaviour of the resource-selection parameter $\beta$ also differed between exponential and logistic weighting function. For the exponential form, $\hat{\beta}$ showed a clear increase with increased subsampling, fitted well by the power-law model (Figure 5.5a). The power $b$ remained similar (ranging 0.105–0.124) across parameter combinations,

increasing slightly with larger $\sigma$ (Figure 5.5b). For the logistic form, estimates $\hat{\beta}$ generally remained closer to the original values for $n = 1$ (Figure 5.5c,d). In most cases, model selection via AIC preferred the power-law model to the ideal constant relationship, however, the estimated values of the power $b$ are small, with 53 out of 60 values being below 0.1 (total range 0–0.156, with one exceptional negative value $b = -0.041$). There was a tendency of $b$ to be smaller and more concentrated under stronger selection (Figure 5.5d).

Estimates of the intercept $\alpha$ in the logistic function showed a slight decline with increased subsampling in most cases (Figure 5.6). This decreasing trend existed no matter whether $\alpha$ was positive, negative, or zero. In general, slopes of the linear fit were all close to zero (ranging -0.047–0.058), and in a few cases the null model with $b = 0$ was chosen. I found a trend in the realized intercept values in the simulated trajectories. With stronger effect of selection (larger $\beta$), the intercept estimate $\hat{\alpha}$ of original trajectories ($n = 1$) was stronger concentrated around the true underlying value, which subsequently led to a stronger concentration of estimates of subsamples (Figure 5.6).

**Results about approximate robustness**

When comparing magnitudes $\delta(\sigma, \alpha, \beta, i)$ of approximate robustness (degree 2) between the two models with exponential and logistic weighting function, I found lower magnitudes for the model with logistic function $w_{\log}$. Magnitudes for the model with $w_{\exp}$ ranged between 0.067 and 1.82, whereas those for the model with $w_{\log}$ ranged between 0.02 and 1.19. The 5% quantile, the median and the 0.95% quantile were $[0.092, 0.34, 0.97]$ ($w_{\exp}$) and $[0.046, 0.21, 0.64]$ ($w_{\log}$).

I found that especially the selection parameter $\beta$ had a strong influence on magnitudes, higher values of $\beta$ leading to higher magnitudes (Figure 5.7). For the model with exponential weighting function, there was a tendency that weighting functions whose underlying landscapes had higher variation $\mathrm{Var}(r(x))$ led to smaller magnitudes (Figure 5.7a). However, I did not find an effect of the parameter $s$ that was used in the simulations to influence the spatial autocorrelation of the landscapes. The model with logistic weighting function did not show such an effect of landscape variation. The logistic model had the additional intercept parameter $\alpha$. I found that higher values of $\alpha$ tended to result in lower magnitudes (Figure 5.7b).

## 5.5 Discussion

I have proposed a new rigorous framework for analyzing movement models' capacities to compensate for varying temporal discretization of data. My framework comprises three definitions of varying strength for robustness of discrete-time movement models. Generally, if a model is robust, it can overcome problems of mismatching temporal scales between different data sets or between data and biological questions. Because my robustness is a very strong condition that holds only for very few and generally more simple models, I have introduced the additional concepts of asymptotic and, most importantly, approximate robustness. While for many movement models it is difficult, or even impossible, to examine the transition densities and their marginals analytically, approximate robustness properties of a model can be calculated numerically also for analytically intractable models. Therefore, I believe that especially approximate robustness will prove a useful new concept for movement analyses.

I have formulated my robustness definitions in terms of the transition densities of Markov models, because these models are often fitted to movement data with likelihood-based methods of statistical inference. For the considered models, we can obtain the likelihood function by multiplying the transition densities of subsequent steps. If a model is robust, the transition densities keep their functional form across varying temporal scales, and parameters are transformed via a well-defined function $g_n$. The likelihood function therefore remains the same but will yield different parameter estimates. However, if the parameter transformation is known, estimates from one scale can be translated to estimates at other scales. If a model is only approximately robust, the likelihood function will not remain exactly but at least approximately the same under a change of scale. Depending on the magnitude of the approximate robustness, the approximation of the likelihood function may be sufficiently good to allow parameter estimates to be reasonably comparable for different scales, especially if the difference in scales is small.

Our concept of robustness for discrete-time movement models is related to the formal concept of robustness in statistics. Generally speaking, robust methods in statistics acknowledge that models are approximations to reality and seek to protect outcomes of statistical procedures (e.g. hypothesis testing, estimation) against deviations from the underlying model assumptions. Classic examples are the arithmetic mean and median as estimates of a population mean: while the median is robust against outliers the mean is not (e.g. Hampel, 1986). Often, robustness is viewed in the context of deviations from assumed probability distributions (distributional robustness; e.g. Huber &

Ronchetti, 2009). For example, data may be contaminated by few observations with heavier tailed distribution than the majority of the observations. In regression analyses, robustness may also relate to the homoscedasticity assumption or the functional form of the response function (Wiens, 2000; Wilcox, 2012). Additionally, robustness has been considered when the assumption of independence is violated and instead observations are correlated (Hampel, 1986; Wiens & Zhou, 1996). In this chapter, I considered robustness in the context of discrete-time movement models with respect to assumptions about the temporal discretization. In view of statistical robustness, I studied violations against the assumption that the temporal resolution of the movement model, a stochastic process, matches the resolution of the data, when in fact the data is only a subsample of the assumed process.

There is also a difference between my robustness of movement models and the well-established robustness in statistics. In my framework, robustness is a direct property of a model. In contrast, classical robustness in statistics is defined for objects such as estimators, test-statistics, or more generally, functionals (real-valued functions of distributions) (Hampel, 1971, 1986). For the type of models I have considered here, parameter estimates cannot be obtained analytically but through numerical optimization of the likelihood function. The likelihood function is build by the model's transition densities, and thus I have defined robustness at a very basic level. A possibility for future research is to investigate whether some of the formal concepts of statistical robustness can be applied to my framework to add further insight. With this chapter, I provide a new perspective for studying effects of temporal discretization of movement processes, and I hope to encourage further research.

My analytical investigations indicate that robustness is a rare property among movement models, especially when behavioural mechanisms such as resource selection are added. Therefore, if we apply models to data without considering this issue, we are in danger of misinterpreting results and drawing erroneous conclusions. However, my analysis also shows positive prospects with respect to approximate robustness. Theorem 5.1 suggests that in slowly varying environments that produce locally linear weighting functions we may find some robustness. Theorem 5.3 and the following examples show that certain smoothness and boundedness conditions on the weighting function can lead to approximate robustness. In addition, Example 5.2 further demonstrates that approximate robustness can be investigated numerically on a case-by-case basis. I have illustrated this with a smooth weighting function $w(x)$ that is a direct function of space. In data applications, an animal's preferences for locations usually do not depend on space *per se* but rather through the type of habitat and available

resources, and the weighting function will be less regular. In the simulation study, I have therefore presented a case with a more realistic model.

While it is difficult to analyze the transition densities and resulting $n$-step densities with analytical calculations, I have demonstrated with the simulation approach how we can still investigate robustness properties of complex models. Sampling from probability distributions instead of calculating them is the key idea of Monte Carlo methods. I have used this method to examine sub-models that have the $n$-step densities as transition densities. With this, I obtained the parameter transformation $g_n$. My approach differs from previous studies that have used subsamples of fine-scale data to establish an empirical relationship between sampling interval and movement characteristics (Pépin *et al.*, 2004; Ryan *et al.*, 2004; Rowcliffe *et al.*, 2012). When using data, it can be difficult to tease apart effects that result from the methodology and effects of actual behavioural changes at different scales. Analyzing model properties as I have proposed here allows us to identify those effects of temporal discretization that are attributable to the methodology.

In the demonstration of the simulation approach, I analyzed spatially-explicit resource selection models. These models have an advantage over traditional resource-selection and step-selection functions. In the traditional, regression-type approach, observed movement steps are compared to potential steps that are obtained separately from an empirical movement kernel (Fortin *et al.*, 2005; Forester *et al.*, 2009). In this approach, movement and resource-selection are treated independently, although it is highly likely that both influence each other. In contrast, when fitting the full random walk with resource selection to data by using the likelihood function (5.18), we can simultaneously estimate parameters both of the general movement kernel and the weighting function.

In my analysis of the resource-selection model, I observed systematic trends in values of parameter estimates with changing temporal discretization of movement trajectories. The main purpose was not to analyze these relationships in full detail but to illustrate that they occur and must not be neglected. Comparing the exponential and logistic form of the spatially-explicit resource selection model, I found that estimates varied more with increased subsampling when the exponential form was used, compared to the logistic form. Using the exponential form, estimates of the kernel standard deviation $\sigma$ decreased with increased subsampling compared to the ideal relationship $\sqrt{n}\sigma$. On the other hand, using the logistic form, $\sigma$ followed the ideal relationship that would occur for a purely Gaussian process very closely, even under additional influence of resource selection. The estimated strength of resource selection,

indicated by $\beta$, increased with the subsampling amount. While this effect was strongly pronounced for the model with exponential weighting function, it was only weak for the logistic model. Therefore, if using the logistic model, one may expect to obtain similar inference results across varying temporal discretization.

When I calculated the magnitudes of approximate robustness for the models used in the simulations, I found that those were in line with the results for the parameter estimates. Overall, the model with logistic weighting function had better robustness properties than the model with exponential weighting function. I also found a matching trend for the movement parameter $\sigma$ with varying true values of $\beta$. Estimates of $\sigma$ were closer to the expected behaviour for weaker resource-selection parameters. This was also reflected in magnitudes of approximate robustness. If selection was weaker in the original model, the model exhibited better robustness properties. These results suggest that numerical calculations of approximate robustness can assist our expectations about changes in parameter estimates. On the other hand, although parameter estimates of the weighting function showed a clear difference in behaviour when comparing between the exponential and logistic weighting function, differences within one model between different parameter combinations were less clear. More analyses would be required to entangle more detailed effects.

Overall, the results from the simulations suggest that depending on the resolution of movement data, we may misinterpret animals' movement tendencies and also may overestimate resource selection effects. It is therefore important that we are aware of the changes to statistical inference that can arise merely from the methodology. Here, I have shown that changes in inference results were stronger for the resource selection model with exponential weighting function compared to the logistic form. A possible explanation may be the additional intercept in the logistic function. With increased subsampling, estimates of $\alpha$ tended to decrease, possibly counteracting the increase in estimates $\hat{\beta}$. This could have led to more stability for the parameter $\sigma$ of the general movement kernel. However, this may not explain why resource selection parameters generally varied less themselves compared to the exponential model. Another possibility is that the different form of the weighting functions causes their different behaviour. While the exponential form of the weighting function greatly enhances differences in landscape values, the logistic form is restricted to values in the interval $(0, 1)$. Theorem 5.3 suggests that variation in the rate of change of the weighting function influences robustness properties. Thus the logistic form may produce more stable inference results for varying temporal resolutions. Lele & Keim (2006) suggested several alternatives to the exponential form of a classic resource selection function. These function could also

be implemented as weighting functions in the spatially-explicit model considered here. My study case showed that the choice of weighting function can have implications for statistical inference, and I encourage to choose these more deliberately.

With my study I have illustrated that the concept of the parameter transformation $g_n$ can help to bridge the gap between different temporal resolutions of data. In the model with exponential weighting function, I found that with increased subsampling estimates of the resource selection parameter $\beta$ deviated strongly from the original values. However, the increase in $\hat{\beta}$ could be fitted with a power-relationship. Thus, using the idea of Monte Carlo sampling, I was able to obtain a parameter transformation $g_n$. Using such transformations when comparing results obtained from data with different temporal resolutions could greatly improve our statistical inference, leading to a better understanding of movement behaviour.

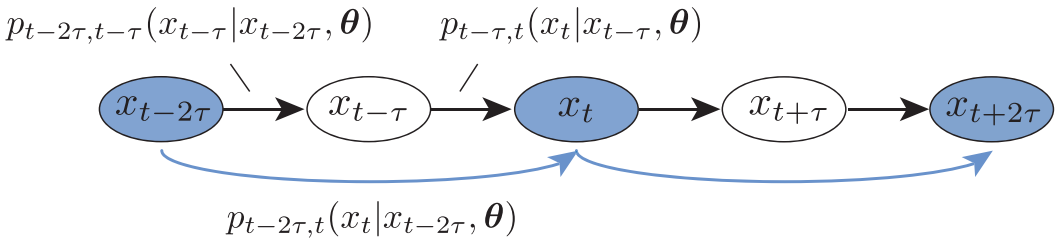



**Figure 5.1.** The second sub-model consists of every second location. The transition densities of the sub-model, which I refer to as 2-step densities, are the marginals over the two intermediate 1-step densities of the original model.

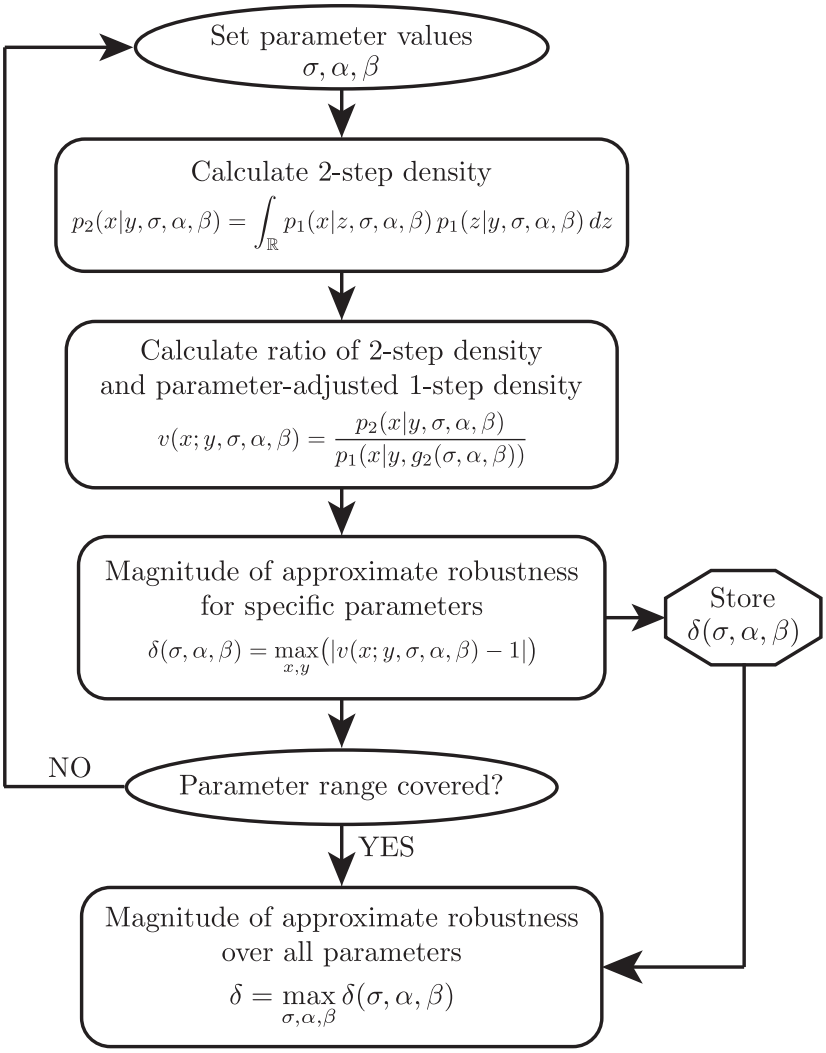



**Figure 5.2.** Steps for calculating the magnitude of approximate robustness of degree 2 for a given model, where $\sigma$ is the parameter of the movement kernel, and $\alpha$ and $\beta$ are parameters of the weighting function. The 1-step density $p_1$ can, for example, be equation (5.14) with the weighting function from Example 5.2, or the resource selection model (5.15) with weighting function (5.16) or (5.17). When the resource selection model is used, the flowchart shows the calculation of the magnitude for one specific resource landscape $r(x)$. When calculating an overall magnitude, practically we do this for a subset of the parameter space.

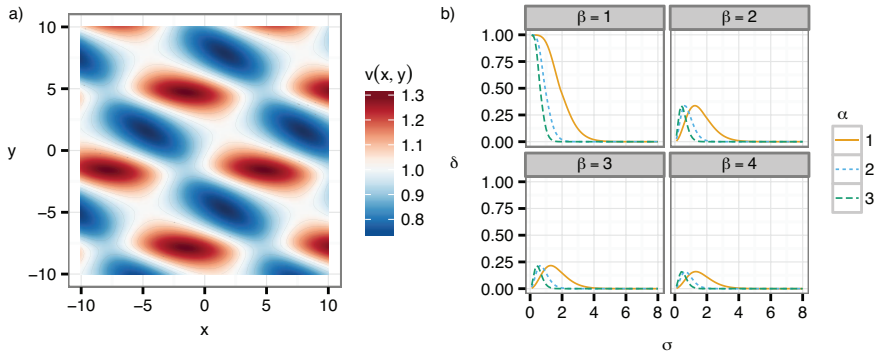



**Figure 5.3. Panel a)**: Numerical calculation of the function $v(x, y)$, which is the ratio of 2-step density $p_{t-2\tau,t}(x|y, \sigma, \alpha, \beta)$ and 1-step density $p_{t-\tau,t}(x|y, g_2(\sigma, \alpha, \beta))$, for the weighting function $w(x) = \beta + \sin(\alpha x)$. Parameter values are $\sigma = 1$, $\alpha = 1$, $\beta = 2$. The function $v(x, y)$ varies roughly between 0.72 and 1.31. **Panel b)**: Numerical calculation of $\delta(\sigma) := \max_{x,y} |v(x, y; \sigma) - 1|$ for the weighting function $w(x) = \beta + \sin(\alpha x)$ for varying values of $\alpha$ and $\beta$. The parameter $\alpha$, which determines the wavelength of the sine, shifts the curve $\delta(\sigma)$ and varies the skewing, while retaining the height of the maximum. The parameter $\beta$ in contrast changes height of the maximum, which is the magnitude $\delta$ of the approximate robustness.

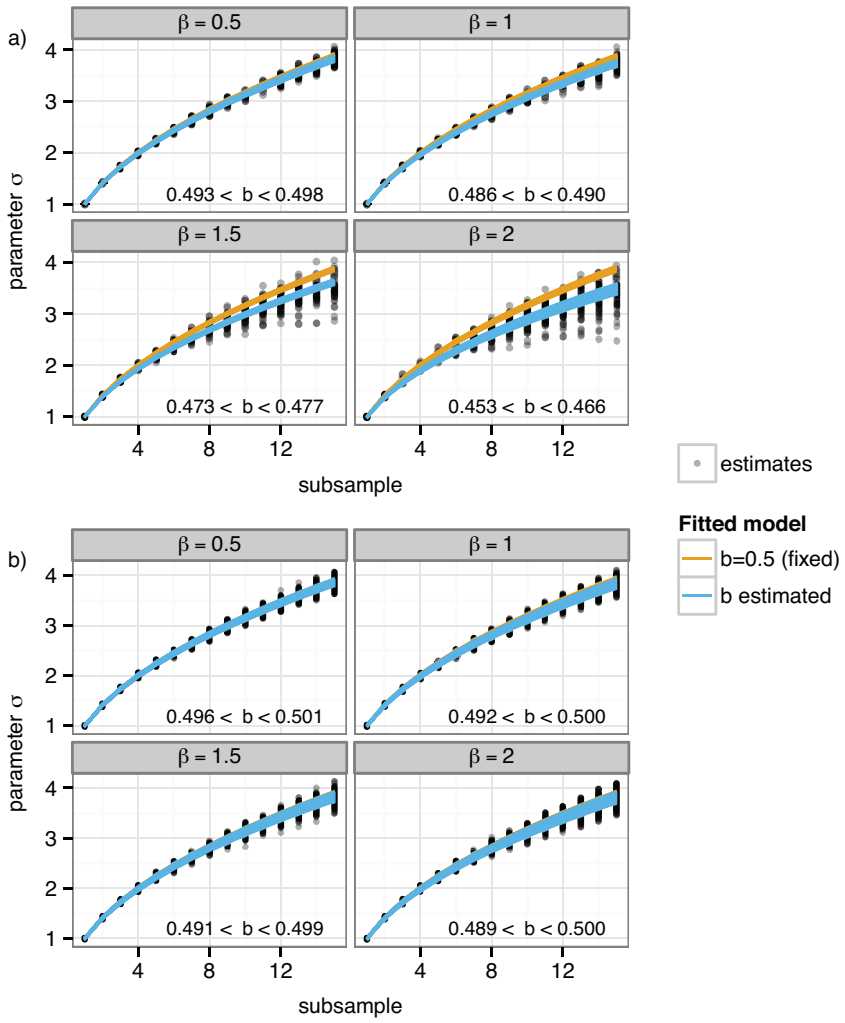



**Figure 5.4.** Simulation results for the kernel parameter $\sigma$: values of $\sigma$ against increasing subsampling amount $n$. Estimates $\hat{\sigma}$ (gray dots) were fitted with a power-relationship, stratified by trajectories, and separately for several combinations of true parameter values ($\sigma$, $\beta$, and $\alpha$ for the model with logistic weighting function). The power $b$ was either fixed at 0.5 (ideal relationship; upper orange lines) or flexible and estimated (lower blue lines). The noted range of $b$ refers to variation for different parameter combinations. Estimates and predictions are standardized by the corresponding true value. **Panel a)**: Model with exponential weighting function. With increasing value of $\beta$, estimates $\hat{\sigma}$ tended to increase less with subsampling compared to the ideal relationship. **Panel b)**: Model with logistic weighting function. The fitted power-relationship was very close to the ideal relationship, such that lines indicating the ideal relationship are overlaid by lines showing the fitted relationship.

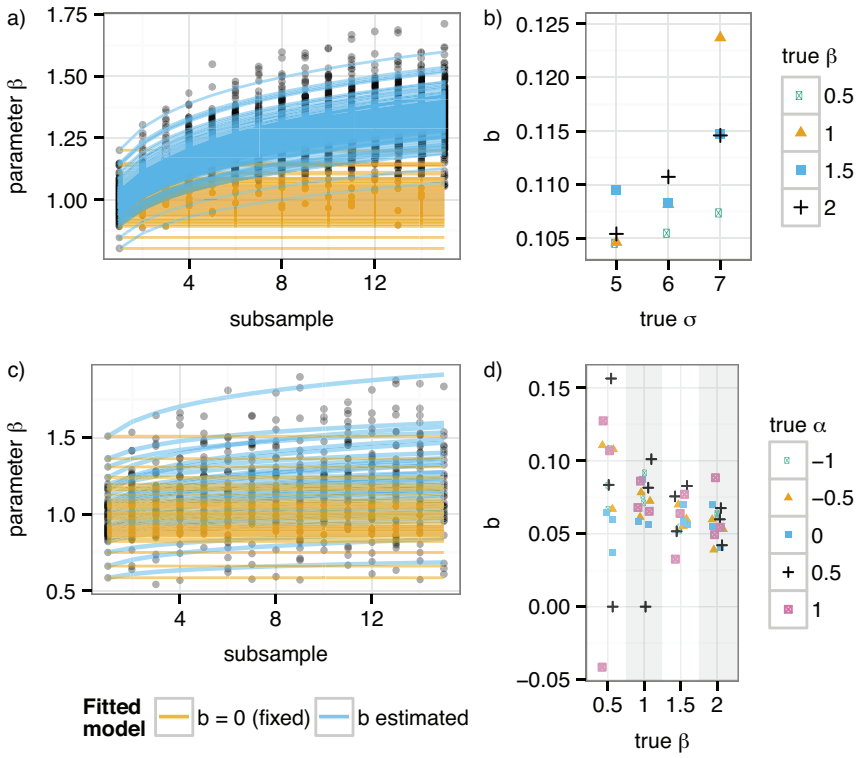



**Figure 5.5.** Simulation results for the resource selection parameter $\beta$ for the model with exponential weighting function $w_{\text{exp}}$ (panels a,b) and logistic weighting function $w_{\log}$ (panels c,d). **Panels a) and c)**: Estimates $\hat{\beta}$ (gray dots) for increasing subsampling amount $n$ were fitted with a power-relationship, stratified by trajectories, and separately for several combinations of true parameter values ($\sigma$, $\beta$, and $\alpha$ for $w_{\log}$). The power $b$ was either fixed at zero, representing the assumption that resource-selection parameters do not change with changing temporal resolution (ideal relationship; straight orange lines), or flexible and estimated (curved blue lines). Estimates and predictions are standardized by the corresponding true value. In panel c), only estimates and predictions for $\alpha = 0$, $\beta = 1$ are shown. **Panel b)**: For $w_{\text{exp}}$, the estimated power $b$ was always above 0.1 and tended to increase with $\sigma$. **Panel d)**: For $w_{\log}$, the estimated power $b$ was mainly below 0.1 and tended to decrease and concentrate more for increasing $\beta$.

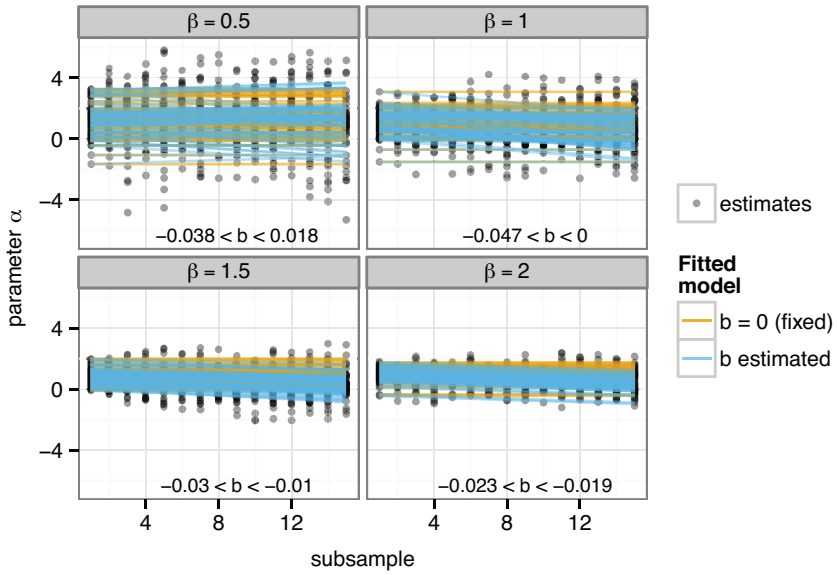



**Figure 5.6.** Simulation results for the resource selection parameter $\alpha$ for the model with logistic weighting function: values of $\alpha$ against increasing subsampling amount $n$. Estimates were fitted with a linear relationship, stratified by trajectories, and separately for several combinations of true parameter values ($\sigma$, $\alpha$, $\beta$). The slope $b$ was either fixed at zero, representing the assumption that resource-selection parameters do not change with changing temporal resolution (ideal relationship; straight orange lines), or flexible and estimated (blue lines). Estimates and predictions are standardized by the corresponding true value and only shown for $\alpha = 0.5$. The noted range of $b$ refers to variation for different parameter combinations.

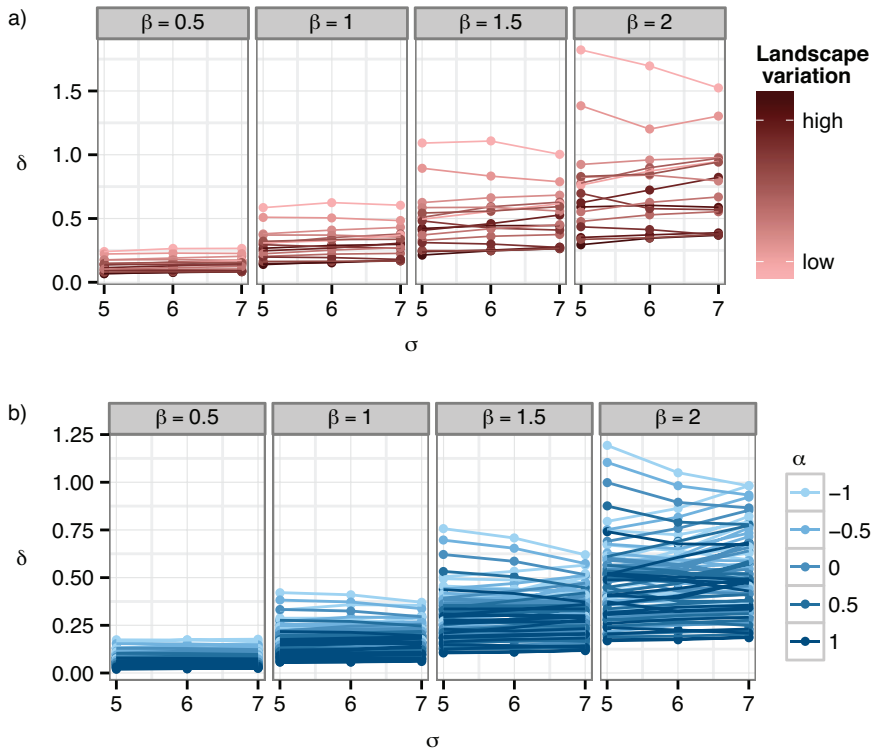



**Figure 5.7.** Magnitudes of approximate robustness for the study case models with exponential and logistic weighting functions. The plots depict $\delta$ for varying values of $\sigma$ and selection parameter $\beta$ (dots). Lines join values for the same landscape $i$, $1 \leq i \leq 16$. **Panel a)**: Magnitudes for the model with exponential weighting function. Values of $\delta$ tend to be lower for landscapes with less variation $\mathrm{Var}(r(x))$. **Panel b)**: Magnitudes for the model with logistic weighting function. Values of $\delta$ tend to be lower for higher values of the additional intercept parameter $\alpha$.

# Chapter 6

# New modelling tools improve qualitative and quantitative understanding of animal movement

In 2008, Nathan *et al.* coined the term "movement ecology" as a paradigm for studying movement of all types and of all organisms. The paradigm emerged from a need to unify a rising number of movement analyses and approaches, fuelled by a significant increase in data availability. Within the paradigm, Nathan *et al.* (2008) proposed a new conceptual framework, in which movement arises from an interplay of three major components: internal state ("why move?"), motion capacity ("how to move?") and navigation capacity ("where to move?") – all possibly affected by biotic and abiotic external factors. The framework is directed towards deciphering both the proximate and the ultimate causes of movement.

With the first part of my thesis, I contributed a new model that implements this conceptual framework. I particularly focused on expanding the navigation process compared to previous approaches. My model has a similar form as spatially-explicit resource selection models (Rhodes *et al.*, 2005; Moorcroft & Barnett, 2008), which formulate movement decisions as a result of general movement tendencies (given by motion capacities) and resource preferences (external factors mediated by navigation capacities). In these models, the navigation process involves an evaluation of the environment, for example with respect to land cover types. In my model, I extended this navigation process by two features. First, I included a new type of information variable that influences movement decisions, which is dynamic. This means that realized movement feeds back to the variable, interactively changing it values. Second, this variable can encode previous experience and be more abstract than directly ob-

servable environmental information. I exemplified this with the variable *time since last visit (TSLV)* to locations. This variables contains information about the spatial and temporal aspect of an individual's travel history, keeping track of temporal distances ("how long ago?") to locations in a spatially-explicit manner. With this new model, I provide a framework for modelling movement strategies in which the realized movement path changes the environment, e.g. due to resources depletion, changes information, e.g. about the travel history, or serves to acquire information, e.g. about temporary availability of resources. Such feedbacks close the loop of the external factors dynamics *sensu* Nathan *et al.* (2008), but also provide a direct link between the movement path and navigation capacities. In Chapters 2 and 3, I verified the new model's suitability for statistical inference, including both model selection and parameter estimation, and demonstrated its applicability with an analysis of wolf movement data. I discuss this in more detail in the following section 6.1.

In the second part of my thesis, I turned towards a more subtle, yet fundamental, methodological problem in movement ecology. Data collection methods, and often modelling approaches as well, discretize the temporal dimension of movement processes, by necessity and also as helpful simplification. However, this discretization challenges data analysis, because data sampling rate may affect results and conclusions. In my thesis, I developed a rigorous and comprehensive new framework for studying this problem. I took the view that data sampling rate is given, but that we may compensate its influence through model choice and modelling tools. I therefore developed the concept of *movement models' robustness against varying temporal resolution*, drawing on the clear language of mathematics and statistics. In Chapters 4 and 5, I introduced a series of definitions for movement models' robustness, which vary in their strength of conditions but all rest on the same requirement that a model can validly be applied to data with varying resolutions while parameters change in a systematic way that can be predicted. I used my new definitions for a thorough analysis of existing models, starting with classic random walks (Chapter 4) and moving on more generally to first-order Markov models that add a spatially-explicit component to random walks (Chapter 5). I showed that true robustness with respect to temporal resolution is rare, thus substantiating the general apprehension that many contemporary movement analyses are tied to their particular resolution, making it difficult to transfer and compare results. However, even more importantly, my work also opens new avenues to solutions, which I discuss in more detail in section 6.2.

## 6.1 Inferring cognition and memory use from movement patterns

In Chapters 2 and 3, I answered the call for considering the role of cognition, including memory, for movement processes. In my model, I implemented that movement can be influenced by TSLV, which keeps track of the spatial and temporal component of an individual's travel history. Information about elapsed times since previous visits to locations can be important for determining return behaviour, for example whether to adopt a "win-stay" (revisit a location) or "win-shift" (shift to alternative location) strategy in a foraging situation (Burke & Fulham, 2003; Janmaat *et al.*, 2006). Optimal behaviour also depends on characteristics of the resources at the focal location, for example with respect to resource depletion or mobility (Pyke, 1984; Sulikowski & Burke, 2011). As an example, I assumed that TSLV influences the probability that an individual will move to a location in two ways. First, for recently visited locations, with low TSLV, the probability of return is small. Second, for locations with long absence, i.e. high TSLV, the probability of return approaches one. Such behaviour may occur particularly in a situation with depleting resources that need time to replenish (Davies & Houston, 1981; Burke & Fulham, 2003). Being able to formulate such movement strategies in movement models that are amenable to statistical inference allows us to test hypotheses about cognitive-based behaviours in free-ranging animals and natural environments.

The dynamic information variable TSLV increases the model's complexity substantially. TSLV is not only an additional, temporally dynamic variable that influences movement but also affected by movement itself. Therefore, I tested in Chapter 2 whether this dynamic mechanisms could be correctly identified, using simulated data. With the simulation study, I showed that classic model selection was able to distinguish the dynamic model, also in mixed form, from traditional models that only contained static environmental effects. Also, parameter estimation generally recovered the values used in the simulation. With a detailed estimability analysis (Lele *et al.*, 2007), presented in Appendix B, I illustrated that Data Cloning estimability diagnostics are a useful complementary tool to understand model fitting problems (Lele *et al.*, 2010). In my analysis, model fitting problems occurred systematically, that is primarily when a model with the dynamic variable TSLV was fitted to data generated without this effect. This attests that model fitting problems, e.g. due to ridges in the likelihood function, should alert and prompt a reconsideration of the model. However, my analysis also showed that when TSLV was a driver of movement decisions, the framework was able

to identify it correctly.

In Chapter 3, I demonstrated the applicability of the new model with a data analysis. I fitted the model to movement paths of gray wolves in south-west Alberta, Canada, to test whether wolves engage in an active prey-management strategy. The aim of such a strategy would be to reduce impacts of behavioural depression of prey through optimal timing of returns to hunting sites, a behaviour that S. L. Lima called "prudent" (Jedrzejewski *et al.*, 2001; Lima, 2002). Jedrzejewski *et al.* (2001) reported observed movement patterns in line with the hypothesis, however, my modelling framework allowed to test the involved mechanisms in more detail. I found that wolves in my analysis did not appear to patrol their territories *per se* as the wolves observed by Jedrzejewski *et al.* (2001). Instead, TSLV only mattered in very specific areas, which were locations close to the territory boundary and with very high prey density. Here, the effect of TSLV was in agreement with the prey-management strategy. However, the combined effect of prey and TSLV, necessary to support the prey-management hypothesis, was only pronounced significantly in one of three wolves. In my Chapter, I only analyzed the movement paths of three wolves in total. To reach stronger conclusions, it will be necessary to analyze more data.

In my modelling framework, I assumed that prey-management is achieved by consideration of time via TSLV. One aspect of the strategy is to leave a site when anti-predator behaviour depresses prey availability, as predicted by optimal foraging theory (Charnov *et al.*, 1976; Pyke, 1984). An open question is whether wolves learn to perceive and react to prey depression, or whether they have learned, evolutionarily, an adaptive movement strategy that uses TSLV as a proxy (Burke & Fulham, 2003; Sulikowski & Burke, 2011).

I further assumed implicitly that information about TSLV is mediated by memory. There is evidence that animals use spatial memory and memory about temporal distances for decision making (Clayton & Dickinson, 1998; Burke & Fulham, 2003; Janmaat *et al.*, 2006; Martin-Ordas *et al.*, 2009). However, it is difficult to inevitable confirm the use of memory, especially when we only have a sampled movement path and relatively coarse environmental data. Instead of drawing on internally stored memory, animals may be able to use externalized memory, a famous example being the slime mould *Physarum polycephalum*, which leaves a trail of extracellular slime and subsequently uses it for navigation (Reid *et al.*, 2012). In case of the wolves, it is possible that they navigate with the help of scent marks. Wolves scent mark not only along the territory boundary but also along common routes and at junctions (Peters & Mech, 1975; Peters, 1979). Scent marks could provide wolves with information about previ-

ous visits by themselves and pack members and thus serve as externalized memory of TSLV (Peters & Mech, 1975). A drawback of relying on scent marks is that they need to be probed, hence they are only useful at their very location. To test whether wolves keep track of TSLV via scent marks or memory, one could analyze wolves' path for goal-orientedness. If animals set target locations for their travels, it is more likely that memory is involved (Asensio *et al.*, 2011; Janson & Byrne, 2007)

With its features, the new model opens new avenues towards future research. In my thesis, I combined effects of TSLV only with other static environmental variables. The model could become even more powerful with inclusion of additional layers that represent specific events or multiple behavioural modes. For example, wolves's movements are certainly influenced by kill events. Handling a kill can require a significant amount of time, ranging from a few hours for small-bodied prey such as deer to 24-48h for large-bodied prey such as moose (Franke *et al.*, 2006). Therefore bouts of extensive search alternate with more stationary phases (Franke *et al.*, 2006; Webb *et al.*, 2008). In Chapter 3, I focused on phases of extensive movement by dismissing short steps indicative of a stationary phase from the analysis. A more complete solution would be to add a second behavioural mode, in which movement is characterized by small steps and use of TSLV is reversed compared to the extensive mode. Additional layers, possibly representing unobservable variables can be realized by placing the model within a hidden Markov model or state-space model (Langrock *et al.*, 2013; McClintock *et al.*, 2012).

In Chapter 2, I accounted for the possibility of such model extensions by using the flexibility of a Bayesian model fitting technique. Such techniques, for example Markov Chain Monte Carlo, provide a better, and sometimes the only, means to fit hierarchical models to data. Software packages such as WinBUGS, JAGS or STAN, provide convenient tools for performing analyses (Lunn *et al.*, 2000; Plummer, 2013; Stan Development Team, 2014). However, when drawing on these tools for my data analysis in Chapter 2, I encountered challenges. The model's likelihood function is composed of probability densities that do not correspond to standard distributions. While the general movement kernel is built from standard distributions (although this is not necessary), the spatially-explicit weighting function modifies it in a highly non-linear way. Additionally, the dynamic nature of TSLV within the spatially-explicit approach requires the processing of large amounts of data. The combination of these two properties strained capacities of both JAGS and STAN. An alternative is to program situation-specific model fitting algorithms in fast languages, e.g. C/C++. Naturally, this reduces the ease with which data analysts can apply complex models. Ultimately,

these technical challenges in movement ecology will be solved by an increasing synergy between disciplines, including statistics, computing science and information technologies (Demšar *et al.*, 2015).

## 6.2 Making movement models more robust against varying temporal discretization

In Chapters 4 and 5, I developed the concept of movement models' robustness against varying temporal discretization of the movement process. This concept is related, not only by name, to the commonly known robustness in statistics. Formal robustness considerations in statistics explicitly acknowledge that statistical models are always approximations of the processes that generate observations and that models make simplifying assumptions, which, however, may be inaccurate (Box, 1980; Huber & Ronchetti, 2009). If reality deviates from assumptions, this can heavily impact results of statistical analyses. For example, it is well known that the sample mean can be affected substantially by outliers, e.g. few observations that originate from a heavier-tailed distribution than the assumed one (e.g. Hampel, 1986; Wilcox, 2012). Robust statistical methods are designed to safeguard results against misspecified assumptions (Hampel, 1986; Huber & Ronchetti, 2009). The same idea underlies my definitions of movement models' robustness. In the case of movement models, we make an assumption about the model's temporal resolution, often based on the data's resolution. However, in fact, the underlying process may be better described by another resolution. Sometimes, the optimal resolution can be determined by scale considerations, for example when we model inter-patch movement at the patch level (Benhamou, 2013). If, in contrast, we are interested in the finer behavioural rules of the inter-patch movement, for example, compared to intra-patch movement, then it may be less clear which resolution to chose. Or, even if we have an idea about a good resolution, it may not match the data's resolution. My robustness of movement models is designed to safeguard statistical inference against varying temporal resolutions.

While my robustness definitions and robust statistical methods share some ideas and objectives, they also differ. Robust methods are constructed to prevent, or limit, outcomes from change due to deviations in model assumptions. In my definitions, I apply this to the model form, via the step distributions, but not necessarily to the model parameters. Instead, I ask model parameters to change systematically with the temporal resolution. This change is described by a well-defined parameter transfor-

mation, by which we can translate parameter values between temporal resolutions. Another specific feature of my definition is that it is defined at the level of the model, because in my view the model is a tool itself to analyze movement in replacement of an experiment. The typical way to link movement models to data is via the likelihood function, whether in a frequentist or a Bayesian framework. Given a robust model, the robustness carries through such likelihood-based statistical inference. An alternative approach may be to consider the problem at the stage of estimation. Here, statistical robustness theory may offer new approaches (Huber & Ronchetti, 2009; Wilcox, 2012).

My analysis of random walks and their spatially-explicit extensions showed that few models have the robustness property. When considering random walks, it is mainly the stable distributions that, as step distributions, lead to robust models. With exception of the normal distribution, stable distributed steps have heavy-tailed step length distributions. These step distributions lead to movement patterns characterized by clusters of small steps, interspersed by few long steps. Such movement patterns, however, arise more likely as a combination of two behavioural modes (Benhamou, 2013; Plank *et al.*, 2013). Therefore it is not clear whether random walks with stable step distributions are useful to describe movement within a single mode, despite their desirable property of being robust. When considering spatially-explicit random walks, I could verify robustness for only very few models. Analytical investigations were limited to models in which the spatially-explicit component was a simple function of space. Finding only few robust models even within this class illustrates that robustness is a very strong condition. This may not come surprisingly, yet it is an important starting point for any further research.

However, in my thesis, I also identified ways to widen the scope of robustness and increase its applicability. In Chapter 4, I illustrated that it is possible to find robust extensions of models that are otherwise not robust. Such an extension contains additional parameters, which we may consider as nuisance parameters, while they allow the focal parameters to remain their original values and validity. However, suitable extensions can be difficult to find or may not admit closed-form solutions. It may still be worthwhile to investigate this new approach further. In Chapter 5, I introduced the definition of approximate robustness. Of all the robustness definitions, this is the most applicable one, especially for models beyond simple random walks. It requires the transferability of a model across temporal resolutions only approximately. I demonstrated both a numerical and a simulation approach for investigating approximate robustness properties of a model. I found that in contrast to exact robustness, approximate robustness shows promise to be more widely present in models. In my simulation study in Chapter 5, I

demonstrated that we can also achieve better robustness properties of a model through specific model components. If we can choose between comparable models, or model components, one choice may perform better with respect to robustness and thus be preferable.

The new concept of movement models' robustness offers possibilities to mitigate the influence of temporal resolution. If a model is (approximately) robust, it is not only valid at a specific resolution but also at coarser resolutions. In the case of random walks with i.i.d. steps, robustness additionally implies validity for finer resolutions. For data analysis, the property of a model to scale up to coarser resolutions is usually more relevant, as data is rather too coarse than too fine. A key element in robustness is the parameter transformation. If we are in a situation where we have a reference resolution for the process of interest but coarser data, possibly that of a second comparable study, the transformation allows us to translate results between the reference and coarser resolution. If, however, we are in a situation where we rely on a model to be robust without any reference to a particular resolution, the parameter transformation should be the identity, at least with respect to the parameters of interest. This is necessary because to use the transformation for translation between resolutions, we need to know not only the transformation itself but also the difference between the resolution. I suggest that further research may be directed towards this issue. In Chapter 5, I showed how to obtain the transformation via simulations. This is similar to previous approaches to account for temporal resolution in estimates of travel distance and path tortuosity (Pépin *et al.*, 2004; Benhamou, 2004). Given that we can calculate the magnitude of approximate robustness numerically, we may also devise a numerical strategy to find a function that minimizes the magnitude and thus constitutes a suitable parameter transformation. Such approaches to find a parameter transformation could also be useful in cases where an assumed "true" resolution is missing.

## 6.3 Closing remarks

With my thesis work, I have contributed new modelling tools for analyzing animal movement. First, my new model stands at the forefront of an enterprise to understand the role of cognition, including memory, for movement processes. The new model makes it possible to test hypotheses about cognitive-based movement strategies, taking into account the history of the realized movement path and how it affects future movement decisions. Such analyses will greatly increase our understanding of the behavioural mechanisms that govern individual movement processes.

I further provided a new rigorous framework for movement models' robustness against varying temporal discretization. My new framework offers many new directions for research, from both a theoretical and an applied perspective. From the application's side, I suggest to particularly expand the concept of approximate robustness, because it will prove useful for data analysis at two stages. First, it will help to determine how strongly we should expect statistical inference results to depend on the assumed temporal resolution. Second, with help of the parameter transformation, it will help us to understand in which way inference results depend on temporal resolution. This will allow us to obtain better estimates of the parameters that shape movement processes.

In conclusion, my work offers new methods to better our understanding of movement behaviour both qualitatively and qualitatively. Ultimately, this does not only increase our general knowledge but is particularly valuable at a time where biodiversity loss has been identified as one of the major threats to the stable environmental state of the Earth during the last 10,000 years (Rockstrom *et al.*, 2009). Increased human land use, industrialization and landscape fragmentation pose major challenges to moving animals (Colchero *et al.*, 2010; Ito *et al.*, 2013; Bull *et al.*, 2013), while on the other hand moving animals may also be able to compensate effects of habitat loss on plant species (Mueller *et al.*, 2014). Therefore, an increased understanding of animal movement processes is one piece of the puzzle of how to maintain species' abundances and distributions.

# Bibliography


Aarts, G., Fieberg, J. & Matthiopoulos, J. (2011) Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods in Ecology and Evolution*, **3**, 177–187.

Abramowitz, M. & Stegun, I.A. (1964) *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.

Achim, A. & Kuruoglu, E.E. (2005) Image denoising using bivariate $\alpha$-stable distributions in the complex wavelet domain. *Signal Processing Letters, IEEE*, **12**, 17–20.

Altizer, S., Bartel, R. & Han, B.A. (2011) Animal migration and infectious disease risk. *Science*, **331**, 296–302.

Amano, T. & Katayama, N. (2009) Hierarchical movement decisions in predators: effects of foraging experience at more than one spatial and temporal scale. *Ecology*, **90**, 3536–3545.

Arthur, S.M., Manly, B.F.J., McDonald, L.L. & Garner, G.W. (1996) Assessing habitat selection when availability changes. *Ecology*, **77**, 215.

Asensio, N., Brockelman, W.Y., Malaivijitnond, S. & Reichard, U.H. (2011) Gibbon travel paths are goal oriented. *Animal Cognition*, **14**, 395–405.

Austin, D., Bowen, W.D., McMillan, J.I. & Boness, D.J. (2006) Stomach temperature telemetry reveals temporal patterns of foraging success in a free-ranging marine mammal. *Journal of Animal Ecology*, **75**, 408–420.

Avgar, T., Baker, J.A., Brown, G.S., Hagens, J., Kittle, A.M., Mallon, E.E., McGreer, M., Mosser, A., Newmaster, S.G., Patterson, B.R., Reid, D.E.B., Rodgers, A.R., Shuter, J., Street, G.M., Thompson, I., Turetsky, M., Wiebe, P.A. & Fryxell,

J. (2015) Space-use behavior of woodland caribou based on a cognitive movement model. *Journal of Animal Ecology*, doi: 10.1111/1365-2656.12357.

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

Balda, R.P. & Kamil, A.C. (1992) Long-term spatial memory in clark's nutcracker, *Nucifraga columbiana*. *Animal Behaviour*, **44**, 761–769.

Bar-Shai, N., Keasar, T. & Shmida, A. (2011) The use of numerical information by bees in foraging tasks. *Behavioral Ecology*, **22**, 317.

Barraquand, F. & Benhamou, S. (2008) Animal movements in heterogeneous landscapes: identifying profitable places and homogeneous movement bouts. *Ecology*, **89**, 3336–3348.

Barraquand, F., Inchausti, P. & Bretagnolle, V. (2009) Cognitive abilities of a central place forager interact with prey spatial aggregation in their effect on intake rate. *Animal Behaviour*, **78**, 505–514.

Bartumeus, F., Da Luz, M., Viswanathan, G. & Catalan, J. (2005) Animal search strategies: a quantitative random-walk analysis. *Ecology*, **86**, 3078–3087.

Benhamou, S. (2004) How to reliably estimate the tortuosity of an animal's path: straightness, sinuosity, or fractal dimension? *Journal of Theoretical Biology*, **229**, 209–220.

Benhamou, S. (2006) Detecting an orientation component in animal paths when the preferred direction is individual-dependent. *Ecology*, **87**, 518–528.

Benhamou, S. (2007) How many animals really do the Lévy walk? *Ecology*, **88**, 1962–1969.

Benhamou, S. (2013) Of scales and stationarity in animal movements. *Ecology Letters*, **17**, 261–272.

Benhamou, S., Sudre, J., Bourjea, J., Ciccione, S., De Santis, A. & Luschi, P. (2011) The role of geomagnetic cues in green turtle open sea navigation. *PLoS ONE*, **6**, e26672.

Bennett, A. (1996) Do animals have cognitive maps? *Journal of Experimental Biology*, **199**, 219.

Berbert, J.M. & Fagan, W.F. (2012) How the interplay between individual spatial memory and landscape persistence can generate population distribution patterns. *Ecological Complexity*, **12**, 1–12.

Bergman, E., Garrott, R., Creel, S., Borkowski, J., Jaffe, R. & Watson, F. (2006) Assessment of prey vulnerability through analysis of wolf movements and kill sites. *Ecological Applications*, **16**, 273–284.

Bestley, S., Patterson, T.A., Hindell, M.A. & Gunn, J.S. (2008) Feeding ecology of wild migratory tunas revealed by archival tag records of visceral warming. *Journal of Animal Ecology*, **77**, 1223–1233.

Börger, L., Dalziel, B.D. & Fryxell, J.M. (2008) Are there general mechanisms of animal home range behaviour? A review and prospects for future research. *Ecology Letters*, **11**, 637–650.

Box, G.E.P. (1980) Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society A*, **143**, 383–430.

Box, G.E.P. & Draper, N.R. (1987) *Empirical model-building and response surfaces*. John Wiley, New York, NY.

Boyce, M., Vernier, P., Nielsen, S. & Schmiegelow, F. (2002) Evaluating resource selection functions. *Ecological Modelling*, **157**, 281–300.

Boyer, D. & Walsh, P.D. (2010) Modelling the mobility of living organisms in heterogeneous landscapes: does memory improve foraging success? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **368**, 5645–5659.

Bradshaw, C.J.A., Sims, D.W. & Hays, G.C. (2007) Measurement error causes scale-dependent threshold erosion of biological signals in animal movement data. *Ecological Applications*, **17**, 628–638.

Breed, G.A., Costa, D.P., Goebel, M.E. & Robinson, P.W. (2011) Electronic tracking tag programming is critical to data collection for behavioral time-series analysis. *Ecosphere*, **2**, art10.

Breed, G.A., Costa, D.P., Jonsen, I.D., Robinson, P.W. & Mills-Flemming, J. (2012) State-space methods for more completely capturing behavioral dynamics from animal tracks. *Ecological Modelling*, **235–236**, 49–58.

Bridge, E.S., Thorup, K., Bowlin, M.S., Chilson, P.B., Diehl, R.H., Fléron, R.W., Hartl, P., Kays, R., Kelly, J.F., Robinson, W.D. & Wikelski, M. (2011) Technology on the move: recent and forthcoming innovations for tracking migratory birds. *BioScience*, **61**, 689–698.

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

Brown, D.D., LaPoint, S., Kays, R., Heidrich, W., Kümmeth, F. & Wikelski, M. (2012) Accelerometerinformed GPS telemetry: Reducing the tradeoff between resolution and longevity. *Wildlife Society Bulletin*, **36**, 139–146.

Brown, J.S., Laundré, J.W., Gurung, M. & Laundré, J.W. (1999) The ecology of fear: optimal foraging, game theory, and trophic interactions. *Journal of Mammalogy*, **80**, 385–399.

Buchmann, C.M., Schurr, F.M., Nathan, R. & Jeltsch, F. (2011) Movement upscaled - the importance of individual foraging movement for community response to habitat loss. *Wildlife Society Bulletin*, **35**, 436–445.

Bull, J.W., Suttle, K.B., Singh, N.J. & Milner-Gulland, E.J. (2013) Conservation when nothing stands still: moving targets and biodiversity offsets. *Frontiers in Ecology and the Environment*, **11**, 203–210.

Burke, D. & Fulham, B.J. (2003) An evolved spatial memory bias in a nectar-feeding bird? *Animal Behaviour*, **66**, 695–701.

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York, 2 edition.

Cagnacci, F., Boitani, L., Powell, R.A. & Boyce, M. (2010) Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. *Philosophical Transactions of the Royal Society B*, **365**, 2157–2162.

Calenge, C. (2006) The package "adehabitat" for the R software: A tool for the analysis of space and habitat use by animals. *Ecological Modelling*, **197**, 516–519.

Casella, G. & Berger, R.L. (2002) *Statistical inference*. Thomson Learning, Pacific Grove, CA, 2 edition.

Charnov, E.L. (1976) Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, **9**, 129–136.

Charnov, E.L., Orians, G.H. & Hyatt, K. (1976) Ecological implications of resource depression. *The American Naturalist*, **110**, 247–259.

Clayton, N.S. & Dickinson, A. (1998) Episodic-like memory during cache recovery by scrub jays. *Nature*, **395**, 272–274.

Codling, E. & Hill, N.A. (2005) Sampling rate effects on measurements of correlated and biased random walks. *Journal of Theoretical Biology*, **233**, 573.

Codling, E.A., Plank, M.J. & Benhamou, S. (2008) Random walk models in biology. *Journal of The Royal Society Interface*, **5**, 813–834.

Colchero, F., Conde, D.A., Manterola, C., Chávez, C., Rivera, A. & Ceballos, G. (2010) Jaguars on the move: modeling movement to mitigate fragmentation from road expansion in the Mayan Forest. *Animal Conservation*, **14**, 158–166.

Collett, M., Chittka, L. & Collett, T.S. (2013) Spatial memory in insect navigation. *Current Biology*, **23**, R789–R800.

Collett, M., Collett, T.S. & Srinivasan, M.V. (2006) Insect navigation: measuring travel distance across ground and through air. *Current Biology*, **16**, R887–R890.

Côrtes, M.C. & Uriarte, M. (2013) Integrating frugivory and animal movement: a review of the evidence and implications for scaling seed dispersal. *Biological Reviews*, **88**, 255–272.

Costa, D.P., Breed, G.A. & Robinson, P.W. (2012) New insights into pelagic migrations: implications for ecology and conservation. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 73–96.

Courbin, N., Fortin, D., Dussault, C., Fargeot, V. & Courtois, R. (2013) Multi-trophic resource selection function enlightens the behavioural game between wolves and their prey. *Journal of Animal Ecology*, **82**, 1062–1071.

Dalziel, B.D., Morales, J.M. & Fryxell, J.M. (2008) Fitting probability distributions to animal movement trajectories: using artificial neural networks to link distance, resources, and memory. *The American Naturalist*, **172**, 248–258.

Davies, N.B., Brooke, M.D.L. & Kacelnik, A. (1996) Recognition errors and probability of parasitism determine whether reed warblers should accept or reject mimetic cuckoo eggs. *Proceedings of the Royal Society B*, **263**, 925–931.

Davies, N.B. & Houston, A.I. (1981) Owners and satellites: the economics of territory defence in the pied wagtail, *Motacilla alba*. *The Journal of Animal Ecology*, pp. 157–180.

de la Giroday, H.M.C., Carroll, A.L., Lindgren, B.S. & Aukema, B.H. (2011) Incoming! Association of landscape features with dispersing mountain pine beetle populations during a range expansion event in western Canada. *Landscape Ecology*, **26**, 1097–1110.

DeCesare, N.J., Hebblewhite, M., Schmiegelow, F., Hervieux, D., McDermid, G.J., Neufeld, L., Bradley, M., Whittington, J., Smith, K.G., Morgantini, L.E., Wheatley, M. & Musiani, M. (2012) Transcending scale dependence in identifying habitat with resource selection functions. *Ecological Applications*, **22**, 1068–1083.

DeMars, C.A., Auger-Méthé, M., Schlägel, U.E. & Boutin, S. (2013) Inferring parturition and neonate survival from movement patterns of female ungulates: a case study using woodland caribou. *Ecology and Evolution*, **3**, 4149–4160.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., Weiskopf, D. & Weibel, R. (2015) Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology*, **3**, 19052.

Deneubourg, J.L., Goss, S., Franks, N. & Pasteels, J.M. (1989) The blind leading the blind: modeling chemically mediated army ant raid patterns. *Journal of insect behavior*, **2**, 719–725.

Edwards, A. (2011) Overturning conclusions of Lévy flight movement patterns by fishing boats and foraging animals. *Ecology*, **92**, 1247–1257.

Egevang, C., Stenhouse, I.J., Phillips, R.A., Petersen, A., Fox, J.W. & Silk, J.R.D. (2010) Tracking of arctic terns *Sterna paradisaea* reveals longest animal migration. *Proceedings of the National Academy of Sciences*, **107**, 2078–2081.

Fagan, W.F., Lewis, M.A., Auger-Méthé, M., Avgar, T., Benhamou, S., Breed, G., LaDage, L., Schlägel, U.E., Tang, W., Papastamatiou, Y.P., Forester, J. & Mueller, T. (2013) Spatial memory and animal movement. *Ecology Letters*, **16**, 1316–1329.

Fleming, C.H., Calabrese, J.M., Mueller, T., Olson, K.A., Leimgruber, P. & 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**, E154–E167.

Forester, J.D., Im, H. & Rathouz, P. (2009) Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. *Ecology*, **90**, 3554–3565.

Forester, J.D., Ives, A., Turner, M., Anderson, D., Fortin, D., Beyer, H., Smith, D. & Boyce, M. (2007) State-space models link elk movement patterns to landscape characteristics in Yellowstone National Park. *Ecological Monographs*, **77**, 285–299.

Fortin, D., Beyer, H., Boyce, M., Smith, D., Duchesne, T. & Mao, J. (2005) Wolves influence elk movements: behavior shapes a trophic cascade in Yellowstone National Park. *Ecology*, **86**, 1320–1330.

Frair, J.L., Merrill, E.H., Allen, J.R. & Boyce, M.S. (2007) Know thy enemy: experience affects elk translocation success in risky landscapes. *Journal of Wildlife Management*, **71**, 541–554.

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

Franke, A., Caelli, T., Kuzyk, G. & Hudson, R.J. (2006) Prediction of wolf (*Canis lupus*) kill-sites using hidden Markov models. *Ecological Modelling*, **197**, 237–246.

Fröhlich, M., Berger, A., Kramer-Schadt, S., Heckmann, I. & Martins, Q. (2012) Complementing GPS cluster analysis with activity data for studies of leopard (*Panthera pardus*) diet. *South African Journal of Wildlife Research*, **42**, 104–110.

Fronhofer, E.A., Hovestadt, T. & Poethke, H.J. (2013) From random walks to informed movement. *Oikos*, **122**, 857–866.

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

Gillies, C.S., Beyer, H.L. & St Clair, C.C. (2011) Fine-scale movement decisions of tropical forest birds in a fragmented landscape. *Ecological Applications*, **21**, 944–954.

Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010) A framework for community interactions under climate change. *Trends in Ecology & Evolution*, **25**, 325–331.

Giuggioli, L. & Kenkre, V.M. (2014) Consequences of animal interactions on their dynamics: emergence of home ranges and territoriality. *Movement Ecology*, **2**, 2–20.

Giuggioli, L., Potts, J.R. & Harris, S. (2011) Animal interactions and the emergence of territoriality. *PLoS Computational Biology*, **7**, e1002008.

Griffin, D.R. (1952) Bird navigation. *Biological Reviews*, **27**, 359–390.

Griffiths, D., Dickinson, A. & Clayton, N. (1999) Episodic memory: what can animals remember about their past? *Trends in Cognitive Sciences*, **3**, 74–80.

Grove, M. (2013) The evolution of spatial memory. *Mathematical Biosciences*, **242**, 25–32.

Hampel, F.R. (1971) A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, **42**, 1887–1896.

Hampel, F.R. (1986) *Robust Statistics: The approach based on influence functions*. Wiley, New York.

Hanks, E.M., Hooten, M.B., Johnson, D.S. & Sterling, J.T. (2011) Velocity-based movement modeling for individual and population level inference. *PLoS ONE*, **6**, e22795.

Haran, M. (2011) Gaussian random field models for spatial data. S. Brooks, A. Gelman, G.L. Jones & X.L. Meng, eds., *Handbook of Markov Chain Monte Carlo*, pp. 449–478. Chapman & Hall/CRC.

Hebblewhite, M. & Haydon, D. (2010) Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. *Philosophical Transactions of the Royal Society B*, **365**, 2303–2312.

Hebblewhite, M. & Merrill, E. (2008) Modelling wildlife–human relationships for social species with mixedeffects resource selection models. *Journal of Applied Ecology*, **45**, 834–844.

Hiebeler, D. (2000) Populations on fragmented landscapes with spatially structured heterogeneities: landscape generation and local dispersal. *Ecology*, **81**, 1629–1641.

Hilborn, R. & Mangel, M. (1997) *The ecological detective: confronting models with data*. Monographs in Population Biology. Princeton University Press, Princeton, NJ.

Holyoak, M., Casagrandi, R., Nathan, R., Revilla, E. & Spiegel, O. (2008) Trends and missing parts in the study of movement ecology. *Proceedings of the National Academy of Sciences*, **105**, 19060–19065.

Horne, J.S., Garton, E.O., Krone, S.M. & Lewis, J.S. (2007) Analyzing animal movements using Brownian bridges. *Ecology*, **88**, 2354–2363.

Huber, P.J. & Ronchetti, E.M. (2009) *Robust statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, N.J., 2 edition.

Ito, T.Y., Lhagvasuren, B., Tsunekawa, A., Shinoda, M., Takatsuki, S., Buuveibaatar, B. & Chimeddorj, B. (2013) Fragmentation of the habitat of wild ungulates by anthropogenic barriers in Mongolia. *PLoS ONE*, **8**, e56995.

James, A., Plank, M.J. & Edwards, A. (2011) Assessing Lévy walks as models of animal foraging. *Journal of The Royal Society Interface*, **8**, 1233–1247.

Janmaat, K.R.L., Byrne, R.W. & Zuberbühler, K. (2006) Evidence for a spatial memory of fruiting states of rainforest trees in wild mangabeys. *Forest Ecology and Management*, **72**, 797–807.

Janmaat, K.R.L. & Chancellor, R.L. (2010) Exploring new areas: how important is long-term spatial memory for mangabey (*Lophocebus albigena johnstonii*) foraging efficiency? *International Journal of Primatology*, **31**, 863–886.

Janmaat, K.R., Ban, S.D. & Boesch, C. (2013) Chimpanzees use long-term spatial memory to monitor large fruit trees and remember feeding experiences across seasons. *Animal Behaviour*, **86**, 1183–1205.

Janson, C.H. & Byrne, R. (2007) What wild primates know about resources: opening up the black box. *Animal Cognition*, **10**, 357–367.

Jedrzejewski, W., Schmidt, K., Theuerkauf, J., Jedrzejewska, B. & Okarma, H. (2001) Daily movements and territory use by radio-collared wolves (*Canis lupus*) in Bialowieza Primeval Forest in Poland. *Canadian Journal of Zoology*, **79**, 1993–2004.

Jerde, C.L. & Visscher, D.R. (2005) GPS measurement error influences on movement model parameterization. *Ecological Applications*, **15**, 806–810.

Johnson, C.J., Parker, K.L., Heard, D.C. & Gillingham, M.P. (2002) Movement parameters of ungulates and scale-specific responses to the environment. *Journal of Animal Ecology*, **71**, 225–235.

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

Jones, R.E. (1977) Movement patterns and egg distribution in cabbage butterflies. *The Journal of Animal Ecology*, **46**, 195–212.

Jonsen, I.D., Flemming, J.M. & Myers, R.A. (2005) Robust state-space modeling of animal movement data. *Ecology*, **86**, 2874–2880.

Kaiser, H. (1976) Quantitative description and simulation of stochastic behaviour in dragonflies (*Aeschna cyanea*, odonata). *Acta biotheoretica*, **25**, 163–210.

Kareiva, P.M. & Shigesada, N. (1983) Analyzing insect movement as a correlated random walk. *Oecologia*, **56**, 234–238.

Klafter, J., Zumofen, G. & Shlesinger, M.F. (1995) Lévy description of anomalous diffusion in dynamical systems. M. Shlesinger, G. Zaslavsky & U. Frisch, eds., *Lecture Notes in Physics*, pp. 196–215. Springer Berlin Heidelberg.

Klenke, A. (2008) *Probability theory: a comprehensive course*. Springer, London.

Knopff, K.H., Knopff, A.A., Kortello, A. & Boyce, M.S. (2010) Cougar kill rate and prey composition in a multiprey system. *The Journal of Wildlife Management*, **74**, 1435–1447.

Kotz, S., Kozubowski, T.J. & Podgórski, K. (2001) *The Laplace distribution and generalizations : a revisit with applications to communications, economics, engineering, and finance*. Birkhäuser, Boston.

Kranstauber, B., Cameron, A., Weinzerl, R., Fountain, T., Tilak, S., Wikelski, M. & Kays, R. (2011) The Movebank data model for animal tracking. *Environmental Modelling & Software*, **26**, 834–835.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D. & Morales, J.M. (2013) Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*, **93**, 2336–2342.

Latham, A.D.M., Latham, M.C., Boyce, M. & Boutin, S. (2011) Movement responses by wolves to industrial linear features and their effect on woodland caribou in northeastern Alberta. *Ecological Applications*, **21**, 2854–2865.

Latombe, G., Fortin, D. & Parrott, L. (2014) Spatio-temporal dynamics in the response of woodland caribou and moose to the passage of grey wolf. *Journal of Animal Ecology*, **83**, 185–198.

Laundré, J.W., Hernández, L. & Altendorf, K.B. (2001) Wolves, elk, and bison: reestablishing the "landscape of fear" in Yellowstone National Park, U.S.A. *Canadian Journal of Zoology*, **79**, 1401–1409.

Lele, S.R. (2009) A new method for estimation of resource selection probability function. *Journal of Wildlife Management*, **73**, 122–127.

Lele, S.R., Dennis, B. & Lutscher, F. (2007) Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters*, **10**, 551–563.

Lele, S.R. & Keim, J.L. (2006) Weighted distributions and estimation of resource selection probability functions. *Ecology*, **87**, 3021–3028.

Lele, S.R., Merrill, E.H., Keim, J. & Boyce, M.S. (2013) Selection, use, choice and occupancy: clarifying concepts in resource selection studies. *Journal of Animal Ecology*, **82**, 1183–1191.

Lele, S.R., Nadeem, K. & Schmuland, B. (2010) Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, **105**, 1617–1625.

Levin, S.A. (1992) The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology*, **73**, 1943–1967.

Lew, A.R. (2011) Looking beyond the boundaries: Time to put landmarks back on the cognitive map? *Psychological Bulletin*, **137**, 484–507.

Lewis, M.A. & Kareiva, P. (1993) Allee dynamics and the spread of invading organisms. *Theoretical Population Biology*, **43**, 141–158.

Lewis, M.A. & Murray, J.D. (1993) Modelling territoriality and wolf–deer interactions. *Nature*, **366**, 738–740.

Lewis, M., White, K. & Murray, J. (1997) Analysis of a model for wolf territories. *Journal of Mathematical Biology*, **35**, 749–774.

Liley, S. & Creel, S. (2007) What best explains vigilance in elk: characteristics of prey, predators, or the environment? *Behavioral Ecology*, **19**, 245–254.

Lima, S.L. (2002) Putting predators back into behavioral predator–prey interactions. *Trends in Ecology & Evolution*, **17**, 70–75.

Lima, S.L. & Dill, L.M. (1990) Behavioral decisions made under the risk of predation: a review and prospectus. *Canadian Journal of Zoology*, **68**, 619–640.

Link, W.A. & Barker, R.J. (2006) Model weights and the foundations of multimodel inference. *Ecology*, **87**, 2626–2635.

Lunn, D.J., Thomas, A., Best, N. & Spiegelhalter, D. (2000) WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–337.

Luschi, P., Sale, A., Mencacci, R., Hughes, G.R., Lutjeharms, J.R.E. & Papi, F. (2003) Current transport of leatherback sea turtles (*Dermochelys coriacea*) in the ocean. *Proceedings of the Royal Society B*, **270**, S129–S132.

Madan, D.B. & Seneta, E. (1990) The Variance Gamma (V.G.) model for share market returns. *The Journal of Business*, **63**, 511–524.

Manly, B.F., McDonald, L.L., Thomas, D.L., McDonald, T.L. & Erickson, W.P. (2002) *Resource selection by animals: statical design and analysis for field studies.* Kluwer Academic Publishers, Dordrecht, 2 edition.

Martin-Ordas, G., Haun, D., Colmenares, F. & Call, J. (2009) Keeping track of time: evidence for episodic-like memory in great apes. *Animal Cognition*, **13**, 331–340.

Masden, E.A., Reeve, R., Desholm, M., Fox, A.D., Furness, R.W. & Haydon, D.T. (2012) Assessing the impact of marine wind farms on birds through movement modelling. *Journal of The Royal Society Interface*, **9**, 2120–2130.

McClintock, B.T., Johnson, D.S., Hooten, M.B., Ver Hoef, J.M. & Morales, J.M. (2014) When to be discrete: the importance of time formulation in understanding animal movement. *Movement Ecology*, **2**, 334.

McClintock, B.T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B.J. & Morales, J.M. (2012) A general discrete-time modeling framework for animal movement using multistate random walks. *Ecological Monographs*, **82**, 335–349.

McClintock, B.T., Russell, D.J.F., Matthiopoulos, J. & King, R. (2013) Combining individual animal movement and ancillary biotelemetry data to investigate population-level activity budgets. *Ecology*, **94**, 838–849.

Mckenzie, H.W., Merrill, E.H., Spiteri, R.J. & Lewis, M.A. (2012) How linear features alter predator movement and the functional response. *Interface Focus*, **2**, 205–216.

McPhee, H.M., Webb, N.F. & Merrill, E.H. (2012) Hierarchical predation: wolf (*Canis lupus*) selection along hunt paths and at kill sites. *Canadian Journal of Zoology*, **90**, 555–563.

Mech, D.L. & Boitani, L., eds. (2006) *Wolves: behavior, ecology, and conservation*. University of Chicago Press, Chicago.

Mech, L.D. (1994) Buffer zones of territories of gray wolves as regions of intraspecific strife. *Journal of Mammalogy*, **75**, 199.

Menzel, R., De Marco, R. & Greggers, U. (2006) Spatial memory, navigation and dance behaviour in *Apis mellifera*. *Journal of Comparative Physiology A*, **192**, 889–903.

Milakovic, B., Parker, K.L., Gustine, D.D., Lay, R.J., Walker, A.B.D. & Gillingham, M.P. (2011) Habitat selection by a focal predator (*Canis lupus*) in a multiprey ecosystem of the northern Rockies. *Journal of Mammalogy*, **92**, 568–582.

Mills, K.J., Patterson, B.R. & Murray, D.L. (2006) Effects of variable sampling frequencies on GPS transmitter efficiency and estimated wolf home range size and movement distance. *Wildlife Society Bulletin*, **34**, 1463–1469.

Misund, O.A. (1993) Dynamics of moving masses: variability in packing density, shape, and size among herring, sprat, and saithe schools. *ICES Journal of Marine Science*, **50**, 145–160.

Moorcroft, P.R., Lewis, M.A. & Crabtree, R.L. (2006) Mechanistic home range models capture spatial patterns and dynamics of coyote territories in Yellowstone. *Proceedings of the Royal Society B*, **273**, 1651–1659.

Moorcroft, P.R. & Barnett, A. (2008) Mechanistic home range models and resource selection analysis: a reconciliation and unification. *Ecology*, **89**, 1112–1119.

Moorcroft, P.R. & Lewis, M.A. (2006) *Mechanistic home range analysis*. Princeton University Press, Princeton, N.J.

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

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

Mueller, T., Fagan, W.F. & Grimm, V. (2010) Integrating individual search and navigation behaviors in mechanistic movement models. *Theoretical Ecology*, **4**, 341–355.

Mueller, T., Lenz, J., Caprano, T., Fiedler, W. & Böhning-Gaese, K. (2014) Large frugivorous birds facilitate functional connectivity of fragmented landscapes. *Journal of Applied Ecology*, **51**, 684–692.

Mueller, T., Olson, K.A., Fuller, T.K., Schaller, G.B., Murray, M.G. & Leimgruber, P. (2008) In search of forage: predicting dynamic habitats of Mongolian gazelles using satellite-based estimates of vegetation productivity. *Journal of Applied Ecology*, **45**, 649–658.

Nadarajah, S. & Kotz, S. (2007) A truncated bivariate cauchy distribution. *Bulletin of the Malaysian Mathematical Sciences Society Second Series*, **30**, 185–193.

Nathan, R., Getz, W., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. & Smouse, P.E. (2008) A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, **105**, 19052–19059.

Neubert, M.G., Kot, M. & Lewis, M.A. (1995) Dispersal and pattern formation in a discrete-time predator-prey model. *Theoretical Population Biology*, **48**, 7–43.

Nolan, J.P. (1997) Numerical calculation of stable densities and distribution functions. *Communications in Statistics Stochastic Models*, **13**, 759–774.

Nolan, J.P. (2013) Multivariate elliptically contoured stable distributions: theory and estimation. *Computational Statistics*, **28**, 2067–2089.

Nouvellet, P., Bacon, J.P. & Waxman, D. (2009) Fundamental insights into the random movement of animals from a single distancerelated statistic. *The American Naturalist*, **174**, 506–514.

O'Keefe, J. & Nadel, L. (1978) *The hippocampus as a cognitive map*. Clarendon Press, Oxford.

Owen-Smith, N., Fryxell, J. & Merrill, E. (2010) Foraging theory upscaled: the behavioural ecology of herbivore movement. *Philosophical Transactions of the Royal Society B*, **365**, 2267–2278.

Patlak, C.S. (1953) Random walk with persistence and external bias. *The Bulletin of Mathematical Biophysics*, **15**, 311–338.

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

Patterson, T.A. & Hartmann, K. (2011) Designing satellite tagging studies: estimating and optimizing data recovery. *Fisheries Oceanography*, **20**, 449–461.

Patterson, T.A., McConnell, B.J., Fedak, M.A., Bravington, M.V. & Hindell, M.A. (2010) Using GPS data to evaluate the accuracy of state-space methods for correction of Argos satellite telemetry error. *Ecology*, **91**, 273–285.

Pépin, D., Adrados, C., Mann, C. & Janeau, G. (2004) Assessing real daily distance traveled by ungulates using differential GPS locations. *Journal of Mammalogy*, **85**, 774–780.

Peters, R.P. (1979) Mental maps in wolf territories. E. Klinghammer, ed., *The Behavior and ecology of wolves*, pp. 119–152. Garland STPM Press, New York.

Peters, R.P. & Mech, L.D. (1975) Scent-marking in wolves: Radio-tracking of wolf packs has provided definite evidence that olfactory sign is used for territory maintenance and may serve for other forms of communication within the pack as well. *American Scientist*, **63**, 628–637.

Piessens, R. (2000) Hankel Transform. A.D. Poularikas, ed., *The Transforms and Applications Handbook*. CRC Press, Boca Raton.

Plank, M., Auger-Méthé, M. & Codling, E. (2013) Lévy or not? Analysing positional data from animal movement paths. M.A. Lewis, P.K. Maini & S.V. Petrovskii, eds., *Dispersal, individual movement and spatial ecology*, pp. 33–52. Springer, Berlin.

Plummer, M. (2013) *rjags: Bayesian graphical models using MCMC*. R package version 3-11. URL http://CRAN.R-project.org/package=rjags

Ponciano, J.M., Taper, M.L., Dennis, B. & Lele, S.R. (2009) Hierarchical models in ecology: confidence intervals, hypothesis testing, and model selection using data cloning. *Ecology*, **90**, 356–362.

Postlethwaite, C.M. & Dennis, T.E. (2013) Effects of temporal resolution on an inferential model of animal movement. *PLoS ONE*, **8**, e57640.

Potts, J.R., Bastille-Rousseau, G., Murray, D.L., Schaefer, J.A. & Lewis, M.A. (2014) Predicting local and non-local effects of resources on animal space use using a mechanistic step selection model. *Methods in Ecology and Evolution*, **5**, 253–262.

Potts, J.R., Harris, S. & Giuggioli, L. (2013) Quantifying behavioral changes in territorial animals caused by sudden population declines. *The American Naturalist*, **182**, E73–E82.

Potts, J.R. & Lewis, M.A. (2014) How do animal territories form and change? Lessons from 20 years of mechanistic modelling. *Proceedings of the Royal Society B*, **281**, 20140231.

Proffitt, K.M., Grigg, J.L., Hamlin, K.L. & Garrott, R.A. (2009) Contrasting effects of wolves and human hunters on elk behavioral responses to predation risk. *The Journal of Wildlife Management*, **73**, 345–356.

Pyke, G.H. (1984) Optimal foraging theory: a critical review. *Annual Review of Ecology and Systematics*, **15**, 523–575.

Pyke, G.H. (2015) Understanding movements of organisms: it's time to abandon the Lévy foraging hypothesis. *Methods in Ecology and Evolution*, **6**, 1–16.

R Core team (2013) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rasmussen, K., Palacios, D.M., Calambokidis, J., Saborío, M.T., Rosa, L.D., Secchi, E.R., Steiger, G.H., Allen, J.M. & Stone, G.S. (2007) Southern Hemisphere humpback whales wintering off Central America: insights from water temperature into the longest mammalian migration. *Biology Letters*, **3**, 302–305.

Recio, M.R., Mathieu, R., Denys, P., Sirguey, P. & Seddon, P.J. (2011) Lightweight GPS-tags, one giant leap for wildlife tracking? An assessment approach. *PLoS ONE*, **6**, e28225.

Reid, C.R., Latty, T., Dussutour, A. & Beekman, M. (2012) Slime mold uses an externalized spatial "memory" to navigate in complex environments. *Proceedings of the National Academy of Sciences*, **109**, 17490–17494.

Reynolds, A.M. & Rhodes, C.J. (2009) The Lévy flight paradigm: random search patterns and mechanisms. *Ecology*, **90**, 877–887.

Rhodes, J.R., McAlpine, C.A., Lunney, D. & Possingham, H.P. (2005) A spatially explicit habitat selection model incorporating home range behavior. *Ecology*, **86**, 1199–1205.

Ripple, W.J. & Beschta, R.L. (2012) Trophic cascades in Yellowstone: The first 15 years after wolf reintroduction. *Biological Conservation*, **145**, 205–213.

Robert, C.P. (1993) Prior feedback: Bayesian tools for maximum likelihood estimation. *Computational Statistics*, **8**, 279–294.

Robert, C.P. & Casella, G. (2000) *Monte Carlo statistical methods*. Springer texts in statistics. Springer, New York, corrected 2. print. edition.

Robinson, W.D., Bowlin, M.S., Bisson, I., Shamoun-Baranes, J., Thorup, K., Diehl, R.H., Kunz, T.H., Mabey, S. & Winkler, D.W. (2009) Integrating concepts and technologies to advance the study of bird migration. *Frontiers in Ecology and the Environment*, **8**, 354–361.

Rockstrom, J., Steffen, W., Noone, K., Persson, A., Chapin, F.S., Lambin, E.F., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., de Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sorlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P. & Foley, J.A. (2009) A safe operating space for humanity. *Nature*, **461**, 472–475.

Rosser, G., Fletcher, A.G., Maini, P.K. & Baker, R.E. (2013) The effect of sampling rate on observed statistics in a correlated random walk. *Journal of The Royal Society Interface*, **10**, 20130273.

Rowcliffe, M.J., Carbone, C., Kays, R., Kranstauber, B. & Jansen, P.A. (2012) Bias in estimating animal travel distance: the effect of sampling frequency. *Methods in Ecology and Evolution*, **3**, 653–662.

Rutz, C. & Hays, G.C. (2009) New frontiers in biologging science. *Biology Letters*, **5**, 289–292.

Ryan, P.G., Petersen, S.L., Peters, G. & Gremillet, D. (2004) GPS tracking a marine predator: the effects of precision, resolution and sampling rate on foraging tracks of African Penguins. *Marine Biology*, **145**.

Samorodnitsky, G.T.M.S. (1994) *Stable non-Gaussian random processes: stochastic models with infinite variance*. Chapman & Hall, New York.

Sawyer, H., Kauffman, M., Nielson, R. & Horne, J. (2009) Identifying and prioritizing ungulate migration routes for landscape-level conservation. *Ecological Applications*, **19**, 2016–2025.

Schick, R.S., Loarie, S.R., Colchero, F., Best, B.D., Boustany, A., Conde, D.A., Halpin, P.N., Joppa, L.N., McClellan, C.M. & Clark, J.S. (2008) Understanding movement data and movement processes: current and emerging directions. *Ecology Letters*, **11**, 1338–1350.

Schlägel, U.E. & Lewis, M.A. (2014) Detecting effects of spatial memory and dynamic information on animal movement decisions. *Methods in Ecology and Evolution*, **5**, 1236–1246.

Schlather, M., Menck, P., Singleton, R., Pfaff, B. & R Core team (2013) *RandomFields: Simulation and Analysis of Random Fields*. R package version 2.0.66. URL http://CRAN.R-project.org/package=RandomFields

Seneta, E. (2004) Fitting the variance-gamma model to financial data. *Journal of Applied Probability*, **41**, 177–187.

Sharp, N.C.C. (1997) Timed running speed of a cheetah (*Acinonyx jubatus*). *Journal of Zoology*, **241**, 493–494.

Shettleworth, S.J. (2010) *Cognition, evolution, and behavior*. Oxford University Press, Oxford; New York, 2 edition.

Skellam, J.G. (1951) Random dispersal in theoretical populations. *Biometrika*, **38**, 196.

Smouse, P.E., Focardi, S., Moorcroft, P.R., Kie, J.G., Forester, J.D. & Morales, J.M. (2010) Stochastic modelling of animal movement. *Philosophical Transactions of the Royal Society B*, **365**, 2201–2211.

Solymos, P. (2010) dclone: Data Cloning in R. *The R Journal*, **2**, 29–37.

Spencer, W.D. (2012) Home ranges and the value of spatial information. *Journal of Mammalogy*, **93**, 929–947.

Squires, J.R., DeCesare, N.J., Olson, L.E., Kolbe, J.A., Hebblewhite, M. & Parks, S.A. (2013) Combining resource selection and movement behavior to predict corridors for Canada lynx at their southern range periphery. *Biological Conservation*, **157**, 187–195.

Stan Development Team (2014) Stan: A C++ Library for Probability and Sampling, Version 2.5.0. URL http://mc-stan.org/

Steutel, F.W. & Van Harn, K. (2004) *Infinite divisibility of probability distributions on the real line*. Marcel Dekker, New York.

Sulikowski, D. & Burke, D. (2011) Movement and memory: different cognitive strategies are used to search for resources with different natural distributions. *Behavioral Ecology and Sociobiology*, **65**, 621–631.

Sutherland, W.J., Freckleton, R.P., Godfray, H.C.J., Beissinger, S.R., Benton, T., Cameron, D.D., Carmel, Y., Coomes, D.A., Coulson, T. & Emmerson, M.C. (2013) Identification of 100 fundamental ecological questions. *Journal of Ecology*, **101**, 58–67.

Tanferna, A., López-Jiménez, L., Blas, J., Hiraldo, F. & Sergio, F. (2012) Different location sampling frequencies by satellite tags yield different estimates of migration performance: pooling data requires a common protocol. *PLoS ONE*, **7**, e49659.

Thorup, K., Bisson, I.A., Bowlin, M.S., Holland, R.A., Wingfield, J.C., Ramenofsky, M. & Wikelski, M. (2007) Evidence for a navigational map stretching across the continental U.S. in a migratory songbird. *Proceedings of the National Academy of Sciences*, **104**, 18115–18119.

Tolman, E.C. (1948) Cognitive maps in rats and men. *Psychological Review*, **55**, 189–208.

Tomkiewicz, S.M., Fuller, M.R., Kie, J.G. & Bates, K.K. (2010) Global positioning system and associated technologies in animal behaviour and ecological research. *Philosophical Transactions of the Royal Society B*, **365**, 2163–2176.

Tracey, J.A., Zhu, J. & Crooks, K.R. (2010) Modeling and inference of animal movement using artificial neural networks. *Environmental and Ecological Statistics*, **18**, 393–410.

Tsoar, A., Nathan, R., Bartan, Y., Vyssotski, A., Dell'Omo, G. & Ulanovsky, N. (2011) Large-scale navigational map in a mammal. *Proceedings of the National Academy of Sciences*, **108**, E718–E724.

Tucker, V.A. (1998) Gliding flight: speed and acceleration of ideal falcons during diving and pull out. *The Journal of Experimental Biology*, **201**, 403–414.

Turchin, P. (1998) *Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants*. Sinauer Associates, Sunderland, Mass.

Van Moorter, B., Visscher, D., Benhamou, S., Borger, L., Boyce, M. & Gaillard, J. (2009) Memory keeps you at home: a mechanistic model for home range emergence. *Oikos*, **118**, 641–652.

Viswanathan, G.M., Buldyrev, S.V., Havlin, S., da Luz, M.G.E., Raposo, E.P. & Stanley, H.E. (1999) Optimizing the success of random searches. *Nature*, **401**, 911–914.

Webb, N. (2009) *Density, demography, and functional response of a harvested wolf population in west-central Alberta, Canada*. Ph.D. thesis, University of Alberta, Edmonton, AB.

Webb, N.F., Hebblewhite, M. & Merrill, E.H. (2008) Statistical methods for identifying wolf kill sites using global positioning system locations. *Journal of Wildlife Management*, **72**, 798–807.

Webb, N.F. & Merrill, E.H. (2012) Simulating carnivore movements: An occupancy-abundance relationship for surveying wolves. *Wildlife Society Bulletin*, **36**, 240–247.

Weron, R. (1996) On the Chambers-Mallows-Stuck method for simulating skewed stable random variables. *Statistics & Probability Letters*, **28**, 165–171.

White, A., Watt, A.D., Hails, R.S. & Hartley, S.E. (2000) Patterns of spread in insect-pathogen systems: the importance of pathogen dispersal. *Oikos*, **89**, 137–145.

White, E.P., Enquist, B.J. & Green, J.L. (2008) On estimating the exponent of power-law frequency distributions. *Ecology*, **89**, 905–912.

Wiens, D.P. (2000) Bias constrained minimax robust designs for misspecified regression models. N. Balakrishnan, V.B. Melas & S. Ermakov, eds., *Statistics for Industry and Technology*, pp. 117–133. Birkhäuser, Boston, MA.

Wiens, D.P. & Zhou, J. (1996) Minimax regression designs for approximately linear models with autocorrelated errors. *Journal of Statistical Planning and Inference*, **55**, 95–106.

Wilcox, R.R. (2012) *Introduction to robust estimation and hypothesis testing*. Academic Press, Boston, 3 edition.

Wittlinger, M., Wehner, R. & Wolf, H. (2006) The ant odometer: stepping on stilts and stumps. *Science*, **312**, 1965–1967.

Wuertz, D., Maechler, M. & Rmetrics core team members (2013) *stabledist: stable distribution functions*. R package version 0.6-6. URL http://CRAN.R-project.org/package=stabledist

Yackulic, C.B., Blake, S., Deem, S., Kock, M. & Uriarte, M. (2011) One size does not fit all: flexible models are required to understand animal movement across scales. *Journal of Animal Ecology*, **80**, 1088–1096.

Yeap, W. (2014) On egocentric and allocentric maps. C. Freksa, B. Nebel, M. Hegarty & T. Barkowsky, eds., *Lecture Notes in Computer Science*, pp. 62–75. Springer International Publishing, Switzerland.

Zub, K., Theuerkauf, J., Jeîdrzejewski, W., Jeîdrzejewska, B., Schmidt, K. & Kowalczyk, R. (2003) Wolf pack territory marking in the Bialowieçza Primeval Forest (Poland). *Behaviour*, **140**, 635–648.

# Appendix A

# Supplemental methods for Chapter 2

## A.1 Data cloning and MCMC in simulation analysis

For all model fits, I used data cloning. Data cloning uses Markov Chain Monte Carlo (MCMC) methods, which calculate a posterior distribution for the model parameters, given the data. This technique is usually employed in Bayesian statistical inference, however, the resulting parameter estimates from data cloning approximate the corresponding maximum likelihood estimates (MLE). This is achieved by applying the Bayesian framework to $K$ copies of the data, which are referred to as clones. Alternatively, the procedure can be viewed as a series of Bayesian updates applied to the same data, each time using the posterior distribution from the previous update as new prior distribution (Robert, 1993). After a movement trajectory has been cloned $K$ times, Bayesian parameter estimation, here via MCMC, is performed on this augmented data. The results of this procedure lead to parameter estimates in the more conventional style of frequentist inference, namely maximum likelihood estimates. However, an important factor to achieve this is a sufficiently large number of clones. If $K$ is large enough, the posterior distribution for the parameters is approximately Normal with mean at the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ of the original (i.e. uncloned) data and with variance $\frac{1}{K} I^{-1}(\hat{\boldsymbol{\theta}}_{\text{MLE}})$, where $I$ is the Fisher information of the original data (Lele *et al.*, 2010). This means that if we choose $K$ large enough, the sample mean of the MCMC is approximately the MLE of the original uncloned trajectory and if we multiply the sample variance by $K$, we obtain an approximation of the inverse Fisher information

(also termed information number; Casella & Berger, 2002). The inverse of the Fisher information is the asymptotic variance of a maximum likelihood estimate and it can be used to calculate Wald-type confidence intervals.

In my analysis, I used $K = 15$ clones. To confirm that this number of clones was sufficient to obtain a good Normal approximation of the posterior distributions and good approximations of the maximum likelihood estimates, I performed test runs with the most complex combination model. I selected a combination trajectory both from the main data set and the supplemental data and iteratively fitted the combination model with increasing number of clones. For each run, I inspected the three diagnostic measures described in Lele *et al.* (2010) and Solymos (2010) (`lambda.max, ms.error, r.squared`). These diagnostics assess whether the Normal approximation of the posterior distributions and the approximation of the sample mean to the MLE are adequate, which is the case if the diagnostics converge to zero. I found that all three diagnostics converged to zero for my test fits, and that they were all close to zero ($< 0.05$) for $K = 15$.

For the MCMC, I used two parallel chains, each running for 7500 iterations, of which I discarded an initialization and burn-in period of 3500 iterations. To assess whether this was sufficient to obtain good mixing properties of the chains and convergence to the stationary distribution, I inspected the Markov chains visually and calculated the *potential scale reduction factor* $\hat{R}$ (Brooks & Gelman, 1998) for each parameter. Using these amounts of MCMC iterations, I obtained good mixing and convergence in matching model fits. In non-matching model fits (model and simulated trajectory mechanism did not match), mixing and convergence problems occurred. To ensure that these problems did not simply occur because of an insufficient number of MCMC iterations, I continued to run some of the non-mixing/non-converging MCMCs for up to 8000 additional iterations. In none of these cases I found that more iterations improved mixing or convergence.

For further analysis, I calculated data cloning estimability diagnostic for selected trajectories. This requires a series of model fits with increasing number of clones, for which I used the functions `dc.fit` and `dc.parfit` from the R data cloning package (Solymos, 2010). I chose number of clones $K = 1, 5, 10, 15$. Because of the high computational needs of the model fits, especially for the most complex combination model, I refrained from increasing the number of clones further. However, as additional test I also examined one- and two-dimensional slices of the corresponding likelihood functions and found that these tests always lead to the same conclusions; see Appendix B. If in a model fit parameters are estimable, their variances should decrease

with increasing number of clones. In particular, the largest eigenvalue of the posterior variance, `lambda.max`, should decrease with rate $\frac{1}{K}$ (Lele *et al.*, 2010; Solymos, 2010). Standardized by its value for the uncloned data, it should converge to zero as $\frac{1}{K}$.

## A.2 Missed observations

In general, there are several approaches how to deal with missed observations in a trajectory statistically. The easiest case is when locations, or steps, are modelled as being independent from each other. However, for models that include autocorrelation, we cannot simply ignore the dependency structure. A possibility is to use some technique of interpolation. Alternatively, we can divide the trajectory into chunks of available data and condition the likelihood function on the first available observation in each chunk. In a correlated random walk, we need three consecutive locations to define one step probability. Therefore, missed locations effectively lead to even larger gaps in the likelihood function.

To avoid any loss of data, we can use the full likelihood based on the entire trajectory $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = (\boldsymbol{x}_{\text{observed}}, \boldsymbol{x}_{\text{missed}})$ and integrate over all missed observations,

$$L(\boldsymbol{\theta}|(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)) = L(\boldsymbol{\theta}|(\boldsymbol{x}_{\text{observed}}, \boldsymbol{x}_{\text{missed}})) = \int p(\boldsymbol{x}_{\text{observed}}, \boldsymbol{x}_{\text{miss}}|\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}_{\text{miss}}. \quad \text{(A.1)}$$

This has the advantage that all original dependencies between locations can be preserved and no information is lost. Calculation of the possibly high-dimensional integral is problematic in common frequentist methods that require optimization of the likelihood function. However, MCMC techniques (and therefore data cloning) circumvent this problem and at the same time provide estimates for the missed variables.

My model is formulated based entirely on locations (intermediate quantities such as step length and bearing are calculated within the model formulation), and therefore implementation of this method is, in principle, straightforward: in the MCMC, missed locations are treated as parameters and their step probabilities serve as priors. I used JAGS for model fitting, which was capable to perform this and to produce converging posterior distributions for missed locations. However, this came at the cost of very high computational needs (both memory requirements and computation time).

The memory model requires reconstruction of time since last visit $\boldsymbol{m}$. For a missed location at time $t$, we accordingly miss $m_t$. Because $m_t$ is a function of $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t-1}$, just as step length and bearing, I could estimate $m_t$ within the model fitting procedure. However, $m_t$ is a high-dimensional variable for each time step and due to computational

restrictions I treated $\boldsymbol{m}$ as known covariate. Therefore, if the location $\boldsymbol{x}_t$ was missing, I did not update $m_t$ for this time step and set $m_t = m_{t-1}$. At the next time step, I updated $m_{t+1}$ via the usual formula, but based on $m_{t-1}$ and $\boldsymbol{x}_{t-1}$. To account for the longer time, I increased the distance $\delta$ from the straight line path($\boldsymbol{x}_{t-1} \to \boldsymbol{x}_{t+1}$), in which locations are considered as visited. If more than one location was missed in a row, I proceeded similarly, starting to update time since last visit at the next available location.

To perform model selection for a completely observed trajectory, it is possible with my models to calculated the likelihood functions and thus BIC. With missed locations, this becomes computationally much more complex due to the integration; compare equation (A.1). To avoid this, we can, as an approximation, instead use estimates of missed locations. Because I treated missed locations as parameters, I obtained posterior distributions and estimates for them. I used these estimates to calculate the likelihood function. A more sophisticated method has been proposed by Ponciano *et al.* (2009). Their method circumvents the problem of integration and uses data cloning itself to obtain estimates of likelihood ratios, which can then be used for AIC or BIC.

## A.3 Simulation of landscapes

I modelled the continuous valued environmental covariate as a Gaussian random field (Haran, 2011). A Gaussian random field is a multivariate Gaussian random variable, indexed by space. Here, the random variable is the resource $r_1$ for each location in the spatial domain, $\{r_1(\boldsymbol{x})\}_{\boldsymbol{x}\in\Omega}$. The covariance between resource values at any two locations $\boldsymbol{x}$ and $\boldsymbol{y}$ is a function of the distance between the locations, so that values of nearby locations are stronger correlated than values of locations that are far apart. I chose the exponential form for the covariance function, $\operatorname{cov}\big(r_1(\boldsymbol{x}), r_1(\boldsymbol{y})\big) = \exp(-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|}{\sigma})$, where $\sigma$ determines the rate at which locations cease to be correlated. I varied $\sigma$ among different landscapes. To simulate such landscapes, I used the R package RandomFields (Schlather *et al.*, 2013).

To generate correlated landscapes of binary variables I used the method and C code provided by Hiebeler (2000). Each landscape is represented by two quantities: $p_0$, the overall proportion of type 0 cells, and $q_{00}$, the probability that a neighbour of a type 0 cell is also of type 0. If $q_{00}$ is high, the landscape is strongly clustered, and vice versa. For my landscapes, I varied both $p_0$, and their degree of clustering, $q_{00}$.

The five landscape pairs I used for my simulations are depicted in Figure A.1

## A.4 Supplemental data

In the main text of the paper, I analyzed a simulated data set of 20 trajectories. This data set was generated using realistic parameter values. However, to test my method in even more scenarios, I generated two additional sets of 20 trajectories, which I refer to as "data set 2" and "data set 3" to separate them easier from the main data set. For these data, I used parameter sets in which I included values that I considered to be potentially more difficult to estimate from data.

For data set 2, I chose relatively small resource selection parameters $\alpha_{\text{res}}$ $\beta_1$, $\beta_2$ and very small interaction parameters $\gamma_1$, $\gamma_2$, which means that I simulated weak effects of the resources. The parameters $\alpha_{\text{mem}}$ and $\beta_{\text{mem}}$ that regulate the influence of time since last visit were chosen so that returns to locations were possible again after short durations of absence. This means that the effect of time since last visit is relatively weak. The parameter values were

| | | | | |
|---|---|---|---|---|
| parameter set 2: | $\kappa = 4$ | $\alpha_{\text{res}} = -0.2$ | $\beta_1 = 0.5$ | $\gamma_1 = 0.008$ |
| | $\lambda = 0.9$ | $\alpha_{\text{mem}} = -3$ | $\beta_2 = 0.8$ | $\gamma_2 = 0.005$ |
| | $\rho = 1.2$ | $\alpha_{\text{com}} = -3.2$ | $\beta_{\text{mem}} = 0.04$ | |

For data set 3, I set one of the interaction parameters in the combination model to zero, so that an interaction between resource values and time since last visit was only present for the binary variable $r_2$. To distinguish this data set further from the main set, I chose $\beta_1 < 0$, so that resource variable $r_1$ had an opposite effect compared to the other data sets. All other parameters are again chosen to be realistic, but different from previous values.

| | | | | |
|---|---|---|---|---|
| parameter set 3: | $\kappa = 4.5$ | $\alpha_{\text{res}} = 0.8$ | $\beta_1 = -1.5$ | $\gamma_1 = 0$ |
| | $\lambda = 1.3$ | $\alpha_{\text{mem}} = -5$ | $\beta_2 = 2.5$ | $\gamma_2 = 0.01$ |
| | $\rho = 1.5$ | $\alpha_{\text{com}} = -5.8$ | $\beta_{\text{mem}} = 0.05$ | |

The simulated resource landscapes were the same as for the main data set. I performed the same analysis on the supplemental data as on the main data and I obtained 160 model fits.

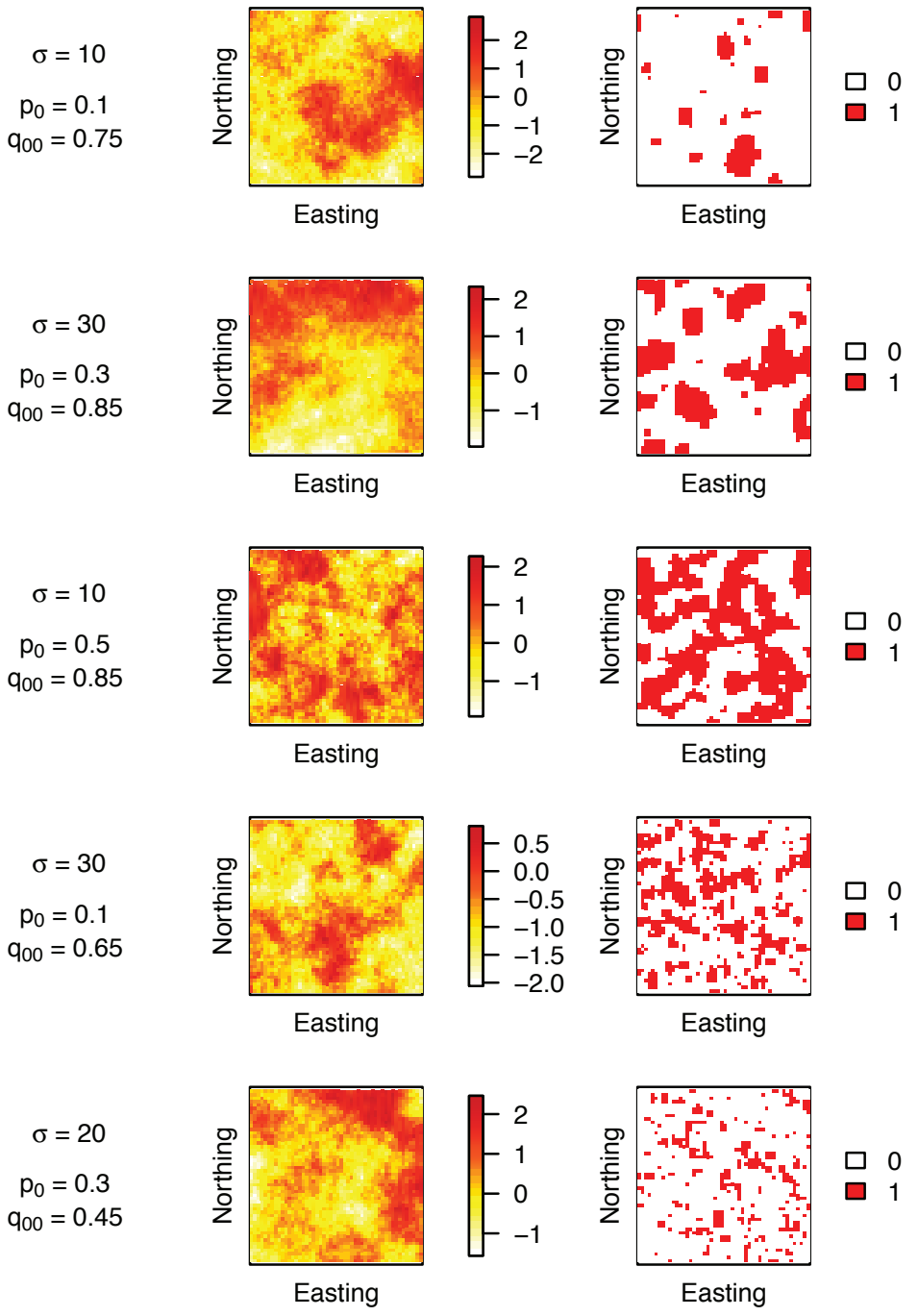



**Figure A.1.** Pairs of simulated landscapes (row-wise). The left side shows the continuous valued resource $r_1$, the left side shows the binary variable $r_2$. Parameter values used to simulate the landscapes are given for every landscape pair.

# Appendix B

# Supplemental results and estimability analyses for Chapter 2

## B.1 Results

### B.1.1 Supplemental data sets

Considering the 160 model fits, 82% had potential scale reduction factor $\hat{R} \leq 1.1$ for all parameters, which means that for those model fits, MCMC runs for all parameters mixed well and converged. If convergence and mixing problems occurred, these were cases where parameters were inapplicable to the analyzed trajectory (Figure B.1).

In contrast to the main data set, there was one instance in the supplemental data for which $\hat{R} > 1.1$ for several parameters in a matching model fit in data set 2. This was a combination trajectory fitted with the matching combination model. I continued to run the Markov chains for more iterations, however, the chains' behaviour remained the same. I therefore looked closer into this model fit, and ultimately ran an entirely new MCMC for this trajectory to calculate estimability diagnostics. In this second run, the parallel chains mixed and converged well and no estimability issues were found. For more details and discussion about this, see section B.2.

Model selection via BIC was able to correctly identify true underlying models for all but one trajectory (Figure B.1). For the fourth resource trajectory of data set 3, BIC was lowest for the null model, followed by the resource model. When I looked closer at the results of the fit with the resource model, I found that selection parameters $\alpha_{\text{res}}$ and $\beta_1$ had very large confidence intervals, and the estimate of the intercept $\alpha_{\text{res}}$ was high. High values of the intercept effectively result in a constant weighting function, thereby mimicking the null model. I discuss this model fit further in section B.2. Additionally,

there was one matching combination model fit that did not mix properly in the first MCMC run (compare previous paragraph), and for which I therefore did not calculate BIC. This lead to the memory model being selected as best model for a combination trajectory. However, when I ran a new MCMC with the combination model, the chains converged and BIC was lowest, followed in order by the memory model ($\Delta$BIC = 178), the resource model ($\Delta$BIC = 676) and the null model ($\Delta$BIC = 800). In the following, I used results from the second MCMC run for the fourth combination trajectory in data set 2.

For the hypothesis test based on confidence intervals, I obtained 139 estimates of selection parameters (I only considered estimates from convergent and well mixing Markov chains). Of these, 69 corresponded to true underlying effects. When I analyzed confidence intervals as to whether they excluded zero and thus suggested covariate effects, I obtained a Type I error rate of 0.01 (a trajectory was simulated without effect, but confidence intervals detect an effect) and a Type II error rate of 0.14 (a trajectory was simulated with an effect, which was not detected). When I pooled supplemental and main data, I obtained a Type I error rate of 0.04, which is close to the expected amount if we use 95% confidence intervals (which corresponds to a 5%-level hypothesis test). For the pooled data, the Type II error rate was 0.09. Hence, overall the hypothesis test gives expected results that include errors, while the model selection via BIC performs better and reliably identifies trajectories' true underlying mechanisms.

Most parameter estimates of matching model fits agreed well with true underlying values. As expected, 95% confidence intervals (n=230) included the true value 0.95% of the time. In data set 2, there was one resource trajectory, for which the estimate of $\alpha_{\text{res}}$ was far away from the true value ($\hat{\alpha}_{\text{res}} = -9.7$, true value was -0.2) and the standard error was very large (sd=21.7). I looked into this further and calculated likelihood slices and data cloning estimability diagnostics. From these, I concluded that there was an estimability problem for $\alpha_{\text{res}}$, while the other parameters were well behaved; for details see section B.2. Therefore, I excluded this estimate of $\alpha_{\text{res}}$. All remaining estimates in data set 2 were balanced around and generally close to their true values (Figure B.2a). In Figure B.2a, I plotted estimates for $\alpha_{\text{res}}$ and $\gamma_2$ separately, using the original unscaled values, because their standardized confidence intervals were larger than for the other parameters. Standardization is sensitive to the size of the standardization constant and may be problematic here, because the true values of $\alpha_{\text{res}}$ and $\gamma_2$ are small, and division by values close to zero results in large values. The unscaled results for $\alpha_{\text{res}}$ and $\gamma_2$ look reasonable (Figure B.2, smaller panels).

In data set 3, there were a few more estimates with large confidence intervals (Figure B.3b), particularly for $\beta_2$ and $\gamma_2$, which I have therefore plotted in separate panels in their original scale. I suspect that these large confidence intervals are due to estimability problems; see also discussion in section B.2. I also plotted $\gamma_1$ separately, because its true value was zero and therefore could not be standardized. All estimates of $\gamma_1$ were close to zero and all confidence intervals overlapped zero. Therefore, the model was able to correctly identify the lacking effect of the interaction parameter.

### B.1.2 Missed observations

For the combination trajectory with missed locations, I performed a matching model fit. I compared parameter estimates and their 95% confidence intervals for the trajectory with missed locations and the corresponding complete trajectory. Parameter estimates for the combination trajectory with missed locations agreed well with true values and were similar to results for the complete trajectory (Figure B.4). Estimates of selection parameters tended to be slightly lower for the incomplete trajectory, but standardized values never deviated by more than 0.15. Parameters $\kappa, \lambda$ of the movement kernel describing step lengths (shape and scale of Weibull distribution) are slightly higher for the incomplete trajectory, resulting in a mean step length of 5.29 compared to 5.0 for the complete trajectory.

## B.2 Convergence and estimability issues

### B.2.1 Estimability in cases of non-convergence

In each data set, about 18-20% of model fits contained one or more parameters, for which $\hat{R}$ was larger than 1.1, indicating non-convergence or non-mixing of the parallel chains. In many cases, I continued to run these chains for the double or triple amount of iterations, without ever seeing a major change in the chains' behaviour. Of course, I cannot exclude the possibility that after many more iterations (tens of thousands) the Markov chains would have finally reached convergence, or in case of non-mixing parallel chains would have switched their behaviour. However, the model fits, especially for the combination model, were both time-consuming (MCMC runs with two parallel chains could take 1-10 days, depending on model) and memory-intense (using approximately 1-5 GB RAM, depending on model). Considering that processing the three presented data sets required in total 240 model fits, I tried to reduce MCMC iterations to a

reasonable amount, which in most cases led to convergent and well-behaved Markov chains.

To understand convergence problems, I calculated data cloning estimability diagnostics and likelihood slices for selected trajectories. To obtain estimability diagnostics for a trajectory, I had to run the data cloning algorithm several times for increasing number of clones. This was even more computationally demanding than running data cloning for a single fixed number of clones. Therefore, I did not calculate estimability diagnostics for all model fits.

I performed estimability analysis on selected trajectories across all three data sets to understand a variety of phenomena. In data set 2, large $\hat{R}$ values occurred in a matching model fit (fourth combination trajectory; see previous section). I inspected MCMC traces and posterior distributions. The two parallel chains in the MCMC did not mix, but each chain appeared to converge on its own (Figure B.5). This resulted in bimodal posterior distributions of the parameters. I calculated estimates from each of the chains separately and calculated their likelihood values. The estimates from one chain (red chain in Figure B.5), say $\hat{\boldsymbol{\theta}}_1$, were close to the true underlying values of the trajectory with $\log L(\hat{\boldsymbol{\theta}}_1) = -6783.94$. In comparison, true parameter values had slightly lower log-likelihood $\log L(\boldsymbol{\theta}_{\text{true}}) = -6788.582$. The estimates from the other chain (black chain in Figure B.5), say $\hat{\boldsymbol{\theta}}_2$, had a lower log-likelihood value of $\log L(\hat{\boldsymbol{\theta}}_2) = -7121.017$. It appears that the likelihood function has a local maximum at $\hat{\boldsymbol{\theta}}_2$. While the first chain found the higher peak, the second chain found the second, lower, peak and failed to move away from it. Because $L(\hat{\boldsymbol{\theta}}_2)$ was distinctly lower than $L(\hat{\boldsymbol{\theta}}_1)$, it did not appear that the chains' behaviour was due to an estimability problem. To confirm this, I calculated estimability diagnostics. This required a new model fit with varying number of clones. In this fit, all Markov chains converged and mixed well. All posterior variances decreased with increasing number of clones and `lambda.max` converged to zero with rate $\frac{1}{K}$, where $K$ is the number of clones (Figure B.6).

I looked into estimability for three more model fits that did not converge during the first run. I analyzed a non-convergent model fit in data set 1, in which the combination model was fitted to a memory trajectory. When I calculated estimability diagnostics, all Markov chains converged and `lambda.max` behaved well and did not indicate any estimability problems. I further analyzed estimability for a non-convergent fit in data set 1, where a memory model was fitted to a resource trajectory. Here, variances of parameter estimates decreased properly for the kernel parameters, however not for selection parameters, $\alpha_{\text{mem}}$ and $\beta_{\text{mem}}$, indicating estimability issues (Figure B.7). This means that the selection parameters of the memory model with respect to time since

last visit could not be determined for the resource trajectory, which indeed did not truly contain an effect of this dynamic variable.

Because most convergence problems occurred when a more complex model was fitted to a null trajectory, I also examined a non-convergent fit of the combination model to a null trajectory in data set 2. Inspection of the non-convergent Markov chains showed that for most parameters, parallel chains did not mix but sampled different regions of the parameter space, resulting in bimodal posterior distributions. I separately calculated estimates and their likelihood values for the two chains, and the likelihood difference was smaller than one. Posterior variances indicated estimability problems for all selection parameters, i.e. those parameters that were not relevant to the null trajectory (Figure B.8).

### B.2.2 Estimability in cases of large confidence intervals

In my analysis of parameter estimates and their confidence intervals in matching model fits, I found that even though MCMC runs converged, parameters of the weighting function occasionally had very large confidence intervals. One example is the matching fit of the fourth resource trajectory in data set 2, in which the estimate of $\alpha_{\text{res}}$ had an unusually high value, together with a large confidence interval. Estimability diagnostics showed decreasing posterior variances for all parameters except $\alpha_{\text{res}}$ (Figure B.9). I suspected that these results were caused by a ridge in the likelihood along $\alpha_{\text{res}}$. I therefore calculated two-dimensional likelihood slices to confirm this. Figure B.10 shows two example slices, in which $\alpha_{\text{res}}$ and $\beta_2$ vary, whereas all remaining parameters are fixed. First, I fixed remaining parameters at their true values. The resulting surface over $\alpha_{\text{res}}$ and $\beta_2$ has a local maximum with log-likelihood value -7099.697 (Figure B.10 (a-i) and (a-ii)). However, when I fixed the other parameters at their MLE values obtained from the model fit, the surface shows a ridge (Figure B.10 (b-i) and (b-ii)). This ridge has a log-likelihood value of -7098.876, which is slightly higher than the local maximum of the other slice. The MCMC explores this area and moves along the ridge. It appears that the ridge has a very subtle maximum between -10 and -9, but it is so subtle that the MCMC extensively moves along the entire ridge.

In my model selection analysis, there was one matching model fit that converged but did not result in lowest BIC. This was a resource trajectory, for which the null model had lower BIC. In the matching fit with the resource model, estimates of selection parameters had high absolute values and large confidence intervals. When I calculated estimability diagnostics for a series of clones $K = 1, 5, 10, 15$, `lambda.max` showed

signs of estimability problems (Figure B.11a). Estimates were very similar to the first MCMC run ($\hat{\alpha}_{\text{res}} = 10.4$, $\hat{\beta}_1 = -24$, $\beta_2 = -3.6$). However, because `lambda.max` generally decreased, I considered the possibility that I had not used enough clones. I therefore calculated estimability diagnostics for $K = 1, 8, 15, 22, 30$. In this run, $\alpha_{\text{res}}$ and $\beta_1$ showed good behaviour, however $\beta_2$ had high value and large confidence interval ($\hat{\alpha}_{\text{res}} = 1.3$, $\hat{\beta}_1 = -1.3$, $\beta_2 = 8.9$). Estimability diagnostics showed potential issues with $\beta_2$ (Figure B.11b,d). To understand this further, I also considered the likelihood function. I calculated log-likelihood values for the true underlying parameter values ($L_{\text{true}} = -6920$), the estimates from the first estimability run for $K = 15$ ($L_{1,k15} = -6915$), and the estimates from the second estimability run for $K = 30$ ($L_{2,k30} = -6918$). I compared these with the log-likelihood value for the model fit in which I fitted the null model to the trajectory ($L_{\text{null}} = -6921$). It appears that estimates from the first run approximate the MLE, whereas estimates from the last run with 30 clones arise from a local maximum with only slightly lower likelihood. When I plotted a likelihood slice for this model fit (fixing all parameters but $\beta_2$ at their estimates), I found a potential ridge in the likelihood for positive large values of $\beta_2$ (Figure B.11c). On the other hand, in the region of the estimates from the first run, I did not see any signs of ridges, however log-likelihood values did not vary much. From these tests, I concluded that the likelihood surface for this resource trajectory has a difficult structure for optimization, but a maximum exists in the region of selection parameters $\hat{\alpha}_{\text{res}} = 10.4$, $\hat{\beta}_1 = -24$, $\beta_2 = -3.6$. These parameter values result in a weighting function that is almost constant (Figure B.12b) and therefore model selection via BIC prefers the more parsimonious null model (Figure B.12d). Estimates from the second estimability run for $K = 30$ are closest to the true values (Figure B.12a,c), however, the likelihood is lower for these values.

## B.3 Conclusions about model fitting

In general, the results for the supplemental data are similar to the results for the data set presented in the main text of the paper. Although I designed the supplemental data set to include potentially more difficult estimation scenarios, my framework was able to detect effects of both resources and the dynamic variable time since last visit.

About 20% of data cloning MCMC runs did not converge the first time, and I did not achieve improvement by increasing the number of MCMC iterations. However, when I re-started certain model fits for estimability analysis, occasionally MCMC runs converged in this second run. From this experience, I recommend to rather re-start

MCMC sampling completely instead of running more iterations, especially when traces show that parallel chains sample distinct regions of the parameter space, leading to bimodal posteriors. I recommend to additionally calculate estimability diagnostics for these model fits. If these indicate estimability problems for certain parameters, this may be an indicator that a model contains covariates that in fact did not influence the movement process. In this case, I recommend fitting alternative models or sub-models and comparing them via model selection.

Convergence problems were related to two different phenomena of the likelihood function. First, the likelihood function had local maxima or ridges but still a unique global maximum. In such cases, single chains could occasionally fail to find the global maximum. This could be a potential difficulty with data cloning. In data cloning, every peak in the likelihood function is enhanced, including local maxima. If a chain by chance, e.g. via a ridge-like structure, reaches a local maximum, it may have difficulty moving away from it. This may also depend on the MCMC algorithm used. Other methods such as standard maximization of the likelihood function are not safe from this problem of local maxima either. It is thus for any method important to use multiple starting points or parallel chains. As second reason for non-convergent chains I found likelihood functions that had ridges or distinct multiple maxima, i.e. global maxima with almost the same likelihood value. These were clear cases of estimability problems, and were detected by data cloning estimability diagnostics. Any other method will fail in these cases too, either through non-convergence or results that indicate multiple possibilities for estimates (multiple maxima are found, bimodality of posteriors). If no problems are detected in these cases, this is even worse, because wrong conclusions are made.

Most convergence problems occurred when a more complex model was fitted to a null trajectory. I suspect that this may be partly due to the form of the weighting function. In the logistic function, $\big(1 + \exp(-\alpha - \beta x)\big)^{-1}$, large values of the intercept $\alpha$ can cause the exponential function to almost vanish, leading to a nearly constant logistic function. Large selection parameters $\beta$ can also have this effect. Therefore, if we fit a model that includes any kind of selection to a null trajectory, we can expect the likelihood function to have multiple maxima, ridges or plateaus, especially in those regions of the parameter space where parameters of the weighting function are large. Therefore, I believe there is no need to be alarmed that many of these model fits did not converge in the analysis. Via estimability diagnostics, we have the ability to detect such situations. As soon as trajectories contained at least one effect (either resources or time since last visit), convergence problems occurred less frequently. However, this

phenomenon of the logistic function may give reason to also consider alternative forms of the weighting function that do not experience this problem.

In matching model fits, I occasionally observed unusually large deviations of estimates from true parameter values or large confidence intervals. Based on my investigations, I suspect that this is mainly due to estimability issues (e.g. ridges in the likelihood). These could occur due to stochasticity in the data simulation. Each trajectory is a realization of a stochastic process. In most cases, we expect trajectories to realize a behaviour according to the parameter values used for the simulations. Still, we must expect to see also cases that are less well behaved.

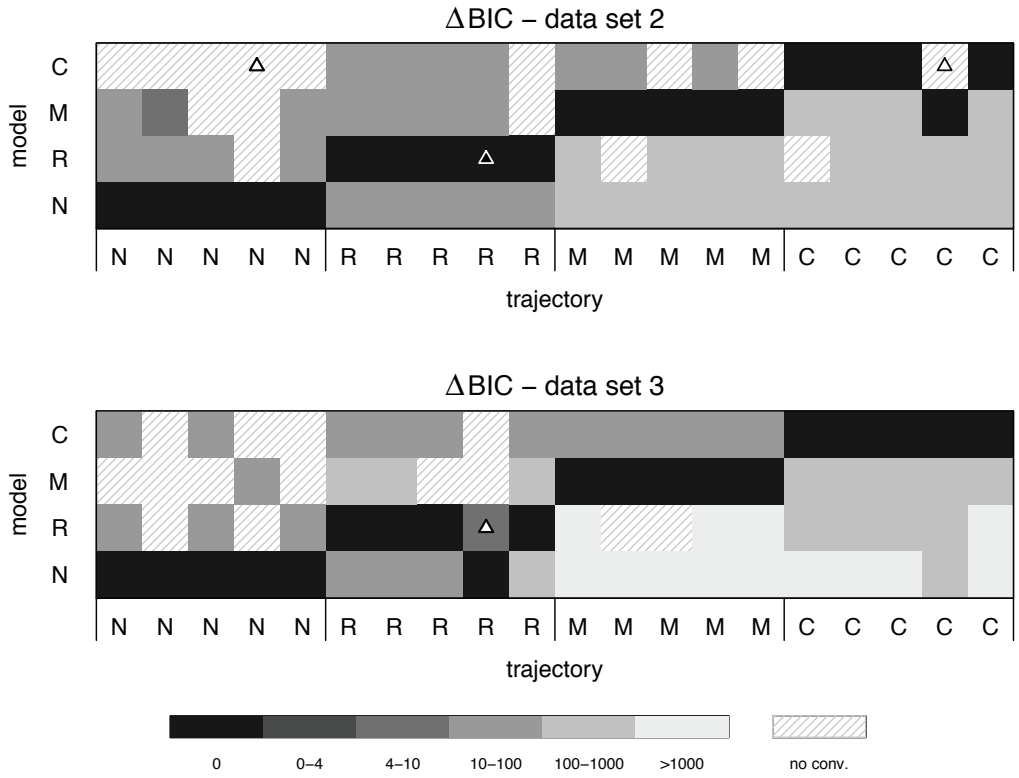



**Figure B.1.** Each column in the two subfigures shows model selection results for one simulated trajectory when it was fitted with the four candidate models (null, resource, memory, combination). For each trajectory, I calculated BIC values for the four fitted models, and the figure shows differences in BIC with respect to the minimal BIC value, that is the model with minimal BIC has $\Delta\text{BIC} = 0$. I excluded model fits with non-convergent MCMC. For coherence, the figures depict the results from the first MCMC run for each trajectory. Triangle indicate those trajectories for which I calculated estimability diagnostics.

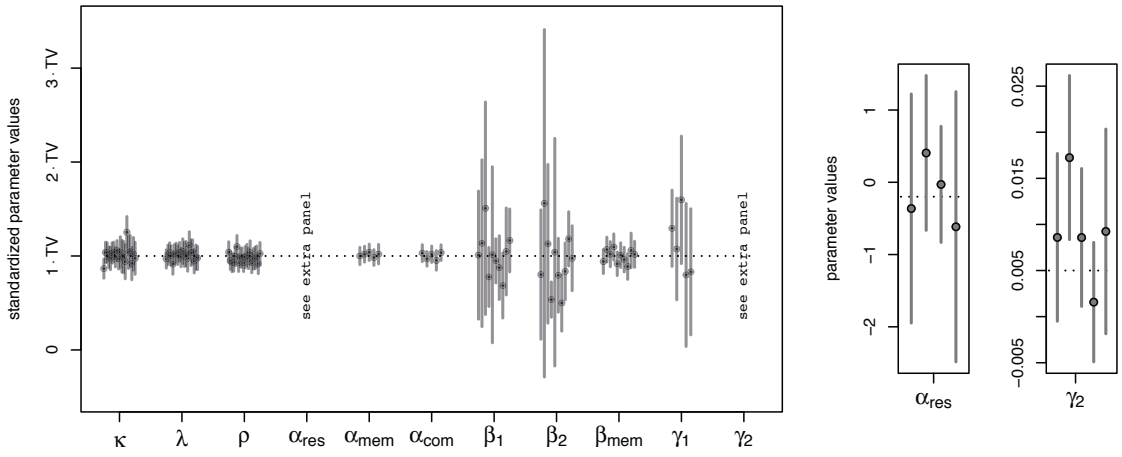



**Figure B.2.** Parameter estimates and their 95% confidence intervals for matching model fits for data set 2. In the large panel, both parameter estimates and Wald-type confidence intervals are scaled by the true parameter values (TV): $\kappa = 4$, $\lambda = 0.9$, $\rho = 1.2$, $\alpha_{\text{res}} = -0.2$, $\alpha_{\text{mem}} = -3$, $\alpha_{\text{com}} = -3.2$, $\beta_1 = 0.5$, $\beta_2 = 0.8$, $\beta_{\text{mem}} = 0.04$, $\gamma_1 = 0.008$, $\gamma_2$=0.005. Smaller panels have unscaled values. In all plots, dotted lines mark true values.

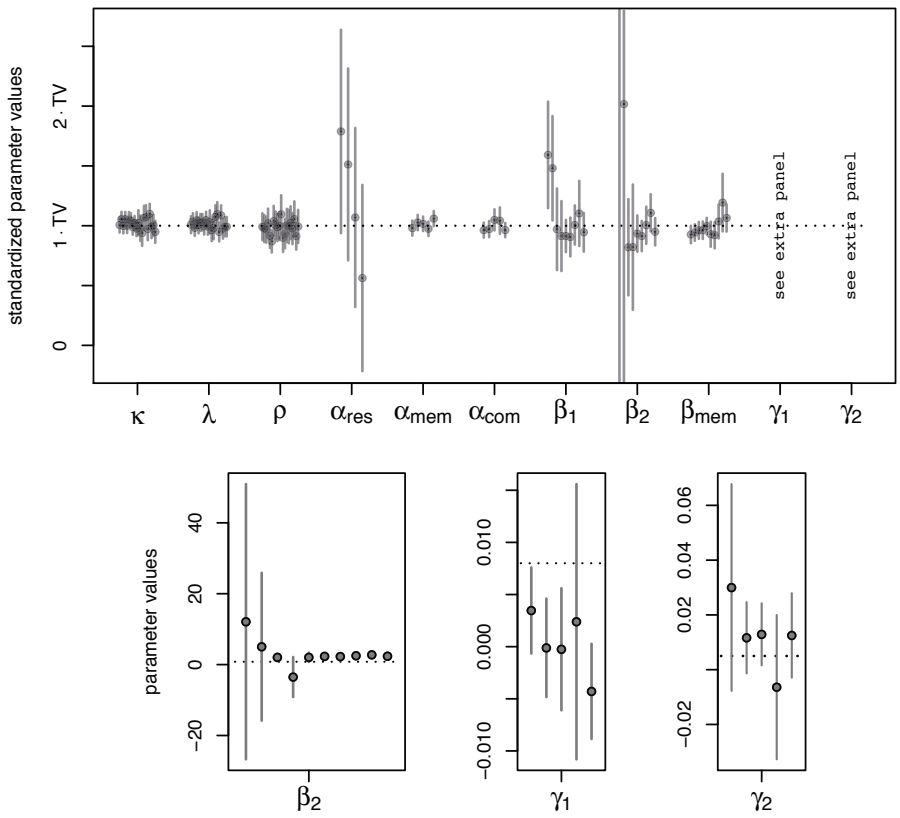



**Figure B.3.** Parameter estimates and their 95% confidence intervals for matching model fits for data set 3. In large panels, both parameter estimates and Wald-type confidence intervals are scaled by the true parameter values (TV): $\kappa = 4.5$, $\lambda = 1.3$, $\rho = 1.5$, $\alpha_{\text{res}} = 0.8$, $\alpha_{\text{mem}} = -5$, $\alpha_{\text{com}} = -5.8$, $\beta_1 = -1.5$, $\beta_2 = 2.5$, $\beta_{\text{mem}} = 0.05$, $\gamma_1 = 0$, $\gamma_2$=0.01. Smaller panels have unscaled values. In all plots, dotted lines mark true values. Estimates from the matching resource model fit in data set 3 that was not selected as best model are excluded.

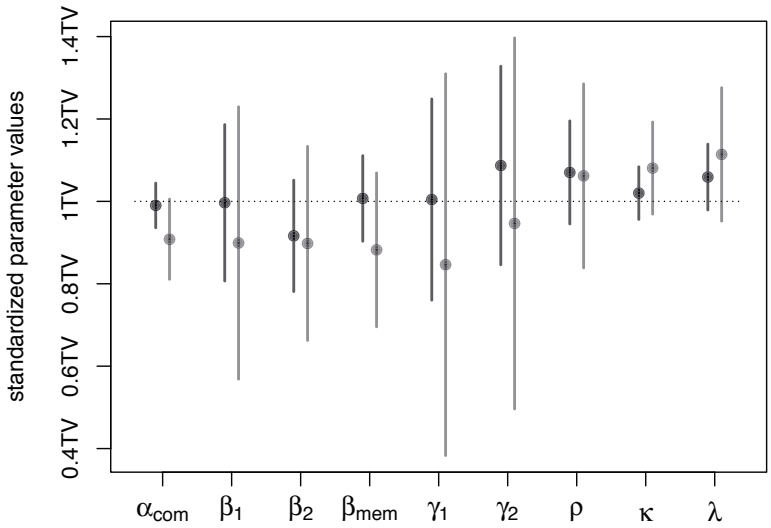



**Figure B.4.** Parameter estimates and their 95% confidence intervals for the combination trajectory on landscape 2, fitted with the combination model. Results for the complete trajectory (dark grey) are compared to results for the same trajectory with 10% missing locations (light grey). Parameter estimates and confidence intervals are scaled by the true parameter values (TV): $\kappa = 5.5$, $\lambda = 1.6$, $\rho = 1$, $\alpha_{\text{res}} = -1$, $\alpha_{\text{mem}} = -4$, $\alpha_{\text{com}} = -5$, $\beta_1 = 1$, $\beta_2 = 2$, $\beta_{\text{mem}} = 0.03$, $\gamma_1 = 0.01$, $\gamma_2$=0.05. Dotted lines mark the true values.

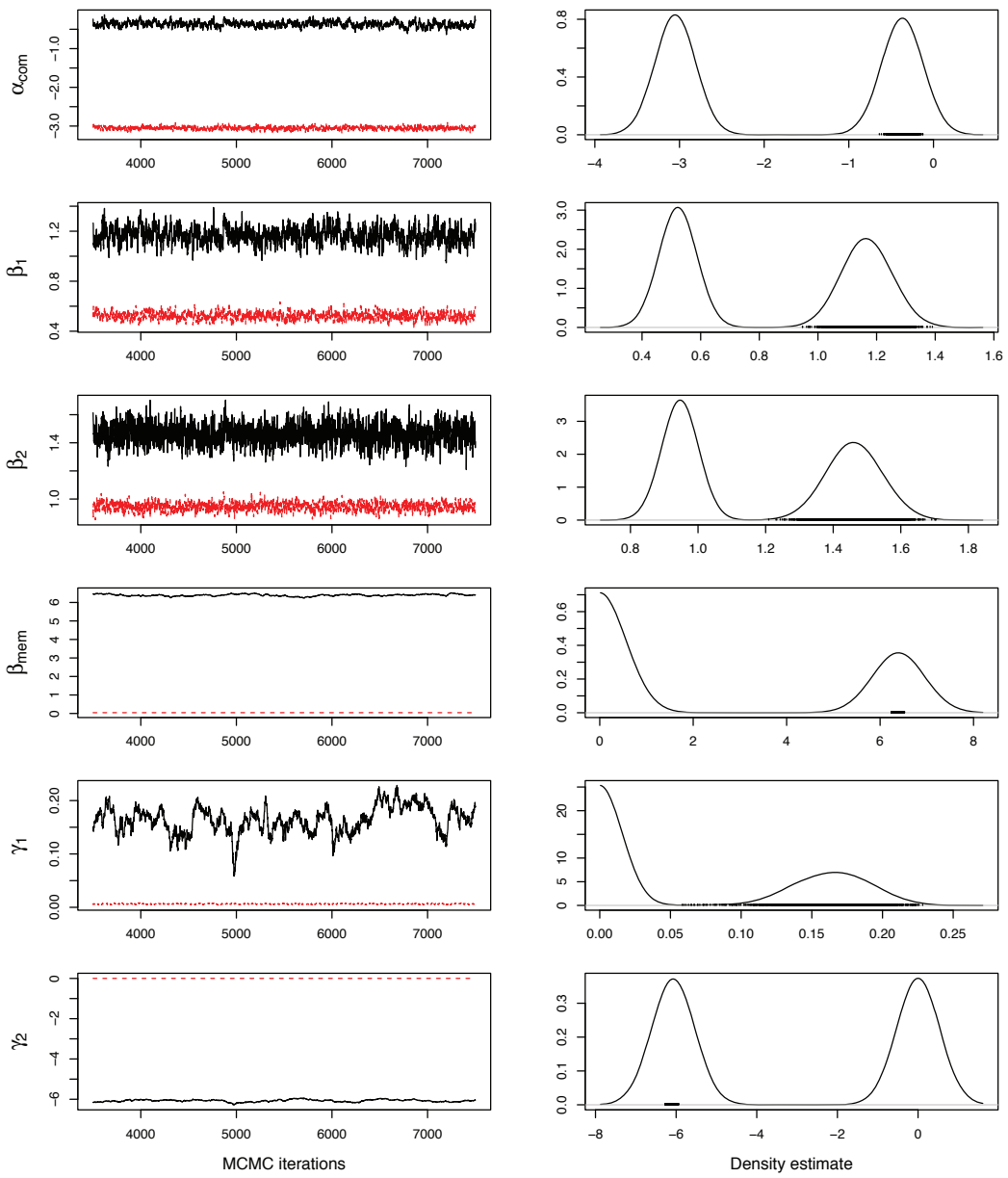



**Figure B.5.** MCMC results for the misidentified combination trajectory in data set 2; compare Figure B.1. The MCMC had mixing problems when the trajectory was fitted with the matching combination model. Trace plots of MCMC iterations and density plots are shown for all parameters of the weighting function. The two parallel chains do not mix, but each appear to converge on their own. Estimates derived only from the red chain have higher likelihood value than estimates derived from the black chain. Estimates from the red chain are close to the true underlying values of the trajectory.

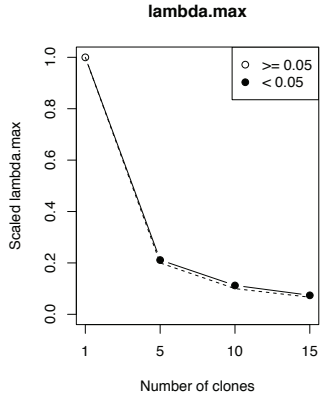



**Figure B.6.** Estimability diagnostics for a second run of the model fit depicted in Figure B.5. In the first MCMC run, mixing problems occurred. However, in a second run to obtain estimability diagnostics, problems did not repeat. The plot shows the comparison of `lambda.max` (points and solid line) and the line $\frac{1}{k}$ (dotted) for the second run without problems.

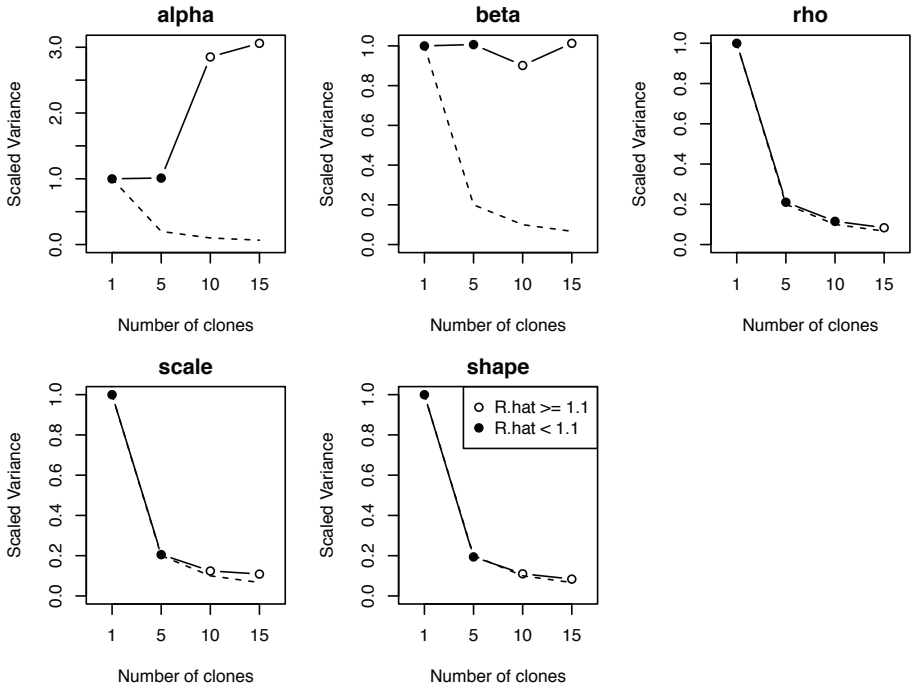



**Figure B.7.** Estimability diagnostics for a memory model fitted to a resource trajectory that did not converge in data set 1; compare Figure 2.4. The plots show variances of the posterior distributions for increasing number of clones. Non-decreasing variances of $\alpha_{\text{res}}$ and $\beta_{\text{mem}}$ indicate estimability issues.

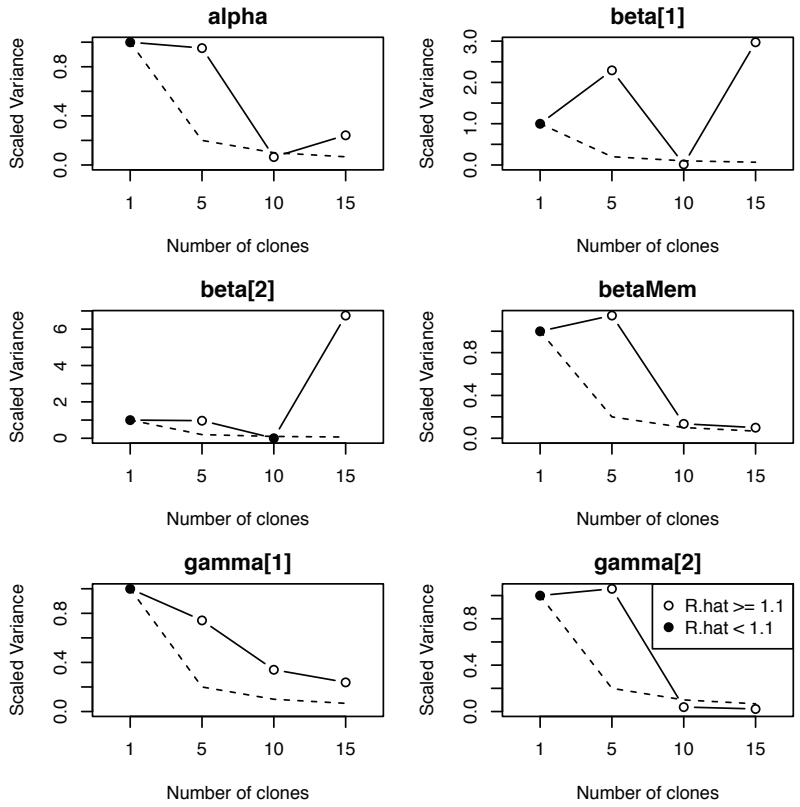



**Figure B.8.** Estimability diagnostics for a combination model fitted to a null trajectory that did not converge in data set 2; compare Figure B.1. The plots show variances of the posterior distributions of selection parameters for increasing number of clones, which all indicate estimability problems.

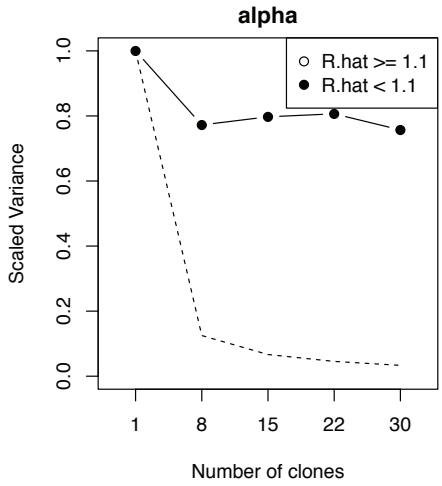



**Figure B.9.** Posterior variance of $\alpha_{\text{res}}$ for a matching resource model fit in data set 2, which showed estimability problems; compare Figure B.1. The dotted line indicates the ideal line $\frac{1}{k}$.

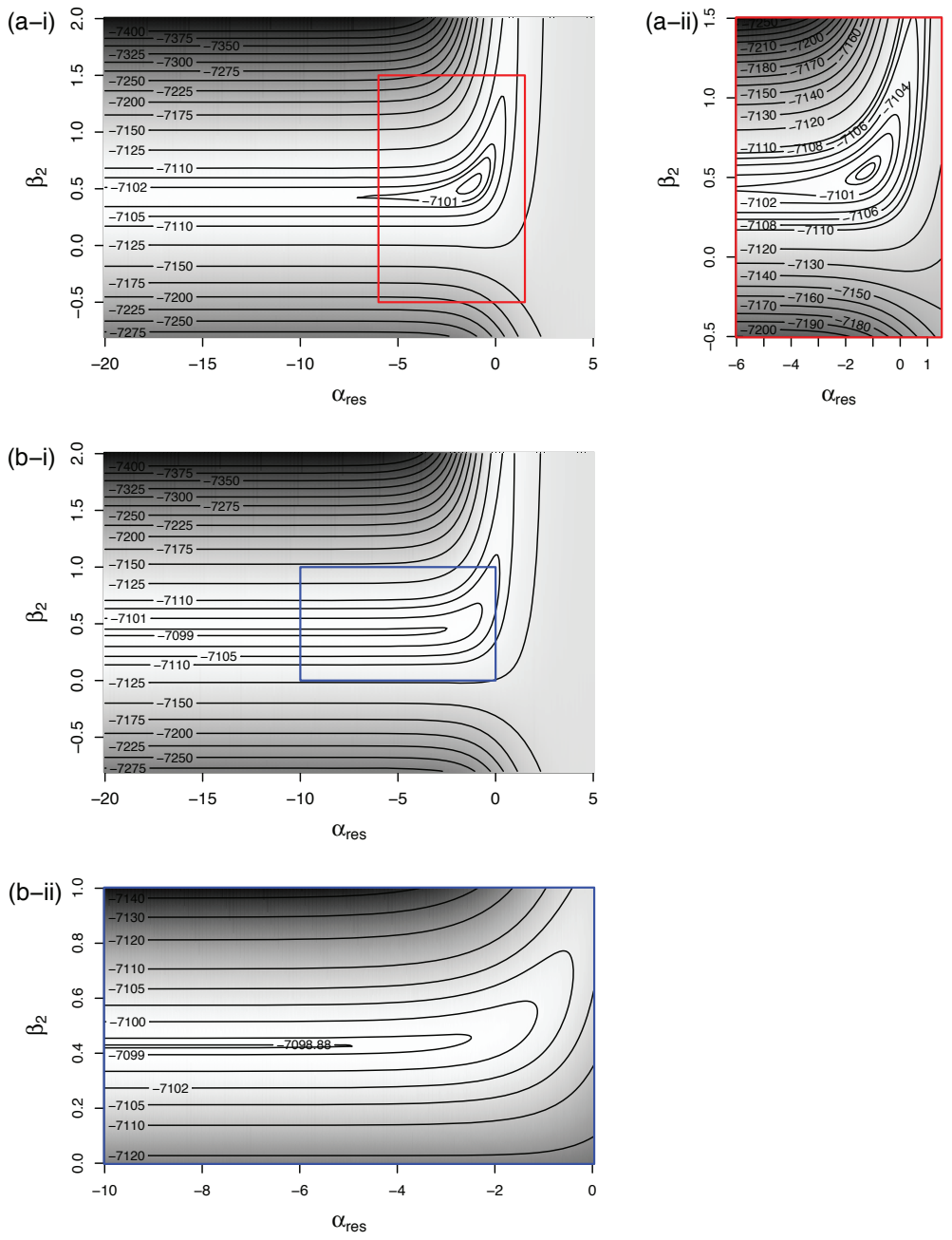
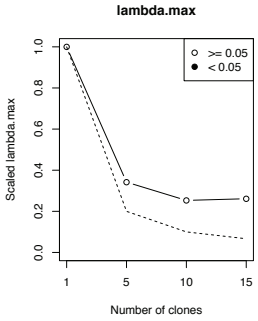
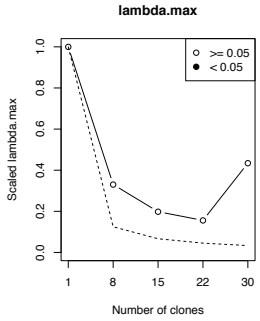


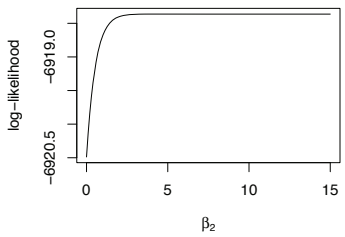

**Figure B.10.** Slices of the log-likelihood function for a resource trajectory with estimability issues in data set 2; compare Figure B.1. **Panels (a-i)** and **(a-ii)** show the log-likelihood surface when all parameters are fixed at their true values except $\alpha_{\text{res}}$ and $\beta_2$. The surface shows a local peak. **Panels (b-i)** and **(b-ii)** show the log-likelihood surface when parameters instead are fixed at their MLE values. This surface shows a ridge and the log-likelihood value of this ridge is slightly higher than the local peak in panels (a-i) and (a-ii). Therefore, $\alpha_{\text{res}}$ cannot be estimated uniquely.



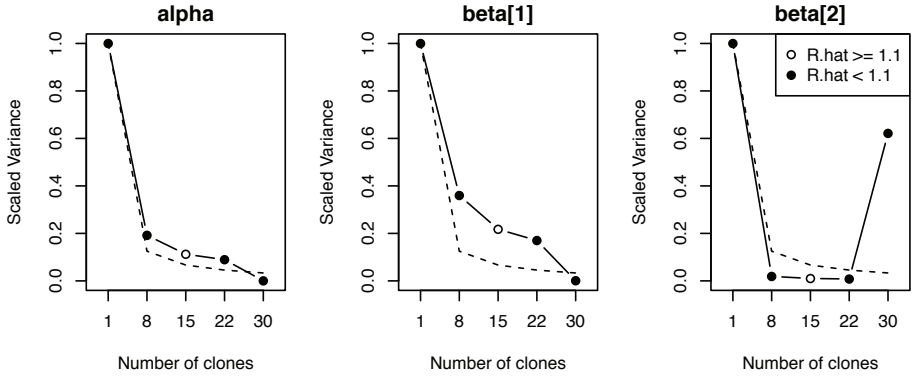

**(a)** Largest posterior variance for the first estimability test.




**(b)** Largest posterior variance for the second estimability test with higher number of clones.




**(c)** Likelihood slice, all other parameters fixed at estimates obtained from second estimability test for $k = 30$.




**(d)** Separate posterior variances for selection parameters for the second estimability test with higher number of clones.

**Figure B.11.** Estimability diagnostics for the fourth matching resource model fit in data set 2, which was misidentified during model selection; compare Figure B.1. Two different series of MCMC runs with varying number of clones suggested difficulties with estimability. Region-wise calculation of the likelihood function confirmed a complex likelihood surface with a potential ridge (subfigure (c)) but nonetheless a slightly higher maximum far away from the true parameter values.

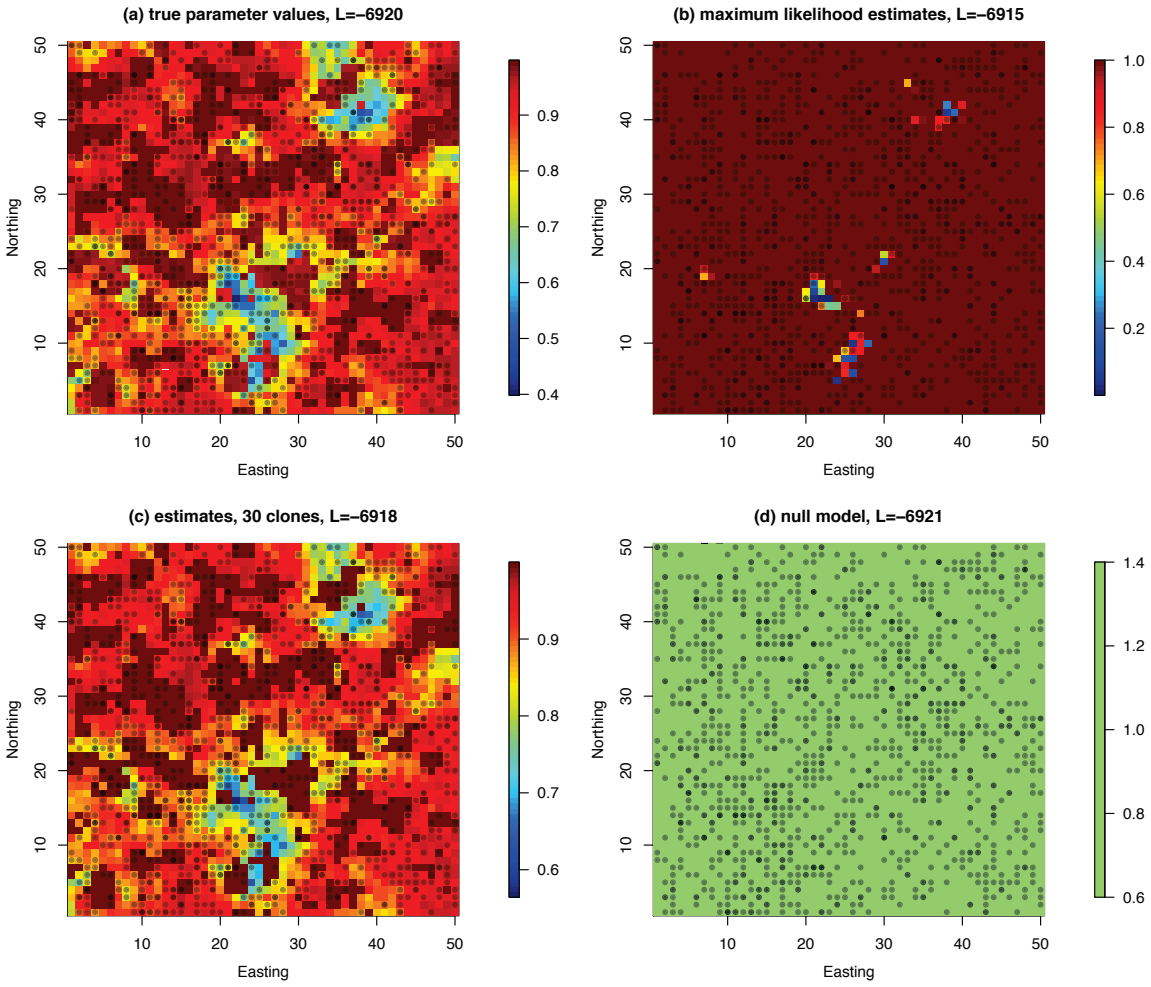



**Figure B.12.** Weighting function for different parameter estimates obtained for the resource trajectory in data set 3 on landscape 4. For the matching fit with the resource model, I obtained different estimates from different MCMC runs due to a complex likelihood surface with local maxima. **Panel (a)**: "True" parameter values used in trajectory simulation. **Panel (b)**: Estimates from a model fit with 15 clones. **Panel (c)**: Estimates from an alternative model fit with 30 clones. **Panel (d)**: Constant weighting function (null model). Gray dots are the locations of the trajectory. Darker dots correspond to multiple visits to a location. Their distribution across the entire home range indicate a rather uniform use of space in accordance with weighting functions (b) and (d).

# Appendix C

# Characteristic function of a radially symmetric random vector

Here, I provide details about the link between the characteristic function of a radially symmetric random vector and the Hankel transform as stated in equation (4.31). The ch.f. of the two-dimensional random vector $\boldsymbol{S}$ with density (4.28) is given by

$$\phi(\boldsymbol{u}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{i\boldsymbol{u}\cdot\boldsymbol{s}}\, p_{S_1,S_2}(s_1, s_2)\,\mathrm{d}s_1\mathrm{d}s_2. \tag{C.1}$$

Because the density is radially symmetric, I switch to polar coordinates via $s_1 = r\cos\beta$ and $s_2 = r\sin\beta$, where the angle $\beta$ is chosen such that the vector $\boldsymbol{u}$ has angle zero. The determinant of the Jacobian for this transformation is $|J| = r$. The dot product of the vectors $\boldsymbol{u}$ and $\boldsymbol{s}$ can be written as $\boldsymbol{u}\cdot\boldsymbol{s} = \|\boldsymbol{u}\|\, r\cos\beta$. With this, we obtain

$$\phi(\boldsymbol{u}) = \int_0^{\infty}\left(\int_0^{2\pi} \mathrm{e}^{i\|\boldsymbol{u}\| r\cos\beta}\,\mathrm{d}\beta\right) p_{S_1,S_2}(r)\, r\,\mathrm{d}r. \tag{C.2}$$

The symmetry of the cosine allows to simplify the inner integral as follows,

$$\int_0^{2\pi} \mathrm{e}^{i\|\boldsymbol{u}\| r\cos\beta}\,\mathrm{d}\beta = 2\int_0^{\pi} \mathrm{e}^{i\|\boldsymbol{u}\| r\cos\beta}\,\mathrm{d}\beta = 2\pi J_0(\|\boldsymbol{u}\| r), \tag{C.3}$$

where $J_0$ denotes the Bessel function of the first kind. The last equation follows from an integral representation of the Bessel function (Abramowitz & Stegun, 1964, 9.1.21). With this, the characteristic function becomes

$$\phi(\boldsymbol{u}) = 2\pi\int_0^{\infty} p_{S_1,S_2}(r)\, r\, J_0(\|\boldsymbol{u}\| r)\,\mathrm{d}r. \tag{C.4}$$

The integral is the Hankel transform of order zero of the density $p_{S_1,S_2}(r)$ evaluated at $\|\boldsymbol{u}\|$.

# Appendix D

# Examples of simulated resource landscapes and trajectories for robustness study

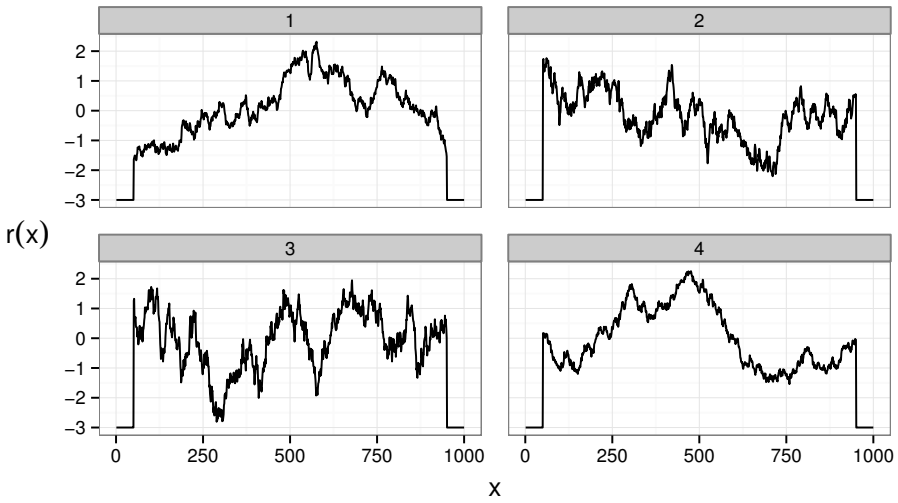



**Figure D.1.** Four of the simulated resource landscapes used for sampling movement trajectories. The depicted landscapes have been generated with spatial autocorrelation $\mathrm{Cov}(r(x), r(y)) = \exp\left(\frac{|x-y|}{s}\right)$ for $s = 200, 300, 400, 500$. I standardized landscapes to range within the interval $(-3, 3)$. At the boundaries, I set values to -3 to avoid movement close to the boundary and resulting boundary effects in the transition densities due to the normalization constant.

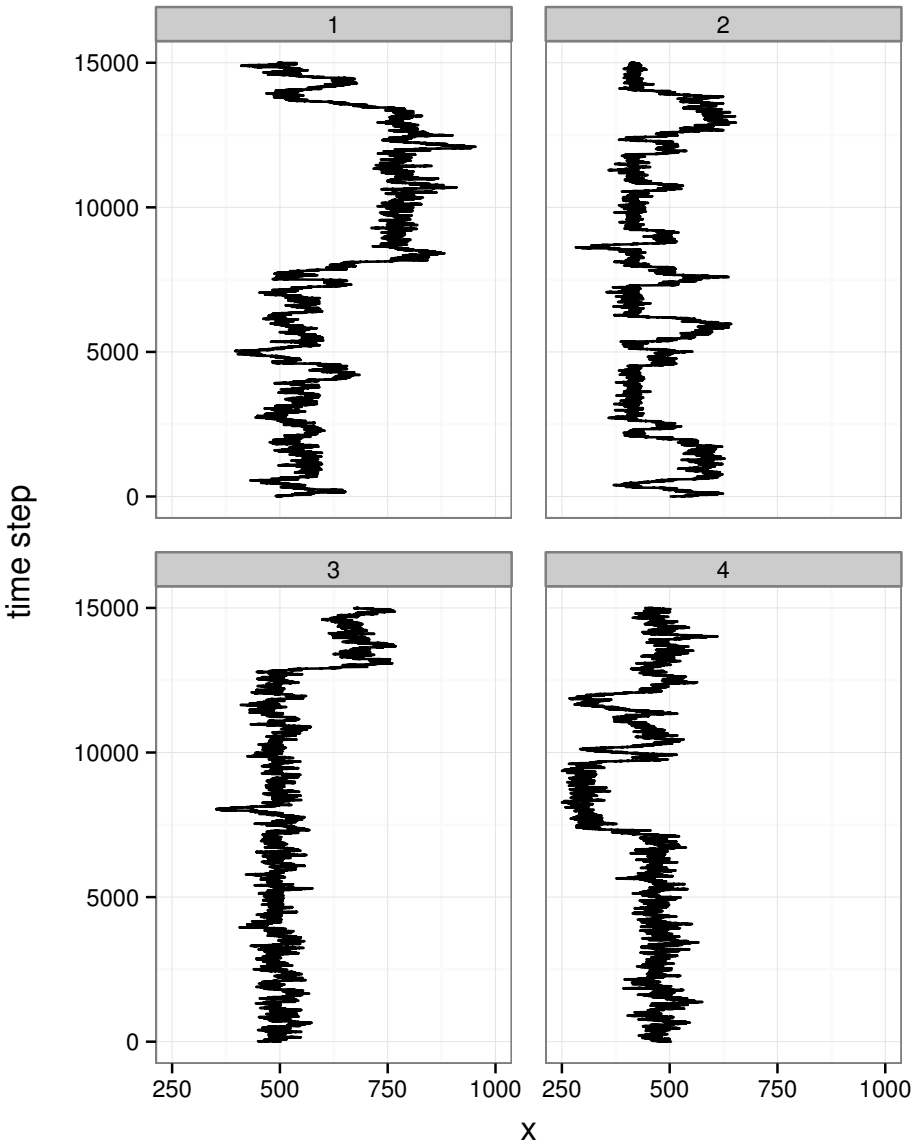



**Figure D.2.** Four of the simulated trajectories from the model with exponential weighting function. The trajectories were generated using the parameter values $\sigma = 6$ and $\beta = 1$. The underlying resource landscapes are the landscapes depicted in Fig. D.1, in the same order.

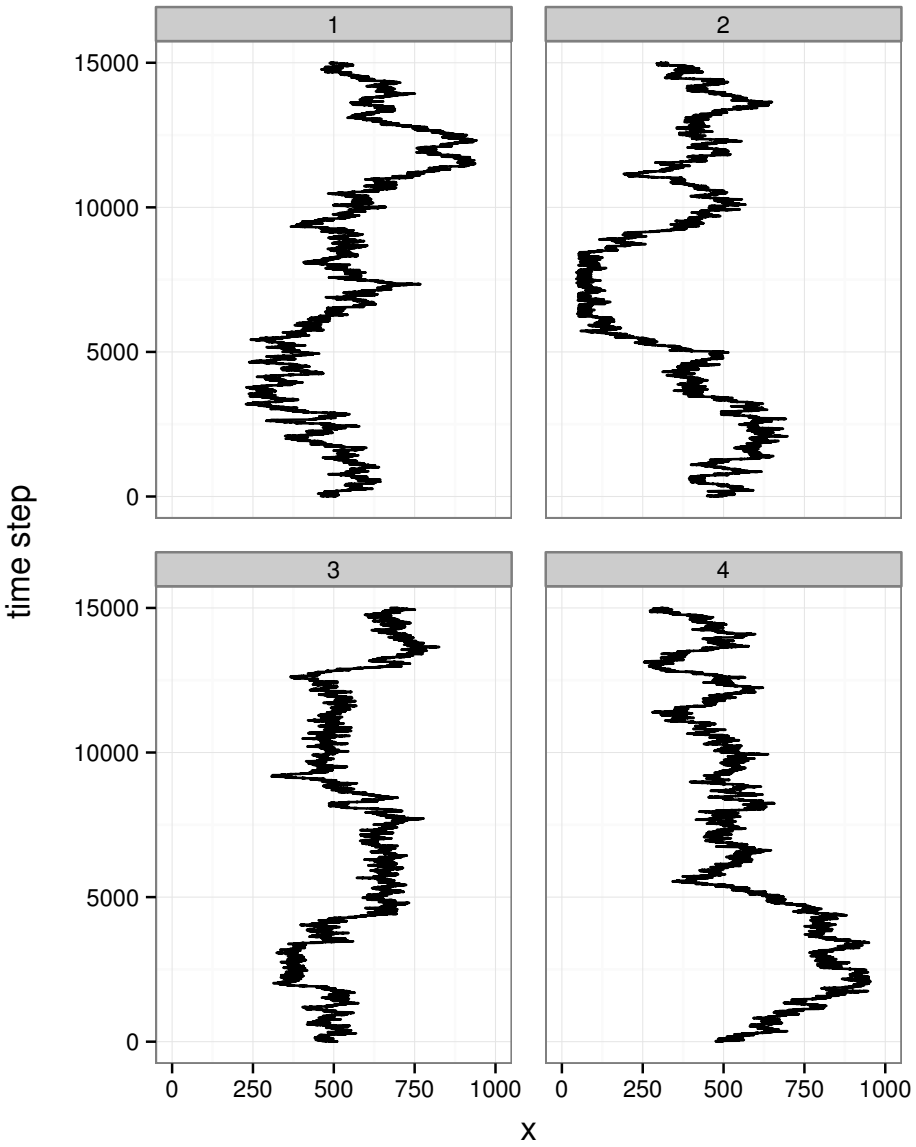



**Figure D.3.** Four of the simulated trajectories from the model with logistic weighting function. The trajectories were generated using the parameter values $\sigma = 6$ and $\beta = 1$ (same as in Fig. D.2) and $\alpha = 0$. The underlying resource landscapes are again the landscapes depicted in Fig. D.1, in the same order.

# Appendix E

# Proofs of robustness results in Chapter 5

## E.1 Proofs of results about exact robustness

*Proof of Theorem 5.1.* First, note that for any standard deviation of the kernel, $\sigma$, the integral $\int_{\mathbb{R}} k_\sigma(y;x)w(y)\,\mathrm{d}y$ reduces to the weighting function evaluated at the kernel's mean,

$$\begin{aligned}\int_{\mathbb{R}} k_\sigma(y;x)w(y)\,\mathrm{d}y &= \int_{\mathbb{R}} k_\sigma(y;x)(ay+b)\,\mathrm{d}y = \int_{\mathbb{R}} k_\sigma(y;x)(a(y-x+x)+b)\,\mathrm{d}y \\ &= (ax+b)\int_{\mathbb{R}} k_\sigma(y;x)\,\mathrm{d}y + a\int_{\mathbb{R}} k_\sigma(y;x)(y-x)\,\mathrm{d}y = ax+b = w(x), \quad \text{(E.1)}\end{aligned}$$

because $k_\sigma(\cdot|y)$ is a Gaussian density integrating to one and with vanishing first central moment. If we consider $w$ as a linear transformation of a Normally distributed random variable with mean $x$, then equation (E.1) reflects a special case of Jensen's inequality, in which equality holds.

I now show robustness of degree $n$ with parameter transformation $g_n(\sigma,a,b) = (\sqrt{n}\sigma,a,b)$ by induction. For $n=1$, we have the trivial transformation $g_1(\sigma,a,b) = (\sigma,a,b)$, and there is nothing to show for robustness of degree 1.

Assume that robustness or degree $n$ holds, that is we have the relationship

$$p_n(x_n|x_0,\sigma,a,b) = p_1(x_n|x_0,\sqrt{n}\sigma,a,b). \quad \text{(E.2)}$$

for all $x_n, x_0 \in \mathbb{R}$. For $n+1$, we use the Chapman-Kolmogorov equation and Markov

property and obtain

$$\begin{aligned} p_{n+1}(x_{n+1}|x_0,\sigma,a,b) &= \int_{\mathbb{R}^n} \prod_{k=1}^{n+1} p_1(x_k|x_{k-1},\sigma,a,b)\,\mathrm{d}x_1 \ldots \mathrm{d}x_n \\ &= \int_{\mathbb{R}} p_1(x_{n+1}|x_n,\sigma,a,b) \left( \int_{\mathbb{R}^{n-1}} \prod_{k=1}^{n} p_1(x_k|x_{k-1},\sigma,a,b)\,\mathrm{d}x_1 \ldots \mathrm{d}x_{n-1} \right) \mathrm{d}x_n \\ &= \int_{\mathbb{R}} p_1(x_{n+1}|x_n,\sigma,a,b)\, p_n(x_n|x_0,\sigma,a,b)\,\mathrm{d}x_n \\ &= \int_{\mathbb{R}} p_1(x_{n+1}|x_n,\sigma,a,b)\, p_1(x_n|x_0,\sqrt{n}\sigma,a,b)\,\mathrm{d}x_n, \end{aligned} \tag{E.3}$$

where the last step follows by induction. We can now insert the model's step probabilities and use equation (E.1) to further calculate,

$$\begin{aligned} p_{n+1}(x_{n+1}|x_0,\sigma,a,b) &= \int_{\mathbb{R}} \frac{k_\sigma(x_{n+1};x_n)\,w(x_{n+1})}{\int_{\mathbb{R}} k_\sigma(y;x_n)w(y)\,dy} \frac{k_{\sqrt{n}\,\sigma}(x_n;x_0)\,w(x_n)}{\int_{\mathbb{R}} k_{\sqrt{n}\,\sigma}(y;x_0)w(y)\,dy}\,\mathrm{d}x_n \\ &= \int_{\mathbb{R}} \frac{k_\sigma(x_{n+1};x_n)\,w(x_{n+1})}{w(x_n)} \frac{k_{\sqrt{n}\,\sigma}(x_n;x_0)\,w(x_n)}{w(x_0)}\,\mathrm{d}x_n \\ &= \frac{w(x_{n+1})}{w(x_0)} \int_{\mathbb{R}} k_\sigma(x_{n+1};x_n)\,k_{\sqrt{n}\sigma}(x_n;x_0)\,\mathrm{d}z. \end{aligned} \tag{E.4}$$

Note that we have assumed that all movement steps are within the domain $\mathcal{I}$, where the weighting function is positive. Since $k_\sigma(x_{n+1};x_n) = k_\sigma(x_{n+1}-x_n;0)$, the integral in the last expression is the convolution of two Gaussian densities with variances $\sigma^2$ and $n\sigma^2$ and with means 0 and $x_0$, respectively. Because of the linearity of Gaussian random variables, this is again a Gaussian density with mean $x_0$ and variance $(n+1)\sigma^2$. Because equation (E.1) holds for the kernel with any standard deviation, we can rewrite the denominator as $w(x_0) = \int_{\mathbb{R}} k_{\sqrt{n+1}\,\sigma}(y;x_0)w(y)\,\mathrm{d}y$. Thus,

$$p_{n+1}(x_{n+1}|x_0,\sigma,a,b) = \frac{k_{\sqrt{n+1}\,\sigma}(x_{n+1};x_0)w(x_{n+1})}{\int_{\mathbb{R}} k_{\sqrt{n+1}\,\sigma}(y;x_0)w(y)\,dy} = p_1(x_{n+1}|x_0,\sqrt{n+1}\,\sigma,a,b). \tag{E.5}$$

□

*Proof of Theorem 5.2.* We proceed analogously to the previous proof. The integral of weighting function and kernel with arbitrary standard deviation $\sigma$ and mean $x$ is here

given by

$$\begin{aligned}\int_{\mathbb{R}} k_\sigma(y;x)\, w(y)\,\mathrm{d}y &= \int_{\mathbb{R}} k_\sigma(y;x)\, Ce^{ay+b}\,\mathrm{d}y \\ &= \frac{C}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} \exp\left(-\frac{(y-x)^2}{2\sigma^2} + ay + b\right)\mathrm{d}y.\end{aligned}$$

By completing the square and using substitution $u = \frac{1}{\sqrt{2}\sigma}(y - x - a\sigma^2)$ we obtain

$$\begin{aligned}\int_{\mathbb{R}} k_\sigma(y;x)\, w(y)\,\mathrm{d}y &= \frac{C}{\sqrt{2\pi}\sigma} e^{\frac{a^2\sigma^2}{2}+ax+b}\int_{\mathbb{R}} \exp\left(-\left(\frac{y-x-a\sigma^2}{\sqrt{2}\sigma}\right)^2\right)\mathrm{d}y \\ &= \frac{C}{\sqrt{2\pi}\sigma} e^{\frac{a^2\sigma^2}{2}+ax+b}\int_{\mathbb{R}} \exp\left(-u^2\right)\sqrt{2}\sigma\,\mathrm{d}u.\end{aligned}$$

The final integral reduces to $\sqrt{2\pi}\sigma$, and therefore,

$$\int_{\mathbb{R}} k_\sigma(y;x)\, w(y)\,\mathrm{d}y = C\, e^{\frac{a^2\sigma^2}{2}+ax+b}. \tag{E.6}$$

Again, I prove robustness of degree $n$ by induction, using parameter transformation $g_n(\sigma, C, a, b) = (\sqrt{n}\sigma, C, a, b)$. In the induction step, we obtain, with help of equation (E.6),

$$\begin{aligned}p_{n+1}(x_{n+1}|x_0,\sigma,a,b) &= \int_{\mathbb{R}} \frac{k_\sigma(x_{n+1};x_n)\, Ce^{ax_{n+1}+b}}{\int_{\mathbb{R}} k_\sigma(y;x_n) Ce^{ay+b}\,dy}\,\frac{k_{\sqrt{n}\,\sigma}(x_n;x_0)\, Ce^{ax_n+b}}{\int_{\mathbb{R}} k_{\sqrt{n}\,\sigma}(y;x_0) Ce^{ay+b}\,dy}\,\mathrm{d}x_n \\ &= \int_{\mathbb{R}} \frac{k_\sigma(x_{n+1};x_n)\, Ce^{ax_{n+1}+b}}{Ce^{\frac{a^2\sigma^2}{2}+ax_n+b}}\,\frac{k_{\sqrt{n}\,\sigma}(x_n;x_0)\, Ce^{ax_n+b}}{Ce^{\frac{na^2\sigma^2}{2}+ax_0+b}}\,\mathrm{d}x_n \\ &= \frac{e^{x_{n+1}}}{e^{\frac{(n+1)a^2\sigma^2}{2}+ax_0}}\int_{\mathbb{R}} k_\sigma(x_{n+1};x_n)\, k_{\sqrt{n}\,\sigma}(x_n;x_0)\,\mathrm{d}z \\ &= \frac{e^{x_{n+1}}}{e^{\frac{(n+1)a^2\sigma^2}{2}+ax_0}}\, k_{\sqrt{n+1}\,\sigma}(x_{n+1};x_0). \\ &= \frac{k_{\sqrt{n+1}\,\sigma}(x_{n+1};x_0)\, Ce^{ax_{n+1}+b}}{\int_{\mathbb{R}} k_{\sqrt{n+1}\,\sigma}(y;x_0)\, Ce^{ay+b}\,\mathrm{d}y} \\ &= p_1(x_{n+1}|x_0,\sqrt{n+1}\,\sigma,a,b)\end{aligned} \tag{E.7}$$

□

## E.2 Proof of result about asymptotic robustness

To highlight the main steps necessary to prove Theorem 5.3, I establish a series of intermediate results. As a first step, I show that the 2-step transition density can be broken up into a product of the form (5.5) in Definition 5.2.

**Proposition E.1.** *The 2-step transition density of model with transitions* (5.14) *can be written as*

$$p_2(x_t|x_{t-2\tau},\sigma,\boldsymbol{\theta}) = p_1(x_t|x_{t-2\tau},\sqrt{2}\sigma,\boldsymbol{\theta})\cdot v(x_t,x_{t-2\tau};\tau), \tag{E.8}$$

*where the function $v$ is given by*

$$v(x_t,x_{t-2\tau};\tau) = \frac{\int_{\mathbb{R}} k_{\sqrt{2}\sigma}(y;x)w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}{\int_{\mathbb{R}} k_{\sigma}(y;x)w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}\int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\Big(z;\frac{1}{2}(x_t+x_{t-2\tau})\Big)\frac{w_{\boldsymbol{\theta}}(z)}{\int_{\mathbb{R}} k_{\sigma}(y;z)w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}\,\mathrm{d}z. \tag{E.9}$$

Note that $v$ depends on $\tau$ through $\sigma$. For later convenience, I define

$$Q(x;\tau) := \frac{\int_{\mathbb{R}} k_{\sqrt{2}\sigma}(y;x)w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}{\int_{\mathbb{R}} k_{\sigma}(y;x)w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y} \tag{E.10}$$

$$I(x_1,x_2;\tau) := \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\Big(z;\frac{1}{2}(x_1+x_2)\Big)\frac{w_{\boldsymbol{\theta}}(z)}{\int_{\mathbb{R}} k_{\sigma}(y;z)w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}\,\mathrm{d}z. \tag{E.11}$$

*Proof.* The proposition can be shown with a straightforward calculation. The 2-step transition density is given by

$$\begin{aligned} &p_2(x_t|x_{t-2\tau},\sigma,\boldsymbol{\theta}) \\ &\quad= \int_{\mathbb{R}} \frac{k_\sigma(x_t;z)\,w_{\boldsymbol{\theta}}(x_t)}{\int_{\mathbb{R}} k_\sigma(y;z)w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}\,\frac{k_\sigma(z;x_{t-2\tau})\,w_{\boldsymbol{\theta}}(z)}{\int_{\mathbb{R}} k_\sigma(y;x_{t-2\tau})w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}\,\mathrm{d}z \\ &\quad= \frac{w_{\boldsymbol{\theta}}(x_t)}{\int_{\mathbb{R}} k_\sigma(y;x_{t-2\tau})w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}\int_{\mathbb{R}} k_\sigma(x_t;z)\,k_\sigma(z;x_{t-2\tau})\,\frac{w_{\boldsymbol{\theta}}(z)}{\int_{\mathbb{R}} k_\sigma(y;z)w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}\,\mathrm{d}z. \end{aligned} \tag{E.12}$$

The product of the two Gaussian densities in the integrand can be transformed as follows

$$k_\sigma(x_t;z)\,k_\sigma(z;x_{t-2\tau}) = k_{\sqrt{2}\sigma}(x_t;x_{t-2\tau})\,k_{\frac{\sigma}{\sqrt{2}}}\Big(z;\frac{1}{2}(x_t+x_{t-2\tau})\Big). \tag{E.13}$$

The two-step density therefore becomes

$$p_2(x_t|x_{t-2\tau}, \sigma, \boldsymbol{\theta}) = \frac{k_{\sqrt{2}\sigma}(x_t; x_{t-2\tau})\, w_{\boldsymbol{\theta}}(x_t)}{\int_{\mathbb{R}} k_\sigma(y; x_{t-2\tau}) w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y} \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\Big(z; \frac{1}{2}(x_t + x_{t-2\tau})\Big) \frac{w_{\boldsymbol{\theta}}(z)}{\int_{\mathbb{R}} k_\sigma(y; z) w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y}\,\mathrm{d}z. \quad \text{(E.14)}$$

The numerator of the first factor is the desired one-step density up to appropriate normalization. If we extend by the required normalization constant,

$$\int_{\mathbb{R}} k_{\sqrt{2}\sigma}(y; x_{t-2\tau}) w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y, \quad \text{(E.15)}$$

we obtain equations (E.8) and (E.9). □

It is now left to show that the function $v - 1$ is in the order of $\tau$ on its entire domain $\mathbb{R}^2 \times \mathbb{R}^+$. In particular, this means that for any fixed $\tau^*$, the function $v(x_1, x_2; \tau^*) - 1$ is bounded on $\mathbb{R}^2$ via $c\tau^*$ for a constant $c$. It turns out to be helpful to analyze $v$ separately on $\mathbb{R}^2 \times (0, \tau_0)$ and $\mathbb{R}^2 \times [\tau_0, \infty)$ for some $\tau_0$. Because the proof is simpler for large $\tau$, I present this result first.

**Lemma E.2.** *Let $w$ be continuous and bounded away from zero, that is there exist $L$ and $U$ such that $0 < L \leq w_{\boldsymbol{\theta}}(x) \leq U$ for all $x \in \mathbb{R}$. Let $w$ further be twice differentiable on $\mathbb{R}$ with $|w''(x)| < M$ for some $M$ and all $x \in \mathbb{R}$. For any $\tau_0 > 0$, we have $v(x_1, x_2, ; \tau) - 1 = \mathcal{O}(\tau)$ on $\mathbb{R}^2 \times [\tau_0, \infty)$.*

*Proof.* Let $\tau_0$ be a number away from zero and fixed. Our goal is to establish bounds on the functions $Q$ and $I$, as defined in (E.10) and (E.11), and to use these to place a bound on $v - 1$. Because $w$ is twice differentiable we can apply Taylor's theorem to obtain a linear approximation for $w$ using any point $x \in \mathbb{R}$,

$$w_{\boldsymbol{\theta}}(y) = w_{\boldsymbol{\theta}}(x) + w'(x)(y - x) + R(y), \quad \text{(E.16)}$$

where $R(y)$ is the remainder term. This leads to

$$\int_{\mathbb{R}} k_\sigma(y; x)\, w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y = w_{\boldsymbol{\theta}}(x) \int_{\mathbb{R}} k_\sigma(y; x)\,\mathrm{d}y + w'(x) \int_{\mathbb{R}} k_\sigma(y; x)\,(y - x)\,\mathrm{d}y + \int_{\mathbb{R}} k_\sigma(y; x)\, R(y)\,\mathrm{d}y, \quad \text{(E.17)}$$

where the first term on the RHS becomes $w_{\boldsymbol{\theta}}(x)$, because the kernel integrates to one, and the integral in the second term is the first central moment of the kernel,

hence vanishes. The remainder $R(y)$, using the Lagrange form, is given by $R(y) = \frac{w''(\xi)}{2}(y-x)^2$, for some $\xi$ between $x_2$ and $y$. Since the second derivative of $w$ is assumed to be globally bounded, we have $|R(y)| \leq \frac{M}{2}(y-x)^2$. We use this to place bounds on the third term, recognizing that the remaining integral $\int_{\mathbb{R}} k_\sigma(y;x)\,(y-x)^2\,\mathrm{d}y$ is the second central moment of the Gaussian kernel $k_\sigma$, which is given by its variance $\sigma^2 = \omega^2\tau$. Therefore,

$$w_{\boldsymbol{\theta}}(x) - \frac{M}{2}\omega^2\tau \leq \int_{\mathbb{R}} k_\sigma(y;x)\, w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y \leq w_{\boldsymbol{\theta}}(x) + \frac{M}{2}\omega^2\tau. \tag{E.18}$$

In general, the lower bound can be arbitrarily close to zero, therefore we cannot simply invert this inequality to obtain an estimate on the inverse of the integral. Instead, we use the bounds on $w$ and again the fact $\int_{\mathbb{R}} k_\sigma(y;x)\,\mathrm{d}y = 1$ for any $\sigma$ and any $x \in \mathbb{R}$ to establish

$$0 < L \leq \int_{\mathbb{R}} k_\sigma(y;x)\, w_{\boldsymbol{\theta}}(y)\,\mathrm{d}y \leq U, \tag{E.19}$$

which can be inverted. Since inequalities (E.18) and (E.19) hold for any $\sigma$ and any $x \in \mathbb{R}$, they allow us to place bounds on both $Q$ and $I$. For $Q$, we obtain

$$\frac{1}{U}\big(w_{\boldsymbol{\theta}}(x) - M\omega^2\tau\big) \leq Q(x;\tau) \leq \frac{1}{L}\big(w_{\boldsymbol{\theta}}(x) + M\omega^2\tau\big) \tag{E.20}$$

for all $x \in \mathbb{R}$, $\tau \in \mathbb{R}^+$. We can avoid the dependency of the bounds on $x$ by again invoking the bounds on $w$,

$$\frac{1}{U}\big(L - M\omega^2\tau\big) \leq Q(x) \leq \frac{1}{L}\big(U + M\omega^2\tau\big). \tag{E.21}$$

For the function $I$, we only make use of the bounds on $w$ and inequality (E.19) and get

$$0 < \frac{L}{U} \leq I(x_1, x_2;\tau) \leq \frac{U}{L} \tag{E.22}$$

for all $x_1, x_2 \in \mathbb{R}$, $\tau \in \mathbb{R}^+$. We can now continue to calculate $v - 1$. An upper bound is immediately given by

$$v(x_1, x_2;\tau) - 1 = Q(x_1;\tau)\, I(x_1, x_2;\tau) - 1 \leq \frac{U^2 - L^2}{L^2} + \frac{MU}{L^2}\omega^2\tau. \tag{E.23}$$

With only few more additional steps, we obtain a lower bound by simply drawing upon $L \leq U$, its squared version and its inverse,

$$-\big(v(x_1, x_2;\tau) - 1\big) \leq \frac{U^2 - L^2}{U^2} + \frac{ML}{U^2}\omega^2\tau \leq \frac{U^2 - L^2}{L^2} + \frac{MU}{L^2}\omega^2\tau. \tag{E.24}$$

Define $C := \frac{U^2-L^2}{L^2\tau_0} + \frac{MU}{L^2}\omega^2$ for the $\tau_0$ chosen up front. Then,

$$\begin{aligned} |v(x_1, x_2; \tau) - 1| &\leq \frac{U^2 - L^2}{L^2} + \frac{MU}{L^2}\omega^2\tau - C\tau + C\tau \\ &= \frac{U^2 - L^2}{L^2} - \frac{U^2 - L^2}{L^2\tau_0}\tau + C\tau \\ &= \left(1 - \frac{\tau}{\tau_0}\right)\frac{U^2 - L^2}{L^2} + C\tau. \end{aligned} \tag{E.25}$$

The product on the RHS is non-positive for $\tau \geq \tau_0$, and hence $|v(x_1, x_2; \tau) - 1| \leq C\tau$ for all $\mathbb{R}^2 \times [\tau_0, \infty)$. □

The bounds on $Q$ and $I$, and thus $v - 1$, established in the preceding proof are not sufficient to conclude the result as $\tau \to 0$, unless $L = U$, which is the trivial case of a constant weighting function. More suitable bounds, however, can be found if inequality (E.18) can be inverted. This can be achieved by assuming $\tau$ to be small enough.

**Lemma E.3.** *Let $w$ be continuous and bounded away from zero, that is there exist $L$ and $U$ such that $0 < L \leq w_{\boldsymbol{\theta}}(x) \leq U$ for all $x \in \mathbb{R}$. Let $w$ further be twice differentiable on $\mathbb{R}$ with $|w''(x)| < M$ for some $M$ and all $x \in \mathbb{R}$. Let $\tau_0 = \frac{2L}{M\omega^2}$. Then $v(x_1, x_2, ; \tau) - 1 = \mathcal{O}(\tau)$ on $\mathbb{R}^2 \times (0, \tau_0)$.*

*Proof.* Here we develop bounds on $Q$ and $I$ such that both $Q - 1$ and $I - 1$ are in the order of $\tau$. Let $\tau \leq \tau_0$ for $\tau_0$ as defined in the lemma. Then the lower bound of equation (E.18) is bounded away from zero,

$$w_{\boldsymbol{\theta}}(x) - \frac{M}{2}\omega^2\tau \geq w_{\boldsymbol{\theta}}(x) - \frac{M}{2}\omega^2\tau_0 > w_{\boldsymbol{\theta}}(x) - \frac{M}{2}\omega^2\frac{2L}{M\omega^2} = w_{\boldsymbol{\theta}}(x) - L \geq 0. \tag{E.26}$$

Hence we can invert the inequality (E.18) and obtain

$$\frac{w_{\boldsymbol{\theta}}(x) - M\omega^2\tau}{w_{\boldsymbol{\theta}}(x) + \frac{M}{2}\omega^2\tau} \leq Q(x; \tau) \leq \frac{w_{\boldsymbol{\theta}}(x) + M\omega^2\tau}{w_{\boldsymbol{\theta}}(x) - \frac{M}{2}\omega^2\tau}. \tag{E.27}$$

Note that the values in the numerators and denominators differ slightly because the variances of the kernel $k$ in the numerator and denominator of $Q$ differ by a factor of 2.

Since $2w_{\boldsymbol{\theta}}(x) - M\omega^2\tau \geq 2L - M\omega^2\tau_0 > 0$, we can conclude

$$\begin{aligned} Q(x;\tau) - 1 &\leq \frac{w_{\boldsymbol{\theta}}(x) + M\omega^2\tau - w_{\boldsymbol{\theta}}(x) - \frac{M}{2}\omega^2\tau}{w_{\boldsymbol{\theta}}(x) - \frac{M}{2}\omega^2\tau} \\ &= \frac{M\omega^2\tau}{2w_{\boldsymbol{\theta}}(x) - M\omega^2\tau} \leq \frac{M\omega^2\tau}{2L - M\omega^2\tau_0}, \end{aligned} \tag{E.28}$$

for all $x \in \mathbb{R}$ and $\tau < \tau_0$. Using $2w_{\boldsymbol{\theta}}(x) + M\omega^2\tau \geq 2w_{\boldsymbol{\theta}}(x) \geq 2L$, we similarly obtain,

$$-(Q(x;\tau) - 1) \leq \frac{3M\omega^2\tau}{2w_{\boldsymbol{\theta}}(x) + M\omega^2\tau} \leq \frac{3M}{2L}\omega^2\tau \tag{E.29}$$

for all $x \in \mathbb{R}$ and $\tau < \tau_0$. If we set

$$C_1 := \max\left(\frac{M\omega^2}{2L - 2\omega^2\tau_0}, \frac{3M\omega^2}{2L}\right), \tag{E.30}$$

it follows that $|Q(x;\tau) - 1| \leq C_1\tau$ on $\mathbb{R}^2 \times (0, \tau_0)$.

Using analogous arguments as before, we can find an upper bound on $I$,

$$\begin{aligned} I(x_1, x_2; \tau) &= \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\Big(z; \frac{1}{2}(x_1 + x_2)\Big) \frac{w_{\boldsymbol{\theta}}(z)}{\int_{\mathbb{R}} k_\sigma(y; z) w_{\boldsymbol{\theta}}(y)\, dy}\, \mathrm{d}z \\ &\leq \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\Big(z; \frac{1}{2}(x_1 + x_2)\Big) \frac{w_{\boldsymbol{\theta}}(z)}{w_{\boldsymbol{\theta}}(z) - \frac{M}{2}\omega^2\tau}\, \mathrm{d}z \\ &= \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\Big(z; \frac{1}{2}(x_1 + x_2)\Big) \frac{w_{\boldsymbol{\theta}}(z) - \frac{M}{2}\omega^2\tau + \frac{M}{2}\omega^2\tau}{w_{\boldsymbol{\theta}}(z) - \frac{M}{2}\omega^2\tau}\, \mathrm{d}z \\ &= \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\Big(z; \frac{1}{2}(x_1 + x_2)\Big)\, \mathrm{d}z + \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\Big(z; \frac{1}{2}(x_1 + x_2)\Big) \frac{\frac{M}{2}\omega^2\tau}{w_{\boldsymbol{\theta}}(z) - \frac{M}{2}\omega^2\tau}\, \mathrm{d}z \\ &\leq 1 + \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\Big(z; \frac{1}{2}(x_1 + x_2)\Big) \frac{\frac{M}{2}\omega^2\tau}{L - \frac{M}{2}\omega^2\tau_0}\, \mathrm{d}z = 1 + \frac{M\omega^2\tau}{2L - M\omega^2\tau_0}. \end{aligned} \tag{E.31}$$

A lower bound is given by

$$\begin{aligned} I(x_1, x_2; \tau) &\geq \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\Big(z; \frac{1}{2}(x_1 + x_2)\Big) \frac{w_{\boldsymbol{\theta}}(z)}{w_{\boldsymbol{\theta}}(z) + \frac{M}{2}\omega^2\tau}\, \mathrm{d}z \\ &= 1 - \int_{\mathbb{R}} k_{\frac{\sigma}{\sqrt{2}}}\Big(z; \frac{1}{2}(x_1 + x_2)\Big) \frac{\frac{M}{2}\omega^2\tau}{w_{\boldsymbol{\theta}}(z) + \frac{M}{2}\omega^2\tau}\, \mathrm{d}z \geq 1 - \frac{M\omega^2\tau}{2L}. \end{aligned} \tag{E.32}$$

Setting $C_2 := \frac{M\omega^2\tau}{2L - M\omega^2\tau_0}$, we obtain $|I(x_1, x_2; \tau) - 1| \leq C_2\tau$ on $\mathbb{R}^2 \times (0, \tau_0)$.

We can now estimate $v - 1$ as follows,

$$\begin{aligned} |v(x_1, x_2; \tau) - 1| = |Q_\tau \, I_\tau - 1| &\leq |Q_\tau - 1| \, |I_\tau - 1| + |Q_\tau - 1| + |I_\tau - 1| \\ &\leq C_1 \, C_2 \tau^2 + (C_1 + C_2) \, \tau \leq \big( C_1 \, C_2 \tau_0 + C_1 + C_2 \big) \, \tau, \end{aligned} \tag{E.33}$$

for all $x_1, x_2 \in \mathbb{R}$ and all $\tau < \tau_0$. □

Lemmata E.2 and E.3, together with proposition E.1 prove Theorem 5.3.