Which Route to Choose? Generating Context-Aware and Empirically Informed Random Trajectories to Study Movement

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Georgios Fanouriou Technitis

aus

Griechenland

Promotionskommission

Prof. Dr. Robert Weibel (Leitung der Dissertation)

Dr. Kamran Safi

Prof. Dr. Ross Purves

Prof. Dr. Patrick Laube

Zürich, 2021

Summary

Understanding the rules that govern movement and its implications to life is the focus of movement ecology. Over the past two decades, the advances in tracking technologies have enabled an unprecedentedly accurate and detailed view of animal behaviour. However, the massive number of unobserved individuals still does not allow conclusive inference. Null models are therefore needed to assist the ecological knowledge generation regarding animal movement, its motivation, and the relationship to the context it takes place in.

The overarching aim of this thesis is to develop algorithms for generating ecologically realistic trajectories that enable hypothesis testing in the movement ecology domain, even in cases when sample data is scarce — which is still frequently the case, given technical constraints such as limited battery life of tracking sensors. The methodology of the thesis is based on the paradigm of movement ecology put forth by Nathan et al. (2008).

I investigate and propose ways to generate individual, context-aware trajectories that connect known endpoints and geometrically similar to observed trajectories. The research work is organised into four research objectives, addressing (1) the conceptual simulation framework design, (2) the generation of a possible trajectory, (3) the generation of the probable trajectory, and (4) ultimately the context incorporation. The main chapters of the thesis cover these four research objectives.

Regarding Research Objective 1, a conceptual framework of movement simulation is proposed that defines the necessary functionality of a movement simulation facility. The framework also maps the relationship of the essential components of the ecological paradigm by Nathan et al. to the movement model and contributed to articulating the trajectory generation strategy. Based on a set of requirements defined together with behavioural biologists, the random walk model was selected as a base movement model. The empirical movement parameters extracted from observed movement (tracking data) as the means for conditioning and evaluating the generated trajectories.

Research Objective 2 involves creating a trajectory that connects two locations separated by a significant geographic distance, while honouring the available time for the move and the speed limits of the mover. This task led to the proposal of an algorithm called the Random Trajectory Generator (RTG). The algorithm inputs the coordinates of the two endpoints, a constant speed value, and the time maximum time available to effectuate the movement. The algorithm handles both planar and spherical coordinates systems (such as the sphere representing the Earth) and generates *possible* trajectories connecting the two endpoints. The generated trajectories consist of points constrained by the 'potential path area', the area that can maximally be reached at the given speed and with the given time budget.

Research Objective 3 was met with the proposal of the Empirical Random Trajectory Generator (eRTG). This algorithm requires one or several observed reference trajectories as input data, used for extracting movement behaviour in the form of movement parameters. The trajectory generation process incorporates the movement parameters into a combination of forward-looking and destination-based (i.e., backward-looking) vector fields. This approach ensures that the mover will reach the predefined destination while expressing movement behaviour statically similar to the input data.

To meet Research Objective 4, context awareness was added to the basic eRTG algorithm. Static spatial context was incorporated in two ways: a binary layer and a ratio-scaled layer of values in the interval [0..1]. With the binary layer, areas can be represented that should be avoided, in the same manner migrating birds avoid large water bodies. With the ratio-scaled layer, the mover's preference is captured as a probability surface with a higher probability on areas more likely for the mover to move on, e.g. migrating birds flying over areas with increased vegetation cover.

Applications of the eRTG in migration and space utilisation studies are addressed, along with a discussion of the implications for landscape connectivity and survival strategy research.

Acknowledgements

The work for this dissertation was funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) through project CASIMO (C09.0167) and carried out at the Department of Geography, University of Zurich. I am thankful to both organisations for supporting my work, but equally so to the people inside and outside of the two, that offered their support to me during the past years. With the next few words, I can only hope to give a glimpse of the gratitude I feel towards them.

I would first like to thank my supervisor, Professor Dr. Robert Weibel, for the time and patience he invested in my work, without which the completion of this dissertation would be uncertain. Rob, I can't thank you enough for being there for me from day one, the good discussions and times we have shared and the inspiring buoyant coolness you deal with things. Truly, you were the first person to make GIUZ home for me.

I would then like to thank my second supervisor, Dr. Kamran Safi, whose expertise was invaluable in formulating the research questions and methodology. The clarity of vision, technical knowledge and most importantly his deep love for ecology and science made me alter my domain significantly and start *tracing* birds, instead of humans. I will always be thankful for that!

Special thanks to the members of my PhD Committee, Professor Dr. Ross Purves and Professor Dr. Patrick Laube for sharing their knowledge throughout the years and their feedback on my dissertation, as well as Professor Dr. Somayeh Dodge for acting as external reviewer and examiner.

To Merlin Unterfinger, for taking a working algorithm, extending it and coding it in an R-package. Your passion, efficiency and dedication, Merlin, are admirable, and your support was instrumental to my work.

To Dr. Walied Ohtman, who made me think more 'spherically' and who spent with me countless hours on the blackboards messing with curvature distortions and Belgian beer qualities.

To the silent hero, Dr. Bart Kranstauber, who helped me move from Python to R, and who brainstormed algorithms and debugged implementations with me.

To my office mates, Ali & Peter — were the next people who turned our office into our home. PhD time would have been boring without you guys! Memories are priceless. Thank you for everything!

To Dani, Alex & Ele for their continuous practical, physical and moral support during debugging, plotting, proofreading and rough Champion's league nights.

To the rest of the folks on the J, K, L floors for all the work, trips, hikes, games, talks, dinners, runs, frisbees, celebrations, laughs and tears we had during these years.

To Professor Dr. Emmanuel Stefanakis, who shared with me the PhD advertisement at first place. If it was not for you, chances are I would have never even seen the Ph.D. ad!

Τέλος, θα ήθελα να ευχαριστήσω την Ελεάννα, Ηρώ & Φάνη για την αγάπη και την υποστήριξη που απολαμβάνω όλα αυτά τα χρόνια. Τα καλύτερα έπονται.

Contents

Abbreviations

ABM Agent-Based Modelling.

BBMM Brownian Bridges Model of Movement.
BCRW Biased Correlated Random Walk.
BRW Biased Random Walk.

CDFs Cumulative Distribution Functions.
CRW Correlated Random Walk.
CsCRW context-sensitive correlated random walk.
CT Complexity Theory.

DES Discrete Event Simulations.
DRW Drifted Random Walk.

eRTG Empirical Random Trajectory Generator.
ETDM Elliptical Time Density Model.

FIFO First-In-First-Out.

GIS Geographic Information Systems.
GIScience Geographic Information Science.
GNSS Global Navigation Satellite System.
GST General Systems Theory.

MBR Minimum-Bounding Rectangle.
MEF Movement Ecology Framework.

NDVI Normalised Difference Vegetation Index.

PD Probability Distribution.
PPA Potential Path Area.
PPA Potential Path Area.
PpA Potential point Area.
PPS Potential Path Space.

RRW Reinforced Random Walk.

RTG Random Trajectory Generator.
RW Random Walk.

SD System Dynamics.
STP Space-Time Prism.

UD Utilisation Distribution.
uERW unconditional Empirical Random Walk.

VHF Very-High-Frequency.

1 Introduction

> *If we knew what it was we were doing, it would not be called research, would it?*
>
> **— Albert Einstein**
> (theoretical physicist)

1.1 Motivation

1.1.1 Importance of movement

Seven processes separate living from non-living entities: movement, respiration, sensitivity, growth, excretion, reproduction, and nutrition [1]. Out of these, movement is the only one that depends on all the others: for example, an animal needs energy from food in order to move, it adapts its movement based on the availability of energy resources, adult individuals have more movement capacity than infants, and during mating season, some species travel great distances to increase the chances for mating encounters. Consequently, by analysing the movement of an individual, we can better understand its physical state, its perception of the environment, and ultimately its life. One could think of movement as a biological signature of an organism, with each movement expresses a set of behaviours that a research can use as a piece in the eco-puzzle of (social) interactions, habitat use, ecological traps, niches, and beyond.

In ecology, the focus on individual behaviour analysis allows for aggregations of and insight into species' survival strategy [2–4]. This information is crucial for conservation biology, as it can provide an early warning of a declining survival rate. Questions such as "*How do animals cope with their environment?*" [5] and "*Which individual behaviours and adaptations help certain species maximise their chances of survival?*" [6–8] can lead to many spatial repercussions, including habitat selection methods, dispersal strategies, and migration patterns. All are, however, linked to the initial question: "*Why does an animal move?*"

1.1.2 Movement in Geographic Information Science and Ecology

Geographic Information Science (GIScience) can extend its analytical capacities from static to dynamic phenomena by understanding the underlying mechanisms of movement.. The theoretical concepts established in the 80s of the temporal Geographic Information Systems (GIS) are now being further developed based on the manifold spatio-temporal data and methods available. This development enables the researcher to analyse large volumes of spatio-temporal (movement) data, identify correlation patterns and ultimately derive real insight into the mover's behaviour. The deductive knowledge we gain on the mover's behaviour brings GIS a step closer to realising its full potential, as Peuquet described in her seminal article: the "*greater promise of spatio-temporal GIS resides ultimately in their capacity to examine causal relationships and their effects in any of four modes of inquiry: exploration, explanation, prediction, and planning*" [9]. This capacity allows us to improve our methodological frameworks further.

The integration of movement in ecology and GIS

Over the last three decades, movement research has gained significant interest in both ecology and GIScience. Nevertheless, Gurarie identified a lack of attention on movement in his review [10] of "Foundations of Ecology: Classic Papers with Commentaries" [11], where only one out of the included fifty articles refers explicitly to movement. In a similar manner within the GIScience domain, despite community consensus on the importance of spatio-temporal system development early on in GIS, the actual research focused for more than half of its lifespan on static data analysis, as captured in Figure 1.1. Jonathan Raper encapsulated the need to promote spatio-temporal research in his keynote speech at the 2nd International Conference on GIScience 2002 in Boulder, CO: "*It's time to abandon the 'fetish of the static' "* as quoted in Laube [12, p. 6].

Understanding why GIScience and Ecology almost synchronously adopted movement as an critical element to explore might prove difficult, though three trends appearing over the same time in both disciplines suggest some of the reasons for the increased interest in movement:

1. technological advances in (animal) movement tracking
2. a shift in focus from static to dynamic analysis
3. increasing interest in the individual movement

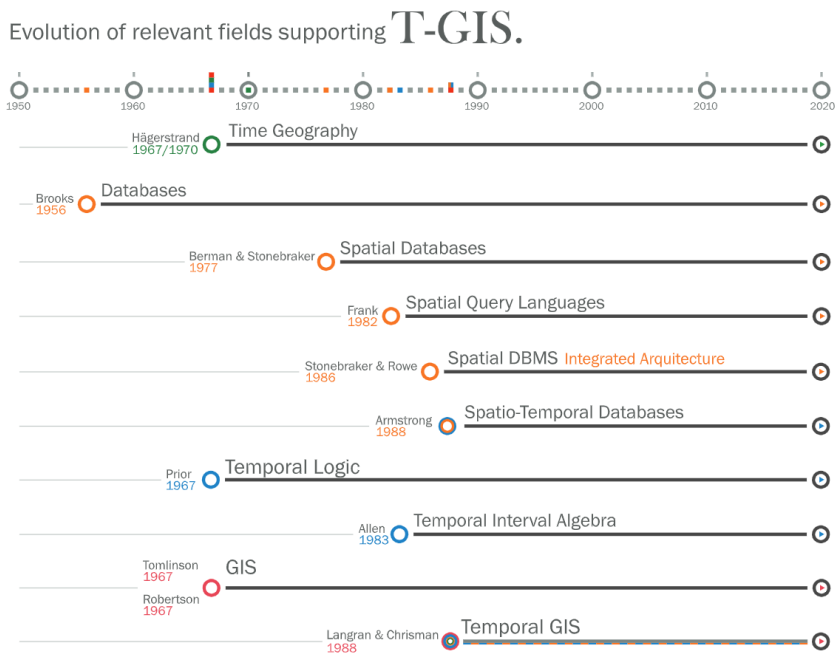

Figure 1.1: Milestones in the development of Temporal GIS, as presented in Siabato et al. [13].

These trends form the starting point of this thesis.

Technological advances in (animal) movement tracking

Technological innovation offers better, faster and more rigorous data collection, often leading to rapid conceptual advances [14]. In order to keep up with the shifts in the theoretical ecological frameworks, one should follow up on the current technological trends.

The past three decades have seen the advent of affordable and ubiquitous technologies that enable tracking moving objects of various kinds at increasing levels of resolution and accuracy (Figure 1.2). Moving from Very-High-Frequency (VHF) to Global Navigation Satellite System (GNSS) such as GPS, tracking has become far more accurate, precise, frequent and light-weight, giving researchers the chance to monitor more refined animal behaviour and allowing them to record processes that had not been recorded before [15–18].

The main limitations of using such technology are a function of available funds and biometrics. From the financial perspective, monitoring large numbers of individuals still requires a sizeable investment into tracking equipment even though now, more than ever, off-the-shelf products claim improved quality for a lower price. Admittedly, while cheaper devices are becoming increasingly present, quality control

during manufacturing is still an issue, especially in inter-study comparison research. Delivering a standard out-of-the-box accuracy and precision in measurement is hard even among devices using the same technology (e.g. GPS). Factors such as the antenna position, the internal clock accuracy, the environment of study, etc., affect the performance of the devices [19], often forcing the researcher to opt for the more expensive equipment when aiming at a stable accuracy and precision.

From the biometric perspective, the researcher needs to balance the data triad, consisting of the weight of the recording device, the volume of the data to be acquired, and the analysis needed to perform on this data. The connection between these three elements is relatively intuitive: more weight means larger battery capacity, which allows for more data recordings, eventually leading to increased data processing demands. This relationship raises the following considerations. In real life, the weight of the device is limited by the animal's carrying capacity, given that the carried equipment's weight should be less than 3-5% of the animal's total body mass not to interfere with its natural behaviour [20–23]. This percentage threshold, in many cases, imposes few tens of grams limit on the weight of the tracking device, forcing the researcher to choose between battery life or frequency of data. When the weight of the tracking device is not an issue and data acquisition has virtually no energy or capacity limitations, the fieldwork naturally results in accumulating large volumes of data. The collected data needs curation throughout their life-cycle, from creation and initial storage to archiving and deletion. Looking at high definition recordings (such as accelerometer, audio or video recordings), the data of a single individual can add up to hundreds of gigabytes in size – the transferring and pre-processing of which often surpasses the memory and processing limits of a single lab computer and requires various parallel processing methods commonly deployed on cluster or cloud solutions. Setting up such data pipelines, and their curation is getting easier by the day, but it adds an extra level of consideration when tracking animals.

Additional considerations for using the collected data include carefully addressing the accuracy, precision and consistency metrics of the derived data [24, 25] and their relevance and utility for the intended study. As Cagnacci et al. [14] pointed out, any GPS location is precisely just that: the mere location of where an individual has been at a specific time, neither its activity nor its use of resources.Any inference should be conducted with caution when no further information accompanies the position record, e.g. contextual data or health status. Nevertheless, overall, technology undoubtedly increased researchers' monitoring capacity, impacting the underlying theoretical framework in animal ecology [18].

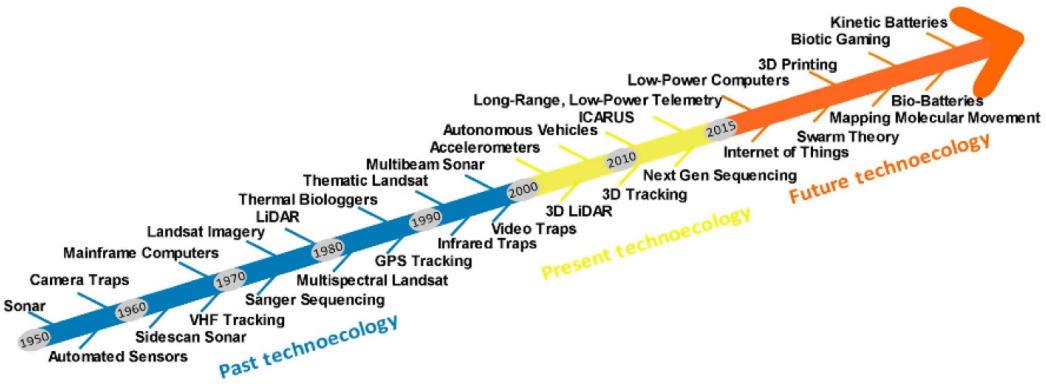

Figure 1.2: Timeline of tracking technologies in Ecology and Environmental Science. [18]

The shift of focus from static to dynamic analysis

Technology advances improved data collection methods, created necessary data-types and altered the information available to researchers, forcing them to revisit their data analysis practices. For many years the lack of appropriate software and hardware impeded to adding the temporal dimension to spatial analysis [26], but new practices enable researchers to account for their data's temporal dimension correctly. New technology allows for both the collection and, more importantly, processing several orders of magnitude more data than before, though this technological advance has not directly translated into scientific insight. Despite improved computational capabilities, it has become apparent that data computation is not the only challenge a researcher should overcome when dealing with the temporal dimension. The low-level structural problems of time [27] make its incorporation into the GIS data models a strenuous task [28].

Spatial queries executed in continuous sequence have added a pseudo-dynamic element to GIS models, enabling them to deal (at least on an elementary level) with state-changing snapshots of a system. A relatively simplified example of such a system is to monitor the vegetation over a given area region. Given consecutive "world-state" snapshots of the region, one can track the evolution of the vegetation's quality [9].

Being able to acquire such temporally sequential snapshots of a dynamic system introduces a first pseudo step towards a more dynamic analysis of spatio-temporal data. Moving from this to an actual dynamic model construct with fully enabled spatio-temporal database management and visual analytics of spatio-temporal data requires more integrated temporal data models and methods [29]. In order to move towards this direction, the GIScience research community sought inspiration in other domains such as physics, mathematics, and computer science. Computational fluid

dynamics [30, 31], cellular automata [32–34] and agent-based modelling [35, 36] are only a few examples of the methodological approaches adopted by GIScience.

Slowly but steadily, the refinement and tailoring of the snapshot models have aligned with the concepts of state-space modelling systems, the space-time cube [37], and time geography [38], all falling under the umbrella term *Dynamic GIS* [28]. Each approach carries its advantages and shortcomings, but all together describe the energetic field of Dynamic GIS in the past three decades, a field that has helped movement analysis and modelling gain significant popularity.

The increasing interest of the individual movement

The importance of accounting for individual behaviour instead of aggregate dynamics has been gaining significant momentum in the ecological community over the past two decades [39–47]. Contrary to the traditional approach that represented an individual's behaviour as noise in the populations' evolutionary signal, and thus considers it unimportant, a strong trend in the community stresses the value of analysing individual-level decisions as an essential element for exploring population dynamics since " *crucial ecological quantities (...) depend critically on these individual-level decisions*" [43].

Moving a step further, analysing the inter-individual movement variation provides insights into how "*the [species'] dispersal correlates with the establishment and colonisation of new areas*" [41], an element crucial to the evolution and survival of a wild population.

Finally, understanding the influence of the individual on group decisions gives us insights into the social structure and dynamics of the group [15], eventually helping us answer not only how collective movement occurs but also why it is initiated in the first place [48].

1.1.3 Thesis rationale

Focusing on the movement of individuals and enriching their trajectories with ecological information can help us understand how the mover reacts to specific contextual changes. Monitoring this reaction allows us to test hypotheses regarding the decision-making mechanism behind the movement and ultimately aim to why the individual moved first place. When a mover is an animate object, gaining even a small amount of knowledge on this motivation provides us with a better understanding of a fundamental element of life itself.

In recent years, the movement ecology community has significantly improved its attempts and capacities to track the movement of individuals (1.1.2, though not all limitations have been surpassed. Field measurements often come in lower than intended quality, with missing records (i.e., data gaps) and irregularities ([49,50].

The research community has developed modelling techniques that allow to improve the quality of the recorded tracking data and therefore gain more information from the raw data, such as statistical filtering to obtain more robust, noise-free positions, or interpolation techniques to fill in data gaps or densify coarse data by additional position estimates. Some of these techniques employ complex statistical or machine learning models that have the advantage of leading to improved data, but often at the cost of reducing the interpretability of the results, particularly when black-box models are used [51–54]. In order to maximise reproducibility, more straightforward and explainable models are required. This requirement is particularly true when the task is to close gaps in fragmentary, incomplete, or coarsely sampled tracking data, because this specific task is completely model-based, and hence it is imperative that the model employed is transparent and understandable to the user. Using transparent and explainable models will allow the researcher to test hypotheses while controlling the ecologic validity of the model's parameters.

The **overall objective** of this thesis is to contribute to this problem area by proposing methods to densify coarsely sampled movement tracking data or fill in gaps in incomplete, fragmentary movement data. This endeavour will be addressed by combining the perspectives of three research domains – GIScience, Ecology, and Computer Science (Figure 1.3) – with the knowledge construction taking place mainly at their intersection: the domain of *movement ecology* [55]. GIScience provides spatio-temporal methods to formalise ecological knowledge, while computer science provides the tools and algorithms needed to address ecological research questions. More specifically, the role of GIScience is to translate the research questions into research objectives and then approach them with tailored, representative and explainable methodologies. Standardising the "tailoring process" allows us to combine nomothetic (law-driven) and idiographic (data-driven) views in a unified approach. The nomothetic part expresses rules that govern the population assuming homogeneity in behaviour among the individuals. For the use cases in which this assumption is violated and where capturing the heterogeneity of the population is important, the idiographic approach has the advantage of serving the user better as it models behaviour on an individual-specific basis. On the other hand, working on an individual level is more prone to over-fitting the system's behaviour to the few individuals used as input data. Having the ability to interplay between the two views adds significant flexibility to the model development process, allowing to tap into the assets that best fit each case study, e.g. having a homogeneous global structure

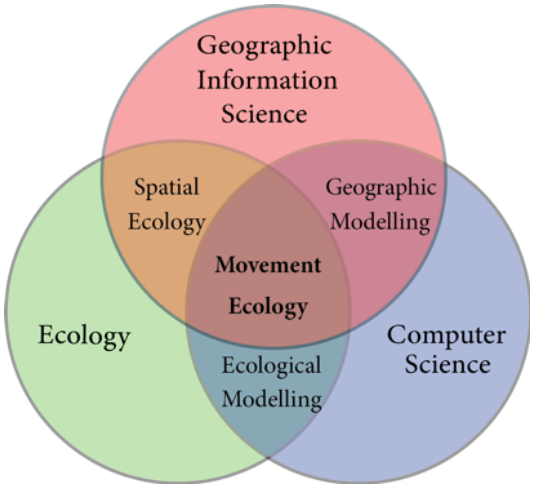

Figure 1.3: The intersection of sciences behind movement ecology.

of behaviour based upon the behaviour of individuals allows to build models that extend the expression capabilities for particular subsets of the population.

Random Walk (RW) was selected as the main underlying framework for movement due to allowing the user to work individual movement in a mathematically extensible and straightforward manner. It can systematically distinguish, control and test underlying movement mechanisms as well as the effect of context on movement from observed data, and last but not least, the RW can introduce the necessary inter- and intra-variability of movement, bridging the gap between the individual data observed and the insights that can be drawn at the population level.

1.1.4 Main research question and research problem

As laid out in the preceding section, this thesis explores how to utilise conceptual and methodological knowledge in movement modelling in a framework that attempts to help researchers answer movement preference questions by testing hypotheses on imperfect, incomplete datasets. This motivation can be cast into the **overarching research question**:

How can we best model movement and context in a simulation framework to better assist in ecological knowledge production?

Specifically, this thesis examines the following **research problem**: *Connect two known endpoints separated by a rather large distance with an ecologically realistic trajectory in a simulation framework that will assist the movement ecologist in testing assumptions and derive explainable results*. In this context, a realistic trajectory is any trajectory geometrically similar to an observed trajectory (see glossary section).

Following the review of state of the art in areas relevant to this research question in Chapter 2, the research gaps will be summarised in Section 2.4 and the specific research objectives addressed by this work will be presented in Section 2.5.

1.2 Thesis structure

The following chapter discusses the related work, providing anchor points for this research in the wider scientific community. Chapters 3 to 6 contain the core scientific contribution of the thesis by introducing the conceptual simulation framework (Chapter 3) and proposing algorithmic simulation approaches of increasing complexity and ecologic realism (Chapters 4 to 6). Chapter 7 discusses the contributions and shortcomings of the methodology developed. Finally, Chapter 8 ends the thesis with the main conclusions and an outlook on potential future work.

2 Background

> *To steal ideas from one person is plagiarism; to steal from many is research.*
>
> **— Steven Wright**
> (stand-up comedian, writer)

Moving on from the overarching research question and main research problem laid out in Section 1.1.4, this chapter aims to review the state of the art in relevant research areas to identify important research gaps and spell out the detailed research questions of this thesis.

First, Section 2.1 presents the history of movement modelling in a nutshell, from Galileo to Random Walk (RW) theory. Next, Section 2.2 is dedicated to the mathematical models commonly used in GIScience and ecology to represent physical movement in ecological studies. Section 2.3 then reviews a conceptual framework of movement in ecology that allows linking the mathematical movement models to the components creating the movement path of a mover. Section 2.4 summarises the research gaps motivating this work. Finally, Section 2.5 revisits this work's overarching research question and formulates the detailed research objectives addressed in the remainder of this thesis.

Note that abbreviations used in this thesis are explained in the Abbreviations section.

2.1 From Galileo to random walks

> "*A large vessel of water [was] placed in an elevated position; to the bottom of this vessel was soldered a pipe of small diameter giving a thin jet of water, which we collected in a small glass during the time of each descent ... the water thus collected was weighed, after each descent, on a very accurate balance. ... This was done with such accuracy that although the operation was repeated many many times, there was no appreciable discrepancy in the results.*" (Galilei 1638, translation by Cupitt 1988)

Galileo's clock used in the measurement of phenomena, described above, was a milestone for separating two sciences: the science of the material and the science of motion. This specific publication inspired Descartes to form his theory on momentum and motion during his time at Leiden University [56].

Van Schooten, who also worked in Leiden at the time, introduced Descartes' work to his student Christian Huygens, who then solved the so-called tautochrone problem (*"finding the curve down which a bead placed anywhere will fall to the bottom in the same amount of time"* [57]) by geometrically proving his *Horologium oscillatorium*, concluding that the requested curve was a cycloid [58].

The sequence of these events created momentum and increased the popularity of the tautochrone problem in the scientific society, finally bringing together two bright young scientists who would create the reference for movement modelling: Lagrange and Euler. Inspired by this work, the two collaborated for almost 20 years to define the equation that describes Huygens' curve, not for just one starting point but for every point on a plane: a primary form of what we today would call a flow field. During this period, each kept a strong stance on the perception of movement in space, concluding in two of the most influential movement modelling approaches to date: aptly termed the Lagrangian and the Eulerian approach, respectively.

The Lagrangian approach describes the movement from the mover's point of view: the researcher follows the mover and can readily obtain information regarding its location and velocity. The Eulerian approach externally monitors movement in its context with fixed locations of observations, acting as 'gates' that the mover crosses. Since the tracking is not continuous, it offers snapshots of information about the movement. To better understand the difference between these two approaches, the following example might be helpful: on a directional, 1D space, an object moves from 0 to 100 in 10 seconds (Figure 2.1). Our observation stations are at points 2 and 98. The Eulerian method measures the speed at $s = 2$ and 98 and returns the average speed, assuming a uniform distribution of speed in-between the points. In contrast, the Lagrangian approach gives information on the movement every second, revealing the underlying distribution of speed and acceleration.

These two different approaches created the basis of two movement ecology traditions [59]. The supporters of the Lagrangian approach moved towards expressing movement as incremental steps and turns with no prior knowledge regarding the context of a movement or the animal's preference. The supporters of the Eulerian perspective, on the other hand, describe movement from fixed external points – a purely contextual perspective [60]. In both cases, movement is often expressed with diffusion equations [43, 49]. If the x and y coordinates are independent, then we have an underlying Markovian diffusion, or if described by a single trajectory: the

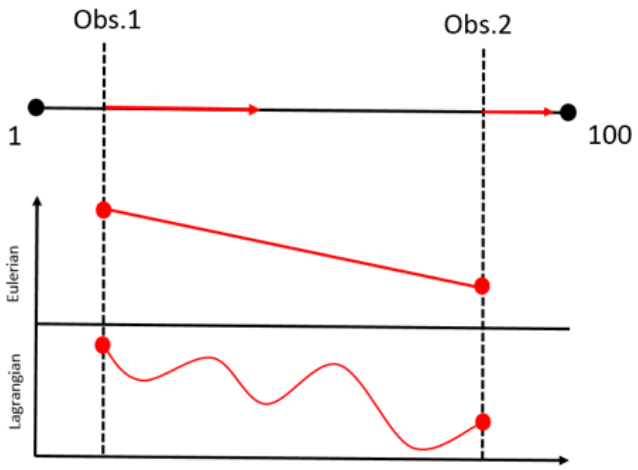

Figure 2.1: Visual representation of the movement as seen through Eulerian and Lagrangian approaches.

Random Walk [61, 62]. The potential of combining the two approaches has been explored mainly as a way to monitor emerging population patterns in the continuum (Eulerian approach) while focusing on controlling the discrete movement of the individual (Lagrangian approach) [63, 64]. The synergy potential can be further explored in cases, for example, of large scale spatial analysis of bird movement, where millions of location data-pairs of individuals (ring data), collected over many years [65] require methods to connect the one location of a particular pair with the other. The synergy between Euler's (ring/encounter data) and Lagrange's (individual trajectory) approach can leverage the significant historical data and population statistics of the former, with realistic ecological trajectories of the latter, as visually represented in Figure 2.1, offering plausible hypotheses for what may have happened between the two observations.

2.2 Models of movement in GIScience and Ecology

2.2.1 Random walk

RW was introduced in Mathematical Biophysics in the early 1950s when Patlak thoroughly investigated the relationship between various forms of correlation and diffusion approximations and ways to explicitly derive the expected displacement and orientation of the individual [66]. In this case, movement is discretised spatially and temporally, with each relocation described as step (s) and turning angle (θ) in the unit of time. Patlak himself, having limited data samples, focused on the theoretical background of the study. He concluded that such a model

"*will provide a systematic method for experimental investigation of the problem of the orientation of organisms ... [and] that by use of the knowledge of the various parameters not only may the behaviour of the organism in its natural environment be predicted but also the search for the underlying physiological mechanisms for this behaviour may be facilitated.*" [67].

This conclusion foresaw that his method could be used to investigate an individual's change of behaviour given a change in its environment. A few years later, Kareiva and Shigesada [68] backed up Patlak's assumptions with one of the most influential publications in the quantitative study of animal movement that included the definition of the Correlated Random Walk (CRW), a walk that involves a correlation between successive step orientations, termed: 'persistence' and its assessment on cabbage white butterflies (*Pieris rapae*) flying on a field. The result confirmed the rather intuitive hypothesis that the butterfly's relocations do not violate CRW assumptions unless a nectar source falls within its perception range:

"*It may be especially fruitful to investigate the foraging or searching movements of animals as CRWs whose parameters (moving angle or length) depend on local ecological conditions. Thus, for example, an organism might always move randomly, but with sharper angles in the presence of food resources.*" [68]

A significant extension of the RW that has been used for incorporating memory in the movement is the Reinforced Random Walk (RRW) model [69, 70], in which the previously visited locations affect the decision for movement in every step. Adding a sequential dependency in each step-pair of the walk is the core attribute of the CRW, making this type of walk widely used in various ecological applications [61,68,71].

In order to ensure that the moving object moves towards a specific direction, one can add a (local) bias on the selection of the next step, leading to a Biased Random Walk (BRW) (also known as Drifted Random Walk (DRW)).

Ahearn et al. [72] in the context-sensitive correlated random walk (CsCRW) interpret the local bias as a probability surface that incorporates the standard CRW process. The probability surface represents a contextual factor (in this case: slope preference) that affects the next step creation process (movement of two tigers). In the specific example a single factor was used but any number of contextual factors can be combined through a joint probability function.

If we combine bias with an average destination point or area, then the bias can be described as a mean-reverting Ornstein-Uhlenbeck process [73,74], which formalises the tendency of the Brownian process to move towards a set long-term (global) mean under the influence of friction [75]. Last but not least, combining the bias

with the correlation creates a Biased Correlated Random Walk (BCRW) , meaning a walk with both local and global preference of direction in its movement.

Empiricists tend to support the CRW, in which the mechanistic RW models are based on the observed parameter distributions of step length and turning angle, while the trajectory recorded describes the sum of the internal decisions taken by the mover. Theoretical ecologists, however, tend to opt for diffusion models, which are also easier to express formally [49, 71, 76–79].

The CRW and its derivatives are considered to provide intuitive ways to describe animal movement. However, a significant limiting factor is that each step only depends on the previous, making the whole movement highly susceptible to the time interval selected for each relocation. Nations and Anderson-Sprecher [80] incorporated autocorrelation in the form of bivariate distributions to tackle this dependency. In other research approaches [81–84], the variogram enabled a practical way to model the data-distance relationship (loosely translated as autocorrelation).

In summary, the RW and its derivatives have been widely used in various movement scenarios [62, 66, 68, 72, 85–99], where they allowed replicating realistic behaviours for a wide range of species.

By adding contextual factors that affect the movement scenario [72], replications tend to become even more flexible in implementation and therefore able to come closer to the real behaviour of the mover. Arguably the most fundamental shortcoming of the RW family, is that while these methods allow modelling the evolution of the path of a mover of time, they do not reach a single point destination given a time budget. Having such control over the destination is a core requirement of the general problem statement (1.1.4) in order to make better use of the recorded movement data. We can directly extract three elements as movement information: the starting location, the end location, and the time difference between the observations. RW fundamentally can only account for two out of the three elements. The two approaches reviewed below – the space-time prism and the Brownian bridges movement model – offer possible solutions to this requirement.

2.2.2 Space-time prism

Following Hägerstrand's theory of time geography [37], the deterministic concept known as the Space-Time Prism (STP) gained significant attention in the GIScience community [100–104]. The STP defines an envelope of accessibility given a specific time budget for a trip and the maximum speed of the mover, resulting in a Potential Path Space (PPS) (confined by a prism in a 3-D cube formed by the two axes of the

plane, and time as the third dimension) or its 2-D projection, the Potential Path Area (PPA), shown in Figure 2.2.

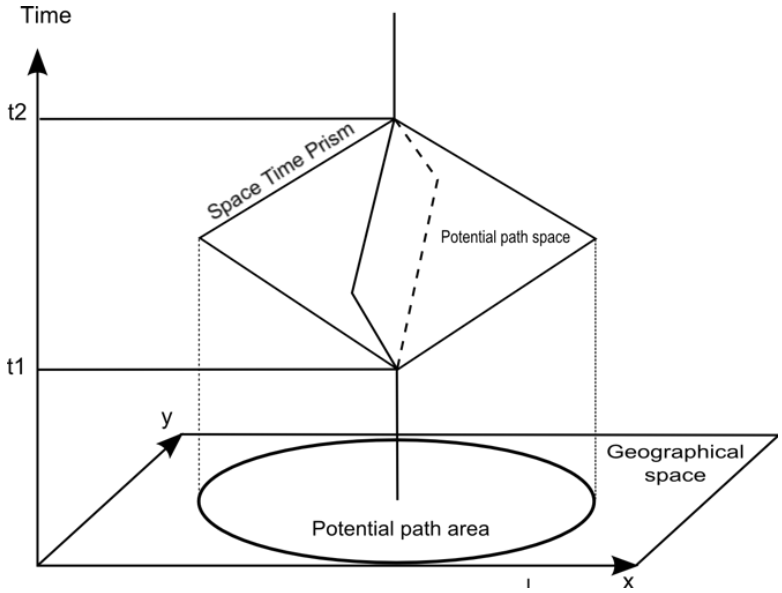

Figure 2.2: Potential path space in a space-time prism and potential path area (based on Miller 1991 [100]).

Various time geography approaches have been proposed to model movement in a planar space given a time budget and maximum speed [105] or in a network space [106], though the literature on kinetic restrictions on space–time prisms is lacking, and the ones existing focus mainly on interpolating the movement. In one-dimensional and two-dimensional space, Kuijpers et al. [107] introduce kinetic prisms and sketch methods for creating their boundaries. Long et al. [108] use skew-normal distributions as a heuristic to represent movement probabilities inside space–time prisms and apply the method to analyse data from wildlife, cyclists, and athletes. For correlated random walks, using caribou, cyclist, hurricane, and athlete tracking data, Long [109] compares methods for kinematic interpolation of movement data based on Bézier and Catmull–Rom splines with standard interpolation methods, concluding that kinematic approaches are superior for applications involving fast-moving objects [109]. Contextual information was introduced in the form of field-based time geography by Long (2018) [110] enabling STP to account for obstacles or limiting visit opportunities.

Although path interpolation algorithms can be utilised in movement research, they are not well suited to observed movement with a highly irregular sampling rate. Moving centroids, accuracy criteria, moving window filtering, and limiting sampling to data recorded can be used to reduce signal noise; however, these strategies have drawbacks, such as being susceptible to outliers and operational factors ([111]. Furthermore, the interpolation tends to force the mover to continuously move towards the endpoint, a phenomenon that is not necessarily realistic; an animal

wearing a GPS device might change the direction of its move in between two observed points entirely, multiple times.

Approaches relying on the STP [101, 103–105, 108, 112, 113] entail generating the envelope of accessibility in the form of the PPS in 3-D time-space and the PPA on the 2-D spatial plane, and can thus solve the problem of bounding the possible movement between two endpoints. However, they can only designate the area that the mover could have reached if constantly travelling at maximum speed. Apart from the fact that such extended maximum-speed movements are biologically relatively uncommon, the resulting PPS or PPA contains no information about the mover's visit preference inside that envelope. Removing this limitation of the basic STP approach, Song and Miller [114] presented an extension, whereby the Brownian bridges movement model (see 2.2.3) was used to generate visit probability distributions that were truncated to the PPS (and PPA) by the use of the STP model, generating zero probabilities in areas that cannot possibly be reached, however, this approach lacks the capability of generating individual trajectories with varying movement parameters (e.g., varying speed and turning angle distributions).

Loraamm ([115]) offered the construction of a behavioural space-time path for estimating visit probabilities in space-time prisms that incorporates animal behaviour and generates individual trajectories, which are informed by context and diurnal period. Loraamm presented the algorithm on a Muscovy Duck data-set in South Florida and expanded the STP algorithmic family significantly by adding a context informed capability to add visit probabilities within the prism. However, by directly connecting the visit probability to the context itself and using the prism as a possible/not possible limiting factor for the maximum speed of the movement, it pays little consideration to the actual movement behaviour observed expressed by speed, turning angle distributions etc. As a result, the model is expected to work best on input data of fine temporal and spatial scales, whereas for input data being irregularly recorded, or displacements taking place over large geographic distances – that is, on a sphere – the task of moving from point to point remains challenging.

2.2.3 Brownian bridges movement model

While the STP provides no information regarding the probability density within the PPA, Brownian motion can be used to generate instances of trajectories whose densities may be used for gaining such information. Brownian bridges 'tie' such motions to two points, therefore enabling the derivation of the likelihood of the mover to be in a specific location (x,y) at a specific point of time (t). The underlying assumption is that the mover has selected the general location to be in and not a specific single point in space-time [116].A bivariate normal distribution represents

the chance to find the mover in any location surrounding a specific point, and a diffusion process reflects the relocation of the mover, as seen in Figure 2.3. The rate in which the mover updates its location is translated into the bandwidth size of a kernel; in the case of a Brownian Bridges Model of Movement (BBMM), into the variance of the Brownian motion (σ_m^2).

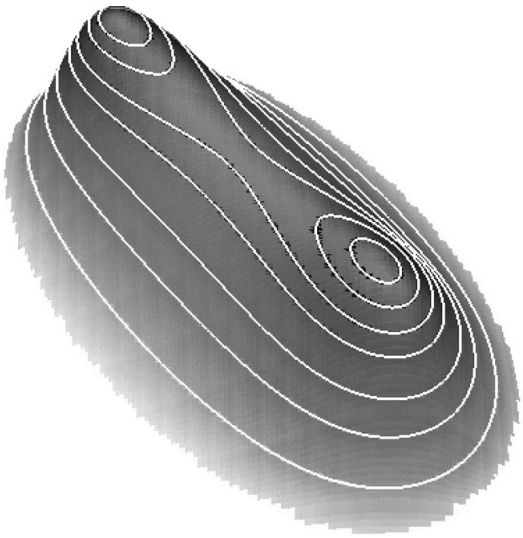

Figure 2.3: Probability density for the fraction of time spent in different locations constructed using the BBMM. The axes of the plane represent the location coordinates in 2-D space, while the elevation of the surface denotes the corresponding density. The two peaks in density correspond to the observed locations that are connected by the BBMM (source: Horne et al. [117]).

The BBMM quantifies the probability of being in a particular location when moving between a start (A) and an end point (B), taking into account the distance between the points and the overall time that it takes to traverse from A to B [117, 118]. Generating multiple Brownian motions tied to fixed endpoints gives a sense of how the mover uses its surrounding space by calculating the probability density of an animal found at the given point in space, also known as a Utilisation Distribution (UD). Ever since Brownian bridges were introduced in the ecological community, this method has been used extensively to define the home range of species [116, 119]. Later, Song and Miller [114] presented a possible combination of the BBMM and the STP, using STP as a truncating mechanism for the BBMM probability. In their case, the resulting UD excludes the areas that the mover cannot reach in a given time frame.

In the same direction but with more focus on biologically interpretable parameters, the Elliptical Time Density Model (ETDM) [120] was created to calculate of the UD of a mover while sampling the movement parameters from Weibull distributions adjusted to the tracking data. ETDM accounts for the mover's biological capacities, but it does not address the mover's biological preference.

In summary, the Brownian bridges family [117–119, 121–126] calculate a probability that is considered to represent the average animal's preference in space. Assuming that the study area has no hard boundaries for the mover – the BBMM does not handle zero likelihood in calculating the probability surface – the results can be quite realistic, or better: biologically explainable. However, as the work of Song and Miller [114] has shown, the basic BBMM has the limitation that non-zero visit probabilities may be generated in areas that cannot possibly be reached, and hence the BB probabilities should be truncated by the STP. Like the STP, however, the BBMM is restricted by the assumption of a constant moving speed, which is only rarely reflected in biological reality. Furthermore, the BBMM lacks the capacity for the user to generate realistic individual trajectories (i.e., paths that represent specific behaviours of an individual) out of the calculated probability surface, with varying speed and turning angle distributions. Additionally, generating Brownian bridges requires calculation for every pixel of the study area, making the process computationally expensive.

2.3 A conceptual framework for movement ecology

The plethora of movement modelling approaches, together with the different needs in the ecological domain, created a prosperous environment for *ad hoc* attempts to solve particular questions for many years. In 2008, Nathan et al. [55] attempted a top-down view by introducing the "emerging paradigm" of *movement ecology* that proposed unification of the formerly fragmented scientific approaches in ecology revolving around the study of movement and dispersal. Bringing together conceptual, methodological and empirical elements, the field of movement ecology can serve as an umbrella under which any research on movement or dispersal of living organisms can be placed.

Importantly, Nathan et al. propose a general conceptual framework that is both simple and intuitive and comprehensive. As shown in Figure 2.4, the conceptual framework incorporates three components associated with the mover (i.e., the focal individual) – its internal state, navigation capacity, and motion capacity – as well as a fourth component, that is, the external triggers exerted by the environment. The resulting movement path in this framework is the expression of all these components and their relationships and interactions, and in turn, feeds back to the internal and external components. The framework does not seek to introduce new elements at the level of individual components [127, 128], yet as a whole, it provides a user- and modelling-friendly view of the domain of movement ecology, it "*integrates eclectic research on movement into a structured paradigm*", and it "*aims at providing a*

basis for hypothesis generation and a vehicle facilitating the understanding of causes, mechanisms, and spatiotemporal patterns of movement and their role in various ecological and evolutionary processes" [55].

Nathan et al.'s conceptual framework can serve the integration of the existing movement models discussed above, linking these to the corresponding components of the framework (i.e. internal state, motion capacity, navigation capacity, external factors, and movement path). Hence, the framework allows integrating the basic methodological approaches and ideas from each movement model and thus the potential synergies among them. For example, Nathan et al. [55] suggested that the strength of their conceptual framework lies in addressing questions relative to the interaction of long-distance movements and given environmental conditions and identified the practical difficulties in quantifying and incorporating specific movements and underlying patterns in their framework.

However, none of the approaches reviewed above can represent the totality of the components and interactions included in Nathan et al.'s conceptual framework. Thus, it appears that any movement simulation solution that would seek to cover the entire conceptual framework comprehensively would need to employ and integrate multiple simulation paradigms representing different capacities to solve the overall problem.

Last but certainly not least, a simulation without the ability to incorporate the context of the movement would be unrealistic by definition, as illustrated by the importance of the *external factors* in the conceptual framework. The importance of incorporating contextual information has been highlighted in computational movement analysis has been highlighted by many authors [129–134], and several authors have incorporated contextual factors in simulation-based studies [61, 72, 92, 124, 126]. However, the simulation methods used have partly limited capacities of adapting to context.

2.4 Research gaps

In summary, the review of state of the art allows to identify of the following research gaps regarding the general problem statement laid out in Section 1.1.4:

- Random walk models have the advantage of generating individual paths of a mover that can be conditioned, if desired, by varying movement parameters (e.g., speed, turning angle). However, they cannot guarantee to reach a particular endpoint.

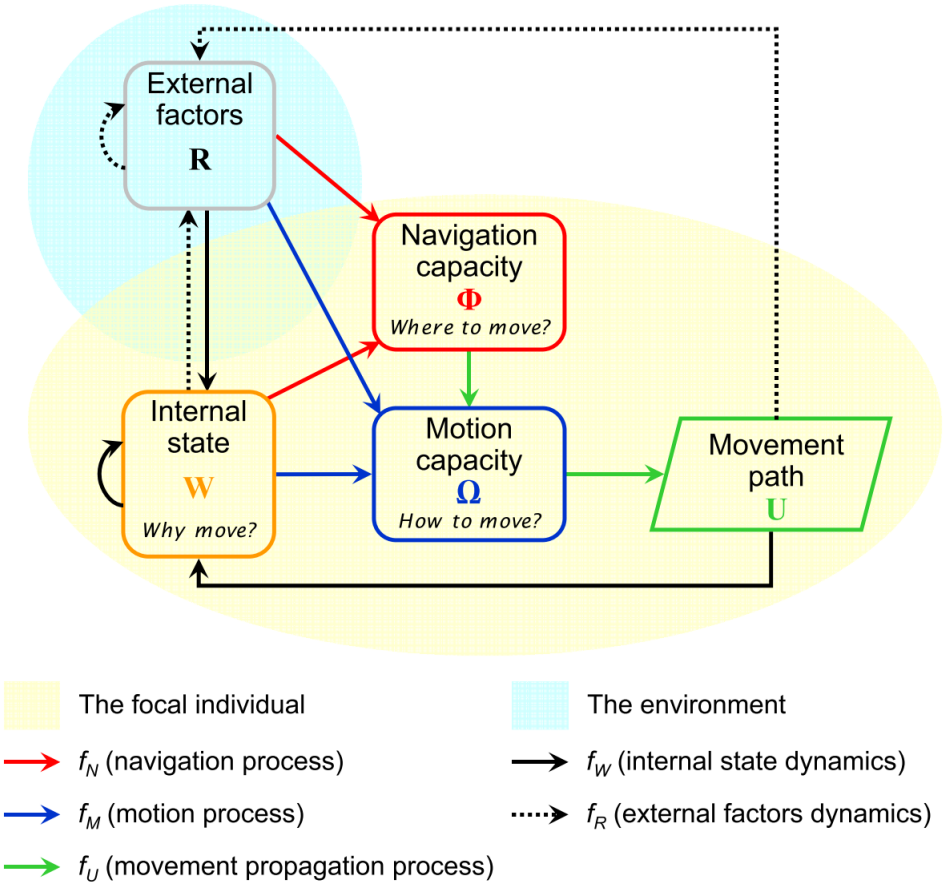

Figure 2.4: A general conceptual framework for movement ecology (source: [55]).

- The Space-Time Prism (STP) and the Brownian Bridges Model of Movement (BBMM) can solve the problem of reaching a given endpoint, but they suffer from other limitations. The basic STP approach lacks the capacity of generating a UD within the PPA. The basic BBMM fails to truncate the UD where visits are not possible given a certain speed and time budget, and no analytical solution exists generating visit probabilities for spherical coordinates. In addition, both the STP and the BBMM share the following limitations: lack of capability of generating individual trajectories with varying movement parameters and a high computational load.
- The consideration of external, contextual factors has partially been addressed by some movement simulation studies but needs further attention.
- No single simulation approach can cover the entirety of the components and relations represented in the conceptual framework of movement ecology proposed by Nathan et al. [55], necessitating the combination of different approaches in an integrated solution.

2.5 Specific research objectives

To conclude of this chapter, I want to revisit the overarching research question driving this work and define the research problem in more detail. As a reminder, the **main research question** stated in Section 1.1.4 is:

How can we best model movement and context in a simulation framework to better assist in ecological knowledge acquisition?

In the same Section 1.1.4, I also defined the general research problem to be addressed here. Following the above review of state of the art in relevant research areas, we are now in a position to state our research problem in more detail, leading to the following **specific research objectives**:

1. Develop an integrated simulation approach that allows covering all components of the conceptual framework of Nathan et al. [55], and their interactions.
2. Develop a simulation approach that allows connecting two known locations, possibly separated by a large spatial and temporal distance, by generating ecologically possible trajectories that comply with the mover's physical limitations and carry the least statistical bias possible.
3. Advance from "ecologically possible trajectories" to ecologically realistic trajectories: Develop statistical methods to empirically add realistic behaviour in the set of trajectories that connect two locations.
4. Embed context to refine the empirically informed and modelled movement further.

The following four chapters contain the main contributions of this thesis. Chapter 3 will address Research Objective 1 (RO1) by proposing a conceptual framework for movement simulation. In Chapter 4, an algorithm called 'Random Trajectory Generator' (RTG) will be presented, addressing RO2. Chapter 5 then revisits the RTG algorithm and extends it into an 'Empirical Random Trajectory Generator' (eRTG), which solves RO3 and meets RO4. Finally, Chapter 6 presents an extended case study that demonstrates how RO4 can be addressed.

3

A conceptual framework for movement simulation

> *The first step is to establish that something is possible then probability will occur.*
>
> **— Elon Musk**
> (engineer, technology entrepreneur, inventor)

While the key objective of this thesis is based on developing a simulation approach that allows to densify coarsely sampled movement tracking data or fill in gaps in incomplete, fragmentary movement data, I would like to start off this chapter[1] by first laying out a conceptual framework for movement simulation that comprehensively covers all components of the Movement Ecology Framework (MEF) by Nathan et al. [55], therefore addressing Research Objective 1 of Section 2.5. Looking first at the 'big picture' will enable us to identify the possible approaches that could be used to implement Nathan et al.'s conceptual model, and it will also allow placing the key contributions of this thesis into perspective.

The importance of reflecting on how the MEF could be implemented in movement ecology research is further stressed by Joo et al. [136], who recently conducted an in-depth literature review on the progress made in movement ecology research over the past decade. The authors clearly state that despite the advances in technology, methods, and increase in publications, research practice has not been exploiting the MEF to its full potential, by utilising only a few of its main components. Joo et al. [136] express this utilisation imbalance by visualising the movement ecology articles' focus topics over the past decade (Figure 3.1).

Against this background, this chapter discusses the role of various types of simulation paradigms when a movement model is applied within MEF. The conceptual simulation framework presented in this chapter aims to assist the research experimentation process by regulating the initiation, execution and evaluation of the statistical models used for simulation [137]. After setting the initial conditions based on the study's hypothesis, running multiple instances of the simulation model and observing its

[1] This chapter is based in large parts on Technitis and Weibel 2012 [135].

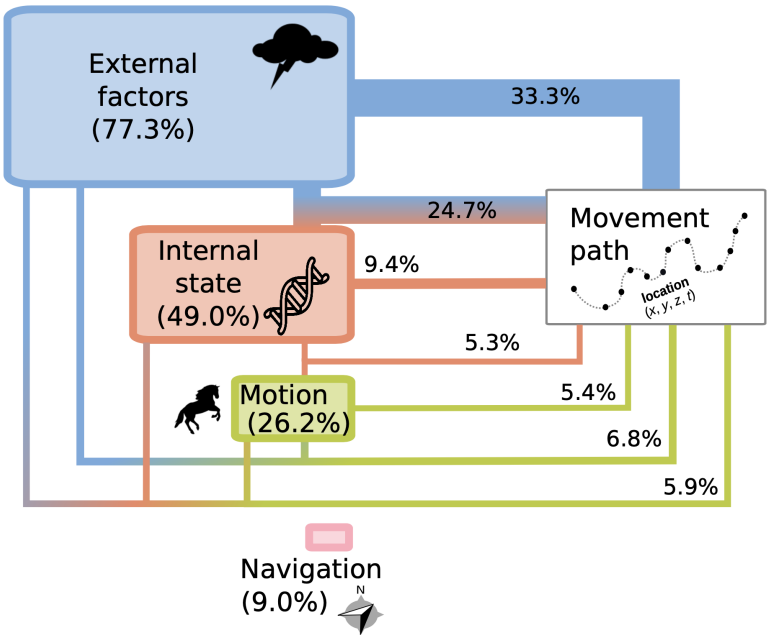

Figure 3.1: Illustration taken from the review by Joo et al. 2020 [136]: *"Representation of the components of the movement ecology framework and how much they were studied in the last decade: external factors, internal state, motion and navigation capacities, whose interactions result in the observed movement path. The size of each component box is proportional to the percentage of papers (in parentheses) tackling them irrespectful [sic!] of whether they are only about this component or in combination with another one. The latter is specified through the segments that join the components to the observed movement path. One fill color corresponds to papers that only studied one component, while two or more colors correspond to papers that tackled two or three components, respectively (the ones from those colors). The width of the segment is proportional to the percentage of papers that studied that combination (or single component). Only combinations corresponding to > 5 % of papers are shown; e.g. combinations involving navigation and papers studying navigation on its own had < 5 % of papers each therefore they are not shown in the graph."*.

performance helps the researcher reduce uncertainty while concluding on how the input selection affects the output variation [138].

In the remainder of this chapter, Section 3.1 gives an overview of three main simulation paradigms and assesses their fitness for movement simulation. Section 3.2 then proposes a logical model of movement that implements the framework of movement ecology by Nathan et al. [55] for simulation. In Section 3.3, four example scenarios serve to showcase the potential usages of each simulation paradigm. In Section 3.4, a combination of the three simulation paradigms is proposed accommodating the logical movement model introduced in Section 3.2. The chapter ends with some concluding remarks in Section 3.5.

3.1 Movement simulation paradigms

3.1.1 Foundations

The attempts to model the plethora of types, scales and complexities of natural movement have led to developing a wide range of simulation methodologies, each of which fits, explains, or relates to the phenomenon under study. Approaches that focus on individual [49,139,140] or collective [15,15,141] movement, dynamic [142–144] or state-driven simulations [145–147] have been separately implemented for many years [148–151]. Nevertheless, the necessity of representing the interconnectivity of natural systems suggests further exploration of the potential synergies between the available approaches.

The need for flexibility in the available simulation options has been raised in several theories, such as chaos theory [152, 153], network theory [154, 155], cybernetics [156], to name but a few. Perhaps the two most compelling integration propositions [157] were made through the General Systems Theory (GST) [158] and the Complexity Theory (CT) [159]. These two theories attempt to create theoretical frameworks that deal with the non-linearity of the various processes incorporated, allowing space for surprising behaviour emerging from the interaction of the system components - or in Aristotle's words, "*the whole is greater than the sum of its parts*" (Aristotle, 350 BCE) [157]. Both theories create an element of hierarchy in their rule systems, and they focus on seeking the principles (behaviours) seen in more than one system [158], that is, principles transferable to multiple domains or sciences. What sets them apart is the perspective under which they define the system itself. GST sets a goal, defines the structure in a top-down manner, monitors the performance, and finally evaluates and connects all the components with the respective feedback loops. In the case of CT, the system emerges from 'bottom-up' interactions of its components.

Three of the most representative simulation paradigms in the two theories are System Dynamics (SD), Agent-based Modelling (ABM), and Discrete Event Simulation (DES). SD is an equation-based approach representing the exchange of matter, energy or information ('flow') between interconnected, aggregated system compartments ('stocks') over time. Complexity theorists, in contrast, have proposed ABM as one of the most important methods for studying multi-individual, complex, (potentially) living systems. Completing the picture, DES has been used to describe the change of state and activity of a system [160], commonly used for simulating changing contextual factors. The remainder of this section reviews the three simulation paradigms as well as their strengths and weaknesses.

3.1.2 Systems dynamics (SD)

The SD setup assists the modeller in gaining a better understanding of the balance among the components of a system and the feedback between them [161], usually expressed as ordinary differential equations [148]. The basic concept of the model can be described by a simple stock and flow diagram (Figure 3.2). The flow affects the stock (in this case, a population) based on an equation, and the stock gives feedback on the flow. The feedback relationships can be either positive or negative. In Figure 3.2 the relationship between the birth rate and population of a species is positive: the more births, the larger the population becomes, and the larger the population becomes, the more births will occur. The feedback relation between deaths and population is negative: the more deaths, the smaller the population.

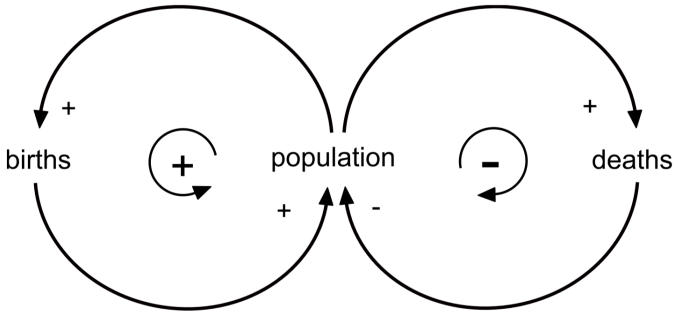

Figure 3.2: A system dynamic feedback loop example.

The key advantage of SD is that it captures the feedback and delay processes inherent to a system to provide the user with the system's behaviour over time [162]. The user may interactively change the balance of these processes and monitor the effect that this change has on the whole system. SD modelling can be very suitable when expressing a closed-loop system where the researcher can fully articulate the relation between two variables. In complex systems, however, a feedback loop is not always so easy to identify, so the adequacy of the simulation may become questionable. SD is best for predicting the evolution of a system qualitatively (e.g. growth vs reduction) rather than for making numerical predictions. Moreover, SD models lack the capacity of systems to modify themselves structurally [161]. Hence, no form of adaptation can be simulated through this paradigm. Last but not least, SD approaches have no spatial awareness. To the best of our knowledge, no SD implementation has incorporated spatial attributes into the model.

3.1.3 Agent-based modelling (ABM)

An agent is defined as "*a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future*" [163]. ABM follows a rule-based approach

representing a modular system as a network of locally connected, intelligent, and often adaptive individual agents. The approach is widely used in movement simulation [164–166] and coupled with GIS and geostatistics, which makes it a powerful tool that can express the moving part of the simulation and the complex dynamics of ecological and social systems [165, 167, 168]. Compared to an equation-based simulation, ABM offers increased flexibility in capturing the variance of behaviours amongst individuals or groups of individuals [169], making it ideal for exploring the social dynamics. Implementations of ABM vary significantly, making a unique definition of an agent challenging [170], but on a general level, an agent is an algorithmically definable entity (Figure 3.3) with the following parameters [171]:

- Internal condition: the sum of all the attributes and capacities describing the agent
- Interaction with external conditions: the type of interaction with the external factors, including reactive, proactive and passive modes
- Interaction with other agents: various social behaviours, such as cooperation, competition and aggression.

The modularity of the approach and the ability to tailor behaviours individually allow for great flexibility, expressiveness and parallel execution of models [169]. This flexibility is based on three main characteristics: *autonomy*, *heterogeneity* and *activity*. *Autonomy* is ensured when the control of the simulation is kept decentralised. An agent is 'free' to interact with other agents and its context without being manipulated directly by the user [153, 172]. *Heterogeneity* is created when agents might have similar or different collective or non-collective behaviour; they are created bottom-up, depending on the characteristics of each agent [173, 174]. *Activity* refers to the fact that the agents are usually active in a simulation. They can be, for example, proactive, reactive, interactive, rationally oriented, mobile or adaptive: that is, agents can follow a predefined logic, act together in a multi-agent system, perceive their context or not, and even have some memory and learning capacity in their behaviour [156, 175].

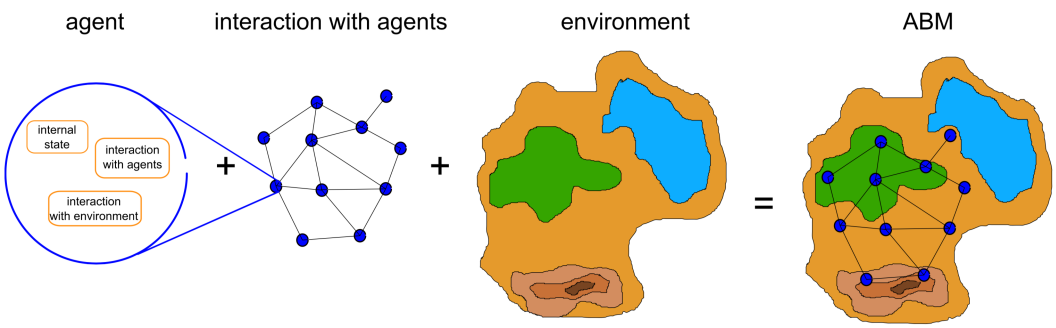

Figure 3.3: An agent-based modelling example.

ABM is rather suitable for modelling complex systems around individual behaviours and their interactions. It has excellent micro-specification capabilities, and each agent or cluster of agents can be independent, so it has autonomous decision-making functionality and obeys tailored behavioural rules. ABM deals with scale, heterogeneity and spatio-temporal discretisation intuitively, encouraging events to emerge unexpectedly, thus giving space for adaptive behaviour and evolution of the overall system. ABM offers an intuitive virtual system to represent the complex real-life system and can be governed by assigning many but straightforward rules to each agent. Each agent can be fully aware of its spatial environment but adapting its behaviour based on social feedback proves more testing as the paradigm focuses on spatial factors [176]. An additional challenge is the validation of this approach, as depending on the freedom allowed and the individuality of each agent, any outcome could be justified. Last but not least, in practical implementations, the distributed parameterisation of the overall model can make the whole system computationally heavy to run for long simulation periods.

3.1.4 Discrete event simulation (DES)

In DES, the 'event' is considered to be the main focus of the modeller. Entities, activities and processes are also defined to assist in working with event data, . An entity has attributes and is capable of changing its state [177–179]. An event is everything that somehow changes the state of an entity. The activity describes the things that happen to an entity for a fixed period. A process is a list of events, activities and delays that define an entity in its life cycle. In some instances, the process replaces the event as the centre focus of the model [171].

The basic components and the flow of the simulation can be described by block charts, queues, delays or resource sharing [180]. The logic behind the operation of DES follows the familiar convention of the standard queuing model [181]: one event takes place only after the previous one is finished, typically in a First-In-First-Out (FIFO) order, like clients in a bank waiting in a queue to be serviced by the cashier. The logical flowchart in the traditional DES approach consists of a discrete chronological sequence of events, which in combination express an activity (e.g., in Figure 3.4, Events 2 and 4 define Activity A). Instants in time when a defined state of a variable, i.e. an event, changes [182]. The inherently ordered structure of DES makes it an appropriate simulation paradigm for representing a system on an operational level [162].

DES is commonly selected to represent more passive entities [177, 183], especially on a macro level. Each entity undergoes only a few and potentially small changes, governed via centralised strict behavioural rules. The latter makes DES a sub-

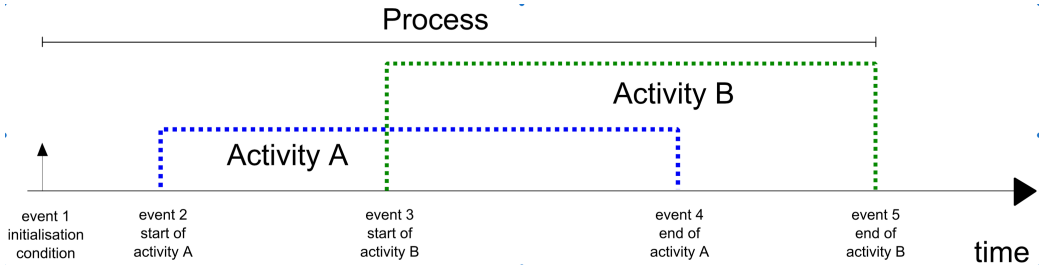

Figure 3.4: A simple discrete event simulation workflow.

optimal approach for cases where individual behaviour is a priority. The rigidity of the DES approach makes it remarkably optimised, but handicaps it in situations where behaviours that have not been observed in the training data needs to emerge out of the system.

3.2 A logical model of movement

A simulation framework that maps all the necessary entities and their interactions of a phenomenon (from here on *the system*) helps the user to select the most suitable simulation methodology, efficiently code models to test hypotheses and, based on their performance, move towards inference and eventually to knowledge generation [184]. With this in mind, I am proposing a simulation framework that aims to translate the established general conceptual framework for movement ecology by Nathan et al. [55] to a *logical model of movement* to identify all of the system's components. In Section 3.4, this logical model will then be subdivided into and expressed with appropriate simulation paradigms.

3.2.1 Overview

Laying out a conceptual architecture of the *system* imposes a certain level of formalisation on the modelling process [78]. The architecture bases the modelling on a modular design setup instead of a reactive, heuristic model building practice. The former gives better control over the testing and complex modelling, while the latter often results in a maze of single runs and isolated results.

With Nathan et al.'s framework of movement ecology shown in Figure 2.4 in mind, the basic components of the proposed logical model of movement are organised into a *static part* and a *dynamic part*, as seen in the overview of Figure 3.5. Perhaps counter-intuitive to its naming, the static part of the model does change throughout the simulation run, but the change is never self-triggered. It receives a status update only following a system change initiated by the dynamic part.

Once the model starts running, the static part is activated, encompassing three main steps: identifying the initial conditions, identifying the possible movements, and selecting the probable movement. The dynamic part then takes over, updating the simulation's location, time, and conditions, to pass the status information back to the static part of the simulation, whose values get updated and continue the processing cycle.

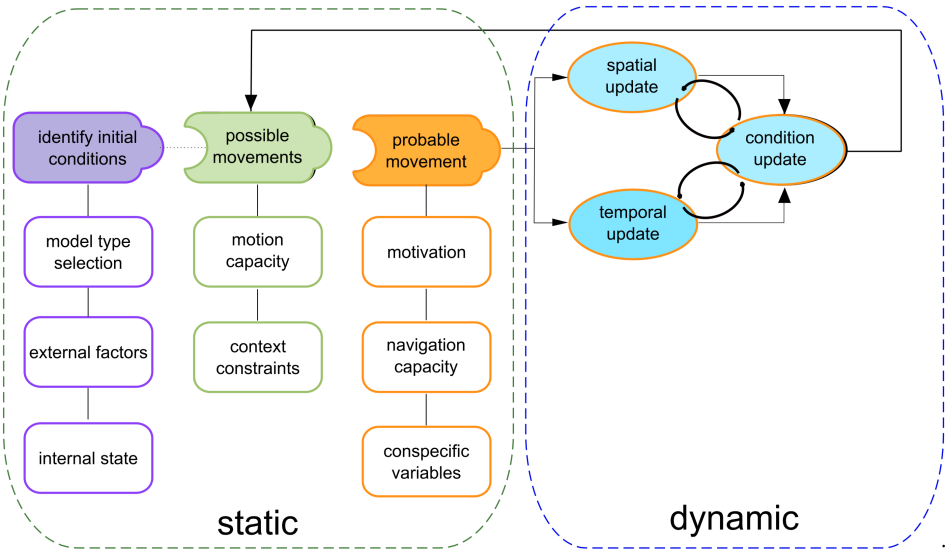

Figure 3.5: Logical model of movement.

3.2.2 Dynamic part

The dynamic part is responsible for keeping track and updating the dynamic variables that change over time as a function of movement, including the spatial position (i.e., spatial update in Figure 3.5), the time (temporal update), and the conditions of movement, such as the internal state of the focal individual or the external factors defining the environment (condition update). Hence, selecting an appropriate time model is key to the dynamic part of the movement modelling. Time, in this case, can be either modelled in a continuously or discretely. When the continuous mode of time is selected, time advances steadily in a fixed increment from the beginning of the simulation until the user-defined ending condition is reached, updating in every time step both the spatial location of the focal individual and the conditions of movement. For example, the weather may have changed (external factor); the energy level of the individual may have dropped below a threshold (internal state), impacting the motion capacity; or the terrestrial individual may have reached a lakeshore (external factor: context constraint). In the discrete-time model, time advances only when the state of a condition or the position of the moving object changes, as in the examples of the preceding sentence. The time will be updated whenever an event such as a specific condition or position change takes place. Thus, the discrete-time model may also be termed the event-driven time model.

Deciding which model of time fits best for a given case is crucial, as it affects the remainder of the modelling flow, in particular (a) the efficiency, (b) the structure of the simulation, and (c) the result, and thus should be adapted to the requirements of the given simulation scenario. Simple examples of effects caused by particular time model decisions are when (a) the user attempts to model one year of movement with spatial updates every millisecond: the volume of the data could then get overwhelming without necessarily being scientifically relevant; or (b) the training data has irregular time intervals: simulating in continuous flow would mean that all the records need to be temporally standardised, and therefore include a spatio-temporal interpolation step early in the simulation process.

3.2.3 Static part

The static part consists of three main steps (Figure 3.5), described below.

Identify initial conditions: This step gives a starting value to all the active parameters of the simulation model. First and foremost, the user selects the model of time (discrete vs continuous) and the spatial movement model (e.g., random walk, Levy flight, etc.) that the simulation will use. After selecting the model, the user may choose which external factors (e.g., temperature, wind speed) and which components of the internal state (e.g., instincts, drive, and reflexes) apply in the specific case.

Possible movements: In this step (green tile in Figure 3.5), the system identifies all feasible movements for the focal individual. Factors that shape the set of possible movements are the motion capacity (e.g., maximum speed and acceleration, turning angle, possible duration at maximum speed) and the context constraints (barriers such as land for (most) fish or wide rivers for gibbons). The possible movements are governed by the mechanistic attributes of the focal individual, that is, the physical part of the movement, as will be discussed in Chapter 4. The capacities and limitations to the movement of a particular individual or species can be defined by the user or automatically extracted from a reference dataset using data mining algorithms [185].

Probable movement selection: Selecting a probable movement (orange tile in Figure 3.5) encompasses the computation of the probability of all the possible movements to select the one with the highest probability. This step is further dealt with in Chapter 5, and it is naturally expressed based on how the individual will likely decide to target and direct its movement.

The calculation of the probable movement requires bringing all the categorical and continuous variables describing the properties of movement to the same scale of

measurement. Such variables describe the navigation capacity of an animal, its motivation, and additional conspecific characteristics that affect an individual's decision to move. The navigation capacity is the individual's capability to perceive its environment; in other words, it requires integrating all the context data that assist an animal to navigate. On the other hand, motivation is usually goal-oriented and a crucial driver of movement activity, associated with a drive such as hunger, thirst or sleep, while remaining closely tied to sensory stimuli. For instance, once the food is available to an animal, it expresses eating behaviour, limiting its mobility. Motivation may also be learned, known as secondary motivation [186]. Various additional conspecific variables might be necessary to better depict an animal's actions, such as inbreeding behaviour or alarm signals.

3.3 Example scenarios

In this section, four example scenarios of increasing complexity are used to illustrate possible usages of the three core simulation paradigms, along with their strengths and weaknesses.

3.3.1 Scenario 1: Single behaviour, fixed move until an end condition

In the first scenario, the focal individual is represented by a non-intelligent agent, which moves based on a simple random walk. Its behaviour is ; for instance, it might constantly be foraging. The context is defined by a single resource, e.g., food. For a more intuitive example, take a short-sighted, memory-less bird looking for food. It starts with fixed energy storage that lasts, for instance, for ten random moves in a given time interval. In each time step, it moves to a new location and checks for resources. If none are present, it moves to the following location. Once it finds food resources, it spends one-time step not moving and gains energy for the next two steps. Ultimately, if no resources are found and the energy storage reduces to 0, the bird dies.

The bird's movement and behaviour, in this example, is sequential and lacks complexity. No social dynamics of any sort have to be taken into consideration since the bird is not interacting with other birds. Thus, the DES paradigm can offer a good and straightforward approach for this simulation scenario.

3.3.2 Scenario 2: Rule-based behaviour selection, variable move

In the second scenario, the bird has two modes to move: either slow (1 move per time step) or fast (3 moves per time step). The bird here has a rule embedded that controls the change of its behaviour. For instance, once the energy level rises over a specific level, the bird will move fast instead of making a single move. It has two different options for its behaviour, e.g., foraging and exploring. The available context remains the food.

The complexity of this model is higher than in the previous case. However, it is still feasible to approach it using DES, and ABM could also be used. Depending on the scale, the required accuracy and the number of individuals to be simulated, these approaches could be used alternatively and complementarily [148]. Alternative usage will be warranted if the researcher feels more confident in expressing the simulation in one of the two paradigms. Complementary use is advisable in order to optimise the available computational resources, given that the performance cost of ABM increases geometrically as the simulated individuals increase.

3.3.3 Scenario 3: Adaptive behaviour, passive context

In the third scenario, the bird has an adaptive character, meaning that given a specific stimulus, it will act accordingly. For instance, if there is a high average concentration of food in six adjacent neighbouring cells (assuming a raster representation of spatial context), then it will keep its movement low for the next move – regardless of having enough energy stored to move fast – in order to better exploit the available food resources.

In this example, the behaviour becomes reactive. the reaction is challenging for the DES paradigm, so an agent-based approach seems more appealing despite the added computing cost. The passive context, on the other hand, may still be expressed as a DES.

3.3.4 Scenario 4: Adaptive and collective behaviour, active context, trigger-based actions

At this point, the simulation model takes a rather complex form, aiming to be more realistic. The moving focal individual expresses both adaptive and collective behaviour. In the bird's example, it may forage, chase prey, follow its mother, flee from a predator or stay in a nest during nighttime. At the same time, the context also

becomes more complex and may include the presence or absence of resources, the risk of predation, atmospheric factors, and others. More specifically, the parameters can take the following forms:

- movement: may be described by moving slowly (1 move per time step); fast (3 moves in a straight line); or fleeing from a predator (3 moves in a zigzag shape);
- behaviour: the bird might be resting, foraging, migrating, etc.;
- context: can consider multiple choices of food with different levels of energy density; external factors (e.g. vegetation, temperature, light intensity); the presence of conspecifics, prey, or predators; and the extent of the home range.

In order to accurately model the behaviours, some new parameters have to be defined in this case. Based on the logical model described in the previous section 2.3, the bird has a dynamic internal state (energy level, age, physical status), motion capacity, and navigation capacity [55] complemented by its motivation. From a qualitative point of view, these parameters are not independent. For instance, once the internal state of the animal records a low energy level, then its motivation for foraging increases. If the animal is threatened (a possible internal state), the agent is likely to behave collectively, and so on. Another relationship concerns the navigation capacity and age: juveniles and older adults have lower navigation capacity, while adults possess the maximum capacity. Finally, a more complex relationship concerns resources and mobility: the more resources available, the less mobility the bird will exhibit.

Taking a detailed look at this scenario, it seems that the birds can be adequately simulated with ABM, while DES would suitably assist in simulating certain external factors. Specific rules — for example, low level of energy, fear and others — activate the corresponding behaviour. The age-dependent evolution of navigation capacity can be simulated straightforwardly, e.g., by using a bell curve equation: until a certain age, the bird has better navigation capacity per year; after that age, the capacity levels off or diminishes.

For modelling the last relationship between the resources and the bird's mobility, let us assume a foraging bird and a fruit tree. The bird is rather mobile while searching for resources up until the time it finds the fruit tree. Once it satisfies its hunger, the bird's mobility drops. In other words, the relationship between a given external factor and its effect on the bird might be bidirectional. Further extending this line of thought, if too many birds are present, then at some point, the behaviour of the next simulated bird that is added should revert to increased mobility, as the available food resources cannot support it. This is a typical form of causality loop seen in

SD that, even though it is computed on the level of each individual, will eventually yield results on a macro level, described by aggregation statistics and population dynamics.

3.4 Exploiting synergies: a hybrid simulation model

Of the four reference scenarios discussed above and summarised in Table 3.1, the first and second scenarios are sufficiently constrained so they can be implemented using a single simulation paradigm (DES or ABM). However, any real-life case would likely be more complex. The third reference scenario, then, takes adaptive (reactive) behaviour of the bird into account and suggests using a combination of simulation paradigms: ABM to implement the reactive behaviour of the bird by a set of rules, and DES for the (possibly changing) passive environment.

Scenario #	**Movement**	**Behaviour**	**Context**
Scenario 1	Steady	Steady	Single resource
Scenario 2	Boolean	Boolean	Single resource
Scenario 3	Adaptive	Adaptive	Single resource expressed by equation
Scenario 4	Adaptive & Collective	Adaptive & Collective	Context

Table 3.1: Summary of the reference scenarios.

The fourth case comes closest to what might be perceived as a 'realistic' setting for simulation in movement ecology. Given the significant complexity of this reference case, a hybrid simulation model that combines all three paradigms (ABM, DES, SD) and that can be overlaid on the logical model of movement introduced in Section 3.2 is proposed, shown in Figure 3.6. In this hybrid approach, ABM and DES simulation models may work in a complementary fashion: when the former becomes too costly, the simulation turns to DES, which results in less depth and quality in the calculations of each individual, but much better overall computational performance. SD may help express the interactions taking place among the various elements of the simulation model at a finer-grained level.

The hybrid approach has the advantage of creating sub-models that run pseudo-synchronously for each time step, computing the effects of the various simulation components on each other. For instance, the more predators are present, the less the bird feels like foraging; the closer the bird is to starvation, the less it cares for the presence of a predator; the more exposed an environment, the less the

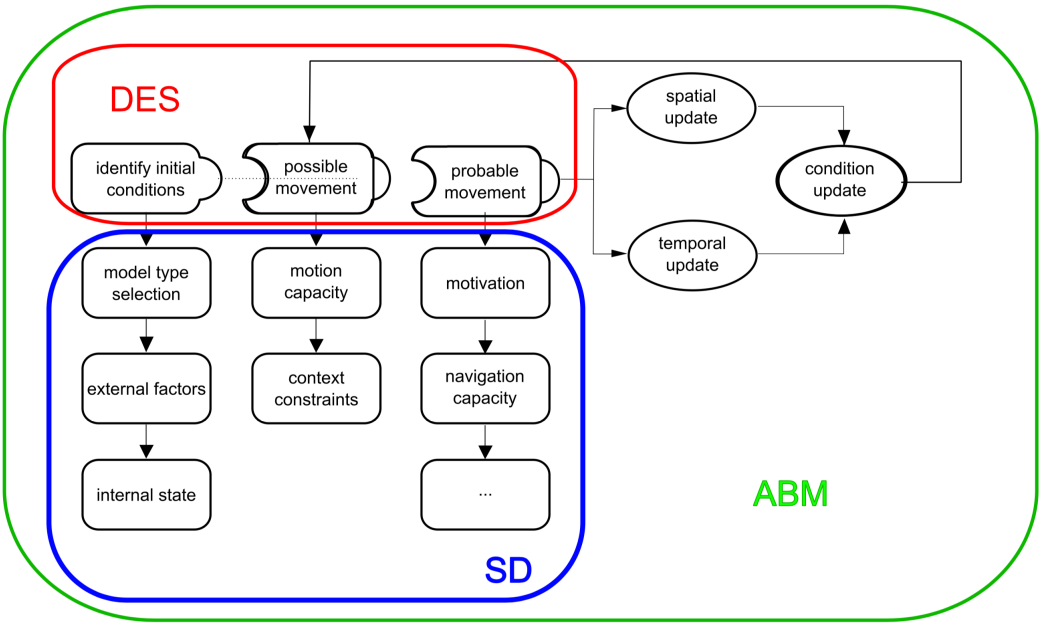

Figure 3.6: Exploiting synergies of the three simulation paradigms: The proposed hybrid movement simulation model.

motivation to forage. This rule-based reasoning can be expressed in the form of causality loops incorporated into an SD model. By conducting a sensitivity analysis on the simulation outcome one can understand how the model responds to given changes of the parameters, and make adjustments on the feedback loops to control the simulation experiment.

This trial-and-error method for adjustment may, on the one hand, reveal important properties of the bird's movement and, on the other hand, assist in quantifying the contribution of each of these properties to the movement itself. Ultimately from an ecological perspective, this approach can help with the testing of ecological hypotheses. In contrast, from the perspective of methods development in GIScience, simulation helps create realistic test data that may be used to evaluate algorithms for movement data mining and analysis.

Realistically simulating movement for ecology also involves multiple spatial as well as temporal scales [55]. The necessity of accounting for scale becomes visible in the fourth reference scenario. A full-scale simulation of movement involves decisions and behavioural expressions that extend beyond the level of the individual, often reaching the level of the entire population. An example scenario is the foraging behaviour that is decided individually, whereas the change of the nesting area or the time to initiate migration is decided collectively. Being able to handle cases of collective behaviour, such as leader-follower behaviour, while at the same time allowing space for expressing individual preferences allows this framework to handle multiple real-life case studies.

3.5 Concluding remarks

In this chapter, a hybrid movement simulation model was introduced that utilises the combination of the benefits offered separately by Agent-Based Modelling (ABM), Discrete Event Simulations (DES) and System Dynamics (SD). This model covers the points above and extends the state of the art in simulation for movement ecology in the described scenarios. Typical movement simulators use a single movement model (e.g. CRW, Levy flight, Brownian motion) at a single spatial and temporal scale and a single simulation paradigm. However, as a growing number of hybrid simulation approaches is being published in multiple domains [172, 187–189], the trend towards a fused, multipurpose and scalable simulation is becoming more prominent. This trend can affect the simulation in movement ecology, and with the proposed hybrid movement simulation model, I hope to provide the conceptual basis for exploring some of these effects.

The proposed hybrid simulation model represents the framework of movement ecology by Nathan et al. [55] comprehensively. However, it does so only in a conceptual way. Nevertheless, apart from providing general guidance for best utilising relevant simulation paradigms to simulate movement, the model also addresses one core issue of movement simulation in detail: It separates the prediction of the next move into two steps, namely the *possible* and the *probable* movement. This separation has two important advantages: First, it renders the simulation process more modular and thus more flexible; second, it allows incorporating motivation in the prediction process.

The following chapters exploit this feature of the proposed conceptual simulation model. For the main research problem of this thesis laid out in Section 1.1.4, the following Chapter 4 addresses the sub-problem of generating ecologically *possible* trajectories that comply with the physical limitations of the mover and carry the least statistical bias possible. This relates to Research Objective 2, as stated in Section 2.5. Chapter 5, then, extends this to solve the sub-problem of generating ecologically realistic trajectories by adding *probabilities*, which relates to Research Objective 3, and by adding spatial context, relating to Research Objective 4.

4

Starting simple: Physically unconstrained, possible movement between two points

> *Simplicity is an exact medium between too little and too much.*
>
> **— Joshua Reynolds**
> (artist)

This chapter introduces the 'Random Trajectory Generator' (RTG), an algorithm that allows creating a physically unconstrained movement trajectory between two given endpoints. Such movement occurs in a homogeneous context, with the only limiting factors being the (constant) speed that the mover can achieve and the time available to complete the move. The chapter starts with the motivation behind the RTG algorithm (Section 4.1) and the challenges that the algorithm has to address (Section 4.2). The core part of the chapter is devoted to the description of the methodology of the RTG algorithm (Section 4.3) followed by an evaluation and discussion of the results generated by the algorithm (Section 4.4). [2]. [190]

4.1 Motivation

The challenge every movement ecologist faces when data gaps occur (e.g., due to GPS signal loss) or simply when the temporal sampling interval is very coarse (e.g., to save battery life), is to decide how to connect the previous with the next available point. Solutions such as shortest distance, curve interpolation or least-cost-path algorithms only marginally solve the problem since the result is biased by the geometric model employed. To assess how biased the result of a particular interpolation method is, one can measure the footprint of the density of a large number of simulated trajectories. The less random the trajectories are, the smaller the footprint. Deterministic methods such as shortest distance and curve interpolation will result in a single line, while

[2] This chapter is based in large parts on Technitis et al. (2015) "From A to B, randomly: a point-to-point random trajectory generator for animal movement," *International Journal of Geographical Information Science*, vol 29, no. 6, pp. 912-934

the cost-path approach has the possibility of relaxing the criterion of least cost but still generates deterministic and thus biased results. Most likely, however, animals will not move in a completely deterministic and predictable way. So, is there a way to generate trajectories that omit this bias but still respect the endpoint constraints? Moreover, to what degree is this possible?

4.2 Research challenges

The review of movement models in Section 2.2 allowed drawing a number of conclusions regarding the fitness of different models for the problem outlined in the preceding section. The Correlated Random Walk (CRW) may be useful from the point of view of ecological modelling, but suffers from a lack of computational efficiency when the destination is a point rather than an area of attraction (i.e. when the hit-to-miss ratio is very poor).

The result of the Brownian Bridges Model of Movement (BBMM) model is a probability surface that may be regarded as representative of reality in terms of the space utilisation of a moving object. However, once an individual trajectory is extracted from a BBMM probability surface, using, for instance, conditional Brownian walk, the result will be a highly unrealistic mode of movement that is biologically unjustifiable for animals. Furthermore, generating a BBMM is computationally expensive and, thus, in practice, not feasible for computing large numbers of trajectories.

The Space-Time Prism (STP) offers an interesting possibility for generating an envelope of accessibility in space and time, that is, the Potential Path Space (PPS). With the work by Song and Miller [114], an approach has been presented for truncating the occupancy probabilities of the BBMM to the more realistic domain of the PPS (and PPA in 2-D space). However, this approach still inherits some of the weaknesses of the BBMM: it lacks the ability to generate individual trajectories with varying movement parameters, it requires a high computational load, and it offers no analytical solution for spherical coordinates.

In order to address these challenges, an algorithm is proposed that:

- is based on a CRW model of movement;
- offers significantly improved efficiency over the CRW model by tying the generated walk to both the origin and the destination;
- provides results as single or multiple trajectories;
- provides space utilisation maps similar to those generated by the BBMM; and
- limits the creation of trajectories to the PPA defined by the space-time prism.

4.3 Methodology

4.3.1 Underlying modelling framework

The conceptual model of movement simulation underlying the proposed algorithm consists of three main steps [135]:

1. initialise the conditions of the simulation;
2. identify the possible area for the next move; and
3. determine the object's next movement with the highest probability based on a specified level of confidence.

In the first step, the conditions defining the simulation are specified. In this case, that implies defining the coordinates of the origin A and the destination B of the movement; the available time budget to move from A to B; the temporal sampling interval used to generate fixes along the trajectory; and the movement object's maximum allowed speed. Step 2 represents the focus of this paper. The proposed algorithm generates realisations of trajectories within the possible area of movement, equivalent to the PPA of the Space-Time Prism (STP) approach, using the control parameters of Step 1. Step 3 is only implicitly dealt with in this chapter. Generating a sufficient number of trajectories approximates an empirical occupancy probability distribution.

For the simulation, the following conditions hold: The moving object performs a CRW at equal time intervals per step and with a fixed maximum speed, in a continuous and homogeneous space (2-D Euclidean space and spherical geometry, respectively).

4.3.2 The algorithm

Computing the Potential point Area (PpA). The trajectory to be generated will consist of a number of intermediate points separated by the temporal sampling interval Δt. To create each of them, we start by constructing the Potential point Area, or short PpA; the lower-case middle 'p' in the acronym is used to differentiate from the space-time prism's term Potential Path Area (PPA). The PpA then denotes the area in the plane where an intermediate point can be possibly placed. The following example describes this step.

Example 1. We assume that the distance *AB* between the origin *A* and destination *B* of the movement is 300 km, as seen in Figure 4.1. Point *A* is the origin of the

moving object's departure at $t_0 = 0$, and point B is the destination of the object's arrival at $t_f = 5$ hours later. The task, then, is to create a trajectory from A to B, given a maximum speed of the moving object $V_{max} = 70$ km/h, within the time budget $t_{total} = 5$ hours, and given the sampling interval $\Delta t = 1$ h.

If the moving object starts at A and travels for one hour as the sampling interval Δt commands, the area where the object would be expected to be is equivalent to a circle of radius r_A, where r_A is the maximum distance that can be covered in Δt given V_{max}, in this case, $r_A = 70$ km. At the same time, B needs to be reached at $t_f = 5$ h. Hence, one hour before $t_4 = 4$ h, the moving object ought to be within its maximum speed radius from B in order to be able to reach the destination in the final hour remaining. Likewise, 4 hours before it reaches point B, the moving object should be within a distance that can be covered within these 4 hours; that is, within a circle around B of radius

$$r_B = 4 * V_{max} = 280km. \tag{4.1}$$

Consequently, the PpA, one hour after the beginning of the movement, is defined by the intersection of the two circles from A and B, respectively (Figure 4.1). Next, we will explain how random placement of an intermediate point location within the PpA is incorporated in the procedure.

Placing points randomly within the PpA. Now that the PpA is defined, we need a specific location to actually place the new point. Since one of the requirements is to generate a random trajectory, the location needs to be selected randomly within the PpA. The point placement is performed in four steps. First, the PpA is calculated using the intersection of the two circles around A and B (Figure 4.1). As a second step, the minimum-bounding rectangle MBR_1 around the PpA is created (shaded in Figure 4.2). This Minimum-Bounding Rectangle (MBR) serves as an approximate candidate region to speed up the following two steps.

In the third step, a potential point $P_{pot}(x_{pot}, y_{pot})$ is temporarily placed within MBR_1. This is performed by drawing two random coordinates *(xpot, ypot)* from two uniform distributions within the ranges [MBR_{Xmin}, MBR_{Xmax}] and [MBR_{Ymin}, MBR_{Ymax}], respectively. Finally, in the fourth step, we test whether P_{pot} actually lies within the PpA. If the distance d_A is less than radius r_A, while at the same time the distance d_B is less than the radius r_B, then P_{pot} is accepted and treated as the origin for the creation of the next point. Otherwise, P_{pot} lies inside MBR_1 but falls outside the PpA; thus, the third and fourth steps of the procedure are repeated until *Ppot* can be accepted.

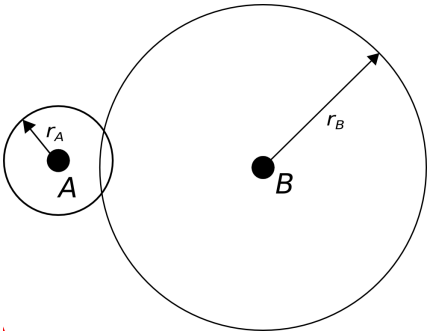

Figure 4.1: The Potential point Area (PpA) after Δt, is defined by the intersection of a circle of radius $r_A = V_{max} = 70$ km around A, and a circle of radius $r_B = 4V_{max} = 280$ km around B.

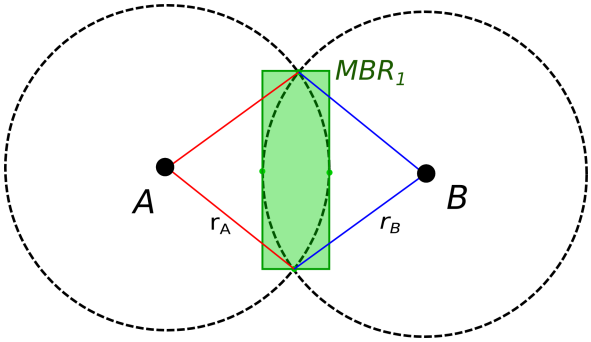

Figure 4.2: Creating the minimum bounding rectangle MBR_1 of the Potential point Area (PpA).

Using the MBR to generate and test P_{pot} greatly simplifies computations and increases efficiency, although this approach, will also generate some misses (4.4.1). Furthermore, compared to an approach that requires sampling of the PpA by some sampling scheme such as a grid, in this approach the resulting point locations and the computational performance of the algorithm are independent of the resolution of the sampling scheme.

Overall, this approach preserves the randomness of the path while observing given constraints (i.e. time budget and maximum speed), efficiently connecting the starting point with the destination point.

Completing the trajectory. Once the point P_{pot} is accepted, it becomes the new origin A_1 from which to initiate the creation of the next intermediate point A_2 of the trajectory. The time budget remaining for the completion of the trajectory is now decreased by Δt. The following example describes the steps necessary to complete the trajectory.

Example 2. The sampling interval is $\Delta t = 1$ h, hence $r_A = V_{max} = 70$ km. However, since the moving object has already travelled for one hour to generate the first point, the time to destination B is now $t_{total} = 2 \times \Delta t = 3$ h, and $r_B = 3 \times V_{max} = 210$ km. With these new values for r_A and r_B , the procedure described in Sections 4.2.1

and 4.2.2 is repeated to create the point A_2 (Figure 4.3), and analogously for the remaining points A_3 and A_4 (Figure 4.4).

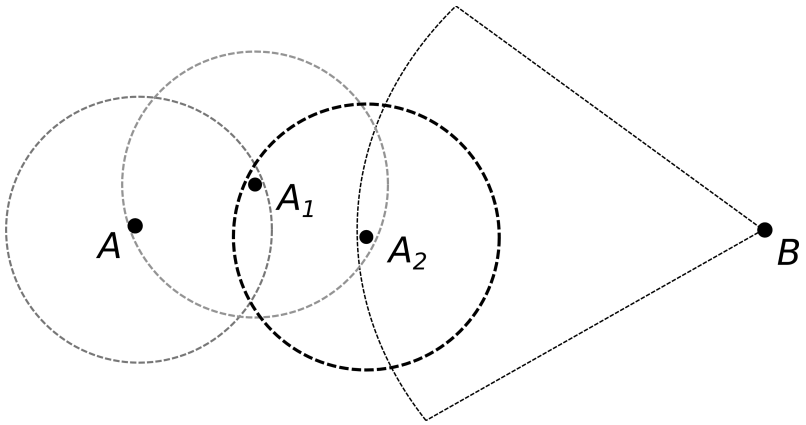

Figure 4.3: Repeating the point generation procedure.

Ultimately this procedure yields a trajectory that consists of the sequence of points $\{A, A_1, A_2, A_3, A_4, B\}$, as shown in Figure 4.4. This sequence represents a random but feasible trajectory with a given origin and destination, while not exceeding the available time budget and the given maximum speed.

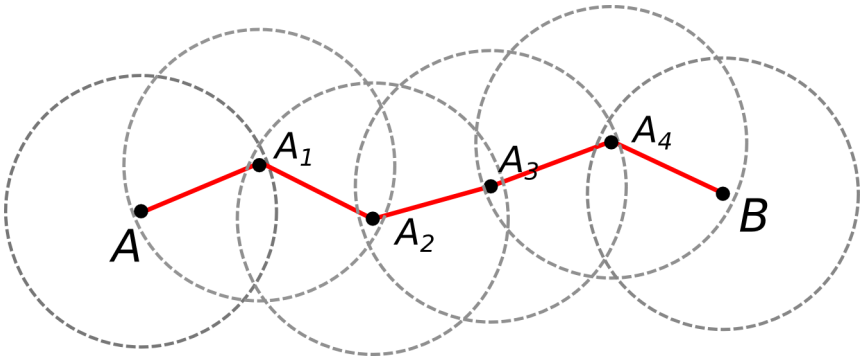

Figure 4.4: Final trajectory.

Limitation in the degrees of freedom. After creating each point, the time budget is updated based on the need to complete the remaining trajectory, thus imposing a growing limitation on the PpA. The two factors affecting the placement of a new point are the last created point and the temporal sampling interval Δt (Figure 4.4). The change of origin affects the position of the circle of all possible locations for the next point to be created (the circle around A_i), while the destination and the decreasing available time budget narrows down the extent of the second circle (the circle around B). In other words, the algorithm as described so far will initially generate largely Brownian motion until the time budget is almost used up, forcing the remaining few time-steps to be spent 'hurrying' towards the destination to reach it on time.

Figure 4.5 shows two example trajectories that both exhibit the bipartite movement regime (near-random vs. heavily drifted, respectively), the consequence of the basic algorithm. It also becomes apparent that the coarser time interval of 50,000 seconds allows for fewer vertices to be placed, thus reducing the extent of the free movement part. It is worth pointing out that this very ability to scale the temporal step of the random walk differentiates RTG from Brownian motion, which by definition ought to converge towards 0.

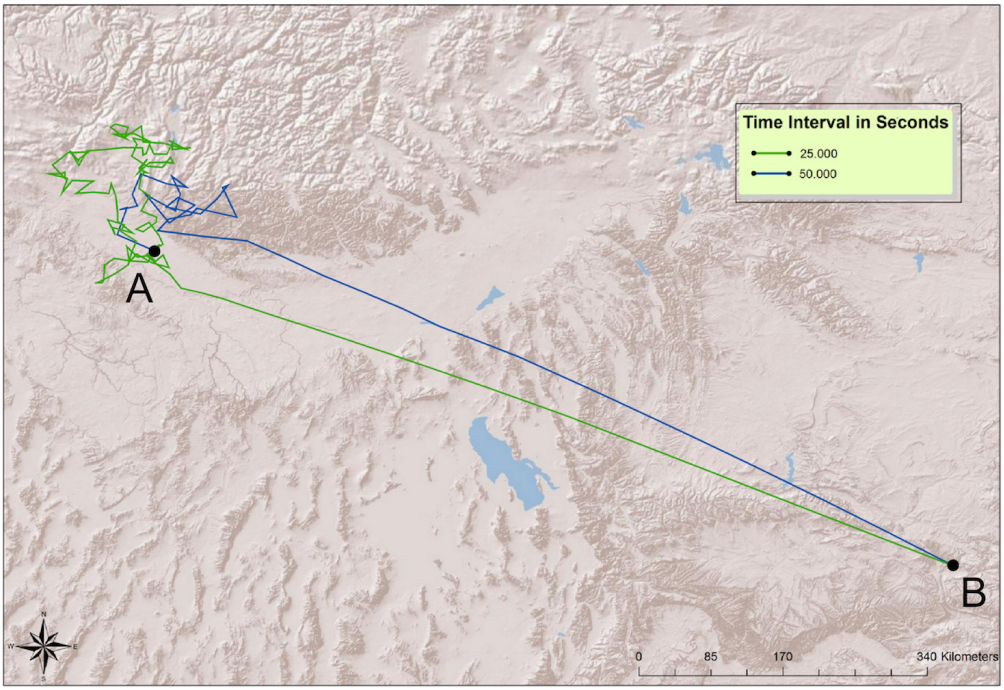

Figure 4.5: Two example trajectories that meet the requirements of Section 1 but are unrealistic, since most of the distance between the origin and the destination is covered by a heavily drifted walk.

In order to avoid the effect of 'wandering' excessively during the initial part of the trajectory generation, the reduction of the degrees of freedom in the PpA should be distributed randomly over the trajectory. This equalised temporal distribution can be achieved by making the order of the point placement no longer sequential. The approach used here is to choose an order of point placements that are random in the time-steps. The algorithm proceeds by first creating a list of time-steps necessary for generating the trajectory. Out of this list, each time-step is randomly selected, one after another, for the placement of the corresponding point. This is illustrated in Figure 4.6, where a set of six intermediate points $\{t_1, t_2, \ldots, t_6\}$ needs to be generated in order to complete the trajectory AB.

Instead of the sequential ordering $(t_1, t_2, \ldots, t_3)$ used initially, we create the points in random sequence, for instance: $(t_5, t_1, t_3, t_6, t_4, t_2)$. Each time, the last created point becomes the new starting point for the algorithm while the destination is

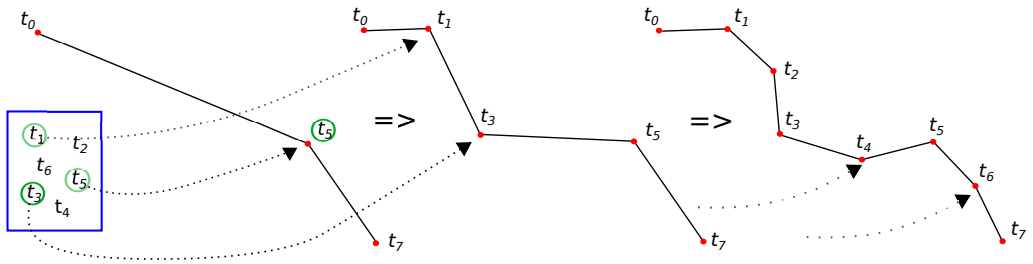

Figure 4.6: Selected random time-steps and their corresponding points.

set to the temporally next available point, if applicable. Once the list is empty, all the necessary points have been created, and the trajectory is compiled by sorting the points into temporally sequential order. A set of trajectories generated by this procedure is depicted in Figure 4.7.It is noticeable that the bipartite pattern of Figure 4.5 is no longer visible; the limitation in the degree of randomness of the walks is now evenly distributed along the trajectories.

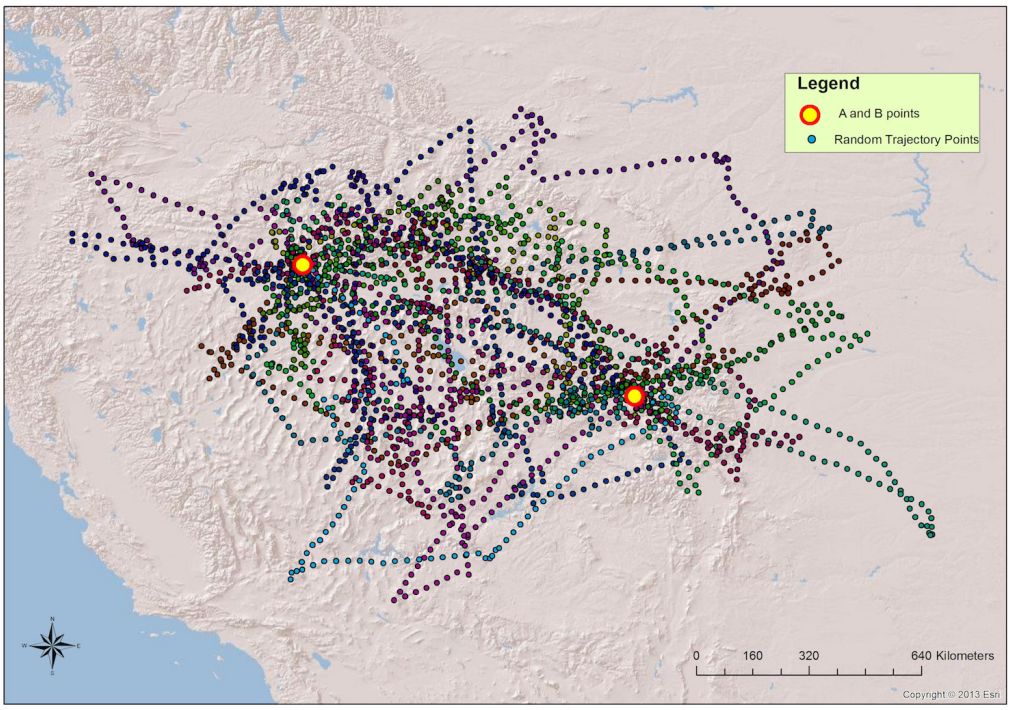

Figure 4.7: A set of 20 trajectories generated using the improved algorithm with random temporal order of point placement. The set-up of the experiment is identical to the one visualised in Figure 4.8.

Spherical version of the algorithm. Owing to the spherical shape of the Earth, using the planar version of the algorithm to create trajectories over large spatial scales bears the problem of spherical distortion, violating the coherence of distance, angles and areas in non-Euclidean space. For the reconstruction of large scale displacements, such as continental or intercontinental migratory movement, it is, therefore, necessary to extend the algorithm to accommodate spherical geometry. It is important to note here that implementing the model on the sphere does not

imply the third dimension of topography. In the present version of the algorithm, the implementation, therefore, does not account for elevation.

In the presented algorithm, we compute a circle-circle intersection and approximate it with a bounding box or accelerated random placement of points. In the spherical version of the algorithm, these circles are circles on a sphere (the Earth). If one ignores the spherical shape of the Earth, this implies that the longitude λ and the latitude ϕ of geographical coordinates are equated with the x and y of a planar coordinate system: x = λ, y = ϕ. This is equivalent to using an equirectangular (or plate carrée) projection, which is known to cause massive distortions towards the poles. Using such a projected map to define circles on the sphere, the projected circles will therefore not retain the properties of a circle. Thus, calculating the bounding box of their intersection at arbitrary positions on the globe would be rather difficult. However, knowing that the distortions of projected circles are minimal along the Equator, we transpose the two circles to the Equator, one of them on the intersection with the Central Meridian, keeping their relative distance the same. Thus, the distortion of the projected circles is minimised, and the bounding box around the intersection can be expressed by a minimum and maximum value for the width and height. This enables us to efficiently select a random point from the bounding box. Once this point is placed, we verify, using the Haversine formula, whether or not the point actually is contained in the circle-circle intersection.

Circle-circle intersections on the sphere, apart from having a different geometry, could be comprised of the same cases seen in Euclidian space. The two circles might be too far apart and have no intersection. One of the circles might be contained in the other, without their boundaries intersecting. The two circles might be touching in a single point, or finally, they might have a proper intersection with two distinct intersection points. The latter case is what we will explore in detail; the other cases are easily accounted for by evaluating inequalities on their radii and distance between the centres.

Figure 4.8 depicts two circles with a proper intersection. Although it contains straight lines, Figure 4.8 is to be interpreted as lying on the sphere, not on the plane. In this setting, the known variables are the circles' radii and the distance between their centres. The unknown variables we wish to compute are the coordinates of the intersection points, specifically the coordinates of C. The other intersection point shares the same x-coordinate but has a mirrored (i.e. negative) y-coordinate.

In 2-D Euclidean space, we would solve a set of Pythagorean equations for the right triangles ΔAMC and ΔBMC. For right triangles on a sphere, we use the spherical analogue of Pythagoras' theorem and this returns the following identities:

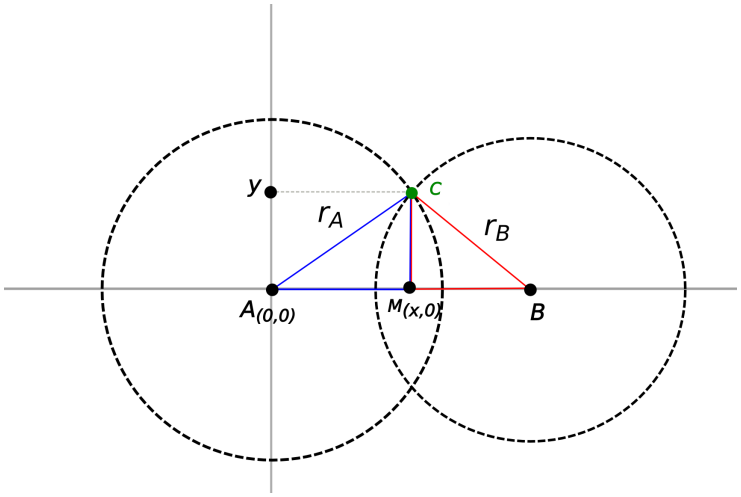

Figure 4.8: Circle-circle intersection on the sphere.

$$
\begin{aligned}
cos\left(\frac{r_A}{R_E}\right) &= cos\left(\frac{x}{R_E}\right) \times cos\left(\frac{y}{R_E}\right) \\
cos\left(\frac{r_B}{R_E}\right) &= cos\left(\frac{\overline{AB}-x}{R_E}\right) \times cos\left(\frac{y}{R_E}\right)
\end{aligned}
\tag{4.2}
$$

where r_A stands for the radius of the circle with centre A, r_B for the radius of the circle with centre B, and R_E for the radius of the Earth. The distance between the two circles is given by AB. The unknowns we wish to solve for are x and y, the coordinates of the upper intersection point.

These equations are solved following a standard procedure. An inverse cosine is only determined up to its sine, and cosine will change its sign when π is added to its argument. So, once we compute one solution of x, we have to account for three others: $-x$, $x + \pi$ and $-x + \pi$. We restrict the circles' radii to that of less than half the Earth's circumference and can thus rule out the latter two solutions. By verifying with the set of equations, we can show that unless x = 0, either x or -x is a solution.

We are, in fact, not interested in the value of x, but rather in the value of y, as it determines (half) the height of the bounding box. The width of the bounding box, defined by x_{min} and x_{max}, respectively, depends entirely on the circles' radii and the distance between their centres (Figure 4.9), as expressed by the following equations:

$$x_{min} = \overline{AB} - r_B$$
$$x_{max} = r_A \tag{4.3}$$

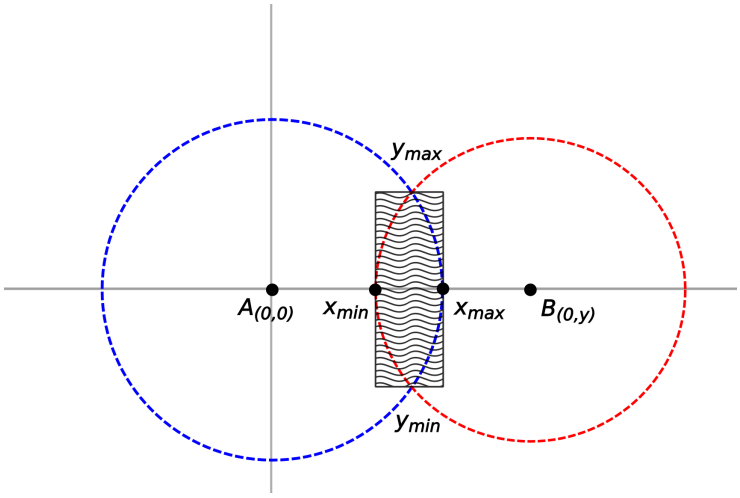

Figure 4.9: The bounding box coordinates.

The height of the bounding box, defined by y_{min} and y_{max}, is obtained by solving the system of equations 4.2 for y. Now that we have the limits of the bounding box, we repeatedly generate a point therein by generating random coordinates within these limits. We then verify the point's distance from the two circle centres using the Haversine formula and accept it if it is also contained in the circle-circle intersection. Otherwise, we generate a new point within the bounding box and repeat as described in Section 4.3.2. Finally, points that have been found to lie within the corresponding circle-circle intersections (i.e. the spherical PpA) are transformed back to their actual location on the globe.

4.4 Evaluation

In order to assess the performance of the proposed RTG algorithm, we proceeded in two steps involving the verification (internal evaluation) and the validation (external evaluation) of the algorithm. The verification step (4.4.1 involved testing the computational performance as well as the robustness of the algorithm. In the validation step (4.4.2, some points were omitted from the input trajectory and then used to assess the accuracy of the simulation. The same inputs were also used with the BBMM approach for comparison. An overall discussion (6.3) concludes this section.

4.4.1 Verification

Performance analysis. The purpose of this analysis was to test the robustness and efficiency of the algorithm under different parameter combinations. In order to obtain more robust estimates, run times were calculated as an average over five generated sample runs. The model was tested against five main parameters: the total number of trajectories created, the number of points within each trajectory, the maximum speed of the moving object, the geographic distance covered, and the total duration of the movement.

Table 4.1 summarises the empirical analysis of the computational efficiency, realised on a standard desktop PC (Intel QuadCore i7 CPU 870 @ 2.93 GHz, with 8 GB RAM, running LinuxMint 16 (petra) 3.11.0-12-generic and Python 2.7). Each scenario is separated by a horizontal line altering one or two parameters at a time, leaving the rest unchanged. It becomes obvious that the RTG algorithm maintains a linear relation between the running time and each parameter tested. As expected, the increase in the number of points, the greater the distance, the total number of trajectories, or the duration of the simulated movement simply increase the load but not the complexity of the calculations.

As described above, the proposed algorithm first approximates the PpA with a bounding box to speed up the point placement step (4.2.2). The price to pay is that misses might occur: points that lie inside the MBR but fall outside the PpA. In the implementation of the algorithm, the maximum number of unsuccessful attempts for placing a random point was limited to 10; otherwise, a miss is reported for the particular point. Misses happened rarely, however, as Table 4.2 shows. The reported accuracy of the algorithm is the percentage of these misses (i.e. the random points created falling outside the PpA) of the total number of points to be created. For example, in the first line, the algorithm failed to create 66 points of the 50,000 requested, yielding an accuracy of 99.868 %.

Boundary cases. Both in the temporal and in the spatial dimensions, the RTG algorithm may encounter boundary cases, that is, cases when the lower or upper bound of an input parameter is reached. The lower temporal bound can be described by the minimum time required to reach the destination. In this case, the moving object has to follow the shortest possible path in order to make it to the destination in time. Thus, the trajectory will form the straight line connecting the two endpoints. This case is handled by creating all the necessary points on the straight line connecting the origin and destination points, as this is the only feasible solution.

Table 4.1: Empirical analysis of computational efficiency

Run	# of Trajs	# of Points	Speed (m/s)	Distance (km)	Duration of Movement (s)	Duration of Runtime (s)
1	5	45	0.4198	2.386	21645	0.478
2	50	45	0.4198	2.386	21645	4.728
3	500	45	0.4198	2.386	21645	47.05
4	5	88	0.4198	2.386	21645	0.89
5	50	88	0.4198	2.386	21645	9.016
6	500	88	0.4198	2.386	21645	90.124
7	10	150	0.4198	2.386	21645	2.909
8	10	218	0.4198	2.386	21645	4.37
9	10	272	0.4198	2.386	21645	5.411
10	10	362	0.4198	2.386	21645	7.205
11	10	410	0.4198	2.386	21645	8.233
12	10	505	0.4198	2.386	21645	10.287
13	1	520	0.4198	0.02	373419	1.135
14	1	1053	0.4198	0.02	373419	2.258

Table 4.2: Robustness of the algorithm

Misses	Trajectories	Points	Distance (km)	Accuracy
66	500	100	200	99.868 %
12	100	100	200	99.88 %
27	500	100	2	99.946 %
2	100	100	2	99.98 %

The lower spatial bound is reached when the intersection area between the two circles (see Figure 4.1) is so small that it may be considered negligible for the purpose of the simulation, that is if it falls below the minimum spatial resolution. As can be seen in Figure 4.10 (I), when the candidate area is larger than the threshold (in this case, 1 m for the width of the bounding box), the point is placed randomly as outlined in Section 4.2.2. If, however, the candidate area is smaller than the threshold (Figure 4.10 (II)), the centroid of the intersection area is selected as the new point.

The opposite case, an upper bound, is reached in the case of long trajectories, such as those covering half or even the entire globe. These represent potential challenges for a trajectory generator.

Thus, to gain a better overview of the algorithm's robustness, such cases and also special areas (i.e. areas around the poles) were tested. For this test, a leading trajectory was needed that could be created with any movement model or parameters as long as it covers distances created than 80-90° on the globe. In the present case, a 1000-step correlated random walk was created starting from the Equator, moving up to the North Pole. More specifically, the wrapped-Cauchy function of the `circular`

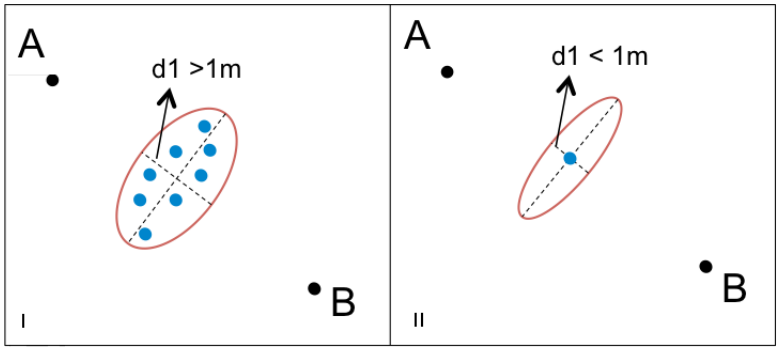

Figure 4.10: Boundary cases of the spatial resolution in point placement.

package in the R statistics language was used, with the mean direction of the distribution set to 0 and the concentration parameter randomly set to 0.8. Once the 1000 steps were calculated, the trajectory was down-sampled, keeping 1 point out of 10, and the RTG algorithm was executed 100 times for each pair of remaining consecutive points. This resulted in 100 trajectories, generating points at the same timestamps as the missing points. Figure 4.11 shows the result of this operation.

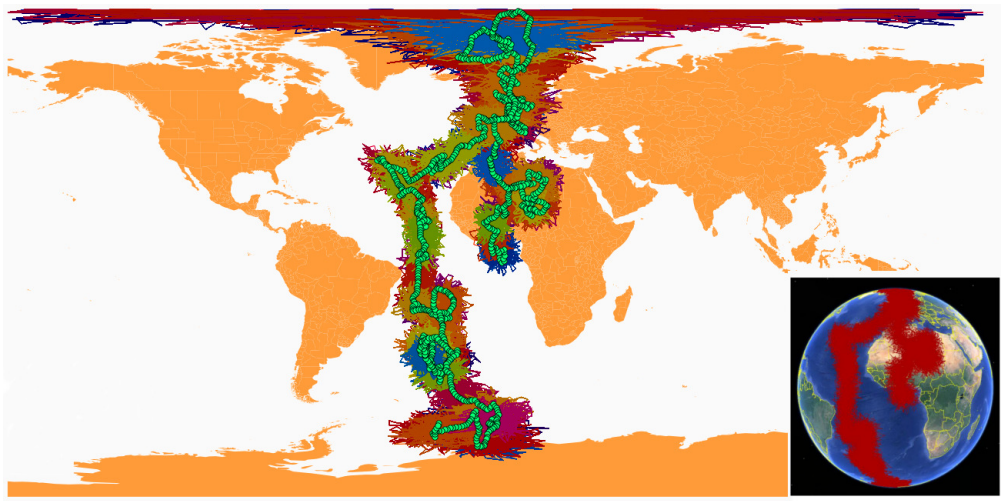

Figure 4.11: 100 runs of the RTG algorithm on a 1000-step, resampled CRW (green points) that starts from the intersection of the Equator with the prime meridian, moving towards the North Pole and then back to the southern hemisphere. The inset map shows all RTG runs in a spherical projection.

The green points denote the CRW that was then down-sampled and used as the input to the RTG algorithm. Each of the 100 simulated trajectories has a randomly assigned colour, with 9 new points created for each pair of consecutive sample points. Owing to the equirectangular projection used for the main map of Figure 4.11, trajectories are particularly distorted towards the North Pole. Thus, the inset map shows all RTG runs on a spherical projection.

4.4.2 Validation

Procedure. The data used for the validation experiment was taken from the Movebank site (see Acknowledgements). The group working on this dataset set up an experiment tracking a flock of white storks that follows a migration path from Poland to South Africa, via Sudan and Tanzania. The birds were tagged with GPS devices that recorded their positions in two modes: a high-frequency mode (average time interval = 49 min) for maximum spatio-temporal resolution and a lower frequency mode (average time interval = 360 min) for making sure that the batteries would last until the end of the journey. The stork trajectory in Figure 4.12 clearly depicts the two different sampling modes. This particular trajectory was used for the evaluation scenario.

In a different application context, the RTG algorithm would be used to realistically densify the coarse part of the data, making it comparable to the high-frequency part. For the purposes of evaluation, though, the focus was placed on the high-frequency part of the trajectory. The initial dataset was down-sampled, omitting 50 % of the points, and then used as input data for simulating the omitted points at $\Delta t/2$, to create one point between each pairs of points. The simulation was run 50 times for each missing part of the trajectory and then rasterised the result, yielding the count of trajectories in each pixel, thus creating a density surface. The raster template used was the same pixel size (1118.113 x 1498.257 m) and extent as with the result the Brownian bridge approach created for easier comparison between the two. These counts can be transformed into relative densities as a proportion of the total trajectories, and after normalisation considered as a probability surface that can be compared against the probability surface extracted from the BBMM, with probability values compared at multiple control points.

RTG algorithm. The RTG algorithm requires three input parameters: two or more spatio-temporal points, speed, and the time interval between two consecutive points. The input points used were the points of the down-sampled high-frequency part of the white stork trajectory (see above). The speed parameter was extracted from the empirical distribution of the given sample. The default speed was the median of the sample (0.026 m/s) for most cases. The reason for selecting the median and not the maximum speed of the sample data was that the latter was an order of magnitude higher than the most commonly used speed, and it was detected in less than 2 % of the instances. In order to approximate the speed behaviour of the object when higher speed was required, the value was randomly selected from the accepted range of the distribution, hence the maximum speed was taken into account. The time interval was calculated from the real-time difference between two consecutive sample points and the number of points to be created. For instance, if two input

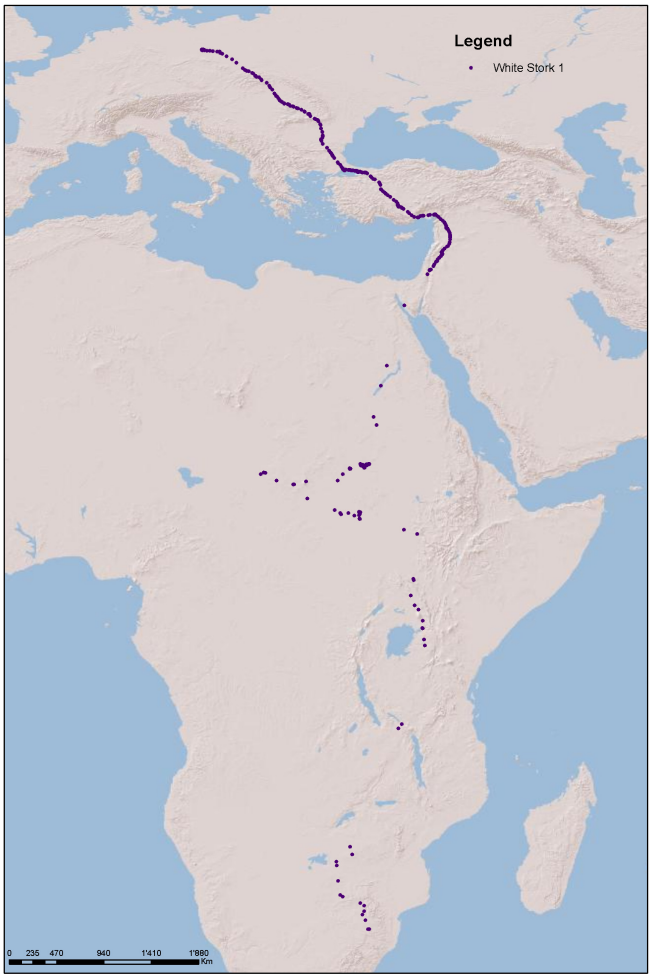

Figure 4.12: Overview of the trajectory of White Stork 1 (violet dots)

points P1 and P3 have a time difference $\Delta t = 30$ min and one intermediate point P2 must be created, the time interval will be $\Delta t/2 = 15$ min; if two points were required, the time interval would be 1 min.

In total, 50 trajectories for each of the 951 pairs of points were created. Figure 4.13 depicts the full result of the algorithm, whereas Figure 4.14 is a close-up of parts of it. Once the trajectories were created, a line density algorithm was used to create the density surface. This was done for visualisation purposes, contrary to the evaluation procedure described above, where the trajectory counts were rasterised directly. The bandwidth for the line density algorithm was set to 0.3 degrees, which covers the maximum recorded distance between two consecutive points on the real trajectory. A snippet of the result is presented in Figure 4.15.

Brownian Bridges Model of Movement. The Utilisation Distribution (UD) of the Brownian Bridges Model of Movement was computed using the move package in R.

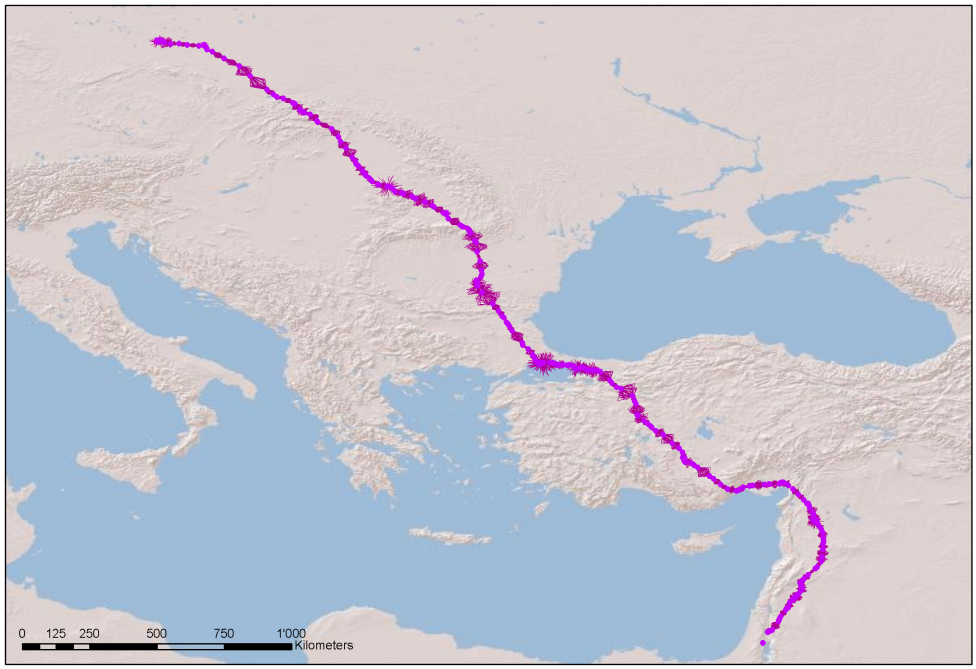

Figure 4.13: Re-sampled and simulated trajectories.

The same down-sampled trajectory was used as input. The location error was set to 10 m, equal to the estimated GPS error. The margin used for behavioural change point analysis was set to 21 m, and the window size along the track was set to 9 m. The calculation extent was set to be 30 % larger than the bounding box of the input points in order to avoid edge effects. The result of the calculation is depicted in the top row of Figure 4.16, including the 99 % contour of the UD probability. For a more differentiated picture of the UD variation, the log transformed version is shown to the right.

Comparison. The BBMM method creates a surface of UD probabilities, whereas the RTG algorithm is conditioned to create single random trajectories. At first sight, there seems to be little use in trying to compare these methods. However, both approaches model the movement between an origin and a destination; they both are derived from the assumption of RWs, and they both account for the total time of travel. For large numbers of trajectories generated with the RTG algorithm, we thus expect to see an increasing similarity between the density of RTG trajectories and the corresponding BBMM occupancy probabilities. That is, while the comparison of the two methods by absolute numbers is not possible, we may nevertheless compare the trajectory density surface generated by the RTG algorithm with the BBMM occupancy probability surface in qualitative terms. For this comparison, the RTG trajectories were imported to R and then rasterised by yielding the line count within each pixel

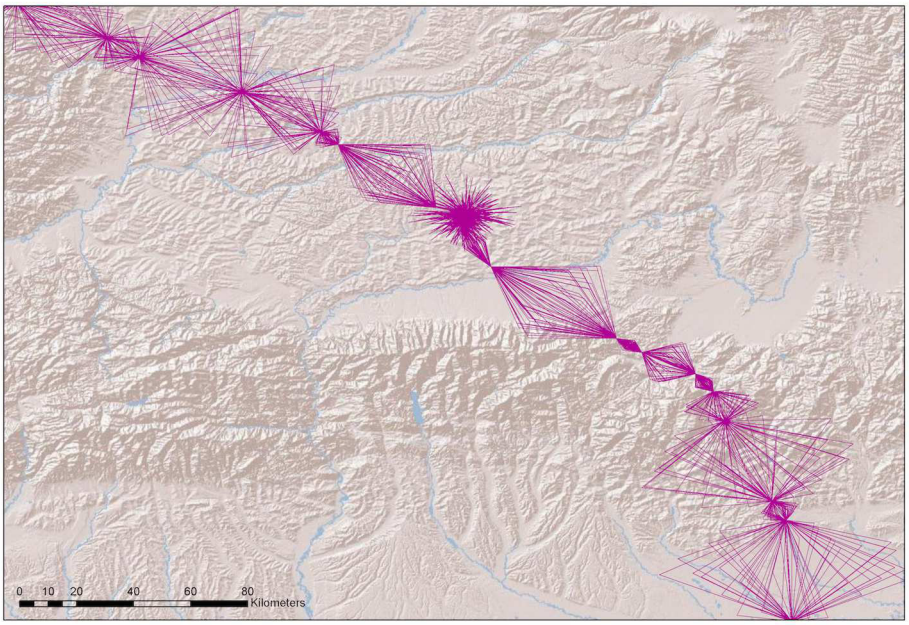

Figure 4.14: Close-up view of Figure 4.13.

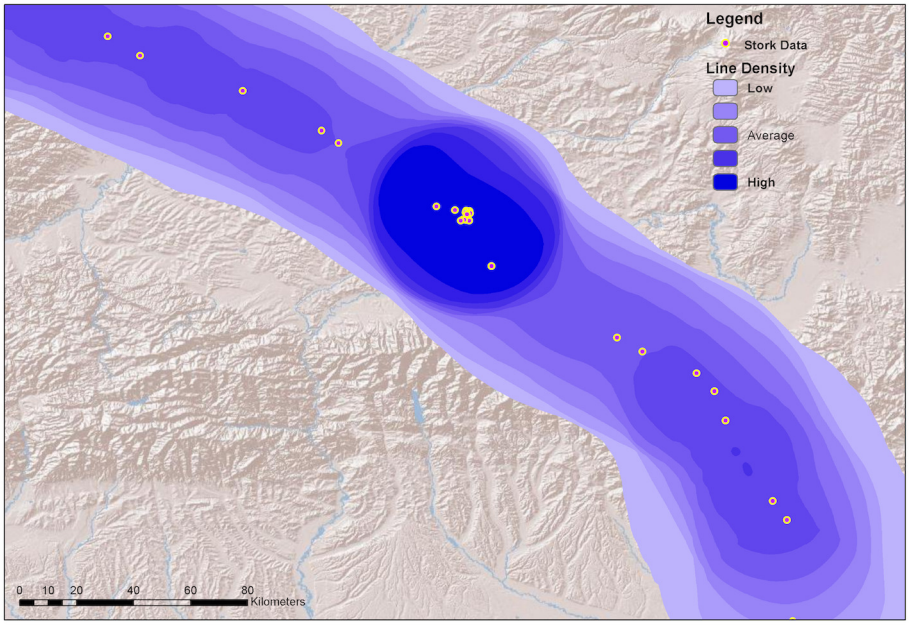

Figure 4.15: Line density surface created over the 50 trajectories of Figure 4.13.

(Figure 4.16). The two rasters (RTG, BBMM) had identical extent and resolution and were normalised before the final comparison.

In each of the 951 control point locations, the UD probabilities were extracted for both approaches. Figure 4.17 (left) visualises the results for a random subset

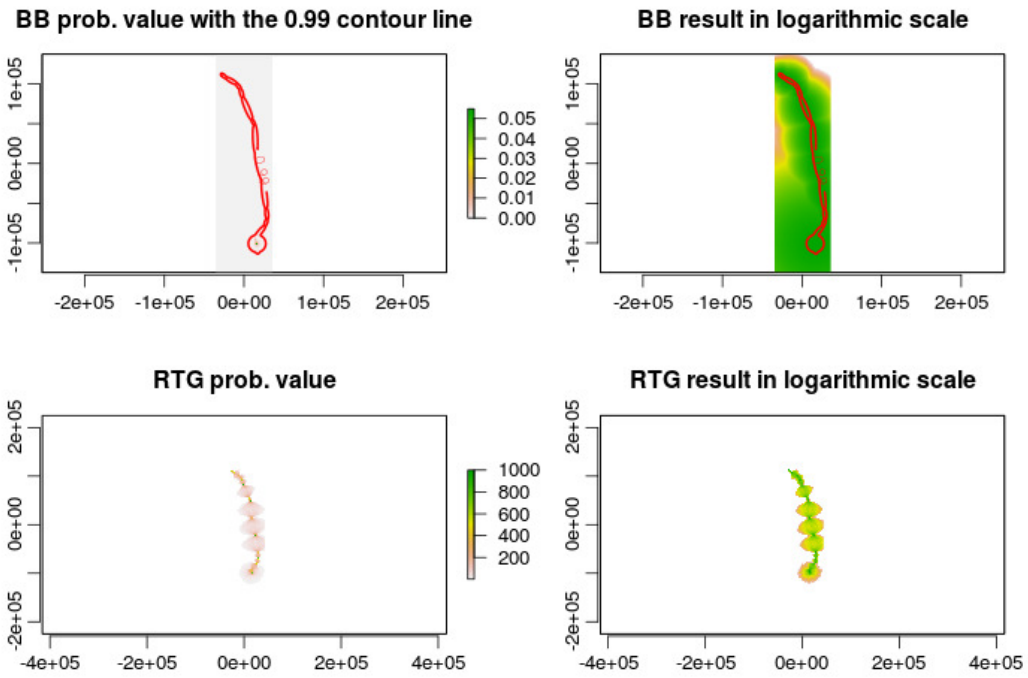

Figure 4.16: Top row: Utilisation distribution using the BBMM, with 99 % contour line, in linear (left) and logarithmic (right) scales. Bottom row: Normalised RTG density distribution in linear (left) and logarithmic (right) scales.

of 30 consecutive control points. The plot shows two distinctive clusters in both approaches: low probabilities for the lower half of point IDs; high probabilities for the upper half. These values are more easily explained if we consider the distances between each consecutive pair of points (Figure 4.17, right). The expected value should be 1.0 for all control points. This is achieved quite well at short distances between points, where both simulated scores are almost perfect (i.e. they almost reach 1.0). However, the greater the distance between points, the more inaccurate the two algorithms, though the RTG algorithm performs significantly better than the BBMM. Overall, the average accuracy for the 30 control points was 62.23 % for the RTG algorithm and 48.38 % for the BBMM, where 100 % would imply a perfect agreement between the expected and the simulated value.

4.4.3 Discussion

In order to draw overall conclusions from the comparative evaluation of the proposed algorithm, let us start by recalling the requirements stated in the methodology in section 4.3. The task at hand was to develop an algorithm that could generate synthetic trajectories between two given points. These trajectories should be based on the principle of random walks, and at the same time, conditioned by control

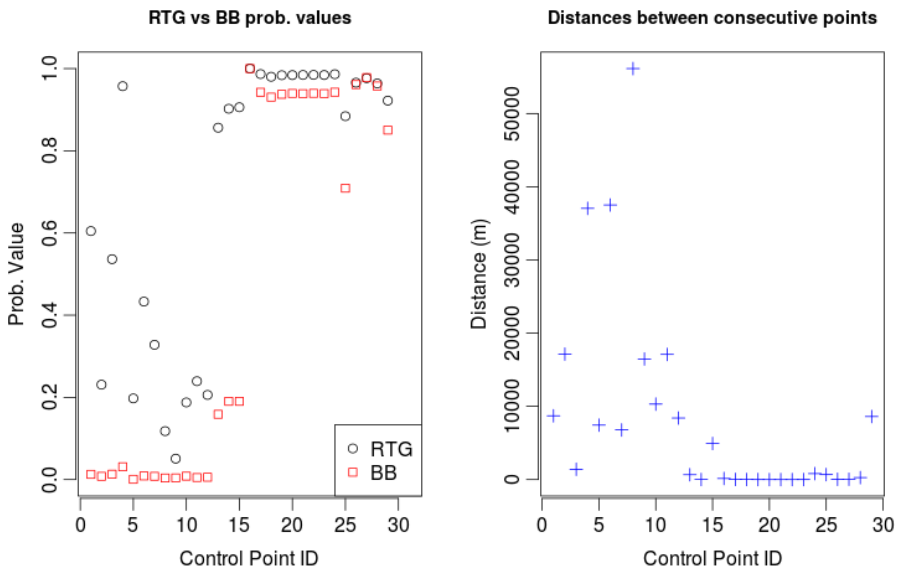

Figure 4.17: Left: The probability values for the RTG and BBMM approaches, respectively, extracted at a random subset of 30 consecutive control points. Right: The distances between each pair of consecutive control points.

parameters such as time budget and maximum speed. The algorithm should be efficient to enable the creation of large numbers of trajectories. And it should be able to cope with the spherical shape of the Earth, as the given endpoints may be separated by hundreds or even thousands of kilometres.

The proposed RTG algorithm clearly outperforms other random walk algorithms. As the above empirical evaluation results have shown, the RTG algorithm maintains the random walk model basis, with the advantage of reliably and efficiently generating random walks between two given endpoints (i.e. the trajectory is guaranteed to reach the destination). Additionally, the RTG algorithm allows us to control the time budget, as opposed to other RW approaches.

The comparison to the BBMM has demonstrated that by generating large numbers of trajectories and creating a density surface out of these, the RTG algorithm produces comparable results. This is not so surprising since both approaches derive from the random walk model, though it should be acknowledged that the distribution used in the RTG to generate vertices within the bounding box is a uniform distribution. In contrast, the BBMM uses a normal distribution perpendicular to the segment connecting the two endpoints. However, note that, due to the constrained travel time between the two endpoints, the empirically generated trajectories are actually not uniformly distributed in space but rather follow a distribution that is not unlike that of the BBMM. In addition, the RTG is free of three shortcomings of the BBMM method. First, the initial outcome is a trajectory and not a probability surface,

allowing the addition of patterns of a specific behaviour in the simulation. Second, in the case of the BBMM, the value of probability does not fully decay to 0 (i.e., each pixel within the study area is assigned a non-zero probability). Violating the maximum speed is thus inevitable if the moving object is to make it to the destination. Third, as the size of the study area becomes larger and the resolution of the analysis becomes finer, the performance of the BBMM algorithm decreases dramatically as a function of the raster surface that must be generated. This overhead is particularly large if the aim is only to extract individual trajectories from the BBMM surface. Conversely, the performance of the proposed RTG algorithm depends linearly on the number of intermediate points that need to be placed. This number, in turn, is dictated only by the time budget, the temporal sampling interval, and the placement of individual points (which are free of complicated numerical simulations, as shown above). Finally, the two approaches could be considered as complementary, as the RTG can be used to enrich the input data for the BBMM approach.

Compared to the space-time prism approaches, including Song and Miller [114], the proposed RTG algorithm avoids the necessity of considering time as a continuous parameter. Instead, RTG focuses on generating individual trajectories for specific time slices in the STP model. This enables this algorithm to handle large geographic distances, accounting for spherical distortion.

5 Empirically informed, probable movement between two points

> *Though this be madness, yet there is method in't.*
>
> — **William Shakespeare, Hamlet**
> (poet, playwright, actor)

The Random Trajectory Generator (RTG) introduced in Chapter 4 models the *possible*, physically unconstrained movement, with no consideration of the mover's preferences, internal state, or spatial environment. In this Chapter, the 'Empirical Random Trajectory Generator' (eRTG) is introduced, initially explaining how the mover's typical movement behaviour can be incorporated into a CRW, followed in Chapter 6 by the addition of the contextual preference (avoidance, attraction) that either decreases or increases the visiting probability of certain areas by the mover.

5.1 Motivation

The ability to create random trajectories with defined geometric properties provides the opportunity to generate null models, which are useful for a wide range of applications in movement analysis [17, 42]. One of the major advantages of generating a null model of spatio-temporally explicit random trajectories is that it provides the possibility to then integrate contextual, in this case ecological, background information to derive contextually informed utilisation distribution probabilities. Being able to derive these probabilities enables the researcher to tackle new questions, for example, how does the movement behaviour of an animal adapt to context changes? Which contextual factors interfere with the animal's movement and in which way?

If the aim is to generate null models for studies in movement ecology, the RTG is insufficient. We need a method that can represent both possible and *probable* movement, which manifests the typical characteristics of a particular mover in a particular behavioural state (e.g., a migrating bar-headed goose; [191]). Also, the method should allow the integration of spatial context, that is, relevant aspects of the spatial environment in which the movement is embedded.

Hence, I will approach the description of the eRTG algorithm in two parts. This Chapter will focus on physically *unconstrained*, *probable* movement, which refers to a context-agnostic movement while maintaining the movement characteristics of empirically observed input data. In the following Chapter 6 I will describe how context-awareness is added in the eRTG algorithm.

5.2 Research challenges

In addition to the requirements stated in Section 4.1 for the initial version of the RTG algorithm — connecting two known endpoints reliably with a CRW, given a constant speed and a limited time budget — three more requirements become essential to the trajectory generation process if we aim to improve ecological realism.

First and foremost, the need for a step-wise decision process is apparent. Given a starting point, the mover needs to 'decide' where to move next based on a set of criteria. Ecologically speaking, this means that the animal would move depending on which need it tries to fulfil or in reaction to its environmental stimuli: if hungry, the probability of selecting a location rich in food resources is higher than any other; if too cold, the probability of migrating is increasing. From a GIS perspective, that would be translated into a Probability Distribution (PD) field out of which the agent will select its next movement.

Second, the decision about the next move itself should be based on movement parameter distributions seen in input data consisting of observed tracks of individuals of the same species and in the same behavioural state as the mover, as these can be considered to represent ecologically 'realistic' values. According to the framework of Nathan et al. [55], the movement path of an animate object should be the result of a sequence of decisions based on the mover's capacities, internal state, and available external opportunities. Since it is in most cases not known which factor affects the movement, in this work, the observed movement is taken as a reference of the range of behaviours that can be realised. Then, any simulated trajectory that replicates the movement parameter distributions of the observed input data is considered 'realistic'.

Third, while movement parameters in the form of step length (i.e., speed), turning angle and persistence would be incorporated in each trajectory, the mover should still make it to the destination without creating unrealistic stretches of movement (i.e., stretches not complying with the given input data distributions) on the way.

As in the RTG algorithm, a random element is introduced, allowing to incorporate the stochastic nature of the empirical data (containing both measurement errors and the arbitrariness of an animal's movement decisions). Adding randomness creates an opportunity to embrace the complex nature of the decision-making process in movement instead of fighting against it. All the unknown factors that affect movement can be hierarchically tested on a 'naive' trajectory against the actual behaviour of the mover, as represented in the recorded data. At the same time, introducing randomness to the movement must be managed cautiously, as it contradicts the ending point condition: the more freedom one allows in the movement, the more difficult it is to end up at a predetermined point.

In summary, an algorithm should be developed that:

- is based on the CRW model of movement;
- creates individual trajectories in a step-wise manner;
- replicates movement patterns of the input data
- allows to move towards, and eventually reach, a specific target within a given time with the minimum necessary corrections, while maintaining the geometric properties of the observed input trajectories; and
- enables incorporation of spatial context in creating the next step, thus allowing the mover to take a movement decision in response to the environment it is exposed to (this point will be considered in more detail in Chapter 6).

5.3 Methodology

5.3.1 Underlying framework

The description of the movement for each step of the trajectory that is being generated, requires selecting the next step's direction and length. Given a previous movement state, there might be limitations in the available range of values due to physical constraints (e.g., the inertia that prevents immediate halts) or biological capacities (e.g., maximum locomotion speed). These limitations, along with preferences, internal state and spatial context, contribute to a mover's decision-making process. Therefore, empirically observed data give a good starting point for extracting the basic characteristics of the movement. The single assumption justifying this extraction is that the recorded sample sufficiently describes the way the mover interacts with its environment at the given temporal and spatial scale. The Empirical Random Trajectory Generator's key principle is to use empirical, observed trajectory data of single or multiple individuals for acquiring the movement parameters

informing the generation of multiple, geometrically similar trajectories between two given endpoints.

Since the distributional characteristics of step length (or speed) and turning angle will be different in different behavioural states — e.g. faster and more directed movement during migration, slower and more tortuous movement during foraging — the raw trajectory data have to be subdivided into segments of homogeneous movement behaviour before generating the empirical movement parameter distributions used as input for the eRTG. A variety of tools and methods are available for this purpose [192–196].

On a higher level, the *movement model* employed in the eRTG is based on the combination of two probabilities: an origin-repulsiveness, forward-looking one and a destination-attractiveness, backward-looking one. On a lower level, the movement itself is reduced to turning angles and step lengths (as basic variables) and their probability distributions (with persistence and heteroscedasticity as their characteristics). Step length forms the positive value expressing the distance of moving from one point to the next. Turning angle refers to the change of direction between consecutive segments. Due to convention, positive turning angle values mean a right turn, negative turning angles indicate a left turn.

Table 5.1 summarises the variables used for describing the movement model. The duration of the movement is derived by the timestamps of the endpoints. Since the time budget is equally distributed between steps, the number of steps required for the movement can then be calculated.

The input data's probability distributions describe the frequency and range of values that the basic variables can take. The variable persistence $R(x)$ expresses the tendency of each variable to maintain its value: 100 % persistence of speed translates to cruising at a constant speed, whereas 90 % means that the next speed value is expected to differ up to 10 % from the current one. The persistence in this implementation refers to the difference between the values in the current and previous time step ($time - lag = 1$), here called *auto-difference*. The heteroscedasticity between two variables expresses the condition that the variance of one variable is not uniformly spread across the range of the second variable's values. A practical example: the faster a bird flies, the less prone it is to change its direction. When its speed decreases, sharper turning angles are more likely. The additional variables qD and qT are used to describe the movement (in terms of distance and turning angle) from the perspective of the destination point.

Table 5.1: Variables used in the eRTG

Variable	Definition
A, B	Starting point,ending point
n	# of points
T	Original trajectory
TA, SL, dt, d, b	Turning angle, step length, time interval, Euclidean distance, bearing (azimuth)
qD, qT	Euclidean distance and turning angle to destination
$f(x)$	Probability density of variable x
$R(x)$	Persistence of x with lag = 1 (auto-difference)
$S(x, y)$	Heteroscedasticity between x and y
P, Q, Z	Forward, backward, total probability

5.3.2 The algorithm

The development of the eRTG based on RTG was done in multiple steps of incremental complication. Adding the probability function in the next-step-selection involves extraction, fitting and incorporation of movement parameter distributions in the trajectory generation. The influence of spatial context is added as a layer that defines avoidance and preference scores, respectively (details to follow in Chapter 6). The simulation is performed on a two-dimensional plane, using projected coordinates, angles in radians, and Euclidean distance. The 3-D extension of the algorithm has been developed by Unterfinger [197].

Calculating the forward probability *P*

The forward probability P aims to describe movement from the mover's perspective, which in its simplest description consists of the starting points, the distance covered (step length) and the absolute direction of the move (bearing). The latter can be easily translated to a relative direction (turning angle) given a previous location (and hence the incoming direction is known). The advantage of using TA over bearing b is crucial, as TA and SL are closely related (the faster the mover moves, the more difficult it is for it to make sharp turns), whereas the relationship between SL and b is more difficult to describe (one can go fast or slow regardless of the bearing).

Calculating $f(TA)$, $f(SL)$ and $S(f(TA, SL)$ requires binning of the data. To select the bin width, the well established Freedman-Diaconis rule is leveraged, aiming to minimise the difference between the area under the empirical probability distribution and the area under the theoretical probability distribution [198]. The density estimate spreads the mass of the empirical distribution function on a regular grid and then uses the fast Fourier transform ([199] to convolve this approximation with

a discretised version of the kernel. The density at the specified points is calculated using linear approximation on the discretised version.

The simulation has a single-step memory, meaning the persistence range equals 1; therefore, each variable's (V) current value V_n is only affected by the previous one V_{n-1}, here called auto-difference. The persistence $R(SL)$ and $R(TA)$ is calculated by fitting a linear model[3] to the kernel density estimate[4] of each auto-difference to approximate the underlying probability density.

Once all the movement parameters have been extracted, the normalised P is calculated:

$$P = f(sL) \times f(ta) \times \sqrt{R(sL) \times R(ta)}, \tag{5.1}$$

where P is the product of all the movement parameters' density functions multiplied by the product of the auto-difference of the two variables. The square root compensates for the covariance of the two variables.

Creating an unconditional empirical random walk walk

The forward probability P enables the user to create random trajectories with similar geometric characteristics as the input data, though with only one fixed point (i.e., the current point). This ability is the stepping stone for calculating the backward probability Q. The calculation of Q requires the parameters qD and qT to be determined in each step lag that is available, from the first one, up to step $n-1$. In essence: How much distance can still be covered given the number of remaining steps to make in the available time budget? Q should know how far a distance, on average, the empirical distribution of the movement parameters permit a mover to cover, given the number of steps still to be simulated. This is central in order to adjust the strength of the pull towards the endpoint. If the number of remaining steps do not permit too much lingering based on the movement parameters, Q must exert strong limitation on the range of the probabilities for P.

Adjusting the power of Q is easier for the last part of the trajectory, as for smaller lags of remaining time-steps, there are more data points regarding how far the mover can travel. Considering the extreme cases of lag $= 1$ and lag $= n-1$ helps to illustrate the problem. In the first case, the observations collected are $n-1$, therefore potentially enough to calculate the statistics of the required movement parameters. In the second case, only a single measurement exists (the full length of

[3] `https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/approxfun`
[4] `https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/density`

the input trajectory), indicating the total distance the individual can cover in n steps, providing almost no information to create a density distribution.

In order to overcome this issue, the concept of the *unconditional Empirical Random Walk (uERW)* was developed: extending the original trajectory while keeping its geometric characteristics intact allows deriving multiple estimates of the distance the mover covers as a function of the number of steps made. By extending a trajectory of n steps, one can estimate how far a mover can get when performing, e.g. 1,000 times n steps, for which empirically we have only one realisation. This extension allows the algorithm to estimate the strength of Q.

In order to generate the uERW from the starting point $A(x, y)$, we randomly sample a pair of turning angle ta and step length sL values from the given probability matrix and add a white noise shift as an error term. The coordinates of the new point are then obtained as follows:

$$\begin{aligned} x_i &= x_{i-1} + cos(b_{i-1} + ta + W_x) \times (sL + W_y), \\ y_i &= y_{i-1} + sin(b_{i-1} + ta + W_x) \times (sL + W_y) \end{aligned} \tag{5.2}$$

where the x-coordinate of the new point equals the x-coordinate of the last point, increased by the product of the cosine of the angle, multiplied by the step length. The angle equals the previous bearing plus the sampled turning angle and the white noise term for x. The step length equals the sampled step length plus the respective white noise. The y-coordinate can be read in the same way.

The new point becomes the starting point for the next selection, and the process is repeated until N points have been generated, where N is a number substantially larger than the original number of points (in this case, 10 * n).

The resulting unconditional Empirical Random Walk (uERW) is illustrated in blue colour in Figure 5.1. It has similar movement parameters as the original trajectory with a larger sample of distances covered for any available time lag, enabling the calculation of statistics for the qD and qT parameters for all the available lags of the movement.

Calculating the backward probability *Q*

The backward probability Q, conceptually speaking, acts as a constant reminder of the final destination, gradually pulling the mover towards the target at any point of the trajectory without creating the excessive drift artefact seen in Figure 4.6.

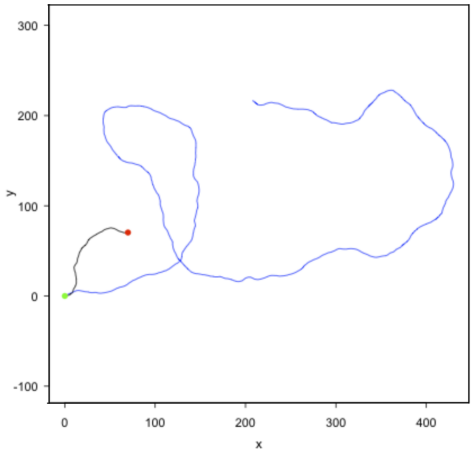

Figure 5.1: Observed input trajectory (black) vs. unconditional Empirical Random Walk (uERW) (blue) with the same movement properties. The latter acts as an extended version (10 times the length) of the first, enabling the user to draw enough samples for the calculation of the backward probability Q.

Calculating Q requires the extraction of reliable descriptive statistics for the qD and qT variables for each movement step, which necessitates the step of the uERW described in the preceding subsection (Figure 5.1).

Given an input trajectory and the uERW the eRTG algorithm extracts the distributions of the distance and the angle towards the target (Figure 5.2) for each time step (Figure 5.3 left), creating frequency distributions for both variables. The end product is a two-dimensional distribution matrix that acts as a probability surface (Figure 5.3 right) selecting of the next point, depending on which step of the movement the mover is at.

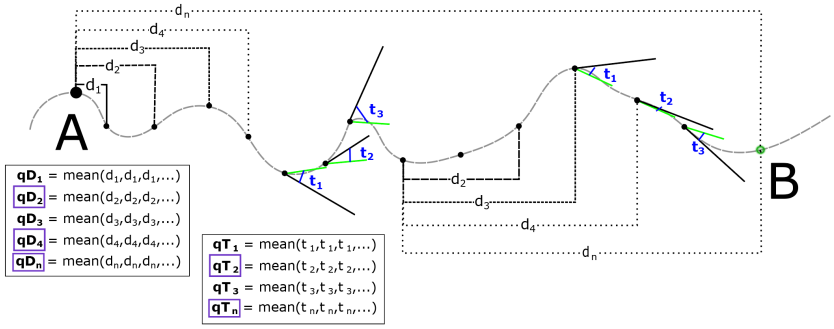

Figure 5.2: Initial step of the calculation of the Q probability. Distribution derivation and calculation of qT, qD.

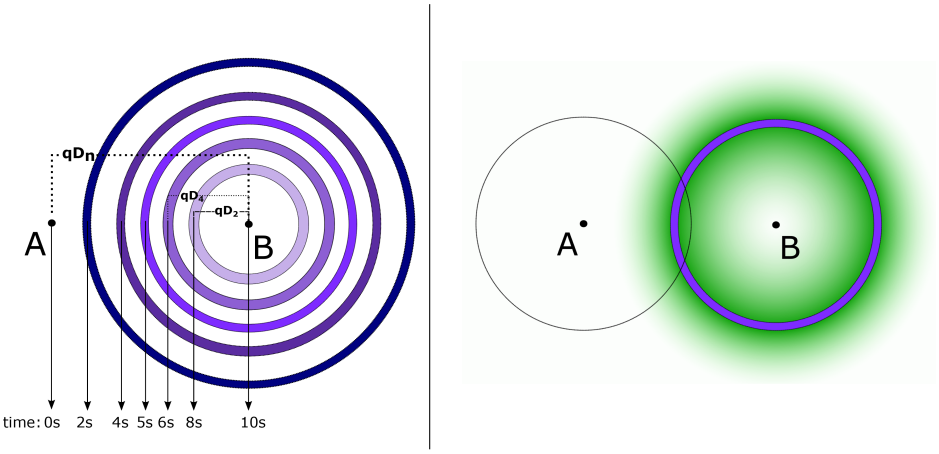

Figure 5.3: Left: Second step of the calculation of Q probability. Creating a global moving window. Right: Final step for calculation of Q probability. Density estimate extraction for qT, qD at time $t = 6$. The black circle around centre A is an indication of the forward probability P.

Generating the trajectories

Starting from the starting point A, with P and Q calculated, their product Z can be obtained, representing the total probability surface of the next movement:

$$Z = P \times Q. \tag{5.3}$$

The next point of movement $p_i(x_i, y_i)$ is selected with a sample drawn out of the Z probability surface, shifted by an error value (ϵ) to represent the error of measurement during the data collection:

$$\begin{aligned} x_i &= x_{i-1} + dx_{z \hookrightarrow \mathcal{Z}} + \epsilon \hookrightarrow \mathcal{N} \\ y_i &= y_{i-1} + dy_{z \hookrightarrow \mathcal{Z}} + \epsilon \hookrightarrow \mathcal{N} \end{aligned} \tag{5.4}$$

The selected point is used as a starting point for the new selection, and the procedure continues *n*-1 times, resulting in the required final trajectory. This step-wise approach is a key element that enables the user to control the incorporation of randomness, noise, and spatial context in every point created. Randomness is introduced by selecting a random point on the product of the P and Q probability surface. The noise slightly shifts this selected location (measurement error), and the spatial context can be introduced by adding a probability surface that affects the point selection in each step, as presented in the following Chapter 6. The overview of the entire methodology can be seen in Figure 5.4.

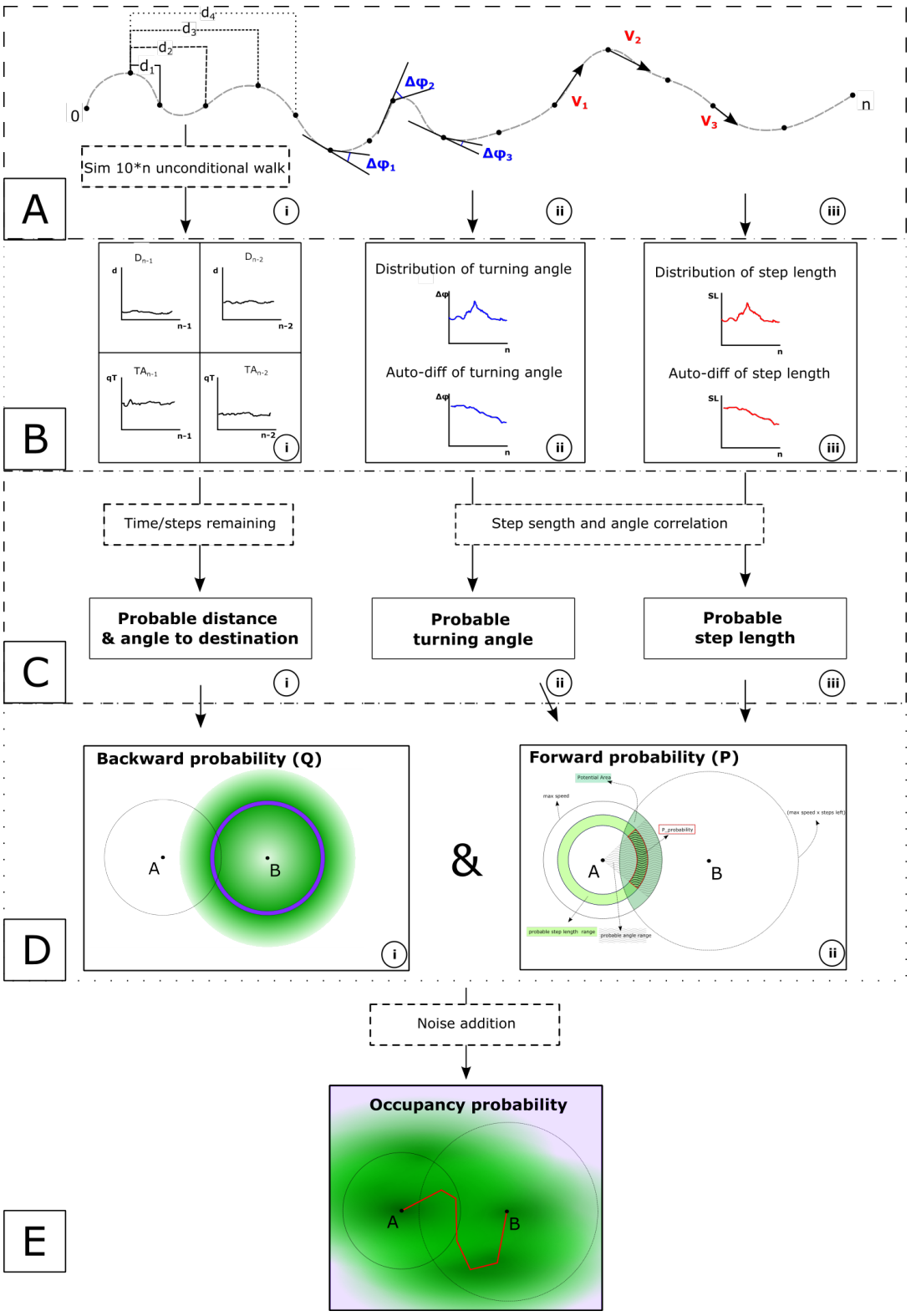

Figure 5.4: Overview of the eRTG methodology.

5.4 Evaluation

5.4.1 Creating the test cases

Multiple synthetic input trajectories (from here on: 'template trajectories') were created to test the methodology's performance. Each template carries significantly different geometric characteristics and generally relates to a physical behaviour of an animal. Figure 5.5 presents three examples of movement: the first one (5.5A) has a 99.99 % correlation in the direction of movement and high persistence in keeping its step length unchanged, resulting in a nearly straight line with a nearly equidistant step length in each time-step. This example is comparable to a migrating bird's movement pattern, which keeps its heading and speed steady. The middle panel of Figure 5.5 exhibits a meandering movement that resembles the foraging behaviour, with smaller, variable step lengths and little to no persistence of the heading — as expected when an animal is searching for food. The step length and bearing persistence, in this case, was 50 %. The third example, shown in the right panel of Figure 5.5, imitates a hybrid behaviour of an animal largely moving along straight segments but occasionally acting opportunistically by altering its speed and deviating from its route. In this case, the persistence of direction is 80 %, and the step length persistence is 50 %.

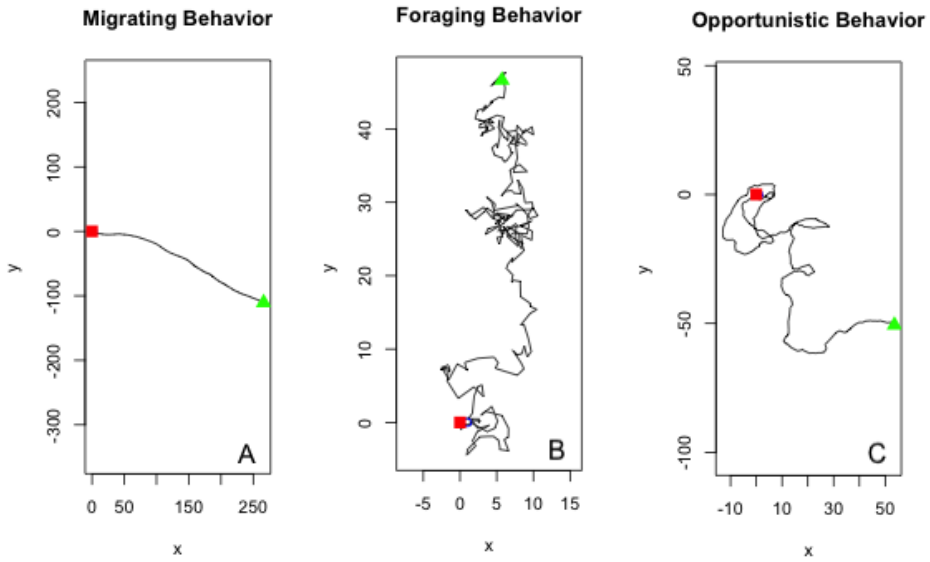

Figure 5.5: Three examples of the synthetic trajectories moving from start (red square) to endpoint (green arrow) used as input data (i.e., template trajectories) for the eRTG simulation.

Following the same procedure, 100 template trajectories were created, and for each, the eRTG was run 100 times. Figure 5.6 shows three of these templates with the respective ten first generated results. For each of the generated trajectories, an

evaluation report is compiled, aiming to give the user the right level of confidence for applying eRTG in each case study. Based on the template trajectory shown in Figure 5.1, the verification process conducted on 10,000 simulated trajectories as described in Section 5.4.2.

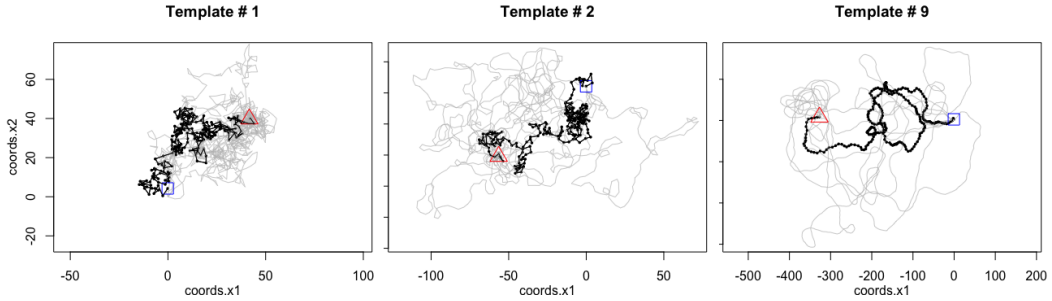

Figure 5.6: Three template trajectories (black) used to test the eRTG and their generated results (grey). The red triangle and blue square depict the start and endpoints, respectively.

5.4.2 Verification of the movement parameters

Dead-end occurrences and selection of initial conditions

A dead-end is created when the mover runs out of options for its next step. Since the eRTG is based on a probabilistic model, the consecutive selection of low probability combinations of movement parameters is possible. Given this constellation, the mover might reach a point where the remaining distance to the destination is greater than its moving capacity. In such a case, the eRTG algorithm cannot draw valid values for the movement parameters anymore, and thus the movement is halted, and the trajectory creation process is terminated.

The expected number of dead-ends varies from case to case, but it is generally related to the degree of variation of the distribution of the movement parameters and the selection of initial conditions. Given the migration example seen in Figure 4.5 (left), the mover travels almost at maximum speed, with minor directional adjustment; therefore, any 'wrong' selection of direction will leave very little room for correction. In contrast, in the foraging case 4.5 (middle), the mover has practically no limitation in the step selection, even towards completing the simulation. Intuitively, the number of dead-ends is expected to be significantly higher in the first scenario than in the second case.

The effect of selecting the initial conditions is a well-discussed point in the simulation literature [200–202]. Before running any simulation procedure, the initial conditions need to be set. In the case of the eRTG simulation algorithm, given a starting point

P_1 the initial conditions are limited to the step length and bearing for the point P_0, enabling to initiate the auto-difference calculation for the point that follows next.

In order to better assess the impact of the selection of initial conditions on the trajectory generation process, a sensitivity analysis was carried out. Three scenarios were used for this assessment, covering the same start and destination location, time budget, and movement behaviour. The only difference was in the way the initial values were selected (mean, median, and random) for the step length and the bearing distributions, respectively. Each of the three scenarios was run 100 times, aiming to create 1,000 trajectories in each run. The number of successful attempts was counted and plotted against each other in Figure 5.7. Only minor deviations were observed in the results, allowing us to conclude that the effect of initial conditions on the number of dead-ends created is limited, with the median selection option performing slightly better than the mean selection option. The random scenario performed worse than the rest, which is explainable if one considers the combination of the limited time budget and the possibility of selecting an entirely disadvantageous initial direction. The dead-ends, in this case, were mainly the trajectories that started towards the opposite rather than the destination direction, and the mover failed to perform a 180-degrees turn while still honouring the movement parameter distributions.

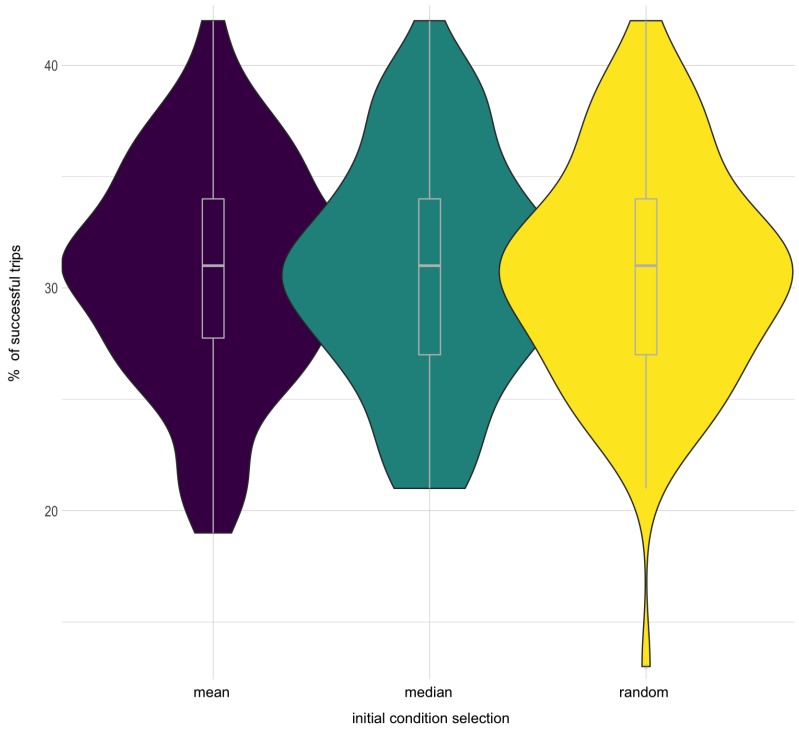

Figure 5.7: The three initial condition scenarios used for sensitivity analysis of the eRTG.

Assessing the completed trajectories

The evaluation of a simulation result is commonly based on validation and verification tests [203]. Validation tests contain the fitness for purpose component, meaning that the outcome of the simulation is good enough to be used for the intended purpose while at the same time being an accurate representation of the real system. While the first part can be conducted by a single check (number of completed trajectories, meaning: reaching the intended destination point), the second is by definition not possible as we have no complete view of the real system that the eRTG is attempting to replicate. As a reminder of the original motivation of this work: the ability to allow for emerging behaviours with the chance to get insights otherwise hidden by the observed data was the main driving power for the development of the algorithm.

The verification tests aim to assess to what degree the movement parameters of the input templates have been preserved throughout the trajectory generation process. Given that each trajectory is random by design and the directional bias applied on each segment is non-linear, coming up with objective, bulletproof metrics for similarity is challenging. Hence, three levels of verification have been used. On the first level, a visual inspection of the statistics of the trajectories is performed, aiming to identify any apparent outliers in the results. On a second level, an intra-trajectory comparison of the starting and ending segments for all the results checks for any unjustified bias. Ultimately, on the third level, a statistical similarity comparison between the distribution shape of the result against the respective distribution of the input data should confirm the preservation of movement parameters.

Level 1: Visual inspection. A qualitative way to compare the input/output data is by superimposing the distributions for both the step length and the turning angle, as seen in Figure 5.8. If the statistics of the completed trajectories do not follow the input data, then it is clear that further inspection of the results is needed before proceeding. The same reasoning was applied on the distribution of the auto-differences. The auto-difference parameter focuses only on $time - lag = 1$, capturing the resistance of the variable to change its subsequent value. The boxplots in Figures 5.9 and 5.10 compare the first 100 trajectories (boxplots of auto-differences) against the mean and standard deviation of the original dataset (blue dashed and dotted lines, respectively). The overlaid histograms reveal a good match between the input and output data.

Level 2: Controlling starting and ending points for bias. The visual inspection suggests that the generated trajectories maintain, on average, the geometric characteristics of the input template trajectory, as intended. Given the 'last minute', unjustified

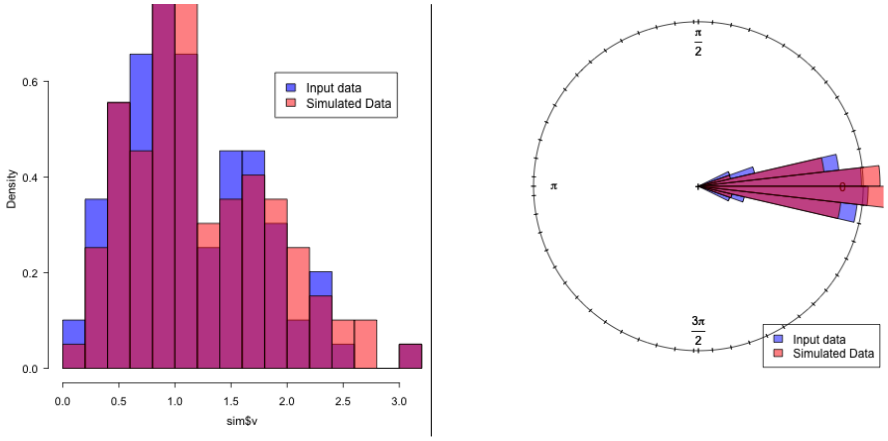

Figure 5.8: Histogram of step length distributions (left): The original distribution (blue) is overlaid on the distribution of generated trajectories (red). Circular plot of the turning angle distributions (right): The original distribution (blue) is overlaid on the distribution of generated trajectories (red).

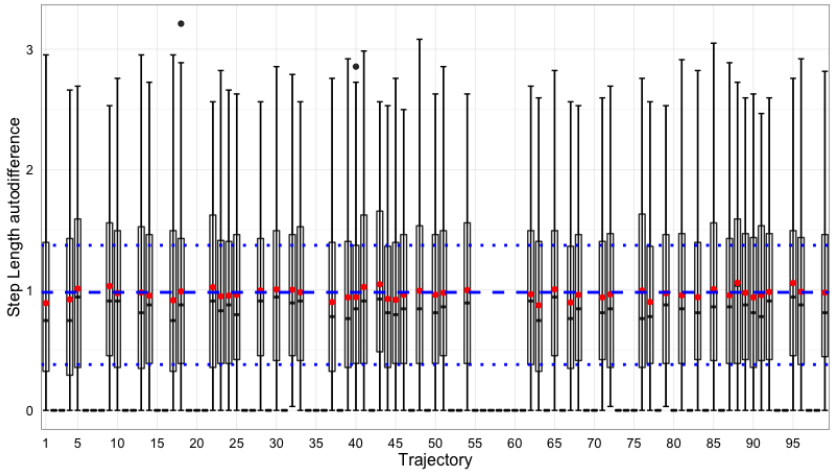

Figure 5.9: Boxplots of the auto-differences of step lengths for the 100 first simulated trajectories in relation to the original mean (blue dashed line) and standard deviation (blue dotted lines). The overlaid plots reveal no apparent discrepancy between input and output data.

stretches noticed in the RTG version of the algorithm (see Figure 4.5), an additional test was put in place to control for potential bias at the beginning or the end of the trajectories generated. Figure 5.11 compares the movement parameters of the initial 15 % against the final 15 % moves of the trajectory (practically the end and the beginning) for all the generated trajectories (boxplots) and the original trajectory's (red dot). The verification tests were run on multiple settings, input data, and temporal resolutions, increasing the confidence in the predictable behaviour of the algorithm: the results are almost identical for both the beginning and final part of the trajectory, confirming that the algorithm behaves consistently regardless of the degrees of freedom available to perform the movement.

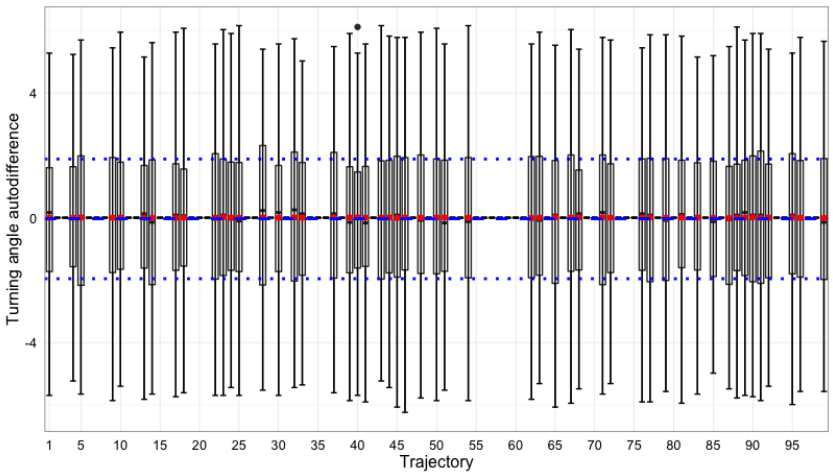

Figure 5.10: Boxplots of the auto-differences of the turning angles for the 100 first simulated trajectories in relation to the original mean (blue dashed line) and standard deviation (blue dotted lines). The overlaid plots reveal no apparent discrepancy between input and output data.

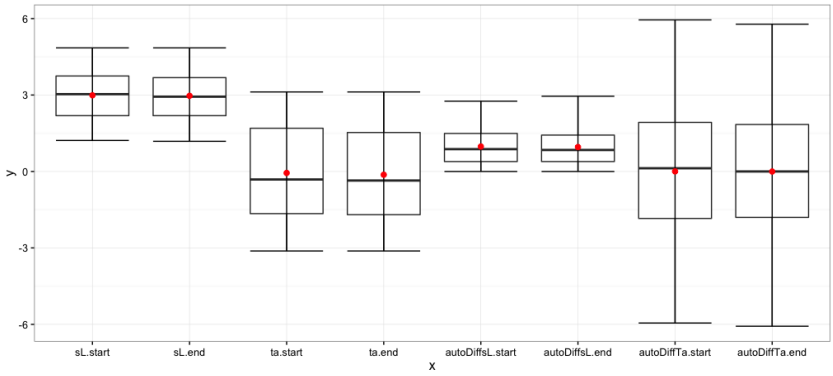

Figure 5.11: Comparison of the first against the last 15% of moves for all generated trajectories (boxplots) to ensure that no bias is present at the beginning and the end of the trajectory. The original trajectory is represented with a red dot.

Level 3: Statistical Evaluation. Formally expressing the aforementioned visual inspection of the simulation results requires testing if the two distributions are statistically significantly different: assuming two independent samples from similar distributions, their Cumulative Distribution Functions (CDFs) are expected to be similar. This relationship also holds in inverse direction: by measuring the differences between the CDFs, the similarity of the distributions, out of which the samples were drawn, can be assessed. Two similarity hypotheses were defined (Table 5.2), each of which was accepted or rejected based on the significance value $p = 0.05$ derived from a suitable statistical test.

In order to select the most appropriate evaluation metric from a collection of standard statistical tests, the different metrics were evaluated for their performance, using samples of known similarity and control. This performance evaluation was carried

out on datasets of different statistical distribution, sample size, and degree of similarity by creating multiple data scenarios.

Table 5.2: The hypotheses used in significance testing.

H0	The two distributions do not differ significantly
H1	The two distributions differ significantly.

The input data scenarios consisted of five reference data samples generated using five different types of distribution (normal, binomial, Tweedie, Weibull, and von Mises). For each reference dataset, nine combinations of simulations of varying sample size (100, 1,000, 10,000 data points) and similarity level (high, moderate, low similarity) were created. Table 5.3 gives a summary of the data scenarios generated. The definition of the first two data properties derives from the dataset available in each case study, whereas the third one was introduced to answer the question: "Does the degree of the underlying similarity affect the performance of a particular test?"

Data distribution	**Sample size**	**Similarity level**	**Statistical test**
Normal	100	low	t-test independent
Binomial	1,000	moderate	t-test related
Tweedie	10,000	high	Chi-squared
Weibull			Cressie-Read power divergence
von Mises			Kolmogorov-Smirnov
			Mann-Whitney rank
			Wilcoxon rank-sum & signed-rank
			Kruskal-Wallis *H*

Table 5.3: Properties of the created data scenarios (Columns 1-3) and the statistical tests (Column 4) assessed with the aim of selecting the most suitable test for evaluating the eRTG.

In total, 45 different input data scenarios were used in the evaluation of statistical tests. For each of these scenarios, 1,000 instances were created and submitted to each statistical test for similarity assessment. The result was then compared against the expected outcome (preselected conditions) and the Δ between the two was taken to express the suitability of each test to serve as a metric for the specific data scenario.

Figure 5.12 illustrates the comparative evaluation result for the scenario where the data follows a normal distribution and the degree of similarity of the testing dataset is set to 99 %. The illustration can be read as follows: When comparing datasets that follow a normal distribution with a sample size of 100, all tests except the Wilcoxon signed-rank test can consistently detect the similarity. As the sample size increases, there is a clear deterioration in the quality of the results for almost all tests. Chi-squared test was more consistent in producing the expected value, with the Kolmogorov-Smirnov test second overall, the Epps–Singleton two-sample test

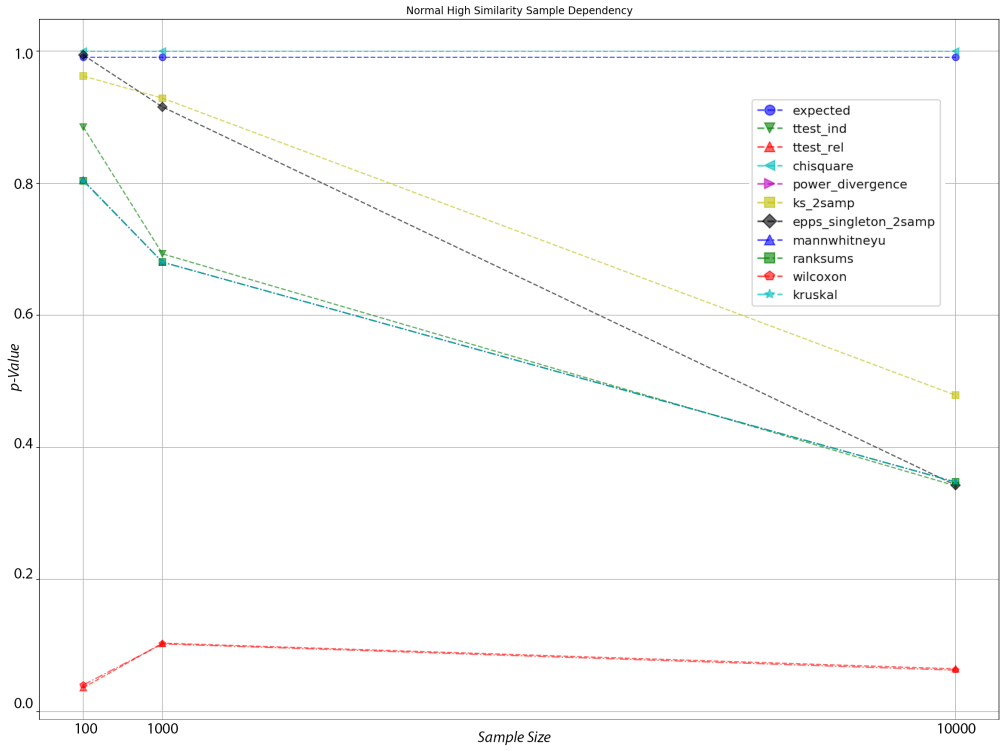

Figure 5.12: Plotting the performance of all statistical tests (and expected values) when comparing datasets that represent the following scenario: normal distribution, high degree of similarity (99 %), sample size between 100 and 10,000.

a close third, and second when the sample size is less than 1,000 data points. It is important to state here that ranking aside, all tests, except Wilcoxon, did capture the similarity, as the p-value is consistently lower than 0.05; so, in every case, H0 of Table 5.2 would be confirmed. Given that this study was empirical, further sources of information were searched to confirm the results and assist in selecting a suitable test. According to Slakter [204], the Kolmogorov-Smirnov test is more conservative than Chi-squared, but it treats each observation individually and therefore handles heterogeneous and small size samples (<50) better. The Chi-squared performs calculations on binned data, therefore depends significantly on how the sample is distributed in each bin[5,6]. All sources considered, the Kolmogorov-Smirnov test was selected as a suitable metric for evaluating the eRTG in the ecological case study application presented in the following Chapter (6), though depending on the case at hand, further exploration may be required.

[5] `https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm`
[6] `https://ned.ipac.caltech.edu/level5/Wall2/Wal4_2.html`

6 Embedding the movement in spatial context

> *Essentially, all models are wrong, but some are useful.*
>
> — **George E. P. Box**
> Statistician

The Empirical Random Trajectory Generator (eRTG) presented in Chapter 5 can replicate the mover's behaviour given as input movement parameter distributions generated from observed trajectory data, but it does so with no contextual awareness. In this Chapter, the role of spatial context is introduced, and an animal movement case study is used as a worked example to demonstrate how spatial context can be integrated into the trajectory generation.

6.1 Motivation

Monitoring how well an animal's migration strategy adapts to the changing environment is the focus of phenological research [205]. Species that fail to adapt the timing and geographic footprint of their migration to the changing conditions of their habitat possess fewer chances of survival [206]. The temporal aspect of bird migration phenology has seen advances, though research on the geographical dimension has only been prioritised recently [205]. Recent studies based on technologically advanced sensors, including radar, give new insights into the relationship between the spatial variability of bird migration and common environmental drivers [207]. At the same time, climate models predict that by 2050 the Arctic Ocean will be ice-free each summer[7], opening new spaces for birds to migrate where they have never been recorded flying before. Predicting their use of space, identifying environmental drivers, and translating them to geographic gradients requires tools for analysing and testing hypotheses on a significant amount of data. Since data is either unavailable (e.g., for possible new migration routes in the Arctic), too expensive or cumbersome

[7] https://climate.esa.int/en/projects/sea-ice/news/simulations-suggest-ice-free-arctic-summers-2050/

to generate, or raises ethical questions[8], simulation, and more specifically trajectory generation, can be used as an alternative way to help the research community advance their knowledge.

Hence, in this chapter, the eRTG algorithm is extended by integrating spatial context. The example of white stork migration is used to demonstrate, in a series of experiments, how spatial context representing environmental variables can be integrated into the basic algorithm and how this affects the trajectories created by the eRTG.

As previously discussed, a comprehensive validation of the result is by definition not possible (Section 5.4.2). Therefore, the evaluation in the examples of this chapter focuses on the verification of the results [203].

6.2 Data

6.2.1 Overview

To demonstrate the potential of the eRTG algorithm, a set of GPS trajectories was selected, recording individual white storks breeding across Central Europe while migrating towards their wintering grounds in Africa in autumn and returning in spring.

The white stork (*Ciconia ciconia*)[9] is arguably one of the easiest to recognise long-migrant bird species due to its approximately 1-metre-long figure, 2-metre wingspan and pronounced long slim legs that are clearly visible from afar even when airborne. The latter allows it to move easily in wetlands, where storks find food for themselves and their young [208–210]. Nowadays, though, the white stork has come under more pressure, mainly due to the intensification of agriculture, the increase of drainage and insecticide use, hunting pressure during stopovers and overwintering, as well as collisions with power lines and wind farms. All of the above call for protecting suitable breeding and wintering habitats and safe migratory corridors to move between habitats. The research and conservation community has been active on the topic, with the Convention on the Conservation of Migratory Species of Wild Animals attempting to formulate a plan for the species' conservation[10] along the eastern migration route from their European breeding grounds to their wintering grounds in middle and southern Africa.

[8] https://psmag.com/environment/the-ethics-of-better-animal-tracking
[9] https://g.co/kgs/Hwvake
[10] https://www.cms.int/en/convention-text

Despite its large size, the white stork is rather energy-efficient, completing the 4,000–5,000 km trip in three to four weeks. The storks travel eight to ten hours per day on average and rest for the next 14 to 16 hours [211]. This travel is possible due to intelligently selecting routes across landscapes with strong winds (for tailwind support) and thermal uplifts, which assist the birds in reaching the desired height and then gliding, without the need of flapping their wings (known as soaring behaviour). The route selection process has not been fully deciphered, and research is currently being conducted to identify the main drivers through hypotheses testing and deductive reasoning [208]. Factors such as vegetation intensity, terrain roughness, population density, presence of coastline, etc. have already entered the equation describing bird migration, but additional factors such as light pollution, ice clearance, drainage systems, etc. remain to be potentially included [212, 213].

6.2.2 Data pre-processing

The dataset used for this study consists of GPS records of the species white stork (*Ciconia ciconia*), similar to the one presented in Section 4.4. In both cases, the data is curated and publicly shared by `https://www.movebank.org/`, an "online platform that helps researchers manage, share, analyse and archive animal movement data". An overview of the data can be seen in Figure 6.1. Working on curated data simplified significantly the amount of pre-processing needed.

Filtering and cleaning of the data. The focus of the study was on the migration of the species; therefore, only trajectories that covered a latitudinal distance larger than $\approx 25^\circ$ were considered relevant to migration and qualified for the study. A data quality check was performed to remove all the trajectories with multiple duplicate records, large temporal gaps, and inconsistent location information (e.g. sudden change of continent within the same minute).

Temporal resampling, segmented, and sample balancing. Given that the migration journey of a stork takes place over several weeks, the available high-frequency data (with a time interval in the order of minutes) was down-sampled to hourly intervals for the first cleaning and then to daily intervals in order to be more compatible with the scale of the behaviour under study. Each stork's migratory trip, southbound and northbound, respectively, was segmented and used as a stand-alone instance while maintaining the ability to trace it back to the specific individual. Last but not least, in order to avoid the statistical bias while fitting the linear models capturing the movement behaviour, all trajectories were tested for keeping stable time intervals between the GPS records.

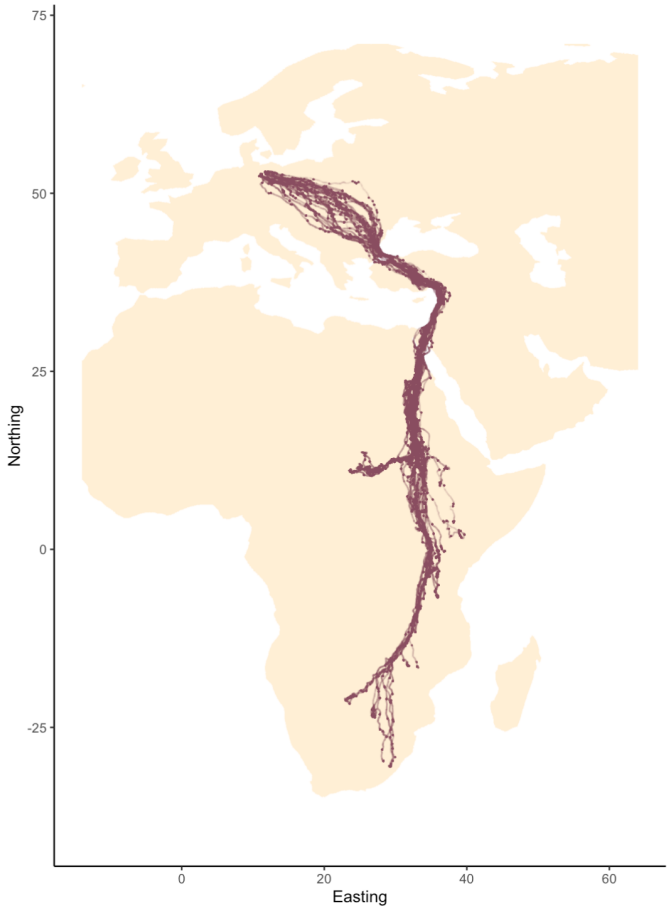

Figure 6.1: Data overview: 500,000 records of 36 white storks migrating in 2012–2016.

Binning of the data. There is significant research being carried out on how to condense a data series into easy-to-interpret groups of points (bins), helpful when dealing with problems caused by categorised continuous variables [214]. The selection of the bin size affects the simulation at its core, as it affects how the basic variable (auto)correlations are calculated. The bin width (w) selection for each variable was calculated based on its interquartile range ($IQR(x)$) according to the Freedman-Diaconis rule [198]:

$$w = 2 \times \frac{IQR(x)}{\sqrt[3]{n}} \tag{6.1}$$

The binning of the data for the density estimations of the simulated trajectories was accomplished based on automatic bandwidth selection in the function *bandwidth.nrd* of the MASS package in R [215].

Identifying a fit-for-training sample. Since migratory behaviour is the main focus of the study, the recorded white stork data was filtered, as mentioned earlier. Only trajectories with a latitudinal difference of greater than 25 degrees within the same

calendar year were selected, denoting that a long-haul flight (migration) took place at least once.

These trajectories incorporated multiple behaviours, which ought to be disentangled to yield input data for the eRTG representing homogeneous behaviour. Hard k-means clustering with $k = 2$ on the recorded speed returned two clusters: one representing the migration behaviour used as input for the eRTG and one representing the non-migratory (e.g., stopover) behaviour, which was omitted, as seen in Figure 6.2.

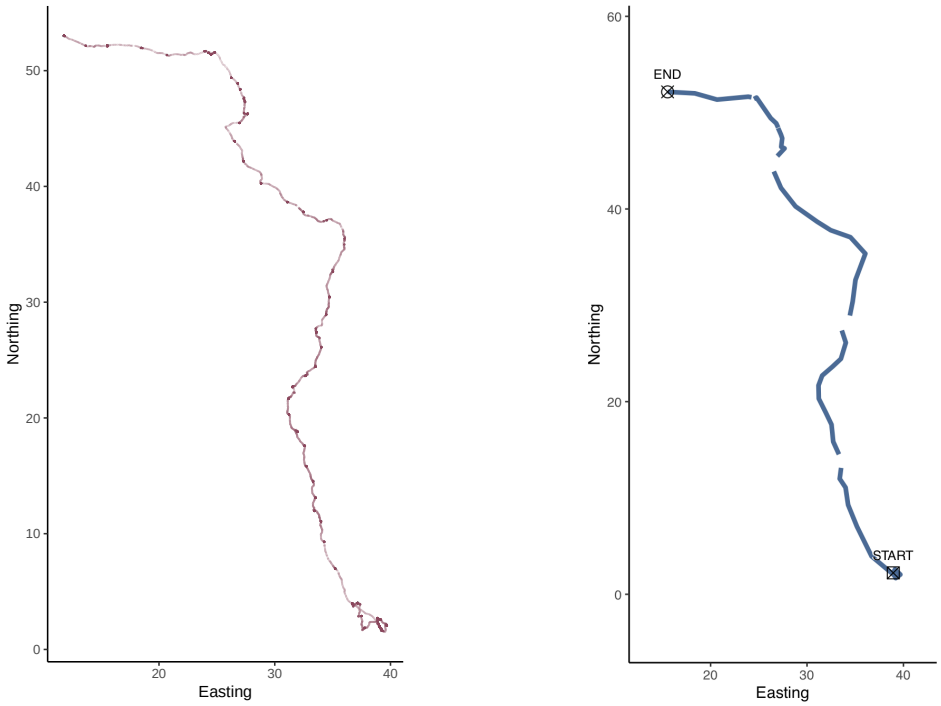

Figure 6.2: Left: Sample observed trajectory of a migrating white stork used in the experiments of this chapter. Right: The same trajectory after pre-processing, with the non-migratory parts removed, ready to be used as input (template) in the eRTG.

Contextual information. The context in each of the experiments (Section 6.3) is introduced as a background raster layer ("*BG*") either carrying a binary mask or continuous probabilities P_BG for the simulation area. P_BG can take values between 0 and 1, denoting no or full contextual attraction to the mover, respectively. The extent of the P_BG raster layer is larger than the expected movement area to ensure that there will be no location left without a probability value attached. The raster's size and resolution may affect the processing time of the simulation. Therefore, a sensible choice between the grid size, resolution and the average step length is advised. Once the raster layer is prepared the eRTG is calling the value for each current location in the Z function (Section 5.3), in order to account for its effect in the generation of the next step, in the form of a joint probability:

$$Z_{BG} = Z * P_{BG} \tag{6.2}$$

6.3 Experiments

6.3.1 Context-agnostic, free-space movement

The first simulation uses a single trajectory (template) as its input (Figure 6.2) and generates synthetic movement trajectories in free space. The eRTG algorithm bases the selection of the next step of the mover solely on intrinsic parameters, with no influence exerted by the environment in which the movement occurs.

Results. Figure 6.3 visualises the first 100 out of 1,000 generated synthetic trajectories (grey lines) based on the template trajectory (red line). The result was generated in free, unconstrained space and is therefore depicted without spatial context (left panel), and overlaid on a global map for geographic orientation (right panel). The majority of the simulated trajectories follow the route of the template trajectory, with some deviations. However, generated outlier trajectories that deviate strongly from the template trajectories are visible, too. It should be noted that according to the evaluation of the basic eRTG algorithm carried out in Section 5.4, all trajectories can be considered statistically similar to the input template trajectory, though they may deviate quite substantially, owing to the given ample time budget available to perform the migration. The density estimation of Figure 6.4 displays the highest values in the area surrounding the start and the end point, as expected since all the trajectories have these in common. At the same time, the density values decrease as they move away from these points, as each agent takes its random route (Figure 6.4, right), though a weak local concentration is noticeable roughly along the template trajectory. All density estimates were computed on identical raster templates by overlaying the simulated results over the raster and counting the trajectory intersections per raster cell.

Evaluation. Upon visual inspection, the generated trajectories seem sensible. Visual inspection revealed no significant dissimilarities between the generated trajectories and the original ones. Given that the movement takes place in free space without physical barriers or constraints, the expected outcome is a result close to the RTG (Figure 4.15) and the BBMM (Figure 2.3) types of probability surfaces. The hypothesis is initially corroborated by the density plot in Figure 6.4 and confirmed by Kolmogorov-Smirnov test results of Figure 6.5. The histogram plots and p-values suggest substantial similarity between the input and the simulated distributions in all four metrics: turning angle, step length, and their respective auto-differences.

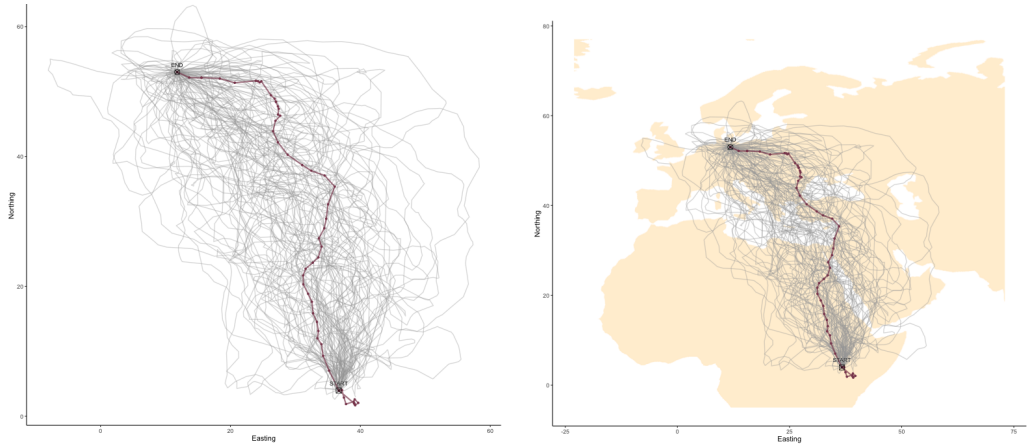

Figure 6.3: Left: Simulation of 100 white stork trajectories (grey lines) derived out of the template trajectory (red line) based on free-space movement. Right: The same simulated trajectories overlaid on a map of continent outlines for geographic orientation.

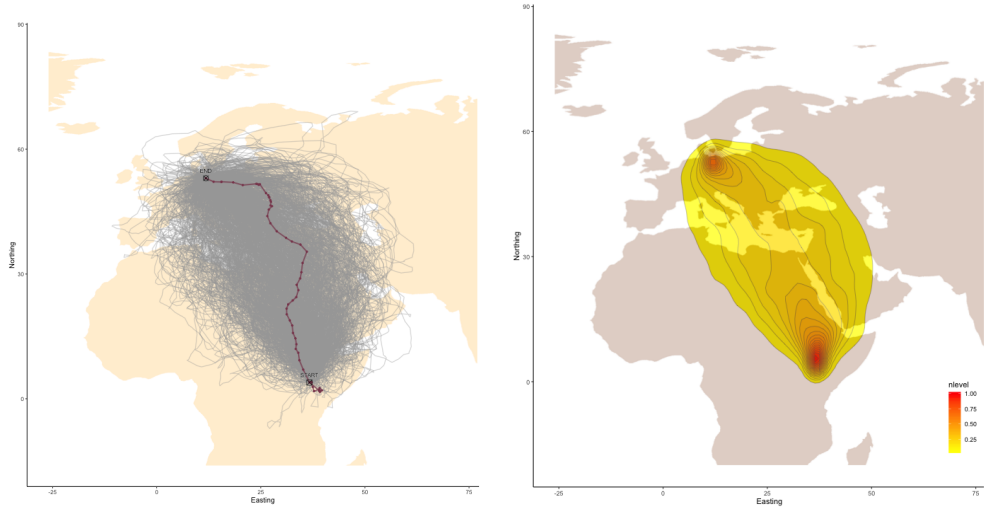

Figure 6.4: Simulation of white stork trajectories based on free-space movement. Left: The template input trajectory (red) and 1,000 generated trajectories (grey) are visualized as lines. Right: The respective line density estimation raster.

6.3.2 Context as an avoidance factor

Incorporating avoidance. Free movement of agents in space can only be partially useful for ecology, as many animals have strict restrictions in their movement repertoire. For example, fish do not commonly move over land, the tree frog can only jump 150 times its body height, and the white stork avoids crossing over large bodies of water. A binary attribute (reachable = 1, not reachable = 0) was introduced to the model to account for these physical limitations and an *avoidance layer* mapping the binary values was added. In the eRTG algorithm, the obstacle falls into the agent's perception precisely one step before the rule is enforced, leaving limited space for manoeuvre, as the movement parameters' distributions are in force.

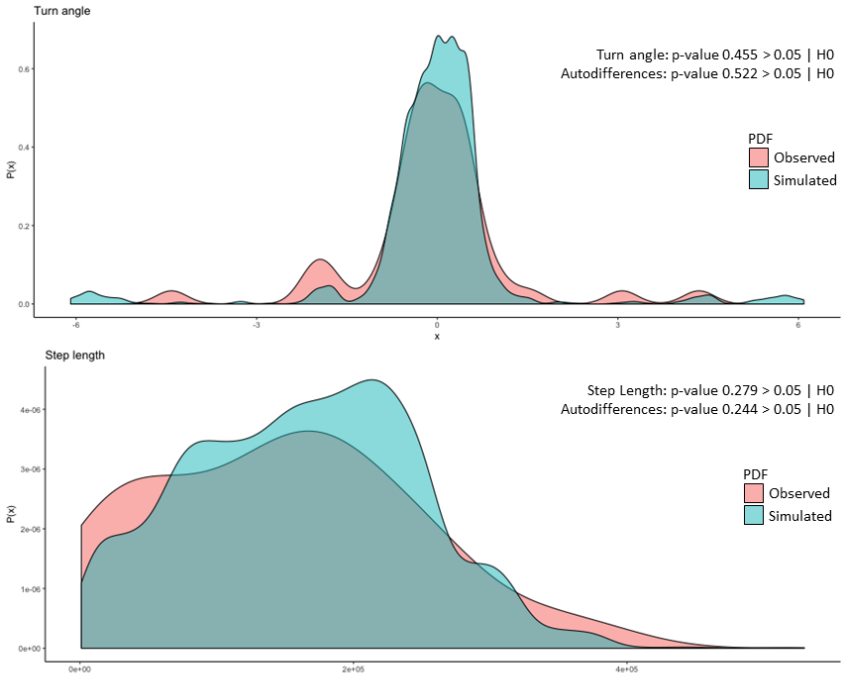

Figure 6.5: Histograms and Kolmogorov-Smirnov test results of the trajectory simulation based on free-space movement for the movement parameters turning angle and step length. The p-values do not show a significant difference between the distributions of observed and simulated data.

In cases where the agent has no option but to violate a given condition to continue its way, the trajectory is considered to have reached a dead-end and is discarded.

Large water bodies as an avoidance factor. As an example of avoidance behaviour, large water bodies are used. The global oceanic layer identifies the areas to be avoided in this experiment. (Figure 6.6). It describes earth surface types (land = 1, water = 0) and is available in the Natural Earth public domain map dataset[11] in a spatial resolution of 110 x 110 metres. The domain experts considered this resolution adequate given the tens of thousands of km the individuals covered in the present case study.

Results. Figure 6.7A depicts the trajectories generated (grey lines) out of the template trajectory (red line) given a single rule: *The mover must not perform a stop on a water body*. This requirement is fulfilled. The simulated trajectories do not cross arbitrarily large water bodies. When zooming into the results (Figure 6.7B), points where a mover marginally managed to jump over water bodies (green arrows) become visible. Locations where a mover had to redirect itself to remain compliant with the simulation rule are also shown (red arrows). Lastly, Figure 6.7C displays the same area of interest for the free-space movement simulation (seen in Figure

[11] `https://www.naturalearthdata.com/downloads/110m-physical-vectors/110m-ocean/`

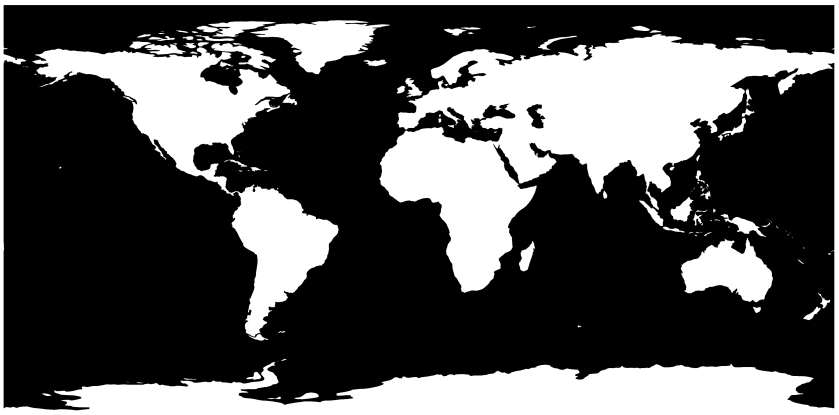

Figure 6.6: The oceanic data layer used for this study as an avoidance layer is publicly available at Natural Earth as a public domain map dataset[11].

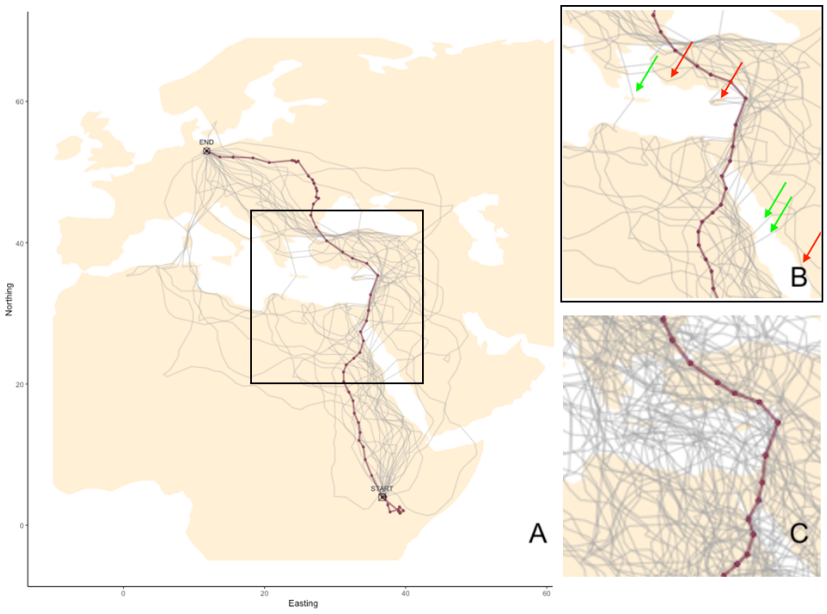

Figure 6.7: A: Trajectory simulations (grey lines) based on the input trajectory (red) and using water bodies as avoidance layer. B: Focus on decision-making points. The green arrows denote the movers that marginally managed to jump over water bodies. The red arrows indicate cases where a mover had to redirect itself to remain compliant with the avoidance constraint. C: For comparison, the result of the free-space movement simulation (cf. Figure 6.3) within the same area of interest.

6.3) for direct comparison of the simulation approaches. Comparing between the constrained (B) and the unconstrained (C) trajectory simulation clearly shows that the constrained movement has been channelled to stay on land, thus exhibiting less lateral deviation from the template trajectory.

The density surface in Figure 6.8 is estimated according to the results seen in Figure 6.7A. The pattern indicates a preferred route due to avoidance of water bodies, manifested in the higher density values, leading from Kenya, Ethiopia, Sudan, Egypt, Israel, Turkey, Ukraine to Poland. The Mediterranean Sea has

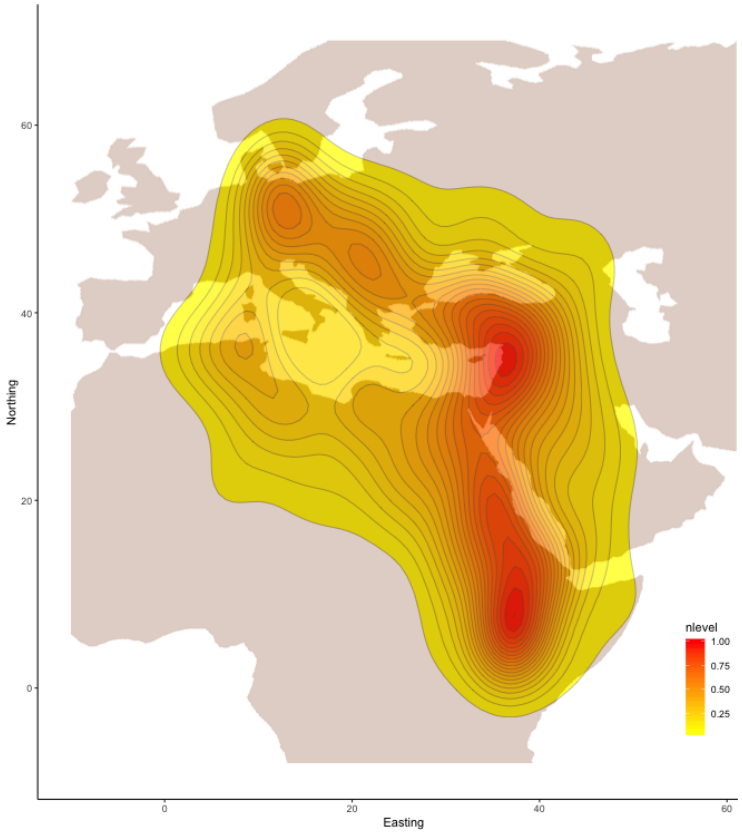

Figure 6.8: Density estimation of the trajectories generated based on the avoidance factor: large water bodies.

the lowest density values. However, some Mediterranean islands such as Sardinia and Corsica serve as stepping stones to cross the Mediterranean Sea and have higher density values. Additionally, it is worth noting that both the known northern migratory corridor (Poland—Ukraine—Turkey) and the southern corridor (Turkey—Syria—Israel—Egypt—Sudan—Ethiopia—Kenya) appeared clearly in the results. Thus, the results suggest that the avoidance of water is one of the key contextual factors constraining the migratory movement of white storks.

Evaluation. Regardless of introducing additional bias in the trajectory generation (avoidance of water bodies), the similarity between the input/output trajectories remained high. The Kolmogorov-Smirnov test suggests no significant difference when comparing the input and output movement parameter distributions against each other (Figure 6.9).

6.3.3 Context as an attraction factor

The limited context-awareness of the agent introduced in the preceding section allows for basic obstacle avoidance without capturing the agent-context interaction

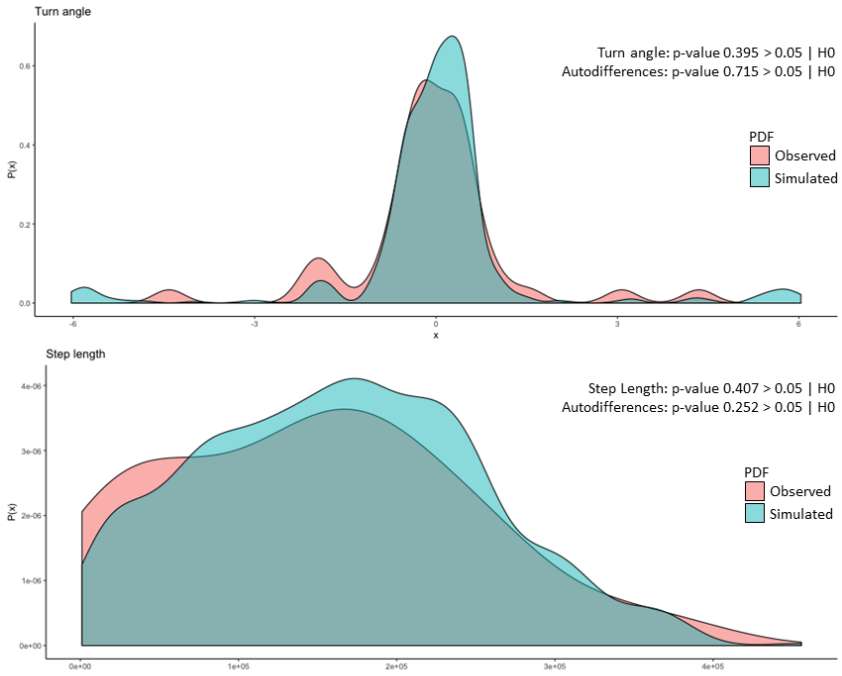

Figure 6.9: Histograms and Kolmogorov-Smirnov test results of the trajectory simulations based on the avoidance layer of large water bodies for the metrics of turning angle, step length and their respective auto-differences.

adequately, since it does not account for the mover's preference, that is, its attraction to certain contextual factors. The preference mechanism in the eRTG is based on user-defined attraction raster layers carrying a ratio-scaled value ranging from 0 to 1. The attraction score reflects the inclination of the mover to move towards a specific point (cell) of the layer, with 0 being unattractive (akin to the avoidance factor seen in Section 6.3.2) and 1 denoting maximal attraction. The attraction score (a) of each cell is factored into the calculation of the total probability Z (Equation 6.3):

$$Z_a = Z * a \tag{6.3}$$

Below, the effect of adding attraction factors is demonstrated on the example of two contextual variables of the physical environment: greenness index and drainage density.

Greenness index

Greenness as an indicator of the live green vegetation available to an animal has been identified as one of the indicators of potential foraging sites [208], both for herbivores and carnivores feeding on herbivores. In practice, the greenness of the vegetation is routinely represented by the Normalised Difference Vegetation Index

(NDVI). This experiment focuses on migration behaviour, which might be affected by vegetation availability. Thus, we are interested in evaluating the utility of the eRTG in testing such a hypothesis.

Attraction layer. The vegetation layer used for mapping the attraction layer is called the NDVI greenness. It is calculated using data collected over the study period (2012) by the moderate resolution imaging spectroradiometer (MODIS) aboard NASA's Terra satellite [216]. NDVI is a robust, empirical measure of vegetation activity on land surfaces, designed to enhance the vegetation signal from measured spectral responses by combining multiple wavebands, often in the red (0.6–0.7 mm) and near-infrared wavelengths (0.7–1.1 mm). The healthier and larger the leaves on a plant, the more near-infrared energy is reflected and transmitted to the satellite sensor.

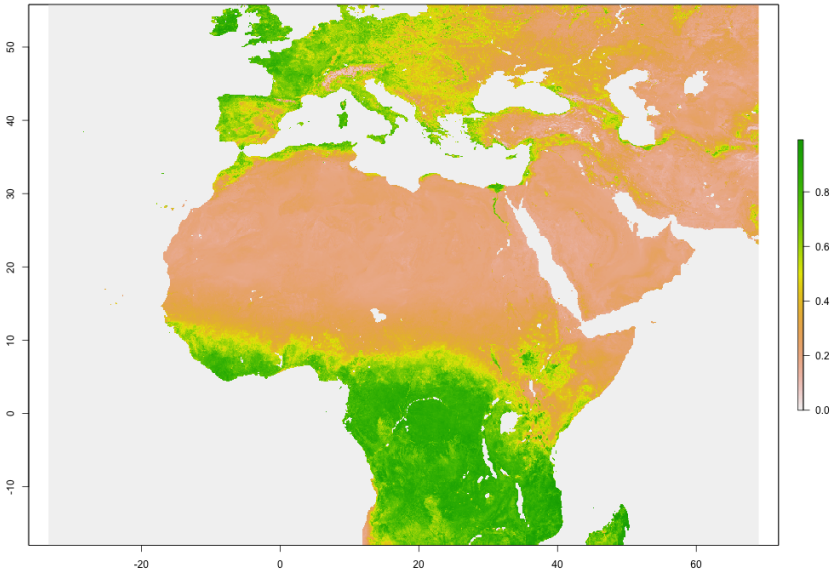

Figure 6.10: Overview of the average Normalised Difference Vegetation Index (NDVI) from February to May 2012. All values less than 0.02 have been set to 0 to remove scarce vegetation footprints.

The accuracy of the NDVI is susceptible to factors beyond the scope of this study, such as the fluctuation in solar irradiance, atmospheric conditions, or the characteristics of the available canopy. For the purposes of studies like this one, however, the NDVI can provide a solid indication of *vegetationless* land cover (values of 0–0.2), and a good enough indication of the intensity of the vegetation (0.20–1). The first cluster was assigned the value of 0 in the preference layer, the second a relative probability reflecting the NDVI value: the higher the preference value, the more probable it is for the specific cell to be selected by the mover. Figure 6.10 shows the average NDVI covering the three months of the study data. The resolution of this raster layer follows the water bodies template.

Results. Figure 6.11 illustrates the 25 first context-aware trajectories generated with NDVI as the attraction factor. The simulated trajectories do not show a distinct preferred migration pattern but vary significantly, with western routes via Tunisia and the Mediterranean islands of Corsica and Sardinia; central routes via Egypt, Israel, Turkey and the Balkans; and eastern routes via Yemen, Saudi Arabia, Iraq, Turkey and Ukraine. In this case, smaller patches of vegetation are used as stepping stones in regions with otherwise little to no vegetation.

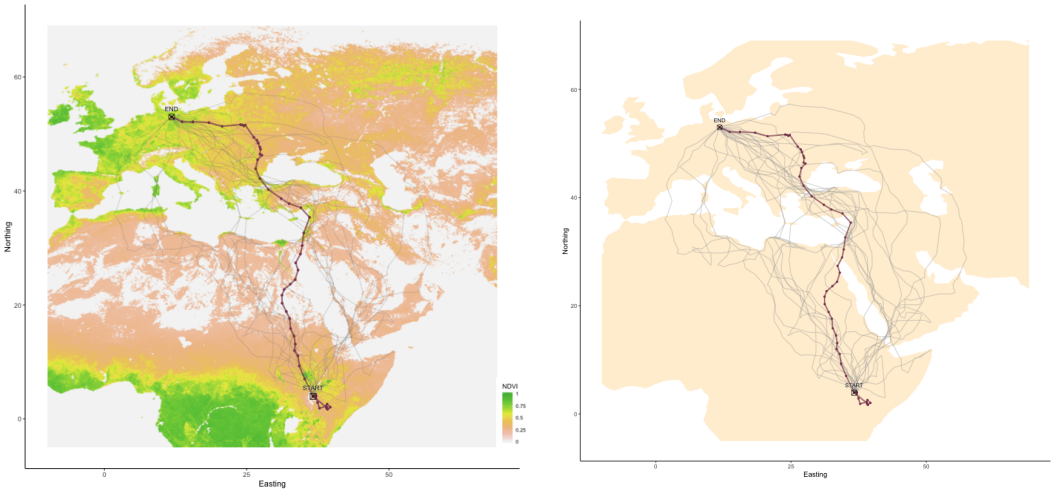

Figure 6.11: Trajectory simulations (grey lines) based on the input trajectory (red) with the NDVI as the attraction layer. Results are overlaid on the NDVI (left) and on a map with continent outlines (right).

The density estimation of these results reveals two preferred simulation routes: from Kenya, Yemen and Saudi Arabia towards Turkey, the Balkans and Poland; and from Kenya and Libya to Greece, the Balkans and Poland (Figure 6.12).

Evaluation. A close visual inspection of the individual trajectories, together with the results of the Kolmogorov-Smirnov test (Figure 6.13), suggests high confidence on the credibility of the results, as the simulated trajectories maintain high similarity in movement metrics compared to the observed ones.

Drainage density

River deltas are a popular stopover amongst migratory birds. The increased biomass-to-water ratio found in these areas results from the substantial amount of nutrients being transferred by the rivers to shallow waters, plus high sun exposure. However, do birds follow the river all the way to the delta? Moreover, if so, do they also use it for navigation on the way back? Such questions led to using the river network as an attraction layer for the final experiment of this chapter.

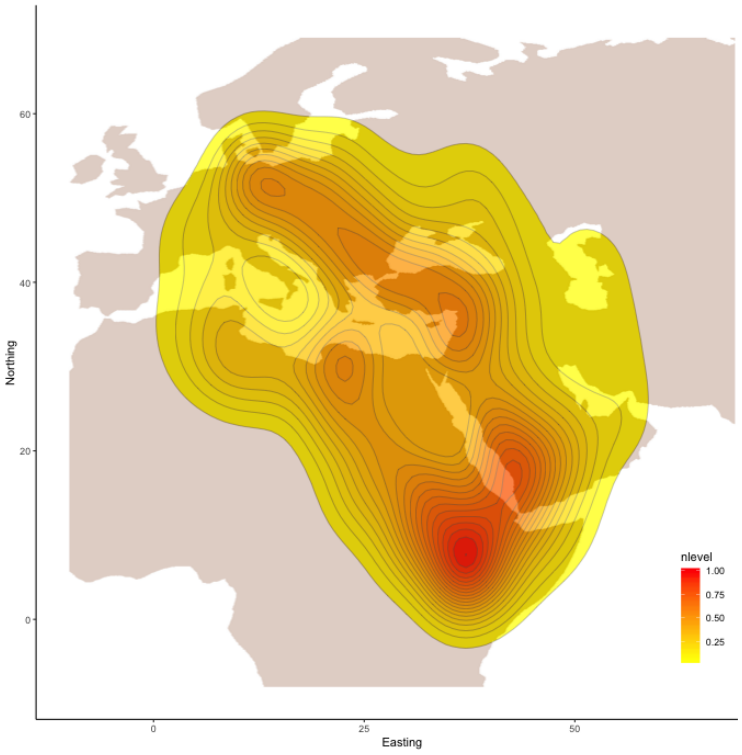

Figure 6.12: Density estimation of the trajectories generated based on the NDVI as the attraction factor.

Attraction layer. The river network for this experiment is represented by a critical drainage density map (Figure 6.14 left), derived from a large-scale river delineation process, conditioned to digital elevation and hydrological models in a study conducted by Schneider et al. [217]. In that study, the authors divided the calculated drainage density layer (as the sum of the total river length) inside a cell by the cell's land area (excluding the ocean fraction) [217]. The resolution of this raster layer follows the water bodies template.

Results. Figure 6.14 (right) presents the trajectories of 100 'river-attracted' agents moving from the given start to the given end point. The saltwater rule is still suggested (given the low drainage density values near and in the sea), but not completely forbidden as in the first experiment (6.3.1).

Evaluation. The simulation results with the river network as attraction factor (Figure 6.14 right) present no apparent bias, though the effect of the water body rule is apparent. The Kolmogorov-Smirnov test (see Figure 6.15) returns strongly positive results, with the p-values on par with the water avoidance experiment (Section 6.3.2. In both the turning angle and the step length distribution, no significant geometric bias is detected. The density presented in Figure 6.16 identifies the com-

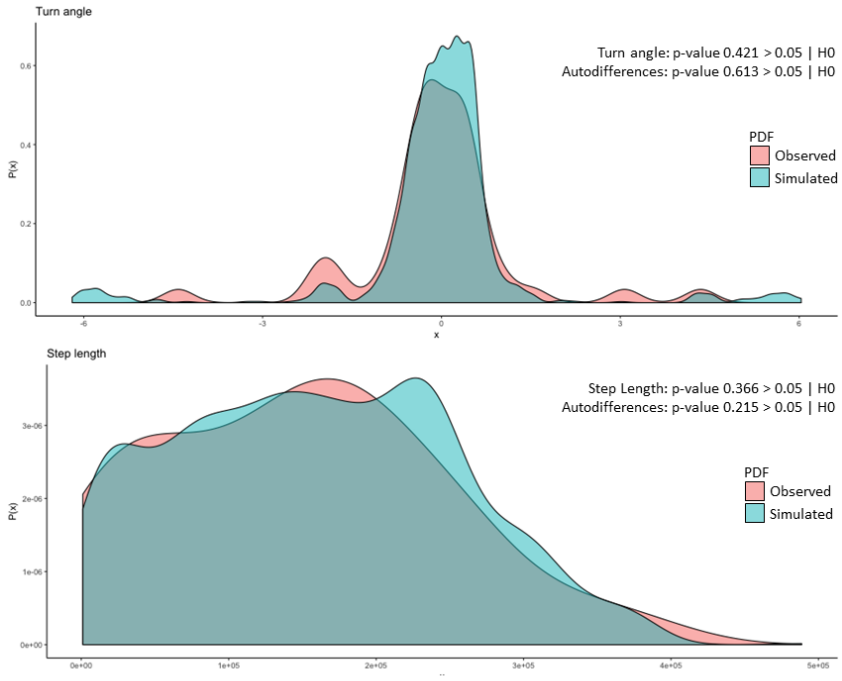

Figure 6.13: Historgrams and Kolmogorov-Smirnov test results of the trajectory simulation based on the NDVI as the attraction factor for the metrics of turning angle and step length, respectively. The p-values do not show a significant difference between the distributions of observed and simulated data.

mon Poland—Ukraine—Turkey—Syria—Israel—Egypt—Sudan—Ethiopia—Kenya passage.

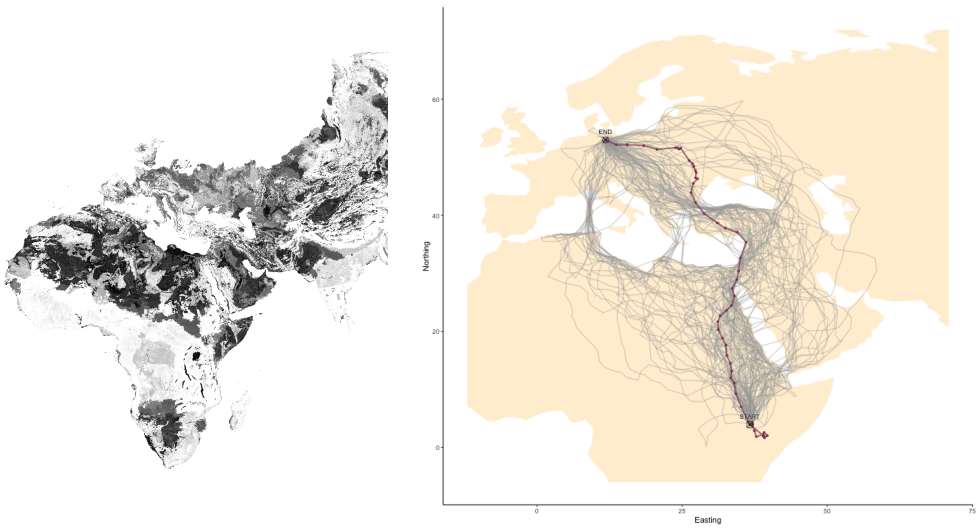

Figure 6.14: Left: The drainage density layer as calculated by Schneider et al. (2017) [217]. The darker the shading, the higher the river length per unit land area. Right: Trajectory simulations (grey lines) based on the input trajectory (red) and the river network as the attraction layer.

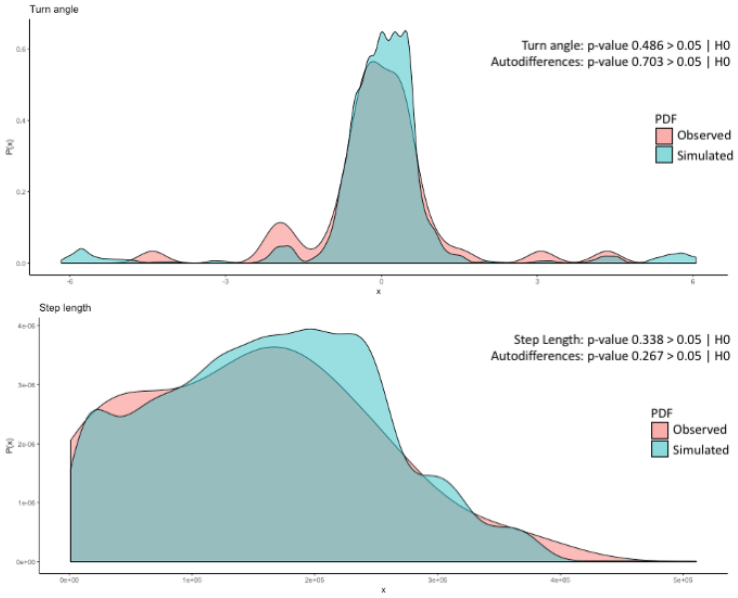

Figure 6.15: Histograms and Kolmogorov-Smirnov test results of the trajectory simulation based on the river network as the attraction factor for the metrics of turning angle and step length, respectively. The p-values do not show a significant difference between the distributions of observed and simulated data.

6.3.4 Comparing the results

Comparing the four simulation scenarios (free-space movement, water body avoidance, greenness index attraction, drainage density attraction) could give the user insightful information regarding which contextual factor is more likely to affect the birds' migration pathway selection. In order to perform this comparison, a 100 x 100 km resolution grid was created and used as a reference, as shown in Figure 6.17(i). The frequency counts per grid cell were then obtained by intersecting the created trajectories of each scenario with the reference grid. Each intersection of a cell with the original trajectory increases that cell's frequency by 1, irrespective of the length of the line segment in the intersection. For every two consecutive locations, all the cells along a straight-line connection were considered as *visited*. Additionally, the start and end point were the only two points excluded from this exercise as all trajectories have these two in common. The frequency profiles for all simulation scenarios are depicted in Figure 6.17(ii).

In this type of profile, a higher number of intersections hints at a stronger positive correlation between the contextual factor and the migratory behaviour of the bird. The unconstrained profile acts as a baseline indicating the average number of intersections without considering the context. Anything lower than the reference profile hints at a negative correlation between the specific contextual factor and the migration behaviour of the bird.

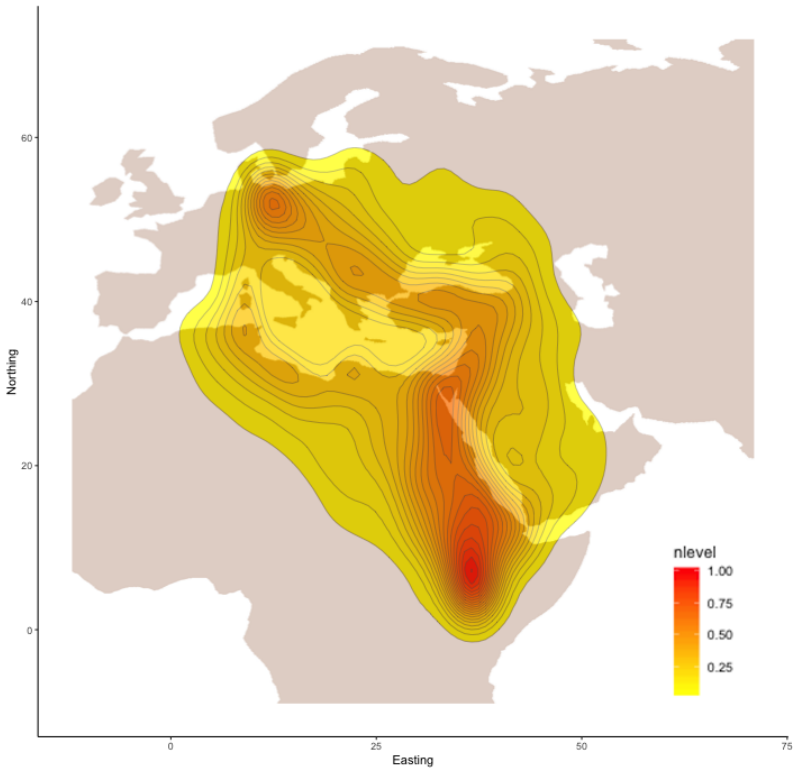

Figure 6.16: Density estimation of the trajectories generated based on the river network as the attraction layer.

Given that the trajectories were generated using a CRW, the absolute number of the intersections is not expected to provide conclusive ecological insight. However, the comparison of the frequency count profiles, when seen on the map, provides multiple starting points for discussion. Starting with the overall view: all context-aware results picked up parts of the movement preference of the stork (peaks), in contrast to the unconstrained walk, which shows no clear tendency towards the original trajectory, apart from the area around the two endpoints. Considering that the unconstrained walk is closer to the RW diffusion principle than the rest of the simulation scenarios, a uniform ($\pm SD$) likelihood of cell selection was expected.

The water body result intuitively peaks at the locations where large water bodies (sea) are within the vicinity. By decreasing the likelihood of the sea cells to be selected during the trajectory generation process, the likelihood of land cells increases, explaining the two peaks seen in the profile graph of Figure 6.17(ii).

The river network profile peaks where rivers are mostly present and reaches a maximum value over the river Nile, comparably to the water bodies case. Simultaneously, some unexpected valleys appear between the 20th and 30th step segment which require further exploration (avoidance behaviour, lack of rivers, alternative dominant attraction factor) before attempting an explanation.

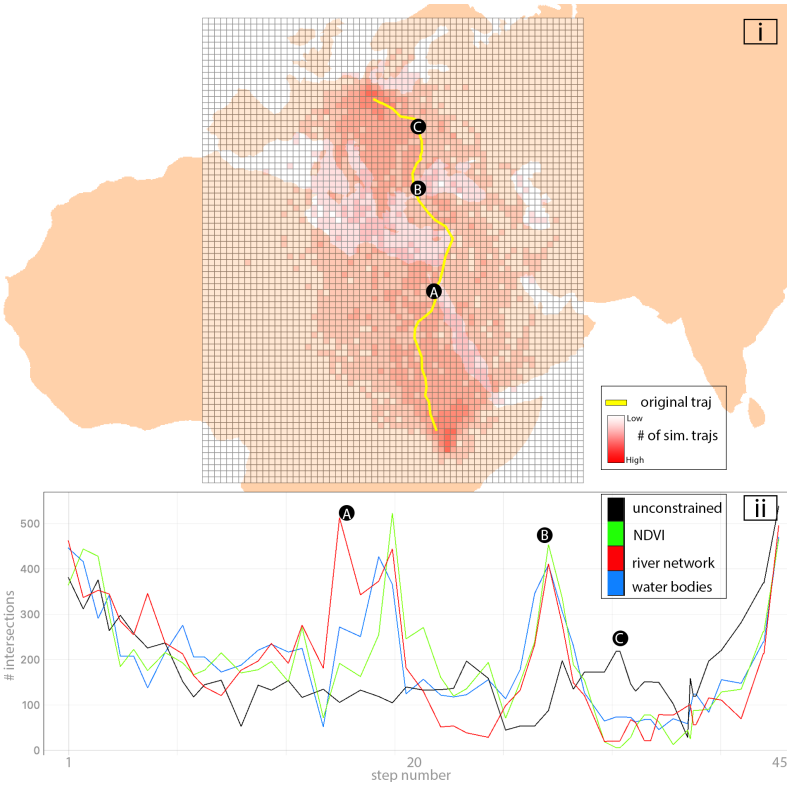

Figure 6.17: (i) Grid with frequency counts per grid cell for the NDVI scenario, with the template trajectory overlaid in yellow. (ii) Profiles of the original trajectory extracted from the frequency grids for each simulation scenario. Labels A, B, C denote corresponding locations in the map and the profile view, respectively.

In the NDVI result, which is shown in the map view of Figure 6.17(i), we see a similar picture in the profile graphs as for the large water bodies and the river network: peaks occur in locations where the attraction factor is stronger (in this case higher NDVI index values). Interestingly, the frequency count drops to almost 0 in an area where vegetation is significantly present (between point C and the end of the trajectory), as seen in Figure 6.17(ii).

Taking a closer look at the area of Southern Egypt – Northern Sudan, denoted by the letter A in Figure 6.17(i), the river drainage network around Lake Nasser, one of the largest artificial lakes in the world seems to affect the migration of birds. In Istanbul (location B), the migratory bird needs to pass between continents in order to continue its migration, so naturally, the water body avoidance factor scores a higher frequency than before, together with the NDVI and river network factors, both of which are quite rich in the specific area. At location C, over Romanian and Ukrainian territory, the river network and NDVI are more homogeneous, and there is no large water body in the vicinity; therefore, none of the three factors affected the navigation of the bird noticeably, giving the edge to the simple unconstrained behaviour scenario. Such observations suggest further research on (a) the creation of an adaptive/hybrid contextual layer and (b) the addition of more contextual

factors that better correlate with the storks' movement (especially in areas where the currently used factors did not sufficiently control the movement, such as in the area around location C).

7 Discussion

> *PICARD: That's all?*
> *DATA: Yes, sir.*
> *PICARD: Run it again.*
>
> **— scene from "Star Trek"**
> (Next Generation)

7.1 Placing this thesis in context

Now more than ever, the democratisation of technology has equipped ecologists with the means to collect unprecedented geospatial and contextual data. Cloud-deployed analytical solutions enable a single university node to handle an infinite amount of data on a broad range of themes. Adding to this, the richness of open-source software suites offers sophisticated statistical tools with minimum scripting requirements. At the same time, this seems insufficient for halting the biodiversity decline rate:

"The overwhelming evidence of Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services Global Assessment (IPBES), from a wide range of different fields of knowledge, presents an ominous picture," said IPBES Chair, Sir Robert Watson [12]. *"The health of ecosystems on which we and all other species depend is deteriorating more rapidly than ever. We are eroding the very foundations of our economies, livelihoods, food security, health and quality of life worldwide."*

The oxymoron that more data and technology accompany less success in conservation can be attributed to points anywhere between the data gathering and the implementation of the conversation efforts. However, it is fair to assume that policymakers must receive ecologically informed advice from domain experts, which requires developing a deep ecological understanding of each domain and species.

The ecological knowledge focuses on understanding patterns found in nature, and this requires a combination of inductive (hard) and deductive (soft) inference

[12] https://www.un.org/sustainabledevelopment/blog/2019/05/nature-decline-unprecedented-report/

methods for knowledge generation. The first is evidence-based: given observed data, a generic model is constructed that expresses the rule and handles the uncertainties, whereas the second is based on falsifying as many hypotheses as possible until a valid argument or conclusion is reached.

This thesis strives to directly act as an enabler of the soft inference, helping to answer questions such as: Is the movement of an animal sensitive to a specific change in its environment? and Where could the animal have been, when not observed? However, the thesis should also indirectly contribute to strong inference by helping to form naive predictions during the initial exploration of a phenomenon that may not be a theory yet to test [218].

It should be clearly stated that this thesis has not produced any ecological insight, as the hypotheses tested served only for evaluation and demonstration purposes. Indeed, in the context of an Empirical Random Trajectory Generator (eRTG), various simulation concepts, paradigms and methods were explored, tested and utilised to extend further the use of the correlated, biased, reinforced random walk.

The core differentiation between eRTG and the orchestra of RW-based algorithms is that the former captures elements of the unknown processes that force the mover to move towards their destination. Navigational efficiency has always been a predictive, forward-looking exercise in the RW models; eRTG introduces a deductive, backward-looking element, encouraging the user to test hypotheses on locomotion, context interaction, taxis, etc. and observe the emerging movement and space-use patterns on an individual and population level.

The result was a movement ecology testbed, based on the distribution statistics of movement variables and subject to a *light, destination-based directional (i.e., gravitational) bias*. The intuitive use of the testbed enables the ecologist to control a movement experiment without necessarily needing to own the respective simulation, geographical or statistical experience.

By using the example of the white stork migration seen in Chapter 6, which relies on just a few trajectories of migrating individuals, the researcher can explore how an entire population of similar storks might migrate, extract the densities of space-use and compare the results to the densities created by applying a contextual factor, as seen, for instance, with the vegetation and river index. Based on the identified differences, new hypotheses can be formulated, either by including new factors or by forming a different meaningful combination of the factors currently under study to explain the results of the respective study.

In the following section, the outcomes of this thesis regarding the research objectives are discussed.

7.2 Discussing the research objectives

7.2.1 Conceptualising the simulation framework

Objective 1: Develop an integrated simulation approach that allows covering all components of the conceptual framework of [55], and their interactions.

This thesis introduced a novel, hybrid conceptual simulation framework (Chapter 3) for developing a range of movement modelling methods in close relationship to the respective ecological paradigm by Nathan et al. [55], emphasising the direct relationship between the elements of the two. As shown in Figure 7.1, the proposed conceptual simulation framework picks up the key components of the paradigm by Nathan et al. and integrates these into the simulation workflow. Thus, using the paradigm of [55] as a reference guide, it was possible to build the theoretical foundation used in the Random Trajectory Generator (RTG) of Chapter 4 and the Empirical Random Trajectory Generator (eRTG) of Chapter 5 while limiting the *ad-hoc assumptions* [219] used in building these simulators.

From the perspective of the simulation architecture used, the proposed simulation framework is hybrid since different processes are delegated to individual simulation models relying on different paradigms (Figure 3.6). Agent-Based Modelling (ABM) is the core orchestrator of the framework, triggering the other modelling components when necessary and performing the detailed simulation on each agent. The Discrete Event Simulations (DES) model focuses more on passive contextual changes. System Dynamics (SD) run on two levels: first as sub-models, adjusting the needs and behaviour of each agent and second as a population balance control mechanism, which decides the age of each new agent, therefore indirectly affecting its capabilities.

By separating the static from the dynamic part of the simulation framework and by walking through cases of increasing complexity in Section 3.3, insights into the three models of simulations became apparent, as discussed below.

Paradigm 1: System Dynamics. SD describes, relatively intuitively, any flow-based structure, be it on a micro-scale such as the motivation, energy consumption, and general preference of an agent, or a macro-scale such as population dynamics. The result of an SD simulation connects a parameter to a systemic fluctuating behaviour developed over time, instead of delivering a single forecast value for a given point in time. For example, more predators lead to more kills, which eventually means less food for the predators, leading to a decrease in their population.

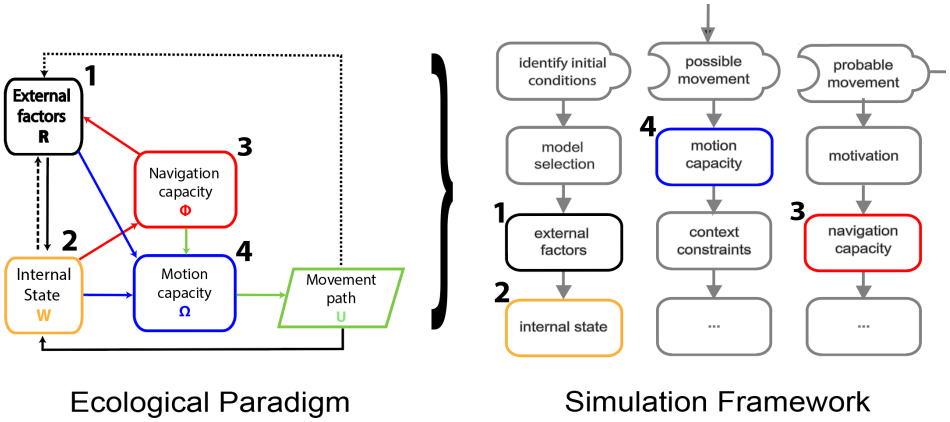

Figure 7.1: Mapping of the ecological paradigm of [55] (left; cf. Figure 2.4) to the static part of the simulation framework (right; cf. Figure 3.5). Numbers shown in the two sub-figures serve to relate the essential components of the model of [55] to the conceptual simulation framework proposed in this thesis.

Paradigm 2: Discrete Event Simulations. DES allows for a relatively strong decision-making process, as it can be mathematically validated, and it gives the powerful ability to answer "what if" questions, allowing even a non-domain expert to follow the decision process. At the same time, adding many rule-based decision layers turns this advantage into a rather complex setup that even an expert might find challenging to connect to a physical/natural process directly.

Paradigm 3: Agent-Based Modelling. ABM has two main advantages over the other techniques. First, it inherently provides a 'natural' description of a system involving animate movers, making it very easy for the user to relate to. Perhaps even more importantly, it allows unobserved patterns to emerge, leading to ecological insights or facilitating the development of hypotheses.

Regarding possible synergies among the simulation model paradigms, the following observations can be made:

- There is a lack of a single best answer for all ecological questions regarding selecting the appropriate simulation framework. Therefore, maintaining flexibility within the framework is important, leading to our hybrid, modular approach.
- All types of modelling can complete any task, though some introduce more complexity in the implementation than others for the same result, favouring a hybrid approach.
- Evaluating a hybrid model is statistically more challenging than evaluating a single-breed one due to the difficulty of isolating the effects of the individual components. Also, the combined uncertainty inherent to the simulation results is different from the sum of individual uncertainties.

The proposed hybrid simulation framework (Figure 3.6) aims to be utilised in the intersection of Ecology, Geography, and Information Science. By strengthening interdisciplinary collaboration, ideas can be tested to reveal key processes underlying movement and touch upon evolutionary, and life history strategy questions [136].

7.2.2 Creating possible trajectories with known endpoints

Objective 2: *Develop a simulation approach that allows to connect two known locations, possibly separated by a large spatial and temporal distance, by generating ecologically possible trajectories which comply with the physical limitations of the mover and carry the least statistical bias possible.*

Chapter 4 introduced the Random Trajectory Generator (RTG), a novel algorithm [190] to create a random trajectory connecting two endpoints, given a specific physical capacity (i.e., maximum speed) and time budget, and to perform this task both on the Euclidean plane and on the sphere. Various existing approaches for creating trajectories that comply with the technical requirements were reviewed prior to the development of the proposed model. The Biased Correlated Random Walk (BCRW), Space-Time Prism (STP) and Brownian Bridges Model of Movement (BBMM) were found to be the most suitable candidates for the purposes of Research Objective 2. Therefore, we used these approaches as a basis for the methodological part of the RTG.

The most important advantage of the RTG over all other types of random walks is handling predefined starting and ending points, given a specific time budget. This ability instantly increases the number of applications this algorithm can facilitate as it can create a connection between any two points, such as is needed in gap-filling tasks, when for example, the quality of the tracking data is low. In comparison, the STP addresses the known endpoints and time budget [100], but it does not provide any information for the space-use inside it. Thus, the probability density surface inside the STP is uniform, meaning that any point within the prism has the same chance of being visited while moving from the start to the destination, an assumption that is not optimal in cases where the focus is on the interaction of a species with its environment [220, 221]. The BBMM [117, 118], on the other hand, does provide the user with a density surface as a result, but it does not create an individual trajectory, and therefore benefits less studies that focus on dispersal, colonisation [41], and interaction among individuals.

The RTG is limited in terms of its degree of extensibility. In particular, the facilities available to condition the movement model are still limited to only two user-defined

parameters: the (maximum) speed of the mover and the available time budget to perform the movement. Thus, the simulation is restricted to generating *possible* trajectories rather than *probable* trajectories, conditioned by the probabilities of speed and directional variations of the movement of actual individuals belonging to a particular species. This additional requirement (engrained in Research Objective 3) thus later led to the development of the eRTG in Chapter 5.

Nevertheless, despite its limitations, we believe that the RTG can find many suitable applications in practice. Typical cases for the usage of the RTG are any cases where two points in space and time (i.e., an origin and a destination) are all the researcher has at hand, and important decisions must be made efficiently.

An example of such cases where the RTG is a perfectly suitable solution is illustrated in Figure 7.2, which uses bird ringing data, typically only two locations, an origin (the ringing location) and a destination (the recovery location) are known. In this figure, the potential of avian influenza spread by two specific duck species is represented by straight lines connecting ringing and recovery locations of each duck. Ringing, in this case, is the first capture where an animal found was virus-free, and the recovery is the second capture where the animal was tested positive for the virus. Using the RTG, in this case, could support any counter-influenza measures and decisions by creating a more realistic probability map, given that the straight lines used in Figure 7.2 substantially misrepresent the actual movement of the birds between the two known endpoints.

In a second example, Figure 7.3 lists a wealth of unique historic ringing observations that remain only partially utilised to date. With no information in-between the observation points, it is challenging to understand individuals' space-use, let alone to draw conclusions on a population level and thus develop a successful conservation plan. Such cases could benefit from the RTG, as it enables delivering trajectories connecting two spatio-temporal points with the least bias possible.

In related work, Kuijpers and Technitis [223] have developed a method to compute the so-called Potential Path Area (PPA) on the sphere, hence determining the potential locations that could have possibly been visited by a mover in large-scale movement on the surface of the Earth. Using the RTG, for which a spherical solution also exists, individual trajectories and trajectory density surfaces could be generated within the spherical PPA. The method of Kuijpers and Technitis can be applied to continental and intercontinental scale movement, such as bird migration or the generation of the paths of adrift containers on the oceans.

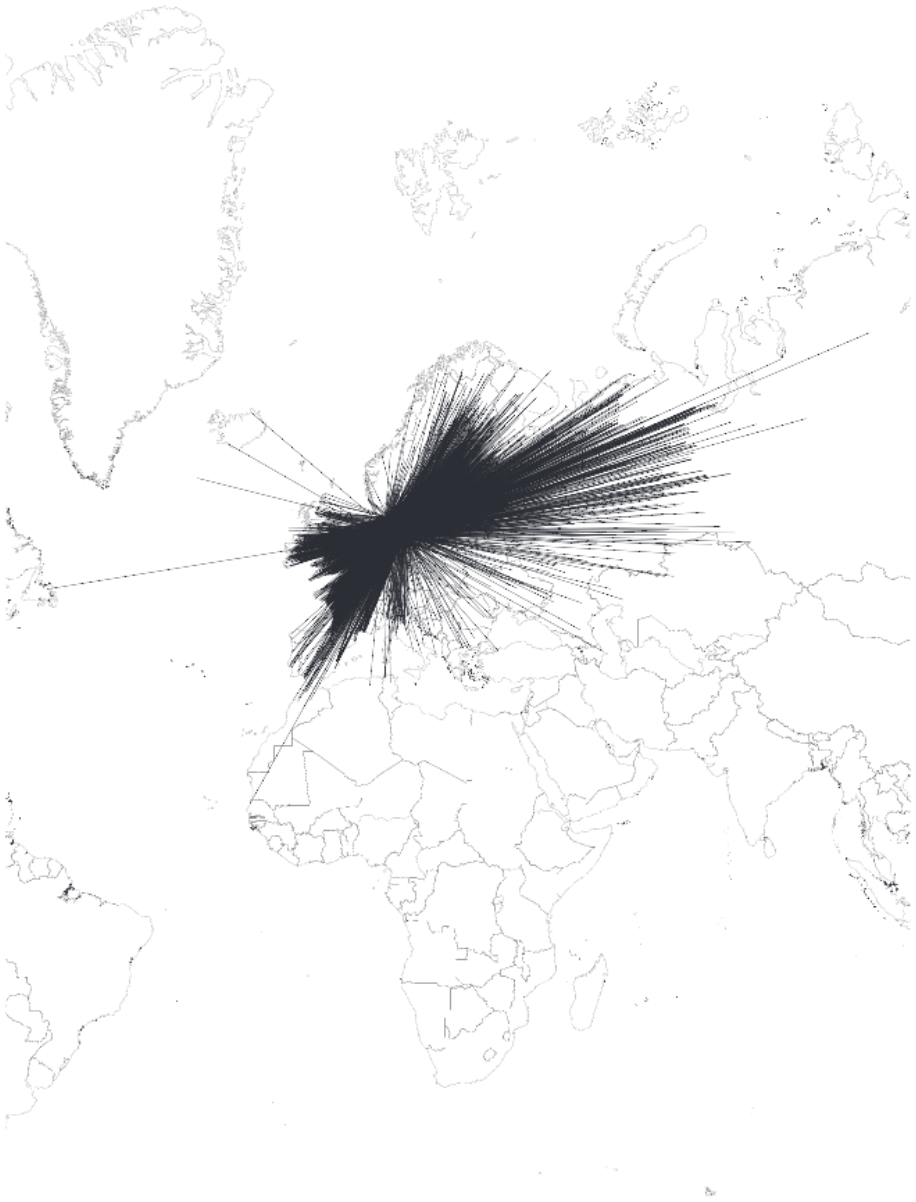

Figure 7.2: Recoveries of individuals belonging to two duck species (*Anas* and *Athya* in a combined data set) ringed in Denmark to show the potential of these species for spreading avian influenza. The lines connect the ringing and recovery locations, respectively [65].

Finally, the RTG algorithm has also been implemented in MoveTk[13], a C++ library offering various functions of computational movement analysis. The library has been developed as part of a collaboration between HERE Technologies, Eindhoven University of Technology and Utrecht University under the Commit2Data program.

[13] `https://github.com/heremaps/movetk`

Scheme code	Country	Ringing centre	1900–19	1920–39	1940–59	1960–79	1980–99	2000 onwards	Total	Percent
BGS[a,b]	Bulgaria	Sofia						45	45	0.00
BLB	Belgium	Brussels		4222	13 106	49 048	48 817	46 187	161 380	4.07
BYM[b]	Belarus	Minsk						39	39	0.00
CIJ	Channel Islands	Jersey		1	402	3713	3841	7326	15 283	0.39
CYC	Cyprus	Nicosia					2	17	19	0.00
CZP[a]	Czech Republic	Praha	163	1122	7133	13 775	27 348	25 415	74 956	1.89
DEH[a]	Germany	Hiddensee				35 100	100 928	166 932	302 960	7.63
DER	Germany	Radolfzell	33	4566	9648	21 474	13 091	39 522	88 334	2.23
DEW[a]	Germany	Wilhelmshaven	4	9642	15 282	48 727	42 833	65 588	182 076	4.59
DKC	Denmark	Copenhagen	579	8425	10 737	58 883	52 871	16 784	148 279	3.74
ESA	Spain	San Sebastián			458	798	223	894	2373	0.06
ESI	Spain	Madrid			38	4129	23 474	30 215	57 856	1.46
ETM	Estonia	Matsalu				1143	3499	8840	13 482	0.34
FRP	France	Paris		19	210	2294	9361	16 094	27 978	0.70
GBT	Britain and Ireland	Thetford	2069	10 404	46 235	234 249	260 687	167 971	721 615	18.18
GRA	Greece	Athens					184	134	318	0.01
HES	Switzerland	Sempach	22	1781	6096	19 145	19 000	30 720	76 764	1.93
HGB	Hungary	Budapest	531	1501	1062	1250	11 564	25 240	41 148	1.04
HRZ	Croatia	Zagreb	33	93	988	1505	2269	11 479	16 367	0.41
IAB[a]	Italy	Bologna		277	734	7718	11 854	13 747	34 330	0.87
ILT[b]	Israel	Tel Aviv				108	230	509	847	0.02
ISR	Iceland	Reykjavik		1	7	81	190	172	451	0.01
LIK	Lithuania	Kaunas	227	86	326	4187	20 895	13 147	38 868	0.98
LVR	Latvia	Riga		255	919	7056	5888	311	14 429	0.36
MLV[a,b]	Malta	Valetta					7975	14 798	22 593	0.56
NAW	USA and Canada	Washington		3	5	105	164	107	384	0.01
NLA[a,b]	Netherlands	Arnhem	193	1329	6325	77 573	293 066	495 930	874 416	22.03
NOS[a]	Norway	Stavanger		72	951	2110	15 777	8049	26 959	0.68
PLG[a]	Poland	Gdańsk		406	366	6218	73 557	79 864	160 411	4.04
POL	Portugal	Lisbon					1498	908	2406	0.06
RSB[a]	Serbia	Belgrade					81	1174	1255	0.03
RUM	Russian Federation	Moscow		1834	7125	9348	10 364	1738	30 409	0.77
SFH[a]	Finland	Helsinki	242	2766	4231	83 519	227 580	337 082	655 420	16.52
SKB[a]	Slovakia	Bratislava				53	524	1825	2402	0.06
SVS	Sweden	Stockholm	57	5524	17 720	47 525	54 774	44 388	169 988	4.28
TUA	Turkey	Ankara					23	313	336	0.01
UKK	Ukraine	Kiev					60	1305	1365	0.03
Total			4153	54 329	150 104	740 834	1 344 312	1 674 809	3 968 541	
Percent			0.10	1.37	3.78	18.67	33.87	42.20		100.00

Figure 7.3: Historical development of ringing in Europe: EURING Data Bank (EDB) holdings by ringing scheme and 20-year period. Numbers of recovery records held by the EDB in June 2015. This includes all birds recovered dead or injured together with recaptures and re-sightings more than 10 km from the original ringing location. Initial ringing records and live recaptures close to the ringing location are excluded [222].

7.2.3 Creating probable trajectories with known endpoints

Objective 3: *Advance from "ecologically possible trajectories" to ecologically realistic trajectories: Develop statistical methods to empirically add realistic behaviour in the set of trajectories that connect two locations.*

As mentioned before, the RTG algorithm is restricted to using a constant speed value to generate the trajectories. However, in reality, the speed of movement and the directional development of a trajectory will vary in species-specific and behaviour-specific ways (e.g., the variation of speed and directional change will be different during foraging or migration movements, respectively). Hence, Chapter 5 introduced the Empirical Random Trajectory Generator (eRTG), an algorithm that relies on the empirical frequency distributions of speed and turning angle extracted from observed tracking data representing a particular species and behaviour. The core eRTG algorithm proceeds in two steps: First, the input behaviour from the observed trajectories is pre-processed into a template trajectory whose geometric characteristics are used to calculate the forward-looking probability (P). Second, a longer, undirected correlated random walk is created based on P, and the parameters

needed to calculate the backward-looking probability are set to (Q) to ensure that the mover reaches the predefined destination.

The eRTG algorithm can deliver trajectories with random variations, which are statistically similar to the observed trajectories, for testing ecological hypotheses, e.g. regarding the influence of spatial contextual factors on particular behaviours. The generated trajectories manifest possible movement (as RTG does), and probable movement, and hence trajectories that are 'realistic' in the sense of reproducing the statistical geometric properties of real trajectories.

The capability of generating individual trajectories is beneficial in studies that look beyond species or meta-population levels, as they require methods that enable movement inference both at the individual and aggregated levels [41, 154]. This information is especially useful for understanding migratory connectivity and population regulation, and critical in cases in which the migratory corridor (routes connecting summer and winter habitats) is narrow, therefore important to be protected to ensure the population's survival [224].

The ring data of barn swallows seen in Figure 7.4 show the locations in which the individual animal was observed but carry no information about the movement between the capture and recapture locations of the bird, making it challenging to locate where the ecological corridor of this species has been for the past years. Using these historical data as input data for the eRTG, in combination with recent higher frequency observed trajectories, such as those found in `https://www.movebank.org/` will allow the researcher to generate trajectories that connect the capture-recapture locations (connectivity mapping), shedding some light on the probable locations of the ecological corridor.

Van Toor et al. [191] extended the connectivity mapping with the eRTG, by combining it with environment-informed suitability models for the case of bar-headed goose migration. The suitability models were used to assess the ecological plausibility of each trajectory and select the ones with positive fitness to be used as relevant null models for animal movement (Figure 7.5).

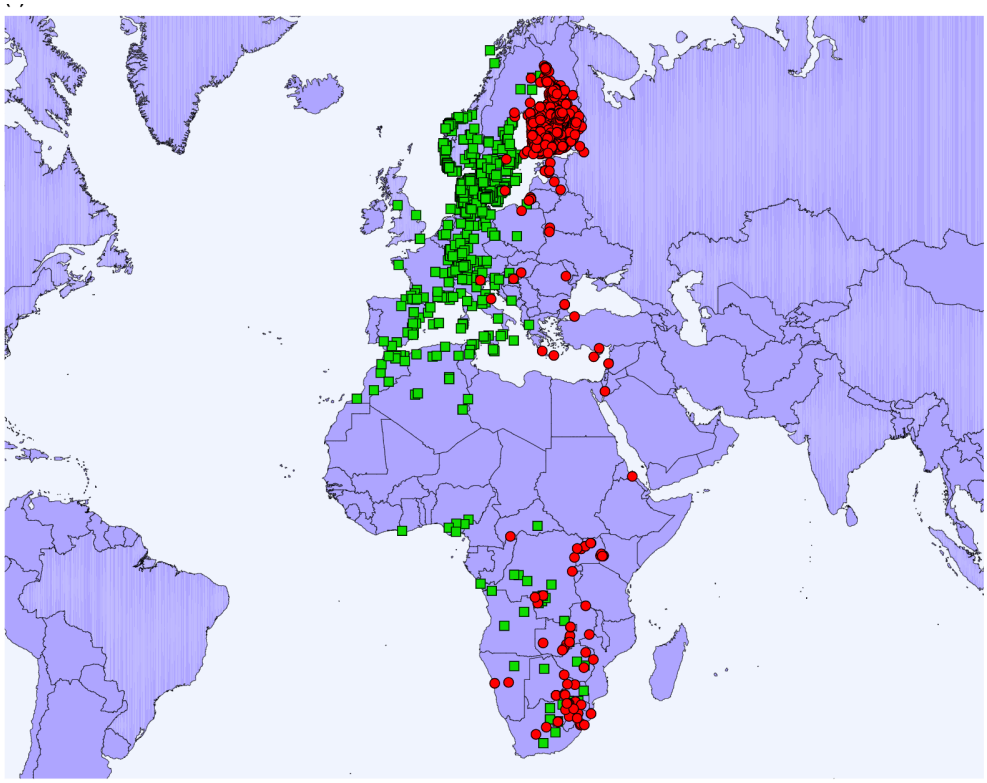

Figure 7.4: Geographic locations of recoveries of barn swallows (*Hirundo rustica*) ringed in Scandinavia (green) and Finland (red), respectively, showing pronounced differences in the number of recoveries between Western and Eastern Europe and between Europe and sub-Saharan Africa, respectively [65].

7.2.4 Contextualising the movement

Objective 4: *Embed context to refine the empirically informed and modelled movement further.*

The basic eRTG algorithm can generate statistically similar trajectories to real, observed trajectories in their speed and geometric characteristics, which can thus be termed 'ecologically realistic'. However, the degree of realism is still limited because the facilitators and inhibitors of movement imposed by the spatial context are not taken into account. The two steps of the basic eRTG algorithm can be extended by including a third step, whereby the spatial context is taken into account in the form of a probability surface to adjust the final selection of the successive displacement accordingly.

Chapter 6 showcased the gradual incorporation of the spatial context in the trajectory generation process. The background information was omitted in the first stage, allowing the agents to move freely in space (representing the basic eRTG). In a second stage, a binary option (water = 0, land = 1) was created to allow the

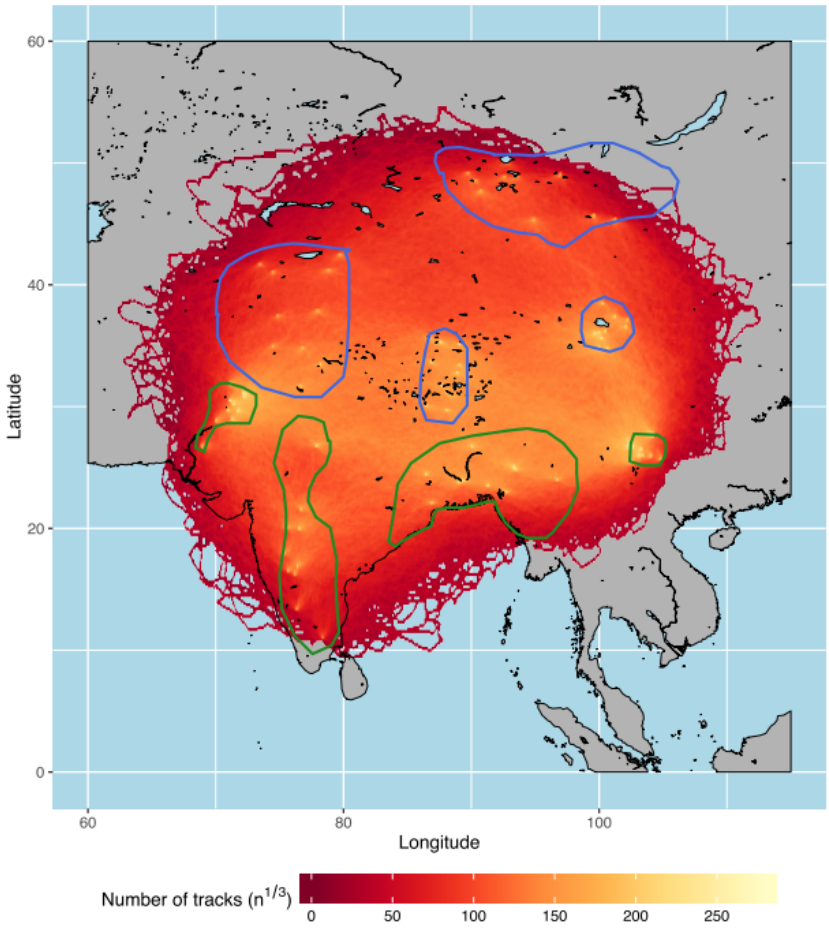

Figure 7.5: Illustration taken from Van Toor et al. [191]: Shown are only tracks with positive fitness, and the colouring illustrates the density of tracks. Blue polygons illustrate the native breeding range, green polygons the native wintering range of the species.

agents to stop only over land, avoiding all the large water bodies. The third stage introduced the preference [0 to 1] of the mover showcased in two examples, one with a vegetation index and one with the river drainage network as attraction elements. The assumption in both cases was that the higher the (vegetation or water) index, the more attractive the location would be as a stop (point). The addition of spatial context as a probability surface allows for easy addition of external factors, at least on an algorithmic level, allowing the user to focus on the challenging ecological side of the work.

Unterfinger [197] extended the contextual awareness and dimensions of the eRTG by introducing movement in the third dimension (eRTG3D) for mapping the collision probabilities between migrating Storks and planes arriving at or departing from Zurich Airport, Switzerland (Figure: 7.6. In order to replicate the 3D movement, two movement behaviours were employed (soaring, gliding), and a rule-based trigger was developed based on the underlying uplift suitability map: if thermal lift exists,

then the soaring behaviour is triggered; otherwise gliding mode is activated. This significant expansion of the eRTG opens the door for applications towards (a) case studies that deal with very high-frequency flying or diving data and (b) extending the transition probabilities that mediate between movement modes, to include environmental conditions, in a state-space manner [225]. Unterfinger's effort is published in the R package *eRTG3D* available on CRAN (`https://CRAN.R-project.org/package=eRTG3D`).

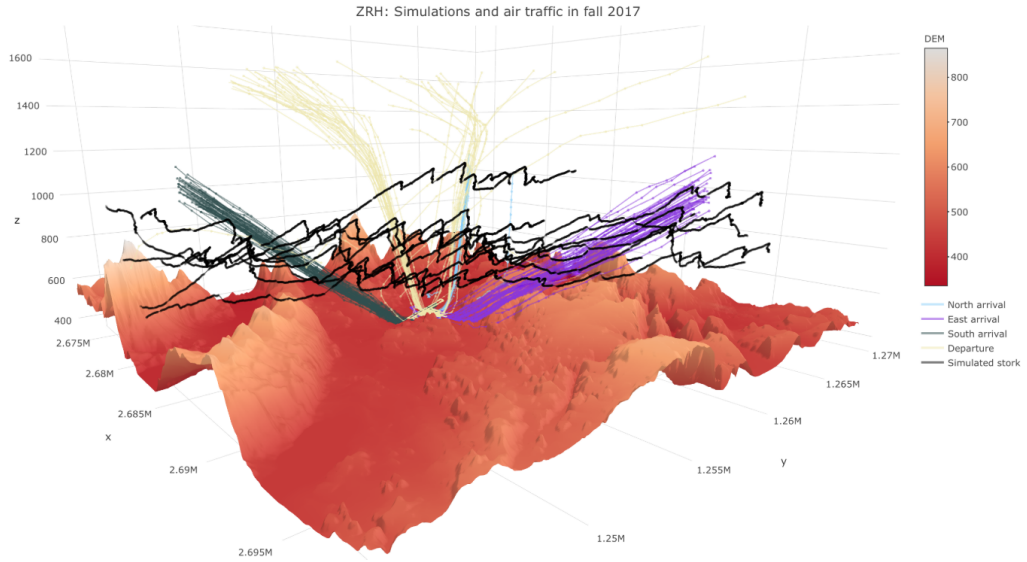

Figure 7.6: An illustration taken from Unterfinger's work [197]: 10 simulation examples and a subset of 50 flight trails per flight corridor at Zurich Airport.

7.3 Answering the main research question

Research Question: *How can we best model movement and context in a simulation framework to better assist in ecological knowledge acquisition?*

The motivation of this research was to find ways to model movement, in the form of simulated trajectories, for use in behavioural movement experiments. Given the ecological perspective of this work, my starting point was to design a simulation modelling framework compatible with the requirements of movement ecology. The proposed hybrid framework became the foundation of the movement model developed, named Empirical Random Trajectory Generator (eRTG). While reviewing methods applied in different domains (time geography, physics, gravitational field simulations, etc.) to use as a potential base model for my work, two fundamental questions kept resurfacing: How can the balance between local and global bias be controlled? And then, how can similar trajectories be created with the exact same endpoints? I selected the Random Walk (RW) as a base movement model due to

its simplicity, transparency, and extensibility. It is easy to incorporate increasing complexity into the basic version of the RW model, starting from a simple RW, to a correlated RW, drifted RW, and finally reinforced RW. The RW model requires no magic numbers in order to be utilised, thus limiting the risk for black-box parts in the methodology, and it can support fundamental mathematical extensions of the basic model, for a wide range of applications, from obstacle avoidance [226, 227] to swarm intelligence [228].

Kareiva and Shigesada's confirmation that random walks can be used to investigate an individual's change of behaviour given a change in its environment [68] formed the backbone of this thesis:

"*It may be especially fruitful to investigate the foraging or searching movements of animals as CRWs whose parameters (moving angle or length) depend on local ecological conditions. Thus, for example, an organism might always move randomly, but with sharper angles in the presence of food resources.*" [68]

Comparing the observed trajectory with a random trajectory that the animal could have taken is the first hint towards a behavioural preference and, therefore, the first step in understanding an animal's reasoning. Transforming the hint into sound ecological hypothesis testing entails creating and cross-validating numerous trajectories. The more detailed the hypothesis, the more realistic the generated data has to be, with 'realistic' in this context, meaning the ability to relate the generated data to the physical choices of the mover.

To increase the ecological interpretability of the proposed simulation algorithms (RTG, and particularly eRTG), the design of the algorithms was based on the development of a conceptual simulation framework, which in turn was grounded in an ecological paradigm [55]. This *grounding* meant that the movement ecology processes were closely followed in the point creation process (possible, probable movements), and the simulated behaviour was extracted from observed trajectories.

One of the core modelling decisions that was made early on was to find the best way to apply bias to a moving object while best preserving its freedom of choice. In essence thisboils down to the question: How can we ensure that the mover will end up in the required location without applying artificial force? Bias and freedom are complementary concepts, as removing one adds more of the other. Accordingly, applying the least bias possible maximises the available freedom of choice. If the freedom of choice is replaced by 'randomness', then my finding can be summarised in a single sentence:

Randomness is best preserved when the application of bias is evenly or randomly distributed along the trajectory in the generating process.

More specifically, given that the mover eventually needs to reach a specific destination, the introduction of bias is inevitable. When the application of the bias is not curated, unrealistic results may be returned, as seen in Figure 4.5. When quantifying the necessary bias and randomising its effect in time, it performs significantly better, as shown in Figure 4.7. Nevertheless, ultimately, extracting the bias from the observed trajectory and applying it to be distributed to all generational processes (as is explained in the calculation of the Q probability in Section 5.3.2) delivers the best results achieved in incorporating the bias in a more *natural* and explainable way.

The application of this method is sufficiently generic to be transferred to and further tested on real ecological questions. Dynamic spatial context can be added in sequential discrete events, and multiple movement modes can be facilitated with simple stochastic or state-space models, as seen in [191, 197]. From a GIScience perspective, the eRTG can be used in combination and extend standard methods such as the Brownian Bridges Model of Movement and the Space-Time Prism. In the first case, the eRTG can be used to calibrate the spread of the visit probabilities given an input behaviour, while in the second, it can be used to introduce a probability surface inside the STP defined boundaries.

7.4 Limitations

The limitations of the RTG algorithm have already been mentioned in Section 7.2.2 as they were the starting point for developing the eRTG algorithm. The following discussion, therefore, concentrates on the limitations of the eRTG algorithm.

Extensive testing of the algorithm revealed some limitations and assumptions of the process. First and foremost, depending on the sinuosity of the input trajectory, the contextual rules enforced, and the time budget restrictions, the agent is trapped at times in the digital dead-end described in Section 5.4.2, not being able to reach the destination under its given movement behaviour. The latter is operationally not critical because running more instances of the eRTG delivers the desired number of results. However, revisiting the ability of the agent to adapt (by changing its behaviour distribution based on the geometric characteristics) and introducing vicinity radius (at this point, the agent sees the water only one step before the dead-end) will optimise the trajectory creation.

Second, all calculations in the eRTG take place on a plane (x,y) projected coordinate system, not on a spheroid (ϕ, λ) geographic coordinate system, imposing distortions on the accuracy of the long segments between points (typically larger than 200km).

Fitting a linear model onto the residuals as a behaviour extraction mechanism rests on various assumptions, including exact linearity of all relationships, normality of residuals or errors from the model, and constant residual variance throughout the range. The "goodness of fit" indicates how much these assumptions hold, though further tests such as outlier detection might be necessary to increase the confidence in the method. Different learning methods should be tested as well, potentially by applying additional clustering methods to identify behaviours on single trajectories and then build deep, convolutional or recurrent neural networks to leverage global datasets of the same or similar species.

The combination of *P*, *Q* & contextual probability is also worth a deeper look as expressing it with the inner product of vectors in Hilbert space might prove sub-optimal in cases. In probability theory a simple multiplication of the two values would suffice to derive the combined probability, given that they are independent. However, in an eRTG case, the joined probability is inverse-distance weighted ($1/D^{\rho}$), explicitly assuming that the closer the agent is to the destination (*D*), the more it is affected by the vector product. The power function adjusts this effect's decay, (ρ or "*rho*", not to be mistaken for the *P-Value* seen in statistics). The selection of ρ requires further empirical research, as bibliographically, no conclusive theoretical justification was found to set its value.

The eRTG only implements part of the conceptual framework proposed in Chapter 3 (cf. Figures 7.2.2 and 3.6). Currently, the eRTG is restricted to a single behavioural mode, captured by one or more observed trajectories representing this particular mode. If the behavioural mode changes, for instance, from migration to foraging, different observed trajectories will have to be fed manually to the eRTG algorithm. Coupling the ABM approach of the eRTG with a simple DES that generates and selects different behavioural modes and changes the initial conditions of the simulation accordingly would allow automating the overall simulation process further.

8 Conclusions

> *Users do not care about what is inside the box, as long as the box does what they need done.*
>
> **— Jef Raskin**
> The Humane Interface

8.1 Contributions and Insights

This Following the detailed discussion of the results for each of the four research objectives of this thesis in Chapter 7, this section provides a condensed summary of the contributions and the insights generated in this work.

8.1.1 Conceptual simulation framework

Develop an integrated simulation approach that allows to cover all components of the conceptual framework of Nathan et al. [55], as well as their interactions.

Contributions
The conceptual hybrid simulation framework proposed in this thesis directly reflects the ecological paradigm proposed by Nathan et al. [55], which is widely seen as the conceptual foundation of movement ecology. The most important difference between the various hybrid simulation frameworks found in the literature ([172, 187–189]) and the proposed one is that the latter accounts for all the basic elements of the ecological paradigm by Nathan et al., allowing for direct connection between the digital and the ecological variables. Expressing an animal as an agent (as seen in Section 3.3), the environmental changes as discrete events in a process (Section 3.4), and the internal state of the mover as a continuously evaluated feedback loop (Section 3.2) assists with the interpretation of the simulation framework. By establishing a consolidated, holistic view of the required components and their interaction, the conceptual simulation framework guided the development of the RTG and eRTG movement models later on.

Insights

The conceptual framework, whose implementation potential is not limited to this thesis, helped me set the broad outlines of functionality and processes that I should focus on. It narrowed down the gap between stating and solving the problem, assisting the discussions with researchers in ecology and articulating the movement strategy. The breakdown of the static part into *initial condition identification*, *possible movement* and *probable movement* marked the larger blocks within which the ecological paradigm is nested (Figure 7.1) and upon which all the developed simulation algorithms were based.

8.1.2 Possible movement

Develop a simulation approach that allows to connect two known locations, possibly separated by a large spatial and temporal distance, by generating ecologically possible trajectories, which comply with the physical limitations of the mover and carry the least statistical bias possible.

Contributions

The Random Trajectory Generator (RTG) was introduced, capable of reliably generating large numbers of trajectories between any two given points on the globe. Collectively, the generated trajectories cover locations that a mover could have *possibly* reached. The trajectories follow the random walk model and can thus be used as a null hypothesis for applications in animal ecology, such as bird migration. The results can be conditioned by two control parameters, the time budget and a constant (usually the maximum) travel speed. The proposed algorithm combines elements from the three leading models in movement simulation: the random walk model, the space-time prism model, and the Brownian bridges movement model. It combines the advantages of all three models, and thus, as demonstrated experimentally, it outperforms them. Since the origin and the destination of the movement may be far apart in space, the generative part of the algorithm was solved on a spherical surface, accounting for the Earth's curvature. In practical terms, the usefulness of this algorithm increases substantially when observations are so scarce that otherwise it would be difficult to make meaningful inferences about movement between measured locations. In cases of connecting capture-recapture locations that are geographically distant, or working with coarse relocation data the RTG can give a starting point for possible space use with minimum assumptions.

Insights

Handling the task of connecting two endpoints that are potentially separated by a large geographic distance carries its practical challenges. Solutions such as kine-

matic interpolation [109] can efficiently close small gaps in high frequency recorded movement but are not so effective in filling in large temporal gaps of varying duration, covering significant geographical distances. The RTG returns consistent, and explainable results and even though its development was primarily rooted in ecology, its usage can be extended to other domains. Confirming its extended applicability, the RTG was included in a library of efficient algorithms for computational movement analysis, targeted at dealing with automotive and human mobility data (`https://movetk.win.tue.nl/`). On the other hand, as discussed in Section 7.2.2, with its assumption of constant speed, the RTG lacked ecological realism and was thus further extended in the subsequent development step.

8.1.3 Probable movement

Advance from "ecologically possible" to "ecologically realistic" trajectories: Develop statistical methods to empirically add realistic behaviour in the set of trajectories that connect two locations.

Contributions

The Empirical Random Trajectory Generator (eRTG) algorithm was developed to generate ecologically realistic trajectories that connect two given endpoints. The realism derives from the use of observed trajectories, whose empirical speed and turning angle distributions are used to condition the generation of trajectory points, rather than simply a constant speed value as in the RTG. The eRTG algorithm solves the given task in two steps by first computing a set of candidate movement directions and then selecting one of them. The selection process is based on two empirically informed vector field histograms computed from observed input trajectory data: one accounting for the forward-looking probability P (of next step selection) and one that expresses the tendency of the move towards the destination (the backward-looking probability Q).

The vector field approach ensures that the mover will reach the predefined destination, while the empirical histograms enforce the resulting trajectories to be geometrically similar to the input trajectories. Given that the input trajectories are real-world, observed data, the simulation delivers realistic trajectories as a result. Being able to generate realistic trajectories aims to complement the efforts of animal tracking taking place in ecology and ultimately to contribute to answering core migratory and landscape connectivity questions.

The essential contribution to the field of ecology goes beyond the practical nature of the simulation and lies in providing a different perspective on the same data. Instead

of attempting to answer the question: "*What does the data show?*", it encourages asking questions such as "*What might the missing data reveal?*", "*Where could the animal have been when we were not observing it?*", "*If we could have tracking devices on all species populations being studied, what would be the variation of movement noted (within an individual, species, or population)?*", and more.

Being able to answer such questions would treat the recorded data points like anchors for exploration instead of boundaries of the analysis, making testing of the ecological hypotheses more resilient to observation bias.

Insights
One of the most important characteristics of eRTG is that it enables non-deterministic yet explainable patterns of movement to emerge. Besides of having identical end-points, every trajectory created shares the statistically same geometric properties with the input data while preserving the random element as the core of trajectory generation. This randomness enables the user to gain an insight into the 'unseen': the range of alternative routes which the observed, or any similarly behaved, mover could have taken if the remainder of the conditions were to be kept unchanged.

8.1.4 Context-awareness

Embed context to further refine the empirically informed and modelled movement.

Contributions
Spatial context was incorporated into the basic eRTG algorithm in two ways: a binary layer and a ratio-scaled layer of values in the interval [0..1]. Thus, two different context-aware movement schemes can be implemented. With the binary layer, areas can be represented that should be avoided, such as defining a mask of water bodies where creating trajectory points is impossible. With the ratio-scaled layer, preference values can be defined, such as a vegetation index denoting the degree of attractiveness of certain areas to be visited by a mover on its path, and hence trajectory points being placed there. These two types of layers were found to be sufficient to represent the *static* spatial context of environmental variables that are stable (e.g., water bodies, topographic barriers) or changing at a slow pace (e.g., vegetation).

Incorporating contextual awareness enables the user to characterise underlying causes that may constrain or limit mobility options. Having this ability added in the eRTG opens the door to investigating wildlife migration and predicting space-usage

patterns across varied habitats with obstacles, attraction and avoidance points, which is the case in almost all applications of movement ecology.

Insights
The effect and usefulness of the incorporation of spatial context in the eRTG are eminent. Adding the extrinsic element to an intrinsically realistic path is necessary for studies that aim to understand how animals move in a landscape [197] even in cases when the observed data is scarce [229].

8.2 Outlook

This thesis has presented algorithms to generate ecologically realistic trajectories connecting two distant endpoints, taking into account static spatial context. A variety of extensions, however, are possible.

First, future work can use the proposed eRTG algorithm in studies addressing core migratory and landscape connectivity questions, as was, for example, done in [191]. Also, using the eRTG, investigations into the inter- and intra-population variability of movement paths will allow the researcher to compare related individual and species movement strategies.

Several points can be improved both regarding the modelling of the capacities of the moving agents and the spatial environment. Adding agent-to-agent interactions will allow testing leader–follower patterns, predator–prey strategies. At the same time, considering the internal state of each individual by altering the movement behaviour of the agent (e.g., switching between foraging, predator evasion and migration) will add further dynamicity into the model, as seen in Unterfinger [197] and Van Toor et al. [191]. Smarter obstacle avoidance capabilities in the form of an obstacle limitation method [230, 231] for controlling the behaviour, or the firefly algorithm [228, 232] for allowing the behaviour to emerge (swarm intelligence) will enhance the applicability of the eRTG in the path-finding domain. The extent and precise shape of space-use depend on the population's demographic composition and can potentially be included and controlled by a systems dynamic approach (Section 3.1.2). Going a step further, emerging or adaptive behaviour enters the picture: latent learning behaviours, changes of behaviour and adaptive penalties can be introduced, and user-defined parameters can be added to enable the controlled handling of the statistical footprint of the simulated trajectories, ultimately allowing for more realistic simulations.

From an engineering perspective, when dealing with very high temporal resolution over long distances, an efficiency drop was noticed in the algorithm as the percentage of dead-end instances increase. A possible explanation of this phenomenon would be the difference in scale between the movement definition (e.g., in seconds and meters) and the behaviour under study (e.g., in migration: weeks and hundreds of kilometres). The evaluation of the results needs to be carefully considered and adapted to the specific characteristics of the data under study (e.g. sample size, or type of distribution).

Boosting the computational performance of the implementation by scalable distributed processing methods and decentralised network topology algorithms [233–235] will enable frequent population-wide simulations, accounting for complex interactions (agent-to-agent and agent-to-environment), dynamic context, and population dynamics. Such simulations will aim to "predict evolutionary, physiological, and life-history consequences of particular strategies" [136].

Last but not least, it is important to address how the insights of such a modelling exercise can find their impact in practice and how decision-makers can be enabled to take an educated decision based on tested results. Some users claim that "a simple model that can be well communicated and explained is more useful than a complex model that has narrow applicability, high costs of data, and more uncertainty" [236], whereas others prefer adding more realism (and with that more complexity) to the model [237]. Through participatory modelling [238], the description of the models can be standardised [239, 240] to assist the process of communicating the results to various stakeholders. Going even a step further would be to align the modelling process to the management process [54]by exposing the model's components and underlying assumptions to the intended user early on to give the user the chance to use the model in an educated rather than a black-box manner.

Bibliography

[1] N. Kadhila, "NSSC Biology Module 1 Student's Book," in *NSSC Biology*, Cambridge University Press, 2nd ed., 2009.

[2] T. Mueller and W. F. Fagan, "Search and navigation in dynamic environments - From individual behaviors to population distributions," *Oikos*, vol. 117, no. 5, pp. 654–664, 2008.

[3] M. Baguette, V. M. Stevens, and J. Clobert, "The pros and cons of applying the movement ecology paradigm for studying animal dispersal," *Movement Ecology*, vol. 2, no. 1, pp. 1–13, 2014.

[4] B. B. Niebuhr, M. E. Wosniack, M. C. Santos, E. P. Raposo, G. M. Viswanathan, M. G. Da Luz, and M. R. Pie, "Survival in patchy landscapes: The interplay between dispersal, habitat loss and fragmentation," *Scientific Reports*, vol. 5, pp. 1–10, 2015.

[5] T. C. Boothby, "Mechanisms and evolution of resistance to environmental extremes in animals," *EvoDevo*, vol. 10, no. 1, pp. 1–12, 2019.

[6] H. W. Fricke, "Behaviour as part of ecological adaptation - In situ studies in the coral reef," *Helgoländer Wissenschaftliche Meeresuntersuchungen*, vol. 24, no. 1-4, pp. 120–144, 1973.

[7] B. B. Wong and U. Candolin, "Behavioral responses to changing environments," *Behavioral Ecology*, vol. 26, no. 3, pp. 665–673, 2015.

[8] J. M. Smith, "The evolution of behavior.," *Scientific American*, vol. 239, no. 3, pp. 176–178, 1978.

[9] D. J. Peuquet, "It's About Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems," *Annals of*

the Association of American Geographers, vol. 84, no. 3, pp. 441–461, 1994.

[10] E. Gurarie, *Models and analysis of animal movements: From individual tracks to mass dispersal*. PhD thesis, University of Washington, 2008.

[11] G. E. Long, "Foundations of Ecology and Classic Papers With Commentaries," *American Entomologist*, pp. 184–184, 1992.

[12] P. Laube, "The low hanging fruit is gone: achievements and challenges of computational movement analysis," *SIGSPATIAL Special*, vol. 7, no. 1, pp. 3–10, 2015.

[13] W. Siabato, C. Claramunt, S. Ilarri, and M. A. Manso-Callejo, "A survey of modelling trends in temporal GIS," *ACM Computing Surveys*, vol. 51, no. 2, 2018.

[14] F. Cagnacci, L. Boitani, R. a. Powell, and M. S. Boyce, "Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges.," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 365, pp. 2157–62, 7 2010.

[15] P. A. Westley, A. M. Berdahl, C. J. Torney, and D. Biro, "Collective movement in ecology: From emerging technologies to conservation and management," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 373, no. 1746, 2018.

[16] A. Verma, R. van der Wal, and A. Fischer, "Imagining wildlife: New technologies and animal censuses, maps and museums," *Geoforum*, 2016.

[17] H. J. Miller, S. Dodge, J. Miller, and G. Bohrer, "Towards an integrated science of movement: converging research on animal movement ecology and human mobility science," *International Journal of Geographical Information Science*, vol. 33, no. 5, pp. 855–876, 2019.

[18] B. M. Allan, D. G. Nimmo, D. Ierodiaconou, J. VanDerWal, L. P. Koh, and E. G. Ritchie, "Futurecasting ecological research: the rise of technoecology," *Ecosphere*, vol. 9, no. 5, 2018.

[19] M. A. Forin-Wiart, P. Hubert, P. Sirguey, and M. L. Poulle, "Performance and accuracy of lightweight and low-cost GPS data loggers according to antenna positions, fix intervals, habitats and animal movements," *PLoS ONE*, vol. 10, no. 6, pp. 1–21, 2015.

[20] M. Teague O'Mara, M. Wikelski, and D. K. Dechmann, "50 years of bat tracking: Device attachment and future directions," *Methods in Ecology and Evolution*, vol. 5, no. 4, pp. 311–319, 2014.

[21] P. López-López, "Individual-Based Tracking Systems in Ornithology: Welcome to the Era of Big Data," *Ardeola*, vol. 63, no. 1, p. 103, 2016.

[22] F. Batsleer, D. Bonte, D. Dekeukeleire, S. Goossens, W. Poelmans, E. Van der Cruyssen, D. Maes, M. L. Vandegehuchte, E. V. D. Cruyssen, D. Maes, and M. L. Vandegehuchte, "The neglected impact of tracking devices on terrestrial arthropods," *Methods in Ecology and Evolution*, vol. 11, no. 3, pp. 350–361, 2020.

[23] C. Scandolara, D. Rubolini, R. Ambrosini, M. Caprioli, S. Hahn, F. Liechti, A. Romano, M. Romano, B. Sicurella, and N. Saino, "Impact of miniaturized geolocators on barn swallow Hirundo rustica fitness traits," *Journal of Avian Biology*, vol. 45, no. 5, pp. 417–423, 2014.

[24] S. J. Cooke, S. G. Hinch, M. Wikelski, R. D. Andrews, L. J. Kuchel, T. G. Wolcott, and P. J. Butler, "Biotelemetry: a mechanistic approach to ecology.," *Trends in ecology & evolution*, vol. 19, pp. 334–43, 6 2004.

[25] S. M. Tomkiewicz, M. R. Fuller, J. G. Kie, and K. K. Bates, "Global positioning system and associated technologies in animal behaviour and ecological research.," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 365, pp. 2163–76, 7 2010.

[26] S. Nadi and M. Delavar, "Spatio-temporal modeling of dynamic phenomena in GIS," *ScanGIS*, no. January 2003, pp. 215–225, 2003.

[27] P. A. Burrough, "Are GIS data structures too simple minded?," *Computers and Geosciences*, vol. 18, no. 4, pp. 395–400, 1992.

[28] J. Albrecht, "Dynamic GIS," in *Handbook of GIScience*, pp. 436–446, Wilson & Fotherringham, 1 ed., 2007.

[29] D. Karssenberg and K. De Jong, "Dynamic environmental modelling in GIS: 1. Modelling in three spatial dimensions," *International Journal of Geographical Information Science*, vol. 19, no. 5, pp. 559–579, 2005.

[30] G. F. Hepner and M. V. Finco, "Modeling dense gaseous contaminant pathways over complex terrain using a geographic information system," *Journal of Hazardous Materials*, vol. 42, no. 2, pp. 187–199, 1995.

[31] C. M. Gold and A. R. Condal, "A spatial data structure integrating gis and simulation in a marine environment," *Marine Geodesy*, vol. 18, no. 3, pp. 213–228, 1995.

[32] D. G. Green, R. E. Reichelt, J. van der Laan, and B. W. Macdonald, "A generic approach to landscape modelling," *Mathematics and Computers in Simulation*, vol. 32, no. 1-2, pp. 237–242, 1990.

[33] N. P. Hegde, I. V. Muralikrishna, and K. V. Chalapatirao, "Integration of Cellular Automata and Gis for Simulating Land Use Changes," *5th International Symposium Spatial Data Quality - ISPRS*, vol. 1, no. 1, pp. 1–2, 2007.

[34] R. Gimblett, "Linking perception research, visual simulations and dynamic modeling within a GIS framework: The ball state experience," *Computers, Environment and Urban Systems*, vol. 13, no. 2, pp. 109–123, 1989.

[35] P. Deadman and R. H. Gimblett, "A role for goal-oriented autonomous agents in modeling people-environment interactions in forest recreation," *Mathematical and Computer Modelling*, vol. 20, no. 8, pp. 121–133, 1994.

[36] M. P. Armstrong and M. P. Armstrong, "Toward the development of a conceptual framework for GIS-based collaborative spatial," in *Proceedings of the 2nd ACM Workshop on Advances in GIS*, (Baltimore,Maryland,USA,December1994), pp. 4–7, 1995.

[37] T. Hägerstrand, "What about people in Regional Science?," *Papers of the Regional Science Association*, vol. 24, no. 1, pp. 6–21, 1970.

[38] N. Thrift, "An introduction to time geography," *Concepts and Techniques in Modern Geography*, vol. 13, pp. 1–37, 1977.

[39] H. Whitehead, "Analysis of animal movement using opportunistic individual identifications: Application to sperm whales," *Ecology*, vol. 82, no. 5, pp. 1417–1432, 2001.

[40] B. T. Mcclintock, D. J. Russell, J. Matthiopoulos, and R. King, "Combining individual animal movement and ancillary biotelemetry data to investigate population-level activity budgets," *Ecology*, vol. 94, no. 4, pp. 838–849, 2013.

[41] K. A. Jønsson, A. P. Tøttrup, M. K. Borregaard, S. A. Keith, C. Rahbek, and K. Thorup, "Tracking Animal Dispersal: From Individual Movement to Community Assembly and Global Range Dynamics," *Trends in Ecology and Evolution*, vol. 31, no. 3, pp. 204–214, 2016.

[42] J. Martin, C. Calenge, P. Y. Quenette, and D. Allainé, "Importance of movement constraints in habitat selection studies," *Ecological Modelling*, vol. 213, no. 2, pp. 257–262, 2008.

[43] G. Andreguetto Maciel and F. Lutscher, "How individual movement response to habitat edges affects population persistence and spatial spread," *American Naturalist*, vol. 182, no. 1, pp. 42–52, 2013.

[44] J. M. Morales, P. R. Moorcroft, J. Matthiopoulos, J. L. Frair, J. G. Kie, R. A. Powell, E. H. Merrill, and D. T. Haydon, "Building the bridge between animal movement and population dynamics," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 365, no. 1550, pp. 2289–2301, 2010.

[45] T. Avgar, A. Mosser, G. S. Brown, and J. M. Fryxell, "Environmental and individual drivers of animal movement patterns across a wide geographical gradient," *Journal of Animal Ecology*, vol. 82, no. 1, pp. 96–106, 2013.

[46] C. M. Buchmann, F. M. Schurr, R. Nathan, and F. Jeltsch, "Movement upscaled - the importance of individual foraging movement for community response to habitat loss," *Ecography*, vol. 35, no. 5, pp. 436–445, 2012.

[47] E. A. Codling, J. W. Pitchford, and S. D. Simpson, "Group navigation and the "many-wrongs principle" in models of animal movement," *Ecology*, vol. 88, no. 7, pp. 1864–1870, 2007.

[48] N. Tinbergen, "Foundations of Animal Behavior- On aims and methods of Ethology," *Z. Tierpsychologie*, vol. 20, no. March, pp. 410–433, 1963.

[49] T. A. Patterson, A. Parton, R. Langrock, P. G. Blackwell, L. Thomas, and R. King, "Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges," *AStA Advances in Statistical Analysis*, vol. 101, no. 4, pp. 399–438, 2017.

[50] J. McGowan, M. Beger, R. L. Lewison, R. Harcourt, H. Campbell, M. Priest, R. G. Dwyer, H. Y. Lin, P. Lentini, C. Dudgeon, C. McMahon, M. Watts, and H. P. Possingham, "Integrating research using animal-borne telemetry with

the needs of conservation management," *Journal of Applied Ecology*, vol. 54, no. 2, pp. 423–429, 2017.

[51] J. S. Clark and A. E. Gelfand, "A future for models and data in environmental science.," *Trends in ecology & evolution*, vol. 21, pp. 375–80, 7 2006.

[52] L. Breiman, "Statistical Modeling: The Two Cultures," *Statistical Science*, vol. 16, no. 3, pp. 199–215, 2011.

[53] M. Ryo, B. Angelov, S. Mammola, J. M. Kass, B. M. Benito, and F. Hartig, "Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models," *Ecography*, vol. 44, no. 2, pp. 199–205, 2021.

[54] N. Schuwirth, F. Borgwardt, S. Domisch, M. Friedrichs, M. Kattwinkel, D. Kneis, M. Kuemmerlen, S. D. Langhans, J. Martínez-López, and P. Vermeiren, "How to make ecological models useful for environmental management," *Ecological Modelling*, vol. 411, no. August, p. 108784, 2019.

[55] R. Nathan, W. M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. E. Smouse, "A movement ecology paradigm for unifying organismal movement research," *PNAS*, vol. 105, no. 49, pp. 19052–19059, 2008.

[56] R. Descartes, *Principia philosophiae*, vol. VIII(i). sumtibus Friderici Knochii & filii, 1644.

[57] E. W. Weisstein, ""Tautochrone Problem." MathWorld–A Wolfram Web Resource.."

[58] A. Elzinga, "Huygens ' Theory of Research and Descartes ' Theory of Knowledge II (Zeitschrift fuer allgemeine Wissenschaftstheorie III)," *Journal for General Philosophy of Science*, vol. 3, no. 1, pp. 9–27, 1972.

[59] P. E. Smouse, S. Focardi, P. R. Moorcroft, J. G. Kie, J. D. Forester, and J. M. Morales, "Stochastic modelling of animal movement.," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 365, pp. 2201–11, 7 2010.

[60] E. Aspillaga, K. Safi, B. Hereu, and F. Bartumeus, "Modelling the three-dimensional space use of aquatic animals combining topography and Eulerian telemetry data," *Methods in Ecology and Evolution*, vol. 10, no. 9, pp. 1551–1557, 2019.

[61] E. A. Codling, M. J. Plank, and S. Benhamou, "Random walk models in biology.," *Journal of the Royal Society, Interface / the Royal Society*, vol. 5, pp. 813–34, 8 2008.

[62] W. F. Fagan and J. M. Calabrese, "The Correlated Random Walk and the Rise of Movement Ecology," *Bulletin of the Ecological Society of America*, vol. 95, no. 3, pp. 204–206, 2014.

[63] G. Wittemyer, J. M. Northrup, and G. Bastille-Rousseau, "Behavioural valuation of landscapes using movement data," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 374, no. 1781, 2019.

[64] R. N. Binny, R. Law, and M. J. Plank, "Living in groups: Spatial-moment dynamics with neighbour-biased movements," *Ecological Modelling*, vol. 415, no. June 2019, p. 108825, 2019.

[65] K. Thorup, F. Korner-Nievergelt, E. B. Cohen, and S. R. Baillie, "Large-scale spatial analysis of ringing and re-encounter data to infer movement patterns: A review including methodological perspectives," *Methods in Ecology and Evolution*, vol. 5, no. 12, pp. 1337–1350, 2014.

[66] C. S. Patlak, "Random walk with persistence and external bias," *The bulletin of mathematical biophysics*, vol. 15, no. 3, pp. 311–338, 1953.

[67] C. S. Patlak, "A mathematical contribution to the study of orientation of organisms," *The Bulletin of Mathematical Biophysics*, vol. 15, no. 4, pp. 431–476, 1953.

[68] P. M. Kareiva and N. Shigesada, "Analyzing insect movement as a correlated random walk," *Oecologia*, vol. 56, no. 2-3, pp. 234–238, 1983.

[69] B. Davis, "Reinforced Random Walk," *Probability Theory and Related Fields*, vol. 229, no. 84, pp. 203–229, 1990.

[70] H. A. Levine and B. D. Sleeman, "A system of reaction diffusion equations arising in the theory of reinforced random walks," *SIAM Journal on Applied Mathematics*, vol. 57, no. 3, pp. 683–730, 1997.

[71] P. Turchin, *Quantitative Analysis of Movement: Measuring and Modelling Population Redistribution in Animals and Plants*. Sinauer Publishers, Sunderland, MA, 1998.

[72] S. C. Ahearn, S. Dodge, A. Simcharoen, G. Xavier, and J. L. Smith, "A context-sensitive correlated random walk: a new simulation model for movement," *International Journal of Geographical Information Science*, vol. 8816, no. September, pp. 1–17, 2016.

[73] H. K. Preisler, A. a. Ager, B. K. Johnson, and J. G. Kie, "Modeling animal movements using stochastic differential equations," *Environmetrics*, vol. 15, pp. 643–657, 11 2004.

[74] H. K. Preisler, D. R. Brillinger, A. A. Ager, J. G. Kie, and R. P. Akers, "Stochastic Differential Equations: A Tool for Studying Animal Movement," in *IUFPRO Conference, Greenwich*, p. 4.11, 2001.

[75] A. M. Yaglom, *Correlation theory of stationary and related random functions: Supplementary notes and references.* Springer Science & Business Media, 2012.

[76] H. G. Othmer, S. R. Dunbar, and W. Alt, "Models of dispersal in biological systems," *Journal of Mathematical Biology*, vol. 26, no. 3, pp. 263–298, 1988.

[77] P. G. Blackwell, "Random Difusion Models for Animal Movement," *Ecological Modelling*, vol. 100, pp. 87–102, 1997.

[78] K. J. Harris and P. G. Blackwell, "Flexible continuous-time modelling for heterogeneous animal movement," *Ecological Modelling*, vol. 255, pp. 29–37, 4 2013.

[79] J. E. Milner, P. G. Blackwell, and M. Niu, "Modelling and inference for the movement of interacting animals," *Methods in Ecology and Evolution*, vol. 12, no. 1, pp. 54–69, 2021.

[80] C. S. Nations and R. C. Anderson-Sprecher, "Estimation of Animal Location from Radio Telemetry Data with Temporal Dependencies," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 11, no. 1, pp. 87–105, 2006.

[81] E. H. Isaaks and R. M. Srivastava, "Applied geostatistics: Oxford University Press," *New York*, p. 561, 1989.

[82] C. H. Fleming, W. F. Fagan, T. Mueller, K. A. Olson, P. Leimgruber, J. M. Calabrese, and E. G. Cooch, "Rigorous home range estimation with movement data: A new autocorrelated kernel density estimator," *Ecology*, vol. 96, no. 5, pp. 1182–1188, 2015.

[83] G. Bohling, "INTRODUCTION TO GEOSTATISTICS And VARIOGRAM ANALYSIS," *Earth*, no. October, pp. 1–20, 2005.

[84] C. H. Fleming, J. M. Calabrese, T. Mueller, K. a. Olson, P. Leimgruber, and W. F. Fagan, "Non-Markovian maximum likelihood estimation of autocorrelated movement processes," *Methods in Ecology and Evolution*, vol. 5, no. 5, pp. 462–472, 2014.

[85] D. S. Fisher, "Random walks in random environments," *The American Physical Society*, vol. 30, no. 2, pp. 960–964, 1984.

[86] R. a. Siegel, "Random walks in biology," *Journal of Controlled Release*, vol. 32, no. 2, pp. 201–202, 1994.

[87] N. A. Hill and D.-P. Häder, "A Biased Random Walk Model for the Trajectories of Swimming Micro-organisms," *Journal of Theoretical Biology*, vol. 186, no. 4, pp. 503–526, 1997.

[88] C. M. Bergman, J. A. Schaefer, and L. S.N., "Caribou movement as a correlated random walk," *Oceanologia*, no. July 1999, pp. 364–374, 2000.

[89] E. A. Codling, "Biased Random Walks in Biology," *Doctor of Philosophy Thesis*, no. August, p. 152, 2003.

[90] J. Kempe, "Quantum random walks: An introductory overview," *Contemporary Physics*, vol. 44, no. 4, pp. 307–327, 2003.

[91] L. Seuront, F. G. Schmitt, M. C. Brewer, J. R. Strickler, and S. Souissi, "From random walk to multifractal random walk in zooplankton swimming behavior," *Zoological Studies*, vol. 43, no. 2, pp. 498–510, 2004.

[92] F. Bartumeus, M. G. E. Da Luz, G. M. Viswanatham, and J. Catalan, "Animal Search Strategies: A Quantitative Random Walk Analysis," *Ecological Society of America*, vol. 86, no. 11, pp. 3078–3087, 2005.

[93] B. R. G. Prasad and R. M. Borges, "Searching on patch networks using correlated random walks: space usage and optimal foraging predictions using Markov chain models.," *Journal of theoretical biology*, vol. 240, pp. 241–9, 5 2006.

[94] S. Benhamou, "How many animals really do the Levy Walk?," *Ecology*, vol. 88, no. 8, pp. 1962–1969, 2007.

[95] Y. Tremblay, P. W. P. W. Robinson, and D. P. Costa, "A parsimonious approach to modeling animal movement data," *PLoS One*, vol. 4, p. e4711, 1 2009.

[96] P. Nouvellet, J. P. Bacon, and D. Waxman, "Fundamental insights into the random movement of animals from a single distance-related statistic.," *The American naturalist*, vol. 174, pp. 506–14, 10 2009.

[97] R. Kawai and S. Petrovskii, "Multi-scale properties of random walk models of animal movement: lessons from statistical inference," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 468, pp. 1428–1451, 2 2012.

[98] T. Duchesne, D. Fortin, and L.-P. Rivest, "Equivalence between Step Selection Functions and Biased Correlated Random Walks for Statistical Inference on Animal Movement," *Plos One*, vol. 10, no. 4, p. e0122947, 2015.

[99] B. T. Mcclintock, R. King, L. Thomas, J. Matthiopoulos, B. J. Mcconnell, and J. M. Morales, "A general discrete-time modeling framework for animal movement using multistate random walks," *Ecological Monographs*, vol. 82, no. 3, pp. 335–349, 2012.

[100] H. J. MILLER, "Modelling accessibility using space-time prism concepts within geographical information systems," *International journal of geographical information systems*, vol. 5, pp. 287–301, 1 1991.

[101] B. Kuijpers, H. J. Miller, T. Neutens, and W. Othman, "Anchor uncertainty and space-time prisms on road networks," *International Journal of Geographical Information Science*, vol. 24, no. 8, pp. 1223–1248, 2010.

[102] T. Neutens, N. Van de Weghe, F. Witlox, and P. De Maeyer, "A three-dimensional network-based space–time prism," *Journal of Geographical Systems*, vol. 10, no. 1, pp. 89–107, 2008.

[103] H. J. Miller, M. Raubal, and Y. Jaegal, "Measuring Space-Time Prism Similarity Through Temporal Profile Curves," in *Geospatial Data in a Changing World: Lecture Notes in Geoinformation and Cartography* (T. Sarjakoski, M. Y. Santos, and L. T. Sarjakoski, eds.), pp. 51–66, Springer, 2016.

[104] B. Y. Chen, Q. Li, D. Wang, S.-L. Shaw, W. H. Lam, H. Yuan, and Z. Fang, "Reliable Space–Time Prisms Under Travel Time Uncertainty," *Annals of the Association of American Geographers*, vol. 103, no. October, pp. 1502–1521, 2013.

[105] J. a. Downs, M. W. Horner, and A. D. Tucker, "Time-geographic density estimation for home range analysis," *Annals of GIS*, vol. 17, no. February, pp. 163–171, 2011.

[106] A. M. El-Sayed, P. Scarborough, L. Seemann, and S. Galea, "Social network analysis and agent-based modeling in social epidemiology.," *Epidemiologic perspectives & innovations*, vol. 9, pp. 1–9, 1 2012.

[107] B. Kuijpers, H. J. Miller, and W. Othman, "Kinetic space-time prisms," *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '11*, p. 162, 2011.

[108] J. a. Long, T. a. Nelson, and F. S. Nathoo, "Toward a kinetic-based probabilistic time geography," *International Journal of Geographical Information Science*, vol. 28, no. February, pp. 855–874, 2014.

[109] J. A. Long, "Kinematic interpolation of movement data Kinematic interpolation of movement data," *International Journal of Geographical Information Science*, vol. 30, no. 5, pp. 854–868, 2016.

[110] J. A. Long, "Modeling movement probabilities within heterogeneous spatial fields," *Journal of Spatial Information Science*, vol. 16, no. 16, pp. 85–116, 2018.

[111] Y. Zheng and X. Zhou, *Compute with spatial trajectories*. Springer New York Dordrescht Heidelberg London, 2011.

[112] J. a. Long and T. a. Nelson, "Time geography and wildlife home range delineation," *Journal of Wildlife Management*, vol. 76, no. 2, pp. 407–413, 2012.

[113] M. W. Horner, B. Zook, and J. a. Downs, "Where were you? Development of a time-geographic approach for activity destination re-construction," *Computers, Environment and Urban Systems*, vol. 36, no. 1, pp. 488–499, 2012.

[114] Y. Song and H. J. Miller, "Simulating visit probability distributions within planar space-time prisms," *International Journal of Geographical Information*, vol. 28, no. 1, 2014.

[115] R. W. Loraamm, "Incorporating behavior into animal movement modeling: a constrained agent-based model for estimating visit probabilities in space-time

prisms," *International Journal of Geographical Information Science*, vol. 34, no. 8, pp. 1607–1627, 2020.

[116] F. Bullard, *Estimating the Home Range of an Animal: A Brownian Bridge Approach*. PhD thesis, University of North Carolina at Chapel Hill, 1991.

[117] J. S. Horne, E. O. Garton, S. M. Krone, and J. S. Lewis, "Analyzing animal movements using Brownian bridges.," *Ecology*, vol. 88, pp. 2354–63, 9 2007.

[118] B. Kranstauber, R. Kays, S. D. Lapoint, M. Wikelski, and K. Safi, "A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement.," *The Journal of animal ecology*, vol. 81, pp. 738–46, 7 2012.

[119] S. Benhamou, "Dynamic approach to space and habitat use based on biased random bridges," *PLoS ONE*, vol. 6, no. 1, p. e14592, 2011.

[120] J. Wall, G. Wittemyer, V. Lemay, I. Douglas-Hamilton, and B. Klinkenberg, "Elliptical Time-Density model to estimate wildlife utilization distributions," *Methods in Ecology and Evolution*, vol. 5, pp. 780–790, 2014.

[121] S. K. Nechaev, A. Y. Grosberg, and A. M. Vershik, "Random walks on braid groups: Brownian bridges, complexity and statistics," *Journal of Physics a-Mathematical and General*, vol. 29, no. 10, pp. 2411–2433, 1996.

[122] Y. Song and H. J. Miller, "Modelling Visit Probabilities within Space-Time Prisms using Directed Random Walk and Truncated Brownian Bridges," in *Giscience Proceedings 2012, Columbus, Ohio*, 2012.

[123] M. M. Fyrillas and K. K. Nomura, "Diffusion and Brownian motion in Lagrangian coordinates," *Journal of Chemical Physics*, vol. 126, no. 16, pp. 1–9, 2007.

[124] N. E. Humphries, N. Queiroz, J. R. M. Dyer, N. G. Pade, M. K. Musyl, K. M. Schaefer, D. W. Fuller, J. M. Brunnschweiler, T. K. Doyle, J. D. R. Houghton, G. C. Hays, C. S. Jones, L. R. Noble, V. J. Wearmouth, E. J. Southall, and D. W. Sims, "Environmental context explains Lévy and Brownian movement patterns of marine predators.," *Nature*, vol. 465, pp. 1066–9, 6 2010.

[125] K. Buchin, S. Sijben, E. E. van Loon, N. Sapir, S. Mercier, T. J. Marie Arseneau, and E. P. Willems, "Deriving movement properties and the effect of the

environment from the Brownian bridge movement model in monkeys and birds.," *Movement ecology*, vol. 3, no. 1, p. 18, 2015.

[126] B. Kranstauber, "Modelling animal movement as Brownian bridges with covariates," *Movement Ecology*, vol. 7, no. 1, 2019.

[127] J. L. Gould and C. G. Gould, *The Animal Mind*. New York: Scientific American Library, 1999.

[128] F. Barraquand and S. Benhamou, "Animal movements in heterogeneous landscapes: identifying profitable places and homogeneous movement bouts.," *Ecology*, vol. 89, pp. 3336–48, 12 2008.

[129] S. Dodge, G. Bohrer, K. Bildstein, S. C. Davidson, R. Weinzierl, M. J. Bechard, D. Barber, R. Kays, D. Brandes, J. Han, and M. Wikelski, "Environmental drivers of variability in the movement ecology of turkey vultures (Cathartes aura) in North and South America," *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, vol. 369, no. 1643, p. 1471–2970, 2014.

[130] B. Tomaszewski and A. M. MacEachren, "Geo-historical context support for information foraging and sensemaking: Conceptual model, implementation, and assessment," *2010 IEEE Symposium on Visual Analytics Science and Technology*, pp. 139–146, 10 2010.

[131] C. Gschwend, *Relating Movement to Geographic Context - Effects of Preprocessing, Relation Methods and Scale*. PhD thesis, University of Zurich, 2015.

[132] M. Buchin, S. Dodge, and B. Speckmann, "Similarity of trajectories taking into account geographic context," *Journal of Spatial Information Science*, vol. 9, no. 2014, pp. 101–124, 2014.

[133] J.-h. Böse, K. Hahn, M. Scholz, H. Schweppe, and A. Voisard, "Using Moving Object Databases to Provide Context Information in Mobile Ad-hoc Networks," *Data Management*, 2006.

[134] B. D. Dalziel, J. M. Morales, and J. M. Fryxell, "Fitting probability distributions to animal movement trajectories: using artificial neural networks to link distance, resources, and memory.," *The American naturalist*, vol. 172, pp. 248–58, 8 2008.

[135] G. Technitis and R. Weibel, "A hybrid simulation model for moving objects," in *AutoCarto 2012*, 2012.

[136] R. Joo, S. Picardi, M. E. Boone, T. A. Clay, S. C. Patrick, V. S. Romero-Romero, and M. Basille, "A decade of movement ecology," *arXiv (preprint)*, pp. 1–17, 2020.

[137] L. Waterhouse, J. White, K. See, A. Murdoch, and B. X. Semmens, "A Bayesian nested patch occupancy model to estimate steelhead movement and abundance," *Ecological Applications*, vol. 30, no. 8, pp. 1–17, 2020.

[138] D. A. Wijeyakulasuriya, E. W. Eisenhauer, B. A. Shaby, and E. M. Hanks, "Machine learning for modeling animal movement," *PLoS ONE*, vol. 15, no. 7 July, pp. 1–30, 2020.

[139] J. D. Bailey, J. Wallis, and E. A. Codling, "Navigational efficiency in a biased and correlated random walk model of individual animal movement," *Ecology*, vol. 99, no. 1, pp. 217–223, 2018.

[140] G. Wallentin, "Spatial simulation: A spatial perspective on individual-based ecology—a review," *Ecological Modelling*, vol. 350, pp. 30–41, 2017.

[141] J. C. Russell, E. M. Hanks, A. P. Modlmeier, and D. P. Hughes, "Modeling Collective Animal Movement Through Interactions in Behavioral States," *Journal of Agricultural, Biological and Environmental Statistics*, vol. 22, 2017.

[142] B. D. Dalziel, J. M. Morales, and J. M. Fryxell, "Fitting dynamic models to animal movement data: The importance of probes for model selection, a reply to Franz and Caillaud," *American Naturalist*, vol. 175, no. 6, pp. 762–764, 2010.

[143] J. C. Russell, E. M. Hanks, and M. Haran, "Dynamic Models of Animal Movement with Spatial Point Process Interactions," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 21, no. 1, pp. 22–40, 2016.

[144] F. H. Jensen, A. K. Scharf, P. E. Smouse, J. M. Eisaguirre, M. Auger-Méthé, C. P. Barger, S. B. Lewis, T. L. Booms, and G. A. Breed, "Dynamic-Parameter Movement Models Reveal Drivers of Migratory Pace in a Soaring Bird," *Frontiers in Ecology and Evolution | www.frontiersin.org*, vol. 1, p. 317, 2019.

[145] I. D. Jonsen, T. A. Patterson, D. P. Costa, P. D. Doherty, B. J. Godley, W. J. Grecian, C. Guinet, X. Hoenner, S. S. Kienle, P. W. Robinson, S. C. Votier,

S. Whiting, M. J. Witt, M. A. Hindell, R. G. Harcourt, and C. R. McMahon, "A continuous-time state-space model for rapid quality-control of Argos locations from animal-borne tags," *arXiv*, pp. 1–13, 2020.

[146] C. M. Albertsen, "Generalizing the first-difference correlated random walk for marine animal movement data," *Scientific Reports*, vol. 9, no. 1, 2019.

[147] F. J. Pérez-Barbería, M. Small, R. J. Hooper, A. Aldezabal, R. Soriguer-Escofet, G. S. Bakken, and I. J. Gordon, "State-space modelling of the drivers of movement behaviour in sympatric species," *PLoS ONE*, vol. 10, no. 11, pp. 1–21, 2015.

[148] C. E. Vincenot, F. Giannino, M. Rietkerk, K. Moriya, and S. Mazzoleni, "Theoretical considerations on the combined use of System Dynamics and individual-based modeling in ecology," *Ecological Modelling*, vol. 222, pp. 210–218, 1 2011.

[149] M. Liedvogel, B. B. Chapman, R. Muheim, and S. Åkesson, "The behavioural ecology of animal movement: reflections upon potential synergies," *Animal Migration*, vol. 1, pp. 39–46, 2013.

[150] G. Catavitello, Y. Ivanenko, and F. Lacquaniti, "A kinematic synergy for terrestrial locomotion shared by mammals and birds," *eLife*, vol. 7, pp. 1–28, 2018.

[151] J. L. Ellis, M. Jacobs, J. Dijkstra, H. van Laar, J. P. Cant, D. Tulpan, and N. Ferguson, "Review: Synergy between mechanistic modelling and data-driven models for modern animal production systems in the era of big data," *Animal*, vol. 14, pp. s223–s237, 2020.

[152] T.-t. Yu, *The Development of a Hybrid Simulation Modelling Approach Based on Agents and Discrete-Event Modelling*. PhD thesis, University of Southampton, 2008.

[153] R. Aziza, A. Borgi, H. Zgaya, and B. Guinhouya, "Simulating Complex systems: Complex system theories, their behavioural characteristics and their simulation," *ICAART 2016 - Proceedings of the 8th International Conference on Agents and Artificial Intelligence*, vol. 2, pp. 298–305, 2016.

[154] E. P. Robertson, R. J. Fletcher, C. E. Cattau, B. J. Udell, B. E. Reichert, J. D. Austin, and D. Valle, "Isolating the roles of movement and reproduction on

effective connectivity alters conservation priorities for an endangered bird," *PNAS*, vol. 115, p. 2021, 2018.

[155] N. Pinter-Wollman, S. M. Fiore, and G. Theraulaz, "The impact of architecture on collective behaviour," 2017.

[156] J. R. Turner and R. M. Baker, "Complexity Theory: An Overview with Potential Applications for the Social Sciences," *Systems*, vol. 7, no. 1, p. 4, 2019.

[157] G. Wallentin and C. Neuwirth, "Dynamic hybrid modelling: Switching between AB and SD designs of a predator-prey model," *Ecological Modelling*, vol. 345, pp. 165–175, 2017.

[158] E. Boulding, Kenneth, "Managemen Scienc," *Management Science*, vol. 2, no. 3, pp. 197–208, 1956.

[159] D. Levy, "Applications and limitations of complexity theory in organization theory and strategy," *Public Administration and Public Policy*, no. 1 990, pp. 67–87, 2000.

[160] A. M. Law and W. D. Kelton, "Simulation Modeling and Analysis (éd. 3rd Edition)," *New York, MA : McGraw-Hill.*, 2000.

[161] H. J. Scholl, "Agent-based and system dynamics modeling: a call for cross study and joint research," *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, vol. 00, no. c, p. 8, 2001.

[162] J. Morgan, S. Howick, and V. Belton, "Designs for the complementary use of System Dynamics and Discrete-Event Simulation," *Proceedings of the 2011 Winter Simulation Conference (WSC)*, pp. 2710–2722, 12 2011.

[163] S. Franklin and A. Graesser, "Is it an agent, or just a program?: A taxonomy for autonomous agents," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1193, pp. 21–35, 1996.

[164] D. G. Brown, R. Riolo, D. T. Robinson, M. North, and W. Rand, "Spatial process and data models: Toward integration of agent-based models and GIS," *Journal of Geographical Systems*, vol. 7, pp. 25–47, 5 2005.

[165] D. C. Parker, S. M. Manson, M. A. Janssen, M. J. Hoffmann, and P. Deadman, "Multi-agent systems for the simulation of land-use and land-cover change:

A review," *Annals of the Association of American Geographers*, vol. 93, no. 2, pp. 314–337, 2003.

[166] W. Tang and D. A. Bennett, "Agent-based Modeling of Animal Movement : A Review," *Geography*, vol. 7, pp. 682–700, 2010.

[167] F. Bousquet and C. Le Page, "Multi-agent simulations and ecosystem management: a review," *Ecological Modelling*, vol. 176, pp. 313–332, 9 2004.

[168] S. F. Railsback, "Concepts from complex adaptive systems as a framework for individual-based modelling," *Ecological Modelling*, vol. 139, pp. 47–62, 3 2001.

[169] D. Helbing and S. Balietti, "How to Do Agent-Based Simulations in the Future : From Modeling Social Mechanisms to Emergent Phenomena and Interactive Systems Design Why Develop and Use Agent-Based Models ?," *Time*, no. 11-06-024, pp. 1–55, 2011.

[170] C. M. Macal, "Everything you need to know about agent-based modelling and simulation," *Journal of Simulation*, vol. 10, no. 2, pp. 144–156, 2016.

[171] S. K. Heath, S. C. Brailsford, A. Buss, and C. M. Macal, "Cross-Paradigm Simulation Modeling: Challenges and Successes," in *Proceedings of the 2011 Winter Simulation Conference*, pp. 2788–2802, 2011.

[172] V. Karbovskii, D. Voloshin, A. Karsakov, A. Bezgodov, and C. Gershenson, "Multimodel agent-based simulation environment for mass-gatherings and pedestrian dynamics," *Future Generation Computer Systems*, vol. 79, pp. 155–165, 2018.

[173] T. J. Grant, H. R. Parry, M. P. Zalucki, and S. P. Bradbury, "Predicting monarch butterfly (Danaus plexippus) movement and egg-laying with a spatially-explicit agent-based model: The role of monarch perceptual range and spatial memory," *Ecological Modelling*, vol. 374, no. September 2017, pp. 37–50, 2018.

[174] K. Rendón Rozo, J. Arellana, A. Santander-Mercado, and M. Jubiz-Diaz, "Modelling building emergency evacuation plans considering the dynamic behaviour of pedestrians using agent-based simulation," *Safety Science*, vol. 113, pp. 276–284, 2019.

[175] F. de Souza, O. Verbas, and J. Auld, "Mesoscopic traffic flow model for agent-based simulation," *Procedia Computer Science*, vol. 151, no. 2018, pp. 858–863, 2019.

[176] Z. Ding, W. Gong, S. Li, and Z. Wu, "System dynamics versus agent-based modeling: A review of complexity simulation in construction waste management," *Sustainability (Switzerland)*, vol. 10, no. 7, 2018.

[177] M. Al-Kaissy, M. Arashpour, S. Fayezi, A. Nezhad, and B. Ashuri, "Process modelling in civil infrastructure projects: A review of construction simulation methods," *Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC 2019*, no. Isarc, pp. 368–374, 2019.

[178] P. Gittins, G. McElwee, and N. Tipi, "Discrete event simulation in livestock management," *Journal of Rural Studies*, vol. 78, no. March, pp. 387–398, 2020.

[179] C. Ruiz-Martin, G. Wainer, and A. Lopez, "Discrete-Event Modeling and Simulation of Diffusion Processes in Multiplex Networks," *ACM Trans. Model. Comput. Simul*, vol. 31, no. 6, 2020.

[180] A. Borshchev and A. Filippov, "From System Dynamics to Agent Based Modeling:," *Simulation*, 2004.

[181] B. Melamed, S. Pan, and Y. Wardi, "Hybrid discrete-continuous fluid-flow simulation," *Computer*, 2001.

[182] S. Robinson, *Simulation : The Practice of Model Development and Use*. Chichester, West Sussex, England: John Wiley & Sons Ltd, 2004.

[183] P. Siebers, C. Macal, J. Garnett, D. Buxton, and M. Pidd, "Discrete-event simulation is dead, long live agent-based simulation!," *Journal of Simulation*, vol. 4, no. 3, pp. 204–210, 2010.

[184] K. V. Knyazkov and S. V. Kovalchuk, "Modeling and simulation framework for development of interactive virtual environments," *Procedia Computer Science*, vol. 29, pp. 332–342, 2014.

[185] P. Torrens, X. Li, and W. a. Griffin, "Building Agent-Based Walking Models by Machine-Learning on Diverse Databases of Space-Time Trajectory Samples," *Transactions in GIS*, vol. 15, pp. 67–94, 7 2011.

[186] C. Dorman and P. Gaudiano, "Motivation," in *Handbook of Brain Theory and Neural Networks* (M. Arbib, ed.), no. 617, ch. Motivation, pp. 1–15, Cambridge MA: MIT Press, 1998.

[187] E. Hunter, B. M. Namee, and J. Kelleher, "An open-data-driven agent-based model to simulate infectious disease outbreaks," *PLoS ONE*, vol. 13, no. 12, pp. 1–35, 2018.

[188] B. Heinzl, P. Raich, F. Preyser, and W. Kastner, "Simulation-based Assessment of Energy Efficiency in Industry: Comparison of Hybrid Simulation Approaches," *IFAC-PapersOnLine*, vol. 51, no. 2, pp. 689–694, 2018.

[189] S. C. Brailsford, T. Eldabi, M. Kunc, N. Mustafee, and A. F. Osorio, "Hybrid simulation modelling in operational research: A state-of-the-art review," *European Journal of Operational Research*, vol. 278, no. 3, pp. 721–737, 2019.

[190] G. Technitis, W. Othman, K. Safi, and R. Weibel, "From A to B, randomly: a point-to-point random trajectory generator for animal movement," *International Journal of Geographical Information Science*, vol. 8816, no. November, pp. 1–24, 2015.

[191] M. L. van Toor, B. Kranstauber, S. H. Newman, D. J. Prosser, J. Y. Takekawa, G. Technitis, R. Weibel, M. Wikelski, and K. Safi, "Integrating animal movement with habitat suitability for estimating dynamic migratory connectivity," *Landscape Ecology*, vol. 33, no. 6, pp. 879–893, 2018.

[192] C. Gaucherel, "Wavelet analysis to detect regime shifts in animal movement," *Computational Ecology and Software*, vol. 1, no. 2, pp. 69–85, 2011.

[193] A. Soleymani, F. Pennekamp, S. Dodge, and R. Weibel, "Characterizing change points and continuous transitions in movement behaviours using wavelet decomposition," *Methods in Ecology and Evolution*, vol. 8, no. 9, pp. 1113–1123, 2017.

[194] M. Teimouri, U. Indahl, H. Sickel, and H. Tveite, "Deriving Animal Movement Behaviors Using Movement Parameters Extracted from Location Data," *ISPRS International Journal of Geo-Information*, vol. 7, no. 3, p. 78, 2018.

[195] S. C. Ahearn and S. Dodge, "Recursive multi-frequency segmentation of movement trajectories (ReMuS)," *Methods in Ecology and Evolution*, vol. 9, no. 4, pp. 1075–1087, 2018.

[196] E. Gurarie, C. Bracis, M. Delgado, T. D. Meckley, I. Kojola, and C. M. Wagner, "What is the animal doing? Tools for exploring behavioral structure in animal movements," *Journal of Animal Ecology*, pp. n/a–n/a, 2015.

[197] M. Unterfinger, *3-D Trajectory Simulation in Movement Ecology : Conditional Empirical Random Walk*. PhD thesis, University of Zurich, 2018.

[198] D. Freedman and P. Diaconis, "On the histogram as a density estimator:L2 theory," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 57, no. 4, pp. 453–476, 1981.

[199] R. A. Becker, J. M. Chambers, and A. R. Wilks, "The New S Language.," *Biometrics*, 1989.

[200] L. Baroudi, M. Kawaji, and T. Lee, "Effects of initial conditions on the simulation of inertial coalescence of two drops," *Computers and Mathematics with Applications*, vol. 67, pp. 282–289, 2 2014.

[201] M. Zozulák, M. Vertal', and D. Katunský, "The influence of the initial condition in the transient thermal field simulation inside a wall," *Buildings*, vol. 9, no. 8, 2019.

[202] A. Madansky, "Optimal Initial Conditions for a Simulation Problem," *Operations Research*, vol. 24, no. 3, pp. 572–577, 1976.

[203] E. Rykiel, "Testing ecological models: the meaning of validation," *Ecological Modelling*, vol. 90, pp. 229–244, 11 1996.

[204] M. J. Slakter, "A Comparison of the Pearson Chi-Square and Kolmogorov Goodness-of-Fit Tests with Respect to Validity," *Journal of the American Statistical Association*, vol. 60, no. 311, p. 854, 1965.

[205] O. Gordo, P. Tryjanowski, J. Z. Kosicki, and M. Fulín, "Complex phenological changes and their consequences in the breeding success of a migratory bird, the white stork Ciconia ciconia," *Journal of Animal Ecology*, vol. 82, no. 5, pp. 1072–1086, 2013.

[206] J. M. Travis, K. Mustin, K. A. Bartoń, T. G. Benton, J. Clobert, M. M. Delgado, C. Dytham, T. Hovestadt, S. C. Palmer, H. Van Dyck, and D. Bonte, "Modelling dispersal: An eco-evolutionary framework incorporating emigration, movement, settlement behaviour and the multiple costs involved," *Methods in Ecology and Evolution*, vol. 3, no. 4, pp. 628–641, 2012.

[207] P. Becciu, M. H. Menz, A. Aurbach, S. A. Cabrera-Cruz, C. E. Wainwright, M. Scacco, M. Ciach, L. B. Pettersson, I. Maggini, G. M. Arroyo, J. J. Buler, D. R. Reynolds, and N. Sapir, "Environmental effects on flying migrants revealed by radar," *Ecography*, vol. 42, no. 5, pp. 942–955, 2019.

[208] F. Oloo, K. Safi, and J. Aryal, "Predicting migratory corridors of white storks, ciconia ciconia, to enhance sustainablewind energy planning: A data-driven agent-based model," *Sustainability (Switzerland)*, vol. 10, no. 5, 2018.

[209] M. Saeedimoghaddam, M. Keyanpour-Rad, H. Shafizadeh-Moghadam, R. Valavi, B. Mirbagheri, A. Shakiba, and A. Matkan, "A probabilistic space-time prism to explore changes in white Stork habitat use in Iran," *Ecological Indicators*, vol. 78, pp. 156–166, 2017.

[210] H. Shimazaki, M. Tamura, and H. Higuchi, "Migration routes and important stopover sites of endangered oriental white storks (Ciconia boyciana) as revealed by satellite tracking," *Memoirs of the National Institute of Polar Research*, vol. 58, no. Special Issue, pp. 162–178, 2004.

[211] P. Berthold, W. V. Bossche, W. Fiedler, E. Gorney, M. Kaatz, Y. Leshem, E. Nowak, and U. Querner, "Der Zug des Weißstorchs (Ciconia ciconia): Eine besondere Zugform auf Grund neuer Ergebnisse," *Journal fur Ornithologie*, vol. 142, no. 1, pp. 73–92, 2001.

[212] M. Clairbaux, J. Fort, P. Mathewson, W. Porter, H. Strøm, and D. Grémillet, "Climate change could overturn bird migration: Transarctic flights and high-latitude residency in a sea ice free Arctic," *Scientific Reports*, vol. 9, no. 1, pp. 1–13, 2019.

[213] B. M. Van Doren, K. G. Horton, A. M. Dokter, H. Klinck, S. B. Elbin, and A. Farnsworth, "High-intensity urban light installation dramatically alters nocturnal bird migration," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 42, pp. 11175–11180, 2017.

[214] F. Harrell, "Vanderbilt Department of Biostatistics," *Problems Caused by Categorizing Continuous Variables Retrieved from: http://biostat.mc.vanderbilt.edu/wiki/Main/CatContinuous*, 2019.

[215] W. N. Venables and B. D. Ripley, "Modern Applied Statistics with S (fourth.). New York: Springer.," *Retrieved from http://www.stats.ox.ac.uk/pub/MASS4*, 2002.

[216] K. Didan, "MOD13C2 MODIS/Terra Vegetation Indices Monthly L3 Global 0.05Deg CMG V006," 2015.

[217] A. Schneider, A. Jost, C. Coulon, M. Silvestre, S. Théry, and A. Ducharne, "Global-scale river network extraction based on high-resolution topography and constrained by lithology, climate, slope, and observed drainage density," *Geophysical Research Letters*, 2017.

[218] C. Loehle, "Hypothesis testing in ecology: psychological aspects and the importance of theory maturation," *Quarterly Review of Biology*, vol. 62, no. 4, pp. 397–409, 1987.

[219] A. Crooks, C. Castle, and M. Batty, "Key challenges in agent-based modelling for geo-spatial simulation," *Computers, Environment and Urban Systems*, vol. 32, no. 6, pp. 417–430, 2008.

[220] S. R. X. Dall, L.-A. Giraldeau, O. Olsson, J. M. McNamara, and D. W. Stephens, "Information and its use by animals in evolutionary ecology.," *Trends in ecology & evolution*, vol. 20, pp. 187–93, 4 2005.

[221] L. Börger, "Stuck in motion? Reconnecting questions and tools in movement ecology," *Journal of Animal Ecology*, vol. 85, no. 1, pp. 5–10, 2016.

[222] C. R. du Feu, J. A. Clark, M. Schaub, W. Fiedler, and S. R. Baillie, "The EURING Data Bank – a critical tool for continental-scale studies of marked birds," *Ringing and Migration*, vol. 31, no. 1, pp. 1–18, 2016.

[223] B. Kuijpers and G. Technitis, "Space-time prisms on a sphere with applications to long-distance movement," *International Journal of Geographical Information Science*, vol. 34, no. 10, pp. 1980–2003, 2020.

[224] J. Berger, J. K. Young, and K. M. Berger, "Protecting migration corridors: Challenges and optimism for Mongolian saiga," 2008.

[225] T. a. Patterson, L. Thomas, C. Wilcox, O. Ovaskainen, and J. Matthiopoulos, "State-space models of individual animal movement.," *Trends in ecology & evolution*, vol. 23, pp. 87–94, 2 2008.

[226] M. F. McCracken and C. S. Peskin, "A vortex method for blood flow through heart valves," *Journal of Computational Physics*, vol. 35, no. 2, pp. 183–205, 1980.

[227] A. J. Chorin, “Vortex sheet approximation of boundary layers,” *Journal of Computational Physics*, vol. 27, no. 3, pp. 428–442, 1978.

[228] B. K. Patle, A. Pandey, A. Jagadeesh, and D. R. Parhi, “Path planning in uncertain environment by using firefly algorithm,” *Defence Technology*, vol. 14, no. 6, pp. 691–701, 2018.

[229] E. C. M. Parsons, “Why IUCN Should Replace “Data Deficient” Conservation Status with a Precautionary "Assume Threatened” Status—A Cetacean Case Study,” *Frontiers in Marine Science | www.frontiersin.org*, vol. 3, p. 193, 2016.

[230] A. Aurbach, B. Schmid, F. Liechti, N. Chokani, and R. Abhari, “Complex behaviour in complex terrain - Modelling bird migration in a high resolution wind field across mountainous terrain to simulate observed patterns,” *Journal of Theoretical Biology*, vol. 454, pp. 126–138, 2018.

[231] J. Minguez, F. Lamiraux, and J.-P. Laumond, “Motion Planning and Obstacle Avoidance,” in *Springer Handbook of Robotics* (B. Siciliano and O. Khatib, eds.), pp. 1177–1202, Cham: Springer International Publishing, 2016.

[232] S. Cheng, X. Wang, and H. Zhang, “Application of Firefly Algorithm to UWB Indoor Positioning,” *2021 IEEE 11th Annual Computing and Communication Workshop and Conference, CCWC 2021*, no. 1, pp. 88–93, 2021.

[233] C. Pasquaretta, T. Dubois, T. Gomez-Moracho, V. P. Delepoulle, G. Le Loc’h, P. Heeb, and M. Lihoreau, “Analysis of temporal patterns in animal movement networks,” *Methods in Ecology and Evolution*, vol. 12, no. 1, pp. 101–113, 2021.

[234] M. Thums, J. Fernández-Gracia, A. M. Sequeira, V. M. Eguíluz, C. M. Duarte, and M. G. Meekan, “How big data fast tracked human mobility research and the lessons for animal movement ecology,” 2018.

[235] T. M. Cheng and A. V. Savkin, “Decentralized control of multi-agent systems for swarming with a given geometric pattern,” *Computers and Mathematics with Applications*, vol. 61, no. 4, pp. 731–744, 2011.

[236] A. Voinov and F. Bousquet, “Modelling with stakeholders,” *Environmental Modelling and Software*, 2010.

[237] A. H. Levis, “Simple versus realistic modeling,” *Computational and Mathematical Organization Theory*, vol. 15, no. 1, pp. 5–7, 2009.

[238] D. O'Sullivan, T. Evans, S. Manson, S. Metcalf, A. Ligmann-Zielinska, and C. Bone, "Strategic directions for agent-based modeling: avoiding the YAAWN syndrome," *Journal of Land Use Science*, vol. 11, no. 2, pp. 177–187, 2015.

[239] V. Grimm, U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, A. Huth, J. U. Jepsen, C. Jørgensen, W. M. Mooij, B. Müller, G. Pe'er, C. Piou, S. F. Railsback, A. M. Robbins, M. M. Robbins, E. Rossmanith, N. Rüger, E. Strand, S. Souissi, R. A. Stillman, R. Vabø, U. Visser, and D. L. DeAngelis, "A standard protocol for describing individual-based and agent-based models," *Ecological Modelling*, vol. 198, no. 1-2, pp. 115–126, 2006.

[240] V. Grimm, U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback, "The ODD protocol: A review and first update," *Ecological Modelling*, vol. 221, no. 23, pp. 2760–2768, 2010.

Appendix

A Info

The work presented in this Thesis was conducted using the following R version and packages.

A.1 R session Info & Packages

```
R version 3.6.0 (2019-04-26)
Platform: x86_64-apple-darwin15.6.0 (64-bit)
Running under: macOS  10.16

Matrix products: default
LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] grid      stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] ggnewscale_0.4.5    eRTG3D_0.6.0        randomcoloR_1.1.0.1 lubridate_1.7.4     ggimage_0.2.5       tmaptools_3.1       scales_1.0.0
 [8] ggmap_3.0.0         sf_0.9-5            forcats_0.4.0       stringr_1.4.0       purrr_0.3.2         readr_1.3.1         tidyr_1.1.0
[15] tibble_2.1.3        ggplot2_3.3.3       tidyverse_1.2.1     osmdata_0.1.2       VIM_4.8.0           data.table_1.12.2   colorspace_1.4-1
[22] dplyr_0.8.5         here_0.1            raster_2.9-5        sp_1.3-2

loaded via a namespace (and not attached):
 [1] Rtsne_0.15          rjson_0.2.20        class_7.3-15        rio_0.5.16          rsconnect_0.8.16    rprojroot_1.3-2     dichromat_2.0-0
 [8] rstudioapi_0.10     ranger_0.11.2       xml2_1.2.0          codetools_0.2-16    robustbase_0.93-5   knitr_1.23          jsonlite_1.6
[15] broom_0.5.2         cluster_2.1.0       png_0.1-7           BiocManager_1.30.10 compiler_3.6.0      httr_1.4.0          rvcheck_0.1.7
[22] backports_1.1.4     assertthat_0.2.1    Matrix_1.2-17       cli_1.1.0           htmltools_0.3.6     tools_3.6.0         gtable_0.3.0
[29] glue_1.3.1          V8_2.3              Rcpp_1.0.1          carData_3.0-3       cellranger_1.1.0    vctrs_0.3.0         nlme_3.1-139
[36] lmtest_0.9-37       lwgeom_0.1-7        xfun_0.8            laeken_0.5.0        openxlsx_4.1.4      rvest_0.3.4         lifecycle_0.2.0
[43] XML_3.98-1.20       DEoptimR_1.0-8      MASS_7.3-51.4       zoo_1.8-7           hms_0.5.0           parallel_3.6.0      RColorBrewer_1.1-2
[50] yaml_2.2.0          curl_3.3            stringi_1.4.3       e1071_1.7-2         boot_1.3-22         zip_2.0.4           RgoogleMaps_1.4.5
[57] rlang_0.4.6         pkgconfig_2.0.2     bitops_1.0-6        evaluate_0.14       lattice_0.20-38     tidyselect_1.1.0    plyr_1.8.4
[64] magrittr_1.5        R6_2.4.0            magick_2.2          generics_0.0.2      DBI_1.0.0           pillar_1.4.2        haven_2.1.1
[71] foreign_0.8-71      withr_2.1.2         units_0.6-5         stars_0.4-3         abind_1.4-5         nnet_7.3-12         modelr_0.1.4
[78] crayon_1.3.4        car_3.0-5           KernSmooth_2.23-15  rmarkdown_2.1       jpeg_0.1-8.1        readxl_1.3.1        vcd_1.4-4
[85] digest_0.6.19       classInt_0.4-3      gridGraphics_0.4-1  munsell_0.5.0       viridisLite_0.3.0   ggplotify_0.0.4
```

A.2 eRTG3D package

The core algorithm of the methodology was coded, extended and released in the eRTG3D package `https://github.com/munterfi/eRTG3D` by Merlin Unterfinger.

The eRTG3D package contains functions to:

- calculate movement parameters of 3-D GPS tracking data, turning angle, lift angle and step length
- extract distributions from movement parameters;
- P probability - The mover's behaviour from its perspective
- Q probability - The pull towards the target
- simulate Unconditional Empirical Random Walks (UERW)
- simulate Conditional Empirical Random Walks (CERW)
- simulate conditional gliding and soaring behaviour of birds between two given points
- statistically test the simulated tracks against the original input
- visualise tracks, simulations and distributions in 3-D and 2-D
- conduct a basic point cloud analysis; extract 3-D Utilisation Distributions (UDs) from observed or simulated tracking data by means of voxel counting
- project 3-D tracking data into different Coordinate Reference Systems (CRSs)
- export data to sf package objects; 'sf, data.frames'
- manipulate extent of raster layers

B List of publications

Journal articles

Technitis, G., Unterfinger, M., Kranstauber, B., Weibel, R. & Safi, K. (in preparation). Generating random trajectories based on empirical paths.

Kuijpers, B., & **Technitis, G.** (2020). Space-time prisms on a sphere with applications to long-distance movement. International Journal of Geographical Information Science, 34(10), 1980–2003. https://doi.org/10.1080/13658816.2020.1738439

van Toor, M. L., Kranstauber, B., Newman, S. H., Prosser, D. J., Takekawa, J. Y., **Technitis, G.**, Weibel, R., Wikelski, M. & Safi, K. (2018). Integrating animal movement with habitat suitability for estimating dynamic migratory connectivity. Landscape Ecology, 33(6), 879–893. https://doi.org/10.1007/s10980-018-0637-9

Technitis, G., Othman, W., Safi, K., & Weibel, R. (2015). From A to B, randomly: a point-to-point random trajectory generator for animal movement. International Journal of Geographical Information Science, 29(6), 912-934. https://doi.org/10.1080/13658816.2014.999682

Conference papers

Technitis, G., Weibel, R., Kranstauber, B., & Safi, K. (2016). An Algorithm for Empirically Informed Random Trajectory Generation Between Two Endpoints. In International Conference on GIScience Short Paper Proceedings (Vol. 1, pp. 292–295). https://doi.org/10.21433/b31194g0c634

Technitis, G., & Weibel, R. (2014). An Algorithm for Random Trajectory Generation Between Two Endpoints, Honoring Time and Speed Constraints. In Proceedings GIScience 2014 (pp. 88–92). https://doi.org/10.5167/uzh-104468

Technitis, G., & Weibel, R. (2012). A hybrid simulation model for moving objects. In: The 2012 AutoCarto International Symposium on Automated Cartography, Columbus, Ohio, USA, 16 September 2012 - 18 September 2012, online. https://doi.org/10.5167/uzh-74665