

Department of Data Analysis and Mathematical Modelling

**Analysis and modeling of biological movement
phenomena based on telemetric observations**

ir. Steffie Van Nieuland

Thesis submitted in fulfillment of the requirements for the degree of
Doctor (Ph.D.) of Applied Biological Sciences

Academic year 2018-2019

Supervisors:

Prof. dr. Bernard De Baets
Department of Data Analysis
and Mathematical Modelling
Ghent University, Belgium

Prof. dr. ir. Jan M. Baetens
Department of Data Analysis
and Mathematical Modelling
Ghent University, Belgium

Examination committee:

Prof. dr. ir. Wim Verbeke (Chairman)

Prof. dr. ir. Peter Goethals
Prof. dr. Nico Van de Weghe
Dr. Urška Demšar
Dr. ir. Stijn Van Hoey

Dean:

Prof. dr. ir. Marc Van Meirvenne

Rector:

Prof. dr. ir. Rik Van de Walle

ir. Steffie Van Nieuland

Analysis and modeling of biological movement phenomena
based on telemetric observations

Thesis submitted in fulfillment of the requirements for the degree of

Doctor (Ph.D) of Applied Biological Sciences

Academic year 2018-2019

Dutch translation of the title:

Analyse en modellering van biologische beweging op basis van telemetriegegevens

Please refer to this work as follows:

S. Van Nieuland (**2018**). Analysis and modeling of biological movement phenomena based on telemetric observations, PhD Thesis, Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium.

The author and the supervisors give the authorization to consult and to copy parts of this work for personal use only. Every other use is subject to the copyright laws. Permission to reproduce any material contained in this work should be obtained from the author.

Dankwoord

Eerst en vooral wil ik mijn promotoren prof. Bernard De Baets en prof. Jan Baetens uitvoerig bedanken voor hun onderzoeksideeën, hun kritische blik en de vele uren die zij gespendeerd hebben aan het nalezen van mijn teksten. Van harte dankuwel voor deze leerrijke rit.

Verder wil ik de organisatie van de Dodentocht, René Janssen en Zachery Ladin bedanken voor het delen van hun tracking data, welke noodzakelijk waren om dit onderzoek uit te voeren. Dank jullie allen voor de interesse, het enthousiasme en de behulpzaamheid die jullie in onze professionele samenwerking getoond hebben.

Alsook wil ik alle (ex-) collega's en in het bijzonder mijn (ex-) bureaugenoten bedanken voor de goede werksfeer. Ik kon steeds bij hun terecht voor het stellen van uiteenlopende onderzoeks- en onderwijsvragen, het uitwisselen van ervaringen en voor een ontspannend praatje.

Tot slot wil ik mijn familieleden en vrienden bedanken voor hun steun en interesse. Het heeft me deugd gedaan. In het bijzonder wil ik Joris bedanken voor de ruimte die hij me gaf om dit werk tot een goed einde te brengen. De ontspannende momenten die hij en Arne me gaven, waren een zaligheid!

Dank aan allen die deze rit samen met mij hebben beleefd.

Steffie Van Nieuland
Gent, oktober 2018

Contents

| | |
|--------------------------------------|--------------|
| Dankwoord | v |
| Contents | vii |
| Summary | xiii |
| Nederlandstalige samenvatting | xvii |
| List of acronyms | xxi |
| List of Figures | xxiii |
| List of Tables | xxxix |
| 1 Introduction | 1 |
| 1.1 Overview | 1 |
| 1.2 Research objectives | 2 |
| 1.3 Roadmap | 3 |
| 2 Movement ecology background | 5 |
| 2.1 Introduction | 5 |

| | | |
|----------|--|-----------|
| 2.2 | Movement | 6 |
| 2.3 | Tracking devices | 8 |
| 2.3.1 | VHF radio, satellite and GPS tracking techniques | 9 |
| 2.3.2 | Other tracking techniques | 13 |
| 2.4 | Tracking data | 15 |
| 2.5 | Tracking data analysis | 18 |
| 2.5.1 | Visualization | 18 |
| 2.5.2 | Quantifying movement | 18 |
| 2.5.3 | Quantifying space use | 20 |
| 2.5.4 | Movement and context | 22 |
| 2.6 | Movement modeling | 25 |
| 2.6.1 | Introduction | 25 |
| 2.6.2 | Continuous time–continuous space models | 26 |
| 2.6.3 | Discrete time–continuous space models | 28 |
| 2.6.4 | Discrete space models | 30 |
| 2.6.5 | Behavioral states | 30 |
| 2.7 | Movement simulation and prediction | 31 |
| 2.8 | Management implications | 32 |
| 3 | Modeling the event ‘100 km Dodentocht’ | 35 |
| 3.1 | Introduction | 35 |
| 3.2 | Data collection and overview | 37 |
| 3.3 | Data analysis | 40 |
| 3.3.1 | RFID tracking data | 40 |
| 3.3.1.1 | Overall statistics | 40 |
| 3.3.1.2 | Correlation analysis | 45 |
| 3.3.2 | GPS tracking data | 50 |
| 3.4 | A spatially explicit marching model | 52 |
| 3.4.1 | Overview | 52 |
| 3.4.2 | Conditional distribution of the marching speed | 53 |
| 3.4.3 | Hazard function | 54 |

| | | |
|----------|---|-----------|
| 3.4.4 | Model simulation | 54 |
| 3.4.5 | Model selection | 55 |
| 3.5 | Model validation | 62 |
| 3.5.1 | Cross-validation | 62 |
| 3.5.2 | Validation with an independent data set | 66 |
| 3.6 | Scenario analysis | 71 |
| 3.6.1 | The copula of starting speed and age | 72 |
| 3.6.2 | The marginal distribution of the starting speed and its modifications | 72 |
| 3.6.2.1 | Varying the mean starting speed of marchers | 74 |
| 3.6.2.2 | Varying the standard deviation of the starting speed of marchers | 77 |
| 3.6.3 | Scenarios | 80 |
| 3.6.3.1 | Encoding | 80 |
| 3.6.3.2 | Analysis | 81 |
| 3.7 | Discussion | 84 |
| 4 | Analytical description of the TIBB | 89 |
| 4.1 | Introduction | 89 |
| 4.2 | Preliminaries | 91 |
| 4.3 | The one-dimensional TIBB | 94 |
| 4.3.1 | Its analytical description | 94 |
| 4.3.1.1 | The inner region: $0 < \beta < 1$ | 94 |
| 4.3.1.2 | The tails: $\beta < 0$ and $1 < \beta$ | 97 |
| 4.3.1.3 | The cases $\beta = 0$ and $\beta = 1$ | 98 |
| 4.3.1.4 | Normalization | 99 |
| 4.3.2 | Proof of symmetry | 101 |
| 4.3.3 | Distribution parameters | 101 |
| 4.3.3.1 | Expected value | 101 |
| 4.3.3.2 | Median and mode | 103 |
| 4.3.3.3 | Variance | 103 |

| | | |
|----------|---|------------|
| 4.3.3.4 | Skewness | 104 |
| 4.3.3.5 | Kurtosis | 105 |
| 4.4 | The two-dimensional TIBB | 106 |
| 4.4.1 | No dependence between both components, a starting location situated at the origin and an ending location situated along the horizontal axis | 106 |
| 4.4.1.1 | Normalization | 107 |
| 4.4.1.2 | Marginal distributions | 108 |
| 4.4.2 | No dependence between both components and a starting loca- tion situated at the origin | 110 |
| 4.4.3 | The two-dimensional TIBB without dependence between both components | 111 |
| 4.4.4 | The two-dimensional TIBB with dependence between both com- ponents | 112 |
| 4.5 | The analytical description of the <i>n</i> -dimensional TIBB | 114 |
| 4.6 | Time efficiency | 115 |
| 4.6.1 | Synthetic experiment | 115 |
| 4.6.2 | Real-world dataset | 116 |
| 4.7 | Discussion | 119 |
| 5 | Possible sources of poisoning for Eagle Owls in Limburg | 123 |
| 5.1 | Introduction | 123 |
| 5.2 | Study area and data set | 125 |
| 5.3 | The Eurasian Eagle Owl | 127 |
| 5.3.1 | Distribution and ecology | 127 |
| 5.3.2 | Prey species | 131 |
| 5.4 | A validated expert-based habitat suitability map | 132 |
| 5.4.1 | Introduction | 132 |
| 5.4.2 | Methods | 132 |
| 5.4.2.1 | Construction | 132 |
| 5.4.2.2 | Validation and interpretation | 135 |
| 5.4.3 | Results | 137 |

| | | |
|----------|--|------------|
| 5.4.3.1 | The habitat suitability map | 137 |
| 5.4.3.2 | The probability density of occurrence | 140 |
| 5.4.3.3 | Validation and interpretation | 142 |
| 5.4.4 | Discussion | 148 |
| 5.5 | Environmental characteristics and locations of possible hunting grounds | 150 |
| 5.5.1 | Introduction | 150 |
| 5.5.2 | Methods | 151 |
| 5.5.2.1 | Selecting fixes linked to possible hunting grounds | 151 |
| 5.5.2.2 | Identifying the important areas | 152 |
| 5.5.2.3 | Identifying the important environmental characteristics | 153 |
| 5.5.3 | Results | 155 |
| 5.5.3.1 | Selecting fixes linked to possible hunting grounds | 155 |
| 5.5.3.2 | Identifying the important areas | 157 |
| 5.5.3.3 | Identifying the important environmental characteristics | 159 |
| 5.5.4 | Discussion | 162 |
| 5.6 | Comparison | 164 |
| 6 | Contrasted post-fledging space use by Neotropical migratory birds within an urbanized landscape | 167 |
| 6.1 | Introduction | 167 |
| 6.2 | Study area and data set | 169 |
| 6.3 | The Gray Catbird and the Wood Thrush | 170 |
| 6.4 | Determining land use | 173 |
| 6.4.1 | Methods | 173 |
| 6.4.2 | Results | 175 |
| 6.4.3 | Discussion | 180 |
| 6.5 | Modeling land use | 182 |
| 6.5.1 | Relevant land cover classes | 182 |
| 6.5.2 | Spatially explicit movement model | 183 |
| 6.5.2.1 | Overview | 183 |

| | | |
|----------|--|------------|
| 6.5.2.2 | Suitability index | 189 |
| 6.5.2.3 | Decision rule | 194 |
| 6.5.3 | Model selection | 196 |
| 6.5.3.1 | Introduction | 196 |
| 6.5.3.2 | Gray Catbird | 197 |
| 6.5.3.3 | Wood Thrush | 199 |
| 6.5.3.4 | Variability among simulations | 202 |
| 6.5.4 | Model performance and cross-validation | 202 |
| 6.5.4.1 | Introduction | 202 |
| 6.5.4.2 | Gray Catbird | 204 |
| 6.5.4.3 | Wood Thrush | 205 |
| 6.5.5 | Discussion | 207 |
| 7 | Conclusions and perspectives | 211 |
| | Bibliography | 221 |
| | Curriculum vitae | 245 |

Summary

Habitat fragmentation and loss, depletion of natural resources and climate change are currently the biggest challenges faced by plants and animals in a world dominated by humans. Nature conservation is crucial to preserve and restore nature's beauty. Development of successful management plans demand in-depth knowledge of the habitats and the species occupying it. Considering knowledge related to how, where, when and why certain locations are visited by animals, is crucial. As such, one can take into account necessary food resources, resting areas, nesting sites, connectivity between nature reserves, possible barriers and so on. In addition to the ecological importance of studying movement phenomena, there is also an economical aspect to it since animals spread pollen, nutrients and diseases whilst moving through a landscape.

In this dissertation, movement is studied to gain knowledge on how, where, when and why individuals move. This study is based upon tracking data. In order to collect this type of data, a tracking device is attached to the individual, which registers its consecutive locations in time. The focus of this research is on animals, but also people are constantly moving. Analysis of human tracking data can be used to manage traffic and guarantee safety in daily life or during (mass) events. Furthermore, tracking data is used for military purposes and for improving sports performances.

This dissertation consists of one more theoretical chapter, involving the analytical description of the time-integrated Brownian bridge (TIBB), and three chapters of a more applied nature, where tracking data sets were used to help answering specific research questions. Each data set contains data of a different type and considers a different species, region and research question. Based upon the available data and research question, appropriate methods and models have been

selected. This selection was not straightforward considering the wide variety of methods and models available. The challenge was in finding the most appropriate ones that extract efficiently the maximum amount of information. In addition to the analysis of the tracking data sets, a mathematical model was built for two of the three studied species, which simulates the consecutive locations of an individual in time. These models helped to summarize the results of a data analysis and to predict the effect of several possible scenarios.

The TIBB, which is studied in detail, is used to determine the intensity of use of an area on the basis of tracking data. Based on this information, important areas to gather food, to find shelter, for migration, and so on, can be identified. In contrast to other methods, where the intensity of use is solely based on the observed locations, the TIBB also includes the time in between observations and the mobility of the considered species, which are essential to make an accurate estimation of the intensity of use. More specifically, the analytical description of the TIBB was derived, as such, improving its computational efficiency. Moreover, the TIBB was used in our study of two out of three considered tracking data sets.

The focus of the dissertation lies on the movement of animals. The very first data set, however, considers people. This data set contains observations of the participants of the '100 km Dodentocht'. This mass event, which takes place annually and has its start in Bornem, Belgium, is a long distance march where participants have to cover a 100 km trail in at most 24 hours. The analysis of this '100 km Dodentocht' was a unique opportunity in the sense that we were able to work in a one-dimensional setting, which is normally not possible for (most) animal tracking data sets. The locations of the checkpoints and time when marchers pass by every checkpoint were available for the editions 2009–2014. These tracking data were analyzed to build a model that can simulate when marchers pass by a certain checkpoint and where marchers retire, based on their age and gender. By simulating a large number of individuals, one can acquire insight into the dynamics of the event. This model was used to predict the effect of a modified starting procedure. To this purpose, three scenarios were tested in order to reduce the crowiness at the start of the event. It was concluded that either the start should be moved to a location outside the town center, where streets are at least 25% wider, or that the marchers should start in two groups at two different locations, and that these groups should ideally merge at about 20 km from the start. The presented framework can be used for analyzing and improving other such events.

The second data set contains tracking data of three pairs of Eurasian Eagle Owls (*Bubo bubo*) that live in three adjoining territories in Zuid-Limburg, the Netherlands. Since the population size had stagnated and extremely high concentrations of poisonous chemical compounds (PCBs and DDEs) were found in owl carcasses, it is suspected that something is jeopardizing the further flourishing of the species in the region. In order to indicate possible sources of pollution, environmental characteristics (ECs) and areas important for the species have been highlighted so

that they can be further investigated and protected. First, a habitat suitability map was constructed based on literature and expert knowledge. Here, the suitability of each selected EC was quantified. The presence or absence of the selected ECs in an area determine the overall suitability of this area. Finally, this allows to identify the most important areas for the owls. In this approach, tracking data have been used to interpret and validate the habitat suitability map.

In a second approach, the tracking data themselves were used to identify important ECs and areas. First, the size of the tracking data set was reduced by eliminating the observations that are probably not related to hunting. Subsequently, the most frequently visited locations were identified by means of a so-called hotspot analysis and the most important ECs were pinpointed. Both approaches demonstrate that the former quarries in the region are very important for the Eurasian Eagle Owls since they are used as nesting sites. Other ECs that attract the species are pastures, differences in altitude within the landscape, forest edges and water bodies. Major roads and town centers will most likely be avoided. Areas and ECs that were highlighted can be used to help prioritizing areas that need to be protected. Furthermore, they can help in setting up a sampling strategy to further investigate possible sources of pollution since samples are best taken in areas that attract the species.

The third and last data set contains tracking data of two American migratory bird species, the Gray Catbird (*Dumatella caroliniensi*) and the Wood Thrush (*Hylocichla mustelina*), which are both studied during the post-fledging period. Locations have been registered on a daily basis upon abandoning of the nest. These two species, while being phylogenetically closely related, share similar life-history traits, though exhibit differential responses to urbanization. The Wood Thrush avoids urban areas, whereas the Gray Catbird is adapted to life in a landscape dominated by people, which results in opposite population trends. In order to estimate the effect of urbanization on post-fledging space use, the use of forest and roads by the bird species, two land uses that determine the degree of urbanization, were investigated. We may conclude that the Gray Catbird spends less time in forests and more time near roads than Wood Thrush, which matches the expectations. Finally, the results of the data analysis were summarized in a model to simulate consecutive locations of individuals. Given the simplicity of the chosen model structure, the performance of the model can be evaluated as good. However, more knowledge should be included in order to better approximate the real behavior of the post-fledgling Gray Catbird and Wood Thrush.

In summary, one theoretical chapter and three chapters of a more applied nature are elaborated in this dissertation. For what concerns the latter, several methods and models helped answering the research questions at stake. In this way, this research offers an added value to the involved parties, it may serve as inspiration to future research with similar research questions and, foremost, it contributes to an improved management and conservation of the environment we live in.

Nederlandstalige samenvatting

Habitatfragmentatie en -verlies, uitgeputte natuurlijke hulpbronnen en klimaatverandering zijn de grootste uitdagingen die planten en dieren momenteel moeten aangaan in een door mensen gedomineerde Aarde. Natuurbeheer is cruciaal voor het behoud en herstel van natuurlijk schoon. Het opstellen van succesvolle beheerplannen vereist kennis van de omgeving en de soorten die erin leven. Het in acht nemen van kennis gerelateerd aan het hoe, waar, wanneer en waarom opeenvolgende locaties door dieren bezocht worden, is hierbij cruciaal. Zo kan onder meer rekening gehouden worden met de nodige voedselbronnen, rustplaatsen, nestgelegenheden, verbindingen tussen natuurgebieden, mogelijke barrières, en zo verder. Naast het ecologisch belang heeft het onderzoek naar beweging bij dieren, tevens ook een economisch belang aangezien zij pollen, nutriënten en ziektes verspreiden terwijl ze door het landschap bewegen.

In dit doctoraat wordt beweging bestudeerd om kennis te vergaren gerelateerd aan het hoe, waar, wanneer en waarom individuen bewegen. Dit gebeurt aan de hand van tracking data. Voor het verzamelen van dit type data wordt een apparaat aan een individu bevestigd waarmee de opeenvolgende locaties van het individu doorheen de tijd geregistreerd worden. De focus van dit onderzoek ligt bij dieren, maar ook mensen zijn voortdurend in beweging. Analyse van menselijke bewegingsdata kan gebruikt worden voor het sturen van het verkeer en voor het garanderen van de veiligheid in het dagdagelijkse leven of tijdens (massa)evenementen. Verder worden tracking data gebruikt voor militaire doeleinden en voor het verbeteren van sportprestaties.

Dit doctoraatonderzoek bestaat uit één theoretisch hoofdstuk waarbij de analytische beschrijving van de tijdsgéïntegreerde Brownse brug (TIBB) werd afgeleid en uit drie meer praktische hoofdstukken waarbij telkens een tracking dataset beschikbaar was om de gestelde onderzoeksvragen te helpen beantwoorden. Elke dataset bevat data van een ander type en betreft een andere omvang, bestudeerde soort, regio en bijhorende onderzoeksvraag. Op basis van de beschikbare data en de onderzoeksvraag, werden geschikte methodes en modellen geselecteerd. Deze selectie was niet evident aangezien er een hele waaier aan methodes en modellen beschikbaar is. De uitdaging lag in het vinden van een zo efficiënt mogelijke aanpak waarmee een maximum aan informatie kan geëxtraheerd worden. Naast de analyse van de datasets, werd bij twee van de drie bestudeerde soorten een wiskundig model opgesteld dat toelaat de opeenvolgende locaties van een individu doorheen de tijd te simuleren. Deze modellen hielpen om de resultaten van de data-analyse samen te vatten en om het effect van verschillende scenario's in te schatten.

De TIBB, een methode die wordt gebruikt om de intensiteit waarmee elk van de delen van een studiegebied gebruikt/bezocht worden, te bepalen aan de hand van tracking data, wordt in dit werk in detail beschouwd. Op basis van de gebruiksin-tensiteit, die uitgedrukt wordt als een probabiliteitsdichtheidsfunctie (PDF), kunnen belangrijke gebieden voor het vergaren van voedsel, het vinden van beschut-ting, migratie, en zo verder, aangeduid worden. In tegenstelling tot andere methodes, waarbij de bepaling van de PDF enkel gebaseerd is op de geobserveerde locaties, wordt bij de TIBB ook de tijd tussen de observaties en de beweeglijkheid van de beschouwde soort in rekening gebracht, wat essentieel is om een goede in-schatting te maken. Meer specifiek wordt in deze thesis de analytische uitdrukking van de TIBB afgeleid waardoor de computationele efficiëntie sterk verbeterd wordt. De TIBB wordt verder gebruikt bij de analyse van twee van de drie bestudeerde datasets.

Ondanks de focus van dit doctoraat op de beweging van dieren, worden er in de eerste dataset mensen beschouwd. Deze dataset bevat observaties van deel-nemers aan de '100km Dodentocht'. Dit massaevenement, waar deelnemers 100 km dienen af te leggen in 24 uur, vindt elk jaar plaats in Bornem, België. De '100 km Dodentocht' bood een mooie gelegenheid om in één dimensie te kunnen werken, wat meestal niet mogelijk is bij dieren. De locaties van de controleposten en doorkomsttijden van de deelnemers aan de controleposten waren beschikbaar voor de edities 2009–2014. Deze gegevens werden geanalyseerd om een model op te stellen dat de opgave en doorkomsttijden van een deelnemer kan simuleren op basis van zijn/haar leeftijd en geslacht. Bij het simuleren van een groot aantal individuen kan men zo een inzicht verwerven in de dynamiek van het evenement. Dit model werd gebruikt om het effect van een gewijzigde startprocedure te voor-spellen. Hiertoe werden drie scenario's getest waarmee de drukte gereduceerd zou kunnen worden. Er werd besloten dat de start beter buiten het centrum zou

liggen, waar de straten minstens 25% breder zijn, of dat men zou kunnen starten op twee verschillende locaties waarbij de twee groepen pas samenkomen 20 km na de start. Het ontwikkelde model zou tevens voor andere massa-evenementen kunnen gebruikt worden om zo een goed verloop en de veiligheid van de aanwezigen te kunnen garanderen.

De tweede dataset bevat tracking data van drie oehoekoppels (*Bubo bubo*) die leven in drie aan elkaar grenzende territoria in Zuid-Limburg, Nederland. Doordat de populatie er in de omgeving niet meer aandikte en hoge PCB-concentraties teruggevonden werden in oehoekarkassen, vermoedt men dat iets de verdere toename van de oehoepopulatie in de regio hypothekeert. Om de mogelijke bronnen van vervuiling aan te duiden, werden de omgevingskarakteristieken en gebieden die belangrijk zijn voor de soort aangeduid met het oog op verder onderzoek en bescherming. Eerst compileerden we een habitatgeschiktheidskaart op basis van literatuur en expertkennis. Hierbij werd eerst de geschiktheid van elke geselecteerde omgevingskarakteristiek gekwantificeerd. De aan- of afwezigheid van de geselecteerde omgevingskarakteristieken in een beschouwd gebied in de regio bepalen dan de globale geschiktheid van dit gebied. Op basis van de globale geschiktheid van elk gebied in de regio konden vervolgens de belangrijkste gebieden voor de uilen aangeduid worden. In deze benadering werden de tracking data enkel gebruikt voor het interpreteren van de resultaten en voor het valideren van de habitatgeschiktheidskaart.

Volgens een tweede benadering werden de tracking data zelf gebruikt om belangrijke omgevingskarakteristieken en gebieden aan te duiden. Hierbij werd de tracking dataset eerst gereduceerd door de observaties die met een grote waarschijnlijkheid niet gerelateerd zijn aan voedselvergaring te verwijderen. Vervolgens werden de belangrijkste gebieden aangeduid door middel van een zogenoemde hotspot analyse, en werden de belangrijkste omgevingskarakteristieken aangeduid. Uit beide benaderingen blijkt dat de voormalige groeves in de regio zeer belangrijk zijn voor de oehoes. Dit komt doordat deze gebruikt worden als nestlocaties. Andere omgevingskarakteristieken die de soort aantrekken zijn weides, hoogteverschillen in het landschap, bosranden en wateroppervlaktes. Drukke wegen en dorpskernen worden eerder vermeden. De gebieden en omgevingskarakteristieken die aangeduid werden als belangrijk kunnen gebruikt worden om percelen te selecteren die opgenomen dienen te worden in beheerplannen. Verder kunnen ze richtinggevend zijn bij staalnamecampagnes. Om mogelijke vervuilingbronnen verder te onderzoeken, worden stalen best genomen in gebieden die de soort aantrekken.

De derde en laatste dataset bevat tracking data van twee Amerikaanse vogelsoorten, de katvogel (*Dumatella caroliniensi*) en boslijster (*Hylocichla mustelina*) waarvan de locaties dagelijks geregistreerd werden vanaf het verlaten van het nest. Beide soorten zijn evolutionair sterk gerelateerd, maar vertonen een differentiële respons ten opzichte van urbanisatie. De boslijster vermijdt urbaan gebied,

terwijl de katvogel aangepast is aan het leven in een door mensen gedomineerd landschap, wat resulteert in tegengestelde populatietrends. Om het effect van urbanisatie op het landgebruik van de nestverlaters in te schatten, werd het gebruik van bos en wegen door de twee vogelsoorten, twee landgebruiken die een maat zijn voor de graad voor urbanisatie, onderzocht. Hieruit blijkt dat katvogels gebieden met minder bos en meer wegen gebruiken dan boslijsters, wat overeenkomt met de verwachtingen. Tot slot werden de resultaten van de data-analyse samengevat in een model dat de opeenvolgende locaties van individuen kan simuleren. Gegeven de eenvoud van de gekozen modelstructuur, mogen de prestaties van het model als goed beoordeeld worden. Echter, meer kennis zou ingebouwd moeten worden om het gedrag van de juveniele katvogels en boslijsters beter te benaderen.

Samengevat bestaat dit doctoraat uit één theoretisch en drie meer praktische hoofdstukken. Bij deze laatste werd gebruik gemaakt van verschillende methodes en modellen om de specifieke onderzoeksvragen te beantwoorden. Op die manier biedt dit werk een meerwaarde voor elke partij waarmee samengewerkt werd, kan het inspirerend zijn voor toekomstig onderzoek met gelijkaardige onderzoeksvragen en levert het een bijdrage tot beter beheer en bescherming van de omgeving waarin we leven.

List of acronyms

| | |
|------|--|
| adv | angular deviation |
| BB | Brownian bridge |
| BBMM | Brownian Bridge Movement Model |
| BCPA | behavioral change point analysis |
| DDE | dichlorodiphenyldichloroethylene |
| DDT | dichlorodiphenyltrichloroethane |
| EC | environmental characteristic |
| EDF | empirical distribution function |
| ENFA | Ecological-niche factor analysis |
| FA | factor analysis |
| GLM | generalized linear model |
| GLMM | generalized linear mixed-effects model |

| | |
|------|---------------------------------|
| GIS | Geographic Information System |
| GPS | Global Positioning System |
| HS | habitat suitability |
| HSI | habitat suitability index |
| IBM | individual based model |
| KDE | kernel density estimate |
| GRCA | Gray Catbird |
| NLCD | National Land Cover Database |
| PCA | principal component analysis |
| PCB | polychlorinated biphenyl |
| PDF | probability density function |
| PMF | probability mass function |
| RFID | Radio Frequency Identification |
| RSF | resource selection function |
| SI | suitability index |
| stdv | standard deviation |
| TIBB | time-integrated Brownian bridge |
| UD | utilization distribution |
| UTM | Universal Transverse Mercator |
| VHF | very high frequency |
| WOTH | Wood Thrush |

List of Figures

| | | |
|------|--|----|
| 2.1 | Serengeti wildebeest migration | 8 |
| 2.2 | A bee (<i>Anthophila sp.</i>) and a deer (<i>Cervidae sp.</i>) equipped with a VHF radio transmitter | 10 |
| 2.3 | A researcher picking up the radio signals of VHF transmitters | 10 |
| 2.4 | A seal equipped with an Argos transmitter | 11 |
| 2.5 | An Eagle Owl (<i>Bubo bubo</i>) and a wolf (<i>Lupus sp.</i>) equipped with a GPS device | 12 |
| 2.6 | Schematic overview of a gps tracking system with upload options . . . | 13 |
| 2.7 | Screenshot of MoveBank | 17 |
| 2.8 | Movement research continuum | 17 |
| 2.9 | Screenshot of the tracking data of two wolves in Mongolia shown in Movebank and an example of a space-time cube | 19 |
| 2.10 | The distribution of the focal species on any ecogeographical variable may differ from that of the whole study area with respect to its mean and the standard deviation | 25 |
| 2.11 | Two parameters and potential functions are used to steer movement of pike in the river Yzer. | 30 |
| 2.12 | Simulated spatial distribution of the pike individuals (arrowheads) within the study area at the beginning of six different days during an annual cycle together with the territories. | 33 |

| | | |
|------|--|----|
| 3.1 | The trail of the 100 km Dodentocht | 38 |
| 3.2 | Relative frequencies of the age of the marchers (men and women) of the 100 km Dodentocht in 2010–2014. | 39 |
| 3.3 | The three weather stations near the trail who provided the weather conditions in 2009–2014 | 41 |
| 3.4 | The average amount of precipitation and temperature per hour during the editions 2009–2014 | 42 |
| 3.5 | The cumulative relative frequency of marchers retiring versus the distance from the start and the time elapsed since the start | 43 |
| 3.6 | The relative frequency of marchers along the trail versus time elapsed since the start | 44 |
| 3.7 | The average marching speed along the trail | 44 |
| 3.8 | Average autocorrelation between the marching speeds at k sections apart. | 45 |
| 3.9 | The cumulative relative frequency of men and women retiring versus the distance from and time since the start | 47 |
| 3.10 | The average marching speed of men and women at every section | 48 |
| 3.11 | Effect of age and gender of a 100 km Dodentocht marcher on the average total marched distance and the average marching speed | 48 |
| 3.12 | Effect of starting speed on the covered distance of a marcher | 49 |
| 3.13 | The average time elapsed since the starting signal before the marchers pass by the starting line as a function of the starting speed. | 49 |
| 3.14 | Visualisation of some statistics derived from the GPS tracking data of one participant | 51 |
| 3.15 | Overview of the locations where participants rest and the duration of the resting periods | 52 |
| 3.16 | The 95% confidence envelope of the simulated marching speeds for M_1 , M_2 and M_3 , together with the relative frequencies of the observed marching speeds | 57 |
| 3.17 | The 95% confidence envelope of the number of marchers passing by at every checkpoint per 0.25 h for M_1 , M_2 and M_3 , together with the observed number of marchers passing by. | 58 |
| 3.18 | The 95% confidence interval of the simulated cumulative relative frequency of marchers retiring at every checkpoint for M_1 , M_2 and M_3 , together with the observed cumulative relative frequency | 59 |

| | |
|---|----|
| 3.19 Match and Kolmogorov-Smirnov distance between the EDFs of the observed and simulated marching speeds for M_1 , M_2 and M_3 | 61 |
| 3.20 Match and Kolmogorov-Smirnov distance between the EDFs of the observed and simulated numbers of marchers passing by at every checkpoint per 0.25 h for M_1 , M_2 and M_3 | 62 |
| 3.21 Match and Kolmogorov-Smirnov distances between the EDFs of the observed and simulated marching speeds at every consecutive section for the five validated editions | 64 |
| 3.22 Match and Kolmogorov-Smirnov distances between the EDFs of the observed and simulated numbers of marchers passing by at every checkpoint per 0.25 h for the five validated edition | 65 |
| 3.23 The 95% confidence envelope of the relative frequencies of the simulated marching speeds for edition 2017, together with the relative frequencies of the observed marching speeds | 68 |
| 3.24 Match and Kolmogorov-Smirnov distances between the EDFs of the observed and simulated marching speeds at every consecutive section for edition 2017 | 69 |
| 3.25 The 95% confidence envelope of the simulated relative number of marchers at every checkpoint for edition 2017, together with the observed number of marchers | 69 |
| 3.26 The 95% confidence envelope of the simulated number of marchers passing by at every checkpoint per 0.25 h for edition 2017, together with the observed number of marchers passing by | 70 |
| 3.27 Match and Kolmogorov-Smirnov distances between the EDFs of the observed and simulated numbers of marchers passing by at every checkpoint per 0.25 h for edition 2017 | 71 |
| 3.28 PMF of the observed starting speed for male marchers, together with its fitted distribution function | 73 |
| 3.29 The 95% confidence interval of the relative frequency of the simulated marching speeds when the mean starting speed is varied. | 75 |
| 3.30 The 95% confidence enveloped of the number of marchers passing by at the consecutive checkpoints per 0.25 h when the mean starting speed is varied. | 76 |
| 3.31 The 95% confidence interval of the relative frequency of simulated marching speeds when the standard deviation of the starting speed is varied. | 78 |
| 3.32 The 95% confidence enveloped of the number of marchers passing by at the consecutive checkpoints per 0.25 h when the standard deviation of the starting speed is varied. | 79 |

| | | |
|------|--|-----|
| 4.1 | The PDFs of a one- and two-dimensional BB and corresponding TIBBs . | 93 |
| 4.2 | Indication of the regions of the PDF of the one-dimensional TIBB. | 95 |
| 4.3 | The marginal distribution in the x- and y-direction of the two-dimensional TIBB | 110 |
| 4.4 | The coordinate system is shown together with the rotated coordinate system | 111 |
| 4.5 | Computing time of the numerically determined PDF of a TIBB versus the discretization step in time, and the computing time of the analytically determined PDF | 117 |
| 4.6 | The analytically determined PDF of the TIBB for the Bechstein's bat dataset | 118 |
| 4.7 | The numerically determined PDF of the TIBB for the Bechstein's bat dataset | 120 |
| 5.1 | Study area together with the GPS tracking data of the tracked Eagle Owls | 126 |
| 5.2 | Study period per owl. | 126 |
| 5.3 | The spatial distribution of the environmental characteristics across the study area | 129 |
| 5.4 | The Eurasian Eagle Owl | 130 |
| 5.5 | Range of the Eurasian Eagle Owl | 130 |
| 5.6 | The HS map for the Eagle Owls in the study area. | 140 |
| 5.7 | The SI map covering the study area for environmental characteristics 1–6. | 141 |
| 5.8 | The SI map covering the study area for environmental characteristics 9–12. | 142 |
| 5.9 | The probability density of occurrence of the six Eagle Owls in the study area. | 143 |
| 5.10 | The probability density of occurrence for each Eagle Owl individual. . . | 144 |
| 5.11 | The probability mass function of the HSI values assigned to the pixels in the entire study area and those assigned to the fixes. | 145 |
| 5.12 | The average HSI value versus I_c and the average HSI value versus \tilde{I}_c | 145 |
| 5.13 | The average HSI value versus M_c and the average HSI value versus \tilde{M}_c | 145 |
| 5.14 | Reclassified HS map using the M_c graph based on the probability of occurrence | 147 |

| | | |
|------|--|-----|
| 5.15 | The mean and range of five movement feature as a function of the fraction of the night. | 156 |
| 5.16 | The convex hulls enclosing the visited locations, together with the visit frequency | 159 |
| 5.17 | The eigenvalues of the 16 factors obtained using ENFA together with the cumulative explained variance. | 161 |
| 6.1 | Range of the Gray Catbird and the Wood Thrush | 171 |
| 6.2 | The Gray Catbird and the Wood Thrush | 172 |
| 6.3 | Time-integrated Brownian bridge-derived normalized PDFs for fledgling Gray Catbirds and Wood Thrushes in both pre- and post-independence periods of parental care during the breeding seasons from 2012–2014 in and around Newark, Delaware, USA. | 176 |
| 6.4 | Comparison of area used within 15 National Land Cover Dataset land cover classes between Gray Catbirds and Wood Thrushes during pre-independence and post-independence periods for post-fledglings during the breeding seasons from 2012–2014 in and around Newark, Delaware, USA. | 177 |
| 6.5 | Comparison of area used within 15 National Land Cover Dataset land cover classes between Gray Catbirds and Wood Thrushes that died or survived during the breeding seasons from 2012–2014 in and around Newark, Delaware, USA. | 177 |
| 6.6 | Comparison of space use associated with road density for Gray Catbirds and Wood Thrushes during pre-independence and post-independence periods for post-fledging during the breeding seasons from 2012–2014 in and around Newark, Delaware, USA. | 178 |
| 6.7 | Comparison of pre- and post-independence periods for Brownian bridge area, path length, total displacement moved, and net displacement . . | 179 |
| 6.8 | Study area together with the VHF tracking data of the post-fledglings during the post-independent period | 185 |
| 6.9 | Study area together with the VHF tracking data of the post-fledglings during the post-independent period | 186 |
| 6.10 | Relative frequency histograms (bin width 0.05) of proportions Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between each pair of consecutive locations in the entire tracking data set. | 190 |

6.11 Average proportions Forest, Road and Urban canopy cover in each bin (bin width 0.05) of the histograms shown in Fig 6.10 in function of the relative frequency of proportions lying in the bin. The solid lines are the graphs of a piecewise function described by Eq. (6.4). Gray corresponds to Gray Catbird and brown to Wood Thrush. 193

6.12 Suitability index values as a function of the proportions of Forest and Road for Wood Thrush (left), and as a function of the proportions of Urban canopy cover and Road for Gray Catbird (right). 194

6.13 HSI values calculated from Eq. (6.6) based on the rank of the original HSI values. 196

6.14 The probabilities of choosing the angle of movement as a function of its rank when there are 4, 8, 16 and 36 possible angles of movement. 196

6.15 The Match and Kolmogorov-Smirnov distances for each land cover class and each model for Gray Catbird. 197

6.16 The sum of the Match and the sum of the Kolmogorov-Smirnov distances for Forest, Road and Urban canopy cover for Gray Catbird. The sums are sorted in ascending order and the plot markers show the model IDs as listed in Table 6.4. 198

6.17 The sum of the Match and the sum of the Kolmogorov-Smirnov distances for Road and Urban canopy cover for Gray Catbird. The sums are sorted in ascending order and the plot markers show the model IDs as listed in Table 6.4. 198

6.18 The relative frequency histograms (bin width 0.05) of the proportions Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between the pairs of consecutive locations for Gray Catbird. Both the histograms based on the observed and on the simulated proportions by Models 3, 6 and 9, are shown. 199

6.19 The Match and Kolmogorov-Smirnov distances for each land cover class and each model for Wood Thrush. 200

6.20 The sum of the Match and the sum of the Kolmogorov-Smirnov distances for Forest, Road and Urban canopy cover for Wood Thrush. The sums are sorted in ascending order and the plot markers show the model IDs as listed in Table 6.4. 200

6.21 The sum of the Match and the sum of the Kolmogorov-Smirnov distances for Forest and Road for Wood Thrush. The sums are sorted in ascending order and the plot markers show the model IDs as listed in Table 6.4. 201

6.22 The relative frequency histograms (bin width 0.05) of the proportions Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between the pairs of consecutive locations for Wood Thrush. Both the histograms based on the observed and on the simulated proportions by Models 21 and 24, are shown. 201

6.23 A box plot for the Match (left) and Kolmogorov-Smirnov (right) distances for Forest, Road and Urban canopy cover that were obtained from simulating 50 times one path per individual. 202

List of Tables

| | |
|--|----|
| 3.1 Distances along the trail between the checkpoints and the start | 38 |
| 3.2 General statistics of the 100 km Dodentocht | 40 |
| 3.3 Match and Kolmogorov-Smirnov distances between the observed and simulated EDFs of the number of marchers retiring at every checkpoint. | 60 |
| 3.4 The average and maximum absolute difference between the average of the 100 simulated and the observed number of people passing by at every checkpoint, per 0.25 h for a 100 km Dodentocht with 10,000 marchers | 62 |
| 3.5 Per edition, Match and Kolmogorov-Smirnov distances between the observed and simulated EDFs of the number of marchers retiring at every checkpoint. | 63 |
| 3.6 Per editon, the average and maximum absolute difference between the average of the 100 simulated and the observed number of people passing by at every checkpoint, per 0.25 h for a 100 km Dodentocht with 10,000 marchers | 66 |
| 3.7 Distances along the 100 km Dodentocht trail between the checkpoints and its start in 2017 | 67 |
| 3.8 Parameters defining the three normal distributions making up the mixture distribution describing V_1 | 73 |

| | | |
|------|--|-----|
| 3.9 | The difference between the observed maximum and the average simulated maximum number of marchers per 0.25 h with a density reduction of 10%, 25% and 50% at the start, obtained by means of wider streets | 83 |
| 3.10 | The difference between the observed maximum and the average simulated maximum number of marchers per 0.25 h with a density reduction of 10%, 25% and 50% at the start, obtained by means of a longer duration of the start | 84 |
| 4.1 | An overview of the distribution parameters of the PDF of the one-dimensional TIBB | 100 |
| 5.1 | Details on the extent of the tracking data set. | 127 |
| 5.2 | Taxonomic classification of the Eurasian Eagle Owl | 129 |
| 5.3 | Some exemplary aggregations of SI values according to Eq. (5.4). | 135 |
| 5.4 | Overview of the considered environmental characteristics (ECs), together with the SI value that typifies each EC e (α_e) and the radius of influence (r_e). | 138 |
| 5.5 | Overview of the considered environmental characteristics (EC). | 154 |
| 5.6 | The mean, median, standard deviation and range of movement features (1)–(3) and the mean, angular deviation and range of movement features (4)–(5) in and outside the breeding season for females B and C. | 157 |
| 5.7 | The mean, median, standard deviation and range of the average speed and the mean, angular deviation and range of the turning angle in between two fixes for every individual and cluster | 158 |
| 5.8 | The relative frequency of occurrence of EC2a, EC2b, EC3, EC4, EC5, EC6, EC9b and EC9c for the fixes and across the study area, and the ratio of the former versus the latter. | 160 |
| 5.9 | The average distance (m) from the fixes to each EC, the average distance (m) from any pixel in the study area to each EC, and the ratio of the former versus the latter. | 161 |
| 5.10 | The loadings of every EC for the first two factors obtained by ENFA. The explained variance of each factor is indicated between parentheses. | 163 |
| 6.1 | Taxonomic classification of the Gray Catbird and the Wood Thrush | 171 |

| | | |
|------|--|-----|
| 6.2 | Mean weighted relative proportions of space use within each National Land Cover Dataset land cover class. | 175 |
| 6.3 | Mean (and standard error) of movement metrics of post-fledging Gray Catbirds and Wood Thrushes during pre- and post-independence periods from 2012–2014 in and around Newark, Delaware, USA | 179 |
| 6.4 | Overview of the constructed models considered for Gray Catbird and Wood Thrush. | 188 |
| 6.5 | The average proportions of Forest, Road and Urban canopy cover, together with the corresponding standard deviation (stdv), that are found across the 95 % TIBB areas defined in between each pair of consecutive locations in the tracking data set. | 189 |
| 6.6 | The values of parameters α_1 and α_2 per land cover class and species. | 193 |
| 6.7 | Calibrated model parameters using the training set for each fold of cross-validation. | 203 |
| 6.8 | The Match and Kolmogorov-Smirnov distances for a five-fold cross-validation, together with the average distances for Gray Catbird. | 205 |
| 6.9 | The Match and Kolmogorov-Smirnov distances for a five-fold cross-validation, together with the average distances. Here, Wood Thrush and Model 21 are under consideration | 206 |
| 6.10 | The Match and Kolmogorov-Smirnov distances for a five-fold cross-validation, together with the average distances. Here, Wood Thrush and Model 24 are under consideration | 207 |

1

Introduction

1.1 Overview

In this dissertation, movement, *i.e.* the act of changing location in time, is studied by means of telemetric observations and mathematical models. Many organisms are on the move on the land, at sea or in the air. Plant species, for example, change locations over several generations, while (most) animals are continuously moving to find food, shelter and mates.

While moving, animals are constantly monitoring the environment they pass by and taking informed decisions. As such, by studying movement data, we are able to analyze and model nature's response to the environment. In general, analyzing movement data of animals can help to understand their behavior in many aspects, to determine the factors influencing them, to identify hotspots and to investigate social interactions. Therefore, studying movement helps to gain knowledge related to where, when and ultimately why the individuals move the way they do. These insights are crucial in the construction of successful management plans so that necessary food resources, resting areas, nesting sites, connectivity between nature reserves, possible barriers and so on can be taken into account. In addition

1 to the ecological importance of studying movement phenomena, there is also an economical dimension to it since animals spread pollen, nutrients and diseases whilst moving. The focus of this research is on animals, but also people are constantly moving. Analysis of human tracking data can be used to manage traffic and guarantee safety in daily life or during (mass) events. Furthermore, tracking data is used for military purposes and for improving sports performances.

Movement phenomena can be studied by means of telemetric observations. In telemetric monitoring campaigns, a tracking device is attached to an individual and used to observe the individual's subsequent locations in time. As such, tracking data, containing information in both space and time, are obtained. Since different field studies have different objectives and logistical or financial constraints, tracking data are diverse. Moreover, they are intrinsically discrete, whereas movement is a continuous phenomenon. Consequently, extracting the desired information is not straightforward and analyzing tracking data brings along several specific challenges. The first challenge arises from the fact that most tracking data are spatially and temporally correlated. Furthermore, it is important to take into account the device error, while also the management of large tracking datasets obtained from monitoring campaigns poses a challenge to researchers.

In all tracking studies, tracking data are transformed into knowledge of the individuals under consideration and their behavior. Two major branches in movement research are distinguished. The first branch involves the understanding of the movement process, and the second branch involves the modeling of movement phenomena and the prediction of possible responses to environmental changes. Here, visualization techniques can be used for exploring tracking data and communicating the results, and, by means of a computational movement analysis, tracking data are used to quantify movement tracks, features and patterns, and their relationship with environmental characteristics. As such, knowledge is extracted from the observations, which can be used subsequently to develop movement models. These models increase our understanding of movement phenomena. Ultimately, the whole process results in models that capture behavioral responses and movement through space and time under varying environmental conditions. In this case, movement models are able to test hypotheses about movement and predict how the species will likely respond to their rapidly changing environment. They serve as tools for making informed decisions that will benefit a sustainable resource management.

1.2 Research objectives

Technological advances in data collection and increased computer power are revolutionizing the field of bioscience engineering. This revolution can also be seen in

movement science. The data-to-decision cycle – encompassing experimental set-up, data collection and analysis, mathematical modeling and decision making – calls for the application and development of convenient approaches. My research is joining this effort. More specifically, the main goal is to select already existing or develop new mathematical methods and models to help answering research questions at stake. To reach this goal, available tracking data were used to gain knowledge and to parameterize, verify and validate models. The data analysis results and models obtained should eventually support a well-founded decision making. For this research, we relied on existing tracking datasets. The specific research objectives are:

1. Build a spatially explicit marching model of the Belgian major endurance event ‘100 km Dodentocht’ to predict the effect of a modified starting procedure;
2. Derive the analytical description of the time-integrated Brownian bridge;
3. Unravel possible sources of poisoning for the Eurasian Eagle Owl (*Bubo bubo*) in Limburg, the Netherlands;
4. Estimate the effects of urbanization on post-fledging space use of two bird species, *i.e.* the Gray Catbird (*Dumatella caroliniensi*) and the Wood Thrush (*Hylocichla mustelina*), which are phylogenetically closely related though exhibit differential responses to urbanization.

1.3 Roadmap

This dissertation consists of one more theoretical chapter, involving the analytical description of the time-integrated Brownian bridge (TIBB), and three chapters of a more applied nature, where tracking data sets are used to address specific research questions. Each data set contains data of a different type and considers a different species, region and research question. Based upon the available data and the research question, appropriate analysis and modeling tools have been selected. To a researcher new in the field, it is hard to find those given the many possibilities. Therefore, an introduction to movement ecology is given in Chapter 2 to get familiar with the unique vocabulary, the particular characteristics of tracking data, the prevailing analysis tools and the available movement models.

If the reader of this dissertation is familiar with movement ecology, Chapter 2 can be skipped and he/she can immediately dive into the research topics at stake, which are all independent of each other and can be read separately.

In Chapter 3, the long distance march ‘100 km Dodentocht’, which takes place annually in Belgium, is considered. Corresponding tracking data are analyzed to

1 build an individual-based spatially explicit marching model that gives insight into the dynamics of this event and allows to evaluate the effect of changes in its starting procedure. Here, we are able to work in a one-dimensional setting, which is not possible for (most) animal tracking data sets.

In Chapter 4, the analytical description of the TIBB is derived. The latter is used to determine the intensity of use of an area, represented by means of a probability density function (PDF), on the basis of tracking data. The TIBB drew our attention in movement ecology literature because it takes into account the registered locations, the animal's mobility and the time between the locations to determine the intensity of use, which is not the case for other approaches. Furthermore, we were struck by the remarkable shapes of the resulting PDFs and the fact that they are computed through the use of numerical integration methods. We demonstrate that it is nevertheless possible to derive the analytical description of the TIBB, thereby making computations more efficient.

In Chapter 5, Eurasian Eagle Owls in Limburg, the southeastern province of the Netherlands are considered since their population size had stagnated and extremely high concentrations of poisonous chemical compounds (PCBs and DDEs) were found in owl carcasses. In order to unravel the possible poisoning sources and to protect the Eagle Owl population, the environmental characteristics and areas to focus on when determining possible sources of poisoning are identified in two ways. First (Section 5.4), the suitability of the study area is summarized in terms of a habitat suitability map based on a literature and expert knowledge. In this approach, the GPS tracking data are solely used for interpretation and validation. Secondly (Section 5.4), the tracking data are used to identify both the important environmental characteristics and areas, rather than basing this identification on literature and expert knowledge. Here, the size of the data set is reduced so that the remaining fixes are linked to possible hunting grounds. Subsequently, we investigate the most frequently visited locations by means of a hotspot analysis and the relative importance of the prevailing environmental characteristics in the owls' hunting grounds by making use of the ecological-niche factor analysis (ENFA).

In Chapter 6, the land use of two American migratory bird species, *i.e.* the Gray Catbird (*Dumetella carolinensis*) and the Wood Thrush (*Hylocichla mustelina*), are studied during the post-fledging period. These two species, while being phylogenetically closely related, share similar life-history traits, though exhibit differential responses to urbanization. To gauge effects of urbanization on post-fledging space use, specific differential responses between the two species during pre- and post-independent periods in their respective use of forest habitat and road area are tested. Furthermore, the obtained responses for the post-independent period are summarized in a simple movement model. Using this model, it is possible to simulate subsequent locations on the basis of the proportions Forest, Road and Urban canopy cover available.

To conclude, an overall discussion can be found in Chapter 7.

2

Movement ecology background

2.1 Introduction

Plants, animals and other organisms are having a hard time in our human-dominated world as a consequence of habitat fragmentation, habitat loss, overexploitation of natural resources and climate change (Wilcove and Wikelski, 2008; Cioc, 2009; Candolin and Wong, 2012). The rapidly changing environment brings along threats that all kinds of organisms have to conquer every day. While moving, animals are constantly monitoring the environment they pass by and taking informed decisions. By studying movement data, we are able to analyze and model nature's response to the environment. Intrinsically, animals are sensors of environmental suitability (Kays et al., 2015). In general, analyzing movement data of animals can help to understand their behavior in many different aspects, to determine the factors influencing them, to learn about hotspots and to investigate social interactions (Gudmundsson et al., 2008). The overall challenge lies in relating movement patterns with the underlying geography and internal state in order to understand

where, when and ultimately why the individuals move the way they do (Gudmundsson et al., 2008; Nathan et al., 2008). These insights can then be used to protect natural heritage and to shape our world with respect for nature in all its forms.

In addition to its ecological importance, studies of movement data are also economically important. While moving, animals disperse pollen, seeds, diseases and nutrients. Moreover, they determine the spatio-temporal distribution of predation (on insects, for example), herbivory and invasive alien species. These processes related to movement are valued at hundreds of billions of dollars per year (Nathan et al., 2008; Wilcove and Wikelski, 2008; Kays et al., 2015; Dingle and Drake, 2007).

The focus of this chapter is on animals, but also people are on the move. Their movement data are used for traffic management (Marković et al., 2018), for surveillance and security in every day life and during events (Versichele et al., 2012), and for military purposes (Gudmundsson et al., 2008). Furthermore, movement data are influencing professional sports. The data are used for analyzing the performance of players and, consequently, for improving their strategy (Gudmundsson et al., 2008; Baca et al., 2009).

2.2 Movement

According to the conceptual framework of Nathan et al. (2008), a moving individual is characterized by three components: an internal state, a motion capacity and a navigation capacity. The internal state refers to the goals the organism wants to achieve, such as finding food, shelter and mates, and handles the question 'why move?'. The motion capacity refers to the way an individual is able to move such as flying, swimming, walking, and addresses the question 'how to move?'. Finally, the navigation capacity refers to the ability to orient in space and time, and stems from the question 'when and where to move?'. A fourth component, the external factors, represents the abiotic and biotic environment influencing movement, such as, physical boundaries and the degree of success in meeting the individuals' needs (Alerstam et al., 2003). Together, the four components and their interactions determine the movement paths of individuals. By investigating these components, an analysis of movement data can be used to unravel underlying movement mechanisms (Nathan et al., 2008).

Most displacements are associated with daily activities like foraging, mate-location and seeking shelter or a resting place. Migration and dispersal, however, are two types of movement that are not associated with such routine activities (Van Dyck and Baguette, 2005).

Migration is the most spectacular type of movement in nature. The word 'migration', however, is often ambiguous and, according to Dingle and Drake (2007),

can refer to four different overlapping concepts: 1) a type of locomotory activity, *i.e.* any physical action that propels an individual from one place to another, that is notably present, undistracted and straightened out; 2) a relocation of the animal that occur on a much greater scale, and involves movement of much longer duration than movement arising in its normal daily activities; (3) a seasonal movement of populations between regions where conditions are alternately favorable or unfavorable (including one region in which breeding occurs); and (4) movements leading to redistribution within a spatially extended population. Migration is thus both defined at an individual and a population level and definitions are based on the process and the outcome.

Migration can be found all over the world in all major branches of the animal kingdom. One of the main iconic examples is the migration of Africa's Great Wildebeests (*Connochaetes taurinus*) (Fig. 2.1) accompanied by large numbers of zebras (*Equus burchelli*) and gazelles (*Nanger granti*), who are searching for food and water through Tanzania and Kenia. Also other mammals, such as caribou (*Rangifer tarandus*) and whales (Humpback Whale (*Megaptera novaeangliae*), Blue Whale (*Balaenoptera musculus*)), and so on) cover large distances. In fish, migration typically involves the transition between spawning, feeding and nursery grounds (Atlantic Bluefin Tuna (*Thunnus thynnus*), South American Pilchard (*Sardinops sagax*), and so on). Moreover, it often involves a transition between salt and fresh water. European Eel (*Anguilla anguilla*), for example, spend most of their life in fresh-water but migrate to the sea to breed, while the Atlantic Salmon (*Salmo salar*) and salmon from the Pacific Ocean (*Salmo Oncorhynchus sp.*) display the opposite pattern. Furthermore, Berthold (2001) states that it is likely that about 50 billion of the 200–400 billion birds existing on this planet migrate annually. For example, the Barn Swallow (*Hirundo rustica*), the Arctic Tern (*Sterna paradisaea*) and many others perform a long distance migration between their breeding and wintering grounds. Among amphibians, newts and toads cover several kilometers to and from their spawning grounds. Among reptiles, the Green Turtle (*Chelonia mydas*) is the most famous migratory example. Also insects migrate: the American Monarch Butterfly (*Danaus plexippus*) migrates up to 4000 km from its most northern habitats in Canada to its winter grounds in Mexico, while swarms of locusts (*Orthoptera sp.*) are known to invade large parts of continents (Berthold, 2001; Candolin and Wong, 2012).

Migratory species need a suitable habitat near the start and the final destination of their journey, and a network of resting places and corridors along the way. Furthermore, migration should not be hindered by barriers. Because of the range of their habitat, migratory species are even more vulnerable to anthropogenic disturbances than non-migratory species (Wilcove and Wikelski, 2008; Wilcove, 2008; Candolin and Wong, 2012). As a consequence, the population size of many migratory species has declined. The Passenger Pigeon (*Ectopistes migratorius*), for example, is extinct and the migration of millions of bison (*Bison bison*) cannot



Figure 2.1: Serengeti wildebeest migration (Source: kuwa-huru.com).

longer be observed (Wilcove and Wikelski, 2008; Wilcove, 2008).

Another type of movement is dispersal which is defined as the movement of individuals away from each other. It often refers to any movement from the birth or breeding site to another location where the individual might establish and reproduce (Van Dyck and Baguette, 2005). Dispersal is a key process in ecology and evolutionary and conservation biology. It may allow for the exploitation of spatially and temporally variable resources, and is essential for most species to persist in the ever changing natural environment. Moreover, it is important to maintain the genetic diversity of wild populations (Clobert et al., 2009).

Finally, it needs to be pointed out that movement can significantly be affected by the motion of the medium – water or air – in which an individual is immersed. Some animals have mechanisms for identifying and exploiting favorably-directed flows, and coping with opposing ones (Chapman et al., 2011). Several studies have shown the importance of favorable wind conditions for the successful migration of birds (Erni et al., 2005) and also fish can benefit from the complex three-dimensional water flows (Liao, 2007)

2.3 Tracking devices

By tracking wildlife, the locations of an animal at consecutive timestamps are obtained. Since the 1960s, individual animals have been equipped with tracking devices to remotely collect the individuals' locations, a process that is referred to as

This section is adapted from Markham (2008).

telemetry. Here, the difficulties in frequently locating individuals over large areas are overcome (Cochran and Lord Jr, 1963). Next to the locations, also physiological (heart rate or body temperature) or activity (head tilt or defecation) information can be acquired, and tracking devices can be combined with sensors logging environmental parameters (temperature, wind or water depth) (Demšar et al., 2015). In this study, we focus on automatic tracking technology to acquire location. The location where an individual was registered at a specific time stamp will be referred to as a 'fix'. Ideally, a tracking device has a small size and low weight, a high sample accuracy and frequency, a global coverage, a long life time, a real-time access to the data and a low cost. Up to this date, a 'one-size-fits-all' tracking device is not available and one needs to compromise between these characteristics. Nowadays, there are three major tracking techniques in use, namely VHF (very high frequency) radio, satellite and GPS (Global Positioning System) tracking.

2.3.1 VHF radio, satellite and GPS tracking techniques

The earliest devices were VHF radio tags designed for rabbits (*Sylvilagus floridanus*), Striped Skunks (*Mephitis mephitis*) and raccoons (*Procyon lotor*) (Cochran and Lord Jr, 1963). As a consequence of improved battery technology, also smaller animals such as birds and insects can be studied nowadays (Hooten et al., 2017). A VHF tracking system consists of two components: the transmitter (attached to the animal) and the receiver, shown in Figs. 2.2 and 2.3. The transmitter emits a regular radio signal at a specific frequency. A beeping sound is heard whenever the signal is picked up by a nearby receiver that is tuned to this frequency, and the beeps get louder as the receiver approaches the transmitter. The transmitters used in one study are tuned to different frequency bands so that individual animals can be identified and tracked. The exact location can be determined by visual identification or by triangulation. As such, VHF tracking studies rely on researchers in the field who need some practice to obtain accurate data and who might affect animal behavior. Sometimes, cars or small aircrafts are used to cover vast or unaccessible regions (Hooten et al., 2017). As an alternative to field visits, one or more fixed stations recording the frequencies of all tracking devices within radio range can be used to automatically determine the individuals' locations (Cagnacci et al., 2010). The location error becomes greater as an animal is located farther from the receiving antenna.

VHF tracking is a simple, relatively low cost technique with low transmitter power needs. The smallest VHF device weighs approximately 0.2 g, but has a very short lifetime (two–three weeks) (Naef-Daenzer et al., 2005). For longer studies, a device that weighs approximately 500 g can transmit a signal for over 4 years. As such, VHF tracking is applicable to a wide range of animal species (Hooten et al., 2017). Due to the limited range of the radio signal, a researcher needs to be in the field to determine the location of an animal, which makes the data collection labor



Figure 2.2: A bee (*Anthophila sp.*) and a deer (*Cervidae sp.*) equipped with a VHF radio transmitter (Source: blog.csiro.au and robertmccaw.com).



Figure 2.3: A researcher picking up the radio signals of VHF transmitters (Source: lesnoezero.villagepages.org).

intensive (Cagnacci et al., 2010). As a consequence, VHF tracking data are often sparse (Kays et al., 2015) and the maximum extent of practical tracking areas varies between 50 and 300 km².

Satellite tracking is similar to VHF radio tracking, but instead of a radio signal being sent to a radio receiver, a signal is sent to a satellite, thereby avoiding field visits. The most popular system is ARGOS, which provides global coverage by two satellites. From the known position of a satellite and the Doppler shifted signals, a location can be determined (Cagnacci et al., 2010) (Fig. 2.4). A drawback of the ARGOS tracking system is the cost, since both the transmitters and data processing are expensive. Moreover, the location accuracy is poor compared to VHF and GPS tracking, and the number of fixes that can be recorded per day is small compared to GPS tracking due to the limited time window where the satellites are within reach. Furthermore, since the device's weight is mainly determined by the power source, ARGOS devices weigh a lot because large batteries are necessary for an extended lifetime (Bridge et al., 2011).

The field of animal tracking studies exploded when the US Department of Defense stopped degrading the accuracy of the Global Positioning System (GPS) in the year 2000 (Kays et al., 2015). Contrary to satellite tracking where the tracking device attached to the animal is a transmitter, the device needed for GPS tracking is a receiver that needs to acquire signals from at least four satellites (of the 24 available) in order to determine its three-dimensional position (Fig. 2.5). The GPS device determines the distance to each available satellite by measuring the time the signal needs to travel from the satellite to the device. From this info and the exact position of the satellites, an animal can be located precisely. Since the receiver needs to be active until a fix is acquired, which can take a few minutes in some cases, there is a high time and power cost per fix. Consequently, it is not practical to operate the receivers continuously for long-term studies. Therefore, they are often in sleep mode, only powered up every other fixed amount of time (typically between 5 minutes and 4 hours) (Tomkiewicz et al., 2010). With solar panels, however, batteries are kept charged so that lifetime tracks can now be collected for some species (Kays et al., 2011).



Figure 2.4: A seal equipped with an Argos transmitter (Source: cnes.fr).



Figure 2.5: An Eagle Owl (*Bubo bubo*) and a wolf (*Lupus sp.*) equipped with a GPS device (Source: Wassink (2014b) and opb.org).

GPS tracking systems are more expensive than VHF systems, but there is minimal time needed in the field, which again reduces the associated costs. The devices are about as expensive as the ones used in ARGOS tracking systems, but there are no costs for data processing. GPS enables a precise (within 5 m) localization of a receiver anywhere in the world, 24 hours a day. Nevertheless, many factors affect the accuracy and availability of GPS-based fixes, such as foliage cover, ionospheric delays, the number of visible satellites and the accuracy of the receiver. The effect of environmentally induced biases must be taken into account to draw valid conclusions.

Considering both GPS and ARGOS tracking systems, two types of data retrieval can be distinguished. The first one stores the gathered information and the second one has upload options. The first type is the most simple and power efficient one. At the end of a study, the device needs to be retrieved so that the information can be downloaded. This implies a risk of data loss and a long time lag between sampling and data retrieval. In the second type, a VHF, GSM or GPS module is available to upload data to researchers in the field, to GSM towers or directly to satellites (Fig. 2.6). As a consequence, the risk of data loss is reduced and data are readily available for analysis. There is, however, an increased power requirement, resulting in larger batteries (Tomkiewicz et al., 2010; Hooten et al., 2017).

In general, devices should weigh less than 4% – 5% of the body weight for mammals and 2% – 3% of the body weight for avian wildlife. Due to these weight limits, every reduction in size increases the range of animals for which the system may be used. In the last few years, tracking technology has improved following the development of smaller (and cheaper) batteries and more energy-efficient micro-processors. Still, 70% of the bird species and 65% of the mammal species cannot be tracked while they are on the move and miniaturization remains a priority (Kays et al., 2015).

GPS tracking systems are ideally used for larger, terrestrial species due to the size of the devices, are applied in open habitat due to the need of satellite coverage, and, are inappropriate for aquatic species due to the need of satellite coverage and

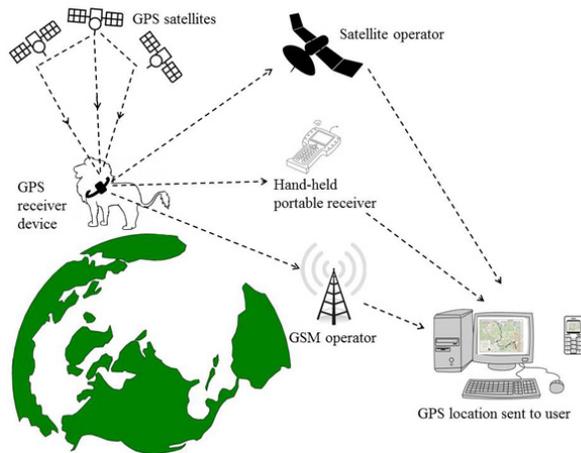


Figure 2.6: Schematic overview of a gps tracking system with upload options (Source: wildtech.mongabay.com)

long acquisition times. Complementing GPS systems, ARGOS is a popular option for marine species and small terrestrial species (Hooten et al., 2017).

2.3.2 Other tracking techniques

Next to the three major techniques in use nowadays, some other techniques are discussed below for the sake of completeness. In 2006, a compromise strategy between GPS and ARGOS was designed under the name FastLoc. It reduces acquisition times, but is less accurate than GPS, though more accurate than ARGOS (Hooten et al., 2017). An other, more sophisticated, technology that can be combined with GPS, estimates the movement path relative to the starting location by combining accelerometer-based speeds and magnetometer headings (bearing angle relative to the North). Although the technology suffers from a rapid error accumulation, GPS can be used for calibration. This technology is referred to as dead-reckoning or path integration (Wilson et al., 2007; Dewhirst et al., 2016).

An other tracking system worthwhile mentioning uses passive RFID (Radio Frequency Identification) tags that involve a battery-less miniature transponder that is attached to an animal. It gets activated when it approaches an antenna that generates a magnetic field and makes the transponder transmit its unique identification number, which is passed to a computer for further processing. Although the readers are expensive, the transponders are not. As such, RFID provides one of the most cost-effective methods to monitor large numbers of animals. Furthermore, the transponders are lightweight and small, and are not affected by water or flesh, making implantation and underwater operation possible. They are useful for presence/absence detection in a certain area, but not for determining individual

movement paths in detail.

A third technique, which is also powered by incident radiation, uses radar to illuminate a tag. A detector records the backscatter radiation to determine the device's position. Also in this case, devices are lightweight and useful for small animals (Mascanzoni and Wallin, 1986; Shamoun-Baranes et al., 2016). Furthermore, radar can also be used to observe flocks of birds and insects without any attached device by analyzing the characteristics of the radar echoes (Drake, 1985; Shamoun-Baranes et al., 2014). Since the infrastructure is already in place and operational for weather surveillance, continental-scale radar networks for monitoring animal movement can likely be established at a relatively low cost. This idea poses a number of challenges for the ornithological and ecological communities: how does one take advantage of this vast data resource, integrate information across multiple sensors and large spatial and temporal scales, and visually represent the data for interpretation (Shamoun-Baranes et al., 2014)?

A monitoring technique used in aquatic environments, is acoustic telemetry. Here, a device is implanted in the belly of a fish that transmits an underwater acoustic signal. When a fish swims in the vicinity of a receiver (typically attached to a buoy, ship wreck and river bank), the information of the unique ID-code of the transmitter is stored together with an exact timing of the event. When signals are simultaneously received by multiple receivers, an exact positioning of the tagged individuals can be derived from the raw data. In this way, movements and habitat use of tagged migratory fish can be tracked (Belgian LifeWatch Infrastructure, 2018).

A fifth technique for determining the approximate location of an individual uses a global location service (GLS) system. Here, a light sensor is used to measure the length of the day in order to determine latitude. Longitude is calculated from the local time of midday or midnight. Due to the inexpensive and light devices, they can be applied to a wide range of animals. The accuracy of this system, however, is poor and the number of fixes a day is restricted to two. Furthermore, a GLS device needs to be retrieved at the end of the study to download the data (Wilson et al., 1992, 2002).

Finally, remote camera traps, which are activated if there is motion or sound in their vicinity, can be used to record the presence of certain animals at various locations. In general, photographs and videos, taken by field scientists, tourists, automated cameras, and incidental photographers make up an important source of data on wildlife today. Although those images are simple and relatively inexpensive to obtain, they require a lot of manual interpretation. Initiatives like Wild Me, however, integrate computer vision and machine learning to scale up processing speed and support new analyses (Berger-Wolf et al., 2017). For example, individual whales are accurately identified by integrating and customizing multiple pattern recognition algorithms (Berger-Wolf et al., 2017) and lions are automatically identified on the basis of their unique coat marking (Burghardt and Čalić, 2006).

This approach is non-invasive, but sensitive to dirt, orientation and ambient lighting (Burghardt and Čalić, 2006).

For tracking people on the move, tracking systems rely on GPS devices in cars, cell phones, smart watches and so on (Van der Spek et al., 2009; Siła-Nowicka et al., 2016; Tang et al., 2016; Brum-Bastos et al., 2018; Marković et al., 2018). Furthermore, cell phone networks (Calabrese et al., 2015), RFID (Baca et al., 2009), Bluetooth (Versichele et al., 2012) and Wifi can provide easy-to-handle data on the location of humans ().

2.4 Tracking data

Tracking devices generate large and, to ecologists and statisticians, largely unfamiliar data sets (Patterson et al., 2017). Basically, tracking data contain the locations of an animal at consecutive timestamps, and, as such, carry spatio-temporal information on the movement of an individual and the underlying mechanisms that drive this movement (Dodge et al., 2016). Tracking data are diverse because different field studies have different objectives and logistical or financial constraints (Hooten et al., 2017). Moreover, the data are intrinsically discrete, whereas movement is a continuous phenomenon. Indeed, from such data one can derive where an individual is located at subsequent time stamps, but in between those, there is no information on its location (Patterson et al., 2017). Consequently, extracting the desired information is not straightforward and analyzing tracking data brings along several specific challenges.

The first challenge arises from the fact that most tracking data are highly spatially and temporally correlated, *i.e.* they are autocorrelated. It can be stated that the higher the frequency of sampling, the more reliable the outlined path is, but the higher the correlation between subsequent observations becomes. Traditional approaches attempt to factor out the dependence by subsampling. However, this leads to considerable data loss. Therefore, it is more appropriate to use all available data and to deal with the autocorrelation. Nowadays, modern approaches integrate space and time to test hypotheses about animal movement (Solla et al., 1999; Cagnacci et al., 2010; Fieberg et al., 2010; Kays et al., 2015; Hooten et al., 2017). There are three common ways to address the autocorrelation in the data analysis stage. The first one models correlation explicitly as an intrinsic and relevant property of the data, the second one disentangles patterns that correspond to behavioral phases or to specific behaviors, and the third one investigates the mechanisms underlying autocorrelation such as memory or internal state (Cagnacci et al., 2010).

Furthermore, it is important to have a measure of the device error. In the field, this error is influenced by environmental conditions and animal behavior, and results

in spatial inaccuracy and missing data. The second challenge lies in assessing the potential implications of inaccuracy and missing data for ecological analyses. When the device error can lead to mistaken inferences on animal behavior, correcting tools should be applied (Cagnacci et al., 2010).

Another challenge researchers have to deal with is the management of large tracking datasets. Good scientific practice requires that data are securely, consistently and efficiently managed. Therefore, a standard data structure is desirable, and a software architecture is needed that automates data acquisition, stores data long term, makes data available for multiple users, integrates different data sources and takes the spatial and temporal information into account (Urbano et al., 2010). Moreover, it would be interesting to integrate tools for data analysis and visualization in order to avoid task replication (Cagnacci et al., 2010). Systems currently in use are MoveBank (movebank.org (Wikelski and Kays, 2014)), TOPP (gtopp.org), ISAMUD (eurodeer.org), BioMap for GPS (lotek.com) and so on (Fig. 2.7). ISAMUD, for example, imports locations from GPS collars and processes the coordinates so that they are stored correctly. Then, the linear distance, time interval, relative and absolute angles are automatically calculated between fixes. Additionally, scripts are available that, for example, perform a spatial analysis (home range, and so on), or connect to the GIS environment GRASS to investigate environmental conditions (Urbano et al., 2010).

In all tracking studies, tracking data are transformed into knowledge of the individuals under consideration and their behavior. In Fig. 2.8, a continuum introduced by Dodge et al. (2016), covering the fundamental areas of movement research, is presented. It has the movement observations at its core. Two major branches in movement research are distinguished. The first branch involves the understanding of the movement process, and the second branch involves the modeling of movement phenomena and the prediction of possible responses to environmental changes. Both branches are connected and provide information to each other.

As Dodge et al. (2016) explain, visualization techniques are used for exploring tracking data and communicating the results to ensure the reliability of methods, models and discovered patterns. By means of a computational movement analysis, movement observations are used to quantify movement tracks, parameters and patterns, and their relationship with the environmental characteristics (context). As such, knowledge is extracted from the observations which subsequently can be used to develop movement models. These models can be used to run simulations and further increase our understanding of movement phenomena. Ultimately, the whole process results in models that capture behavioral responses and movement through space and time under varying environmental conditions. These models can be parameterized, calibrated and validated using real movement observations and the obtained insights. Some relevant analysis and modeling approaches are considered in more detail in Sections 2.5 and 2.6, respectively.

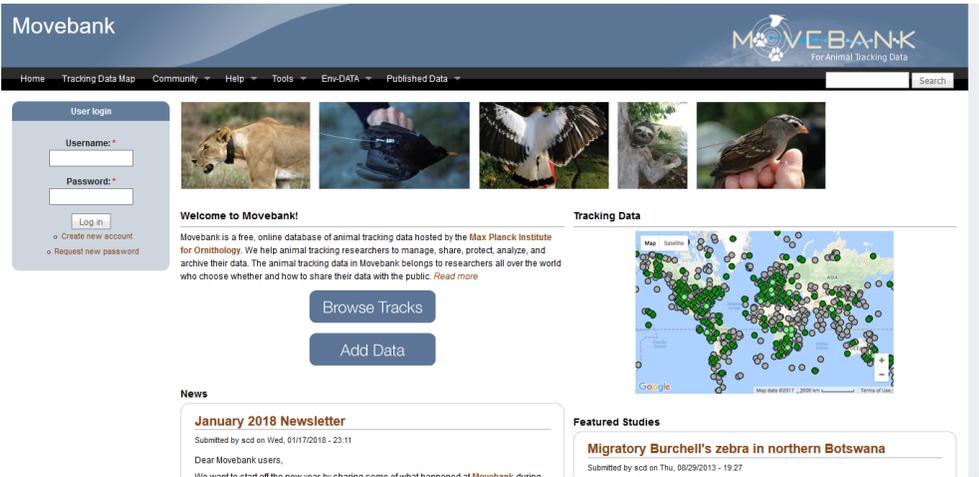


Figure 2.7: Screenshot of MoveBank

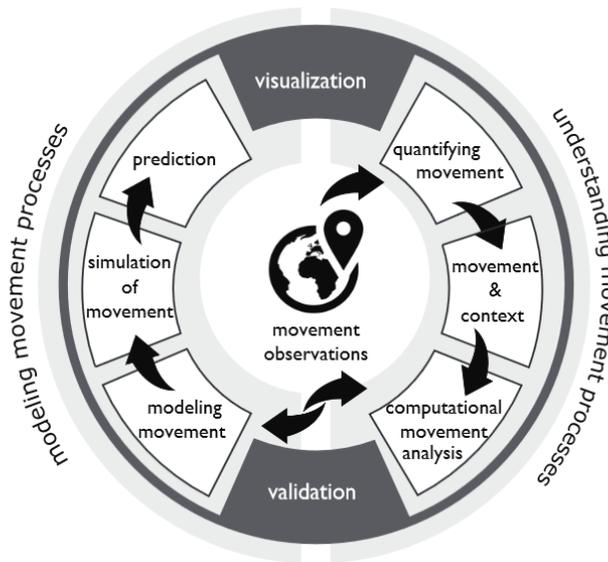


Figure 2.8: Movement research continuum (Dodge et al., 2016).

There have already been a lot of attempts to handle spatio-temporal (movement) data. Since movement ecologists are mainly using R for data analysis, most attempts resulted in R packages. Some useful R packages to analyze spatio-temporal data can, for example, be found at

<https://CRAN.R-project.org/view=SpatioTemporal> and there exist R packages written specifically for the analysis of animal movement. Some of these relevant R packages include: `adehabitat`, `rsMOVE`, `moveHMM`, `ctmm`, `GeoLight`, `move`, `trip`, `wildlifeDI`, `wildlifeTG`, `PathInterpolatR`, `BBMM`, `UvA-BiTS ETL`, `ETN` and so on. All packages are well documented and help with accessing and processing data to address movement ecology questions. `Adehabitat`, for example, provides tools for the analysis of habitat selection by animals, `moveHMM` and `ctmm` supply tools for animal movement modeling and `rsMOVE` supports the combined use of animal movement and remote sensing data. Furthermore, Geographic Information System (GIS) software (`ArcGIS`, `QGIS` or `GRASS`) is used for geospatial data management and analysis, image processing, graphics and maps production, spatial modeling and visualization.

2.5 Tracking data analysis

Since tracking data are diverse and many different questions are posed (Kays et al., 2015), there is not a single recipe for processing tracking data.

2.5.1 Visualization

Usually, visualizing the fixes and the trajectories are good ways to start exploring tracking data. It allows for preliminary observing and interpreting (known) patterns and structures, linking environmental characteristics, detecting problems with the data and postulating hypotheses. As animals often stay in relatively small areas over extended periods of time, some regions have overlapping trajectories. As a consequence, plotting the fixes and trajectories on a map or by means of a three-dimensional space-time cube often results in visual clutter. This gets even worse when visualizing trajectories of several individuals at the same time. Therefore, animations might be more effective to visually explore the data, and computational movement analysis is needed to extract knowledge (Demšar et al., 2015; Dodge et al., 2016).

2.5.2 Quantifying movement

Further analysis can focus on features derived from the fixes that embody the physical features of movement, such as the distance between successive fixes,

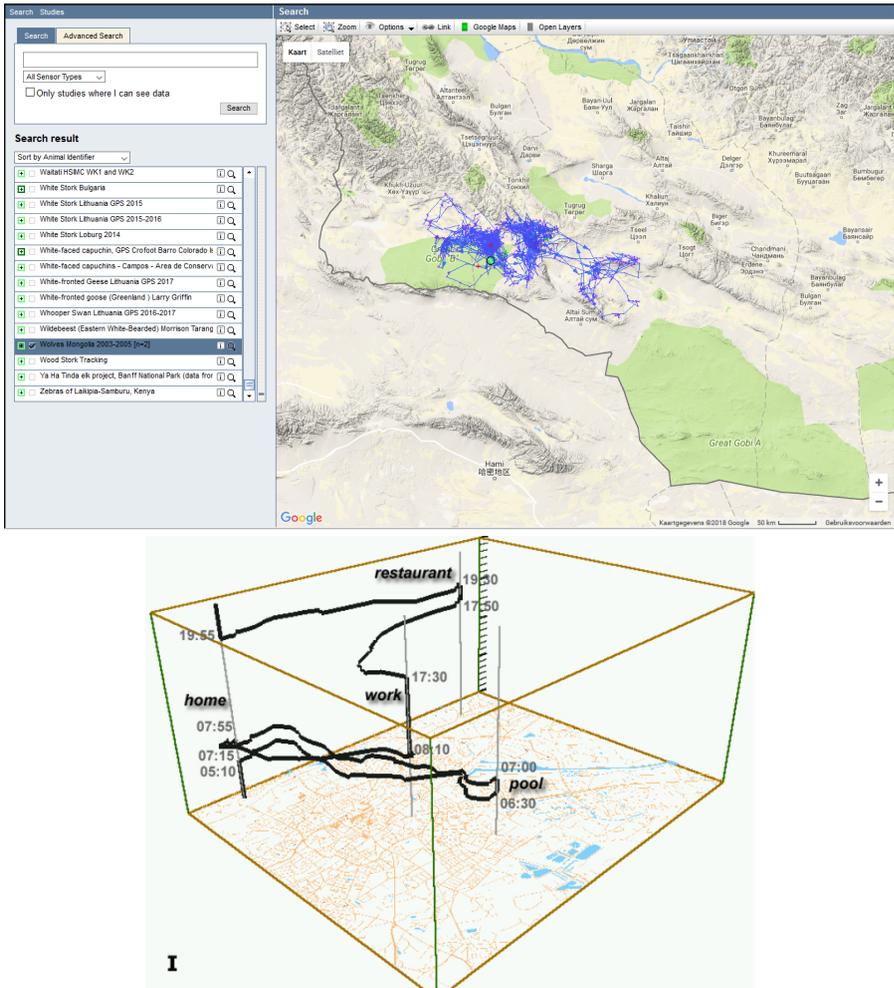


Figure 2.9: Screenshot of the tracking data of two wolves in Mongolia shown in Movebank (top) and an example of a space-time cube (bottom) (Kraak, 2003).

compass direction, displacement, turning angle, speed and acceleration, or features that describe the path shape, such as curvature, sinuosity and tortuosity. These movement and path features are used to describe movement quantitatively, to uncover structure in tracking data that can help identifying routine, to define similarities between individuals and to decompose trajectories into segments of homogeneous movement. Each segment is then presumably related to distinctive movement behaviors since foraging, escaping predators, sitting in the nest and searching for prey, all intuitively give rise to different movement characteristics (Demšar et al., 2015; Dodge et al., 2016).

This decomposition of trajectories can be achieved by means of expert knowledge or basic data analysis. Additionally, there exist several methods for (semi-) auto-

2 matically classifying tracking data into behavioral classes. Gurarie et al. (2016) distinguish four categories of methods. The first one includes metric-based methods where one calculates a movement feature, after which a clustering algorithm partitions the results into behavioral classes. The second one contains classification methods where one a priori sets a discrete number of behavioral classes for a given movement feature, after which a state is assigned to each visited location. The Bayesian partitioning of Markov models is such an approach. It assumes that the state transition is Markovian and that movement features are independent between subsequent observations (Calenge, 2006). The third category encloses the phenomenological methods identifying structure or periodicity in movement feature values, such as behavioral change point analysis (BCPA) and wavelet decomposition (Polansky et al., 2010; Gurarie et al., 2009). The former is free of a priori assumptions on the number of states (Gurarie et al., 2009). The fourth and last category uses mechanistic movement models to estimate transitions between states using a statistical framework. They involve fitting a comprehensive model of movement behavior, and are as such meant to address explanatory questions. Multi-state random walk and state space models belong to this category (Morales et al., 2004; Patterson et al., 2008).

Considering the movement and path features enables us to gain an understanding of the fundamental nature of movement. In many applications, studying movement features is more relevant than studying features of movement paths, as the former better describe the physical and biological aspects of movement (Dodge et al., 2016).

2.5.3 Quantifying space use

In addition to movement and path features, the focus may be on the spatial component of the tracking data. Here, the aim is often at better understanding where an individual spent its time.

A utilization distribution (UD) depicts the intensity of use of an area by an animal within a study area. Estimating the UD might help identifying important areas for food resources, escape cover, travel routes, mates and so on, when a set of fixes is the only available information on an individual (Kie et al., 2010). Currently, the kernel density estimate (KDE) is the most popular statistical approach to determine the UD. The UD at a location \mathbf{x} in a two-dimensional space is computed as

$$\bar{f}_h(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h^2} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right). \quad (2.1)$$

Here, $\bar{f}_h(\mathbf{x})$ is the UD at \mathbf{x} , \mathbf{x}_i represents one of the n locations visited by the individual(s), h is a smoothing parameter and K is a unimodal symmetrical bivariate

probability density function (PDF) (Gaussian for example). The actual shape of K has limited effects on the resulting UD. Fixed KDEs have a fixed h , while h varies in the case of adaptive KDEs. Choosing an appropriate h is very important, but, up to this date, no single best method for doing this has been developed (Solla et al., 1999; Kie et al., 2010).

Alternatively, Brownian bridges (BB) can be used to estimate the UD. A BB assumes Brownian motion, *i.e.* random motion of particles or individuals, between two consecutive fixes (Ross, 1996). Here, the particle's position at any moment in time is described by a normal distribution. The (time-independent) marginal probability density at each point is obtained by averaging these normal distributions over time. The resulting PDF is called the time-integrated Brownian bridge (TIBB). By calculating the weighted average of TIBBs between every consecutive pair of observations, a PDF (a UD) is obtained describing the probability density that an animal is located at a certain point at an arbitrary moment in time within the considered time interval. It is determined by the registered locations, the time between those measurements and the animal's mobility, which are essential though not accounted for in other approaches such as probability density estimates based on the density of observations (Bullard, 1991; Horne et al., 2007). The use of BBs in animal movement studies, has already been considered in depth and adapted to account for changes in behavior (Kranstauber et al., 2012), habitat preferences (Benhamou, 2011; Kranstauber et al., 2014) and moving-resting processes (Yan et al., 2014). TIBBs are preferred due to recent comparisons among space use estimators suggesting that the incorporation of temporal components leads to more reliable results than traditional kernel-based estimators (Walter et al., 2015). In Chapter 4, the analytical description of the TIBB is derived.

From the UD, the home range of an individual can be derived. The definition of a home range most referred to is the one of Burt (1943). It states that a home range is the area transversed by an individual in its normal activities of food gathering, mating and caring for young. Occasional visits outside the area should not be considered as home range (Hooten et al., 2017). The home range can be delineated on a map, consisting of a raster of pixels (or cells) organized into rows and columns, by considering the area enclosed by, for example, the 95 % contour of the UD. Therefore, a threshold is calculated so that the subset of pixels having a PDF value in their center that is greater than or equal to this threshold, represents 95 % of the volume under the PDF. As such, a contour is obtained that delineates the area having the highest possible PDF values where the individual is expected to be located during 95 % of the studied time interval (Kie et al., 2010; Hooten et al., 2017). Analogously, the core area can be defined by the 50 % contour (Hooten et al., 2017). When using the UD, the importance of areas rarely used by an animal, is minimized. Relevant travel corridors, for example, between important feeding and resting areas may not be identified. Furthermore, areas of little value that are situated close to important areas, are, on the contrary, unnecessarily labeled as

important (Kie et al., 2010).

Alternatively, a home range can be estimated by drawing a convex hull, by employing spatial statistics and trend fitting with longitude and latitude as the only covariates or by fitting mechanistic models. The latter recognize that space use is a function of the underlying patterns of movement behavior that are based on the spatial distribution of critical and limiting resources. As such, not only the fixes, but also environmental information is considered for delineating the home range (Kie et al., 2010). Therefore, species' preferences towards environmental characteristics need to be studied to identify variables that affect space use (Hooten et al., 2017).

To this day, there is no consensus on which approaches are better in delineating the home range (Hooten et al., 2017). However, traditional ones such as the KDE and convex hull will remain popular due to their simplicity (Kie et al., 2010).

2.5.4 Movement and context

Movement is often driven by the environment where the individual is moving through, which can be described by means of its networks such as roads, obstacles such as lakes and rivers, landscapes such as land cover and vegetation, ambient attributes such as weather conditions, and presence of other individuals. These environmental characteristics (ECs) can trigger a specific behavior (*i.e.* hunting, walking, running, and so on) and hence, can induce or limit movement. Therefore, movement observations alone cannot be used to unravel all mechanisms behind movement processes. It is essential to relate observations to the physical environment (Dodge et al., 2016). This is nowadays possible thanks to remotely-sensed data layers available at scales from several kilometers to less than a meter (Kays et al., 2015).

The presence of a species in an area depends on three constraints. The first one is the ecological niche (Hirzel and Le Lay, 2008). In the absence of immigration, a species can only survive and produce offspring if the combination of environmental variables is locally suitable. The number of environmental variables defines the dimension of the environmental space. The volume in this space where the species can persist and reproduce, is called its ecological niche (Hutchinson, 1957). The second constraint is the interaction with other species, which determines the amount of predation, competition and food availability. The third constraint is the accessibility, which determines the areas available for the species. Theoretically, it is possible to reconstruct a realized niche for a species from the ECs observed at the locations it occupies. This reconstruction and, as such, the determination of the important ECs and, ultimately, the prediction of presence or abundance at some location, is the main goal of several approaches (Hirzel and Le Lay, 2008).

Some popular ones, *i.e.* habitat suitability index models, resource selection functions and ecological-niche factor analysis, are considered in more detail.

Habitat suitability index (HSI) models, for example, are widely used. As a first step, candidate ECs are selected on the basis of expert knowledge or literature, and the tolerance of the species under consideration to an EC is represented by a function relating the EC to suitability, the so-called suitability index (SI). By convention, a completely unsuitable, respectively completely suitable, EC state typically corresponds to an SI value of zero, respectively of one. The SI and the state of each EC in the considered area lead to an SI value in the unit interval. Finally, the SI values of all ECs are aggregated into a single overall HSI value by taking the minimum, sum, (weighted) average or by applying another operator (U.S. Fish and Wildlife Service, 1981). This HSI value represents the overall suitability of the considered area. The SIs can either express a subjective opinion or, upon assuming a direct relationship (generally linear) between the carrying capacity and suitability (Oldham et al., 2000), can be derived from a dataset with information on the presence and absence or abundance of the species in the study area (Ortigosa et al., 2000; Ottaviani et al., 2004). The applicability of these approaches depends on data availability (Ottaviani et al., 2004).

Resource selection functions (RSFs) are popular to investigate (the characteristics of) the visited areas as compared to the environment that is available (Boyce et al., 2002). Resources often refer to all components that can be assimilated by organisms (food and water) (Guisan and Thuiller, 2005). Nevertheless, any EC of interest can be included. An RSF-based model is a kind of HSI model that uses statistical methods instead of expert knowledge or other approaches where model parameters are not directly estimated from data. It is usually derived from presence/absence observations of the species or, since it is not possible to define non-use in telemetry studies, from presence/availability data (Boyce et al., 2002). In an RSF-based model, the ECs are used to define the PDF describing the probability density of use of an area within the study area. Therefore, the availability and selection of the ECs are both modeled as nonnegative functions with the states of the ECs at the considered area as independent variables. The product of both functions is proportional to the spatial density of visits in the area. If the product is normalized (by dividing the product of both functions by the sum of the products determined at all areas in the study area), a PDF is obtained that serves as a probabilistic model. In basic studies, the availability of an EC is described by a uniform distribution function, as such assuming that an individual can be located anywhere in the study area with equal probability. The selection function can have any form as long as it yields positive values. The exponential and logistic functions having a linear combination of the states of the ECs as input are the most popular ones. More specifically, these functions are used as mean functions in a generalized linear model (GLM) having the EC states as independent variables. The exponential function is used if the number of observations in an area is modeled. In this case,

the outputs are strictly positive integer values, and they are Poisson distributed. Alternatively, the logistic function is used if the probability of presence is modeled. Here, the output values lie between zero and one, and they are Bernoulli distributed. In order to obtain the coefficients of the linear function combining the EC states, maximum likelihood or Bayesian analysis is used (Hooten et al., 2017).

2 When only presence data is available and the states of the ECs are likely non-independent, a principal component analysis (PCA) or factor analysis (FA) may be performed to reduce the number of variables by combining them in so-called latent variables. Furthermore, such analyses indicate which of these latent variables explain most of the variance (Jolliffe, 1990; Yong and Pearce, 2013). Here, we discuss the ecological-niche factor analysis (ENFA), a factor analysis building upon Hutchinson's concept of an ecological niche (Hirzel et al., 2002; Basille et al., 2008), which is used to determine the most important ECs (Hirzel et al., 2002; Basille et al., 2008). In Hutchinson (1957), the ecological niche is defined as the hypervolume in the multidimensional EC space where the species can maintain a viable population. ENFA involves computing directions in the space of variables in such a way that the first direction measures the deviation of the species' niche from the average prevailing conditions, the so called marginality, while the others quantify the narrowness of the niche, the so called specialization (Fig. 2.10). The first direction is chosen to account for all marginality, the others to maximize specialization (Hirzel et al., 2002). Every axis of the multidimensional niche (one per EC) can be characterized by an index of marginality and specialization. As such, ENFA allows to extract the combination of the original variables for which the species shows most of the marginality and specialization (Hirzel et al., 2002). Given the fact that a study area is used as reference in ENFA, its results will strongly depend on the delineation of this area (Hirzel et al., 2002), and the latter should be decided by keeping the research objectives in mind.

Other approaches are generally referred to as habitat selection models, habitat/species distribution models, ecological niche models or gradient analyses (Hirzel and Le Lay, 2008). The variety of statistical techniques used in these approaches, such as generalized regressions, classification techniques, environmental envelopes, ordination techniques, Bayesian techniques, other techniques and even a combination of the aforementioned techniques are discussed in Guisan and Zimmermann (2000).

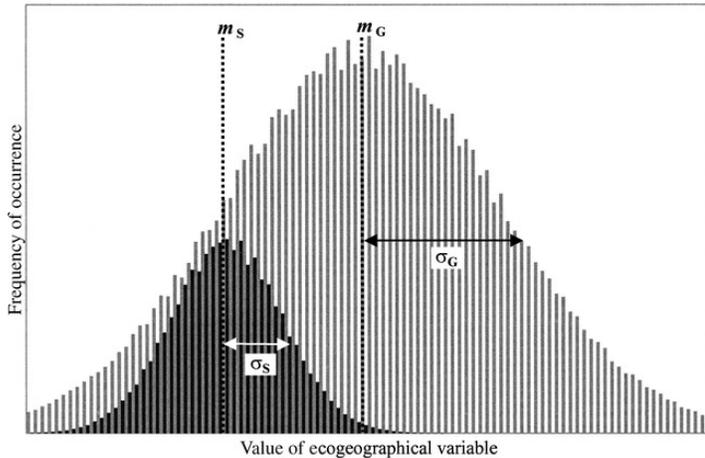


Figure 2.10: The distribution of the focal species on any ecogeographical variable (black bars) may differ from that of the whole study area (gray bars) with respect to its mean and standard deviation, thus allowing marginality and specialization to be defined (Hirzel et al., 2002).

2.6 Movement modeling

2.6.1 Introduction

Movement models are used for analyzing, describing, simulating and predicting movement. Animal movement can be modeled in many ways, depending on the conceptualization of space, time and state as discrete or continuous domains. Moreover, the movement process can be represented by the locations themselves, or by derived quantities, such as step lengths and turning angles (McClintock et al., 2012).

Two classical modeling approaches to study animal movement can be distinguished. The first one is the Lagrangian modeling approach where (the location of) an individual is followed in time. The Eulerian approach, on the other hand, focuses on how the probability of occurrence changes over time at some location in the study area. It is location-based instead of individual-based, and it is used to predict expected patterns of space use by an individual or a population in time. It typically involves diffusion models that make use of the Fokker-Planck equation for space use (Smouse et al., 2010). This equation describes the UD of an individual under the influence of drag forces and random forces. If only random forces occur, the Fokker-Planck equation describes Brownian motion (Codling et al., 2008). Alternatively, the UD might, for example, be determined by means of the KDE or RSFs (Smouse et al., 2010). Lagrangian models are often (discrete time) individual-based models (IBMs) and Eulerian models (continuous time) population-based model (Hooten et al., 2017). In this dissertation, the Lagrangian approach is

of interest. In McClintock et al. (2014), references to Lagrangian animal movement models are given.

Movement is, like many other processes in nature, too complex to be described accurately in every aspect of space, time and state in a single model. There is often a trade-off between model generality, sense of reality and precision. Depending on the focus, three types of models can be distinguished (Guisan and Zimmermann, 2000). When the focus is on generality and precision, the model is called analytical. Such models are designed to accurately predict the response of a system under consideration. In order to be accurate, reality often needs to be simplified and limited. When the focus is on reality and generality, the model is called a mechanistic, causal or process model. Such models base predictions on cause-effect relationships and are evaluated on the basis of the theoretical correctness of the predicted responses, rather than on the precision. When focus is on precision and reality, the model is called empirical, statistical or phenomenological. These models do not describe cause-effect relationships between model parameters and responses and, as such, do not inform about underlying ecological mechanisms.

In practice, it is difficult to classify models according to this classification. Moreover, generality, sense of reality and precision are not always mutually exclusive (Guisan and Zimmermann, 2000). In this section, both basic statistical models and models including context, movement behavior, expert knowledge, and so on are considered. Since many models have been developed, our intention is not to provide an exhaustive list of all modeling approaches, but rather to give an idea about basic statistical and more mechanistic models. In the following, we assume no (or little) measurement error associated with the telemetry data so that locations can be modeled directly. As a consequence, state-space models, accounting for telemetry error (Hooten et al., 2017), are not considered.

2.6.2 Continuous time–continuous space models

The process of movement occurs in continuous time and space. Continuously in time means that for any time, a location can be observed. Therefore, continuous time–continuous space models are theoretically the most appealing. Moreover, they allow for a compact notation and facilitate a mathematical analysis using tools from calculus. They are, however, often very technical for non-statisticians (McClintock et al., 2012; Hooten et al., 2017).

In this setting, Brillinger et al. (2004) and Preisler et al. (2004) use stochastic partial differential equations to describe the incremental step an animal takes at

time t at location $\mathbf{x}(t) = (x(t), y(t))^T$:

$$\begin{pmatrix} dx(t) \\ dy(t) \end{pmatrix} = \begin{pmatrix} \mu_x(\mathbf{x}, t) \\ \mu_y(\mathbf{x}, t) \end{pmatrix} dt + \mathbf{D}(\mathbf{x}, t) \begin{pmatrix} d\psi_x(t) \\ d\psi_y(t) \end{pmatrix}. \quad (2.2)$$

Here, $dx(t)$ and $dy(t)$ are the incremental step sizes in the x and y directions respectively, $\mathbf{u}(\mathbf{x}, t) = (\mu_x(\mathbf{x}, t), \mu_y(\mathbf{x}, t))^T$ is a drift parameter, \mathbf{D} is the diffusion matrix that describes the correlation between steps in the x and y direction. $\psi_x(t)$ and $\psi_y(t)$ are random processes with expected values zero that introduce variability. These random processes determine, together with \mathbf{u} and \mathbf{D} , the traveled distance and direction.

From this general model, several particular cases can be obtained (Preisler et al., 2004):

- If $\mathbf{u} = (0, 0)^T$, $\mathbf{D} = \sigma^2 I_2$, with I_2 the 2x2 identity matrix, and $\psi_x(t)$ and $\psi_y(t)$ are independent Brownian processes, where the increments follow a normal distribution with mean zero and variance dt , a continuous uncorrelated random walk, *i.e.* Brownian motion, is obtained. Here, animals travel at random.
- If $\mathbf{u} = \mathbf{A}$, $\mathbf{D} = \sigma^2 I_2$, with I_2 the 2x2 identity matrix, and $\psi_x(t)$ and $\psi_y(t)$ are independent Brownian processes, Brownian motion with drift is obtained. Here, animals drift in the direction of the vector \mathbf{A} or are attracted towards a central location, such as a food patch and home range. The latter is called an Ornstein-Uhlenbeck process where the location (instead of the direction) of attraction stays unchanged. Alternatively, potential functions depending on space or on space and time, can be used to highlight areas of strong attraction and steer movement towards them. They allow for a more realistic way of guiding animal movement, and are often referred to as force fields (Preisler et al., 2004).
- If $\mathbf{u} = (0, 0)^T$, $\mathbf{D} = \sigma^2 I_2$, with I_2 the 2x2 identity matrix, and $\psi_x(t)$ and $\psi_y(t)$ are positively correlated increments, correlation is introduced. Since the individual's consecutive moves are correlated, it tends to move with persistence in a given direction with a given speed.
- Other models can be obtained, for example, by specifying non-Markovian random processes $\psi_x(t)$ and $\psi_y(t)$, or by involving time lags in the description of the drift term, \mathbf{u} .

By integrating Eq. (2.2) in time from 0 to time instance t , $\mathbf{x}(t)$ is obtained (Hooten et al., 2017):

$$\mathbf{x}(t) = \mathbf{x}(0) + \int_0^t \mathbf{u}(\mathbf{x}, \tau) d\tau + \int_0^t \mathbf{D}(\mathbf{x}, \tau) \begin{pmatrix} d\psi_x(\tau) \\ d\psi_y(\tau) \end{pmatrix} d\tau. \quad (2.3)$$

As such, the location of an animal at time t equals the starting location plus the cumulative drift and cumulative diffusion.

Equation (2.2) offers several modeling opportunities going from a purely statistical to a more mechanistic approach. After choosing the appropriate model structure, the model parameters are defined by including a priori knowledge and by using tracking data to fit the model. Since tracking data are not continuously available, discretization is needed for fitting the model (Hooten et al., 2017). A common practice to parameterize mechanistic models involves the iterative adjustment of the model parameters so that the model output reproduces the real system (*i.e.* inverse modeling). This requires an extensive number of simulation runs, which might be impractical for long simulations. Forward-modeling techniques, on the other hand, assess the model parameters directly from the available data (Latombe et al., 2014). This is, for example, done by maximizing a likelihood function (Patterson et al., 2017).

2.6.3 Discrete time–continuous space models

Since the dynamics of movement are easily conceptualized in discrete time by means of a series of turning angles and step lengths (covered distance per time interval or speed), a discrete time context is valuable, practical, simple and easy to interpret biologically. Also during simulations performed by computers, discretization is unavoidable (McClintock et al., 2014; Hooten et al., 2017).

In discrete time–continuous space models, a random walk is the basic model assuming independent increments $\Delta \mathbf{x}(t + \Delta t) = \mathbf{x}(t + \Delta t) - \mathbf{x}(t)$, following a normal distribution with mean zero. In the simplest case, the covariance matrix is specified as $\sigma^2 I_2$ with σ^2 the variance and I_2 the 2x2 identity matrix. This model implies that an animal moves in a random direction with step lengths drawn from a Weibull distribution, where σ^2 controls the step length. Generally, we can write

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \epsilon(t + \Delta t), \quad (2.4)$$

where $\epsilon(t + \Delta t) \sim \mathcal{N}(0, \sigma^2 I)$ (Hooten et al., 2017).

Several extensions providing insights into the dynamics and behavior of moving animals can be added to the random walk. A location of attraction, \mathbf{x}^* , for example, can be included as

$$\mathbf{x}(t + \Delta t) = M \mathbf{x}(t) + (I - M) \mathbf{x}^* + \epsilon(t + \Delta t), \quad (2.5)$$

where the propagator matrix M controls the dynamics towards \mathbf{x}^* and is, in the simplest case, equal to ρI . In the latter case, the dynamics in latitude and longitude are the same and mutually independent. The smoothness is controlled by

ρ , and, in order to have, on average, attraction point \mathbf{x}^* , it must hold that ρ lies between zero and one. If there are several points of attraction, \mathbf{x}^* can arise from a mixture of these points, each associated with a probability. Moreover, the location \mathbf{x}^* can change in time to allow for a changing attraction dynamics (Hooten et al., 2017).

The models of Tracey et al. (2005) and Hubbard et al. (2004), involve modeling turning angles and step lengths, and use the following basic movement equation for determining consecutive locations:

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + v(t) a(t) \Delta t, \quad (2.6)$$

where $v(t)$ represents movement speed and $a(t)$ movement direction:

$$a(t) = \begin{pmatrix} \cos(\theta(t)) \\ \sin(\theta(t)) \end{pmatrix}. \quad (2.7)$$

Here, $\theta(t)$ represents the angle of movement relative to the positive x -axis. Tracey et al. (2005) assume that speed and direction follow a chosen distribution function, while Hubbard et al. (2004) calculate the movement direction $a(t)$ as a weighted average of two terms, the first one representing the tendency to follow the group (described by the average direction of the neighbors) and the second one the tendency to visit highly suitable areas (implemented by means of a potential function). In both models, individual variability is introduced, which is a valuable asset when performing simulations and predictions (Huston et al., 1988). Moreover, the model of Hubbard et al. (2004) also offers several modeling opportunities since more terms can be included into the weighted average, the behavioral state can be time-dependent, speed can be variable, memory can be considered, and so on.

We adapted the model of Hubbard et al. (2004) to build an individual-based movement model for Pike (*Esox Lucius*) in the River Yser, Belgium (Baetens et al., 2013). Here, the behavior of an individual is determined by a seasonally varying attraction towards its territory and spawning grounds. The degree of attraction is determined by two parameters α and β , where α steers the attraction towards the spawning grounds and β the attraction towards the individual's territory (Fig. 2.11(a)). The potential function describing the attraction towards the three spawning grounds along the river stretch, is shown in Fig. 2.11(b).

Also in the case of discrete time models, the model parameters need to be defined after choosing the appropriate model structure. This can be done by including a priori knowledge and by using tracking data to fit the model (Patterson et al., 2017). Both inverse and forward techniques can be used for this purpose (Latombe

The individual-based movement model for Pike (*Esox Lucius*) in the River Yser, Belgium was built during my master thesis, after which my supervisors performed a model validation (Baetens et al., 2013).

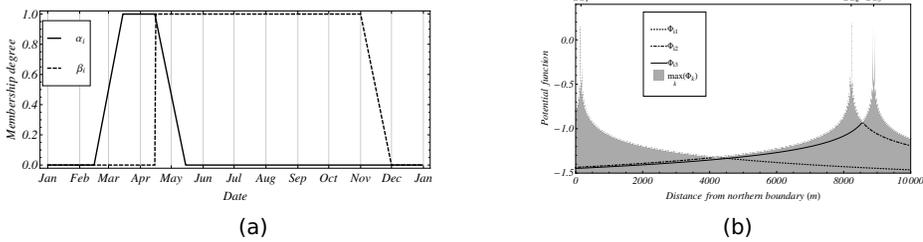


Figure 2.11: (a) Two parameters $\alpha_i(t)$ and $\beta_i(t)$, defined as trapezoidal fuzzy sets, steer the attraction of pike i towards the spawning grounds and towards its territory. (b) Potential functions Φ_1 , Φ_2 and Φ_3 corresponding with three spawning grounds SG_1 , SG_2 and SG_3 together with the overall potential function $\max_k(\Phi_k)$ (shaded area).

et al., 2014). As mentioned previously, it is often more intuitive for ecologists to think about movement in a discrete time setting as a series of step lengths and turning angles. However, during the discretization of the movement path into a finite number of equidistant time points, the specification of the time step length is critical. The time step length needs to match the scale at which the behavioral decisions are made (Hooten et al., 2017).

2.6.4 Discrete space models

Space can be conceptualized as a discrete domain by overlying the study area with a lattice of discrete identical sites. Since we are considering Lagrangian models, we are not interested in the state of the sites themselves, but rather in the consequent sites visited by an individual and the corresponding residence times. This type of conceptualization might be more appropriate when the focus is on the effect of environmental conditions, rather than on the individual characteristics and behavior, since environmental information is typically provided in a discrete way, as a map (Hanks et al., 2015).

2.6.5 Behavioral states

As mentioned earlier, the model parameters can vary over time in order to capture behavioral dynamics. Each set of parameters then corresponds to a behavioral state. Usually, two or three states are defined a priori as part of the model structure. A hidden Markov model, for example, is a discrete time movement model where each state (typically two) is considered as a distinct random walk. Alternatively, behavioral states can be conceptualized as continuous. Here, model parameters are allowed to vary smoothly in time (Breed et al., 2012; McClintock et al., 2014; Hooten et al., 2017).

Discrete time models that simultaneously incorporate multiple behavioral states, Markov-state switching, correlated movements, and directed movements have already been developed and fitted to data. Similar applications of multi-state mixture models have not yet appeared in continuous time. Here, fitting the models is challenging because of the increase in mathematical complexity. Because multi-state models are of great practical importance when investigating animal movement, this currently remains an advantage of discrete time models. However, once the continuous time models have caught up with their discrete time counterparts for what concerns the incorporation of multiple behavioral states, continuous models are expected to become more widely used because of their computational efficiency (McClintock et al., 2014; Hooten et al., 2017).

2.7 Movement simulation and prediction

A constructed movement model can be used to imitate the real-world movement phenomenon under consideration by performing simulations. Simulation is a computational approach which allows for an explicit representation of the movement process underlying the observed movement paths. Because research is often focused on individual movement, computing technology advanced and a lot of tracking data are nowadays available, individual-based models (IBMs) are widely used to uncover complex movement behavior (Tang and Bennett, 2010). Such models represent distinct entities, such as individual animals. In the case of a mechanistic IBM, the dynamic nature of the internal state and the individual's relation with the environment are described. The in-silico individuals are goal driven and try to fulfill specific objectives, they are aware of and can respond to the environment, they can move and they can be designed to learn and adapt their state and behavior in response to stimuli of other agents. Moreover, there is a differentiation in the individuals' behaviors which allows the simulation of emergent patterns at population scale. As such, our ecological understanding of the species is increasing (McLane et al., 2011; Latombe et al., 2014; Daly et al., 2016; de Ulzurrun et al., 2017). As an illustration, some simulation results of the individual-based movement model for pike in the river Yzer (cfr. Subsection 2.6.4 and Baetens et al. (2013)) are shown in Fig. 2.12.

Mechanistic IBMs are also useful for predicting the impacts of changing environmental conditions and management plans. Predictions can be made to new sites within the range of the environment and time frame sampled. This is usually sufficiently reliable for conservation planning and resource management, and for identifying suitable habitats for rare species. Prediction can, on the other hand, also be made to new sites outside the sampled range or time frame. Such an extrapolation or forecasting is risky, but often the only tool available to assess future events (Elith and Leathwick, 2009). When the focus of movement modeling

is on predicting, rather than on increasing the understanding of the species under consideration, the trade-off between generality and fit to the training data has to be well balanced so that overfitting is avoided and reliable predictions can be made (Elith and Leathwick, 2009; Latombe et al., 2014).

Before using a model, its performance must be checked. Usually, this is done by statistical tests of model fit and comparison with existing knowledge. Moreover, predictive performance can be verified by data resampling techniques, such as cross-validation, or by making use of independent data sets (Elith and Leathwick, 2009).

2.8 Management implications

Wildlife species are under pressure from both natural and anthropogenic influences, including climate change, pollution, habitat loss and fragmentation. In wildlife management, critical habitats to conserve a certain species are identified. However, delineating areas to be protected does not suffice for an effective conservation. Management actions may be limited by costs, stakeholders interest, human-wildlife conflicts and policy enforcement. Taking these limitations in account, the most cost-effective action needs to be selected that maximizes the benefits. Here, incorporating movement ecology in decision making helps prioritizing the most effective management action that has the highest chance of success (Allen and Singh, 2016).

Tracking data itself can be valuable for management purposes. For example, virtual fencing reduces human-wildlife conflicts by sending alerts when tagged animals cross predefined areas. In this way, crop raids by, for example, tagged elephants, are reduced. Furthermore, live data provide a powerful tool to engage public interest. In this way, citizens are more involved and motivated to help conserving wildlife (Kays et al., 2015).

Analyzing movement data helps gathering knowledge about critical habitat patches, resources, stop-over sites, corridors and wintering sites, and about biogeography and barriers (Hirzel and Le Lay, 2008; Kays et al., 2015). This knowledge allows ecologists to define ecological requirements of species and their limiting factors, to identify reintroduction sites, to find unknown populations and to design conservation plans and reserves. Furthermore, the effectiveness of these plans can be increased by linking movement ecology and applying strategies that are flexible in time and/or space. This results in actions along the entire movement path, dynamic protected areas, time-restricted harvesting practices and management actions to overcome barriers and loss of important sites (Allen and Singh, 2016).

In addition to movement analysis, movement modeling is used for further increasing the understanding of the species under consideration. This is achieved by

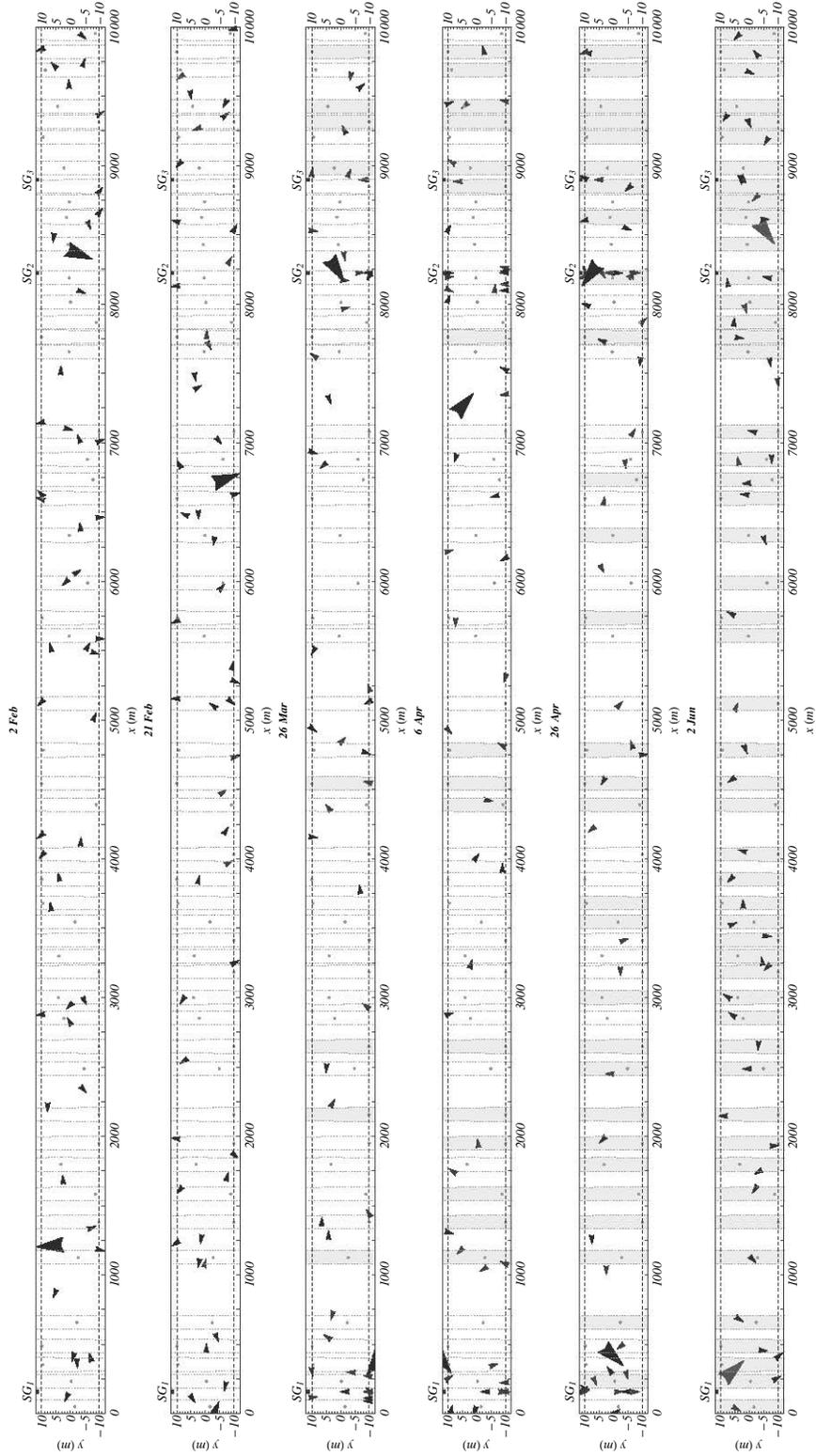


Figure 2.12: Simulated spatial distribution of the individuals (arrowheads) within the study area at the beginning of six different days during an annual cycle together with the territories (dotted lines). The individuals, of which two are shown by large arrowheads for tracking purposes, are represented by black arrowheads outside the spawning season and gray arrowheads otherwise. The spawning grounds are indicated as SG_k.

considering movement behavior in a more quantitative way and summarize it in a set of rules. Furthermore, when the model describes the underlying movement behavior of the species under consideration, movement models can be used to test hypotheses about movement and predict how the species will likely respond to their rapidly changing environment. In order to assess the impact of multiple scenarios, a scenario analysis can be performed. This is valuable for assisting decision making (McLane et al., 2011). In particular, IBMs have proven to be useful. They have been used to study the differential effects of predator sterilization versus removal, of habitat selection and population demographics of animals under predicted future human development and habitat loss, of impacts of different management options to reduce road-site mortality (McLane et al., 2011).

As a consequence of the huge amount of data and knowledge gathered from animal tracking in the last years, Demšar et al. (2015) and Kays et al. (2015) believe that we are moving from a point where tracking data is used to learn us about animals to a point where tracking data is used to teach us about the world. They suggest a new approach that views animals as sensors of the environment that potentially help us to monitor the planet. Although further miniaturization of sensors is therefore needed, forecasting on the basis of predictable animal movements can help identifying and solving environmental problems. For example, animals may be able to anticipate upcoming natural disasters and change their movement behavior accordingly. If this change in behavior can be observed by humans, one is also able to anticipate and to be prepared for that (Kays et al., 2015).

3

Spatially explicit modeling of the Belgian major endurance event '100 km Dodentocht'

3.1 Introduction

The importance of a healthy life incites people to practice sports. Some are trying to push back their boundaries increasingly further, while others are looking for new challenges. In this context, endurance sports events like (ultra)marathons, triathlons and bicycle races are nowadays becoming very popular. Generally speaking, more and more such events are organized and their number of participants is steadily growing. For example, the number of US marathon finishers has grown from about 25,000 in 1967 to approximately 541,000 in 2013 (Running USA with

This chapter was published as Van Nieuland et al. (2016). Additionally, this chapter is extended with weather info, analysis of gps tracking data and model validation on the basis of the RFID tracking data from the 2017 edition.

Athlinks (2009-13) and Active.com (2005-08), 2014), and the total number of participants in the Antwerp 10 Miles, Antwerp, Belgium increased tenfold over 10 years, with over 40,000 participants during the last edition (Vananderoye, 2014). Also long distance marches, such as the Kennedy March in the Netherlands, whose participants have to cover 80 kilometers in at most 20 hours (Kennedy-Mars Sittard, 2015), and the ultra challenge series in the United Kingdom, where participants have to walk 100 kilometers within approximately 20–30 hours (Actionchallenge, 2015), reflect this trend.

The event '100 km Dodentocht', which takes place annually and has its start in Bornem, Belgium, is yet another example of a long distance march. Its marchers have to cover a 100 km trail in at least 10 and at most 24 hours. The first edition took place in 1969 and had only 65 brave youngsters at the start, which clearly contrasts the approximately 11,000 marchers that appeared at the start of the most recent editions. First and foremost, the organizing committee aims at getting people on the move and tries to provide the marchers with a never-to-be-forgotten experience, rather than engaging them in a competition for arriving first at the finish (100 km Dodentocht Kadee Bornem, 2015).

The perception of marchers in endurance sports events is very important for maintaining the success of such events. Information technology has enhanced the way athletes experience these events: online registration systems avoid long queues, tracking systems help with scoring and allow marchers, as well as their friends and family, to check their performance immediately after – or even during – races, and compare it with their results in previous editions or with those of other marchers (Malkinson, 2009).

Using tracking data collected during the editions 2009–2014, the typical dynamics of the 100 km Dodentocht is studied in this chapter and a spatially explicit marching model is built to mimic the dynamics of marchers, as such providing a means to the organizers to assess the effect of potential modifications to the starting procedure on the event dynamics. More specifically, this model allows for simulating the speed of every marcher individually along the trail. These speeds can then be used to calculate the passing times of the participating marchers at the consecutive checkpoints in order to allocate the facilities and staff at these checkpoints. To the best of our knowledge, no studies have addressed the facts and figures of endurance events from this point of view.

In Section 3.2, an overview of the tracking data set is given, after which we present its analysis and build and validate the spatially explicit marching model in Sections 3.3 and 3.4. Finally, we analyze how a modified starting procedure might change the dynamics of the endurance event and formulate guidelines to the organizers for optimizing the first phase of the 100 km Dodentocht in Section 3.6. The discussion can be found in Section 3.7.

3.2 Data collection and overview

Like with many other endurance events, the marchers in the 100 km Dodentocht are tracked along the trail by making use of passive radio frequency identification (RFID) (Grillmayer, 2013). More specifically, a battery-less miniature transponder is attached to the participant and gets activated when its carrier steps over an antenna mat that generates a magnetic field and makes the transponder transmit its unique identification number, which is passed to a computer for further processing (Malkinson, 2009). Along the 100 km Dodentocht trail, the mats are located only at the entrance of every checkpoint. Once the presence of a 100 km Dodentocht marcher at such a checkpoint has been registered, the time of this registration and the average marching speed between the present and the previous checkpoint is passed in real-time to his/her online profile in order to maximize the involvement of his/her followers (100 km Dodentocht Kadee Bornem, 2015). Since there is only one registration per checkpoint, it is important to note that it is impossible to identify the duration of possible resting periods at checkpoints.

Unlike the use of video or GPS technology or the use of health wearables to improve the performances of individual athletes or sports teams (Baca et al., 2009), RFID data have, to the best of our knowledge, not yet been exploited to gain insight into the dynamics of endurance sports events like long-distance marches, let alone the use of such data to develop a dedicated spatially explicit marching model. This is notable because such an approach could support the organizers in making well-founded decisions on how to improve the experience for their marchers and optimize the allocation of staff and logistics to the different checkpoints along the trail.

The organizing committee of the 100 km Dodentocht provided the RFID data of all marchers who participated in the 100 km Dodentocht marches that were held between 2009 and 2014, as well as the outline of the trail. In addition, the organizing committee provided the gender and birth date of the marchers in the editions 2010–2014. Since no other personal details were registered at the time of registration, characteristics such as overall health condition, fitness level, amount and type of food are not known and their effect could not be investigated.

Leaving aside minor changes, the trail and the locations of the checkpoints remained the same over the years covered. As such, we may assume that the influence of the trail configuration did not have an impact on the dynamics of the participating marchers across the subsequent editions. The distances between the checkpoints and the start are given in Table 3.1. As usual, the start and finish of the 100 km Dodentocht were located in Bornem, Belgium and the loop trail passed through several villages and cities in the provinces of Antwerp, Flemish Brabant and East Flandres (cfr. Fig. 3.1). At every checkpoint, marchers were offered snacks and drinks, while a warm meal was offered at checkpoint 7 (50.0 km).

Each checkpoint had opening and closing times that were determined on the basis of the minimum and maximum allowed marching speed of about 4.2 km h^{-1} and 10 km h^{-1} , respectively. Marchers arriving before the opening time of a checkpoint had to wait at its entrance until its predefined opening time, while the ones arriving past closure time were disqualified. As a consequence of the imposed marching speed limits, checkpoints at a larger distance from the start had longer opening hours than those located along the first part of the trail and therefore required more operational efforts. In the remainder of this paper, the segment of the trail located in between two checkpoints is referred to as a section. Every section is assigned an ID, being the number of the checkpoint at the beginning of the section. For every individual marcher, the RFID data was collected until the last recorded checkpoint, *i.e.* the checkpoint at which a marcher was disqualified or finished. The last recorded checkpoint of marcher i is referred to with the symbol N_i .

Table 3.1: Distances along the trail between the checkpoints and the start in 2009–2014. Also the average distances were calculated.

| Checkpoint ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|-----------------------|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|-------|-------|
| Average Distance (km) | 1.2 | 7.5 | 17.6 | 24.6 | 31.7 | 39.8 | 50.0 | 56.2 | 63.8 | 69.1 | 74.2 | 80.3 | 84.6 | 89.7 | 94.4 | 100.0 |
| Distance (km) in 2009 | 1.2 | 7.6 | 17.6 | 24.7 | 33.4 | 40.1 | 50.2 | 57.5 | 66.7 | 71.8 | 76.0 | 81.2 | 85.5 | 90.6 | 94.6 | 100.0 |
| 2010 | 1.2 | 7.4 | 17.3 | 24.3 | 33.1 | 39.8 | 49.2 | 56.5 | 65.7 | 70.8 | 75.0 | 80.2 | 84.6 | 89.7 | 94.6 | 100.0 |
| 2011 | 1.2 | 7.4 | 17.3 | 30.4 | 39.3 | 49.5 | 56.8 | 65.0 | 70.4 | 74.6 | 80.3 | 84.3 | 89.3 | 94.2 | 100.0 | |
| 2012 | 1.2 | 7.4 | 17.3 | 24.4 | 30.4 | 39.3 | 49.5 | 54.5 | 60.8 | 66.3 | 72.9 | 79.4 | 84.5 | 89.8 | 94.6 | 100.0 |
| 2013 | 1.2 | 7.6 | 17.8 | 24.7 | 31.4 | 39.9 | 50.3 | 55.1 | 61.5 | 66.8 | 73.6 | 80.6 | 84.6 | 89.7 | 94.5 | 100.0 |
| 2014 | 1.3 | 7.7 | 18.1 | 25.1 | 31.7 | 40.5 | 51.1 | 57.0 | 63.4 | 68.8 | 73.1 | 80.2 | 84.2 | 89.3 | 94.2 | 100.0 |



Figure 3.1: The red flags indicate the checkpoints (labeled by their ID) and the thick line indicates the trail of the 2014 edition. (Source: National Geographic, Esri and AGIV)

The entire data set contains RFID data of 65,552 marchers, corresponding with 790,230 scans. Across the editions covered in this study, there were on average 22,81% female and thus 77,19% male marchers.

Figure 3.2 depicts the relative frequencies of the ages of all marchers included in the data set. Marchers between 25 and 30 years old were the most common, while there were fewer marchers aged between 31 and 50 years, though the latter were equally present. Marchers older than 50 and younger than 20 were under-represented. The relative frequencies were somewhat skewed to the right for male marchers as compared to women, which means that relatively more male marchers were older than 50. The same conclusion can be drawn from the relative frequencies of the individual years, which are not shown for the sake of brevity.

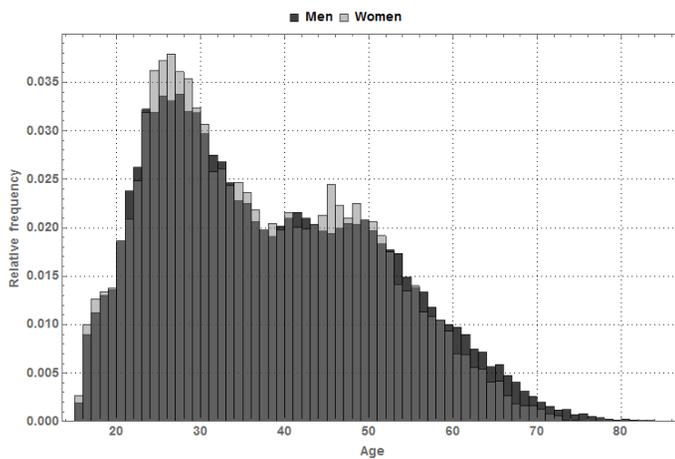


Figure 3.2: Relative frequencies of the age of the marchers (men and women) of the 100 km Dodentocht in 2010–2014. The bar plots are transparent and plotted on top of each other. As such, colors are mixed where the bar plots for men and women coincide.

After the construction and cross-validation of the model using the data of 2009–2014, the model was used for determining the dynamics of the upcoming edition of 2017. Afterwards, the organizing committee provided the RFID tracking data so that we were able to perform an extra validation of our model with an independent data set.

Next to the RFID tracking data, we also had GPS tracking data of 711, respectively 740 participants of edition 2013, respectively 2014 at our disposal. The GPS data were derived from a smartphone app, developed for the event, with which participants could record their positions at several time instances. The latitude and longitude of each location, the time stamp, the battery level and the accuracy were registered.

The meteorologic conditions are external conditions that might influence the dynamics of the event. Rain (showers) and cold nights are, according to participants,

a hard nut. The average temperature and amount of precipitation at three weather stations near the trail (Kapelle-op-den-Bos at 51°00' N and 4°21' E, Bornem at 51°05' N and 4°13' E and Moerzeke at 51°04' N and 4°10' E, cfr. Fig. 3.3) during the editions 2009–2014 were available. Only in 2011 and 2014 there were rain events. For what concerns the temperature, 2011, 2013 and 2014 were mild with a relatively small difference between the lows and highs. On the contrary, in 2010 and 2012 lows dropped to 10°C and 12°C, respectively, while the edition 2009 stands out by high temperatures during the entire march (Fig. 3.4).

3.3 Data analysis

3.3.1 RFID tracking data

3.3.1.1 Overall statistics

On average, 10,925 marchers showed up at the start of which 61% reached the finish in about 21 h. An overview of the number of marchers, the average marching times and the success rate per edition can be found in Table 3.2. These numbers were calculated from all the available scans (790,230). In the remainder, however, only marchers with a complete set of scans, *i.e.* a set without missing registrations and without registrations after retiring, are considered. As such, the number of considered marchers is lower than the total number of marchers (cfr. Table 3.2).

Table 3.2: General statistics of the 100 km Dodentocht in 2009–2014. The number of marchers, the average marching time and success rate per edition and the number of marchers with a complete set of scans.

| Edition | # Marchers | Average marching time | Success rate (%) | # Marchers with a complete set of scans |
|---------|------------|-----------------------|------------------|---|
| 2009 | 10,428 | 20 h29 min | 61.4 | 10,210 |
| 2010 | 10,605 | 20 h22 min | 61.4 | 7,441 |
| 2011 | 10,507 | 20 h25 min | 58.3 | 8,462 |
| 2012 | 10,964 | 20 h28 min | 62.2 | 9,258 |
| 2013 | 11,181 | 20 h30 min | 62.4 | 10,237 |
| 2014 | 11,867 | 20 h20 min | 58.0 | 11,149 |

The cumulative relative frequency of men and women retiring as function of the distance from the start and the time elapsed since the start can be found in Figs. 3.5(a) and 3.5(b), respectively. The chosen bin width for the first one corresponds to the distance between the consecutive checkpoints (the vertical grid-lines represent the locations of the checkpoints), while for the latter it was set to 0.5 h. Most people retired at checkpoint 6 (39.8 km) and 7 (50.0 km). A similar conclusion can be drawn from Fig. 3.5(b). Indeed, some marchers retired after 3

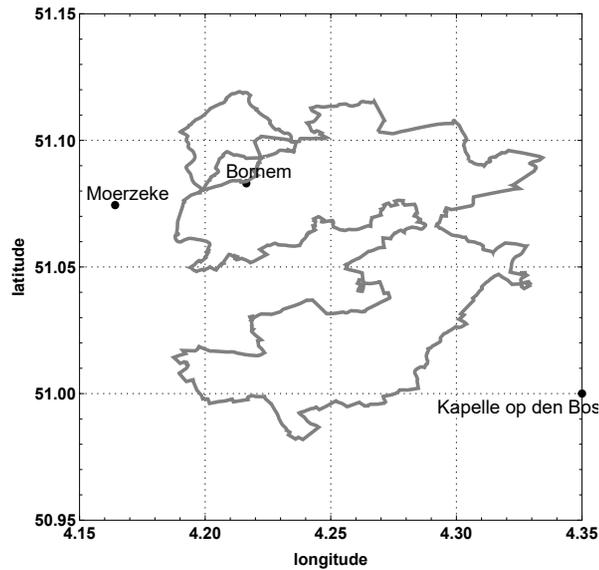


Figure 3.3: The weather conditions in 2009–2014 were available from three weather stations near the trail: Kapelle-op-den-Bos at $51^{\circ}00'$ NB and $4^{\circ}21'$ OL, Bornem at $51^{\circ}05'$ NB and $4^{\circ}13'$ OL and Moerzeke at $51^{\circ}04'$ NB and $4^{\circ}10'$ OL.

to 5 hours, but most marchers retired after 7 to 12 hours. Once having marched 15 hours, most marchers kept on going.

Although the overall shape of the curves is similar across the editions, it can nonetheless be noted from Fig. 3.5 that in 2014, people retired earlier. Probably, this was caused by the bad weather conditions at the beginning of the march. Furthermore, there is a notable increase in the number of marchers retiring near the end of the 2011 edition, which can be understood by the fact that it started raining 13 hours after the start of this edition (cfr. Fig. 3.4). Hence, we can conclude that rain might be an important factor inciting marchers to give up. Likewise, cold nights (2010) induced retiring earlier and at shorter distances from the start. Finally, as illustrated by the curves for the editions 2012 and 2013, mild temperatures and dry conditions seem to be the most ideal conditions to engage in the 100 km Doodentocht. Still, as the overall success rate was approximately the same across the editions 2009–2014 (Table 3.2), we concluded that the weather conditions mainly influence the time when and the distance at which participants retire rather than the success rate. Furthermore, as it is impossible to quantify the effect of the weather conditions, we did not distinguish any further between the different editions in the remainder of this chapter.

The dispersion of all marchers included in the data set along the trail and over time is illustrated in Fig. 3.6. In red, the opening and closing times of the checkpoints are shown. In the beginning of the march, marchers were still clustered, but as they moved at a speed between 4.2 km h^{-1} and 10 km h^{-1} , this clustering became

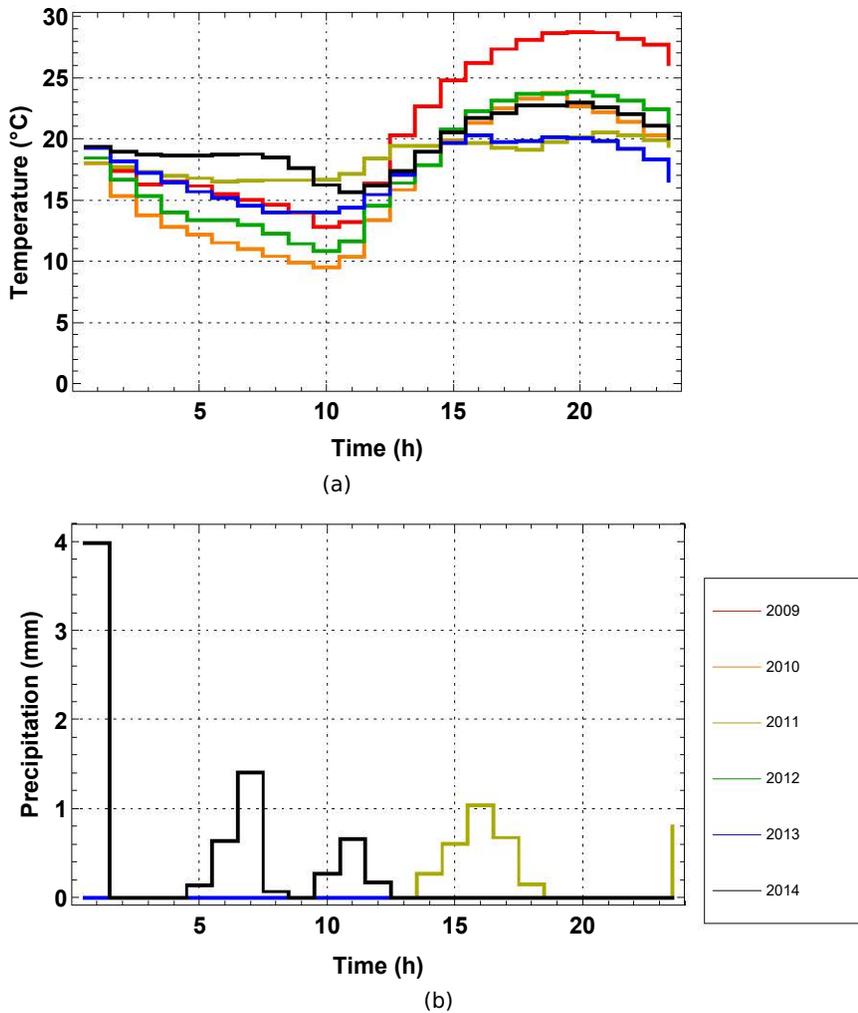


Figure 3.4: Temperature (a) and amount of precipitation (b) per hour during the editions 2009–2014 based on observations at three weather stations near the trail.

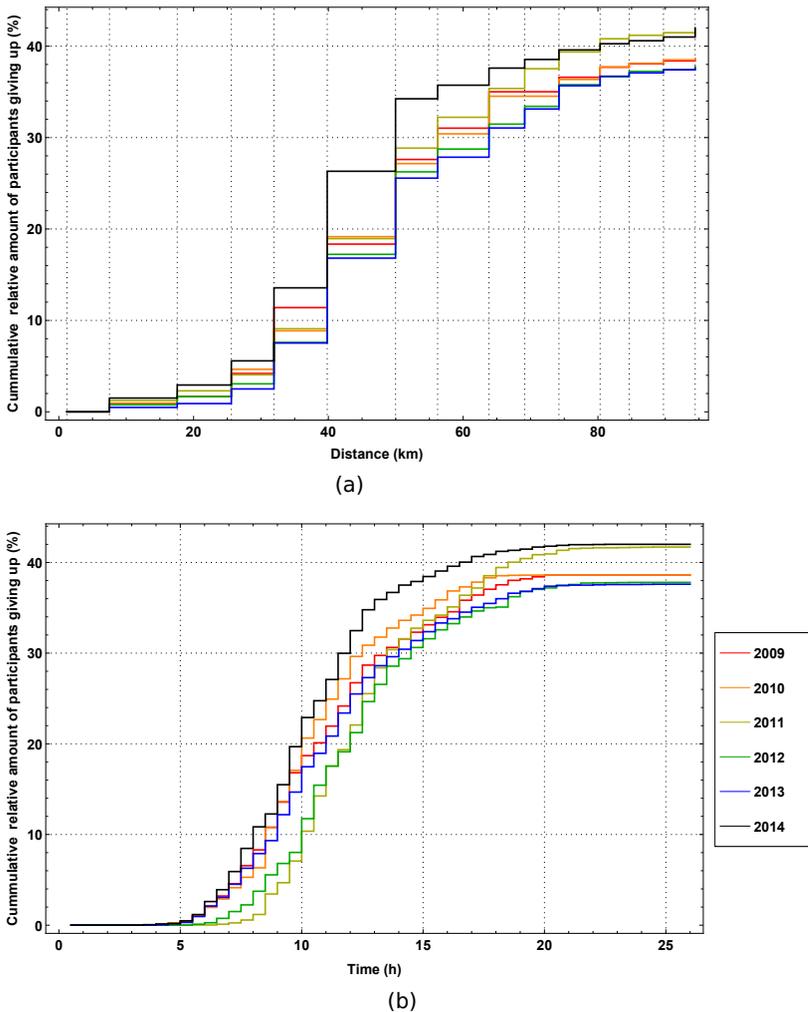


Figure 3.5: The cumulative relative frequency of marchers retiring versus the distance from the start (a) and the time elapsed since the start (b) during the editions of the 100 km Dodentocht in 2009–2014.

much less pronounced as time elapses. This indicates why checkpoints at larger distances from the start required a lot of operational effort. Since most people progressed at a relatively low speed, the busiest periods at the checkpoints were typically near their closing times.

As a last step of the data analysis, the average marching speed along the sections is studied. As can be seen in Fig. 3.7 (the vertical gridlines represent the locations of the checkpoints), it appears to decrease steadily along the trail, but there is a pronounced drop noticeable at checkpoint 7 (50.0 km). Yet, given the fact that a warm meal was offered at this checkpoint, this sudden drop is undoubtedly a consequence of relatively long resting periods at this checkpoint, rather than a

3

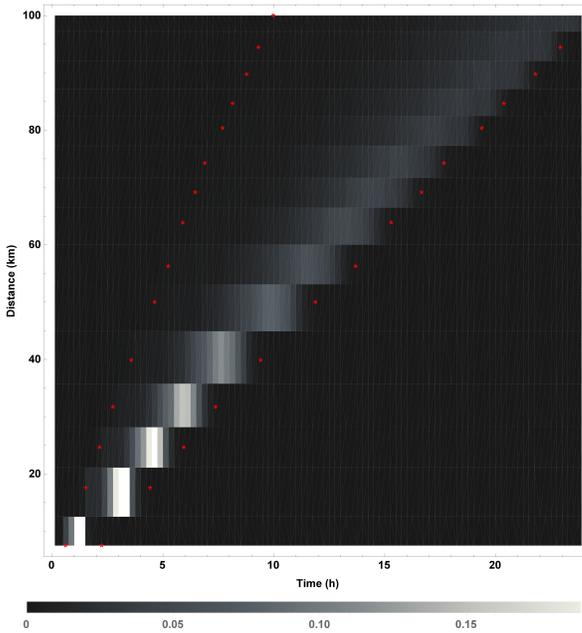


Figure 3.6: The relative frequency of marchers along the trail versus time elapsed since the start. In red, the opening and closing times of the checkpoints are shown.

lower marching speed. Unfortunately, the duration of these resting periods cannot be taken into account since the marchers were registered only once at every checkpoint (cfr. Section 3.2).

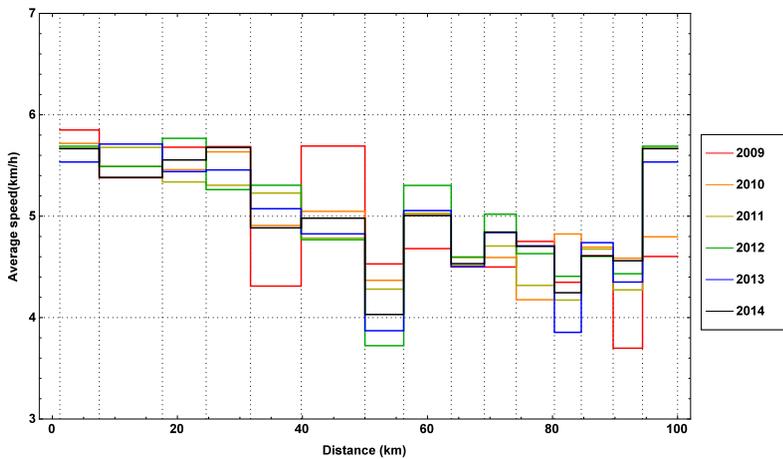


Figure 3.7: The average marching speed along the trail during the editions of the 100 km Dodentocht in 2009-2014.

3.3.1.2 Correlation analysis

Since it is likely that the speed of a marcher along a given section is related to his/her speed at the preceding and subsequent sections, the autocorrelation between observations separated by k sections, with $k = 1, \dots, N_i - 1$, was calculated (Chatfield, 2004):

$$r_k^i = \frac{\sum_{s=1}^{N_i-k} (v_s^i - \bar{v}_{(1)}^i)(v_{s+k}^i - \bar{v}_{(2)}^i)}{\left[\sum_{s=1}^{N_i-k} (v_s^i - \bar{v}_{(1)}^i)^2 \right]^{1/2} \left[\sum_{s=k+1}^{N_i} (v_s^i - \bar{v}_{(2)}^i)^2 \right]^{1/2}}, \quad (3.1)$$

where $\bar{v}_{(1)}^i$ and $\bar{v}_{(2)}^i$ are the mean of the first $N_i - k$ and last $N_i - k$ registered marching speeds of marcher i , respectively. Positive autocorrelation occurs when the deviation of the speed from the mean tends to be followed by a deviation of the same sign.

The average autocorrelation over all marchers and all years is shown in Fig. 3.8. From this figure, it follows that a positive autocorrelation exists between marching speeds up to five sections apart, while a negative one can be seen between marching speeds along sections further apart. The autocorrelation is especially pronounced between marching speeds along two consecutive sections ($k = 1$) and along sections that are two sections apart ($k = 2$), while it is negligibly positive between marching speeds along sections that are four or more sections apart.

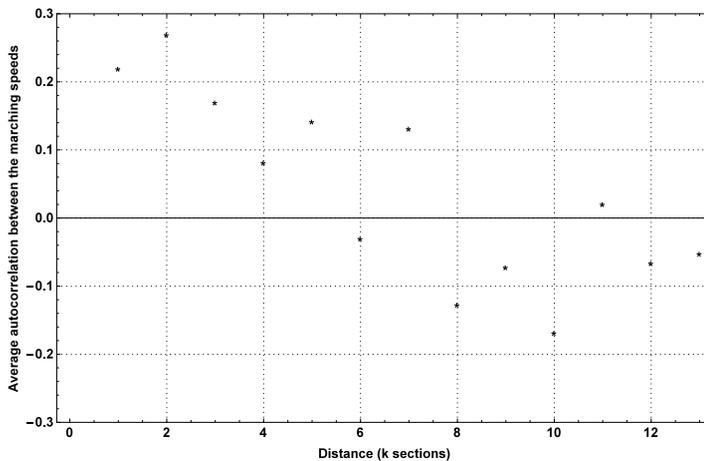


Figure 3.8: Average autocorrelation between the marching speeds at k sections apart.

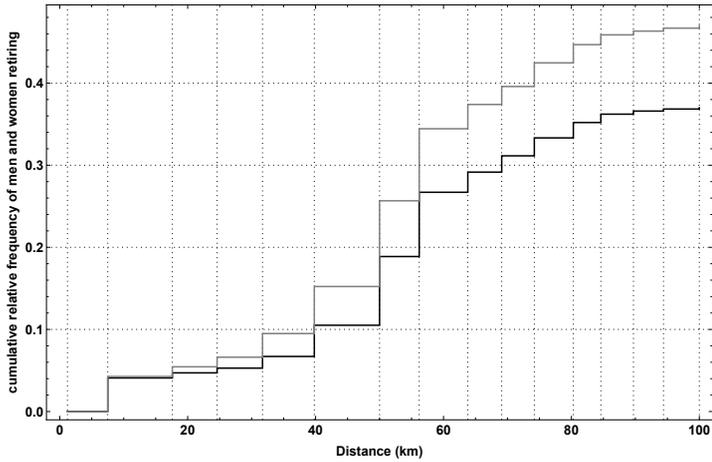
As a next step in the correlation analysis, the influence of age and gender on the marching speed and covered distance was examined. Taking into account that age and gender of the marchers are missing for 2009, only the data of editions

2010–2014 were used for this purpose. First, the effect of gender is considered. In 2010, 21.64% of the marchers were women, while this was 22.54%, 23.28%, 22.45% and 24.16% in 2011, 2012, 2013 and 2014, respectively. Over these editions, the success rate of women was on average 53%, while it was 63% for men. The cumulative relative frequency of men and women retiring as function of the distance from (a) and time since the start (b) can be found in Fig. 3.9, while Fig. 3.10 illustrates the differences between the distributions of the marching speeds of men and women along the trail. Generally speaking, female marchers tended to march somewhat slower than men, while the former also tended to retire at shorter distances from the start and in greater numbers than their male counterparts. As both retiring and marching speed are influenced by gender, this variable can be considered as an important factor. This was also confirmed by a t-test applied to both the observed covered distances and the average marching speeds for men and women.

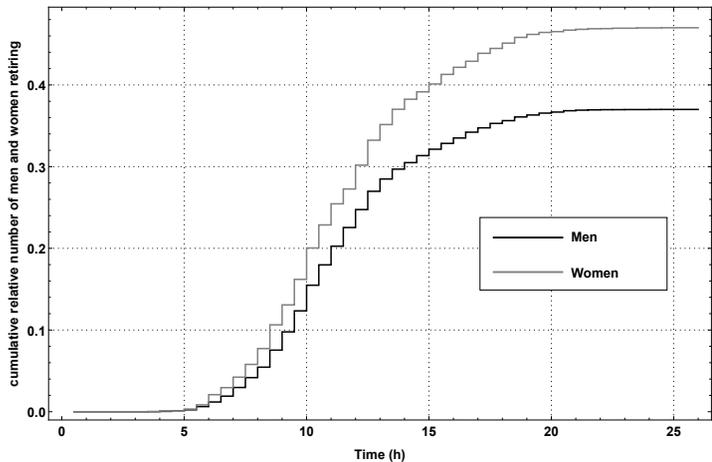
The influence of age on the average marched distance and average marching speed is summarized in Fig. 3.11, which depicts the average total marched distance versus the average marching speed per age for men (black) and women (gray). For men, two clusters can be discerned from this plot, a first one containing marchers of at most 30 years old, while the second one contains marchers who were 46 or older. Located between these two clusters are the marchers aged between 32 and 46 who show a steady increase in average marched distance and average marching speed as function of their age. It is remarkable that older marchers tended to walk faster and keep on going longer, though we argue that this might be explained by the fact that younger marchers typically decide to participate in a more impulsive way. Consequently, they appear less prepared at the start of the 100 km Dodentocht than many of the older marchers. For women, age seems to particularly affect the covered distance, while the effect on average speed is less pronounced. Still, a cluster containing relatively younger women (younger than 28), a second cluster containing older women (older than 40), and a transition region are also present in the case of female marchers. Since there is a significant age range where the average marched distance increases as function of the age, a clear-cut clustering of marchers with highly similar age-speed characteristics is not evident.

Finally, the relation between the covered distance and the starting speed is displayed in Fig. 3.12. As can be seen in this figure, the covered distance increases steadily with a higher starting speed, provided that the starting speed is higher than 5 km h^{-1} . Otherwise, this trend is not discernible for men. Nevertheless, the starting speed may serve as an indication of the success rate. In general, it can be stated that marchers starting at a high marching speed cover greater distances.

After the starting signal of the 100 km Dodentocht had been given, approximately 11,000 marchers started together. As a consequence of the size of the crowd, it took up to half an hour before the last marcher passes by the starting line. In



(a)



(b)

Figure 3.9: The cumulative relative frequency of men and women retiring versus the distance from (a) and time since the start (b) of the 100 km Dodentocht in 2010–2014.

Fig. 3.13, one can find the average time elapsed since the starting signal before the marchers pass by the starting line versus the starting speed. As one can see, people intended to march fast, tried to start up front.

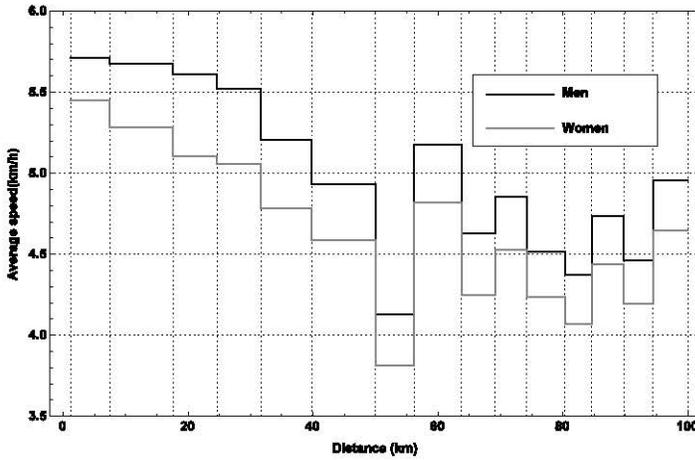


Figure 3.10: The average marching speed of men and women at every section of the 100 km Dodentocht in 2010–2014.

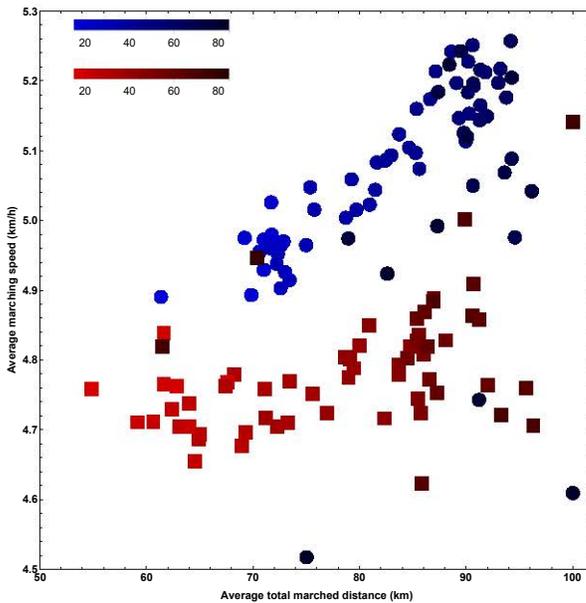


Figure 3.11: Effect of age and gender on the performance of a 100 km Dodentocht marcher. Each marker represents the average total marched distance and the average marching speed. Men are presented by blue disks and women by red squares. The darkness of the markers is determined by the age.

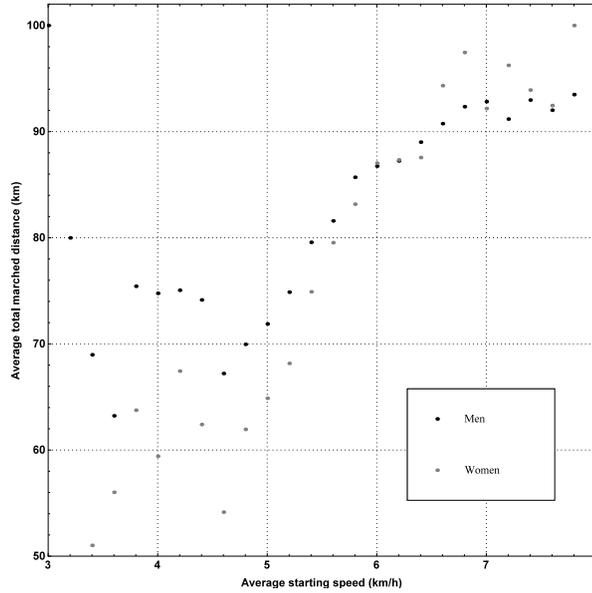


Figure 3.12: Effect of starting speed on the covered distance of a 100 km Dodentocht marcher. The average total marched distance versus the starting speed of a 100 km Dodentocht marcher for men and women.

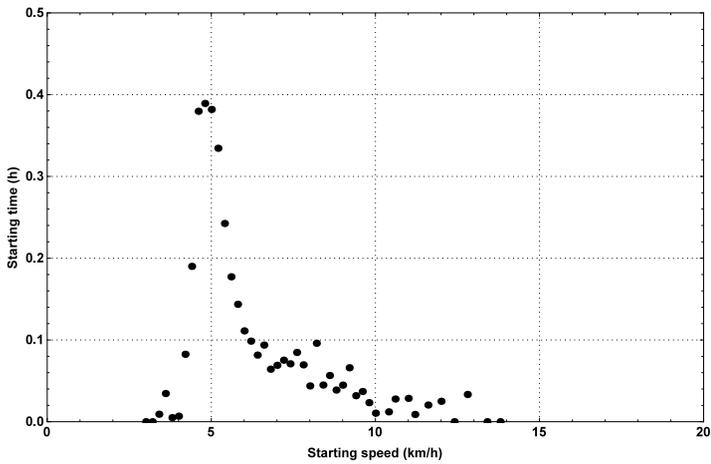


Figure 3.13: The average time elapsed since the starting signal before the marchers pass by the starting line as a function of the starting speed.

3.3.2 GPS tracking data

Before performing any data analysis, the latitude and longitude of each registered location was transformed into a distance from the start along the trail (the covered distance) by making use of a reference trail. The reference trail for 2013 contained 1791 locations and the one for 2014 contained 1379 locations along the trail. Since these locations are situated close to each other, the cumulative distances along the consecutive locations were calculated in order to transform the latitudes and longitudes to corresponding covered distances.

As a preprocessing step, the registered locations far away from the reference trail, recorded before the start or after retiring and participants with less than 20 registrations were deleted. As such, only 375 participants for 2013 and 259 for 2014 were used in the data analysis. On average, an observation was made every 17 min, which on average corresponds to 1.35 km. The accuracy of the measurements was on average 24 m with a minimum of 5 m and maximum of 159 m.

After the dataset clean up, the latitude and longitude of each registered location were converted to a covered distance by adopting the distance corresponding to the location of the reference trail that is most near.

From the GPS tracking data set, instantaneous marching speeds could be calculated. The latter was not possible from the RFID tracking data set, where average marching speeds between the consecutive checkpoints were obtained. However, these average marching speeds would be more informative if the breaks could be identified. Therefore, the GPS tracking data were analyzed to determine some valuable information on the duration of the resting periods along each section.

In Fig. 3.14(a), the progress of one of the participants along the trail, together with the speed versus the covered distance (b) and time since the start (c) is shown. Around 20 km, 40 km, 50 km, 60 km and 85 km, the speed decreases, and the participant probably rested. As expected, these distances correspond to the locations of some of the checkpoints. The drops in marching speed were used to identify resting periods automatically. More precisely, a resting period of a participant was defined as the period where the marching speed is lower than the speed defined by the 10% decile of all his/her registered speeds. As such, the participant considered above took a break at about distances 20 km, 40 km, 50 km, 60 km and 85 km.

To determine the duration of a resting period, the difference in time between the observations at one location was calculated. Observations are assumed to originate from the same locations if the distance between them is less than 1 km. If there is only one observation, it is not possible to give a decision on the duration and it is set to zero. In this way, we underestimate the total duration of the resting periods. The duration of the resting periods of our participant under considera-

tion versus the distance from the start, can be seen in Fig. 3.14(d). The vertical gridlines represent the locations of the checkpoints.

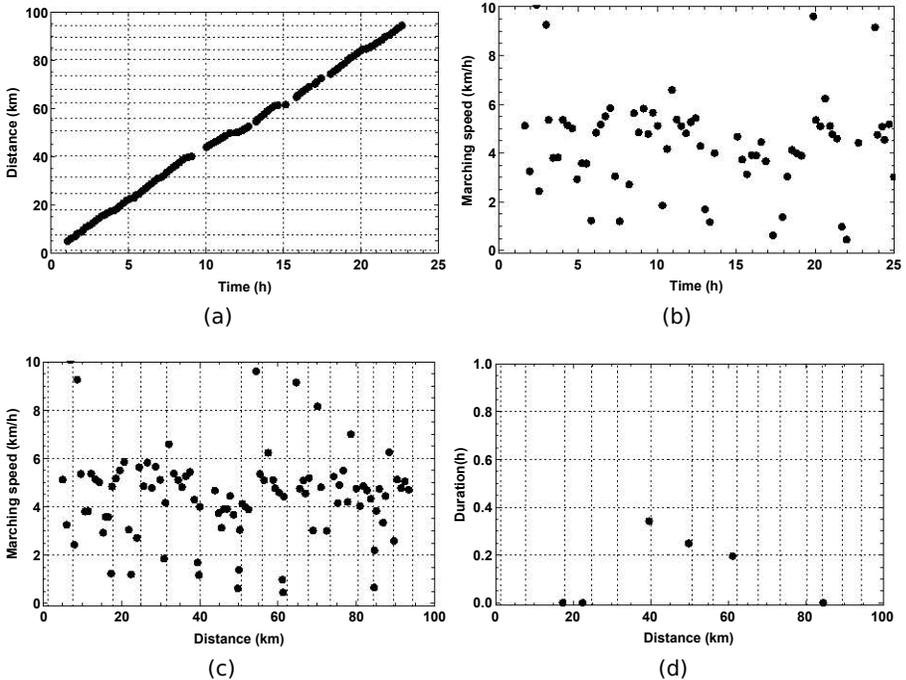


Figure 3.14: The progress of one participant along the trail (a), together with its speed versus the time since the start (b) and covered distance (c). The duration of the resting periods of this participant versus covered distance, can be seen in (d).

In order to get a general overview of the locations where participants rest and the duration of the resting periods, all the GPS tracking data were considered. In Fig. 3.15(a), one can see the relative number of marchers taking a break along the trail (discretization step of 1 km). This number is relatively higher near checkpoints. At most 25% of the marchers took a break at the same location. In between the checkpoints, only few participants took a break. For calculating the average duration of the resting periods, only locations where more than 5% of the participants were resting, are taken into account. The results can be seen in Fig. 3.15(b). From this figure, it can be concluded that checkpoints located at distances of about 40 km and 50 km hosted the most breaks and thus for the longest period in time. This figure also confirms the presumption made in the previous section that the drop in marching speed at checkpoint 7 (50.0 km) is the result of resting.

In the next section, the spatially explicit marching model is built. Since there are more radio tracking data available compared to the GPS tracking data, the former are used for model construction. Since not all participants took a rest at the same time and since the resting periods vary, we could not draw general conclusions

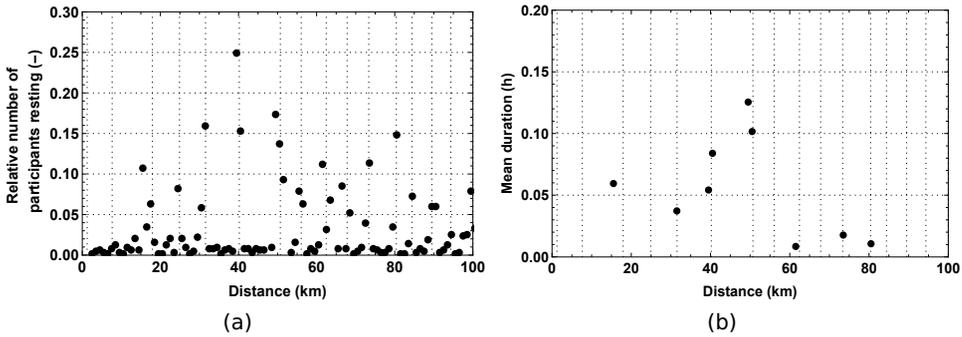


Figure 3.15: The relative number of marchers taking a rest at every kilometer along the trail (a) and the average duration of the resting period at distances where more than 5% of the participants rested (b).

about resting from the GPS tracking data that might be taken into account in the marching model. As such, the results of this section are not further used in the remainder of this chapter.

3.4 A spatially explicit marching model

3.4.1 Overview

The spatially explicit marching model presented here aims to mimic the speed of an individual *in silico* marcher along every section of the 100 km Dodentocht trail. It was developed in such a way that it is able to reflect the spatio-temporal variability in the marching dynamics among the marchers. Ultimately, it should allow the organizers to assess the potential effects of modifications to the starting procedure and the trail architecture. Given the limited size of the GPS tracking data set, only the RFID tracking data set was used for building the spatially explicit marching model.

Here, we aimed at simulating an average edition of the 100 km Dodentocht, thereby not accounting explicitly for the governing weather conditions. This approach is justified as the spatially explicit marching model should allow for assessing the possible effects of modifications to the starting procedure and the trail architecture, which are assumed to be independent of the weather conditions. Essentially, by using the data from all editions, and hence indirectly accounting for variability, and finally resorting to Monte Carlo simulations, it is possible to derive confidence intervals that – amongst other things – reflect variations due to external factors.

To grasp the variability in marching speed along a given section s , the marching

speed along this section v_s has to be conceived as a random variable V_s , described by a probability density function (PDF) Ω_{V_s} . Since age \mathcal{A} and gender \mathcal{G} (cfr. Figs. 3.10 and 3.11), as well as the marching speed along previous sections \mathcal{P} (Fig. 3.8) affect V_s , these variables should be taken into account when simulating the dynamics of an individual marcher. More precisely, this means that the speed along the sections under consideration should be sampled from a PDF $\Omega_{V_s|\mathcal{G},\mathcal{A},\mathcal{P}}(v_s | g, a, p)$ that is conditioned on \mathcal{G} , \mathcal{A} and \mathcal{P} . Here, \mathcal{G} is a binary variable with either man or woman as value, $\mathcal{A} \in \mathbb{N}$ is a discrete variable, and $\mathcal{P} \in \mathbb{R}^+$ is a continuous variable which is specified in the Section 3.4.5. The values under consideration for V_s , \mathcal{G} , \mathcal{A} and \mathcal{P} are denoted by v_s , g , a and p , respectively.

In addition to the marching speed along every section, the total covered distance by a marcher is estimated. This is done by making use of a hazard function (Be-
wick et al., 2004). The hazard function $h(n)$ describes the conditional probability of retiring at checkpoint n , giving that the participant already marched to checkpoint n . In order to construct a realistic hazard function, the checkpoint numbers until marchers retire are considered together with the age and gender of marchers, as Figs. 3.9 and 3.11 clearly indicate that these individual characteristics are important for understanding the dynamics of retiring. Furthermore, Figs. 3.11 and 3.12 indicate that there exists a relationship between the starting speed and retiring. As such, the starting speed is also taken into account.

In the remainder of this section, details on the conditional distribution of V_s , the exact formulation of \mathcal{P} , and the hazard functions are given. This section is concluded by a comprehensive discussion on the selection of the most appropriate model structure.

3.4.2 Conditional distribution of the marching speed

The conditional distribution $\Omega_{V_s|\mathcal{G},\mathcal{A},\mathcal{P}}(v_s | g, a, p)$ was constructed experimentally by counting the entries in the radio tracking data set \mathcal{D} fulfilling the conditions on gender, age and marching speed at previous sections. More specifically, the empirical distribution function (EDF) $\hat{F}_{V_s|\mathcal{G},\mathcal{A},\mathcal{P}}(v_s | g, a, p) : \mathbb{R} \rightarrow [0, 1]$ was determined:

$$\hat{F}_{V_s|\mathcal{G},\mathcal{A},\mathcal{P}}(v_s | g, a, p) = \frac{\#\{i | v_s^i \in \mathcal{D} \wedge v_s^i \leq v_s \wedge \mathcal{G} = g \wedge \mathcal{A} = a \wedge \mathcal{P} = p\}}{\#\text{ of marchers in } \mathcal{D} \text{ with } \mathcal{G} = g \wedge \mathcal{A} = a \wedge \mathcal{P} = p}, \quad (3.2)$$

where v_s^i is observation i of v_s in \mathcal{D} . Essentially, the EDF $\hat{F}_{V_s|\mathcal{G},\mathcal{A},\mathcal{P}}(v_s | g, a, p)$ may be considered as an approximation of the underlying cdf. In contrast to age and gender, the marching speed is a continuous variable, so that a discretization is

needed to construct the EDF. Here, a bin width of 0.2 km h^{-1} was chosen and the speeds were rounded to 0.2 km h^{-1} .

3.4.3 Hazard function

The Cox model (Cox and Oakes, 1984) was used to determine the hazard function for the 100 km Dodentocht. Cox models were determined for both men and women separately, with explanatory variables being age and starting speed. The instantaneous risk of retiring can be written as (Cox and Oakes, 1984):

$$\log(h(n)) = \log(h_0(n)) + b_1 a + b_2 v_1, \quad (3.3)$$

where $h(n)$ is the hazard at checkpoint n , a and v_1 are the values of the explanatory variables age \mathcal{A} and starting speed V_1 , $h_0(n)$ is the baseline hazard, \log is the natural logarithm and b_1 and b_2 are model parameters. The coefficients b_1 and b_2 were estimated together with the baseline hazard $h_0(n)$ on the basis of the RFID data from the editions 2010–2014. The models were fit using Breslow's partial likelihood method (Breslow, 1975) in Mathematica (Version 9.0.1, Wolfram Research Inc., Champaign, US). The EDF of the marchers' retiring checkpoints based on the cumulative hazard function $h(n)$ was then used to sample the individual total covered distances using the inverse transform method (Banks, 1998).

3.4.4 Model simulation

Simulations of average 100 km Dodentocht editions were performed by simulating 10,000 marchers, of which 22.81% were female in accordance with the values reported in the Section 3.2. Although there were on average 10,925 marchers at the start of the 100 km Dodentocht (cfr. Table 3.2), we opted to simulate only 10,000 marchers for the ease of interpretability of the simulation results. When the simulation results are compared to the observations, the latter were scaled in such a way that the reported differences are valid for an edition with 10,000 marchers.

A simulation generally consists of the following steps. At the beginning of a simulation, every marcher is assigned an age and a starting speed. These values are drawn from the joint probability distribution of \mathcal{A} and V_1 , which is derived from the RFID data for men and women separately. Subsequently, given an age, gender and starting speed of an *in silico* marcher, the checkpoint at which he/she retires is derived from the EDF of the marchers' last registered checkpoints (cfr. Section 3.4.3). Then, for every *in silico* marcher individually, the speeds at the consecutive sections are determined up to the one at retiring. These marching

speeds are obtained by sampling the EDF of V_s given by Eq. (3.2). The algorithm describing the sampling of the marching speeds of the participants of an average 100 km Dodentocht can be found in Algorithm 1.

As a next step, the times at which the participants pass by each checkpoint are calculated. For each participant, the passing time at the first checkpoint (the starting time) is sampled from the EDF of the starting times conditioned on the sampled starting speed V_1 . The passing times at the other checkpoints up to the sampled checkpoint of retiring are calculated on the basis of the starting time, the obtained marching speeds and the average distances between the consecutive checkpoints.

3.4.5 Model selection

Based on the autocorrelation between marching speeds at consecutive sections, we can conclude from Fig. 3.8 that the speed along up to the three preceding sections should be taken into account when determining the marching speed along a given section s . Still, it should be verified whether a simpler model that is based only on, for instance, the previous marching speed, could be as effective as its counterpart based on the last three registered marching speeds. For that reason, the performance of the spatially explicit marching model was checked for three different forms of the conditional distribution of the marching speed. In the remainder, M_1 is used to refer to the spatially explicit marching model based on an EDF involving only V_{s-1} in addition to age and gender, while M_2 and M_3 are used to refer to the models based upon an EDF involving $(V_{s-1} + V_{s-2})/2$ in \mathcal{P} and $(V_{s-1} + V_{s-2} + V_{s-3})/3$ in addition to the age and gender, respectively.

In order to compare the performance of the spatially explicit marching model across the EDFs, the models were run 100 times according to the procedure outlined in Algorithm 1. The relative frequencies of the simulated marching speeds along the trail for the different models, together with the ones of the observed marching speeds can be found in Fig. 3.16, while the number of marchers passing by at every checkpoint per 0.25 h are shown in Fig. 3.17. In Fig. 3.16, respectively in Fig. 3.17, a bin width of 0.2 km h^{-1} , respectively 0.25 h is used. Furthermore, the simulated cumulative relative frequency of people retiring at every checkpoint is shown in Fig. 3.18. By comparing the simulation results with the corresponding distributions derived from the RFID data, one can visually evaluate the goodness of fit.

Algorithm 1: Pseudo code of the spatially explicit marching model of the 100 km Dodentocht

Pseudo code to build the model

Determine the joint EDF of \mathcal{A} and V_1 for men and women, *i.e.* $\hat{F}_{\mathcal{A}, V_1 | \mathcal{G}}(a, v_1 | g)$ for all values of $a \in [10, 90]$ and $v_1 \in [0, 20]$, using a discretization step of respectively 1 and 0.2;

for $s = 2$ **to** $s = 15$ **do**

Determine the EDF of V_s conditioned on \mathcal{A} and \mathcal{P} for men and women, *i.e.*

$\hat{F}_{V_s | \mathcal{G}, \mathcal{A}, \mathcal{P}}(v_s | g, a, p)$ for all values of $a \in [10, 90]$ and all values of v_s and $p \in [0, 20]$, using a discretization step of respectively 1, 0.2 and 0.2 resp.;

Determine the EDF of V_s conditioned on \mathcal{P} for men and women, *i.e.*

$\hat{F}_{V_s | \mathcal{G}, \mathcal{P}}(v_s | g, p)$ for all values of v_s and $p \in [0, 20]$, using a discretization step of 0.2;

Determine the EDF of V_s for men and women, *i.e.* $\hat{F}_{V_s | \mathcal{G}}(v_s | g)$ for all values of $v_s \in [0, 20]$, using a discretization step of 0.2;

end for

Determine the EDF of the retiring checkpoints based on the cumulative hazard function including the explanatory variables \mathcal{A} and V_1 for men and women separately;

Pseudo code to simulate marchers

for all marchers (For an average 100 km Dodentocht: 10,000) **do**

Determine the gender g of the marcher; (For an average 100 km Dodentocht: 22.81% women)

Sample a and v_1 from $\hat{F}_{\mathcal{A}, V_1 | \mathcal{G}}(a, v_1 | g)$;

Sample the retiring checkpoint n from the EDF of the retiring checkpoints given that $\mathcal{A} = a$, $\mathcal{G} = g$ and $V_1 = v_1$;

for $s = 2$ **to** $s = n - 1$ **do**

if (# of marchers in \mathcal{D} with $\mathcal{G} = g \wedge \mathcal{A} = a \wedge \mathcal{P} = p) \neq 0$ **then**

Sample v_s from $\hat{F}_{V_s | \mathcal{G}, \mathcal{A}, \mathcal{P}}(v_s | g, a, p)$;

else

if (# of marchers in \mathcal{D} with $\mathcal{G} = g \wedge \mathcal{P} = p) \neq 0$ **then**

Sample v_s from $\hat{F}_{V_s | \mathcal{G}, \mathcal{P}}(v_s | g, p)$;

else

Sample v_s from $\hat{F}_{V_s | \mathcal{G}}(v_s | g)$;

end if

end if

end for

end for

return g, a and v_s with $s = 1, \dots, n - 1$

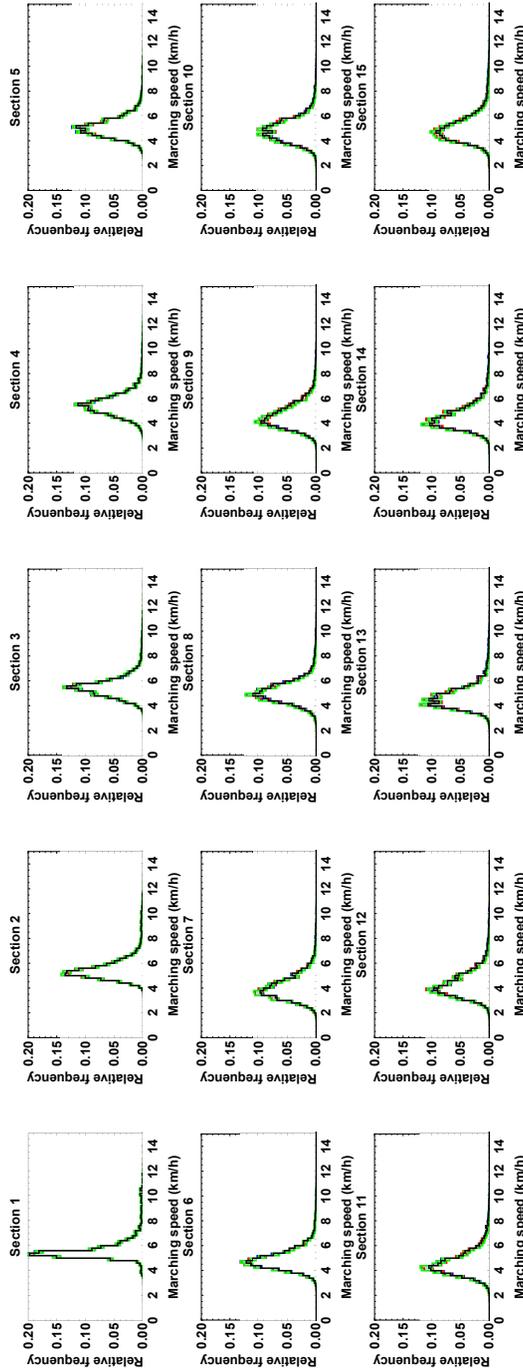


Figure 3.16: The 95% confidence envelope of the simulated marching speeds for M_1 (blue), M_2 (red) and M_3 (green), together with the relative frequencies of the observed marching speeds (black).

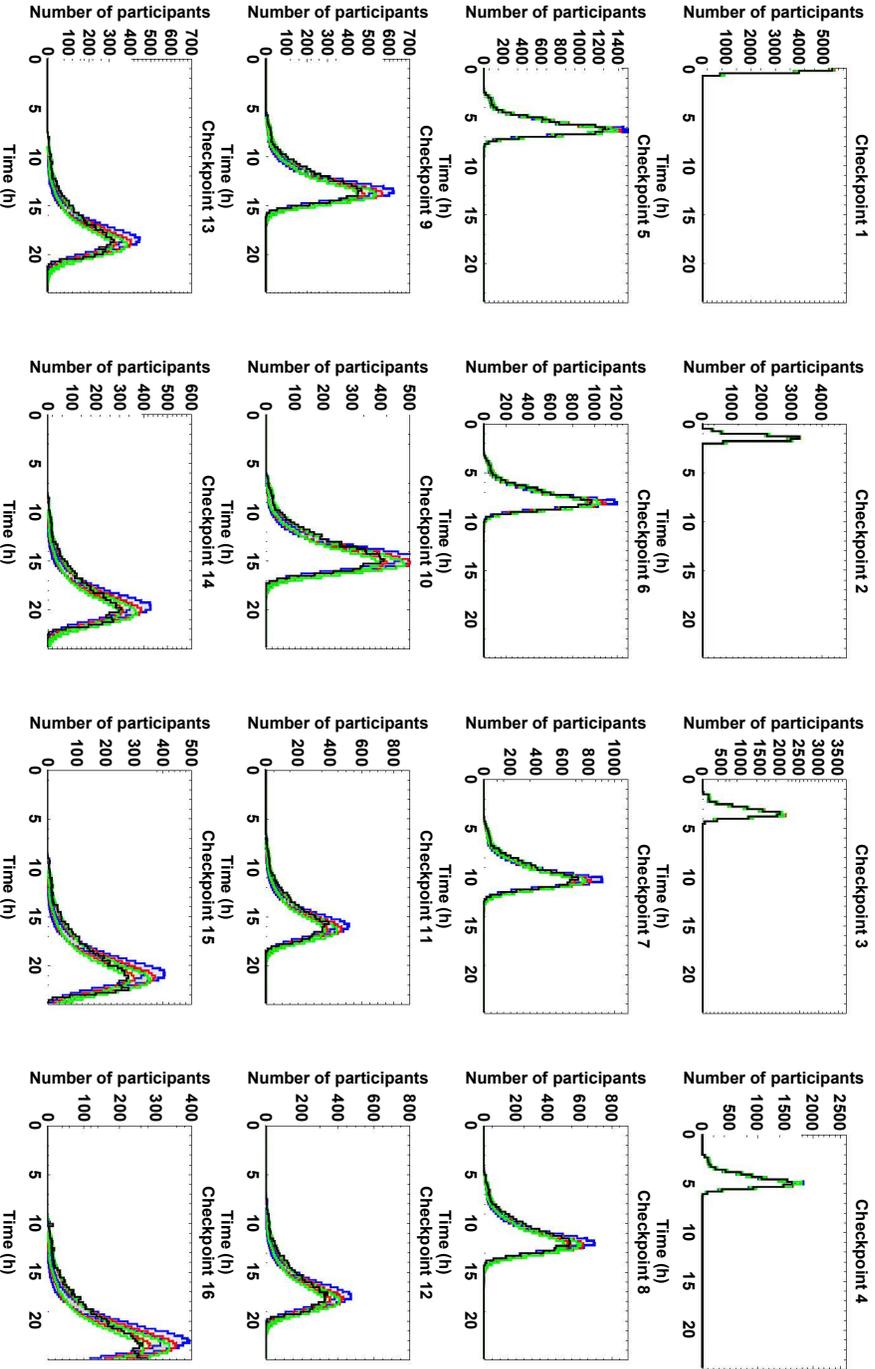


Figure 3.17: The 95% confidence envelope of the number of marchers passing by at every checkpoint per 0.25 h for M_1 (blue), M_2 (red) and M_3 (green), together with the observed number of marchers passing by (black).

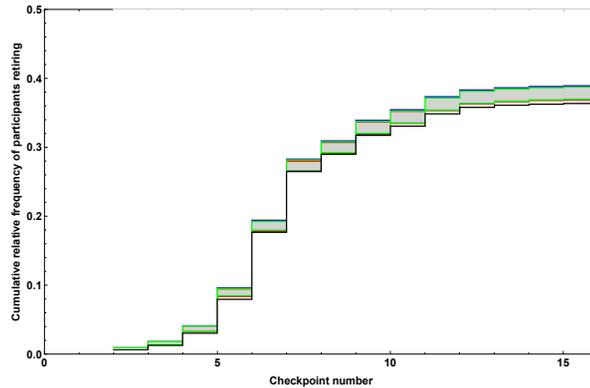


Figure 3.18: The 95% confidence interval of the simulated cumulative relative frequency of marchers retiring at every checkpoint for M_1 (blue), M_2 (red) and M_3 (green), together with the observed cumulative relative frequency (black).

As can be inferred from Fig. 3.16, the marching speeds are approximated well, irrespective of the governing EDF. Nevertheless, there are substantial differences between the models for what concerns the estimation of the passing times (cfr. Fig. 3.17). The shapes of these relative frequencies are similar to the one for the observations but, starting from checkpoint 7, the peaks are overestimated while the tails towards the left (earlier passing times) are underestimated. Although all three models suffer from this, it is clear that M_1 performs the worst, while M_2 and M_3 yield reasonable passing times.

For what concerns the cumulative relative frequency of marchers retiring at every checkpoint, Fig. 3.18 shows that it is overestimated by the Cox model, irrespective of the involved EDF. The retiring dynamics, with a high chance of retiring at checkpoints 6 and 7, is however simulated reasonably well.

In order to select the most suitable model, the performance of the three models was evaluated objectively, and preference was given to the simplest model that still mimics the dynamics of an average 100 km Dodentocht reasonably well. More precisely, the Match distance and the Kolmogorov-Smirnov distance were considered (Rubner et al., 2000; Werman et al., 1985; Frank and Massey, 1951), two distance functions summarizing the discrepancies between observed values and the values simulated under the model in question (Rubner et al., 2000). The Match distance quantifies the total difference between the corresponding EDFs as follows (Rubner et al., 2000):

$$d_M(h, k) = \sum_i \left| \hat{h}_i - \hat{k}_i \right|, \quad (3.4)$$

where h and k are the two EDFs under consideration. The Kolmogorov-Smirnov distance only considers the maximum difference between the corresponding EDFs

(Rubner et al., 2000):

$$d_{KS}(h, k) = \max_i \left| \hat{h}_i - \hat{k}_i \right|. \quad (3.5)$$

The average Match and Kolmogorov-Smirnov distances between the observed and simulated marching speeds and passing times are shown in Figs. 3.19 and 3.20 respectively. The average distances for the cumulative relative frequency of marchers retiring at every checkpoint can be found in Table 3.3. From Fig. 3.19, it follows that the distances between the observed and simulated EDFs of the marching speeds are low, that they increase from checkpoint 6 on, and that there are differences between the models. More precisely, M_1 is the best at simulating the marching speeds, while M_3 performs the worst. Concerning the passing times, the distances between the observed and simulated numbers of marchers passing by at every checkpoint per 0.25 h are low if the speeds at the consecutive sections are composed in such a way that the simulated speed at a section is chosen well considering the simulated speeds at all other sections (cfr. Fig. 3.20). For example, a marcher marching fast along the first few sections, is likely to maintain high speeds along the following sections, even if the marcher takes a break along one of these sections. If the model would consistently assign low speeds in this case, these speeds would not match with the previous high speeds, causing a bad composition and high Match and Kolmogorov-Smirnov distances. The distances increase from checkpoint 6 on, with the largest increase at checkpoints 7 and 8, where most people took a break due to the meal that was offered causing a decline of recorded marching speeds. Because of that, fast marchers taking a pause and slower marchers not taking a break had the same marching speeds along the considered sections, which hinders an unambiguous determination of their subsequent marching speeds by sampling the EDF of V_s given by Eq. (3.2). For what concerns the passing times, M_2 and M_3 are performing the best. Only at checkpoint 16, at the very end of the trail, M_3 outperforms the other models, while M_2 is the best at checkpoints 7–15. For what concerns the observed and simulated numbers of marchers retiring at every checkpoint, the three models lead to similar distances, with only slightly lower distances for M_2 .

Table 3.3: Match and Kolmogorov-Smirnov distances between the observed and simulated EDFs of the number of marchers retiring at every checkpoint.

| Model | Match distance (-) | Kolmogorov-Smirnov distance (-) |
|-------|--------------------|---------------------------------|
| M_1 | 0.16 | 0.017 |
| M_2 | 0.15 | 0.015 |
| M_3 | 0.15 | 0.016 |

Considering these results, M_2 was finally selected as optimal model describing the dynamics of the 100 km Dodentocht. This model was preferred to M_1 since it leads to a better agreement between the observed and simulated passing times, while it was preferred to M_3 because its construction is simpler and it leads at the same

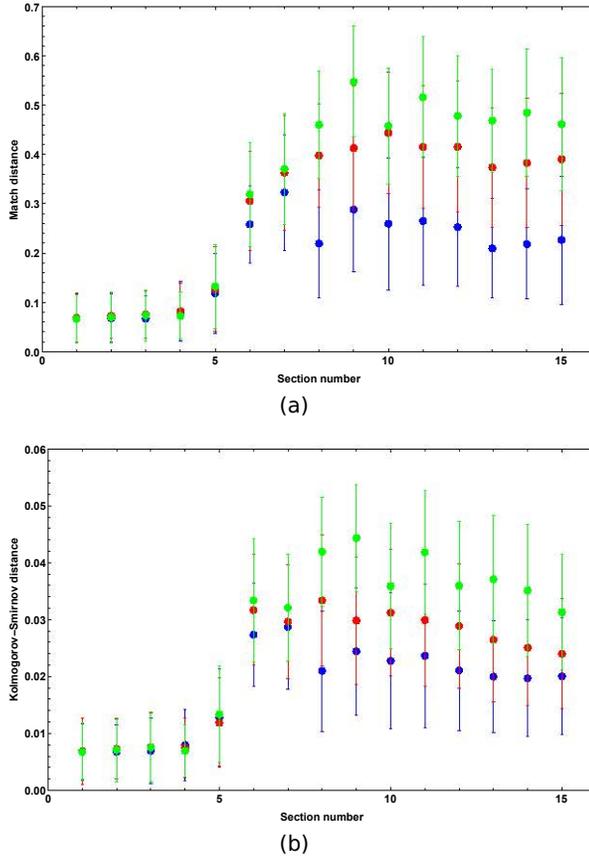


Figure 3.19: Match (a) and Kolmogorov-Smirnov (b) distance between the EDFs of the observed and simulated marching speeds for M_1 (blue), M_2 (red) and M_3 (green). The average distances are located in the middle, while the error bars denote twice the standard deviation.

time to an equally good – and sometimes even better – agreement between the observed and simulated quantities.

In view of organizational affairs, the number of people passing by every checkpoint at every 0.25 h is considered in more detail. The mean and maximum absolute differences between the average simulated and the observed number of people passing by can be found in Table 3.4. Since marchers are clustered at the start, a difference of for example 100 marchers has less impact at the start than towards the end. Therefore, the differences should be evaluated together with the average number of people passing by at the considered checkpoint, which can be estimated from Fig. 3.17. Moreover, since the shape of the cdfs of the passing times are estimated well, it can be concluded that busy and calm periods are simulated accurately. Therefore, and due to the discretization of time, the differences in number of marchers passing by has to be accounted for in nearby time bins.

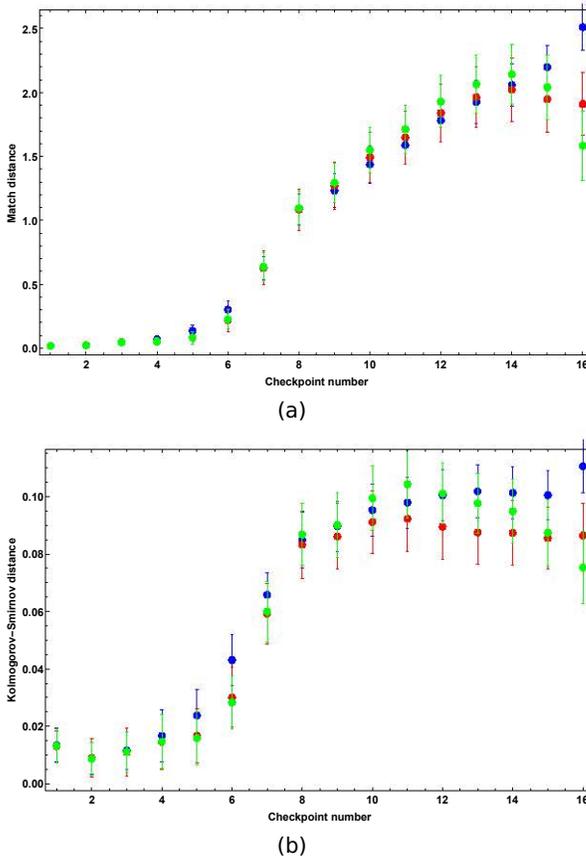


Figure 3.20: Match (a) and Kolmogorov-Smirnov (b) distance between the EDFs of the observed and simulated numbers of marchers passing by at every checkpoint per 0.25 h for M_1 (blue), M_2 (red) and M_3 (green). The average distances are located in the middle, while the error bars denote twice the standard deviation.

Table 3.4: The average and maximum absolute difference between the average of the 100 simulated and the observed number of people passing by at every checkpoint, per 0.25 h for a 100 km Dodentocht with 10,000 marchers. The simulations were performed with M_2 .

| Checkpoint ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|-----------------------------|----|----|-----|-----|-----|----|----|-----|----|-----|-----|----|----|----|----|----|
| Mean absolute difference | 0 | 0 | 3 | 4 | 5 | 6 | 9 | 12 | 12 | 12 | 12 | 12 | 12 | 13 | 13 | 12 |
| Maximum absolute difference | 12 | 10 | 130 | 108 | 119 | 84 | 97 | 115 | 98 | 109 | 106 | 78 | 90 | 85 | 73 | 90 |

3.5 Model validation

3.5.1 Cross-validation

Models are valuable only if they have been validated with respect to their purpose. The purpose of the spatially explicit marching model is to gain insight into the dy-

namics of the 100 km Dodentocht. Knowing the number of marchers passing by the checkpoints at every time instance is important for optimizing the operational efforts along the trail and is, as such, the most important aspect. However, since the simulated number depends on how the consecutive marching speeds of individual marchers are combined and on the number of marchers retiring at every checkpoint, also these two aspects are considered.

We resorted to a predictive validation strategy (Sargent, 2011). In our setting, this involves assessing the performance of the spatially explicit model for a given edition (*i.e.* a given year) when the model is constructed from the dataset obtained by deleting all entries of that particular edition. Given the fact that the RFID dataset presented in the Section 3.2 contains entries from five editions, such a validation should be done fivefold. For every edition, the Match and Kolmogorov-Smirnov distances between the observed and simulated EDFs and the corresponding confidence intervals of these distances were calculated (Sargent, 2011).

The simulations were set up as follows. The number of simulated marchers was equal to the number of marchers with a complete set of scans during the event under consideration. The percentage of male marchers and the conditional distributions $\hat{F}_{V_s|G,A,P}$, $\hat{F}_{V_s|G,P}$ and $\hat{F}_{V_s|G}$ were determined on the basis of the tracking data of the other editions. Per edition, 100 Monte Carlo simulations were performed.

The average Match and Kolmogorov-Smirnov distances between the EDFs of the simulated and observed marching speeds at the consecutive sections and passing times along the consecutive checkpoints are shown in Figs. 3.21 and 3.22, respectively. The asterisks represent the average distances for every edition, while the error bars represent the corresponding 95% confidence interval over all editions. The average Match and Kolmogorov-Smirnov distances per edition between the EDFs of the simulated and observed number of marchers retiring at every checkpoint can be found in Table 3.5. When considering all editions, these distances are 0.35 and 0.04, respectively, while the 95% confidence intervals are [0.02, 0.68] and [0.01, 0.07].

Table 3.5: Per edition, Match and Kolmogorov-Smirnov distances between the observed and simulated EDFs of the number of marchers retiring at every checkpoint.

| Validated edition | Average Match distance (-) | Average Kolmogorov-Smirnov distance (-) |
|-------------------|----------------------------|---|
| 2010 | 0.60 | 0.065 |
| 2011 | 0.40 | 0.045 |
| 2012 | 0.21 | 0.022 |
| 2013 | 0.37 | 0.037 |
| 2014 | 0.18 | 0.039 |

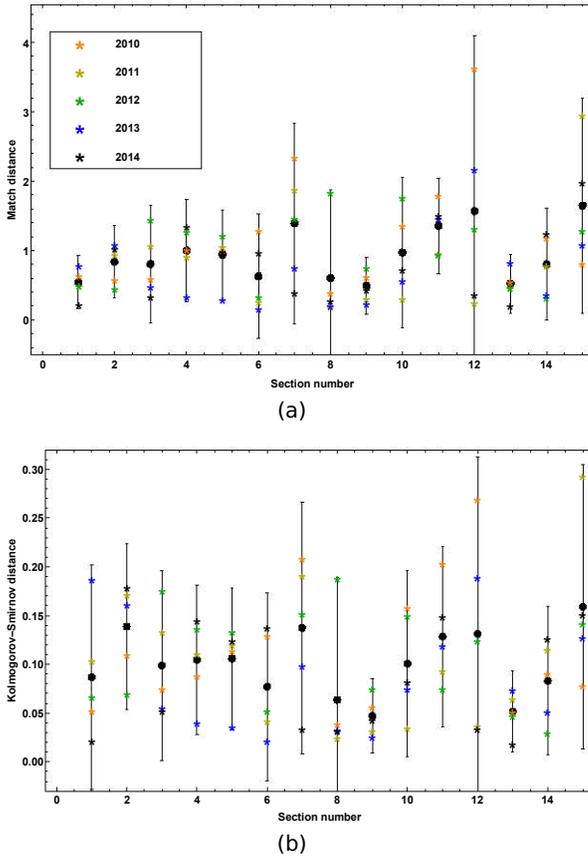


Figure 3.21: Match (a) and Kolmogorov-Smirnov (b) distances between the EDFs of the observed and simulated marching speeds at every consecutive section. The average distances for the five validated editions are represented by colored asterisks, while the error bars represent the corresponding 95% confidence interval over all the editions.

From Fig. 3.21, it can be seen that the 95% confidence interval of the average Match distance is always located in $[0, 4]$, while the one of the Kolmogorov-Smirnov distance is located in $[0, 0.3]$. Comparing this observation to the 95% confidence intervals obtained when an average edition was simulated on the basis of all available data (see Section 3.4.5), which were $[0, 0.7]$ and $[0, 0.06]$, respectively, it is apparent that the simulated marching speeds differ more from the recorded ones if the dynamics of a particular edition was simulated using the spatially explicit marching model that was constructed from the other editions. For what concerns the performance of the model across the editions, there is not a single one that stands out.

Considering the distances between the observed and simulated passing times, it can be seen from Fig. 3.22 that the 95% confidence interval of the average Match distance is always located in $[0, 3]$, while the one of the Kolmogorov-Smirnov

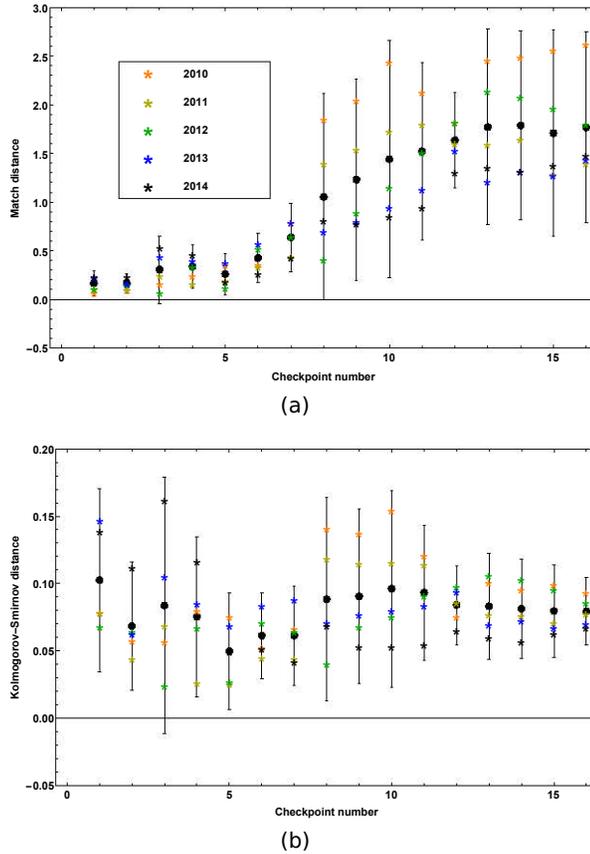


Figure 3.22: Match (a) and Kolmogorov-Smirnov (b) distances between the EDFs of the observed and simulated numbers of marchers passing by at every checkpoint per 0.25 h for the five validated editions. The average distances for the 5 validated editions are represented by colored asterisks while the error bars represent the corresponding 95% confidence interval over all the editions.

distance always lies in $[0, 0.2]$. Comparing these intervals with the ones obtained when an average edition of the 100 km Dodentocht was simulated on the basis of all data (see Section 3.4.5), which are $[0, 2.5]$ and $[0, 0.11]$, respectively, it may be concluded that the passing times estimated for an edition on the basis of tracking data of other editions are equally well simulated as when all tracking data were used to build and validate the model. The average distances of edition 2010 are the largest.

For what concerns the distances between the distributions of the retiring dynamics, the average distances are approximately three times smaller when all data were used to build the model. Again, edition 2010 appears to be the most difficult one to simulate realistically (Table 3.5).

Given its importance from an organizational point of view, the differences between the average simulated and observed number of marchers that pass at consecutive



checkpoints per 0.25 h are given in Table 3.6. For the sake of comparability, these numbers were scaled in such a way that the reported differences are valid for an edition with 10,000 marchers. The average absolute differences per 0.25 h are low, but the maximum differences can be high, especially at checkpoints 2 and 3. As mentioned in Section 3.4.5, the effects of these errors are not that drastic since the marchers are still clustered at the beginning of the trail. Moreover, the shape of the cdfs of the passing times is approximated well, so it can be concluded that busy and relatively calm periods are simulated well. Also in terms of the number of marchers passing at the consecutive checkpoints, edition 2010 stands out, as the maximum differences are relatively high for all checkpoints. This observation is in agreement with our observations on the basis of the Match and Kolmogorov-Smirnov distances (cfr. Figs. 3.21 and 3.22).

Table 3.6: The average and maximum absolute difference between the average of the 100 simulated and the observed number of people passing by at every checkpoint, per 0.25 h for a 100 km Dodentocht with 10,000 marchers. For every edition, the model was constructed from the dataset obtained by deleting all entries of that particular edition.

| Checkpoint ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|----------------------------------|------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Average absolute difference 2010 | 17 | 19 | 21 | 23 | 27 | 17 | 20 | 22 | 18 | 20 | 17 | 17 | 17 | 18 | 17 | 14 |
| Maximum absolute difference 2010 | 804 | 586 | 469 | 638 | 668 | 370 | 182 | 185 | 210 | 275 | 259 | 287 | 246 | 223 | 194 | 143 |
| Average absolute difference 2011 | 16 | 9 | 17 | 9 | 8 | 12 | 10 | 18 | 17 | 17 | 16 | 13 | 14 | 13 | 13 | 12 |
| Maximum absolute difference 2011 | 796 | 207 | 372 | 180 | 120 | 158 | 87 | 188 | 149 | 123 | 103 | 107 | 78 | 84 | 72 | 95 |
| Average absolute difference 2012 | 14 | 13 | 9 | 14 | 7 | 13 | 11 | 9 | 11 | 11 | 13 | 14 | 15 | 16 | 16 | 14 |
| Maximum absolute difference 2012 | 681 | 583 | 237 | 193 | 105 | 195 | 104 | 82 | 67 | 67 | 70 | 70 | 76 | 89 | 91 | 149 |
| Average absolute difference 2013 | 31 | 13 | 22 | 18 | 14 | 17 | 16 | 13 | 13 | 13 | 13 | 13 | 14 | 14 | 15 | 14 |
| Maximum absolute difference 2013 | 1473 | 547 | 517 | 290 | 203 | 178 | 199 | 122 | 98 | 87 | 87 | 73 | 87 | 70 | 102 | 159 |
| Average absolute difference 2014 | 29 | 24 | 34 | 25 | 14 | 13 | 8 | 10 | 8 | 9 | 8 | 10 | 10 | 11 | 11 | 10 |
| Maximum absolute difference 2014 | 1387 | 1131 | 848 | 504 | 245 | 197 | 76 | 104 | 84 | 86 | 74 | 59 | 85 | 77 | 76 | 68 |

Considering the results for model M_2 presented in Section 3.4.5 and the validation results presented throughout this section, we may conclude that the developed spatially explicit marching model enables us to reliably simulate the dynamics of an average 100 km Dodentocht edition, especially for what concerns the passing times as the predictive power for the latter is the highest.

3.5.2 Validation with an independent data set

Here, the model is validated using data from the 2017 edition. Moreover, prior to the event, the estimated flow of participants was simulated with our marching model and used in the event control room in case of emergencies, for better guaranteeing the safety of the participants.

For this edition, participants needed to pre-enroll (for the first time), so it was known that there would be 13,974 registrations, approximately 3,000 more than the average number of participants in 2009–2014. Besides, there were relatively more women than during the editions of 2010–2014, *i.e.* 29.32%, while the age distribution of the 2017 participants was in line with the ones of 2010–2014.

The distances between the checkpoints and the start for edition 2017 and the average for editions 2009–2014 can be found in Table 3.7. The major differences

between edition 2017 and the editions used to train our model are found at checkpoints 7 and 8. In 2017, checkpoints 6 and 7, respectively checkpoints 7 and 8, were 4.9 km, respectively 11.2 km apart, while this was 10.2 km, respectively 6.2 km in the training set. As such, in 2017, checkpoint 7 was located before the symbolic distance of 50 km, in contrast to its location during editions 2009–2014, where it was located at about 50 km. Furthermore, in 2017, warm meals were served at checkpoint 8, while this was at checkpoint 7 during editions 2009–2014.

Table 3.7: Distances along the 100 km Dodentocht trail between the checkpoints and its start point in 2017. Also the average distances of the editions 2009–2014 are shown.

| Checkpoint ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|-----------------------|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|-------|
| Average Distance (km) | 1.2 | 7.5 | 17.6 | 24.6 | 31.7 | 39.8 | 50.0 | 56.2 | 63.8 | 69.1 | 74.2 | 80.3 | 84.6 | 89.7 | 94.4 | 100.0 |
| Distance (km) in 2017 | 1.6 | 12.5 | 18.1 | 25.5 | 30.9 | 37.4 | 42.3 | 53.8 | 62.2 | 67.8 | 73.2 | 79.4 | 82.8 | 86.5 | 94.2 | 100.0 |

Since both age and gender were available, we first determined $\hat{F}_{V_1|G,A}(v_1 | g, a)$ to sample a starting speed for every participant, after which Algorithm 1 was applied. In total, 100 Monte Carlo simulations were performed to calculate a confidence envelope of the simulated marching speeds and of the number of participants passing by at every checkpoint per 0.25 h.

After the event, the organizing committee provided the RFID tracking data to validate our model with this data set. All together, there were 13,030 participants with a meaningful set of scans. The success rate was 67.6% for men, 61.7% for women and 65.9% for all participants. This overall rate is high compared to the ones for the editions 2009–2014 (61%).

In Fig. 3.23, the relative frequencies of the simulated and observed marching speeds are shown along every section. It can be seen that the observed starting speed is shifted towards higher values, while from section 2 on, this is no longer the case. As expected, the major differences between the histograms can be found along sections 7 and 8, as the locations of the corresponding checkpoints were changed as compared to the editions 2009–2014. The marching speeds in section 7 are shifted towards higher values and the drop in marching speeds is now noticeable in section 8 because a warm meal was served at checkpoint 8 now, instead of at checkpoint 7. Also in section 10, there is a clear difference noticeable with unknown cause. These discrepancies can also be seen in Fig. 3.24, where high Match and Kolmogorov-Smirnov distances are noticeable at sections 7, 8 and 10. Compared to the distances calculated in Section 3.5.1, their ranges are now $[0, 5]$ and $[0, 0.5]$ respectively, which is more than double the ranges found for the editions 2009–2014.

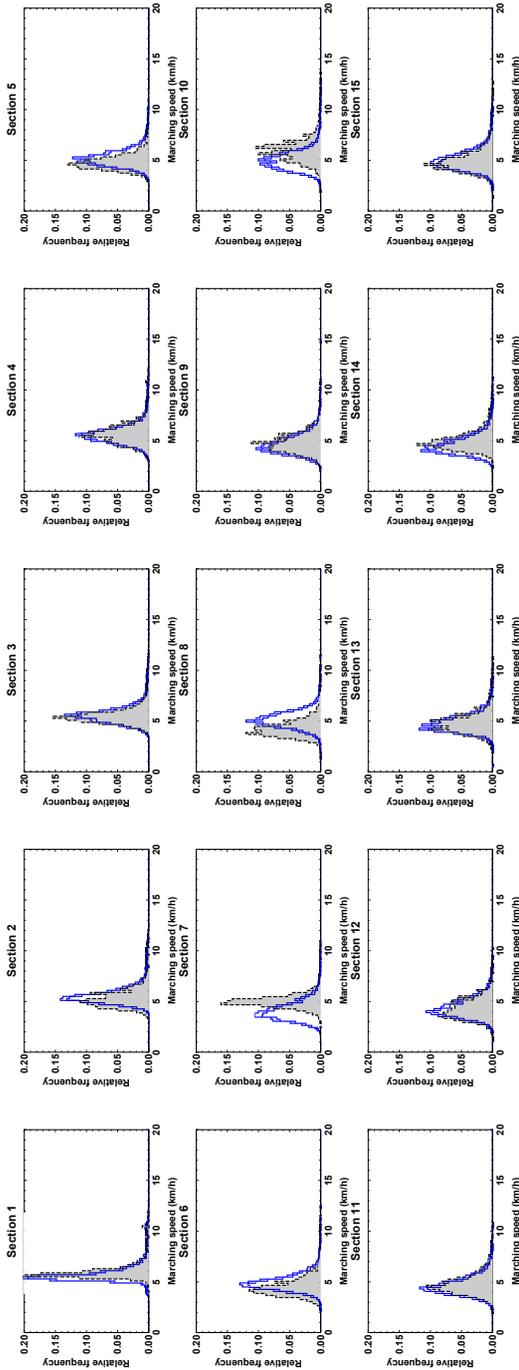


Figure 3.23: The 95% confidence envelope of the simulated marching speeds for edition 2017 (blue), together with the relative frequencies of the observed marching speeds (black and filled with gray).

In Fig. 3.25, the relative number of marchers versus the covered distance is shown. Apparently, retirement is underestimated until checkpoint 6 (37.4 km), after which it is overestimated. In reality, most people retired at checkpoints 7 (42.3 km) and

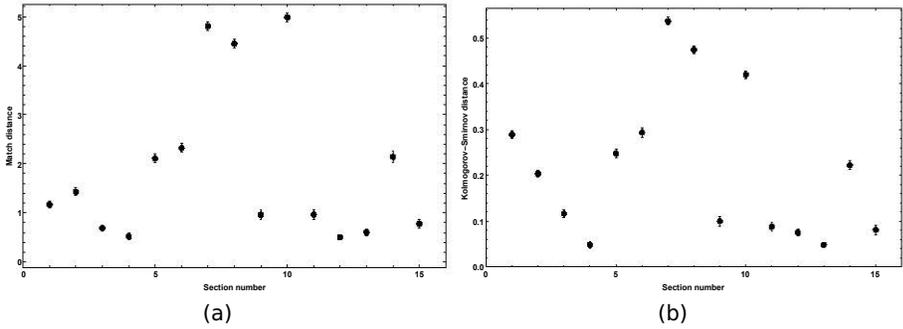


Figure 3.24: Match (a) and Kolmogorov-Smirnov (b) distances between the EDFs of the observed and simulated marching speeds at every consecutive section for edition 2017. The error bars represent the corresponding 95% confidence interval over all simulations.

9 (62.2 km), while *in silico*, most people do this at checkpoints 7 and 8 (53.8 km). This is probably caused by the change in trail layout, where the symbolic distance of 50 km and the warm meals were moved from checkpoint 7 to 8.

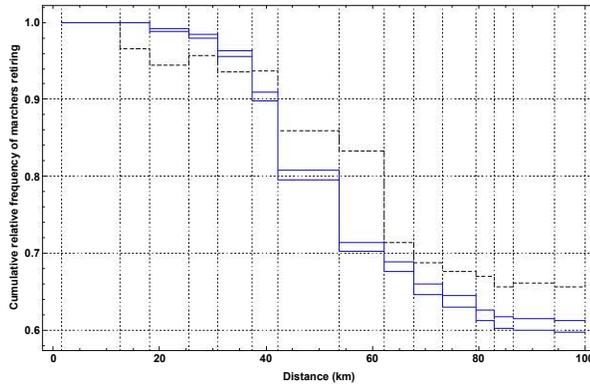


Figure 3.25: The observed (black and dashed) relative number of marchers at every checkpoint and the corresponding 95% confidence interval of the simulations of edition 2017 (blue).

In Fig. 3.26, the observed and simulated numbers of marchers passing by at every checkpoint per 0.25 h is shown. To account for the 994 participants without tracking data, the histograms of the observed numbers were rescaled. In general, one can say that the shape of the curves agrees well until checkpoint 7. From checkpoint 8 on, closing times are not respected as simulated participants still arrive after closure, while the observed curves are more skewed towards earlier time instances. The cause for this could, however, not be identified. The Match and Kolmogorov-Smirnov distances between the EDFs of the observed and simulated numbers of marchers passing by at every checkpoint per 0.25 h for edition 2017 are shown in Fig. 3.27. Compared to the distances calculated in Section 3.5.1, the ranges for edition 2017 are [0, 4] and [0, 0.2] respectively, which are double or



the ones for the editions 2009–2014.

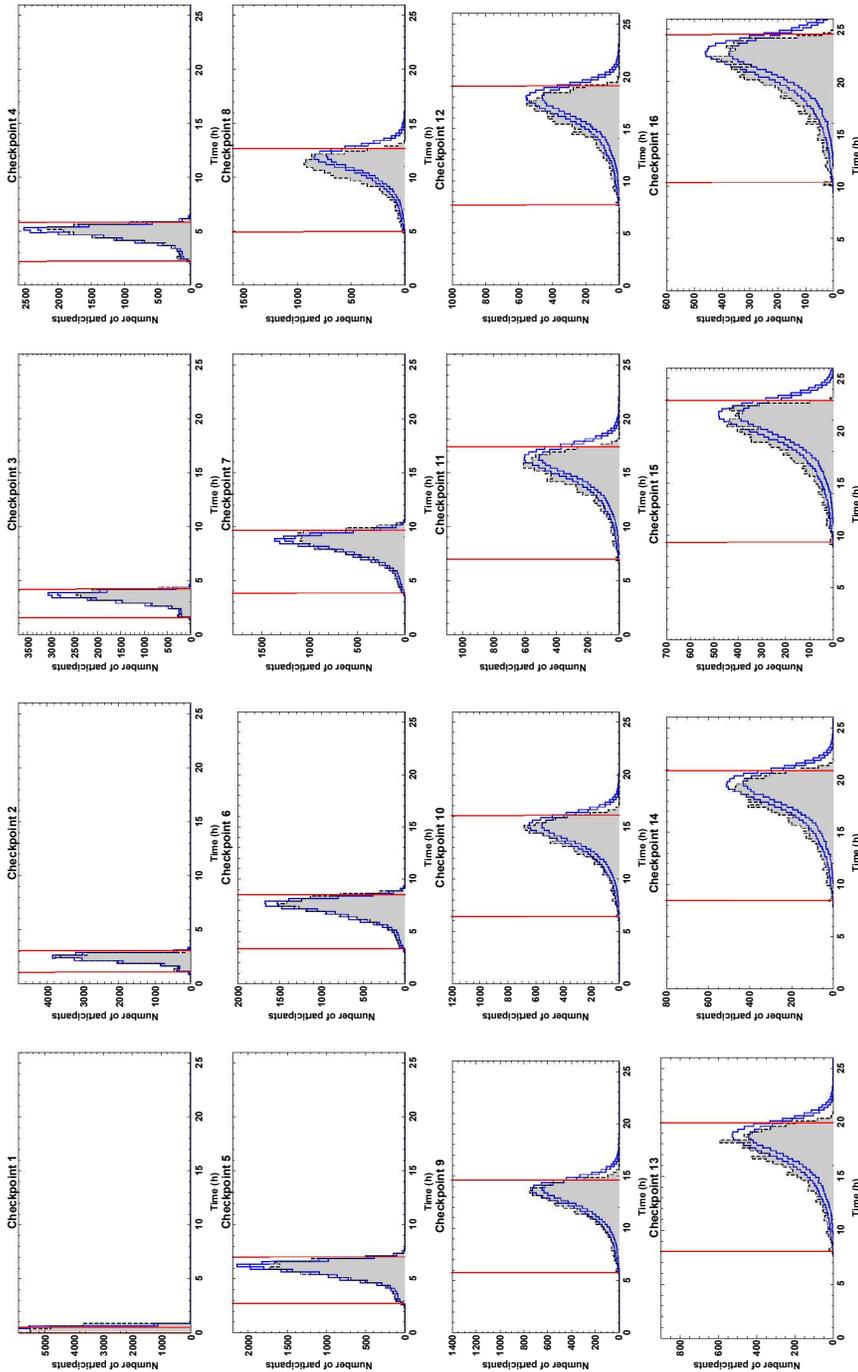


Figure 3.26: The 95% confidence envelope of the simulated number of marchers passing by at every checkpoint per 0.25 h for edition 2017 (blue), together with the observed number of marchers passing by (black and filled with gray). In red, the opening and closing times of the checkpoints are shown.

The comparison of the simulation results and the observed data for edition 2017, shows that our model is not perfect. The modified trail layout and the increase in

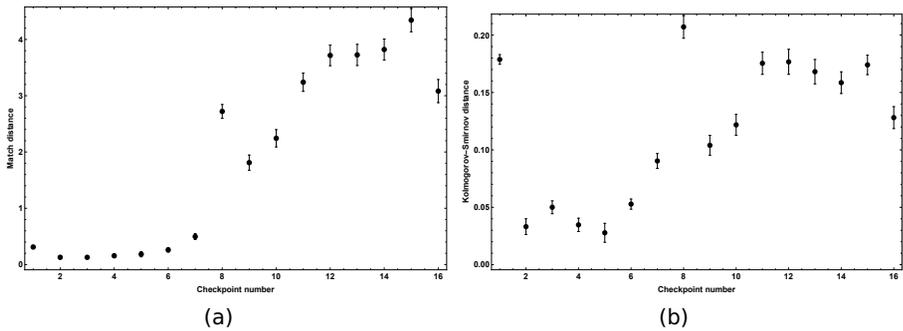


Figure 3.27: Match (a) and Kolmogorov-Smirnov (b) distances between the EDFs of the observed and simulated numbers of marchers passing by at every checkpoint per 0.25 h. The error bars represent the 95% confidence interval over all the simulations of edition 2017.

number of participants could partially explain this. As such, it is illustrated that the model may only be used in a setting similar to the one prevailing during calibration, unless it is adapted accordingly. Nevertheless, based on our model simulations, one can get an idea of the busy and calm periods, but it is not possible to predict the exact number of participants passing by.

3.6 Scenario analysis

After the starting signal of the 100 km Dodentocht, approximately 11,000 marchers start together. The first part of the trail passes through the narrow streets in the town center of Bornem, Belgium, where the marchers are motivated by their followers. It is so crowded along this part of the trail that the marchers have to progress in a flock, which is a situation that the organizers of the endurance event would like to overcome. For that reason, the organizing committee proposed three scenarios on the basis of their experience that are aimed at giving the marchers more space along the first part of the trail. Scenario 1 involves a starting location outside the town center, where the streets are wider, so that the marchers will be scattered more along the trail by the time they enter the town center. In Scenario 2, the starting marchers are divided into two or more groups that start at different locations. Eventually, these groups will meet one another when the marchers are more scattered along the trail. Finally, in Scenario 3 the start is organized in such a way that fewer people pass the starting line at once, which implies that the start will take longer than half an hour. This could be achieved by installing a bottleneck at the start.

In order to assess the effect of the proposed scenarios on the event dynamics, our spatially explicit model was adapted accordingly. Since these modifications imply changes to the starting conditions, the distribution of V_1 needs to be adapted

according to the scenario under consideration. Yet, for doing so, V_1 has to be sampled without using the empirical joint distribution function $\hat{F}_{\mathcal{A}, V_1 | \mathcal{G}}(a, v_1 | g)$, but by relying on a distribution function describing V_1 solely. The latter is conditioned on \mathcal{G} , for which a value is chosen at the beginning of the simulation of a marcher. Since age and speed are correlated (cfr. Fig. 3.11), both variables cannot be sampled independently. Here, we used a copula so that V_1 can be sampled from its marginal distribution.

3.6.1 The copula of starting speed and age

A bivariate copula C , *i.e.* a bivariate probability distribution function on $[0, 1]^2$, was determined for describing the dependence between V_1 and \mathcal{A} . This copula can be used together with the two marginal distribution functions of V_1 and \mathcal{A} , *i.e.* F_{V_1} and $F_{\mathcal{A}}$, to estimate the joint distribution function (Sklar, 1959):

$$F_{V_1 \mathcal{A}}(v_1, a) = C(F_{V_1}(v_1), F_{\mathcal{A}}(a)). \quad (3.6)$$

By sampling a copula that links V_1 and \mathcal{A} , a value between 0 and 1 is obtained for both variables. Subsequently, the inverse transform method (Banks, 1998) was used to transform this value into a sample of V_1 and \mathcal{A} .

Many parametric copula families have been reported in literature (Nelsen, 1998), of which we chose the Frank (Frank, 1979), Clayton (Clayton, 1978) and Gaussian (Salvadori et al., 2007) copula. Each of them was fitted to all pairs of V_1 and \mathcal{A} , after which Kendall's tau (Nelsen, 1998) was used to select the most appropriate one. On the basis of 10,000 observations of V_1 sampled from each copula, we concluded that the Frank and Gaussian copula have a Kendall's tau that is similar to the one of the data. More precisely, they led to a Kendall's tau of respectively 0.110 and 0.095 for men, whereas it was 0.103 when computed on the basis of the data set. Similarly, a Kendall's tau of 0.123 and 0.117 was computed for female marchers for which the data gave rise to a Kendall's tau of 0.124. Since the maximum likelihood of the Frank copula was higher than the one of the Gaussian copula, the former was finally selected to describe the relationship between V_1 and \mathcal{A} . The relevant copula parameters were 0.368 and 0.298, respectively for male and female marchers.

3.6.2 The marginal distribution of the starting speed and its modifications

In order to implement the outlined scenarios, the distribution of V_1 has to be modified accordingly. To make meaningful adaptations, the distribution of V_1 should

be described analytically. On the basis of a preliminary analysis of the data, we chose to fit a mixture of three normal distribution functions \mathcal{M} to the data for men. The parameters of the involved distributions were fit by using the maximum likelihood method in Mathematica (Version 9.0.1, Wolfram Research Inc., Champaign, US). They are listed in Table 3.8, while the fitted mixture distribution is shown in Fig. 3.28, together with the probability mass function (PMF) of V_1 . Similarly, a mixture distribution for women was obtained. The first normal distribution of \mathcal{M} , \mathcal{N}_1 , reflects the group of marchers of the 100 km Dodentocht, the third normal distribution, \mathcal{N}_3 , represents the group of joggers that gives rise to a right tail in the relative frequencies of the marching speeds and finally, the second normal distribution, \mathcal{N}_2 , relates to fast marchers whose characteristics are intermediate between those of joggers and marchers.

Table 3.8: Parameters defining the three normal distributions making up the mixture distribution describing V_1 .

| Denotation | Men | | | Women | | |
|-----------------|-----------------------------|---|--------|-----------------------------|---|--------|
| | Mean (km h^{-1}) | Standard deviation (km h^{-1}) | Weight | Mean (km h^{-1}) | Standard deviation (km h^{-1}) | Weight |
| \mathcal{N}_1 | 5.27 | 0.31 | 75 | 5.21 | 0.26 | 72 |
| \mathcal{N}_2 | 6.26 | 0.43 | 15 | 5.96 | 0.33 | 14 |
| \mathcal{N}_3 | 7.57 | 2.32 | 10 | 6.42 | 2.17 | 14 |

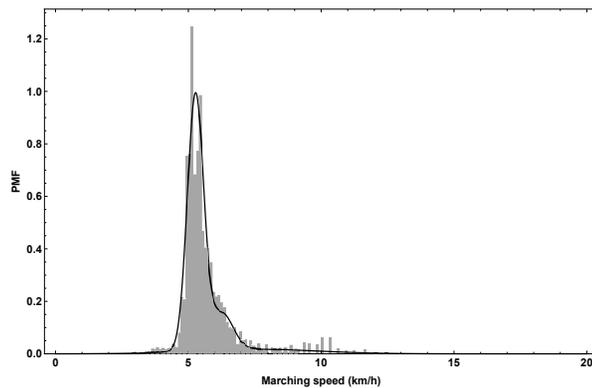


Figure 3.28: Probability mass function of the observed starting speed and the mixture distribution function that was fitted to the corresponding data for male marchers. The mixture distribution involves three normal distribution functions with parameters listed in Table 3.8.

Since the groups of joggers and fast marchers mainly start up front and, as such, suffer less from the crowd along the first part of the trail, it may be anticipated that the proposed scenarios will mostly affect the dynamics of the group of marchers. As such, only the parameters of \mathcal{N}_1 should be adapted when implementing the scenarios. Reducing the number of marchers per unit area along the first part of the trail will increase the starting speed of some marchers because they are less hindered by others, on the one hand, but might decrease the starting speed of others since they are no longer forced to move along with the crowd. In this way, marchers will be able to march at their preferred speed. The net effect will

depend on the ratio between both types of marchers. Ultimately, these effects will imply changes to both the mean and the standard deviation of the distribution of the marchers' marching speeds. For comprehensiveness, we first conducted a sensitivity analysis in order to assess the impact of a varying mean and standard deviation of \mathcal{N}_1 on the event dynamics, after which we turned to an encoding of the scenarios under consideration.

3.6.2.1 Varying the mean starting speed of marchers

3 In order to examine the effect of a varying mean, the mean of \mathcal{N}_1 was successively replaced by 4 km h^{-1} , 4.5 km h^{-1} , 5 km h^{-1} , 5.5 km h^{-1} and 6 km h^{-1} for both men and women, as opposed to the ones derived from the RFID data (cfr. Table. 3.8). The relative frequency of the simulated marching speeds at every section and the passing times at the corresponding checkpoints can be found in Figs. 3.29 and 3.30, respectively. One can see that the effect on the marching speed is the most pronounced along the first few sections and gently fades out along sections at greater distances from the start. If the average marching speed is higher than the one of the benchmark situation (*i.e.* as given by the RFID data), the difference between the distribution of the marching speeds of the marchers and the ones of the fast marchers and joggers becomes smaller.

The higher the mean marching speed, the earlier marchers pass the checkpoints. As the average marching speed increases, the maximum number of people passing per 0.25 h at the first three checkpoints becomes much higher than in the benchmark situation. This can be understood because the characteristics of the marchers gradually align with those of the fast marchers and joggers. Consequently, we concluded that it will become busier near the checkpoints when the average starting speed increases, which is the opposite of the situation pursued by the organizers.

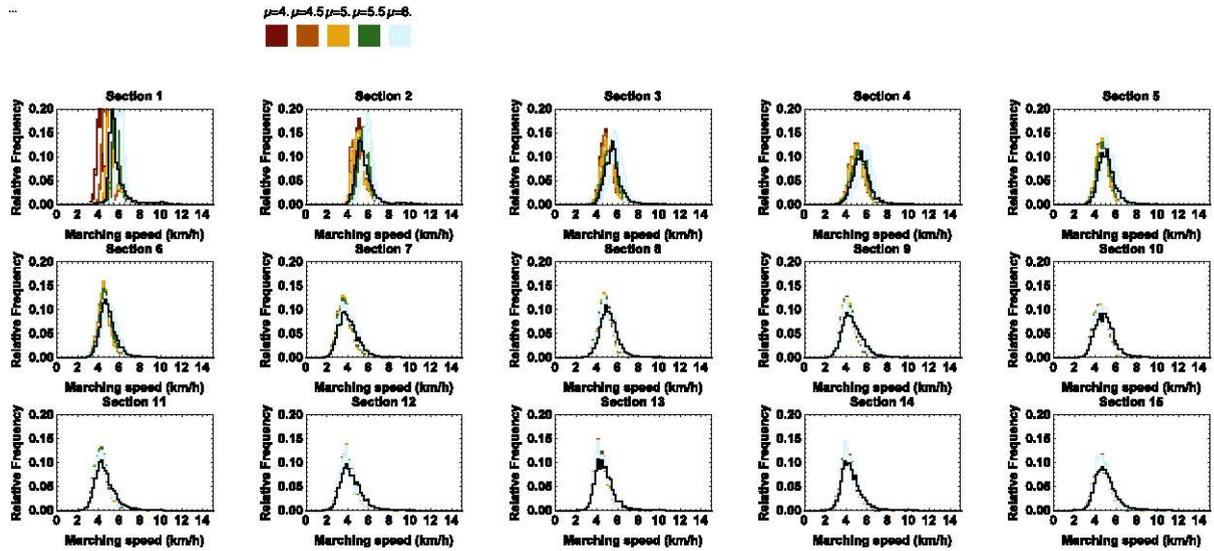


Figure 3.29: The 95% confidence interval of the relative frequency of the simulated marching speeds when the mean of N_1 varied between 4 km h^{-1} and 6 km h^{-1} . The black line represents the benchmark situation.

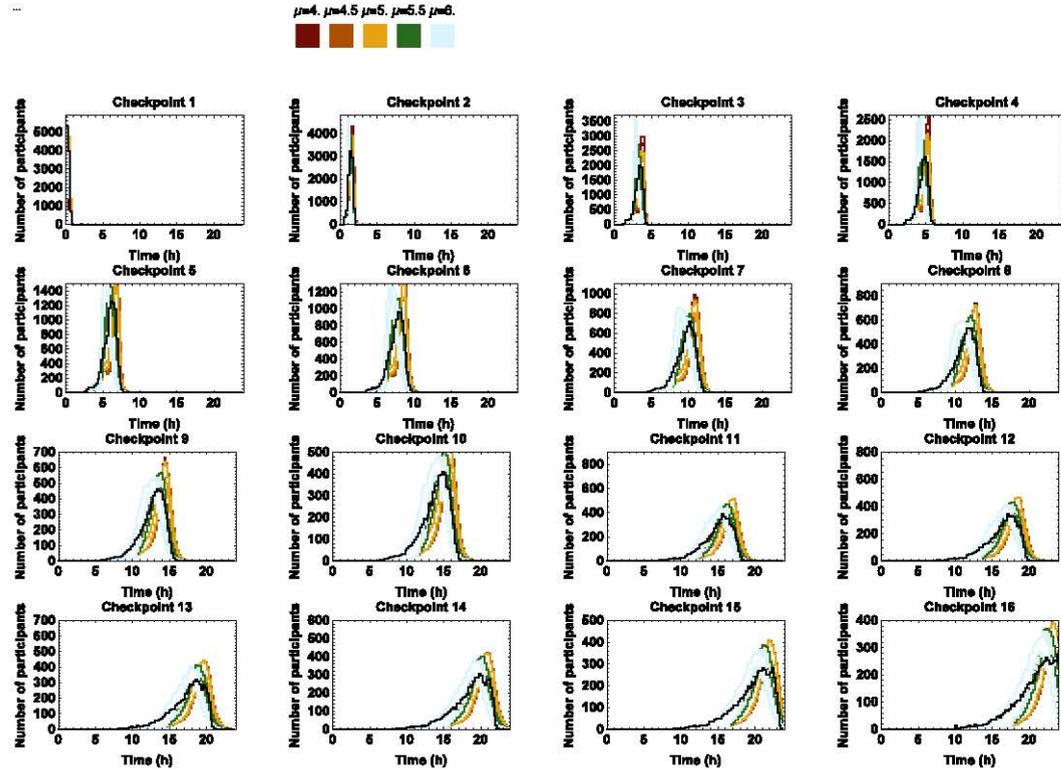


Figure 3.30: . The 95% confidence enveloped of number of marchers passing by at the consecutive checkpoints per 0.25 h when the mean of \mathcal{N}_1 varied between 4 km h^{-1} and 6 km h^{-1} . The black line represents the benchmark situation.

3.6.2.2 Varying the standard deviation of the starting speed of marchers

In order to examine the effect of a varying standard deviation, the standard deviation of \mathcal{N}_1 was successively replaced by 0.3 km h^{-1} , 0.4 km h^{-1} , 0.5 km h^{-1} , 0.6 km h^{-1} and 0.7 km h^{-1} for both men and women, as opposed to the ones derived from the RFID data (cfr. Table 3.8). We chose to restrict our analysis to the effect of values greater than or equal to the ones of the data since the amount of variation in the starting speed was considered relatively low given the much more pronounced variability along sections further away from the start. The histograms of the simulated marching speeds along the subsequent sections and the passing times at the corresponding checkpoints can be found in Figs. 3.31 and 3.32, respectively.

Apparently, increasing the standard deviation only affects the marching speeds along the first two sections of the trail. The shape of the histograms of the passing times closely resembles the shape of the distributions of the benchmark situation from checkpoint 3 on. Only at the first few checkpoints, it becomes slightly calmer. Increasing the standard deviation would thus give marchers more space to move freely along the first sections, as such improving the starting situation.

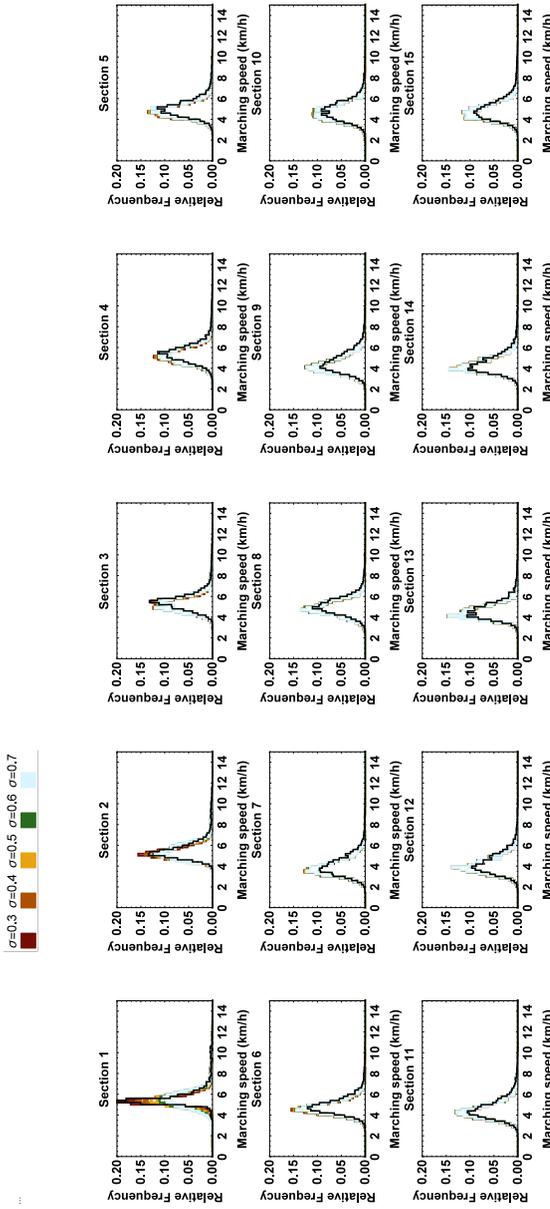


Figure 3.31: The 95% confidence interval of the relative frequency of simulated marching speeds when the standard deviation of \mathcal{N}_1 is varied between 0.3 km h^{-1} and 0.7 km h^{-1} . The black line represents the benchmark situation.

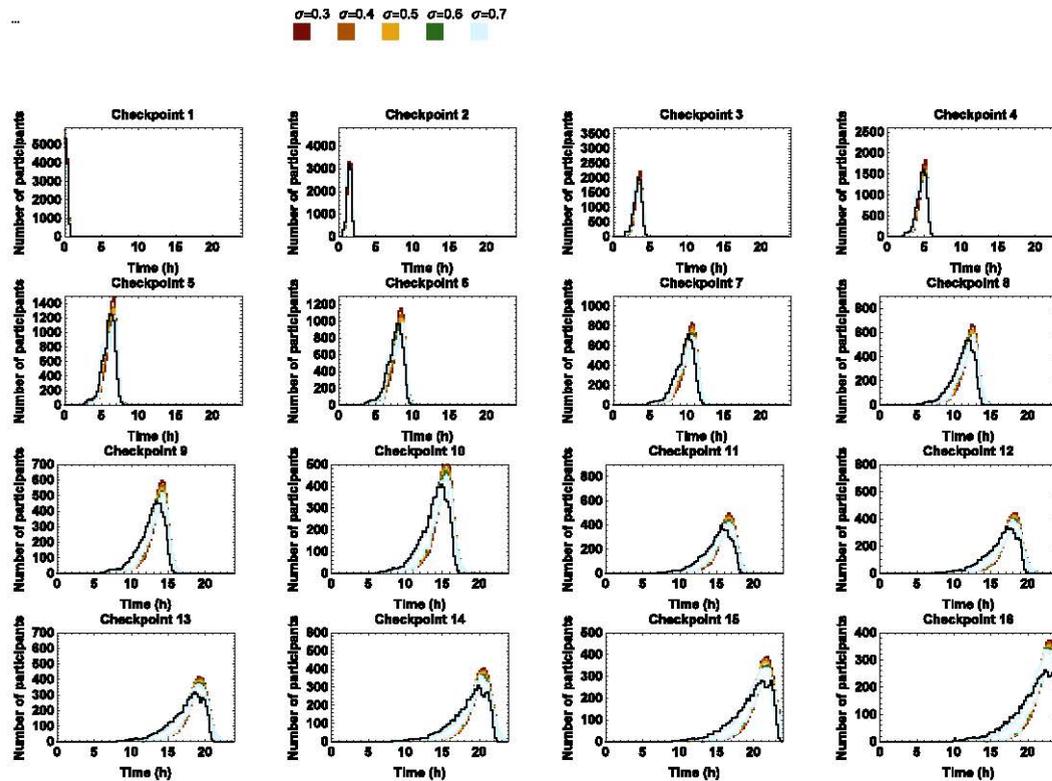


Figure 3.32: The 95% confidence interval of the number of marchers passing by at every checkpoint per 0.25 h when the standard deviation of \mathcal{N}_1 is varied between 0.3 km h^{-1} and 0.7 km h^{-1} . The black line represents the benchmark situation.

3.6.3 Scenarios

3.6.3.1 Encoding

The scenarios outlined at the beginning of this section aim at reducing the density of marchers along the first part of the trail. For example, starting in two groups will reduce the density by 50% (Scenario 2). Furthermore, a density reduction of 10%, 25% and 50% could be the result of starting outside the town center, or by installing a bottleneck at the starting line (Scenarios 1 and 3). In order to assess the impact of these scenarios on the event dynamics, the parameters of \mathcal{N}_1 were changed in order to reflect them appropriately.

The Kladek formula (Weidmann, 1993) expresses that the average marching speed increases with decreasing density. We used the revised formula from Venuti and Bruno (2007) which is valid for rush hour situations, to assess the effect of a decreasing density on the starting speed:

$$v = v_M \left(1 - \exp \left(-\gamma \left(\frac{1}{u} - \frac{1}{u_M} \right) \right) \right). \quad (3.7)$$

Here, v refers to the marching speed, v_M is the free speed, which corresponds to a mean maximum velocity of 6.08 km h^{-1} , u is the marchers' density and u_M is the jam density, which corresponds to the density where the speed becomes 0 km h^{-1} , and equals 3.93 m^{-2} . The parameter γ equals 1.07 (Venuti and Bruno, 2007).

In Venuti and Bruno (2007), the critical density u_c that corresponds to the upper bound on the marchers' density beyond which unconstrained marching becomes impossible, is calculated as

$$u_c = \frac{1}{S(v_M)}, \quad (3.8)$$

with $S(v)$ the required surface area for a marcher with speed v , which is obtained from:

$$S = w(1.075 l_p + 0.55 d_s), \quad (3.9)$$

where $w = 0.45(1 + 0.62 v/1.34)$ equals the lateral width of a pedestrian, $l_p = 1/(0.35 v^2 - 1.59 v + 2.93)$ equals its step length and $d_s = 0.36 + 1.06 v + b v^{10} (2.08 v_M - 0.36)/v_M^{10}$ is the distance needed between two consecutive marchers in the marching direction. Using Eqs. (3.8) and (3.9) with $v = v_M$, it follows that u_c equals 0.33 m^{-2} (Venuti and Bruno, 2007). Using the critical density u_c in Eq. (3.7), it follows that people moving faster than 5.77 km h^{-1} may be considered as marching freely. As such, considering the parameters in Table 3.8, we observe that intended scenarios primarily affect the group of marchers described by \mathcal{N}_1 . Hence, it is justified to modify the parameters of this distribution only. In literature, it has been reported that the marching speed

follows a normal distribution (Daamen and Hoogendoorn, 2006; Weidmann, 1993), that the standard deviation of freely marching individuals is 1.33 km h^{-1} (Daamen and Hoogendoorn, 2006) but, to the best of our knowledge, there are no reports on the relation between the standard deviation and the density of the marchers.

Equation (3.7) was used to estimate the density of the marchers at the start of the 100 km Dodentocht on the basis of the mean marching speeds given by \mathcal{N}_1 for men and women (cfr. Table 3.8). We found a density of 0.72 m^{-2} and 0.70 m^{-2} , respectively, resulting in an average density of 0.71 m^{-2} . Since this density is much larger than $u_c = 0.33 \text{ m}^{-2}$, it is obvious that marchers cannot move freely along the first part of the trail of the 100 km Dodentocht.

Since the variability in starting speeds is relatively limited compared to the variability in marching speeds along sections further away from the start, as most marchers did not retire along sections 2 and 3 and knowing that the speeds along those sections are still correlated with the starting speed (cfr. Section 3.3), we studied the distribution parameters of the marching speed along sections 2 and 3 (V_2 and V_3). A mixture distribution of three normal distributions was also fitted to V_2 and V_3 . For men, the first normal distribution of V_2 and V_3 has a standard deviation of 0.59 km h^{-1} and 0.6 km h^{-1} respectively. For women, these standard deviations are 0.5 km h^{-1} and 0.51 km h^{-1} respectively. Comparing these values with the ones in Table 3.8, it is obvious that the standard deviation along the first section is nearly half of the ones along sections 2 and 3. As such, the average standard deviation of the distributions of V_2 and V_3 was used (*i.e.* 0.60 km h^{-1} for men and 0.51 km h^{-1} for women) to replace the standard deviation of \mathcal{N}_1 .

When running simulations for scenarios involving an increased average starting speed, marchers are expected to cross the starting line earlier than in the benchmark situation (cfr. Fig. 3.13), as they align with fast marchers and joggers. Consequently, all marchers would have crossed the starting line in a shorter time span, which means that the density of marchers along the first section would again increase. This should be avoided by managing the number of people passing the starting line in such a way that the duration of the starting procedure does not change. In Scenario 3, however, the duration of the start should be longer as fewer people are allowed to pass the starting line at once compared to the benchmark situation. For that purpose, the sampled passing times were rescaled.

3.6.3.2 Analysis

The impact of the scenarios on the event dynamics and operational efforts was assessed by means of *in silico* experiments. More precisely, 100 events were simulated, with 10,000 marchers each, per scenario. The obtained *in silico* event dynamics were compared with the benchmark situation.

Scenario 1 Scenario 1 involves a start that is located outside the center of Bornem, Belgium. Obviously, to what extent the density of the marchers is reduced, depends on the amount of additional space that the wider streets outside the town center bring along. For that reason, we evaluated the effect of a reduction of the marcher density at the start by 10%, 25% and 50%. Using the revised Kladek formula, these reductions correspond to an average starting speed of 5.44 km h^{-1} , 5.69 km h^{-1} and 5.99 km h^{-1} , respectively. Besides, the standard deviation of \mathcal{N}_1 was chosen to be 0.60 km h^{-1} for men and 0.51 km h^{-1} for women. Decreasing the density at the start even further would not change the distribution parameters substantially, since the marchers can almost move freely when it is reduced by 50%. In agreement with the sensitivity analysis presented in Section 3.6.2, these reductions only affect the marching speeds along the first two sections, while the passing times are affected between checkpoints 2 and 6. Moreover, the maximum frequency of marchers passing the checkpoints per 0.25 h occurs around the same time as in the benchmark situation, but the average maximum number of people passing per 0.25 h at checkpoints 2 to 3 is higher than in the benchmark situation. This increase becomes more pronounced as the density decreases more, and can also be observed at checkpoints farther away from the start (Table 3.9). At those checkpoints, the organization should deploy more resources.

This observation can be explained by the fact that the characteristics of the marchers align with those of joggers and fast marchers, as such jeopardizing the mitigating effect of wider streets. At checkpoint 2, for example, the maximum number of marchers registered per 0.25 h increases by 15%, 20% and 32% as compared to the benchmark situation if the density is reduced by 10%, 25% and 50%, respectively. Assuming that the streets are still 10%, 25% or 50% wider at checkpoint 2, this implies a net density increase of 5%, a decrease of 5% and a decrease of 18%, respectively, at that particular checkpoint, which implies that the streets should be at least 25% wider in order to have a beneficial effect on the marcher density along the trail.

The passage of the event through Bornem is considered as a true town fest. The presence of the marchers and their supporters has a substantial commercial value for local businesses in Bornem. Therefore, it is important that the participants pass as early as possible through the town center so that their supporters visit the town shops, restaurants and pubs. In the benchmark situation, at most 5,332 marchers pass at checkpoint 1 per 0.25 h, and a density reduction of at least 50% is needed to ensure that the marchers can move freely along that part of the trail (cfr. Section 3.6.3.1). Hence, at most 2,666 may pass per 0.25 h by the checkpoint located in the town center. Consulting Table 3.9, we may conclude that, despite the obtained density reduction, the maximum number of people passing by checkpoint 3 (*i.e.* 17.6 km) in a 0.25 h interval has decreased to such an extent that it may be located in the town center, while still not causing any discomfort for the marchers.

Table 3.9: The difference between the observed maximum and the average simulated maximum number of marchers per 0.25 h with a density reduction of 10%, 25% and 50% at the start, obtained by means of wider streets. The maximum number of people passing per 0.25 h interval in the benchmark situation is given in the first row.

| Checkpoint ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|--------------------------|-------|-------|-------|-------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Benchmark situation | 5,332 | 3,228 | 1,992 | 1,615 | 1,262 | 971 | 721 | 537 | 464 | 410 | 394 | 344 | 322 | 308 | 281 | 276 |
| Density reduction of 10% | -226 | 482 | 330 | 31 | -49 | -46 | -43 | 5 | 27 | 29 | 14 | 29 | 33 | 36 | 48 | 41 |
| Density reduction of 25% | -226 | 639 | 539 | 260 | 79 | -34 | -87 | -37 | -8 | 1 | -10 | 11 | 19 | 23 | 39 | 31 |
| Density reduction of 50% | -222 | 1031 | 805 | 517 | 327 | 109 | -35 | -47 | -30 | -12 | -22 | 2 | 12 | 15 | 32 | 23 |

Scenario 1 implies that a new route has to be laid out taking into account the findings above. As such, this implies that one or more new locations for checkpoints have to be identified and, since the start would be outside the town center, additional drinking and eating stands have to be installed on top of to the usual equipment. Consequently, introducing this scenario would require organizational efforts. Still, once the structural changes have been introduced, they can remain unchanged for several years. Moreover, the efforts have to be made prior to the event, while there would not be an extra work load during the event since opening hours and number of participants passing by at the consecutive checkpoints would only slightly change. Nevertheless, since marchers would only pass through the town center after 17.6 km, this scenario jeopardizes the revenues of the local businesses and the popularity among the locals. A starting location at the outskirts of the town so that the marchers leave the town into a wide street, might be a compromise since they and their followers could then still linger around town at the beginning of the event.

Scenario 2 In Scenario 2, the marchers are divided into two or more groups that start at different locations. Eventually, these groups will meet when the marchers are already more scattered along the trail. Considering two groups, a start that takes up to half an hour and a street design as in the benchmark situation, it follows that the density at the start reduces with 50%, so that the simulation results boil down to the ones for Scenario 1. Yet, the number of people passing per 0.25 h as reported in Table 3.9 should be divided by two. Doing so, it may be concluded that the two groups ideally merge at a distance of about 17.6 km (checkpoint 3) because the maximum number of people passing by is lower than 2,666, as such allowing free movement. Considering even more groups would not have a much larger effect because the marchers can move almost freely as soon as they are divided into two groups.

Scenario 2 would involve extra work for the organizers both prior to and during the event. An alternative, parallel route for the first 17.6 km has to be found, which then has to be made free of traffic, equipped with toilets, etc. Furthermore, an extra start and second checkpoint have to be chosen. On the other hand, however, less people pass by these checkpoints, so less volunteers have to be present per checkpoint. In order to avoid disadvantages for the local businesses

(cfr. Scenario 1), the organizers should install both starting locations in the town center.

Scenario 3 The third scenario aims at reducing the density by installing a bottleneck at the start in order to limit the number of marchers that can pass per unit of time. Again, we considered a reduction of the density at the start by 10%, 25% and 50%. As opposed to Scenarios 1 and 2, the start will take longer, which was mimicked by rescaling the starting times drawn from the distribution function of the experimental starting times, *i.e.* by dividing them by 0.9, 0.75 and 0.5, respectively. The maximum number of people passing per 0.25 h at checkpoint 1 is lower, irrespective of the density reduction, as compared to the benchmark situation, with the lowest frequencies occurring for the highest density reduction (Table 3.10).

Table 3.10: The difference between the observed maximum and the average simulated maximum number of marchers per 0.25 h with a density reduction of 10%, 25% and 50% at the start, obtained by means of a longer duration of the start. The maximum number of people passing per 0.25 h interval in the benchmark situation is given in the first row.

| Checkpoint ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|--------------------------|--------|-------|-------|-------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Benchmark situation | 5,332 | 3,228 | 1,992 | 1,615 | 1,262 | 971 | 721 | 537 | 464 | 410 | 394 | 344 | 322 | 308 | 281 | 276 |
| Density reduction of 10% | -613 | 244 | 248 | 28 | -56 | -49 | -49 | 4 | 23 | 29 | 14 | 27 | 30 | 33 | 47 | 40 |
| Density reduction of 25% | -1,180 | 80 | 303 | 152 | 44 | -41 | -89 | -42 | -10 | 1 | -9 | 10 | 20 | 22 | 38 | 31 |
| Density reduction of 50% | -2,209 | -395 | -99 | 7 | 49 | 6 | -61 | -54 | -36 | -19 | -32 | -6 | 3 | 8 | 28 | 22 |

Similar to the alignment of marchers with fast marchers and joggers observed in Scenario 1, the mitigating effect of a longer duration of the start is partly compensated by the higher average marching speed. Apparently, the density should be reduced by 50% in order to arrive at lower densities at checkpoints 2 and 3.

The longer duration of the start implies that the closing times of the checkpoints should be postponed, while the opening times should stay the same so their total operating time would increase. Based on the *in silico* results, we concluded that checkpoints should be opened for an extra one and a half hour along the first part of the trail and for one extra hour near the end of the trail. The flow of people would be more gradual, but the peaks in the number of participants would remain high from checkpoint 4 on, so we expect that the operational efforts will be somewhat higher.

3.7 Discussion

In this chapter, we presented a validated spatially explicit model for simulating the endurance marching event 'the 100 km Dodentocht'. To the best of our knowledge, this work is the first of its kind aiming at building a spatially explicit model

that mimics the dynamics of such an endurance event on the basis of RFID tracking data. The model is data driven because it relies on the patterns that are uncovered in the RFID dataset and finally encoded in terms of conditional distribution functions and copulas.

Although we had GPS tracking data at our disposal that were translated into distances from the start, we built our model on the RFID tracking data set because this dataset covered five years and contained a huge number of participants. Furthermore, since RFID data were sampled at specific locations, they were easy to work with. For the GPS tracking data, on the contrary, we only had two years at our disposal and few participants were tracked. Moreover, after the data set clean up, we noticed that the time steps between two consecutive locations and the accuracies were highly variable.

As can be seen from Figs. 3.16 and 3.17, the proposed model is able to mimic the dynamics of the 100 km Dodentocht realistically, which comes forward by the fact that the relative frequencies of the observed and simulated marching speeds along the consecutive sections of the trail are similar. For what concerns the number of people passing per 0.25 h at the consecutive checkpoints, the corresponding histograms have a similar shape, but from checkpoint 7 on, the peaks are overestimated, while the tails towards the left (earlier passing times) are underestimated. This discrepancy is caused by a variety of features having an influence on the event dynamics, of which only a few (age, gender and previous marching speeds) could be taken into account. The influence of weather conditions and small changes in trail layout, for example, could not be taken into account since, based on the five editions available, which all had slightly different weather conditions and trail layout, we were not able to quantify their specific effects. For that purpose, including some similar ones concerning weather and trail layout, are needed. Another important cause for the discrepancy is the fact that pauses could not be identified using the RFID data. The pauses make that there are sudden drops in the observed marching speeds that cannot be grasped by the spatially explicit marching model. More explanatory variables and information about the resting behavior are needed to incorporate this correctly. GPS tracking data were analyzed for this purpose, but since not all participants take a rest at the same time and since the resting periods vary, we could not draw general conclusions about resting that could be included in the model. The latter can, for example, be easily obtained by installing two mats, one in front of and one just after the checkpoint. Furthermore, the participants themselves and their training, food supply and support cause, amongst other features, variability in the event dynamics. Since there was no information or data on their influence, they could not be taken into account in our study.

All together, our model is not able to predict the exact number of participants passing by at every time interval, but we think it is nevertheless useful to estimate busy and calm periods at each checkpoint. When using the model for practical

purposes, one should take into account possible errors that might lead to organizational issues (Table 3.6). Moreover, the model may only be used in a setting similar to the one considered during calibration, unless it is adapted accordingly. This was illustrated when the simulation results and the observed data for edition 2017 were compared. Higher Match and Kolmogorov-Smirnov distances were obtained in comparison with the ones obtained from cross-validation mainly due to the fact that the locations of checkpoints 7 and 8 differed from the ones during the 2010–2014 editions (cfr. Section 3.5.2).

3 The spatially explicit model may be used to assess the possible effects of changes to the trail, the organization and the characteristics of the marchers on the marching speeds and the resulting passing times, on the condition that the considered modifications can be translated into changing model parameters. Often, it is crucial to know the impact of such modifications on the event dynamics before implementing them, but aside from relying on a mathematical model, there is no way to accomplish this. In this chapter, we examined the effects of structural changes to the trail (Scenario 1), and the starting procedure (Scenarios 2 and 3) attempting at a decrease of the marchers' density. Based on our simulation results, we may conclude that a decrease of the density at the start does not necessarily induce such a decrease at subsequent checkpoints. A density reduction at the start of at least 50% implies lower densities at all subsequent sections, irrespective of the specific measure under consideration. As such, it is important to estimate the density reduction that matches the measure under consideration in order to distribute the resources efficiently over the trail. The developed spatially explicit marching model can also be used to assess the effects of changes in the age distribution of the marchers or the proportion of female marchers. This is an especially interesting perspective given the changing demography in Belgium and many other countries. The effect of a totally different trail layout, or other scenarios that are hard to translate into appropriate model parameter values, cannot be researched with the presented model.

Similar to the 100 km Dodentocht, most endurance (walking, running, cycling) events have a predefined route with several checkpoints and make use of RFID tracking systems to follow their participants. Therefore, the presented modeling methodology – including model construction, model selection and (cross-) validation – can be used for analyzing and improving other such events. Yet, sufficient tracking data should be available in order to grasp the variability among participants, and finally build reliable EDFs that determine the speeds on the basis of some known features. A data analysis revealing the features influencing the speed of the participants should always be performed prior to the determination of the conditional EDFs. During such an analysis, one should have a look at the possible influences of all features for which data is available. Typically, some personal characteristics of the participants, trail layout, weather conditions and the auto-correlation between the speeds at each section can be considered. However, also

other features might influence the dynamics and should be taken into account if their influence can be quantified. Next to the analysis of available data, it is also important to get familiar with the event by doing field visits and communicate with the organizers, participants, supports and residents. The latter is helpful when interpreting analysis results and deciding on the inclusion of extra features into the movement model.



4

An analytical description of the time-integrated Brownian bridge

4.1 Introduction

A Brownian motion, conditioned by a starting and ending location, is called a Brownian bridge (BB) (Ross, 1996). The particle's position at any moment in time is described by a normal distribution. The (time-independent) marginal probability density of occurrence at each location is obtained by averaging these normal distributions over time. The probability density function (PDF) of this so-called time-integrated Brownian bridge (TIBB) is used in animal movement research to identify important regions on the basis of tracking data describing the locations of individuals at certain moments in time (Bullard, 1991; Horne et al., 2007). Such

This chapter was published as Van Nieuland et al. (2015). Additionally, this chapter is extended with the derivation of distribution parameters of the one-dimensional TIBB. Applications of the TIBBs can be found in Chapters 5 and 6.

an identification of important regions is realized by the Brownian Bridge Movement Model (BBMM) through the use of TIBBs (Horne et al., 2007; Nielson et al., 2012). More specifically, the BBMM constructs the PDF of a weighted average of TIBBs between every consecutive pair of observations (Bullard, 1991; Horne et al., 2007). The resulting PDF describes the probability density that an animal is located at a certain point at an arbitrary moment in time within the considered time interval. It is determined by the registered locations, the time between those measurements and the animal's mobility (Bullard, 1991; Horne et al., 2007), which are essential for the delineation of important areas and are not accounted for in other approaches such as probability density estimates based on the density of observations (Bullard, 1991; Horne et al., 2007).

The BBMM, and more specifically the TIBB, has already been studied in depth and adapted to account for changes in behavior (Kranstauber et al., 2012), habitat preferences (Benhamou, 2011; Kranstauber et al., 2014) and moving-resting processes (Yan et al., 2014). It has been used to perform home range and animal movement analyses. For example, the migration route of caribou (*Rangifer tarandus*) (Horne et al., 2007) and mule deer (*Odocoileus hemionus*) (Sawyer et al., 2009), as well as the utility distribution of the herbivorous fish *Sarpa sarpa* (Pàges et al., 2013) and the home ranges of black and turkey vultures (*Coragyps atratus* and *Cathartes aura*) and other species (Buchin et al., 2012; Bullard, 1991; Byrne et al., 2014; Kranstauber et al., 2012; Yan et al., 2014), have been investigated through the use of the BBMM or one of its extensions. The latter are implemented and available within the 'BBMM' and 'move' packages for R, which makes the model accessible and easy to use (Kranstauber et al., 2014; Nielson et al., 2012).

Yet, to the best of our knowledge, the integration of the normal distributions that is required to obtain the PDF of the TIBBs in the BBMM is thus far always performed through the use of numerical integration methods. Here, we demonstrate that it is nevertheless possible to derive an analytical expression that avoids inaccurate results and long computing times. It contains exponential, error and Bessel functions, which are implemented efficiently in mathematical software. As increasingly more human and animal tracking data sets are becoming available to researchers (Miller, 2010), the deployment of accurate and efficient data analysis techniques is of utmost importance.

This chapter contributes to the efforts in this domain by making numerical integration methods redundant when dealing with the TIBB and the BBMM. Although less important for animal movement, the PDF of the TIBB in a one-dimensional Euclidean space is derived in Section 4.3, while a two-dimensional Euclidean space is considered in Section 4.4. A three-dimensional Euclidean space is considered in Section 4.5 as it allows us to introduce the reasoning and steps that are needed for tackling problems in higher dimensions. A comparison of the time efficiency of the analytically and numerically determined PDFs of the TIBB can be found in Section 4.6, which is followed by a short discussion in Section 4.7. First, an overview

of the mathematical notations is given in Section 4.2.

4.2 Preliminaries

Let $X(t)$ describe the position of a particle or individual in a one-dimensional Euclidean space at time t such that its starting location can be represented as $X(0) = a$ and its ending location as $X(T) = b$. If Brownian motion is assumed between these two locations, we have that $X(t) \sim \mathcal{N}(\mu(t), \sigma^2(t))$ for any $t \in [0, T]$. Here, $\mu(t) = a + t(b - a)/T$ represents the mean and $\sigma^2(t) = \sigma_m^2 t(T - t)/T$ the variance of the normal distribution. The mobility factor σ_m^2 embodies the mobility of the particle or individual and is related to the ease with which it moves (Horne et al., 2007). Further, it holds that $a, b \in]-\infty, +\infty[$, $T \in]0, +\infty[$ and $\sigma_m^2 \in]0, +\infty[$. The probability density function (PDF) of $X(t)$ is then given by (Bullard, 1991)

$$f_X(x | t) = \frac{1}{\sqrt{2\pi\sigma^2(t)}} \exp\left(-\frac{(x - \mu(t))^2}{2\sigma^2(t)}\right), \quad (4.1)$$

for any $t \in [0, T]$. Note that $f_X(\cdot | 0) = \delta_a$ and $f_X(\cdot | T) = \delta_b$, where δ_x is the Delta Dirac function at x . The Brownian bridge can also be considered in an n -dimensional Euclidean space. There, $\mathbf{X}(t)$ is a vector of size n that describes the position of a particle or individual at time t . The starting and ending location can now be represented by the n -dimensional vectors $\mathbf{X}(0) = \mathbf{a}$ and $\mathbf{X}(T) = \mathbf{b}$, such that $\boldsymbol{\mu}(t) = \mathbf{a} + t(\mathbf{b} - \mathbf{a})/T$. If Brownian motion is assumed between these two locations, we have that $\mathbf{X}(t) \sim \mathcal{N}(\boldsymbol{\mu}(t), \sigma^2(t)I_n)$, where I_n is the identity matrix of size n . The PDF of $\mathbf{X}(t)$ can be written as

$$f_{\mathbf{X}}(\mathbf{x} | t) = \frac{1}{(2\pi\sigma^2(t))^{n/2}} \exp\left(-\frac{(\mathbf{x} - \boldsymbol{\mu}(t))^{\top}(\mathbf{x} - \boldsymbol{\mu}(t))}{2\sigma^2(t)}\right). \quad (4.2)$$

The above assumes that the different components of $\mathbf{X}(t)$ are mutually independent. Dependence can be accounted for by including a covariance matrix, denoted as $\Sigma(t)$, leading to

$$f_{\mathbf{X}}(\mathbf{x} | t) = \frac{1}{\sqrt{(2\pi)^n |\Sigma(t)|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}(t))^{\top}(\Sigma(t))^{-1}(\mathbf{x} - \boldsymbol{\mu}(t))\right), \quad (4.3)$$

where $|\Sigma(t)|$ is the determinant of $\Sigma(t)$ (Walck, 1996). In this case, we have that $\mathbf{X}(t) \sim \mathcal{N}(\boldsymbol{\mu}(t), \Sigma(t))$. Note that by putting $\Sigma(t) = \sigma^2(t)I_n$, Eq. (4.3) reduces to Eq. (4.2).

The expression given by Eq. (4.1), (4.2) or (4.3) describes the PDF of a particle's position at a specific moment in time. In order to compute the (time-independent)

marginal probability density at a position \mathbf{x} , the probability densities are averaged over time (Bullard, 1991):

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{T} \int_{t=0}^T f_{\mathbf{x}}(\mathbf{x} | t) dt. \quad (4.4)$$

This PDF, the PDF of the TIBB, describes the position \mathbf{X} of a particle or individual (Bullard, 1991). Considering an interval $[\mathbf{x}_1, \mathbf{x}_2]$ (in two dimensions, such an interval represents a rectangle), the integral of this PDF over this interval reflects the expected value of the proportion of time a particle resides in that interval (Horne et al., 2007). This proportion of time is referred to in literature using various names such as sojourn, local and occupation time (Borodin, 1989; Chi et al., 2015; Hooghiemstra, 2002). If we assume that $\Sigma(t) = \sigma^2(t)I_n$, the parameters that are needed to describe the PDF of the TIBB are \mathbf{a} , \mathbf{b} , T and σ_m^2 , such that we may introduce the notation $\mathbf{X} \sim \mathcal{B}(\mathbf{a}, \mathbf{b}, T, \sigma_m^2)$.

To numerically obtain the PDF of the TIBB, the midpoint integration method can be used to compute Eq. (4.4). In view of the singularities at $t = 0$ and $t = T$, numerical integration cannot be performed easily. However, by introducing a location error σ_e^2 , the singularities at $t = 0$ and $t = T$ vanish and the numerical integration can be performed (Bullard, 1991). Since our main focus is on the movement of the particle as such, we did not account explicitly for this location error. Moreover, the aforementioned singularities can be tackled in a rigorous way when deriving an analytical expression for Eq. (4.4).

To settle the mind, let us consider Fig. 4.1(a) which depicts the PDFs of a one-dimensional BB, starting at $X(0) = 0$ and ending at $X(10) = 10$, at $t = 2$, $t = 5$ and $t = 8$. Here, the discretization step in space, *i.e.* Δx , is 0.01. Figure 4.1(b) depicts the PDF of the TIBB obtained using the midpoint integration method with discretization step in time, *i.e.* $\Delta t = 0.0001$. The PDFs of a two-dimensional BB at $t = 2$, $t = 5$ and $t = 8$ are shown in Fig. 4.1(c) and the corresponding PDF of the TIBB is shown in Fig. 4.1(d). The two-dimensional Brownian bridge starts at $\mathbf{X}(0) = (0, 0)^T$ and ends at $\mathbf{X}(10) = (10, 10)^T$. Here, \mathbf{a}^T represents the transpose of a vector \mathbf{a} . Here, $\Delta x = 0.5$ and $\Delta t = 0.0001$. In both one and two dimensions, the location error is very small, namely 10^{-8} , $\sigma_m^2 = 1$ and the PDFs were obtained through using the midpoint integration method. As can be seen in Figs. 4.1(a) and (c), the mean of the BB shifts gradually from \mathbf{a} to \mathbf{b} , proportional with the amount of elapsed time. Furthermore, the variance is zero at $t = 0$ and steadily increases until $t = T/2$, where it reaches its maximum, after which it decreases and approaches zero again at $t = T$. These observations are quite intuitive since the locations at $t = 0$ and $t = T$ are known and estimated positions at times farther away from the observed ones, are less certain and have higher variances.

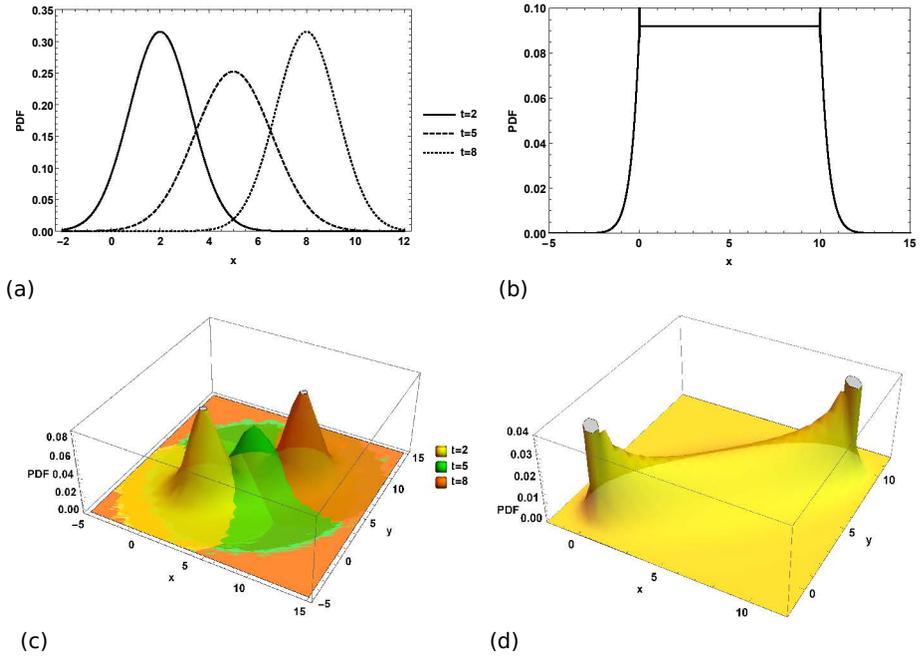


Figure 4.1: (a) The PDFs of a one-dimensional BB at $t = 2$, $t = 5$ and $t = 8$, together with (b) the numerically determined PDF of the TIBB. The bridge starts at $X(0) = 0$, ends at $X(10) = 10$ and has a mobility factor $\sigma_m^2 = 1$. (c) The PDFs of a two-dimensional Brownian bridge with $\mathbf{X}(0) = (0, 0)$, $\mathbf{X}(10) = (10, 10)$ and $\sigma_m^2 = 1$ and (d) the numerically determined PDF of the TIBB.

4.3 The one-dimensional TIBB

4.3.1 Its analytical description

To obtain an analytical description of the PDF of the one-dimensional TIBB, the factor $(2\pi\sigma_m^2 T)^{-1/2}$ in Eq. (4.4), where $f_{\mathbf{x}}(\mathbf{x} | t)$ is given by Eq. (4.1), is discarded during calculations for the sake of simplicity. Hence, in the one-dimensional setting, the integral to be computed is

$$I_{1D}(x) = \int_0^T \frac{1}{\sqrt{t(T-t)}} \exp\left(-\frac{(x-a-\frac{t}{T}(b-a))^2}{2\sigma_m^2 \frac{t(T-t)}{T}}\right) dt. \quad (4.5)$$

As a first step, we introduce $\alpha := (b-a)^2/(2\sigma_m^2 T)$, $\beta := (x-a)/(b-a)$ and $u := t/T$, so that Eq. (4.5) can be rewritten as:

$$I_{1D}(x) = \int_0^1 \frac{1}{\sqrt{u(1-u)}} \exp\left(-\alpha \frac{(\beta-u)^2}{u(1-u)}\right) du. \quad (4.6)$$

Introducing β implies that the starting and ending location correspond to $\beta = 0$ and $\beta = 1$, respectively, irrespective of whether $a < b$ or $a > b$. After replacing u by $z/(1+z)$, Eq. (4.6) becomes

$$I_{1D}(x) = \int_0^{+\infty} \frac{1}{\sqrt{z}(1+z)} \exp\left(-\frac{\alpha}{z}(\beta - (1-\beta)z)^2\right) dz. \quad (4.7)$$

In Fig. 4.2, the generic shape of a one-dimensional TIBB is shown, in which the symmetry at $\beta = 0.5$ and the two tails at both side of a plateau are notable. When deriving the analytical description, the plateau, *i.e.* the case $0 < \beta < 1$, is considered first. Subsequently, the cases $\beta < 0$ and $\beta > 1$ are studied and, finally, the probability densities at $\beta = 0$ and $\beta = 1$ (singularities) are determined.

4.3.1.1 The inner region: $0 < \beta < 1$

First, let us substitute z by $\beta v/(1-\beta)$:

$$I_{1D}(x) = \sqrt{\beta(1-\beta)} \int_0^{+\infty} \exp\left(\frac{-\alpha(1-\beta)\beta(1-v)^2}{v}\right) \frac{dv}{\sqrt{v}(1-\beta+\beta v)}. \quad (4.8)$$

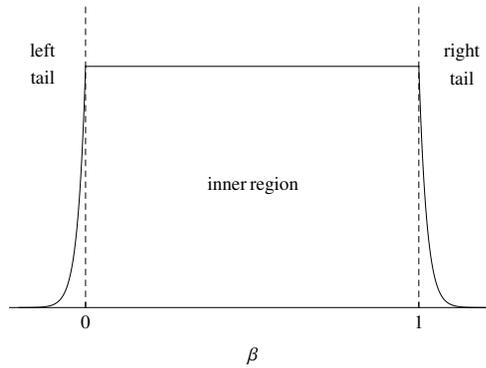


Figure 4.2: Indication of the regions of the PDF of the one-dimensional TIBB.

After substituting v by $\tan^2(\theta/2)$,

$$I_{1D}(x) = \sqrt{\beta(1-\beta)} \int_0^\pi \exp(-4\alpha(1-\beta)\beta \cot^2 \theta) \frac{d\theta}{(1-\beta) \cos^2 \frac{\theta}{2} + \beta \sin^2 \frac{\theta}{2}}, \quad (4.9)$$

is obtained. By making use of the half-angle formulas $\cos^2 \frac{\theta}{2} = \frac{1+\cos \theta}{2}$ and $\sin^2 \frac{\theta}{2} = \frac{1-\cos \theta}{2}$, we arrive at

$$I_{1D}(x) = 2 \sqrt{\beta(1-\beta)} \int_0^\pi \exp(-4\alpha(1-\beta)\beta \cot^2(\theta)) \frac{1}{1+(1-2\beta)\cos(\theta)} d\theta. \quad (4.10)$$

Splitting the integration interval into two parts (from 0 to $\pi/2$ and from $\pi/2$ to π) and relying on the symmetry properties of $\cos(\theta)$, *i.e.* $\cos(\pi-\theta) = -\cos \theta$, we find

$$I_{1D}(x) = 2 \sqrt{\beta(1-\beta)} \int_0^{\pi/2} \exp(-4\alpha(1-\beta)\beta \cot^2 \theta) \left(\frac{1}{1+(1-2\beta)\cos \theta} + \frac{1}{1-(1-2\beta)\cos \theta} \right) d\theta, \quad (4.11)$$

which can be reformulated as

$$I_{1D}(x) = 4 \sqrt{\beta(1-\beta)} \int_0^{\pi/2} \exp(-4\alpha(1-\beta)\beta \cot^2(\theta)) \frac{d\theta}{1-(1-2\beta)^2 \cos^2(\theta)}. \quad (4.12)$$

Now substituting $\cot(\theta)$ in Eq. (4.12) by w leads to

$$I_{1D}(x) = 4 \sqrt{\beta(1-\beta)} \int_0^{+\infty} \exp(-4\alpha(1-\beta)\beta w^2) \frac{1}{(1+w^2) \left(1 - (1-2\beta)^2 \frac{w^2}{1+w^2}\right)} dw. \quad (4.13)$$

Although Eq. (4.13) can be evaluated symbolically using mathematical software such as Mathematica (Version 9.0.1, Wolfram Research Inc., Champaign, US), the procedure leading to an analytical expression is shown for the sake of comprehensiveness. Equation (4.13) can be simplified considerably by substituting w by $\frac{s}{2\sqrt{\beta(1-\beta)}}$. Moreover, recasting the integrand by multiplying it with $\exp(\alpha)/\exp(\alpha)$ leads to:

$$I_{1D}(x) = 2 \exp(\alpha) \int_0^{+\infty} \exp(-\alpha(1+s^2)) \frac{1}{1+s^2} ds. \quad (4.14)$$

We will now show that Eq. (4.14) equals

$$I_{1D}(x) = \exp(\alpha)\pi(1 - \operatorname{erf}(\sqrt{\alpha})), \quad (4.15)$$

where erf is the error function (Abramowitz and Stegun, 1972). As a first step, $J(\mu)$ is introduced:

$$J(\mu) = 2 \int_0^{+\infty} \exp(-\mu(1+s^2)) \frac{1}{1+s^2} ds. \quad (4.16)$$

Obviously, it holds that

$$J(\alpha) = J(0) + \int_0^\alpha \frac{\partial J(\mu)}{\partial \mu} d\mu. \quad (4.17)$$

Here, $J(0)$ is determined first by substituting s by $\tan(\theta)$:

$$J(0) = 2 \int_0^{+\infty} \frac{1}{1+s^2} ds = 2 \int_0^{\frac{\pi}{2}} d\theta = \pi, \quad (4.18)$$

and $\frac{\partial J(\mu)}{\partial \mu}$ follows from:

$$\frac{\partial J(\mu)}{\partial \mu} = -2 \int_0^{+\infty} \exp(-\mu(1+s^2)) ds, \quad (4.19)$$

which becomes, after substituting $\sqrt{\mu}s$ by l ,

$$\frac{\partial J(\mu)}{\partial \mu} = -2 \frac{\exp(-\mu)}{\sqrt{\mu}} \int_0^{+\infty} \exp(-l^2) dl \quad (4.20)$$

$$= -\exp(-\mu) \frac{\sqrt{\pi}}{\sqrt{\mu}}. \quad (4.21)$$

$J(\alpha)$ then becomes

$$J(\alpha) = \pi + \sqrt{\pi} \int_0^{\alpha} -\exp(-\mu) \frac{1}{\sqrt{\mu}} d\mu. \quad (4.22)$$

After substituting μ by t^2 in Eq. (4.22), we get

$$J(\alpha) = \pi + \sqrt{\pi} \int_0^{\sqrt{\alpha}} -\exp(-t^2) 2dt \quad (4.23)$$

$$= \pi + \sqrt{\pi}(-\sqrt{\pi} \operatorname{erf}(\sqrt{\alpha})) \quad (4.24)$$

This expression can now be used in Eq. (4.14).

4.3.1.2 The tails: $\beta < 0$ and $1 < \beta$

Here, the right and left tail of the distribution are analytically derived. The two tails are discussed together since the derivation of their analytical description is similar. Taking into account that either $\beta < 0$ or $1 < \beta$, substituting z by $-\beta/(1-\beta)$, we get from Eq. (4.7):

$$I_{1D}(x) = \sqrt{|\beta(1-\beta)|} \int_0^{+\infty} \exp\left(\frac{-\alpha|(1-\beta)\beta|(1+v)^2}{v}\right) \frac{dv}{\sqrt{v(1-\beta-\beta v)}}. \quad (4.25)$$

Therein, v is replaced by $\tan^2(\theta/2)$, and recalling that $\cos^2 \frac{\theta}{2} = \frac{\cos\theta+1}{2}$, we get

$$I_{1D}(x) = 2\sqrt{|\beta(1-\beta)|} \int_0^{\pi} \exp\left(-4\alpha|(1-\beta)\beta| \frac{1}{\sin^2(\theta)}\right) \frac{1}{(1-2\beta)+\cos(\theta)} d\theta. \quad (4.26)$$

Similar to the transformation of Eq. (4.10) into Eq. (4.13), the integration interval is split, symmetry properties are invoked and $\cot(\theta)$ is substituted by w to arrive at

$$I_{1D}(x) = 4 \sqrt{|\beta(1-\beta)|} (1-2\beta) \int_0^{+\infty} \exp(-4\alpha|(1-\beta)\beta|(1+w^2)) \frac{1}{(1-2\beta)^2 + 4|(1-\beta)\beta|w^2} dw. \quad (4.27)$$

As a last step, w is replaced by $\frac{1-2\beta}{2\sqrt{|\beta(1-\beta)|}} s$, so that Eq. (4.27) simplifies to:

$$I_{1D}(x) = 2 \exp(-4\alpha|(1-\beta)\beta|) \int_0^{+\infty} \exp(-\alpha(1-2\beta)^2 s^2) \frac{1}{1+s^2} ds. \quad (4.28)$$

4 Following the same reasoning as the one used to compute the integral in Eq. (4.14), it can finally be shown that

$$I_{1D}(x) = \begin{cases} \pi \exp(\alpha) (1 - \operatorname{erf}((1-2\beta)\sqrt{\alpha})) & , \text{ if } \beta < 0, \\ \pi \exp(\alpha) (1 - \operatorname{erf}((2\beta-1)\sqrt{\alpha})) & , \text{ if } 1 < \beta. \end{cases} \quad (4.29)$$

4.3.1.3 The cases $\beta = 0$ and $\beta = 1$

The probability density of the TIBB at $\beta = 0$ and $\beta = 1$ is derived by considering the limits of Eqs. (4.15) and (4.29) at these points. To determine the probability density at $\beta = 0$, the limits

$$\lim_{\beta \rightarrow 0^-} \pi \exp(\alpha) (1 - \operatorname{erf}((1-2\beta)\sqrt{\alpha}))$$

and

$$\lim_{\beta \rightarrow 0^+} \pi \exp(\alpha) (1 - \operatorname{erf}(\sqrt{\alpha}))$$

are evaluated. Since both limits yield $\pi \exp(\alpha) (1 - \operatorname{erf}(\sqrt{\alpha}))$, the probability density at $\beta = 0$ is given by $\pi \exp(\alpha) (1 - \operatorname{erf}(\sqrt{\alpha}))$. Similarly, it can be shown that the probability density at $\beta = 1$ equals $\pi \exp(\alpha) (1 - \operatorname{erf}(\sqrt{\alpha}))$.

4.3.1.4 Normalization

Summarizing, the analytical description of the PDF of the TIBB, up to a constant factor, is thus given by:

$$I_{1D}(x) = \begin{cases} \pi \exp(\alpha) (1 - \operatorname{erf}((1 - 2\beta)\sqrt{\alpha})) & , \text{ if } \beta < 0, \\ \pi \exp(\alpha) (1 - \operatorname{erf}(\sqrt{\alpha})) & , \text{ if } 0 \leq \beta \leq 1, \\ \pi \exp(\alpha) (1 - \operatorname{erf}((2\beta - 1)\sqrt{\alpha})) & , \text{ if } 1 < \beta, \end{cases} \quad (4.30)$$

where $\beta = (x - a)/(b - a)$ and $\alpha = (b - a)^2/(2\sigma_m^2 T)$.

In order to comply with Kolmogorov's second axiom, the PDF given by Eq. (4.30) should be normalized such that $\int_{-\infty}^{+\infty} I_{1D}(x) dx = 1$ (Kolmogorov, 1956). Therefore, we first of all compute

$$C = \int_{-\infty}^{+\infty} I_{1D}(x) dx = (b - a) \int_{-\infty}^{+\infty} I_{1D}((b - a)\beta + a) d\beta, \quad (4.31)$$

or, equivalently, upon substituting the expression for $I_{1D}(x)$,

$$C = (b - a)\pi \exp(\alpha) \left(\int_{-\infty}^0 (1 - \operatorname{erf}((1 - 2\beta)\sqrt{\alpha})) d\beta + \int_0^1 (1 - \operatorname{erf}(\sqrt{\alpha})) d\beta + \int_1^{+\infty} (1 - \operatorname{erf}((2\beta - 1)\sqrt{\alpha})) d\beta \right) \quad (4.32)$$

As a first step, the first integral (T_1) in Eq. (4.32) is considered. Therein, $(1 - 2\beta)\sqrt{\alpha}$ is replaced by γ , so that

$$T_1 = \int_{\sqrt{\alpha}}^{+\infty} (1 - \operatorname{erf}(\gamma)) \frac{1}{2\sqrt{\alpha}} d\gamma, \quad (4.33)$$

which leads to (Noog and Geller, 1969):

$$T_1 = \frac{1}{2\sqrt{\pi\alpha}} \exp(-\alpha) - \frac{1}{2} (1 - \operatorname{erf}(\sqrt{\alpha})) \quad (4.34)$$

upon evaluating the upper and lower limits of integration. The second integral (T2) in Eq. (4.32) can be computed easily, and the third integral (T3) in Eq. (4.32) corresponds to the right-hand side of Eq. (4.34) since the PDF of the one-dimensional TIBB has an axis of symmetry at $(a + b)/2$ and as such the tails have the same

shape. Equation (4.32) then finally becomes

$$C = (b - a) \frac{\sqrt{\pi}}{\sqrt{\alpha}}. \quad (4.35)$$

Its reciprocal value can now be used to normalize and hence, the PDF of the TIBB reads explicitly:

$$f_X(x) = \sqrt{\frac{\pi}{2\sigma_m^2 T}} \exp\left(\frac{(b-a)^2}{2\sigma_m^2 T}\right) \left(1 - \operatorname{erf}\left(\sqrt{\frac{1}{2\sigma_m^2 T}} (|x-a| + |b-x|)\right)\right). \quad (4.36)$$

The proof of symmetry and the distribution parameters for the PDF of the one-dimensional TIBB can be found in Sections 4.3.2 and 4.3.3, respectively. In Table 4.1, an overview of these parameters is given. In statistical literature, the PDF of a time-integrated stochastic process is also referred to as the local time of this process (Borodin, 1989; Pitman, 1998). It should be mentioned that the analytical expression of the local time of the one-dimensional BB in case $X(0) = 0$, $X(1) = b$ and $\sigma_m^2 = 1$ has been mentioned before, although obtained in a different way (Borodin, 1989; Pitman, 1998). To the best of our knowledge, the expression has not yet been reported in animal movement research. Furthermore, the analytical expression of the PDF of the general one- (although obvious) and two-dimensional TIBB is not yet known. However, such an analytical expression for two- or higher dimensional cases would certainly be beneficial for animal movement researchers in order to speed up the calculations and render the results more accurate. Furthermore, we never came across the term 'local time' in ecological literature and researchers are probably not aware of the accomplishments that are present in the 'statistical' field. In Section 4.4, the expression for the PDF of the two-dimensional TIBB is derived.

Table 4.1: An overview of the distribution parameters of the PDF of the one-dimensional TIBB, where $\alpha = (b - a)^2 / (2\sigma_m^2 T)$

| Parameter | Formula |
|-----------|--|
| Mean | $\frac{a+b}{2}$ |
| Median | $\frac{a+b}{2}$ |
| Mode | $[a, b]$ |
| Variance | $(b-a)^2 \frac{1+\alpha}{12\alpha}$ |
| Skewness | 0 |
| Kurtosis | $-\frac{6}{5} + \frac{9}{5(1+\alpha)^2}$ |

4.3.2 Proof of symmetry

Because the distribution of the one-dimensional TIBB has two tails with a similar shape at both sides of a plateau, it is expected that the PDF has $x = (a+b)/2$ as an axis of symmetry. Since the PDF $f_X(x)$ is symmetrical if and only if there exists a value x_0 such that $f_X(x_0 - \delta) = f_X(x_0 + \delta)$ for $\delta \in \mathbb{R}$, we first substitute x by $x^* + \frac{a+b}{2}$ to show our PDF's symmetry. Since $\frac{dx}{dx^*} = 1$, the PDF hereby becomes

$$f_X(x) = \sqrt{\frac{\pi}{2\sigma_m^2 T}} \exp\left(\frac{(b-a)^2}{2\sigma_m^2 T}\right) \left(1 - \operatorname{erf}\left(\sqrt{\frac{1}{2\sigma_m^2 T}} \left(\left|x^* + \frac{b-a}{2}\right| + \left|-x^* + \frac{b-a}{2}\right|\right)\right)\right). \quad (4.37)$$

It can now be seen easily that $I(-\delta) = I(\delta)$, and consequently that $x^* = 0$ is the symmetry axis of the distribution, from which it automatically follows that $x = (a+b)/2$ is the axis of symmetry of $f_X(x)$ in the original coordinate system.

4.3.3 Distribution parameters

4.3.3.1 Expected value

In the case of a symmetrical distribution, the expected value, $E[X]$, equals the x value where symmetry occurs. As such, $E[X] = \frac{a+b}{2}$.

The expected value can also be derived analytically:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx, \quad (4.38)$$

First, α and β are again introduced so that Eq. (4.30), normalized using the expression given by Eq. (4.35), can be used:

$$E[\beta] = \sqrt{\pi\alpha} \exp(\alpha) \left(\int_{-\infty}^0 \beta (1 - \operatorname{erf}((1-2\beta)\sqrt{\alpha})) d\beta + \int_0^1 \beta (1 - \operatorname{erf}(\sqrt{\alpha})) d\beta + \int_1^{\infty} \beta (1 - \operatorname{erf}((2\beta-1)\sqrt{\alpha})) d\beta \right). \quad (4.39)$$

In the first term (T_1) of this expression, $(1 - 2\beta)\sqrt{\alpha}$ is replaced by γ :

$$T_1 = \int_{-\infty}^0 \beta - \beta \operatorname{erf}((1 - 2\beta)\sqrt{\alpha}) d\beta \quad (4.40)$$

$$= \int_{\sqrt{\alpha}}^{\infty} \left(\frac{1}{4\sqrt{\alpha}} - \frac{\gamma}{4\alpha} - \frac{\operatorname{erf}(\gamma)}{4\sqrt{\alpha}} + \frac{\gamma \operatorname{erf}(\gamma)}{4\alpha} \right) d\gamma, \quad (4.41)$$

which corresponds to (Noog and Geller, 1969):

$$T_1 = \left(\frac{\gamma}{4\sqrt{\alpha}} - \frac{\gamma^2}{8\alpha} - \operatorname{erf}(\gamma) \left(\frac{\gamma}{4\sqrt{\alpha}} - \frac{\gamma^2}{8\alpha} + \frac{1}{16\alpha} \right) + \frac{\exp(-\gamma^2)}{4\sqrt{\pi}\sqrt{\alpha}} \left(\frac{\gamma}{2\sqrt{\alpha}} - 1 \right) \right) \Bigg|_{\gamma=\sqrt{\alpha}}^{\gamma=+\infty}. \quad (4.42)$$

Since $\lim_{\gamma \rightarrow \infty} \operatorname{erf}(\infty) = 1$, T_1 is

$$T_1 = \frac{-1}{16\alpha} - \frac{1}{8} + \operatorname{erf}(\sqrt{\alpha}) \left(\frac{1}{8} + \frac{1}{16\alpha} \right) + \frac{\exp(-\alpha)}{8\sqrt{\pi\alpha}}, \quad (4.43)$$

The second integral (T_2) in Eq. (4.39) can be computed easily:

$$T_2 = \left(\frac{\beta^2}{2} (1 - \operatorname{erf}(\sqrt{\alpha})) \right) \Bigg|_{\beta=0}^{\beta=1} \quad (4.44)$$

$$= \frac{1}{2} - \frac{\operatorname{erf}(\sqrt{\alpha})}{2}, \quad (4.45)$$

and the third integral (T_3) in Eq. (4.39) is computed in a similar way as T_1 . For that purpose, $(2\beta - 1)\sqrt{\alpha}$ is first substituted by γ ,

$$T_3 = \int_1^{\infty} \beta [1 - \operatorname{erf}((2\beta - 1)\sqrt{\alpha})] d\beta \quad (4.46)$$

$$= \int_{\sqrt{\alpha}}^{\infty} \left(\frac{1}{4\sqrt{\alpha}} + \frac{\gamma}{4\alpha} - \frac{\operatorname{erf}(\gamma)}{4\sqrt{\alpha}} - \frac{\gamma \operatorname{erf}(\gamma)}{4\alpha} \right) d\gamma, \quad (4.47)$$

which equals

$$T_3 = \left(\frac{\gamma}{4\sqrt{\alpha}} + \frac{\gamma^2}{8\alpha} - \operatorname{erf}(\gamma) \left(\frac{\gamma}{4\sqrt{\alpha}} + \frac{\gamma^2}{8\alpha} - \frac{1}{16\alpha} \right) - \frac{\exp(-\gamma^2)}{4\sqrt{\pi}\sqrt{\alpha}} \left(1 + \frac{\gamma}{2\sqrt{\alpha}} \right) \right) \Bigg|_{\gamma=\sqrt{\alpha}}^{\gamma=+\infty}, \quad (4.48)$$

or upon evaluating the limits of integration:

$$T_3 = \frac{1}{16\alpha} - \frac{3}{8} + \operatorname{erf}(\sqrt{\alpha}) \left(\frac{3}{8} - \frac{1}{16\alpha} \right) + \frac{3 \exp(-\alpha)}{8\sqrt{\pi}\sqrt{\alpha}}. \quad (4.49)$$

The expected value then becomes:

$$E[\beta] = \sqrt{\pi\alpha} \exp(\alpha) \left(\frac{-1}{16\alpha} - \frac{1}{8} + \operatorname{erf}(\sqrt{\alpha}) \left(\frac{1}{8} + \frac{1}{16\alpha} \right) - \frac{\exp(-\alpha)}{8\sqrt{\pi}\sqrt{\alpha}} + \frac{1}{2} - \frac{\operatorname{erf}(\sqrt{\alpha})}{2} + \frac{1}{16\alpha} - \frac{3}{8} + \operatorname{erf}(\sqrt{\alpha}) \left(\frac{3}{8} - \frac{1}{16\alpha} \right) + \frac{3 \exp(-\alpha)}{8\sqrt{\pi}\sqrt{\alpha}} \right), \quad (4.50)$$

which equals

$$E[\beta] = \frac{1}{2}, \quad (4.51)$$

so that

$$E[X] = \frac{a+b}{2}. \quad (4.52)$$

4.3.3.2 Median and mode

The median also equals the point where symmetry occurs, *i.e.* $X = \frac{a+b}{2}$.

Furthermore, it can be seen from Eq. (4.36) that $f_X(x)$ reaches its maximum, *i.e.* $\frac{\sqrt{\pi\alpha}}{(b-a)} \exp(\alpha) \left(1 - \operatorname{erf} \left(\frac{(b-a)}{\sqrt{2\sigma_m^2 T}} \right) \right)$, if $x \in [a, b]$, the mode of the distribution.

4.3.3.3 Variance

To derive the variance of X , *i.e.* $\operatorname{Var}(X)$, α and β are again introduced and Eq. (4.30), normalized with respect to Eq. (4.35), is used:

$$\operatorname{Var}(\beta) = E[\beta^2] - (E[\beta])^2 = \int_{-\infty}^{+\infty} \beta^2 I_{1D}(x) d\beta - \left(\frac{1}{2} \right)^2. \quad (4.53)$$

Following the same procedure as the one for deriving $E[\beta]$, we get

$$\text{Var}(\beta) = \frac{1 + 4\alpha}{12\alpha} - \frac{1}{4}, \quad (4.54)$$

so in terms of the original variable:

$$\text{Var}(X) = (b - a)^2 \frac{1 + \alpha}{12\alpha}. \quad (4.55)$$

4.3.3.4 Skewness

The skewness of $f_X(x)$ is zero because $f_X(x)$ is a symmetric function. Nevertheless, it can also be calculated by computing the third standardized moment:

$$S[X] = E \left[\left(\frac{X - E[X]}{\sigma} \right)^3 \right]. \quad (4.56)$$

As a first step, α and β are again introduced and the skewness is written as a function of the non-central third moment $E[\beta^3]$:

$$S[\beta] = \frac{E[\beta^3] - E[\beta]^3 - 3E[\beta]\text{Var}(\beta)}{\text{Var}(\beta)^{\frac{3}{2}}}, \quad (4.57)$$

Here, $E[\beta^3]$ is the only unknown term which can be calculated in a similar way as the first and second moment. As such,

$$E[\beta^3] = \frac{1}{4} + \frac{1}{8\alpha}, \quad (4.58)$$

and the skewness is:

$$S[\beta] = \frac{\frac{1}{4} + \frac{1}{8\alpha} - \left(\frac{1}{2}\right)^3 - 3\left(\frac{1}{2}\right)\left(\frac{1+\alpha}{12\alpha}\right)}{\left(\frac{1+\alpha}{12\alpha}\right)^{\frac{3}{2}}}, \quad (4.59)$$

which results, as expected and mentioned before, into zero. Similar to the procedure followed to describe the effect of a linear transformation on the variance (Veerarajan, 2008), it has been proven that such a transformation does not have an impact on the skewness, and as such, that $S[\beta] = S[X]$.

4.3.3.5 Kurtosis

The kurtosis is calculated using the fourth standardized moment:

$$K[\beta] = E \left[\left(\frac{\beta - E[\beta]}{\sigma_\beta} \right)^4 \right] - 3 = \frac{E[\beta^4] - 4E[\beta]E[\beta^3] + 6E[\beta]^2E[\beta^2] - 3E[\beta]^4}{\text{Var}(\beta)^2} - 3. \quad (4.60)$$

The minus 3 at the end of this formula is often introduced as a correction term to ensure that the kurtosis of an univariate normal distribution is zero (Westfall, 2014). The only unknown term is the fourth moment $E[\beta^4]$ which can be calculated in a similar way as the second and third moments:

$$E[\beta^4] = \frac{1 + 6\alpha + 8\alpha^2}{40\alpha^2}. \quad (4.61)$$

Consequently, the kurtosis is given by

$$K[\beta] = -\frac{6}{5} + \frac{9}{5(1+\alpha)^2}. \quad (4.62)$$

Since it has been shown that a linear transformation does not have an impact on the kurtosis (Veerarajan, 2008), we finally obtain in terms of the original coordinate system $K[\beta] = K[X]$.

4.4 The two-dimensional TIBB

4.4.1 No dependence between both components, a starting location situated at the origin and an ending location situated along the horizontal axis

Having derived the analytical description of the PDF of the one-dimensional TIBB, the derivation of its two-dimensional counterpart is considered in this section. First, a direction-independent mobility factor, is assumed. For the sake of simplicity, the factor $(2\pi\sigma_m^2 T)^{-1}$ in Eq. (4.4), where $f_{\mathbf{x}}(\mathbf{x} | t)$ corresponds to Eq. (4.2), is omitted from the presented computations, the starting location \mathbf{a} is assumed to be located at the origin, while the ending location \mathbf{b} is situated along the horizontal axis, so that $\mathbf{a} = (0, 0)^T$ and $\mathbf{b} = (b_x, 0)^T$. Under these assumptions, the integral to be evaluated is

$$I_{2D}(x, y) = \int_0^T \frac{T}{t(T-t)} \exp\left(-\frac{(\mathbf{x} - \frac{t}{T}\mathbf{b})^2}{2\sigma_m^2 \frac{t(T-t)}{T}}\right) dt, \quad (4.63)$$

where $\mathbf{x}(t) = (x, y)^T$. This expression becomes

$$I_{2D}(x, y) = \int_0^1 \frac{1}{u(1-u)} \exp\left(-\frac{\alpha((\beta-u)^2 + \gamma^2)}{u(1-u)}\right) du, \quad (4.64)$$

after using $\alpha := b_x^2/(2T\sigma_m^2)$, $\beta := x/b_x$, $\gamma := y/b_x$ and $u := t/T$. Further, substituting u by $z/(1+z)$ gives rise to:

$$I_{2D}(x, y) = \int_0^{+\infty} \frac{1}{z} \exp\left(-\frac{\alpha((\beta - (1-\beta)z)^2 + (\gamma(1+z))^2)}{z}\right) dz. \quad (4.65)$$

Using formula 8.432.7 in Gradshteyn and Ryzhik (2007), it can be verified that this integral collapses to:

$$I_{2D}(x, y) = 2 \exp(-2\alpha(-\beta + \beta^2 + \gamma^2)) K_0\left(2\sqrt{\alpha(\beta^2 + \gamma^2)}\sqrt{\alpha(1 - 2\beta + \beta^2 + \gamma^2)}\right), \quad (4.66)$$

where K_0 is the modified Bessel function of the second kind of order zero. Since $\lim_{k \rightarrow 0} K_0(k) = +\infty$, the probability density approaches $+\infty$ as $(x, y) = (0, 0)$ or

$(x, y) = (b_x, 0)$. This integral can also be computed using mathematical software such as Mathematica.

4.4.1.1 Normalization

In order to retrieve the factor that enables a normalization of Eq. (4.66), which is needed to comply with Kolmogorov's second axiom, we integrate Eq. (4.63) over \mathbb{R}^2 , i.e.

$$C = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} I_{2D}(x, y) dx dy = b_x^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} I_{2D}(b_x \beta, b_x \gamma) d\gamma d\beta. \quad (4.67)$$

This corresponds to

$$C = b_x^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_0^1 \frac{1}{u(1-u)} \exp\left(\frac{-\alpha((\beta-u)^2 + \gamma^2)}{u(1-u)}\right) du d\gamma d\beta, \quad (4.68)$$

which may be recasted into

$$C = b_x^2 \int_0^1 \frac{1}{u(1-u)} \left[\int_{-\infty}^{+\infty} \exp\left(\frac{-\alpha(\beta-u)^2}{u(1-u)}\right) d\beta \right] \left[\int_{-\infty}^{+\infty} \exp\left(\frac{-\alpha\gamma^2}{u(1-u)}\right) d\gamma \right] du. \quad (4.69)$$

In the first inner integral, $\frac{\sqrt{\alpha}(\beta-u)}{\sqrt{u(1-u)}}$ is substituted by t_1 , while in the second one $\frac{\sqrt{\alpha}\gamma}{\sqrt{u(1-u)}}$ is replaced by t_2 . Eq. (4.69) then becomes

$$C = b_x^2 \int_0^1 \frac{1}{u(1-u)} \left[\int_{-\infty}^{+\infty} \exp(-t_1^2) \frac{\sqrt{u(1-u)} dt_1}{\sqrt{\alpha}} \right] \left[\int_{-\infty}^{+\infty} \exp(-t_2^2) \frac{\sqrt{u(1-u)} dt_2}{\sqrt{\alpha}} \right] du. \quad (4.70)$$

Since $\int_{-\infty}^{+\infty} \exp(-t^2) dt = \sqrt{\pi}$, we finally arrive at

$$C = b_x^2 \frac{\pi}{\alpha}. \quad (4.71)$$

Hence, the PDF of the two-dimensional TIBB is

$$f_{\mathbf{X}}(x, y) = \frac{2\alpha}{\pi b_x^2} \exp(-2\alpha(-\beta + \beta^2 + \gamma^2)) K_0\left(2\sqrt{\alpha(\beta^2 + \gamma^2)}\sqrt{\alpha(1 - 2\beta + \beta^2 + \gamma^2)}\right), \quad (4.72)$$

where $\beta = x/b_x$ and $\gamma = y/b_x$, or in terms of the original coordinate system:

$$f_{\mathbf{X}}(x, y) = \frac{1}{\pi \sigma_m^2 T} \exp\left(-\frac{1}{\sigma_m^2 T}(-x b_x + x^2 + y^2)\right) K_0\left(\frac{1}{\sigma_m^2 T} \sqrt{x^2 + y^2} \sqrt{b_x^2 - 2x b_x + x^2 + y^2}\right). \quad (4.73)$$

4.4.1.2 Marginal distributions

To conclude this section, we would like to point out one noteworthy property of the PDF of the two-dimensional TIBB. More specifically, the marginal distribution of the x -component, respectively the y -component of the random variable \mathbf{X} corresponds to the PDF of the one-dimensional TIBB, *i.e.* $\mathcal{B}(0, b_x, T, \sigma_m^2)$, respectively $\mathcal{B}(0, 0, T, \sigma_m^2)$ (Chi et al., 2015).

In order to compute a marginal distribution, Eq. (4.64), normalized using the expression given by Eq. (4.71), is resumed:

$$I_{2D} = \frac{\alpha}{b_x^2 \pi} \int_0^1 \frac{1}{u(1-u)} \exp\left(\frac{-\alpha((\beta - u)^2 + \gamma^2)}{u(1-u)}\right) du. \quad (4.74)$$

This expression is integrated over the direction perpendicular to the direction of interest. The marginal distribution in the x -direction is considered first after which the marginal distribution in the y -direction is derived. The former is:

$$I_{2D_x} = b_x \int_{-\infty}^{+\infty} I_{2D} d\gamma, \quad (4.75)$$

which equals

$$I_{2D_x} = \frac{\alpha}{b_x \pi} \left(\int_0^1 \frac{1}{u(1-u)} \exp\left(-\frac{\alpha((\beta - u)^2)}{u(1-u)}\right) \left(\int_{-\infty}^{+\infty} \exp\left(-\frac{\alpha\gamma^2}{u(1-u)}\right) d\gamma \right) du \right). \quad (4.76)$$

The inner integral was already computed (cfr. Eqs. (4.69)-(4.71)) so this expression becomes

$$I_{2D_x} = \frac{\sqrt{\alpha}}{b_x \sqrt{\pi}} \int_0^1 \frac{1}{\sqrt{u(1-u)}} \exp\left(-\frac{\alpha((\beta-u)^2)}{u(1-u)}\right) du. \quad (4.77)$$

The obtained integral is nothing else than the one-dimensional TIBB (cfr. Eq. (4.6)).

Now, the marginal distribution in the y -direction is considered.

$$I_{2D_y} = b_x \int_{-\infty}^{+\infty} I_{2D} d\beta, \quad (4.78)$$

which equals

$$I_{2D_y} = \frac{\alpha}{b_x \pi} \left(\int_0^1 \frac{1}{u(1-u)} \exp\left(-\frac{\alpha\gamma^2}{u(1-u)}\right) \left(\int_{-\infty}^{+\infty} \exp\left(\frac{-\alpha(\beta-u)^2}{u(1-u)}\right) d\beta \right) du \right). \quad (4.79)$$

The inner integral was computed before (cfr. Eqs. (4.69)-(4.71)), so we get

$$I_{2D_y} = \frac{\sqrt{\alpha}}{b_x \sqrt{\pi}} \int_0^1 \frac{1}{\sqrt{u(1-u)}} \exp\left(\frac{-\alpha\gamma^2}{u(1-u)}\right) du. \quad (4.80)$$

To compute the remaining integral, we use $u = \sin^2 \theta$:

$$I_{2D_y} = \frac{\sqrt{\alpha}}{b_x \sqrt{\pi}} \int_0^{\frac{\pi}{2}} \exp\left(\frac{-4\alpha\gamma^2}{\sin^2(2\theta)}\right) \frac{2 \sin \theta \cos \theta}{\sin \theta \cos \theta} d\theta. \quad (4.81)$$

which eventually yields

$$I_{2D_y} = \frac{\sqrt{\alpha \pi}}{b_x} (1 - \operatorname{erf}(2\sqrt{\alpha}|\gamma|)). \quad (4.82)$$

To settle the mind, the marginal distributions in the x - and y -direction of the two-dimensional TIBB with starting location $(x, y) = (0, 0)$ at time $t = 0$, ending location $(x, y) = (10, 0)$ at time $t = 10$ and $\sigma_m^2 = 1$ are shown in Fig. 4.3.

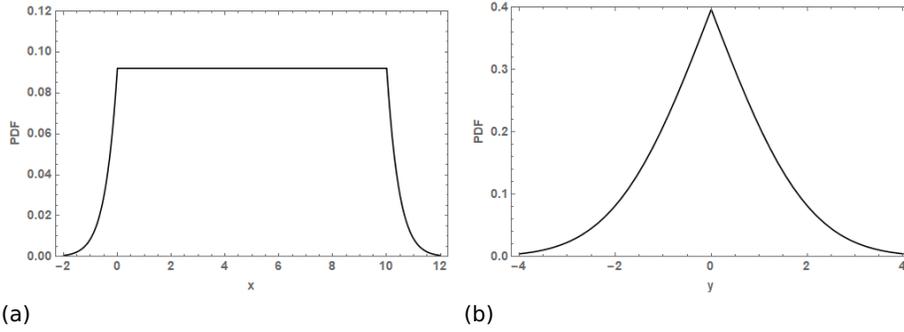


Figure 4.3: The marginal distribution in (a) the x - and (b) y -direction of the 2D TIBB with starting location $\mathbf{a} = (0, 0)$ at time $t = 0$, ending location $\mathbf{b} = (10, 0)$ at time $t = 10$ and $\sigma_m^2 = 1$.

4.4.2 No dependence between both components and a starting location situated at the origin

If the ending location is not situated along the horizontal axis, the coordinate system should be rotated in such a way that movement occurs along the horizontal axis of the rotated coordinate system, such that Eqs. (4.72) and (4.73) can again be invoked to compute the PDF of the TIBB. The original coordinate system is referred to as the xy -system, whereas the rotated system is referred to as the $x'y'$ -system. Furthermore, the coordinates of the ending location in the xy -system are $\mathbf{X}(T) = \mathbf{b} = (b_x, b_y)^T$ (Fig. 4.4). In the $x'y'$ -system, the starting location is also situated at the origin and the ending location is situated at a distance b_r measured along the x' -axis. In this way, the movement in a general direction may be conceived as movement along the horizontal axis of an appropriately rotated coordinate system. Consequently, Eq. (4.73) can be used to determine the PDF of the two-dimensional TIBB after replacing x by x' , y by y' , b_x by b_r and α by $\alpha' = b_r^2 / (2T\sigma_m^2)$. As can be inferred from Fig. 4.4, the line segment with length b_r is the hypotenuse in the triangle $\widehat{b_x O b_y}$ with opposite b_y and adjacent b_x . Here, the length b_r can be calculated using the Pythagorean theorem, *i.e.* $b_r = \sqrt{b_x^2 + b_y^2}$.

Transforming a point in the rotated coordinate system to a point in the original coordinate system can be achieved by considering the following rotation matrix:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad (4.83)$$

where θ is the angle of rotation, which follows easily from $\theta = \arctan2(b_y/b_x)$. Substituting the expressions for α' , x' and y' as given by Eq. (4.83) into Eq. (4.73), leads to an analytical expression of the PDF of the TIBB with a starting location

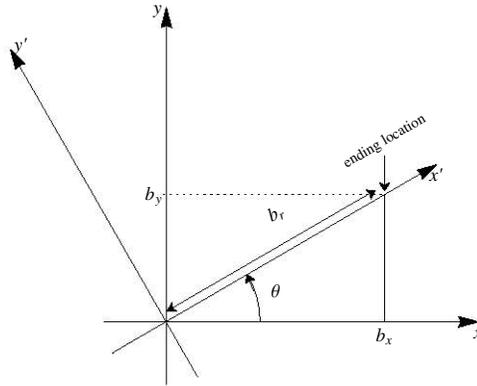


Figure 4.4: The coordinate system (with x - and y -axis) is shown together with the rotated coordinate system (with x' - and y' -axis). In the rotated system, $\mathbf{x}(0)$ is situated at the origin and $\mathbf{x}(T)$ is situated at a distance b_r along the x' -axis.

situated at the origin and an arbitrary ending location:

$$f_{\mathbf{x}}(x, y) = \frac{1}{\pi T \sigma_m^2} \exp \left(-\frac{1}{T \sigma_m^2} (-(\cos(\theta)x + \sin(\theta)y) b_r + (\cos(\theta)x + \sin(\theta)y)^2 + (\cos(\theta)y - \sin(\theta)x)^2) \right) K_0 \left(\frac{1}{\sigma_m^2 T} \sqrt{R} \right), \quad (4.84)$$

with

$$R = ((\cos(\theta)x + \sin(\theta)y)^2 + (\cos(\theta)y - \sin(\theta)x)^2) (b_r^2 - 2 b_r (\cos(\theta)x + \sin(\theta)y) + (\cos(\theta)x + \sin(\theta)y)^2 + (\cos(\theta)y - \sin(\theta)x)^2). \quad (4.85)$$

4.4.3 The two-dimensional TIBB without dependence between both components

The case where the starting and ending location are arbitrary, is the most general one and involves the expression that is of particular use in animal movement research. Here, a rotation and a translation of the coordinate system (axes x and y) is needed to ensure that the movement starts at the origin of the transformed coordinate system (axes x' and y') and occurs along the transformed x -axis. In this way, Eq. (4.73) can again be used to determine the PDF of the TIBB. The coordinates of an arbitrary starting location in the xy -system are $\mathbf{X}(0) = \mathbf{a} = (a_x, a_y)^T$ and that of the ending location $\mathbf{X}(T) = \mathbf{b} = (b_x, b_y)^T$. In Eq. (4.73) x is replaced

by x' , y by y' , b_x by b_r and α by $\alpha' = b_r^2/(2\sigma_m^2 T)$. In this case, b_r is again the distance between the starting and ending location: $b_r = \sqrt{(b_x - a_x)^2 + (b_y - a_y)^2}$.

Transforming a point to the original coordinate system can be achieved by considering the following rotation matrix:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x - a_x \\ y - a_y \end{pmatrix}, \quad (4.86)$$

where θ is the angle of rotation, *i.e.* $\arctan2((b_y - a_y), (b_x - a_x))$. Substituting the expressions for x' and y' as given by Eq. (4.86) into Eq. (4.73), results in an analytical expression of the PDF of the TIBB with an arbitrarily situated starting and ending location.

4.4.4 The two-dimensional TIBB with dependence between both components

In this section, the mobility factor is assumed to depend on the direction of movement, such that both spatial components are mutually dependent. For the sake of simplicity, the starting location is assumed to be located at the origin and the ending location is located along the x -axis, *i.e.* $\mathbf{X}(T) = (b_x, 0)^T$. The mobility factor in the x -direction is denoted by σ_x^2 and the one in the y -direction by σ_y^2 . In order to derive an analytical description of the PDF of the TIBB, let us recall the general formulation of a two-dimensional Brownian bridge containing the covariance matrix $\Sigma(t)$. In the two-dimensional case, Eq. (4.3) becomes

$$f_{\mathbf{X}}(x, y | t) = \frac{1}{2\pi\sqrt{|\Sigma(t)|}} \exp\left(-\frac{1}{2}(x - \mu, y)(\Sigma(t))^{-1}(x - \mu, y)^T\right), \quad (4.87)$$

where

$$\Sigma(t) = \frac{t(T-t)}{T} \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}, \quad (4.88)$$

$$|\Sigma(t)| = \left(\frac{t(T-t)}{T}\right)^2 \sigma_x^2 \sigma_y^2 (1 - \rho^2), \quad (4.89)$$

and

$$(\Sigma(t))^{-1} = \frac{1}{\left(\frac{t(T-t)}{T}\right) \sigma_x^2 \sigma_y^2 (1 - \rho^2)} \begin{pmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{pmatrix}. \quad (4.90)$$

Substituting Eqs. (4.89) and (4.90) into Eq. (4.87) leads to

$$f_{\mathbf{x}}(x, y | t) = \frac{1}{2\pi \frac{t(T-t)}{T} \sigma_x \sigma_y \sqrt{1-\rho^2}} \exp \left(\frac{-1}{2 \frac{t(T-t)}{T} \sigma_x^2 \sigma_y^2 (1-\rho^2)} \left(\left(x - \frac{t}{T} b_x \right)^2 \sigma_y^2 - 2 \left(x - \frac{t}{T} b_x \right) \rho \sigma_x \sigma_y + y^2 \sigma_x^2 \right) \right). \quad (4.91)$$

In order to calculate the PDF of the TIBB, the factor $(2\pi T \sigma_x \sigma_y \sqrt{1-\rho^2})^{-1}$ is discarded in further calculations, $u := t/T$ is introduced and the factors b_x^2 and σ_y^2 are introduced in the common numerator of the argument of the exponential function:

$$I_{2D}(x, y) = \int_0^1 \frac{1}{u(1-u)} \exp \left(\frac{-\sigma_y^2 b_x^2}{2T u(1-u) \sigma_x^2 \sigma_y^2 (1-\rho^2)} \left(\left(\left(\frac{x}{b_x} - u \right) - \rho \frac{\sigma_x}{\sigma_y} \frac{y}{b_x} \right)^2 + \frac{y^2 \sigma_x^2}{b_x^2 \sigma_y^2} (1-\rho^2) \right) \right) du. \quad (4.92)$$

After introducing $\beta := \frac{x}{b_x} - \rho \frac{\sigma_x}{\sigma_y} \frac{y}{b_x}$, $\gamma := \frac{y \sigma_x}{b_x \sigma_y} \sqrt{1-\rho^2}$ and $\alpha := \frac{b_x^2}{2T \sigma_x^2 (1-\rho^2)}$, this expression becomes

$$I_{2D}(x, y) = \int_0^1 \frac{1}{u(1-u)} \exp \left(\frac{-\alpha}{u(1-u)} ((\beta - u)^2 + \gamma^2) \right) du. \quad (4.93)$$

This integral is structurally similar to the one encountered in the case of uncorrelated movement (cfr. Eq. (4.64)), though α , β and γ are defined differently. The PDF of the two-dimensional TIBB thus becomes

$$f_{\mathbf{x}}(x, y) = \frac{2\alpha \sigma_x \sqrt{1-\rho^2}}{\pi b_x^2 \sigma_y} \exp(-2\alpha(-\beta + \beta^2 + \gamma^2)) K_0 \left(2\sqrt{\alpha(\beta^2 + \gamma^2)} \sqrt{\alpha(1 - 2\beta + \beta^2 + \gamma^2)} \right). \quad (4.94)$$

Since the expression is the same as in the case of uncorrelated movement, the expressions describing the PDF in the case without dependence between both spatial components can be used immediately when the spatial components are dependent, provided the proper expressions for α , β and γ are used. The case where the starting and ending locations are arbitrary can be handled analogously as in the case without dependence between the spatial components. Here, the mobility factor will be determined in the direction of movement and in the direction perpendicular to it. The mobility factor was already split in two directional components in Kranstauber et al. (2014). As such, our analytical expression may

again be useful to speed up the calculations and increase the accuracy of the PDF of the TIBB.

4.5 The analytical description of the n -dimensional TIBB

In order to complete the analytical description of the PDF of the TIBB, higher dimensional spaces are considered. For the sake of simplicity, it is assumed that the starting location is situated at the origin, the ending location is situated at a distance b along the x_1 -axis and the mobility factor is the same in all directions. Other cases will not be considered, but their derivation is similar to the one for the two-dimensional case. Let us first recall Eq. (4.2), which describes an n -dimensional BB in an appropriate coordinate system (with axes x_1, x_2, \dots, x_n). Discarding the factor $(2\pi\sigma_m^2)^{-n/2}$ during our calculations, the integral leading to the PDF of the TIBB, up to a constant factor, becomes

$$I_{nD}(\mathbf{x}) = \int_0^T \frac{1}{\sqrt{(t(T-t))^n}} \exp\left(\frac{-1}{2\frac{t(T-t)}{T}\sigma_m^2} \left(\left(x_1 - \frac{t}{T}b\right)^2 + x_2^2 + \dots + x_n^2\right)\right) dt. \quad (4.95)$$

Now, let $\alpha := b^2/(2\sigma_m^2 T)$ and $\beta := x_1/b$, while γ defines $\sqrt{x_2^2 + \dots + x_n^2}/b$. After introducing these substitutions in Eq. (4.95), the expression that has to be computed is:

$$I_{nD}(\mathbf{x}) = \int_0^T \frac{1}{\sqrt{(t(T-t))^n}} \exp\left(\frac{-\alpha T^2}{t(T-t)} \left(\left(\beta - \frac{t}{T}\right)^2 + \gamma^2\right)\right) dt. \quad (4.96)$$

Substituting t/T by $z/(1+z)$, we obtain

$$I_{nD}(\mathbf{x}) = \int_0^{+\infty} \frac{(1+z)^{n-2}}{\sqrt{z^n}} \exp\left(\frac{-\alpha}{z} \left((\beta + \beta z - z)^2 + (1+z)^2 \gamma^2\right)\right) dz, \quad (4.97)$$

which can be used to compute the PDF of the TIBB in an n -dimensional space if $n \geq 2$. For odd n , formula 3.472.5 in Gradshteyn and Ryzhik (2007) can be used, whereas formula 8.432.7 can be applied for even n . For the sake of illustration, the integral is computed in three dimensions, such that we finally arrive at:

$$I_{3D}(\mathbf{x}) = \frac{\left(1 + \sqrt{\frac{1-2\beta+\beta^2+\gamma^2}{\beta^2+\gamma^2}}\right) \sqrt{\pi}}{\sqrt{\alpha(1-2\beta+\beta^2+\gamma^2)}} \exp\left(-2\alpha\left(-\beta + (\beta^2 + \gamma^2)\left(1 + \sqrt{\frac{1-2\beta+\beta^2+\gamma^2}{\beta^2+\gamma^2}}\right)\right)\right). \quad (4.98)$$

Higher dimensions are not explicitly considered since they are not of interest to animal ecology.

4.6 Time efficiency

4.6.1 Synthetic experiment

The computing times required for determining the PDF of the TIBB either analytically and numerically were first examined by considering an artificial dataset containing only two observations, $\mathbf{X}(0) = (0, 0)^\top$, $\mathbf{X}(10) = (10, 10)^\top$ and mobility factor $\sigma_m^2 = 1$. The two-dimensional space Ω is defined as

$\Omega = [\min(\mathbf{X}) - 2 \text{stdv}(\mathbf{X}), \max(\mathbf{X}) + 2 \text{stdv}(\mathbf{X})] \times [\min(\mathbf{Y}) - 2 \text{stdv}(\mathbf{Y}), \max(\mathbf{Y}) + 2 \text{stdv}(\mathbf{Y})]$ (Nielson et al., 2012),

so $\Omega = [-14, 24] \times [-14, 24]$. Mathematica (Version 9.0, Wolfram Research Inc., Champaign, US) was used to perform the calculations and all results were obtained using the Delcatty cluster on Stevin, the high-performance computing (HPC) infrastructure of Ghent University. This cluster consists of 160 computing nodes, each of which contains a Dual Intel Xeon CPU E5-2670 octo-core processor with 64 GB of physical memory per node.

The PDF of the TIBB was obtained analytically according to Section 4.4.3 at several locations in Ω . More specifically, a square grid \mathcal{G} with mesh width $\Delta x = 0.5$ was chosen and the function values of the PDF were computed at the nodes of this grid. It took on average 1.51 s (10 repetitions) to compute the PDF of the TIBB analytically at all nodes.

To obtain the PDF of the TIBB numerically, a discretization step in time Δt and a location error σ_e^2 (Bullard, 1991; Horne et al., 2007) had to be selected in addition to the spatial mesh width. For the sake of uniformity, the same Δx was used as for the evaluations of the analytically obtained PDF. Furthermore, σ_e^2 was chosen very small, *i.e.* $\sigma_e^2 = 10^{-7}$, and the midpoint integration method was used. Since the numerical PDF depends on the choice of Δt , we always considered both the computing time and the values of the PDF. Here, the maximum value of the relative

difference between $f_{\mathbf{x}}^A$ and $f_{\mathbf{x}}^N$, i.e.

$$D(f_{\mathbf{x}}^A, f_{\mathbf{x}}^N) = \max_{i,j} \left(\left| \frac{f_{\mathbf{x}}^A(i,j) - f_{\mathbf{x}}^N(i,j)}{f_{\mathbf{x}}^A(i,j)} \right| \right) \quad (4.99)$$

where $f_{\mathbf{x}}^A(i,j)$, respectively $f_{\mathbf{x}}^N(i,j)$, denotes the value of the analytically determined, respectively numerically determined, $f_{\mathbf{x}}$ at location (i,j) in \mathcal{G} , was used to compare the PDFs.

As can be seen in Fig. 4.5(a), the average computing time required to determine the PDF of the TIBB numerically decreases exponentially as Δt increases. In Fig. 4.5(b), the maximum value of the relative difference between the analytical and numerical PDF in function of Δt is shown. Since we may assume that the evaluation of the analytical expression is accurate because its components are implemented efficiently in mathematical software, we concluded from Fig. 4.5(b) that the numerical PDF of the TIBB approaches the analytical one as Δt decreases.

The horizontal line in Fig. 4.5(a) indicates the average computing time required to compute the PDF of the TIBB analytically. Approximately at $\Delta t = 0.04$, the analytical and numerical PDF of the TIBB have the same computing time, but the maximum relative difference then is 8%. So if one uses the midpoint integration method, one should always take into account that the choice of Δt influences the resulting PDF. More specifically, one should choose a Δt for which the maximum relative difference between the analytical and numerical PDF of the TIBB does not exceed a fixed, small threshold. To obtain, for example, a maximum relative difference of 1.13%, Δt should have been at most 0.016, but the computing time then becomes 3.55 s, which is about 2.35 that of the analytical PDF.

4.6.2 Real-world dataset

To conclude this section, we consider a dataset of Bechstein's bat (*Myotis bechsteinii*) living in Flanders (Belgium) (Janssen, 2011; Wikelski and Kays, 2014). The dataset contains 29 observations of one individual at September, 13, 2009 from 17.45 PM until 21.45 PM. As such, there is an observation every 15 min (0.25 h). We chose a grid \mathcal{G} with 250×250 nodes covering the concerned geographical extent and calculated the analytical and numerical PDF of the TIBB. The observations are available as UTM-coordinates, but for representing our results, the lower left corner of our grid starts at $(0,0)$. There are multiple (N) data points collected during a measurement campaign so they are handled as a sequence of consecutive starting and ending locations (Bullard, 1991). First, the PDF of an analytical TIBB for every successive observation pair i , i.e. $f_{\mathbf{x}_i}$, has to be calculated. Then, the

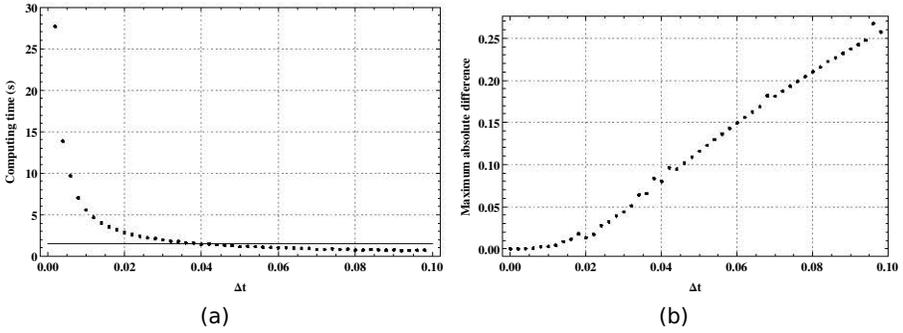


Figure 4.5: Computing time of the numerically determined PDF of a TIBB versus Δt , where the horizontal line indicates the computing time of the analytically determined PDF (a). The maximum value of the relative difference between the analytical and numerical PDF of the TIBB versus Δt (b). The Brownian bridge starts at $\mathbf{X}(0) = (0, 0)^T$, ends at $\mathbf{X}(10) = (10, 10)^T$ and has a mobility factor $\sigma_m^2 = 1$.

results are aggregated using a weighted average:

$$f_{\mathbf{x}}(x, y) = \sum_{i=1}^{N-1} \frac{T_i}{\sum_{i=1}^{N-1} T_i} f_{\mathbf{x}_i}(x, y). \quad (4.100)$$

The weight of a TIBB is proportional to the time elapsed between its starting and ending location (Bullard, 1991; Horne et al., 2007). Evaluating the analytical PDF took on average 75.16 s and its result can be seen by means of a surface plot and a contour plot in Figs. 4.6(a) and 4.6(b). The origin is located in the south-west corner of the region at stake. For obtaining the contour plot, we determined 10 thresholds, one threshold τ per decile \mathcal{D} , so that the subset of pixels having a PDF value in their center that is greater than or equal to this threshold represents $(100 \times \mathcal{D})\%$ of the volume under the PDF. For each threshold, the pixels having PDF values equal to this threshold are connected, so that a contour is obtained that delineates the area having the highest possible PDF values where the individual is expected to be located during $(100 \times \mathcal{D})\%$ of the studied time interval.

The numerical PDF of the TIBB was calculated for three different step sizes Δt , being $1/20 \times 0.25$ h, $1/250 \times 0.25$ h and $1/25000 \times 0.25$ h. The resulting contour plots are shown in Fig. 4.7. As can be inferred from this figure, these plots show more detail as Δt becomes smaller. For $\Delta t = 1/25000 \times 0.25$ h, the resulting PDF looks qualitatively the same as the analytically determined one, while the difference between the contour plots of the numerical and analytical PDF is obvious for $\Delta t = 1/250 \times 0.25$ h and $\Delta t = 1/20 \times 0.25$ h. This is also confirmed by the maximum values of the relative difference between the analytically and numerically PDF, which equals 0.15% for $\Delta t = 1/25000 \times 0.25$ h, while it is as large as 24.19% for $\Delta t = 1/20 \times 0.25$ h. On the other hand, the computing time increases from 37.25 s to 13 398.45 s as Δt decreases from $1/20 \times 0.25$ h to $1/25000 \times 0.25$ h. If a value of 1% is allowed for the maximum value of the relative difference, the

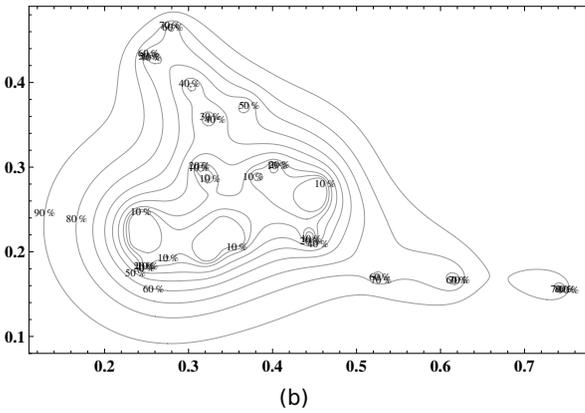
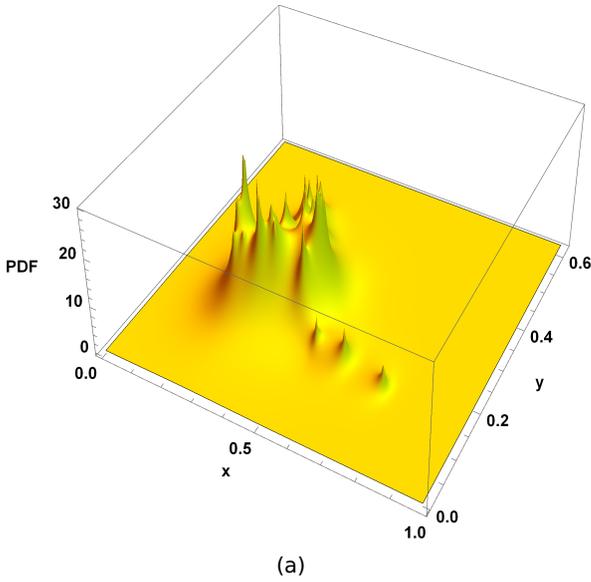


Figure 4.6: The analytically determined PDF of the TIBB for the Bechstein's bat dataset: (a) surface plot and (b) contour plot. The lower left corner of our grid is (0,0) and has km as unit.

computing time of the numerical PDF is 2.8 times that of the analytical PDF.

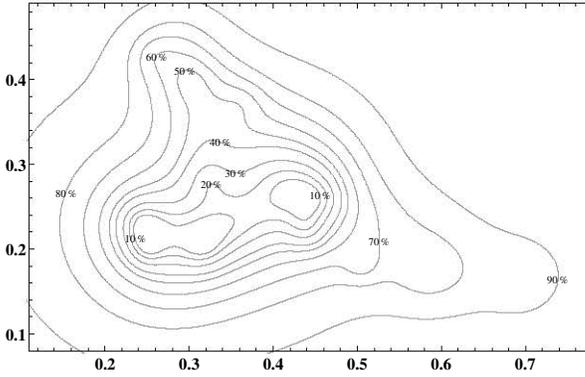
From the two experiments, it can be concluded that the analytical description of the TIBB makes it way more practicable to use as it is less time consuming, which is especially important if an accurate result is required. The datasets used in this paper are very small and as such, the computing times for the PDF of the TIBB are still relatively limited but as more and more animal tracking data become available and the typical dataset sizes increase, the analytical expression will definitely be useful in animal movement research. Moreover, there is no more need of selecting the discretization step in time and the results are always accurate.

4.7 Discussion

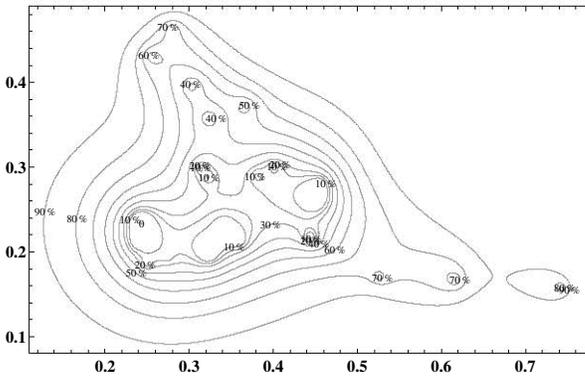
In this chapter, the position \mathbf{X} at time t of a particle performing a BB was treated as a random variable $\mathbf{X}(t)$, such that the BB can be interpreted as a stochastic process. Although the analytical description of the PDF of a TIBB with $X(0) = 0$, $X(1) = b$ and $\sigma_m^2 = 1$ was already derived in Borodin (1989), we first showed how it can be obtained using basic calculus as this helps to unravel the higher dimensional cases. Subsequently, the two-dimensional counterpart was derived. Starting from the simplest case, which involves only movement in the x -direction and independence between both directions, we moved on to the setting involving movement in a random direction and/or dependence. The closed-form expression of the PDF of the two-dimensional TIBB is interesting for animal movement studies since our results enable one to rely on efficiently implemented, well-known functions when applying BBMMs to delineate important regions for a species and avoid numerical integration methods. More specifically, the presented expression can be used when considering basic Brownian bridges (Horne et al., 2007), dynamic Brownian bridges (Kranstauber et al., 2012), biased random bridges (Benhamou, 2011) and bivariate Gaussian bridges (Kranstauber et al., 2014). In this way, the results will be more accurate, the calculations less time consuming and the selection of an appropriate discretization step in time unnecessary. This was illustrated in Section 4.6 for small datasets. Obviously, the merits of introducing the analytical expression would be even clearer with larger datasets.

As such, we hope that, in the near future, the BBMM- and move-packages available in R (Nielson et al., 2012; Kranstauber and Smolla, 2014) will be adapted to our findings so that researchers will prefer the analytical description laid bare in this chapter over the use of numerical integration methods.

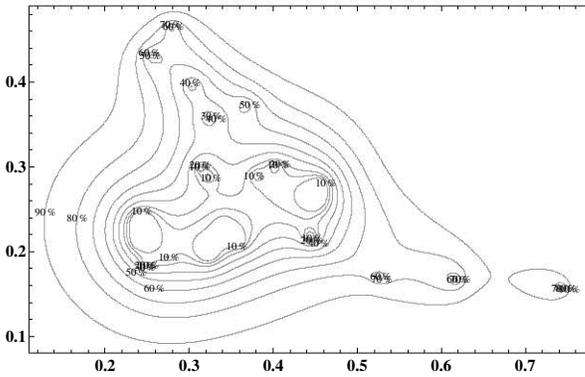
In some tracking studies, the measurement error is an important parameter that should be incorporated in the TIBB according to ecologists. In such a setting, Bullard (1991) and Horne et al. (2007) proposed to redefine the variance $\sigma^2(t)$ as $\sigma_m^2 t(T-t)/T + ((T-t)^2 + t^2)/T^2 \sigma_e^2$, where the first term corresponds to



(a)



(b)



(c)

Figure 4.7: The numerically determined PDF of the TIBB for the Bechstein's bat dataset: (a) $\Delta t = 1/20 \times 0.25h$, (b) $\Delta t = 1/250 \times 0.25h$, and (c) $\Delta t = 1/25000 \times 0.25h$. The lower left corner of our grid is $(0,0)$ and has km as unit.

the $\sigma^2(t)$ used in the TIBB without error and the second term quantifies the impact of the measurement error. Pozdnyakov et al. (2014) however states that the Markov property invoked to arrive at this expression does not hold and defines $\sigma^2(t)$ alternatively. Since we did not consider measurement errors, our analytical expression is only applicable if these errors are negligible. Yet, measurement errors can also be accounted for a posteriori by considering the starting and ending locations as random variables with probability density functions $f_{\mathbf{a}}(\mathbf{x}_{\mathbf{a}})$ and $f_{\mathbf{b}}(\mathbf{x}_{\mathbf{b}})$. Here, $\mathbf{x}_{\mathbf{a}}$ and $\mathbf{x}_{\mathbf{b}}$ are n -dimensional position variables (Horne et al., 2007). The PDF of the TIBB with error can then be computed as follows:

$$f_{\mathbf{x}}^e(\mathbf{x}) = \int_{\mathbf{x}_{\mathbf{a}}=-\infty}^{+\infty} \int_{\mathbf{x}_{\mathbf{b}}=-\infty}^{+\infty} f_{\mathbf{x}}(\mathbf{x}) f_{\mathbf{a}}(\mathbf{x}_{\mathbf{a}}) f_{\mathbf{b}}(\mathbf{x}_{\mathbf{b}}) d\mathbf{x}_{\mathbf{a}} d\mathbf{x}_{\mathbf{b}}. \quad (4.101)$$

The derivation of the analytical expression of Eq. (4.101) is not trivial. Although Eq. (4.101) has to be computed numerically, the computation time will still be favorable since the analytical expression of the PDF of the regular TIBB can be used.

5

Unraveling possible sources of poisoning for the Eurasian Eagle Owl (*Bubo bubo*) in Limburg, the Netherlands

5.1 Introduction

In the 1970s, the Eurasian Eagle Owl (*Bubo bubo*) disappeared from most areas in Central Europe as a consequence of human activities, such as shooting, the use of agrochemicals and resource exploitation (Herrlinger, 1973). Between 1964 and 1999, the Eagle Owl was reintroduced in Germany (Dalbeck and Heg, 2006; Herrlinger, 1973). Since 1997, this species has also been present in Limburg, the southeastern province of the Netherlands, most likely following the dispersion of

This chapter was converted to two papers. The first one was submitted to the European Journal of Wildlife Research (Van Nieuland et al., 2018a) and the second one was accepted for publication in Ardea (Van Nieuland et al., 2018b).

individuals from Germany (Wassink, 2010b). Based on the presence of suitable breeding habitat, Limburg has the potential to sustain up to ten breeding pairs, but up till 2010 at most five pairs were recorded (Voskamp, 2004; Wassink, 2010b). Suitable breeding locations are situated in the steep slopes of former quarries in the region.

Since 2010, there has been a continuous population increase in neighboring regions, while populations in Limburg have stagnated. This suggests that certain environmental factors in this region are constraining the Eagle Owl population size below its estimated maximum (Wassink, 2010b). Moreover, during the period 1998–2010, 11 carcasses were found in Limburg, where the cause of death could not be identified in eight cases. Two of these individuals were examined and extremely high concentrations of PCBs (polychlorinated biphenyls) and DDEs (dichlorodiphenyldichloroethylene) were found (van den Brink and Jansman, 2005). The latter chemical compound is a breakdown product of dichlorodiphenyltrichloroethane (DDT), the use of which has been illegal since the 1970s in both the Netherlands and Belgium (UNEP and FAO, 1991). Another noteworthy observation concerning the Eagle Owl population in Limburg relates to the high turnover rate at which adult Eagle Owls were replaced with another one (Wassink, 2010b; Voskamp, 2004). For example, in 1999, a male individual disappeared from one of the quarries, and was rapidly replaced, and in 2009, such a rapid replacement occurred in three out of four territories (Wassink, 2010b).

5 The high concentrations of PCBs and other poisonous compounds in the examined owl carcasses, as well as the stagnation in population size and high turnover rate, indicated that the presence of toxic compounds in the environment might be jeopardizing the expansion of the Eagle Owl population in Limburg. In order to locate possible sources of poisoning and implement an appropriate conservation policy, a GPS tracking study was conducted on six adult Eagle Owls in 2010–2011.

In order to locate the possible poisoning sources and to protect the Eagle Owl population, it is important to identify the most frequently visited areas and to get insight into the motives for visiting a particular area (Martínez et al., 2003). Therefore, the environmental characteristics (ECs) and areas to focus on when determining possible sources of poisoning need to be identified. Here, we considered two approaches to do this.

In the first approach (Section 5.4), the suitability of the study area was summarized in terms of a habitat suitability (HS) map. The first step in the construction of such a map consists of selecting the ECs that are determinants for the species' dynamics (U.S. Fish and Wildlife Service (1981), Section 2.5.4). Here, this was done on the basis of literature and expert knowledge. Furthermore, the most important areas in the study area were identified by interpreting and reclassifying the resulting HS map. In this approach, the tracking data were solely used for interpretation and validation.

In the second approach (Section 5.5), the focus was on determining the areas and ECs visited most often during hunting in order to better explain the high concentrations of PCBs and DDE. This was accomplished by reducing the size of the data set on the basis of literature and expert knowledge so that the remaining fixes are linked to possible hunting grounds. Subsequently, a hotspot analysis was conducted to locate the most visited and, consequently, most important areas for the tracked individuals. Furthermore, we investigated the relative importance of the prevailing ECs in the owls' hunting grounds by making use of ecological-niche factor analysis (ENFA, Section 2.5.4). Contrary to the first approach, the tracking data were thus used to identify both the important ECs and areas, rather than literature and expert knowledge.

Both approaches were compared in terms of the areas and ECs that are identified as important (Section 5.6). The intention of this chapter is to provide managers with insights in order to better protect the owls and their habitat. Next to this identification, additional field studies are still needed to unravel the details about possible poisoning sources, which is, however, beyond the scope of this thesis. As a first step, the study area, the tracking dataset and the species are described in Sections 5.2 and 5.3.

5.2 Study area and data set

Tracking data of three breeding pairs, referred to in the remainder as pairs A, B and C, settled in (former) quarries of the province of Limburg, the Netherlands, are considered. Figure 5.1 shows the study area together with the tracking data that were collected between June 25, 2010 and July 20, 2011. The six adults were caught by a spring net baited with dead pigeons and equipped with tracking devices (GPS-UHF logger, E-obs, Gruenwald, Germany). Pair A lived in the surroundings of the village Meerssen, pair B was settled close to the city Maastricht, and finally, pair C lived in the surroundings of the village Cadier en Keer. The rectangular area enclosing all data points is considered. In a Universal Transverse Mercator (UTM) coordinate system, this corresponds to y-coordinates varying between 5,629,628.10 m and 5,642,918.10 m, and x-coordinates between 682,237.82 m and 698,047.82 m in UTM-zone 31N. As such, the study area measures 15.61 km by 13.08 km. We used the UTM projected coordinate system since it keeps the distortion minimal and leads to distances in meters (Langley, 1998).

Figure 5.2 displays the periods during which the different owls were tracked together with some important time stamps defining their behavior, while Table 5.1 shows more information on the extent of the tracking data set. During these periods, the owls' locations were recorded every three nights, from between 4–5 PM until about 5–7 AM (GMT +1), with a time resolution of 20 minutes. The male member of pair A was only tracked from 8 PM onwards. Missing data of females B

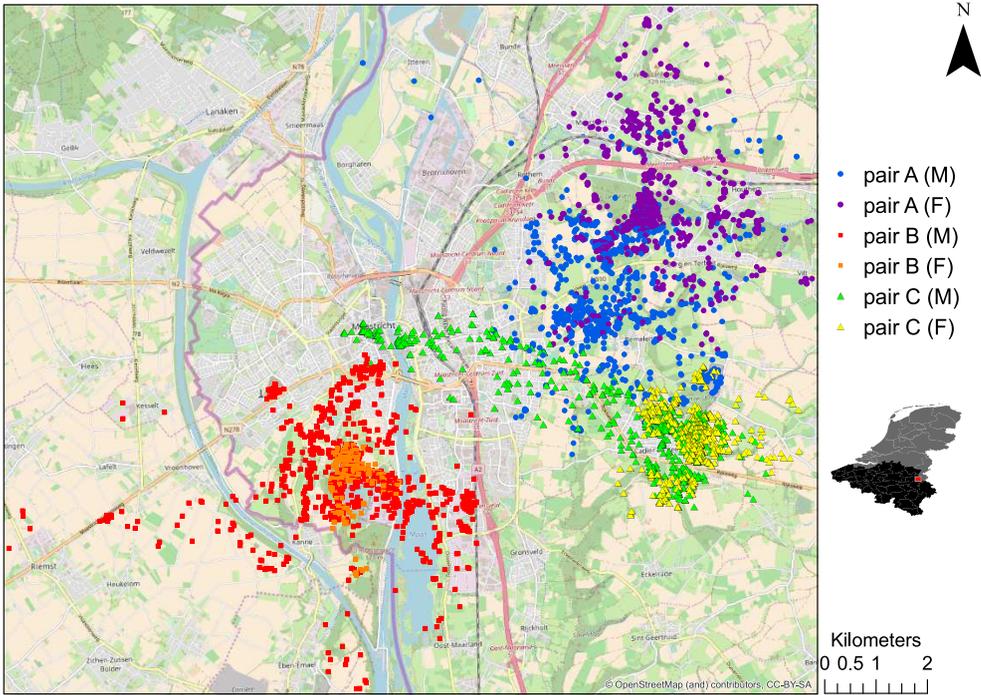


Figure 5.1: Study area together with the GPS tracking data of the tracked Eagle Owls (Source: OpenStreetMap contributors, UTM geographic coordinate system). The location of the study area is shown in the inset by a red rectangle.

5

and C can be linked to breeding, since there was no GPS coverage at the nesting site.

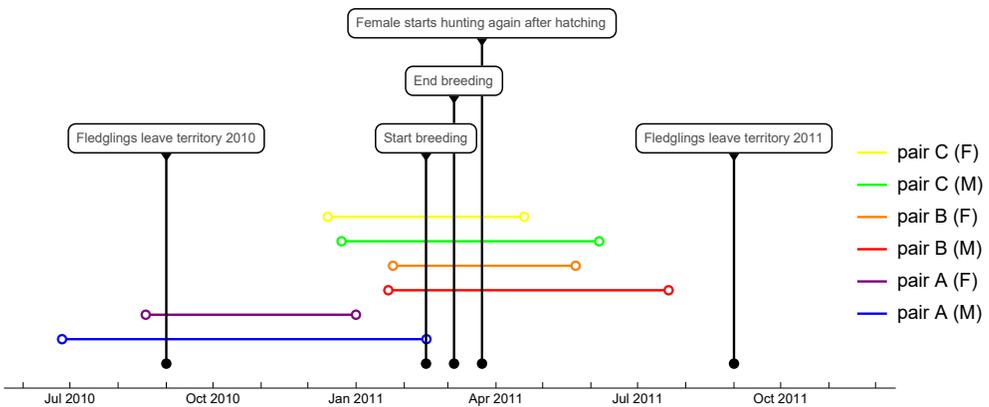


Figure 5.2: Study period per owl.

The topographical maps of the Netherlands (Het Kadaster, 2016) and Belgium (Na-

Table 5.1: Details on the extent of the tracking data set.

| Individual | Number of time-stamped locations | Missing data (%) | Total number of fixes |
|------------|----------------------------------|------------------|-----------------------|
| A Male | 2306 | 2.73 | 2243 |
| A Female | 1966 | 11.40 | 1741 |
| B Male | 2412 | 9.87 | 2174 |
| B Female | 1649 | 49.60 | 831 |
| C Male | 2710 | 8.30 | 2484 |
| C Female | 2048 | 37.20 | 1287 |

tionaal Geografisch Instituut, 2010) are used to extract geographical information. They are available at a scale of 1:5,000 and 1:10,000, respectively, and were merged appropriately. In our study, a pixel represents an area of 5 m by 5 m. The relevant available geographical information can be seen Figs. 5.3(a)–5.3(c). Next to several villages and the city of Maastricht, our study area is made up of half-open cultural landscape with structural elements. It contains pastures and arable land intersected by linear vegetation structures, isolated vegetation, differences in altitude, water courses and roads, and are alternated by woody patches, orchards and other land uses related to trees.

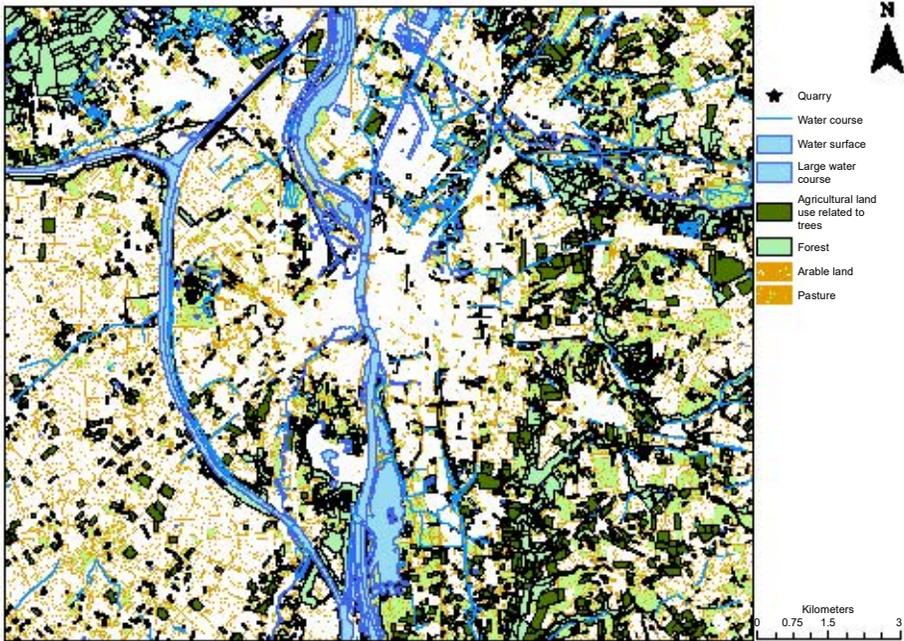
5.3 The Eurasian Eagle Owl

5.3.1 Distribution and ecology

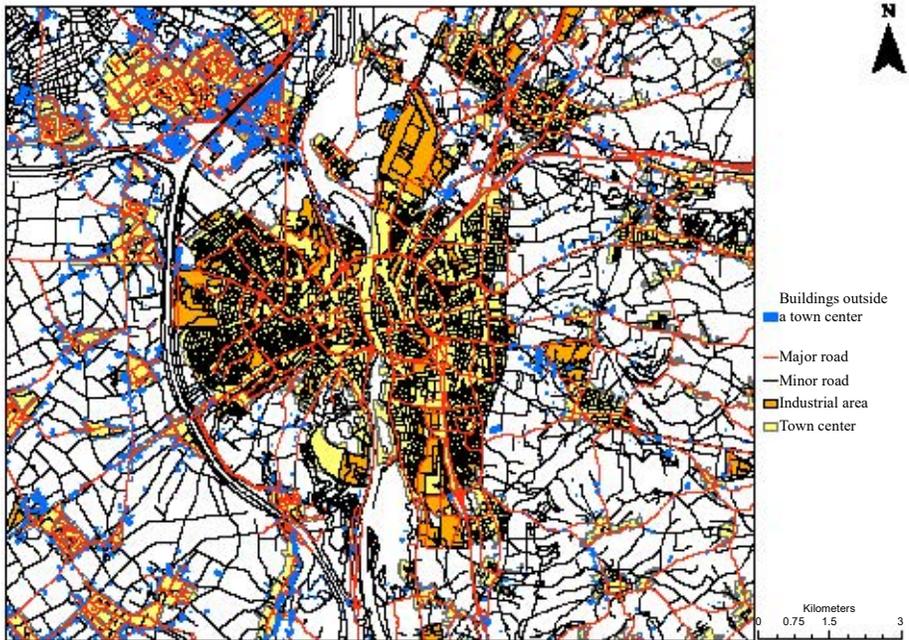
The Eurasian Eagle Owl (*Bubo bubo*) is a predatory bird known for its impressive size, weight and looks (König et al., 2010). It is easily recognized by its long black ear-tufts, fully feathered talons and bright orange-red eyes. Its rusty brown plumage serves as camouflage (Heintzenberg, 2008; König et al., 2010) (Fig. 5.4). The taxonomic classification of this species is given in Table 5.2. Together with the American horned owls, the Eagle Owls make up the genus *Bubo*. The species *Bubo bubo* has about 16–26 subspecies with northern subspecies often being bigger and having a darker plumage than their southern counterparts (Heintzenberg, 2008; König et al., 2010).

Eurasian Eagle Owls are endemic to a large part of Europe and Asia, but not to the British Islands, north-west France, northern Scandinavia and the islands of the Mediterranean Sea (Fig. 5.5) (Gilruth, 2016). The total range in Europe and Asia covers about 51 million square kilometers, while the total population size is estimated to be between 100,000 and 500,000 individuals (Gilruth, 2016).

Eagle Owls are active from dusk till dawn, with peak activity typically one hour



(a)



(b)

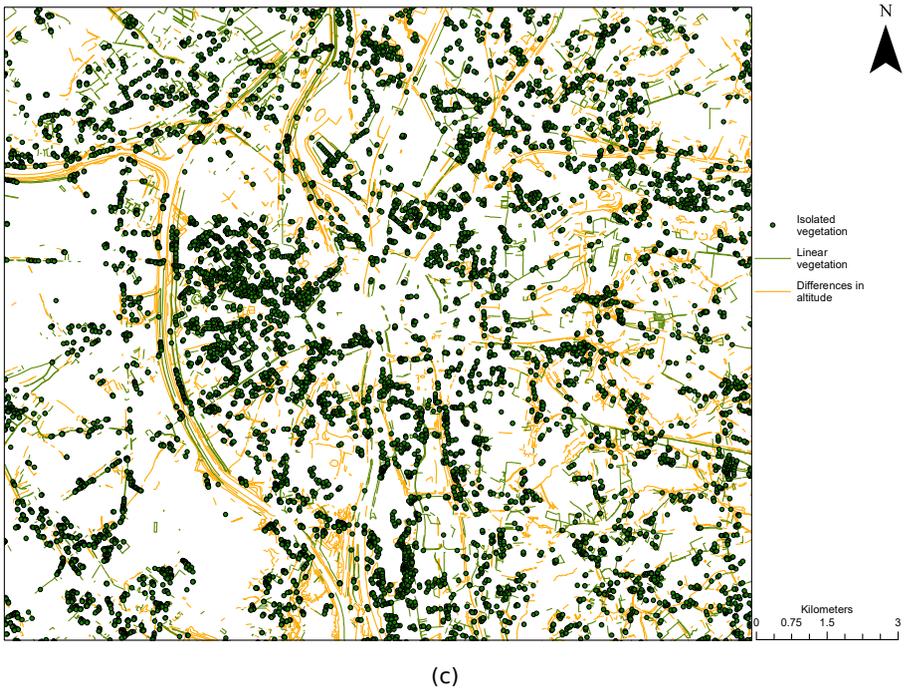


Figure 5.3: The spatial distribution of quarries, arable land, pasture, forest, agricultural land use related to trees and water bodies (a), of town centers, industrial areas, buildings outside town centers and roads (b) and of linear and isolated vegetation and differences in altitude across the study area (c).

Table 5.2: Taxonomic classification of the Eurasian Eagle Owl (Birdlife international, 2018a).

| | |
|---------|----------------------|
| Kingdom | Animalia |
| Phylum | Chordata |
| Class | Aves (Birds) |
| Order | Strigiformes (Owls) |
| Family | Strigidae (True Owl) |
| Genus | Bubo (Eagle-Owl) |
| Species | Bubo bubo |

before sunset and during sunrise (Penteriani, 2002; König et al., 2010). During the day, they rest out of sight high in the trees or on rocks. Adult Eagle Owls are strictly territorial and are known to live in the same area for years. They prefer structured landscapes, with steep slopes, hills and valleys, and a high prey availability throughout the year. Moreover, they prefer elevated structures in half-open landscapes, such as trees or ridges, where they can sit and wait until prey pass by (Wassink, 2010a, 2011a, 2012, 2014b).

Despite their imposing posture and advanced hunting techniques, Eagle Owls are



Figure 5.4: The Eurasian Eagle Owl (Karunaratne, 2012).

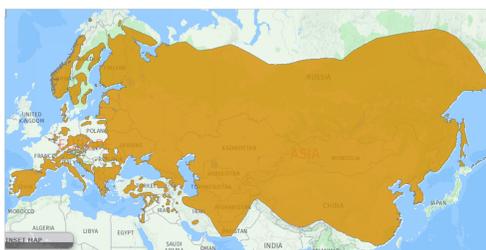


Figure 5.5: Range of the Eurasian Eagle Owl (Gilruth, 2016).

5 tolerant towards other Eagle Owls. When these enter their territory, they will be warned by the song of the resident owl and the latter becomes aggressive only when the intruders come too close to the nesting site (Heintzenberg, 2008). Eagle Owls tend to avoid interaction with humans, though several breeding pairs have been spotted in major cities, such as Helsinki, due to their high food availability (Alerstam et al., 2003). Still, they are very sensitive to disturbances near their nesting sites, and there are reports of humans being attacked when approaching the fledglings (Heintzenberg, 2008; König et al., 2010).

Miosga et al. (2015) state that certain flight paths and observation posts reoccur in observations of Eagle Owls' hunting behavior. For example, long flights are rare since most flights are short displacements, both in distance and time, alternated with long breaks. This indicates that Eagle Owls move through the landscape from one observation post to another, guided by structural elements in the landscape, such as hedgerows, trees, differences in altitude and so on.

Generally speaking, individuals spend most of their time near their nesting site. During breeding season, the movement patterns of male and female owls are very different. Only the latter incubate the eggs, while the males look for food (Voous and Cameron, 1990; König et al., 2010; Mikkola, 2014). Consequently, it are only the males that hunt and feed the females and chicks from mid-February until the

end of March (Heintzenberg, 2008). However, once the food requirements of the chicks increases, females also start hunting.

In addition, März (1958) mentioned that the hunt of Eagle Owls' hunting is versatile and hard to observe due to the fact that they are nocturnal, opportunistic predators that plunder nests, surprise sleeping birds or just sit and wait at their observation posts for prey to pass by.

5.3.2 Prey species

Eagle Owls are at the top of the food chain and do not have natural enemies. The owls have been endangered in many parts of Europe due to human persecutions and habitat destruction. Furthermore, since the Eagle Owl eats species from higher trophic levels, there is an increased health risk due to bioaccumulation (Lourenço et al., 2011). From 1960 onwards, strict protective EU legislation has saved the Eagle Owl from extinction. Besides, successful reintroductions helped to increase population sizes (Heintzenberg, 2008). Currently, the bird's conservation status is 'being of least concern' (International Union for Conservation of Nature and Natural Resources, 2018).

An Eagle Owl kills its prey with its powerful talons and by biting their heads with its bill. Small animals are consumed at once, whereas larger ones are taken to a perch, where they are torn into pieces. Birds are first plucked before they are eaten (van den Brink and Jansman, 2005; Heintzenberg, 2008; König et al., 2010). Since owls are opportunistic predators, their space use is mainly determined by the presence of prey. In the next paragraph, an overview of the most important prey species for the Eagle Owls in Limburg, the Netherlands, and the habitat preferences of these prey species, is given.

The most common prey species are the common Wood Pigeon (*Columba palumbus*) and the Rock Pigeon (*Columba livia*). The former's density is high in urban areas and half-open landscapes (Penteriani et al., 2002), which are characterized by pastures, arable land, trees and hedgerows (Wassink, 2014b). Wood Pigeons often sleep in those trees, hedgerows or along the forest edges, whereas they are absent in treeless landscapes and rare inside large forests. Rock Pigeons can be found year-round near urban areas (SOVON, 2012). In addition to Carrion Crows (*Corvus corone*) and Rook (*Corvus frugilegus*), which are generally present in large numbers (SOVON, 2012), also rats (*Rattus ssp.*), mice (*Mus ssp.*), European Hedgehogs (*Erinaceus europaeus*), Eurasian Coots (*Fulica atra*) and European Rabbits (*Oryctolagus cuniculus*) are preyed upon. Rats and mice are omnivores and can easily adapt to several habitats, but they prefer lush verge vegetation, and agricultural or urban areas (Marchesi et al., 2002; Strachan and Moorhouse, 2006; Uhlenbroek, 2008). In the Netherlands, hedgehogs are found near urban areas, such as gardens, parks and orchards. They are known to avoid coniferous forests

and humid areas (Young et al., 2006; Morris, 2006; Uhlenbroek, 2008; van de Poel et al., 2015). Rabbits also prefer dry areas to dig their holes, and areas with sufficient protection against predators, such as half-open landscapes, gardens, edges of forests and so on (Lombardi et al., 2003). To a lesser extent, other raptors and owls are also eaten.

5.4 A validated expert-based habitat suitability map

5.4.1 Introduction

To what extent an area is suitable for the survival and reproduction of a certain species, is often summarized in terms of an HS map. This map consists of a raster of pixels (or cells) organized into rows and columns, where each pixel is assigned a value in the unit interval, the value of the so-called habitat suitability index (HSI). By convention, non-suitable pixels are assigned a value of zero and fully suitable pixels a value of one (U.S. Fish and Wildlife Service, 1981). In order to demonstrate the accuracy of an HS map, validation is crucial, preferably using an independent set of data on the sites where the species is present and where it is absent (Roloff and Kernohan, 1999; Ottaviani et al., 2004; Hirzel et al., 2006). In this section, we constructed an HS map for the Eurasian Eagle Owl (*Bubo bubo*) in the province of Limburg to summarize the knowledge found in literature and expert knowledge.

5.4.2 Methods

5.4.2.1 Construction

The first step in the construction of an HS map usually consists of selecting the ECs that are determinants for the species' dynamics (U.S. Fish and Wildlife Service, 1981). These ECs are typically related to topography, human influence, land use and landscape diversity (Hirzel et al., 2006). The tolerance of the species to an EC can be represented by a function relating the EC to suitability, the so-called suitability index (SI). The SI and the spatial distribution of each EC lead to a pixel-based SI value in the unit interval. As such, each EC gives rise to a suitability map. Finally, the SI values of all ECs are aggregated per pixel into a single overall HSI value (U.S. Fish and Wildlife Service, 1981), on whose basis an HS map can be generated.

The SIs can either express a subjective opinion or, upon assuming a direct relationship (generally linear) between the carrying capacity and suitability (Oldham

et al., 2000), can be derived from a dataset with information on the presence and absence of the species in the study area (Ortigosa et al., 2000; Ottaviani et al., 2004). The applicability of these approaches depends on data availability (Ottaviani et al., 2004). In this study, GPS tracking data of six individuals were available. Since this data set is relatively small and we do not have information on species absence, we adopted a knowledge-based approach to design the HS map on the basis of literature and expert knowledge.

The information coming from literature and experts mainly describes the Eagle Owl's behavior in a qualitative way. To transform this into specific SI values, we considered each EC individually and chose 0.5 as a neutral SI value, meaning that a pixel having this SI value has no effect on the species, while it has a positive, respectively negative, effect if the SI value is higher, respectively lower, than 0.5. Depending on the intensity of influence exerted by a certain EC, SI values are closer to zero or to one. Essentially, the scale used in this section fits with the meaning that is typically assigned to SI values (U.S. Fish and Wildlife Service, 1981). Indeed, a completely unsuitable, respectively completely suitable, environment typically corresponds to an SI value of zero, respectively of one, whereas an environment with a low, respectively high, suitability is assigned an SI value between zero and 0.5, respectively between 0.5 and one (Wakeley, 1988).

Throughout this section, the SI value of a pixel is either based on the presence or absence of EC e in this pixel p , or on the distance from this pixel p to the nearest pixel where EC e is present. Moreover, for the sake of simplicity, a linear relationship between distance and suitability is assumed. To mathematically define the function relating EC e to an SI value, we first define an indicator function $\mathcal{E}_e(p)$:

$$\mathcal{E}_e(p) = \begin{cases} 1 & , \text{ if EC } e \text{ is present in pixel } p, \\ 0 & , \text{ otherwise,} \end{cases} \quad (5.1)$$

where $p = 1, \dots, N$, with N the number of pixels in the raster \mathcal{R} covering our study area. Next, the set E_e is defined as the subset of pixels in \mathcal{R} where EC e is present:

$$E_e = \{p \in \mathcal{R} | \mathcal{E}_e(p) = 1\}. \quad (5.2)$$

Furthermore, let $d(p, E_e)$ be the minimum Euclidean distance from a pixel p to E_e . Mathematically, the SI for EC e in pixel p is then given by:

$$S_{\alpha_e, r_e}^e(p) = \begin{cases} \alpha_e - (\alpha_e - 0.5) \frac{d(p, E_e)}{r_e} & , \text{ if } 0 \leq d(p, E_e) \leq r_e, \\ 0.5 & , \text{ otherwise,} \end{cases} \quad (5.3)$$

where r_e is the radius of influence of EC e and α_e is the numerical index that

represents the suitability of EC e if $d(p, E_e) = 0$. The SI value is different from 0.5 only if $d(p, E_e) \leq r_e$. If $r_e = 0$, the SI value only depends on the presence or absence of EC e in pixel p (Ortigosa et al., 2000).

There exist several functions to aggregate n SI values of a pixel p into one HSI value. They influence the expected mean of the obtained HSI values, as well as the degree to which variables compensate one another; that is, the degree to which an increase in one EC can substitute for a decrease in another EC while maintaining the same HSI level. All kind of functions can be chosen but the choice should lead to HSI values in the unit interval and it should reflect the relationship between the environmental characteristics (U.S. Fish and Wildlife Service, 1981; Van Horne and Wiens, 1991). For example, the minimum of the SI values is taken if the EC with the lowest SI value overrides all other ECs. In this case, there is a limiting EC and the other ECs are not relevant. Here, the value 1 is a neutral SI value, *i.e.* one does not influence the resulting HSI value (U.S. Fish and Wildlife Service, 1981). A weighted sum or average is taken if the combined effect of several SI values is considered. In this case, zero is the neutral value (Van Horne and Wiens, 1991). A product allows individual ECs to have a large potential influence on lowering the overall HSI value. Often, the resulting HSI value is lower than the lowest SI value and it decreases steeply with the reduction of an SI value. Here, the neutral value is one (U.S. Fish and Wildlife Service, 1981). In our case, we chose 0.5 as neutral value. Therefore, the uninorm aggregation function given by De Baets and Fodor (1999):

5

$$\mathcal{H}(p) = \begin{cases} 0 & , \text{ if } |\{e \mid S_{a_e, r_e}^e(p) = 1\}| \geq 1 \text{ and } |\{e \mid S_{a_e, r_e}^e(p) = 0\}| \geq 1, \\ \frac{\prod_{e=1}^n S_{a_e, r_e}^e(p)}{\prod_{e=1}^n S_{a_e, r_e}^e(p) + \prod_{e=1}^n (1 - S_{a_e, r_e}^e(p))} & , \text{ otherwise,} \end{cases}$$

(5.4)

was used for the pixel-based aggregation. This has, to the best of our knowledge, never been done before in this context. This aggregation function is symmetric and associative (Yager and Rybalov, 1996; Grabisch et al., 2009), and has 0.5 as neutral value (De Baets and Fodor, 1999). Table 5.3, showing a few exemplary computations of $\mathcal{H}(p)$, indicates that ECs with a positive influence compensate for ECs with a negative influence, and that synergistic effects occur when combining ECs having positive, respectively negative influences.

Table 5.3: Some exemplary aggregations of SI values according to Eq. (5.4).

| $S_{a_e, r_e}^{e1}(p)$ | $S_{a_e, r_e}^{e2}(p)$ | $\mathcal{H}(p)$ |
|------------------------|------------------------|------------------|
| 0.9 | 0.5 | 0.9 |
| 0.2 | 0.9 | 0.69 |
| 0.2 | 0.8 | 0.5 |
| 0.9 | 0.7 | 0.95 |

5.4.2.2 Validation and interpretation

After constructing an HS map, validation is needed to confirm its credibility. If an HS map is well constructed, pixels with high HSI values are more likely to be visited by the species. Validation is preferably done by making use of an independent dataset on the sites where the species is present and where it is absent (Roloff and Kernohan, 1999; Ottaviani et al., 2004; Hirzel et al., 2006). In this study, we used the available GPS tracking data for validation. In GPS tracking studies, individuals are followed in time, resulting in discrete observations of the continuous movement process. Since there is no information on species absence, one needs appropriate tools to deal with presence-only data (Ottaviani et al., 2004; Hirzel et al., 2006). Furthermore, GPS tracking data are auto-correlated. This means that common approaches focusing on hypothesis-testing cannot be used (Boyce et al., 2002). Therefore, the HSI values of the pixels that were probably visited by the species were compared with those available in the landscape (Boyce et al., 2002). More specifically, we combined the methodologies presented in Reynolds-Hogland and Mitchell (2007) and Hirzel et al. (2006).

In the paper of Reynolds-Hogland and Mitchell (2007), habitat selection is estimated on the basis of GPS tracking data. Since we wanted to evaluate the preferential selection of high HSI values using GPS tracking data, we could use their methodology. In this paper, fixes are first transformed into a PDF reflecting the probability density that an individual visits a location in the study area, and are summarized in terms of a home range. The PDF or the home range can then be used, instead of the individual fixes, to analyze HSI selection and validate the HS map (Mitchell et al., 2002; Reynolds-Hogland and Mitchell, 2007). Here, we used the BBMM to determine the PDF value in the center of each pixel (Bullard (1991); Horne et al. (2007); Van Nieuland et al. (2015) and Chapter 4). As a next step, the home range was defined on the basis of the resulting PDF. More precisely, we calculated a threshold τ so that the subset of pixels having a PDF value in their center that is greater than or equal to this threshold, represents 90% of the volume under the PDF. This subsets represents the area with the highest possible PDF values where the individuals are expected to be located during 90% of the studied time interval (Kie et al., 2010; Pop et al., 2018). In this way, the PDF itself and the home range could be used for validation.

On the basis of the PDF and the home range, two indices were calculated to compare the pixels that were probably visited with those available in the study area: Ivlev's electivity index (Ivlev, 1960; Reynolds-Hogland and Mitchell, 2007) and Manly's habitat selection index (Manly et al., 2002; Desbiez et al., 2009; Hirzel et al., 2006). In both cases, the HSI range is partitioned in b bins, after which two proportions are calculated for every bin $c = 1, \dots, b$. The first one is the proportion of pixels in the study area that belong to HSI bin c and, as such, quantifies the area available for the species:

$$A_c = \frac{a_c}{N}, \quad (5.5)$$

with a_c the number of pixels in the study area belonging to HSI bin c and N the total number of pixels in the study area. The second one is the proportion of pixels in the home range that belong to HSI bin c and, as such, describes the area used by the species:

$$U_c = \frac{u_c}{\tilde{N}}, \quad (5.6)$$

with $\tilde{N} \leq N$ the number of pixels in the home range, *i.e.* the number of pixels having centers where the PDF value is greater than or equal to a threshold τ , and u_c the number of pixels in the home range falling into HSI bin c .

In order to quantify the area used by the species more accurately, the PDF itself can be used (Reynolds-Hogland and Mitchell, 2007):

$$\tilde{U}_c = \sum_{i=1}^{a_c} t_i s^2, \quad (5.7)$$

with t_i the PDF value in the center of pixel p and s the size of a pixel. As such, $t_i s^2$ represents the volume under the PDF and, thus, the probability of occurrence in pixel p . Consequently, \tilde{U}_c embodies the probability of occurrence in HSI bin c .

Finally, these proportions are used to define two indices. The first one, Ivlev's electivity index (Ivlev, 1960; Mitchell et al., 2002; Reynolds-Hogland and Mitchell, 2007) is given by:

$$I_c = \frac{U_c - A_c}{U_c + A_c} \quad (5.8)$$

for $c = 1, \dots, b$. The values of this ratio range from -1 (avoidance) to 1 (strong selection). If the HS map is well designed, $I_c < 0$, resp. $I_c > 0$, in HSI bins corresponding to a lower, resp. higher, habitat suitability. In this expression, U_c can be replaced by \tilde{U}_c (Reynolds-Hogland and Mitchell, 2007), to obtain

$$\tilde{I}_c = \frac{\tilde{U}_c - A_c}{\tilde{U}_c + A_c}. \quad (5.9)$$

The second validation index, Manly's habitat selection index (Manly et al., 2002;

Desbiez et al., 2009), reads:

$$M_c = \frac{U_c}{A_c} \quad (5.10)$$

for $c = 1, \dots, b$. This ratio is similar to the one underlying the Boyce index (Boyce et al., 2002; Hirzel et al., 2006). If the HS map is well designed, U_c should be lower than A_c , and thus $M_c < 1$, in HSI bins corresponding to a lower habitat suitability. Again, U_c can also be replaced by \tilde{U}_c in Eq. (5.10) in order to calculate the ratio on the basis of the probability of occurrence of the species:

$$\tilde{M}_c = \frac{\tilde{U}_c}{A_c}. \quad (5.11)$$

Ideally, the indices as function of the average HSI value of each bin c should increase linearly, so that the average HSI value is proportional to the calculated indices (Hirzel et al., 2006). It should, however, be mentioned that the magnitude of the validation indices are sensitive to the extent and the shape of the study area.

Instead of choosing b HSI bins, Hirzel et al. (2006) used a moving window of width W that shifts from HSI value 0 to 1 in order to arrive at smoother graphs of the indices versus the average HSI value. As such, the shape of the graphs can be used, next to the indices values themselves, to better evaluate and interpret the resulting HS map and to decide which pixels are suitable for the studied species and which are not (Hirzel et al., 2006).

5.4.3 Results

5.4.3.1 The habitat suitability map

As a first step in constructing the HS map, we inventoried the ECs that influence the behavior of the Eagle Owls. Then, for every EC, we defined the appropriate a_e and r_e , after which an SI value for every EC was assigned to every pixel. The 12 ECs that are assumed to influence the Eagle Owls in Limburg, the Netherlands were selected on the basis of literature and expert knowledge (Table 5.4). The importance of each selected EC can be assessed by means of a_e . The considered ECs and their corresponding SIs are thus 100% subjective. Nevertheless, we relatively weighted our choices and motivated them as much as possible.

An important feature when selecting a habitat is the presence of potential nesting sites. In Limburg, so far the Eagle Owls only nest in former quarries (EC1) (van Lierop and Janssen, 2014; Wassink, 2014c), whose presence may as such be designated as of utmost importance in our study area ($a_e = 1$). When the owls have to cover large distances to find their prey, this consumes a lot of energy and thereby

Table 5.4: Overview of the considered environmental characteristics (ECs), together with the SI value that typifies each EC (α_e) and the radius of influence (r_e).

| EC | Description | α_e | r_e (m) |
|------|--|------------|-----------|
| EC1 | Quarry | 1 | 3500 |
| EC2 | Arable land or pasture | 0.7 | 0 |
| EC3 | Forest edge | 0.8 | 0 |
| EC4 | Agricultural land use related to trees | 0.8 | 0 |
| EC5 | Town center | 0.6 | 0 |
| EC6 | Industrial area | 0.6 | 0 |
| EC7 | Vegetation structures | 0.9 | 75 |
| EC8 | Difference in altitude | 0.7 | 75 |
| EC9 | Water bodies | 0.8 | 75 |
| EC10 | Minor roads | 0.6 | 75 |
| EC11 | Major roads | 0.2 | 1000 |
| EC12 | Buildings outside town center | 0.7 | 75 |

makes locations farther away from quarries less suitable. The average distance between 28 territories of Eagle Owls nesting in the Netherlands, Belgium and Northern Germany was found to be 6906 m (Wassink, 2007). For that reason, we chose $r_e = 3500$ m as the maximal distance from the quarry edges up to which the quarry has a positive influence.

In the remainder of this section, we first study the ECs for which the influence radius is zero ($r_e = 0$). These ECs are mainly related to land use. Secondly, we study the ECs for which $r_e > 0$ and that are related to physical elements in the landscape.

5
 Pixels where arable land or pastures (EC2) are present, were assigned a suitability of 0.7 ($\alpha_e = 0.7$) since they do not only provide prey, but also bring along threats, such as barbed wire and poisonous pesticides and herbicides. They cover most of the study area and form, together with the vegetation present, a half-open landscape where Eagle Owls live and hunt. No distinction in suitability between arable land and pasture could be made on the basis of literature. Forests are potential hunting grounds (Wassink, 2014b,a). The center of a forest, however, is less suitable due to the higher vegetation density hindering hunting. This problem is much less of an issue along forest edges (EC3), which serve as ideal sleeping grounds for pigeons and habitats for rabbits (Wassink, 2014b,a). The extent of the positive influence of such edges, depends on the area and shape of the forest, but typically ranges from 100 to 300 m inwards (Rusak, 2003). In our study area, the forests are relatively small, so we chose the former as forest edge width. Here, a suitability of 0.8 was assigned to pixels located in the forest edge ($\alpha_e = 0.8$). Pixels corresponding to locations in the center of the forest or outside the forest were assigned a suitability of 0.5 for EC3. Pixels containing agricultural land uses related to trees (EC4), such as orchards, tree and fruit nurseries and poplar plantations, were assigned an SI value of 0.8 since they are attractive for prey and consist of sparse vegetation

serving as observation post ($a_e = 0.8$). In regions where the Eagle Owls can avoid contact with humans, they live in non-urban areas (Donázar, 1989; Martínez and Calvo, 2000). However, more and more breeding pairs can be found in major cities (such as Helsinki) (Alerstam et al., 2003). Likewise, in our study area, some preys are caught near urban areas (EC5) (Voskamp, 2004; Wassink, 2010a), so we set $a_e = 0.6$ for this EC. This SI value is lower than the one assigned to agricultural land use, because it was assumed that the more natural setting provided by the latter is an added value as compared to urban areas. Furthermore, industrial areas (EC6) are good hunting grounds for the species since, for example, pigeons often rest there (Wassink, 2011b). In these areas, however, there is also human disturbance (König et al., 2010; Miosga et al., 2015; Wassink, 2016). Because of this human disturbance, we chose $a_e = 0.6$, which equals the value chosen for urban areas.

The first considered physical elements whose $r_e > 0$ are the isolated and linear vegetation structures (EC7) serving as observation posts for which we chose $a_e = 0.9$ and $r_e = 75$ m. This radius was chosen based on the findings reported in Mortenson (1971) for a related owl species. More precisely, they investigated the number of attempts to catch prey versus the distance to the prey. We linearly extrapolated the results of their study and assumed that from a distance of 75 m on, there would no longer be an attempt to catch a prey. The same radius of influence is used for EC8, EC9, EC10 and EC12. Differences in altitude (EC8) also serve as observation posts, but do not necessarily imply an increased prey availability. Although, they often occur together with EC7. So, we set $a_e = 0.7$ and $r_e = 75$ m. Water bodies (EC9) positively impact the presence of Eagle Owls since their banks often provide shelter to prey (Martínez et al., 2003; Strachan and Moorhouse, 2006; Wassink, 2007; Heintzenberg, 2008; König et al., 2010). Consequently, $a_e = 0.8$ and $r_e = 75$ m. However, pixels completely covered with water bodies were assigned a neutral SI value. Next, roads are considered which might be avoided due to traffic, but their verges can contain prey. Moreover, it is easy to sit and wait in a tree along a road until a prey crosses so that it can be caught. Yet, this is unlikely to hold for busy roads. As such, we distinguished between minor roads (EC10) and major roads (EC11). The former are paved and more than 4 m wide, and have a negative effect on the species, so that $a_e = 0.2$ and $r_e = 1000$ m on either side of the road. This radius has been reported as the distance within which breeding birds are effected by roads (Reijnen and Foppen, 2006). For what concerns minor roads (EC10), we assumed that only the ones located outside town centers have a positive effect, since there is less human disturbance and it has been reported that the species diversity in their verges is higher outside these centers (CBS, Den Haag; Planbureau voor de Leefomgeving, Den Haag/Bilthoven en Wageningen UR, Wageningen, 2015). Consequently, $a_e = 0.7$ and $r_e = 75$ m for minor roads outside town centers. Finally, buildings outside town centers (EC12) are considered because they are typically located relatively far from other human activities. Moreover, these structures often relate to farms, which represent po-

tential food sources (Tucker et al., 2018). Consequently, $a_e = 0.7$ and $r_e = 75$ m for this EC.

The resulting HS map for Eagle Owls in our study area is shown in Fig. 5.6, while the SI maps for the individual ECs can be found in Figs. 5.7 and 5.8.



Figure 5.6: The HS map for the Eagle Owls in the study area.

5.4.3.2 The probability density of occurrence

The weighted average of the time-integrated Brownian bridges (TIBBs) defined in between each pair of consecutive fixes in the tracking data were calculated per individual owl to determine the probability density of occurrence of every individual at every location in the study area. Next, the average PDF of the six Eagle Owls in our study area was calculated. For each PDF, we determined 10 thresholds, one threshold τ per decile \mathcal{D} , so that the subset of pixels having a PDF value in their center that is greater than or equal to this threshold represents $(100 \times \mathcal{D})\%$ of the volume under the PDF. For each threshold, the pixels having PDF values equal to this threshold are connected, so that a contour is obtained that delineates the area having the highest possible PDF values where the individual is expected to be located during $(100 \times \mathcal{D})\%$ of the studied time interval. These contours are shown in Fig. 5.9 and 5.10. If the contour corresponding to the threshold of $\mathcal{D} = 1$,

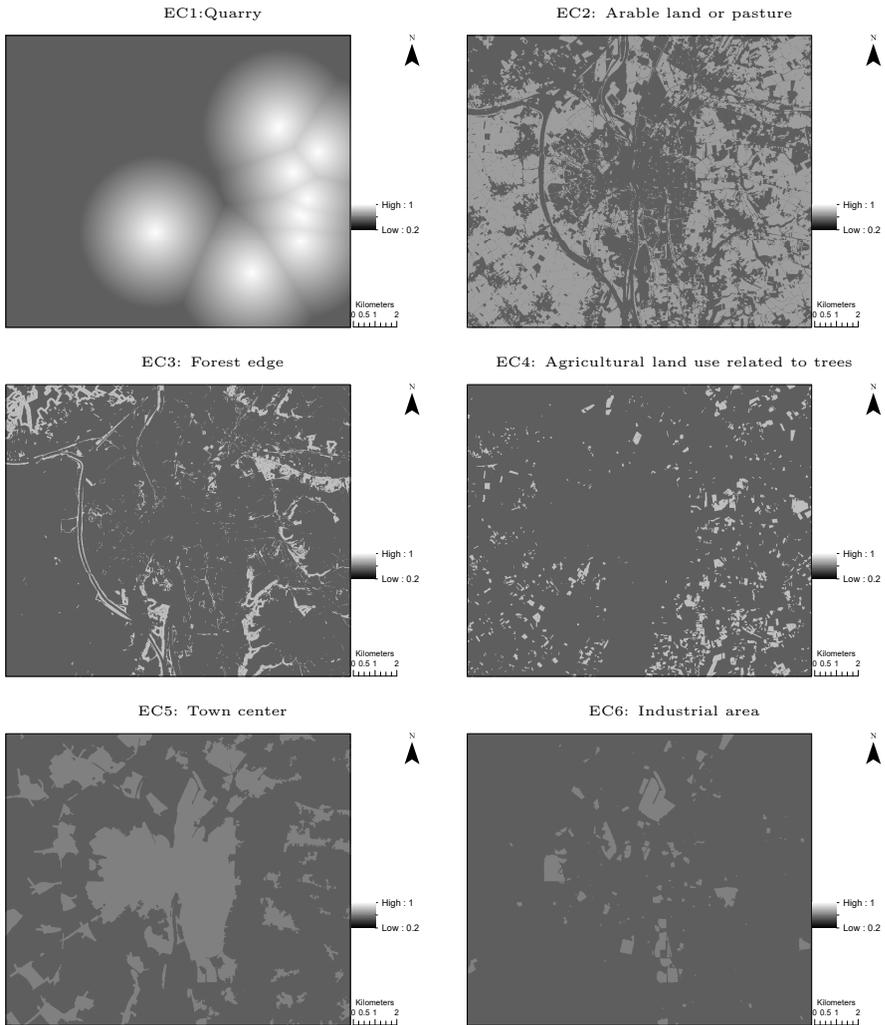


Figure 5.7: The SI map covering the study area for environmental characteristics 1–6.

i.e. the contour enclosing the area where the individuals were certainly observed within the studied time interval, is not visible, it corresponds to the boundary of the study area. The area enclosed by the contour corresponding to the threshold of, for example, $\mathcal{D} = 0.1$ is smaller than the area enclosed by the one corresponding to the threshold of $\mathcal{D} = 1$, but contains the most visited areas, with the highest probability densities.

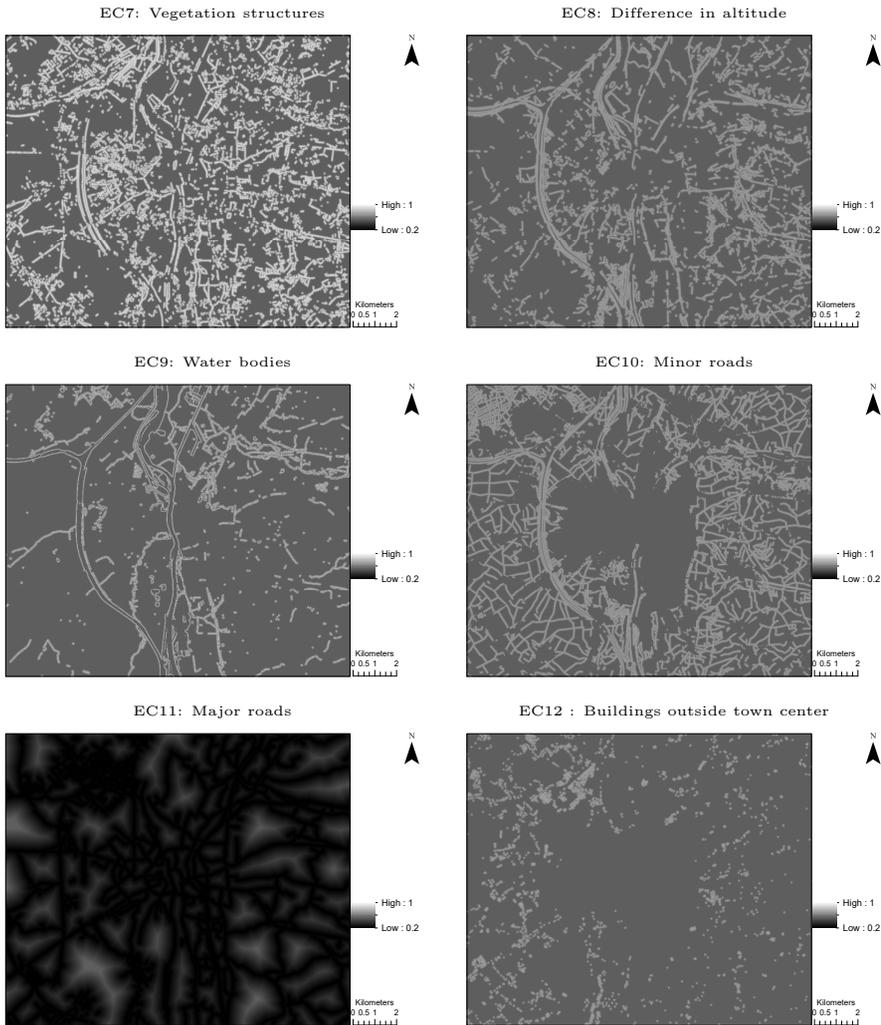


Figure 5.8: The SI map covering the study area for environmental characteristics 9–12.

5.4.3.3 Validation and interpretation

In Fig. 5.11, the probability mass function (PMF) of the HSI values assigned to the pixels in the entire study area (black) and those assigned to the fixes (gray) are shown (bin width= 0.05). The HSI values in the study area range from 0.2 to 1 with relatively more pixels with a higher suitability. However, this skewness is much more pronounced in the PMF for the fixes. The large majority of the pixels are located in highly suitable areas with an HSI value greater than 0.95. This

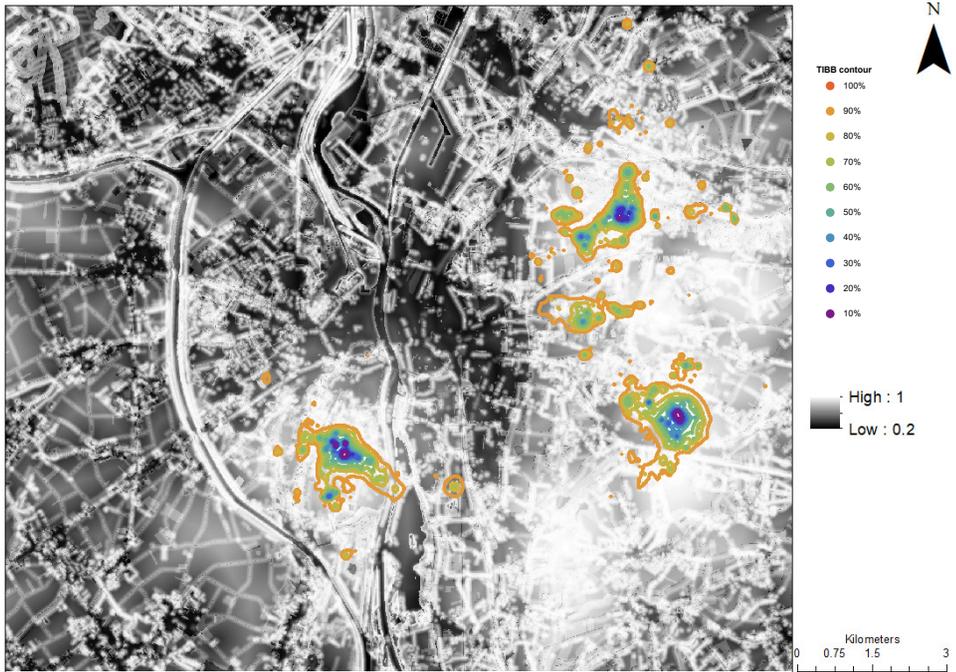


Figure 5.9: The probability density of occurrence of the six Eagle Owls in the study area. Ten thresholds were determined for this PDF, one threshold τ per decile \mathcal{D} , so that the subset of pixels having a PDF value in their center that is greater than or equal to this threshold, represents $(100 \times \mathcal{D})\%$ of the volume under the PDF. For each threshold, the pixels corresponding to this threshold are connected so that a contour is obtained that delineates the area having the highest possible PDF values where the individual is expected to be located during $(100 \times \mathcal{D})\%$ of the studied time interval. The HS map is used as basemap.

probably stems from the fact that many fixes are located near the nesting sites. Nevertheless, also locations farther away from nesting sites are of interest since they can, for example, serve as hunting grounds.

Ivlev's electivity index (Ivlev, 1960; Mitchell et al., 2002) and Manly's habitat selection index (Manly et al., 2002; Desbiez et al., 2009) were both calculated using a moving window of width $W = 0.1$ that was shifted along the HSI range in steps of 0.01. The average HSI value versus I_c (Eq. (5.8)) and \tilde{I}_c (Eq. (5.9)) are shown in Fig. 5.12, while the average HSI value versus M_c (Eq. (5.10)) and \tilde{M}_c (Eq. (5.11)) are shown in Fig. 5.13. It can be seen that these indices do not increase linearly as a function of the HSI value, but there is a sharp increase towards the end of the HSI range. This is due to the fact that many fixes in our dataset are located close to each other in a limited number of pixels, which consequently all have a very high probability of occurrence. Furthermore, they are, as mentioned before, located in highly suitable pixels, close to nesting sites and bias the results.

As can be seen in Fig. 5.12, I_c is about -1 for HSI values lower than 0.55, while it is

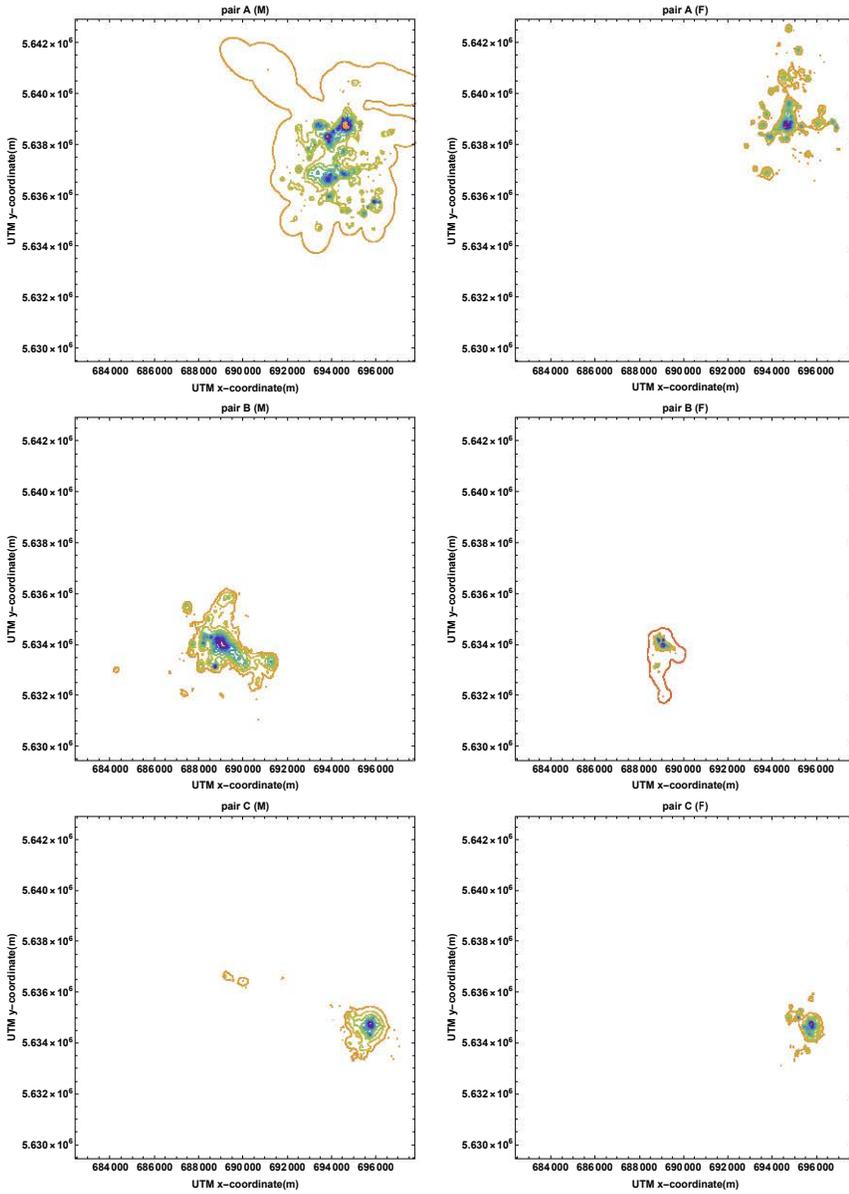


Figure 5.10: The probability density of occurrence for each Eagle Owl individual. Ten thresholds were determined for this PDF, one threshold τ per decile \mathcal{D} , so that the subset of pixels having a PDF value in their center that is greater than or equal to this threshold, represents $(100 \times \mathcal{D})\%$ of the volume under the PDF. For each threshold, the pixels corresponding to this threshold are connected so that a contour is obtained that delineates the area having the highest possible PDF values where the individual is expected to be located during $(100 \times \mathcal{D})\%$ of the studied time interval. The legend as in Fig. 5.9 and M indicates male and F female.

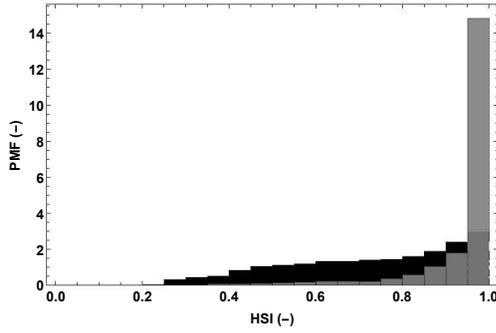


Figure 5.11: The probability mass function of the HSI values assigned to the pixels in the entire study area (black) and those assigned to the fixes (gray) (bin width= 0.05).

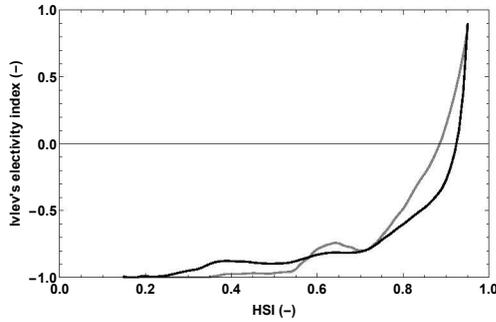


Figure 5.12: The average HSI value versus I_c (gray) and the average HSI value versus \tilde{I}_c (black). In both cases, a moving window of width 0.1 that is shifted along the HSI range in steps of 0.01, was used.

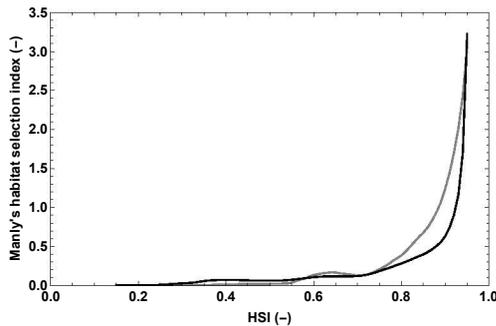


Figure 5.13: The average HSI value versus M_c (gray) and the average HSI value versus \tilde{M}_c (black). In both cases, a moving window of width 0.1 that is shifted along the HSI range in steps of 0.01, was used.

about -0.75 for HSI values between 0.6 and 0.7, after which it increases gradually. In contrast, \tilde{I}_c increase gradually, though with a steeper slope from about 0.7 on, after which (at about 0.9) a very steep increase follows. If $I_c = 0$ would be chosen as a threshold to distinguish between suitable and non suitable areas (according to Mitchell et al. (2002); Reynolds-Hogland and Mitchell (2007)), only areas with HSI values greater than 0.89 would be identified as suitable for the owls. For \tilde{I}_c , this corresponds with HSI values greater than 0.93. These high HSI values result from the bias in the fixes' locations described earlier. Therefore, we have a look at the shape of the graphs to interpret the results in an alternative way. On the basis of I_c , we may conclude that pixels with an HSI value lower than 0.55 are unsuitable (where I_c is about -1), pixels with an HSI value between 0.55 and 0.7 are suitable (just before the steep increase), while pixels with an HSI value greater than 0.7 are highly suitable. Alternatively on the basis of \tilde{I}_c , we may conclude that pixels with an HSI value lower than 0.7 are unsuitable (where \tilde{I}_c is low), pixels having an HSI value between 0.7 and 0.9 are suitable (just before the steep increase), while pixels with an HSI value greater than 0.9 are highly suitable, and are probably located close to a possible nesting site.

As can be seen in Fig. 5.13, M_c is about zero for HSI values lower than 0.55, while it is about 0.15 for HSI values between 0.6 and 0.7 after which it increases gradually. For \tilde{M}_c , which is computed on the basis of the probability of occurrence, the graph gradually increases from about 0.7 on, after which (at about 0.9) there is a steep increase. If $M_c = 1$ would be chosen as a threshold for distinguishing between suitable and non suitable areas (according to Hirzel et al. (2006)), only areas with HSI values greater than 0.89 would be labeled as suitable for the owls, *i.e.* that it holds $U_c \geq A_c$. For \tilde{M}_c , this corresponds with HSI values greater than 0.93. Given these HSI values and given that the slope of the graph of M_c , respectively of \tilde{M}_c , changes at the same HSI values as the ones of I_c , respectively of \tilde{I}_c , the same conclusions can be drawn from the shapes of the graphs when using Manly's habitat selection index instead of Ivlev's electivity index.

Given the fact that Eagle Owls almost exclusively use areas with HSI values greater than 0.7, and mostly areas with HSI values greater than 0.9, we may conclude that our HS map is well designed. It clearly reflects the suitability of the study area for the Eagle Owls. Since both indices lead to the same results, we cannot point towards the better one. However, we advise to use the indices based on Eq. (5.7), *i.e.* \tilde{I}_c and \tilde{M}_c , since these assign higher weights to more frequently visited locations than to less frequently visited ones. As such, it better represents the species' preferences.

To conclude, HSI thresholds of 0.7 and 0.9 were used to reclassify the HS map (Fig. 5.6) into three suitability classes: very (highly) suitable, suitable or less suitable (Fig. 5.14). Since suitable areas are more likely to be visited by the owls, they can be assigned as most important for the species.

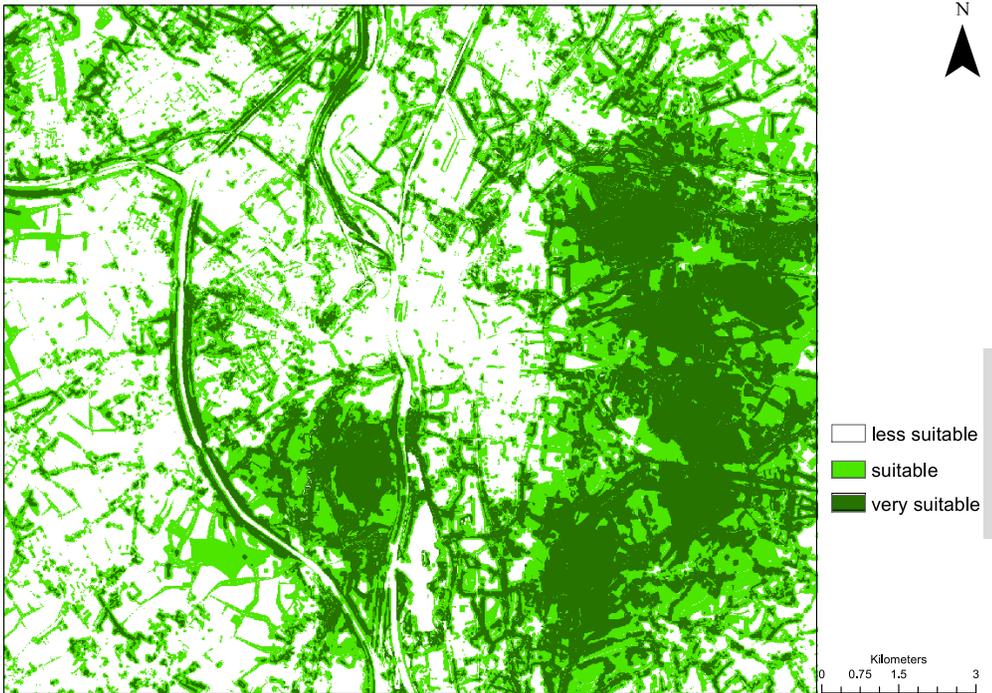


Figure 5.14: Reclassified HS map using the M_c graph based on the probability of occurrence (Eq. (5.7)) (HSI thresholds of 0.7 and 0.9).

5.4.4 Discussion

We selected 12 ECs that are important for the Eagle Owls in the study area and selected appropriate radii of influence (Table 5.4). Literature and expert knowledge are underpinning the resulting HS map. In order to consistently transform the qualitative information into SI values, we chose 0.5 as neutral SI value. As can be seen in Table 5.4, 11 out of the 12 ECs can lead to SI values greater than 0.5, so we mainly accounted for positive impacts on the Eagle Owls. Given the available GIS data, only major roads could be assigned an SI value lower than 0.5. For what concerns the other ECs, there is always some potential positive influence, mainly prey related. The highest α_e values were assigned to quarries and vegetation structures, followed by forest edges, agricultural land use related to trees and water bodies. It is important to note that one has to carefully select and motivate the considered ECs and their corresponding SIs, but a sound, objective underpinning of these choices is mostly not possible (Ray and Burgman, 2006; Zajac et al., 2015). Still, we motivated our choices as much as possible. To keep a neutral value of 0.5 when aggregating the SI values per pixel, we relied on a uninorm aggregation function. This has, to the best of our knowledge, never been done before in this context.

5 GPS tracking data were used to evaluate the resulting HS map. This type of data is not commonly used for that purpose. Typically, validation is based on an independent set of data on the sites where the species is present and where it is absent. Unfortunately, this type of data was not available for our species in the region. Therefore, we combined two established methodologies. In the one of Reynolds-Hogland and Mitchell (2007), habitat selection was estimated on the basis of GPS tracking data. First, the fixes were transformed into a PDF and a home range, after which, the Ivlev's electivity index was calculated to quantify the habitat selection. Since we wanted to evaluate the selection of pixels corresponding to each HSI bin, we could use this methodology. In Hirzel et al. (2006), validation of HS maps are performed on the basis of (independent) presence-only data. Here, the Manly's habitat selection index is considered. Therefore, we also considered this index in our analysis. Moreover, in Hirzel et al. (2006), the idea of a moving window instead of choosing a specific number of HSI bins to arrive at smoother graphs is presented. The latter are eventually used to interpret and validate the obtained HS map. As such, we used already existing methodologies which were carefully selected and combined for validation.

Many gps fixes in our dataset, however, are located close to each other in a limited number of pixels, which consequently all have a very high probability of occurrence. Furthermore, they are located in highly suitable pixels, close to a possible nesting site. As a consequence, the results are biased towards high HSI values so that we considered the shape of the graphs to interpret the results. Moreover, since we compared the pixels that were probably visited with those available in the

landscape, the values of the indices depend on the definition of the study area. We defined our study area as the rectangular area enclosing all data points. If information on habitat suitability is needed in more detail, the study area might be limited to, for example, the neighborhood of the quarries. In this case, the interpretation of the HS map and the definition of (un)suitable can be done in function of the area of interest.

When considering the graphs of the average HSI value versus Ivlev's electivity index I_c and Manly's habitat selection index M_c (Fig. 5.12 and 5.13), we may conclude that pixels with an HSI value lower than 0.55 are unsuitable, those with an HSI value between 0.55 and 0.7 are suitable, while those with an HSI value greater than 0.7 are highly suitable. When considering the graphs of average HSI value versus Ivlev's electivity index \tilde{I}_c and Manly's habitat selection index \tilde{M}_c , which are directly based on the probability of occurrence of the species, we alternatively concluded that pixels with an HSI value lower than 0.7 are unsuitable, those with an HSI value between 0.7 and 0.9 are suitable, while those with an HSI value greater than 0.9 are highly suitable, and are probably located near a possible nesting site.

As such, both Ivlev's electivity index and Manly's habitat selection index lead to the same conclusions. Yet, one arrives at a different interpretation when using the probability of occurrence (Eq.(5.7)) instead of the home range (Eq.(5.6)) to describe the area used by the species. In the latter case, the influence of the individual fixes is reduced since less frequently visited pixels are weighed equally than more frequently visited ones. Consequently, the indices based on the home range increase more gradually. We believe, however, that the use of the probability of occurrence instead of the home range is more appropriate since it better accounts for the species' preferences.

Given the fact that Eagle Owls almost exclusively use areas with HSI values greater than 0.7, and mostly areas with HSI values greater than 0.9, we may conclude that our HS map is well designed. The HS map clearly reflects the suitability of the study area for the Eagle Owls.

HS maps are useful for representing the spatial distribution of the ECs that have the strongest influence on the occurrence and abundance of a species in a simple and understandable way (Ortigosa et al., 2000). Still, they cannot fully explain the presence and abundance of a species in its territory, mainly because the ECs are not the only factors that determine species occurrence. Consequently, their value lies in describing both the quality and quantity of the available habitat for the species under consideration (Ortigosa et al., 2000), while the SIs help to understand the species niche requirements (Hirzel et al., 2006).

From a policy-making point of view, the developed HS map can first of all be used to identify the areas of importance to the species in the study area. The HS map of the study area can be used to select the parcels where restoration and/or conservation measures should preferably be enforced, as it indicates where land use

changes, for instance, would substantially affect the population at stake. Moreover, it could be used to guide field studies for locating possible sources of poisoning in the study area, motivated by the high concentrations of toxic substances that have been found in the carcasses of the owls (van den Brink and Jansman, 2005). Here, samples are best taken in areas that are highly suitable and found to attract the owls.

5.5 Environmental characteristics and locations of possible hunting grounds

5.5.1 Introduction

It is known that top predators accumulate toxic substances in their body and are, as such, indicator species for environmental pollution (Lourenço et al., 2011). Hence, the high concentrations of PCBs and DDEs could be explained by the Eagle Owls' presence at the top of the local food chain (van den Brink and Jansman, 2005). Since identifying possible sources of poisoning can only be done after gaining a full understanding of this species' hunting behavior, one should try to identify the locations where and the periods during which the owls typically hunt. This was done by analyzing the tracking data and consulting literature.

Given the relatively low temporal resolution of the tracking data and the fact that the owls are relatively inactive during the day, it is hard to identify hunting behavior on the basis of a spatio-temporal analysis of the tracking data. Yet, in light of the facts mentioned in Section 5.3, non-hunting behavior can be identified since it is likely to occur when females are incubating their eggs, during the day or when a large, directed movement is registered. Therefore, we first investigated the features of the female owls' movements during the breeding season. Secondly, we studied these features for all tracked individuals and how they change during the night. This allowed us to select the periods when (some) individuals do not hunt, as such enabling us to reduce the number of data points used for subsequent analysis. Thirdly, we removed the fixes belonging to a long, directed movement. After reducing the size of the data set by making use of this kind of expert knowledge (Subsection 5.5.2.1), we assumed that the remaining fixes are linked to possible hunting grounds.

When considering the home range of animals, there are areas that are used more frequently than others, typically in the surroundings of the nesting sites and hunting areas (Delgado and Penteriani, 2007). Since only fixes linked with possible hunting grounds are now remaining in our data set, a hotspot analysis was conducted in order to locate these areas for the tracked individuals (Environmental

Systems Research Institute (ESRI), 2016) to identify focal areas for further study (Subsection 5.5.2.2).

We also investigated the relative importance of the prevailing ECs of the owls' hunting grounds (Subsection 5.5.2.3). Still, additional field studies are needed to unravel the details of possible poisoning sources, which is, therefore, beyond the scope of this thesis.

5.5.2 Methods

5.5.2.1 Selecting fixes linked to possible hunting grounds

We computed five features of the owls' movement behavior, namely (1) the total covered distance during one night, (2) the distance between the nesting site and every registered location, (3) the covered distance per time interval (average speed in between two fixes), (4) the direction from a registered location to the nesting site and (5) the turning angle. Throughout this manuscript, the unit of distance is meter, the unit of time is hour and the unit of angles is degree. The direction varies between 0° and 360° , where 0° points to the east and 90° to the north, while the turning angle varies between 0° and 180° .

As a first step, the fixes corresponding to females B and C were split into two subsets, a first one containing the fixes during the breeding season, and a second one containing those outside the breeding season. This allowed us to compare those two behavioral stages. To make sure that the samples are independent, we calculated the nightly average values for features (2)–(5). This is straightforward for features (2) and (3), but special attention has to be given to features (4) and (5). We followed the procedure outlined in Cain (1989), where each angle θ is envisaged as a vector with unit length in a fixed reference coordinate system. Therefore, every θ was first transformed to a coordinate vector $(\cos(\theta), \sin(\theta))$. These pairs were then averaged per night and the resulting average was subsequently transformed to polar coordinates $(\bar{r}, \bar{\theta})$, where \bar{r} represents a nightly average radius and $\bar{\theta}$ a nightly average direction (Cain, 1989). Then, per subset, the mean, median, standard deviation and range were calculated for features (1)–(3) and mean, angular deviation and range for features (4) and (5). The mean for features (4) and (5) was again computed according to Cain (1989), while the angular deviation equals $\sqrt{2(1 - \bar{r})}$. To decide whether the means and medians are significantly different across the behavioral stages, a t-test and Mann Whitney test respectively, were performed for features (1)–(3) (Student, 1908) and a Mardia test, designed for circular data, was performed for features (4) and (5) (Batschelet, 1972; Mardia, 1972).

As a next step, we investigated whether the time of the day influences the movement features. Since owls are nocturnal predators, hunting their prey only in dark-

ness (März, 1958), we may expect that they are mainly active from sunset to sunrise. Since the time of sunset and sunrise varies during the course of the year, we expressed the time stamp of each location as the fraction of the night that had elapsed between sunset and sunrise in steps of 0.05, with zero representing sunset and one representing sunrise.

We chose to resort to metric-based methods to identify the locations where the owls just pass through and are hence not related to foraging or resting. Any further distinction between the latter two types of behavior is hard to make since owls are equally inactive during both. These metric-based methods were preferred over any of the others in the categories mentioned in Subsection 2.5.2, since it is hard to define behavioral states a priori, while BCPA was not an option since the dataset does not constitute a continuous time series of fixes, but is made up of fixes collected during every third night (cfr. Section 5.2). Likewise, mechanistic movement models would also suffer from the latter limitations and fitting a comprehensive model to the data set is not relevant given the scope of this section.

We assumed that an owl is not foraging or resting if it moves with a relatively high speed and continues its movement in its current direction, which can be verified by studying the covered distance per time interval (v) and the turning angle (α) (features (3) and (5)) (Gurarie et al., 2009). To eliminate the influence of scale, both variables were normalized by dividing them by their maximum value recorded during the entire study period. The normalized (v, α) couples were clustered for every individual separately using DBSCAN (Ester et al., 1996), a density-based clustering method, and taking the Euclidean distance as distance metric. We used the implementation of this method available in Mathematica (Wolfram Research Inc., 2018c).

By relying on expert knowledge and DBSCAN, we were able to reduce the number of fixes so that only fixes likely to be linked with foraging or resting remained in the data set. These fixes were therefore assumed to be located in possible hunting grounds and were used for further analysis.

5.5.2.2 Identifying the important areas

Since the owls frequently visit the same locations, it is interesting to perform a hotspot analysis of the visited locations to quantify their intensity of use (Environmental Systems Research Institute (ESRI), 2016). Here, the study area is overlain with a raster and the number of fixes in each cell are counted. The appropriate cell size was automatically determined by ArcGis. The outcome of such analysis allows us to identify areas for further investigation, based on the pollution of the possible hunting grounds for the tracked animals. This relies on the expectation that the most frequently visited locations have, relatively speaking, the largest impact on the prey - and hence on the poisonous compound - consumption by the

Eagle Owls. To get an idea of the space used by each individual, the convex hull enclosing its fixes is given. Because convex hulls are sensitive to outliers, sample size and autocorrelated data (Millsbaugh and Marzluff, 2001), they were, however, not considered for further analysis.

5.5.2.3 Identifying the important environmental characteristics

In order to explain either a high or low visit frequency, we investigated the ECs of the remaining fixes. The considered ECs (Table 5.5) are similar to the ones selected for the construction of the HS map (cfr. Section 5.4 and Table 5.4). Therefore, we refer to Sections 5.3 and 5.4.3 to motivate our choice. Here, however, EC2, EC7 and EC9 are subdivided into two or three more specific ECs. This division was not meaningful when selecting the ECs for constructing the HS map since we then relied on literature and expert knowledge that is too coarse-grained to underpin such a distinction. Nevertheless, more detail can be added during the analysis of tracking data. We made, for example, a distinction in water bodies because banks of small water courses and lakes are more natural, and as such might be more suitable for the owls.

We are particularly interested in identifying the ECs that are determinants for the presence of the tracked individuals. Therefore, we resort to ecological-niche factor analysis (ENFA), a factor analysis building upon Hutchinson's concept of an ecological niche, to determine the most important ECs (Hirzel et al., 2002; Basille et al., 2008). In Hutchinson (1957), the ecological niche is defined as the hypervolume in the multidimensional EC space where the species can maintain a viable population. For a detailed mathematical description of ENFA, we refer the reader to Hirzel et al. (2002). Here, we only outline its major characteristics.

Since ENFA is based on numerical computations, categorical information cannot be processed easily, so that every category has to be considered separately. More precisely, the study area was overlaid with a raster consisting of pixels (or cells) organized into rows and columns. In our study, a pixel represents an area of 5 m by 5 m. First, a Boolean map was constructed for each EC where the pixel value equals one if the EC is present and zero if it is not, after which a new map was created where every pixel value represents the minimum distance to the EC under consideration. This was done for the 16 ECs listed in Table 5.5. Thus, the minimum distances from each pixel to each of the 16 ECs are the variables used in ENFA. Next, the fixes of the owls were encoded in a raster data layer, where a pixel value of one was assigned if it contained at least one fix (*i.e.* a visited pixel), and zero otherwise.

ENFA involves computing directions in the space of variables in such a way that the first direction is defined by the point obtained by averaging per EC the distances between the visited pixels and each EC, and the point obtained by averaging per

Table 5.5: Overview of the considered environmental characteristics (EC).

| EC | Description |
|------|--|
| EC1 | Quarry |
| EC2a | Pasture |
| EC2b | Arable land |
| EC3 | Forest edge |
| EC4 | Agricultural land use related to trees |
| EC5 | Town center |
| EC6 | Industrial area |
| EC7a | Isolated vegetation |
| EC7b | Linear vegetation |
| EC8 | Difference in altitude |
| EC9a | Small water course |
| EC9b | Large water course |
| EC9c | Water surface |
| EC10 | Small road |
| EC11 | Non-small road |
| EC12 | Building outside town center |

5 EC the distances between all pixels in the study area and each EC. The other directions maximize the ratio of the variance of distances to the ECs from the visited pixels and the variance of available distances to the ECs across the study area. The first direction is chosen to account for all marginality, the others are chosen to maximize specialization (Hirzel et al., 2002). Hence, marginality measures the deviation of the species' niche from the average prevailing conditions, while specialization quantifies the narrowness of the niche (Fig. 2.10). The higher the specialization, the more restricted the niche is in that dimension. The specialization directions are defined by vectors that are orthogonal to the marginality vector, though not mutually, and extract most of the specialization (Basille et al., 2008). If several environmental variables are considered simultaneously, ENFA embodies the concept of an ecological niche. Here, this niche is a subset of pixels where the species has a probability to occur. Every axis of the multidimensional niche (one per EC) can be characterized by indices of marginality and specialization. As such, ENFA allows to extract the combination of the original variables (the ECs) for which the species shows most of the marginality and specialization (Hirzel et al., 2002). The relationships between these variables and both the marginality and specialization factors are expressed by so-called loadings. The larger the absolute value, the more the EC contributes to the considered marginality or specialization factor, while the sign only has meaning for the marginality factor. Given the fact that a study area is used as reference in ENFA, its results will strongly depend on the delineation of this area (Hirzel et al., 2002), and the latter should be decided by keeping the research objectives in mind. We thus limited our study area to a rectangle enclosing all data points.

5.5.3 Results

5.5.3.1 Selecting fixes linked to possible hunting grounds

Female B was tracked for 12 nights during the breeding season and for a total of 22 nights outside the breeding season, while this is respectively 10 and 24 nights for female C. Female A was not considered during this analysis since fixes for this individual were only obtained outside the breeding season. As can be seen in Table 5.6, the total distance covered during the breeding period was about half of the distance covered outside that period, indicating a much lower hunting appetite during the breeding season. Likewise, the average speed between two consecutive fixes was significantly lower during the breeding season. Female C tended to stay closer to the nesting site during the breeding season, whereas female B was always located near its nesting site during the entire study period. The direction towards the nesting site was significantly affected by the season while the turning angle was not.

Since both females stayed near the nesting site and the total covered distances were rather low during the breeding season, our findings support those of Heintzenberg (2008), in the sense that females do not hunt when incubating their eggs or taking care of their fledglings. As such, all data points of females B and C collected from February 15, 2011 until the end of March, 2011 were excluded from further analysis, reducing the number of locations from 13,091 to 10,070.

The mean and range of each feature as a function of the fraction of the night can be seen for every individual in Figs. 5.15(a)–5.15(e) when tracking data from at least five nights were available. It is obvious that the owls became active at sunset since at that moment the distance to the nesting site changed and both the average speed between two fixes and the total covered distance became positive. Right before or after sunrise, the owls returned to the vicinity of their nesting site and their average speed between two fixes dropped. Probably because only a few observations were used to calculate the mean and range near the ends of the considered time interval, the total covered distance does not reach a clear plateau. At sunset and sunrise, a preferential direction from the nesting site can be noticed for pair B (male) and pair C (female and male), indicating that these individuals preferentially explored certain sectors surrounding their nesting site at those times of the day. The turning angle did not depend on the elapsed fraction of the night, as their ranges enclose all possible values, irrespective of time.

The outcomes of this analysis confirm the findings of März (1958), *i.e.* that owls are most active from sunset to sunrise. As such, all fixes that were collected before sunset and after sunrise were excluded from further analysis, which further reduced the total number of fixes to 8,620.

Using DBSCAN with the remaining (v, α) pairs, we ended up with two clusters

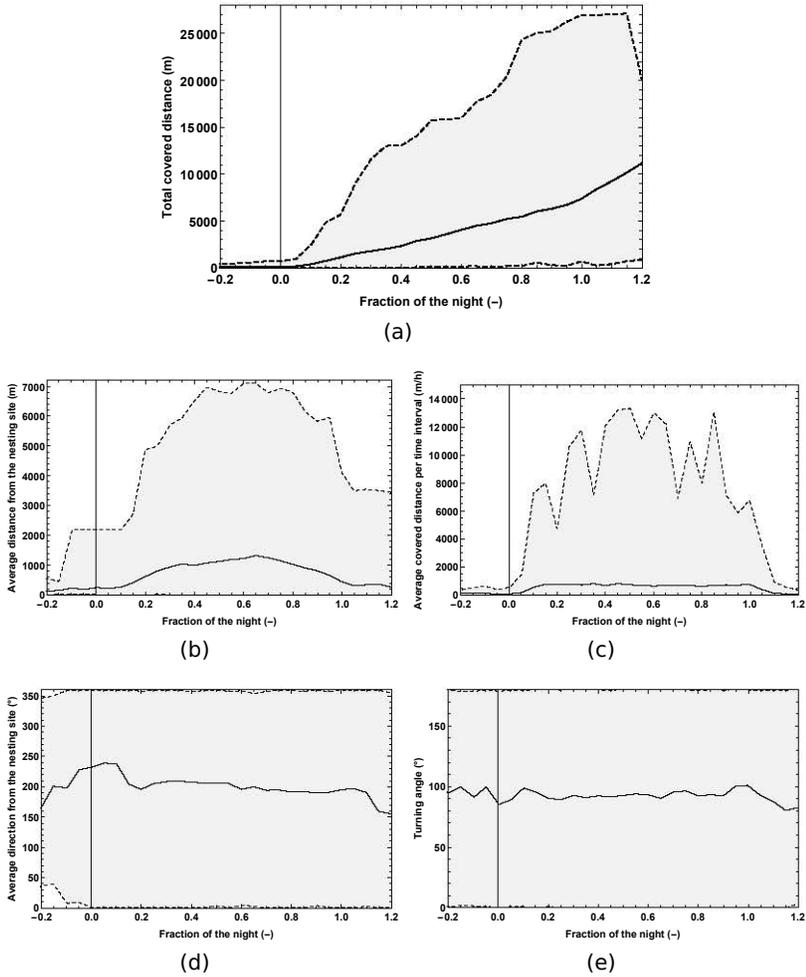


Figure 5.15: The mean (solid line) and range (dashed lines) of five movement feature ((a) total covered distance, (b) distance from the nesting site, (c) covered distance per time interval, (d) average distance from the nesting site and (e) turning angle) as a function of the fraction of the night. The mean and range were calculated over all individuals when tracking data from at least five nights were available.

5

Table 5.6: The mean, median, standard deviation (stdv) and range of movement features (1)–(3) and the mean, angular deviation (adv) and range of movement features (4)–(5) in and outside the breeding season for females B and C. * indicate that means/medians are significantly different between both seasons.

| Individual | | pair B Female | | pair C Female | |
|---|--------|------------------|---------------------|------------------|---------------------|
| Movement feature | | breeding season | non-breeding season | breeding season | non-breeding season |
| (1) Total covered distance during one night (m) | mean | 1140.82 * | 2421.51 * | 2435.29 * | 5454.30 * |
| | median | 1001.19 * | 2108.42 * | 2277.14 * | 5646.20 * |
| | stdv | 1089.03 | 1681.47 | 2171.63 | 3334.36 |
| | range | [31.28, 2923.39] | [175.66, 5328.46] | [42.99, 5156.94] | [87.50, 14013.6] |
| (2) Average distance from the nesting site (m) | mean | 154.49 | 244.71 | 198.53 * | 466.19 * |
| | median | 130.95 | 183.89 | 137.74 * | 363.05 * |
| | stdv | 88.2 | 238.97 | 158.92 | 305.06 |
| | range | [43.39, 291.86] | [37.27, 1099.2] | [62.04, 530.37] | [89.92, 1157.66] |
| (3) Average covered distance per time interval ($m h^{-1}$) | mean | 127.14 * | 242.13 * | 178.19 * | 366.54 * |
| | median | 119.00 * | 182.03 * | 182.13 * | 374.98 * |
| | stdv | 92.49 | 154.96 | 131.66 | 206.96 |
| | range | [5.21, 245.80] | [34.25, 527.90] | [3.92, 349.42] | [12.49, 932.53] |
| (4) Average direction towards the nesting site ($^{\circ}$) | mean | 170.42 * | 192.61 * | 341.07 * | 268.93 * |
| | adv | 50.34 | 64.37 | 60.49 | 70.57 |
| | range | [114.99, 243.15] | [107.14, 282.92] | [66.14, 354.14] | [114.16, 350.51] |
| (5) Average turning angle ($^{\circ}$) | mean | 112.07 | 111.15 | 100.05 | 96.23 |
| | adv | 48.85 | 51.34 | 50.11 | 52.02 |
| | range | [83.76, 156.81] | [86.08, 149.53] | [90.94, 112.80] | [73.47, 120.54] |

for each individual. As can be seen in Table 5.7, cluster 1 is characterized by a relatively low mean and median value of the covered distance per time interval v , while the values of turning angle α are somewhat higher, although rather similar to those in cluster 2. For what concerns α , all possible values occur in both clusters.

A t-test could not be performed to verify whether the mean values of v differed significantly between the clusters, since v is not normally distributed. The median values, however, were found to be significantly different while the values of α differed significantly between the clusters for pairs A and C. As such, we may conclude that cluster 1 contains (v, α) pairs with a low average speed and a somewhat higher turning angle compared to cluster 2, so that foraging or resting can be related to cluster 1 and locations where the owl was just passing through correspond to cluster 2 .

As such, we only considered the fixes corresponding to cluster 1 for locating possible hunting grounds, which reduced the number of data points further from 8,620 to 8,057.

5.5.3.2 Identifying the important areas

The convex hulls enclosing the remaining visited locations are shown for every individual separately in Fig. 5.16. Furthermore, the results of the hotspot analysis, indicating the number of visits in each cell, is shown. The most frequently visited

Table 5.7: The mean, median, standard deviation (stdv) and range of the average speed (v) and the mean, angular deviation (adv) and range of the turning angle (α) in between two fixes for every individual and cluster. * indicate that means/medians are significantly different between both clusters, while (^{NA}) indicates that one or more assumptions of the statistical test were not fulfilled.

| Movement feature | | v (m h ⁻¹) | | α (°) | |
|------------------|--------|--------------------------|-----------------------|----------------|----------------|
| Individual | | cluster 1 | cluster 2 | cluster 1 | cluster 2 |
| A Male | mean | 267.62 ^{NA} | 2293.01 ^{NA} | 96.43 * | 70.06 * |
| | median | 110.03 * | 1948.96 * | | |
| | st/adv | 354.82 | 1155.63 | 52.42 | 50.15 |
| | range | [0.50,1928.27] | [7003.39,7125.23] | [0.01, 179.98] | [0.60, 179.74] |
| A Female | mean | 192.08 ^{NA} | 1863.23 ^{NA} | 101.63 * | 63.13 * |
| | median | 88.12 * | 1634.72 * | | |
| | st/adv | 241.38 | 1003.98 | 52.89 | 51.01 |
| | range | [0.11, 1463.64] | [555.24, 5713.15] | [0.36, 179.90] | [0.07, 179.42] |
| B Male | mean | 637.83 ^{NA} | 3783.68 ^{NA} | 84.58 | 84.32 |
| | median | 367.85 * | 3480.79 * | | |
| | st/adv | 741.68 | 1619.22 | 51.70 | 57.06 |
| | range | [0.90, 3872.61] | [1332.81,11767.9] | [0.15, 179.99] | [0.78, 179.94] |
| B Female | mean | 64.24 ^{NA} | 749.96 ^{NA} | 114.75 | 87.65 |
| | median | 31.20 * | 629.29 * | | |
| | st/adv | 81.86 | 498.97 | 50.19 | 54.16 |
| | range | [0.64,498.21] | [92.83,2842.38] | [0.46,179.84] | [0.02,179.22] |
| C Male | mean | 364.26 ^{NA} | 3263.32 ^{NA} | 98.34 * | 65.79 * |
| | median | 171.12 * | 2400.74 * | | |
| | st/adv | 510.47 | 2572.77 | 53.71 | 49.20 |
| | range | [0.54, 3428.03] | [465.14,13315.80] | [0.10, 180.] | [0.60, 179.92] |
| C Female | mean | 177.33 ^{NA} | 933.70 ^{NA} | 97.46 * | 93.02 * |
| | median | 100.53 * | 854.22 * | | |
| | st/adv | 186.15 | 624. | 55.19 | 45.62 |
| | range | [0.05, 1012.50] | [3.42,3839.26] | [0.25, 179.90] | [0.25, 179.55] |

areas were more or less situated near the nesting sites. In the outskirts of the convex hull, one finds locations that were visited only once or twice, though there are also some more frequently visited areas (3–50 times visited). Especially the latter might have been the more distant hunting grounds of the tracked individuals.

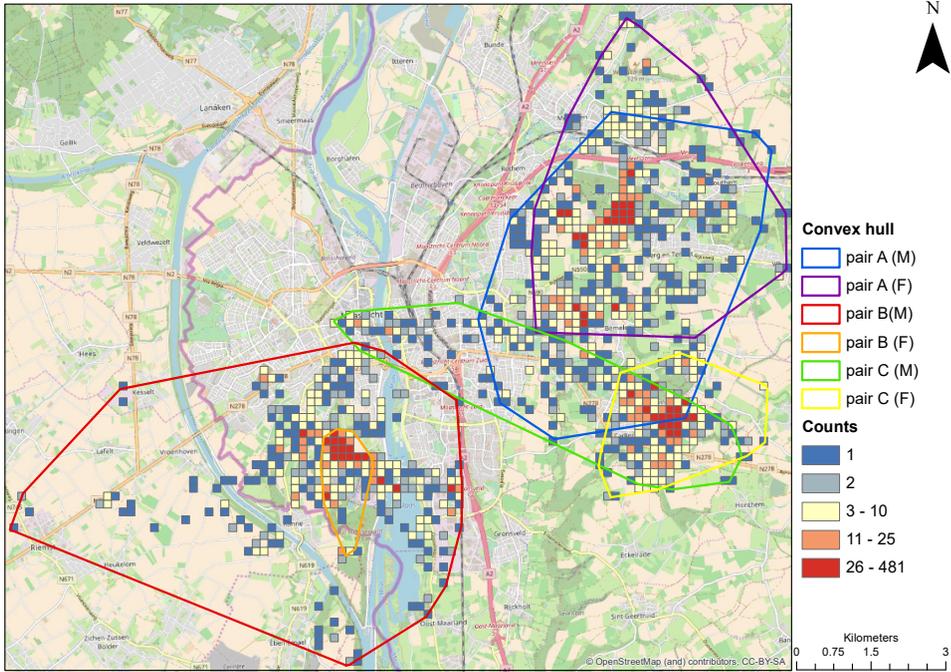


Figure 5.16: The convex hulls enclosing the visited locations, together with the visit frequency (Source: OpenStreetMap contributors). M indicates male and F female.

5.5.3.3 Identifying the important environmental characteristics

Since ENFA uses the study area as a reference, we first studied the distribution of the ECs in the study area and compared it to that of the fixes. More specifically, we computed the relative frequency of occurrence of EC2a–EC6, EC9b and EC9c. These ECs are represented by polygons so they have an area (Table 5.8). For the study area, this was done by dividing the area covered by the EC by the total area. The ECs were sometimes overlapping, and in some regions (7.6%), none of the considered ECs was present. Consequently, 14.4% of the fixes could not be linked with EC2a–EC6, EC9b and EC9c. The average distances to each EC from both any pixel in the study area and the fixes can be seen in Table 5.9.

Table 5.8 indicates that pastures (EC2a), arable land (EC2b) and town centers (EC5) are the main land uses in the study area. However, pastures (EC2a) and

forest edges (EC3) were used most intensively. From Table 5.9, it can be seen that the average distances from quarries (EC1), pastures (EC2a), forest edges (EC3) and differences in altitude (EC8) to the fixes are noticeably smaller compared to those to an arbitrary pixel. Consequently, it could be concluded that these ECs attracted the owls. Likewise, agricultural land use related to trees (EC4), industry (EC6), isolated vegetation (EC7a) and water bodies (EC9a, EC9b and EC9c) might also have been of interest to the owls. Town centers (EC5), linear vegetation (EC7b) and roads (EC11), on the other hand, were avoided. This was, however, unexpected for EC7b. The other ECs, *i.e.* arable land (EC2b), small roads (EC10) and buildings outside town centers (EC12), neither attracted, nor repelled the owls.

Tables 5.8 and 5.9 give a preliminary idea about the importance of each EC. ENFA, however, also takes into account the correlation between the ECs. Before turning to ENFA, the correlation matrix of the 16 ECs was investigated since any dimensionality reduction is only useful if there exists a relationship between the ECs (Hirzel et al., 2002). Given that correlation coefficients of up to 0.69 were found, we determined that conducting ENFA should be useful.

Each marginality and specialization factor has a corresponding eigenvalue, which represents the variance explained by this factor (Fig. 5.17). The first eigenvalue corresponds to the marginality and the others to the specialization factors. In order to select the factors with relevant information, we should look for a sharp drop in the curve showing the cumulative explained variation (Fig. 5.17) (Yong and Pearce, 2013). Such a clear drop is present between the second and third eigenvalue, so we considered the first two factors for further analysis, *i.e.* the marginality and the first specialization factor. Together, they explain about 50% of the variance in the data. The loadings of each EC for the first two factors can be found in Table 5.10. To judge whether an EC is important for a given factor or not, we mutually compared the loadings.

Marginality measures the deviation of a niche from the average available conditions (Basille et al., 2008). Considering the marginality loadings in Table 5.10, the marginality factor is mainly determined by the quarries (EC1) so that we may conclude that the latter functioned as the main attractor for the tracked individuals. Their importance can be explained by the fact that the nesting sites were located

Table 5.8: The relative frequency of occurrence of EC2a, EC2b, EC3, EC4, EC5, EC6, EC9b and EC9c for the fixes and across the study area, and the ratio of the former versus the latter.

| | Fixes | Study area | Ratio |
|--|-------|------------|-------|
| Pasture (EC2a) | 0.305 | 0.193 | 1.58 |
| Arable land (EC2b) | 0.153 | 0.308 | 0.50 |
| Forest edge (EC3) | 0.295 | 0.095 | 3.11 |
| Agricultural land use related to trees (EC4) | 0.044 | 0.065 | 0.68 |
| Town center (EC5) | 0.154 | 0.289 | 0.53 |
| Industry (EC6) | 0.031 | 0.047 | 0.66 |
| Large water course (EC9b) | 0.011 | 0.013 | 0.85 |
| Water surface (EC9c) | 0.007 | 0.033 | 0.21 |

Table 5.9: The average distance (m) from the fixes to each EC, the average distance (m) from any pixel in the study area to each EC, and the ratio of the former versus the latter.

| Distance to | Fixes | Study area | Ratio |
|---|--------|------------|-------|
| Quarry (EC1) | 709.13 | 3713.79 | 0.19 |
| Pasture (EC2a) | 46.81 | 98.48 | 0.48 |
| Arable Land (EC2b) | 152.04 | 163.90 | 0.93 |
| Forest edge (EC3) | 60.06 | 230.67 | 0.26 |
| Agricultural land use related to trees (EC 5) | 236.84 | 298.23 | 0.79 |
| Town center (EC5) | 320.68 | 231.81 | 1.38 |
| Industry (EC6) | 666.59 | 866.41 | 0.77 |
| Isolated vegetation (EC7a) | 143.32 | 162.71 | 0.88 |
| Linear vegetation (EC7b) | 180.33 | 143.99 | 1.25 |
| Difference in altitude (EC8) | 62.37 | 137.42 | 0.45 |
| Small water course (EC9a) | 339.47 | 406.87 | 0.83 |
| Large water course (EC9b) | 950.17 | 1273.28 | 0.75 |
| Water surface (EC9c) | 294.49 | 460.81 | 0.64 |
| Small road (EC10) | 58.54 | 61.69 | 0.95 |
| Road (EC11) | 329.93 | 211.94 | 1.56 |
| Building outside town center (EC12) | 310.79 | 342.07 | 0.91 |

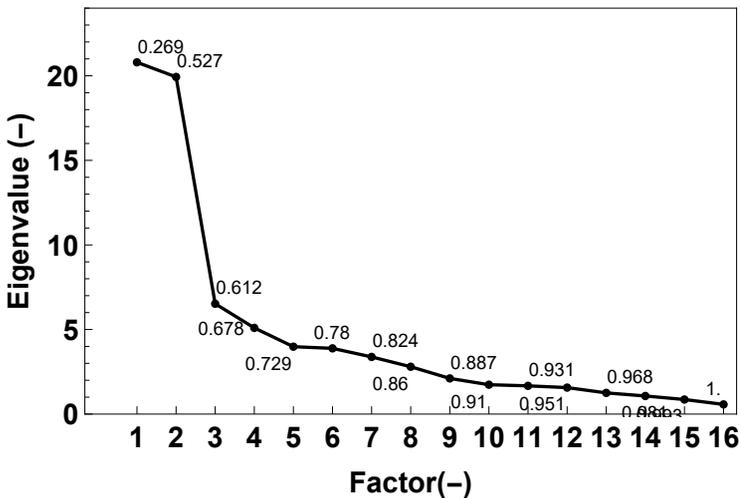


Figure 5.17: The eigenvalues of the 16 factors obtained using ENFA together with the cumulative explained variance.

there and as such, their surroundings were often visited and provided nearby hunting opportunities. Pasture (EC2a), forest edges (EC3), differences in altitude (EC8) and water surfaces (EC9c) also have negative loadings, albeit to a lesser extent, and seems to have attracted the tracked Eagle Owls, which agrees with the main finding on the basis of Tables 5.8 and 5.9. Additionally, agricultural land use related to trees (EC4), industry (EC6), isolated vegetation (EC7a) and water courses (EC9a and EC9b) seems to have been of interest, although less pronounced. The presence of arable land (EC2b), small roads (EC10) and buildings outside the town

center (EC9b) did not have any influence (values close to zero), whereas town centers (EC5), linear vegetation (EC7b) and roads (EC11) were avoided (positive loadings). As such, it is clear that pastures (EC2a) were preferred over arable land (EC2b) and that water surfaces (EC9c) were relatively more important than water courses (EC9a and EC9b). Furthermore, areas with a lot of human activity (town centers (EC5) and roads (EC11)) were avoided. Industrial areas (EC6), however, attracted the owls and small roads did not have any influence. Our results also suggest that differences in altitude (EC8) and forest edges (EC3) were frequently visited, and were probably used as observation posts.

Specialization measures the narrowness of the niche (Basille et al., 2008). The first specialization factor is mainly determined by forest edges (EC3). Quarries (EC1), pastures (EC2a) and agricultural land uses related to trees (EC4) also significantly narrow the ecological niche.

5.5.4 Discussion

In this section, we identified the ECs that are the most important for the possible hunting grounds of a population of Eagle Owls in the province of Limburg on the basis of tracking data of six individuals. First, we removed the fixes corresponding to locations where owls were probably not hunting, on the basis of an analysis of five relevant movement features. Given the relatively low temporal resolution of the data set and the fact that the movement features are similar for different kinds of behavior (*i.e.* foraging and resting), it was difficult to determine the owls' behavior exclusively on the basis of this analysis. More precisely, we were only able to distinguish the locations where the individuals were just passing through from the ones where they were foraging or resting. The latter locations were assumed to be located at possible hunting grounds and were used to gain insight into the ECs that are important for the hunting grounds.

After removing the fixes where the owls were likely not hunting, we studied the use of space of the individuals. The areas near the nesting sites were visited most often, but there were also other, more distant, frequently visited locations that are scattered across the study area. These locations probably belong to more distant hunting grounds. Studying frequently visited locations might help to locate possible poisoning sources within the study area because it may be expected that prey originating from such areas make up a relatively large portion of the owl's diet. As such, it is recommended that any further investigation is guided by the outcomes of the hotspot analysis in combination with the results of ENFA.

ENFA was performed to pinpoint the most important ECs which determine the selection of hunting grounds by the tracked individuals. Generalizing conclusions drawn for these individuals to the population of Eagle Owls in the study area, we infer from the resulting loadings that the (former) quarries (EC1) are of utmost impor-

Table 5.10: The loadings of every EC for the first two factors obtained by ENFA. The explained variance of each factor is indicated between parentheses.

| | Marginality factor (26.9%) | Specialization factor 1 (25.8%) |
|--|----------------------------|---------------------------------|
| Quarry (EC1) | -0.70 | -0.31 |
| Pasture (EC2a) | -0.22 | -0.17 |
| Arable land (EC2b) | -0.01 | -0.04 |
| Forest edge (EC3) | -0.28 | 0.91 |
| Agricultural land use related to trees (EC4) | -0.10 | -0.14 |
| Town center (EC5) | 0.16 | 0.01 |
| Industry (EC6) | -0.15 | 0.00 |
| Isolated vegetation (EC7a) | -0.10 | -0.04 |
| Linear vegetation (EC7b) | 0.13 | 0.09 |
| Differences in altitude (EC8) | -0.30 | -0.02 |
| Small water course (EC9a) | -0.12 | 0.03 |
| Large water course (EC9b) | -0.17 | 0.07 |
| Water surface (EC9c) | -0.27 | -0.01 |
| Small road (EC10) | -0.04 | -0.05 |
| Road (EC11) | 0.30 | -0.04 |
| Building outside town center (EC12) | -0.04 | 0.08 |

tance since they are the only nesting locations found in our study area (Wassink, 2014c; van Lierop and Janssen, 2014). Furthermore, forest edges (EC3), and to a lesser extent, pasture (EC2a), stand out as the most important land uses. Moreover, observation posts are most likely located in the neighborhood of differences in altitude (EC8), and roads (EC11) clearly are avoided by the owls. It might be worthwhile to take into account these findings when conducting field campaigns to identify possible sources of poisoning.

In addition, it can also be concluded that the neighborhoods of water bodies are visited relatively often. More specifically, in our study area, water surfaces (EC9c) are relatively more interesting for the species than water courses (EC9a and EC9b). Agricultural land use related to trees (EC4) has a negative loading for the marginality factor and a relatively high loading for the first specialization factor (Table 5.10), so that we may conclude that this land type is also important to the Eagle Owls' hunting. From the ENFA, it emerges that areas with a lot of human activity (town centers (EC5) and roads (EC11) are avoided. This contrasts with sightings of Eagle Owls in cities (Alerstam et al., 2003), but in our study area, there might already be sufficient numbers of prey, so that there is no need for the owls to hunt in urbanized areas. Industrial areas (EC6) seem to attract the owls, while small roads (EC10) do not seem to have any influence. This is probably because it is rather quiet at night near the premises, so that prey can be easily spotted. From our analysis, it also becomes clear that the owls can frequently be found in the vicinity of isolated vegetation (EC7a) and differences in altitude (EC8), whereas linear vegetation (EC7b) structures are not selected as possible hunting grounds. This was not expected since the latter might also serve as observation posts. The results obtained from ENFA are similar to the conclusions drawn from comparing the relative frequencies of land use in the study area with those of the fixes (Table 5.8), and from comparing the average distance to each EC from any pixel in the study

area and those of the fixes (Table 5.9).

5.6 Comparison

In this chapter, we considered two approaches to identify important ECs and areas for the Eagle Owls in Limburg, the Netherlands so that these areas and these ECs can be prioritized when further investigating possible sources of poisoning. Here, we mainly used already existing methodologies which were carefully selected and combined. The first approach exploits literature and expert knowledge to determine the importance of several ECs whereas, in the second approach, the tracking data is used. Concerning the ECs, both approaches show that quarries are, next to forest edges and water bodies, very important for the presence of the owls and that major roads are definitely avoided. Moreover, differences in altitude and vegetation structures are used as observation posts. From ENFA, it follows that the presence of differences in altitude is more important than the presence of vegetation. This is, however, assessed the other way around on the basis of literature and expert knowledge. Although to a lesser extent, town centers, industry and agricultural land uses related to trees attract the owls. According to the first approach, also pastures, arable land, small roads and buildings outside a town center are expected to have a positive influence. This, however, only holds for pastures according to the second approach, while the other ECs are not influencing the tracked individuals, neither positive nor negative.

Next to ECs, areas that are possibly linked to poisoning sources are determined. They are delineated by interpreting the HS map in the first approach and by performing a hotspot analysis based on the fixes related to possible hunting grounds in the second approach. In both approaches, the neighborhood of nesting sites is highlighted.

From a policy-making point of view, any new research project to locate the possible poisoning sources should be based on expert knowledge, involve an investigation of the history of the region and its land uses, and will require field visits and sampling campaigns of soil, water, vegetation or prey. Experts might, for example, know the available prey linked to several ECs and high PCB concentrations or simply know the hazards in the region. Their knowledge can also help to identify ECs, different from those identified using ENFA and literature, that attract the owls. Pair B, for example, often visited the recycling center and the waste water treatment plant in the study area. Historical records on the area might also help to locate the poisoning sources. Former dumping sites, for example, can still affect the natural environment up to this day.

Finally, to set up a sampling strategy, we advise that the results of the HS map, hotspot analysis and the relative importance of the ECs should be taken into account. More specifically, samples should preferably be taken near the nesting sites

or in areas that are highly suitable and frequently visited, and in areas with ECs that were found to be determinants for the Eagle Owl presence. For instance, samples should be taken near the quarries, near differences in altitude, in pastures and along forest edges, maybe even starting at locations where several ECs are present. Prey animals are the most interesting to sample. Hedgehogs, for example, have a small home range, are related to a small area and eat Earthworms (*Lumbricidae ssp.*), which absorb PCBs easily.

Even though a more detailed analysis of this kind is beyond the scope of this chapter, our analysis gives insight into the ECs that are linked with possible hunting grounds and, as such, with possible poisoning sources. Additionally, it provides managers with information on how to prioritize their future research. Such research may reveal that specific locations or ECs are responsible for the high PCB concentrations or that, on the contrary, PCBs are accumulated over the whole of the study area and no single point source of pollution can be identified.

Lastly, our results allow us to assess the suitability of regions with similar ECs, which would be useful given the increasing number of breeding pairs that has been reported over the last years in the Netherlands, Belgium and Germany (Natuurpunt, 2016; SOVON, consulted in 2017; Abo Wind, 2017). So, when conservation and/or resettlement measures would be implemented, this study may serve as a guideline for selecting and prioritizing the areas where their implementation would have the biggest impact. It should, however, be emphasized that the presented results are tailored to the land use and land cover occurring in the study area, so that the impact of potentially important ECs that are absent in the study area will not be accounted for in such an approach. Moreover, the results are tailored to the owls in the study area, who typically inhabit quarries and mainly prey on pigeons, crows, rats, mice and hedgehogs. Although quarries serve as nesting sites in many regions, also other nesting sites can be exploited, such as old nests of buzzards (*Buteo buteo*), industry buildings, forests, and so on (Wassink, 2011b). Furthermore, other prey, and, as such, other habitats, might be important for the species in other regions. Still, our study provides a well-founded basis, but limited adaptation might be needed for areas differing significantly from the one at stake in this chapter.

6

Contrasted post-fledging space use by Neotropical migratory birds within an urbanized landscape

6.1 Introduction

Persistent alarming declines in migratory songbird populations continue to motivate research exploring likely contributing factors that can inform conservation efforts. Previous studies on Neotropical migratory species have demonstrated negative effects from resultant habitat loss and fragmentation due to increasing urbanization on populations throughout the annual cycle.

This study was conducted in collaboration with the University of Delaware, which is data owner. In this study, Zachery Ladin analyzed the differential responses towards urbanization, while we calculated the TIBBs used in this analysis and built the simulation model. Section 6.4 was accepted for publication in *Movement Ecology* (Ladin et al., 2018).

Movement data have been used successfully to analyze space and resource use patterns in birds during breeding (Cooper et al., 2014; Boelman et al., 2015), non-breeding (Grzybowski, 1983; Hutto, 1985), and specifically, in response to anthropogenic factors (Masden et al., 2009; Boggie and Mannan, 2014). These studies have led to important insights into migration routes and migratory connectivity (Trierweiler et al., 2014; Hobson and Kardynal, 2015; Cooper et al., 2017), which are helping to identify population-limiting factors (Hewson et al., 2016), and improve conservation efforts throughout the annual cycle (Sorte et al., 2015; Runge et al., 2015). Additionally, studies on fledgling movements and habitat use provide key insights into this particularly vulnerable life-stage (Lens and Dhondt, 1994; King et al., 2006; Jenkins et al., 2017).

While more studies are now being conducted to help fill information gaps throughout the annual cycle during migratory (Rockwell et al., 2017; Cohen et al., 2017; Gómez et al., 2017; Paxton and Moore, 2017), and over-wintering periods (Ruiz-Gutierrez et al., 2016; Goodenough et al., 2017; Colorado and Rodewald, 2017), there still remain important gaps concerning the post-fledging period (Cox et al., 2014). During this post-fledging period, birds generally have high mortality rates (Yackel Adams et al., 2006; Gruebler and Naef-Daenzer, 2010; Eng et al., 2011). As such, the post-fledging period is critical in the annual cycle and we must continue to investigate avian responses to urbanization during this period (Cox et al., 2014).

To improve our understanding of how anthropogenic effects of urbanization can influence space use during the post-fledging period, we conducted a comparative study using movement data from juvenile Wood Thrushes (*Hylocichla mustelina*) and Gray Catbirds (*Dumatella caroliniensi*) during the post-fledging period. These two species, while being closely-related phylogenetically (Prum et al., 2015), share similar life-history traits (Evans et al., 2011; Roth et al., 1996), though exhibit differential responses to urbanization. Currently, there are opposing annual population trend estimates for the Gray Catbird (0.42 %) and the Wood Thrush (-2.77 % (Sauer et al., 2017)).

Our objectives of the movement data analysis were to compare 1) space use patterns and 2) movement metrics of Gray Catbirds and Wood Thrushes during both pre- and post-independent temporal periods, which we defined here as ≤ 20 days and > 20 days from leaving the nest. To gauge effects of urbanization on post-fledging space use, we were interested in testing for specific differential responses between species during pre- and post-fledging periods in their respective use of forest habitat and road area. Loss and fragmentation of forest habitat are directly related to urbanization and increasing developed land cover, and similarly, increases in road density within urbanized landscapes can pose myriad threats to wildlife (Gibbs and Shriver, 2002; Brady and Richardson, 2017), and birds in particular (Erritzoe et al., 2003; Benítez-López et al., 2010).

To finish, the obtained responses towards urbanization during the post-independent period were summarized in a model that simulates the consecutive locations of in-

dividual fledglings on a daily basis. Here, the selection of the subsequent locations depends on the available proportions of forest habitat and road area. Ultimately, the model will be used to assess the potential effect of land cover changes. In this way, the responses of both species under consideration towards (locally) increased or decreased urbanization can be studied in detail, as such, creating a tool for making data-driven informed decisions that will benefit sustainable resource management.

In Section 6.2, an overview of the tracking data set is given, after which we describe the studied species in Section 6.3. The movement data analysis is performed in Section 6.4, and the construction and validation of the spatially explicit marching model is presented in Section 6.5.

6.2 Study area and data set

The study area in and around Newark, Delaware (39.6837° N, 75.7497° W) and Landenberg, Pennsylvania (39.7778° N, 75.7716° W) is located in the eastern mid-Atlantic United States. During the 2012–2014 breeding seasons (i.e., May–August), we systematically searched 23 forest fragments, 13 of which contained nests of catbirds and Wood Thrushes. The forest fragments ranged in area from 5 to 163 ha, and consisted of dominant canopy tree species including *Fagus grandifolia*, *Acer rubrum*, *Quercus* spp., *Liriodendron tulipifera*, and *Liquidambar styraciflua*. Additionally, both native and non-native understory woody species including *Lindera benzoin*, *Viburnum* spp., *Clethra alnifolia*, *Rosa multiflora*, *Eleagnus umbellata*, and *Rubus* spp could be found. Once a nest was detected, it was monitored every 3–4 days (Martin and Geupel, 1993).

The nestling Gray Catbirds and Wood Thrushes were banded (with aluminum US Geological Survey bands; Permit number: 23475), and 1–3 days prior to fledging, 1 or 2 nestlings were selected and equipped with a VHF radio-transmitter (weighing < 1g; transmitters ranging in frequencies between 150–151 MHz (Blackburn transmitters, Nacogdoches, Texas)). The nestlings were extracted by hand from nests, transmitters were attached using a 1 mm elastic thread fitted with a figure-8 loop harness (Rappole and Tipton, 1991), after which the nestlings are returned to their nests. As such, Gray Catbird and Wood Thrush fledgling movement was tracked using VHF radio transmitters fitted on individuals that originated from 69 nests within 13 discrete forest fragments in a generally, urbanized landscape.

Using hand-held Yagi antennas attached to programmable receivers (Advanced Telemetry Systems, Isanti, MN), each individual was located once per day throughout the study period until they either died, the radio-transmitter battery failed, or they dispersed from the study area. When a bird was successfully located, GPS units were used to record location coordinates (latitude and longitude). The

field researchers deemed events as mortality events if transmitters were recovered containing part or all of a dead individual's body, blood, or feathers. In some circumstances, where the locations of birds were unable to be determined if an individual was out of range compared to its previous location, a roof-mounted omnidirectional antenna on top of a vehicle was used to search an increasingly larger area beginning at the bird's previous known location. These opportunistic searches continued for all missing birds each day lasted until the bird was located or up to five days beyond the estimated date of transmitter battery failure.

All together, location data for 52 fledgling Gray Catbirds and 60 Wood Thrushes that consisted of 4,066 unique locations are available. These data were further subset into pre-and post-independence periods from 44 and 32 Gray Catbirds, and 56 and 49 Wood Thrushes, respectively. On average, individuals were tracked for 41 days, and the length of the tracking period ranged between 2 to 88 days. During the duration of the study, 69% of Gray Catbirds and 86% of Wood Thrushes survived.

To quantify relative space use of Gray Catbirds and Wood Thrushes during pre-and post-independent periods, land use/land cover data from the National Land Cover Database (NLCD) (30 x 30 m resolution, Albers Equal Area projection) was used. These 2011 NLCD data include the following land cover/land use classes: Developed Open Space, Developed Low Intensity, Developed Medium Intensity, Developed High Intensity, Barren Land, Deciduous Forest, Evergreen Forest, Mixed Forest, Scrub/Shrub, Grassland/Herbaceous, Pasture/Hay, Cultivated Crops, Woody Wetlands, Emergent Herbaceous Wetlands and Open Water (Homer et al., 2015). A Road map (30 x 30 m resolution, Albers Equal Area projection) was obtained by rasterizing Topologically Integrated Geographic Encoding and Referencing (TIGER) data (2016 TIGER/Line Shapefiles).

The NLCD, has proven extremely useful in identifying priority landscapes in need of conservation and restoration. However, these data often lack the resolution needed to identify and examine fine scale changes in land cover or to make fine scale decisions about how and where to focus restoration efforts. Therefore, we also considered the high-resolution land cover map produced by the Chesapeake Conservancy (1 x 1 m resolution, Albers Equal Area projection). Twelve land cover classes are defined therein: Background, Water, Emergent Wetlands, Tree Canopy, Scrub/Shrub, Low Vegetation, Barren, Structures, Other Impervious Surfaces, Roads, Tree Canopy over Structures, Tree Canopy over Other Impervious Surfaces and Tree Canopy over Roads (Chesapeake Conservancy, 2017).

6.3 The Gray Catbird and the Wood Thrush

The Gray Catbird and the Wood Thrush are passerine birds (Table 6.1) that are widely distributed across eastern USA and south-eastern Canada, where they breed.

During winter, they can be found in southern Mexico and Central America, south to Panama (Fig. 6.1) (Evans et al., 2011; Roth et al., 1996; Birdlife international, 2018b).

Table 6.1: Taxonomic classification of the Gray Catbird and the Wood Thrush.

| | | |
|---------|---------------------------------------|------------------------------------|
| Kingdom | Animalia | Animalia |
| Phylum | Chordata | Chordata |
| Class | Aves (Birds) | Aves (Birds) |
| Order | Passeriformes | Passeriformes |
| Family | Mimidae | Turdidae |
| Genus | Dumetella | Hylocichla |
| Species | Dumetella carolinensis (Gray Catbird) | Hylocichla mustelina (Wood Thrush) |

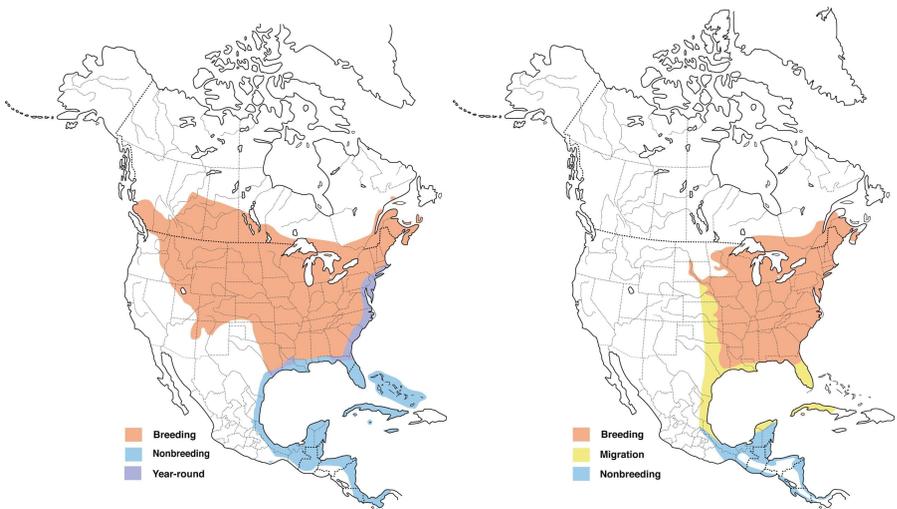


Figure 6.1: Range of the the Gray Catbird (left) and the Wood Thrush (right) (Evans et al., 2011; Roth et al., 1996).

The Gray Catbird can be recognized by its gray plumage with black cap and tail, and by a rufous-brown patch under the tail. It has black bill, eyes, legs and feet (Evans et al., 2011). The Wood Thrush is warm reddish-brown above and have white with bold black spots on their underparts. It has a bold, white eyering (Roth et al., 1996) (Fig. 6.2).

Gray Catbirds live and nest in dense shrubs and thickets of young trees. These habitat can often be found in a human dominated landscape in the form of clearings, roadsides, fencerows, abandoned farmland, and residential areas. On tropical wintering grounds, catbirds spend more time in forests than they do while they are in North America. Catbirds are shy but full of energy, hopping and fluttering from branch to branch. They are reluctant to fly across open areas, preferring quick, low flights over vegetation. In summer, Gray Catbirds eat mainly ants, beetles, grasshoppers, midges, caterpillars and moths. When fruits are available, they also



Figure 6.2: Gray Catbird (left) and the Wood Thrush (right) (by Bob Moul).

eat a variety of berries (The Cornell Lab of Ornithology, 2018). Furthermore, Gray Catbirds are diurnal, which means they are active during the day, and migratory, but they migrate at night. Pairs are territorial during the breeding season and in winter, but during migration they travel in flocks of about 10 to 15 birds. Spring migration ranges from March to May, and fall migration ranges from late August to November (Dunne, 2006).

Wood Thrushes breed throughout mature deciduous and mixed forests, preferring those with a moderate shrub/subcanopy layer, a fairly open forest floor, shade, moist soil, decaying leaf litter and water nearby. In their winter range, Wood Thrushes live in the understory of old tropical forests. However, they may also be found along the edges of forests, and in younger forests (Birdlife international, 2018b; The Cornell Lab of Ornithology, 2018). Wood Thrushes forage by hopping through leaf litter on the forest floor, tossing leaves to expose insects. Wood Thrushes feed mostly on leaf-litter invertebrates and fruits from shrubs, which in particular are important for meeting the demands of migration (Roth et al., 1996; The Cornell Lab of Ornithology, 2018). Wood Thrushes are diurnal and usually solitary, defending their territory, though they sometimes join flocks of other birds in the winter. The average distance that Wood Thrushes migrate is 2200 km. Wood Thrushes usually arrive at the breeding grounds by mid-April and begin leaving for the fall migration around mid-August. Wood Thrushes also migrate at night (Roth et al., 1996; Birdlife international, 2018b).

Gray Catbirds are monogamous. Breeding pairs form soon after the catbirds arrive on the breeding grounds in spring. They will almost never return to the same breeding site year after year. Gray Catbirds breed between April and early August and usually raise two broods per season. The female lays 1 to 5 (usually 3 or 4) eggs and incubates them for 12 to 14 days. The chicks leave the nest after 10 to 11 days, but the parents continue to feed them for up to 12 days more.

Wood Thrushes are also monogamous. Breeding pairs form in mid-April and early-May. Most Wood Thrushes find a new mate each year. They breed from early May

to late August and typically raising two broods. The female lays 2 to 4 eggs, and incubates them for about 13 days. The chicks leave the nest when they are 12 to 15 days old. The parents continue to feed them until they are about 21 to 31 days old (Roth et al., 1996).

Gray Catbirds have been uplisted to 'Least Concern' since the species has a wide range and a stable population size. Wood Thrushes, on the contrary, have been uplisted to 'Near Threatened' since the species has undergone a moderately rapid population decline over the past three generations (International Union for Conservation of Nature and Natural Resources, 2018). The primary threat to the species is likely to be the clearance and fragmentation of forests in both its breeding and non-breeding ranges. Fragmented habitats may offer poorer food and increase the exposition of nests to predators such as raccoons, jays, crows, and domestic or feral cats, and to the Brown-headed Cowbird, which is a nest parasite (Birdlife international, 2018b; The Cornell Lab of Ornithology, 2018). Demographic models suggest that population declines are primarily driven by loss and fragmentation of non-breeding habitat in Central America (Taylor and Stutchbury, 2016). Gray Catbird, on the contrary, are less vulnerable for these human induced habitat changes since they often benefit from the habitats created in a human dominated world (The Cornell Lab of Ornithology, 2018). Moreover, Grey Catbirds are not a victim of nest parasitism since they learn to recognize their own eggs and the cowbird's eggs, mistakes are rare (Cimprich and Moore, 1995).

6.4 Determining land use

6.4.1 Methods

Despite the increasing numbers of studies focusing on previously-understudied post-fledging ecology, contextual research linking post-fledging survival and space use to anthropogenic effects remain limited. Here, in a comparative study, we examined space use of post-fledgling Gray Catbird and Wood Thrushes within an urbanized landscape to test for differential responses to urbanization.

Starting from daily location data, we used TIBBs to generate PDFs of the species' probability of occurrence, and estimated the relative space use among land cover types. Here, the PDF was constructed for each individual during both pre- and post-independence using the analytical solution proposed in Chapter 4. Since fledglings are studied, who are exploring the habitat, the assumption of a random walk in between two fixes appears justified.

On the basis of the PDF, the home range was defined for each individual during pre- and post-independent periods by calculating a threshold so that the subset

of pixels having a PDF value in their center that is greater than or equal to this threshold, represents 95 % of the volume under the PDF. This subset equals the area with the highest possible PDF values where the individuals are expected to be located during 95 % of the studied time interval (Kie et al., 2010).

To estimate the relative space use among land cover types from the NLCD, the total number of pixels assigned to each land cover class within the home range of an individual during pre- or post-independent period was extracted, and the proportion of each land cover class within the home range was calculated. Additionally, the weighted relative proportion of each land cover class was calculated by taking the sum of the PDF values of the pixels that were assigned the land cover class under consideration and are located within the home range. Finally, the weighted relative proportions (which all summed to 1) were multiplied by the total area (km²) of the home range to obtain the respective area (km²) of each land cover class. For both the weighted relative proportions and area (km²) of each land cover class for both species during pre- and post-independence, the mean and standard error were calculated. Since none of the individual birds from our study used either Barren Land or Open Water, these two land cover classes were removed from all analyses. Furthermore, the respective road area (km²) was computed within each individual's home range during pre- and post-independence using similar methods described above. Additionally, we categorized the fate of individuals as either survived or died independently within pre- and post-independence periods.

To gain an understanding into how underlying movement metrics were related to observed patterns in space use between species and within species between pre- and post-independence periods, we computed several movement data features for each individual. We calculated 1) the area of the home range (km²), 2) path length (m), with each path being defined as the Euclidean distance between a unique pair of consecutive locations where the individual was observed, 3) total displacement (m), which we defined as the sum of distances moved per individual, per species and time-period, and 4) net displacement (m), which we defined as the Euclidean distance between the first and final recorded locations for an individual. All movement metrics were calculated for individuals separately during pre- and post-independence periods.

To test the null hypothesis that catbirds and Wood Thrushes respond similarly in their space use and movement behavior, the R package 'lme4' (Bates et al., 2014) was used to fit generalized linear mixed-effects models (GLMM) to test for differences in main effects of the response variables respective land cover area (km²) and road area (km²), and included all interactions between 1) species, 2) land cover class, 3) fate (survived or died), and 4) post-fledging periods. When significant correlations were detected from mixed-effects models ($\alpha \leq 0.05$) for variables with more than two factor-levels, the 'multcomp' package was used to implement Tukey's HSD post hoc methods to test for pair-wise differences (Hothorn et al., 2008).

We also fit generalized linear mixed-effects models to test for differences between species (along with all interactions with fate and post-fledging period) for the following movement metric response variables: home range area, path length, total displacement, and net displacement.

To account for potential effects from fledglings originating from the same nest, and differences in sampling effort among years, we included unique nest ID nested within Year as a random effect. To ascertain significant differences, we used parametric bootstrapping (PBtest) to test for differences between full and reduced models (*i.e.* to test for significant contributions of each independent covariate to variance explained) with the ‘pbkrtest’ package (Halekoh and Højsgaard, 2014). We tested all data for departures from normality using visual inspection of quantile-quantile residual plots of variables, and all statistical analyses were conducted using R (version 3.4.3; R Core Team 2017).

6.4.2 Results

TIBB-derived normalized PDFs for fledgling Gray Catbirds (GRCA) and Wood Thrushes (WOTH) in both pre- and post-independence periods of parental care during the breeding seasons from 2012–2014 in and around Newark can be found in Fig 6.3.

The weighted relative proportions of space use among NLCD land cover types can be found in Table 6.2. In general, the mean weighted relative proportions of space use differed among land cover classes, as well as between species and post-fledging periods.

Table 6.2: Mean weighted relative proportions of space use within each National Land Cover Dataset land cover class. Means (and standard errors) are shown.

| Land cover class | Gray Catbird | | Wood Thrush | |
|------------------------------|--------------------------------------|---------------------------------------|--------------------------------------|---------------------------------------|
| | Pre-independence (44 individuals) | Post-independence (32 individuals) | Pre-independence (56 individuals) | Post-independence (49 individuals) |
| Deciduous Forest | 0.53 (0.04) | 0.22 (0.03) | 0.85 (0.03) | 0.71 (0.02) |
| Mixed Forest | 0.02 (0.01) | 0.01 (0.003) | 0.01 (0.004) | 0.02 (0.004) |
| Evergreen Forest | 0.001 (0.001) | 0.002 (0.002) | 0.001 (0.001) | 0.014 (0.01) |
| Developed Open Space | 0.09 (0.02) | 0.2 (0.03) | 0.06 (0.01) | 0.11 (0.01) |
| Developed Low Intensity | 0.02 (0.01) | 0.14 (0.02) | 0.01 (0.004) | 0.04 (0.01) |
| Developed Medium Intensity | < 0.01 (< 0.01) | 0.09 (0.02) | 0.01 (0.003) | 0.02 (0.004) |
| Developed High Intensity | < 0.01 (< 0.01) | 0.03 (0.01) | < 0.01 (< 0.01) | < 0.01 (< 0.01) |
| Pasture/Hay | 0.18 (0.03) | 0.17 (0.03) | 0.01 (0.01) | 0.03 (0.01) |
| Cultivated Crops | < 0.01 (< 0.01) | 0.02 (0.01) | < 0.01 (< 0.01) | < 0.01 (< 0.01) |
| Grassland/Herbaceous | 0.00 (0) | 0.00 (0) | 0.00 (0) | < 0.01 (< 0.01) |
| Scrub/Shrub | 0.01 (0.01) | 0.004 (0.002) | 0.01 (< 0.01) | 0.01 (< 0.01) |
| Emergent Herbaceous Wetlands | < 0.01 (< 0.01) | < 0.01 (< 0.01) | 0.00 (0) | 0.00 (0) |
| Woody Wetlands | 0.14 (0.02) | 0.12 (0.03) | 0.05 (0.02) | 0.04 (0.01) |
| Barren Land | 0.00 (0) | 0.00 (0) | 0.00 (0) | 0.00 (0) |
| Open Water | 0.00 (0) | 0.00 (0) | 0.00 (0) | 0.00 (0) |

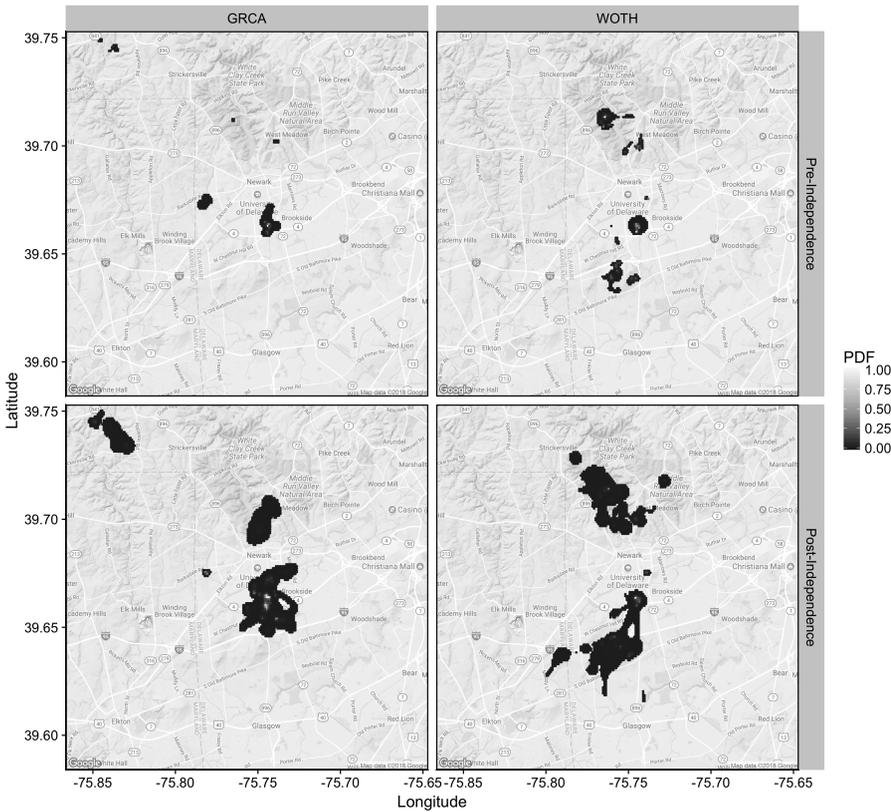


Figure 6.3: Time-integrated Brownian bridge-derived normalized PDFs for fledgling Gray Catbirds (GRCA) and Wood Thrushes (WOTH) in both pre- and post-independence periods of parental care during the breeding seasons from 2012–2014 in and around Newark, Delaware, USA.

6 More specifically, it can be shown that there are significant differences between both species among land cover classes (PBtest = 180.2, nsim =1000, $P < 0.0001$). The use of deciduous forest area, for example, was 5 times greater by Wood Thrushes ($0.66 \pm 0.10 \text{ km}^2$) compared to catbirds ($0.013 \pm 0.018 \text{ km}^2$).

Moreover, we detected a significant (post-fledging period x land cover class x species) interaction (PBtest = 138.2, nsim =1000, $P < 0.0001$), where the area of deciduous forest used by Wood Thrushes was 5.3 times greater than for catbirds during post-independence (Fig. 6.4).

Additionally, a significant (fate x land cover class x species)-interaction (PBtest = 23.0, nsim =1000, $P < 0.04$) was detected, which showed that, of birds that survived, Wood Thrushes used deciduous forest area 4.9 times more often than catbirds (Fig. 6.5).

Although we detected a marginally significant (post-fledging period x fate x species) interaction (PBtest = 3.15, nsim =1000, $P = 0.08$), post-hoc pairwise com-

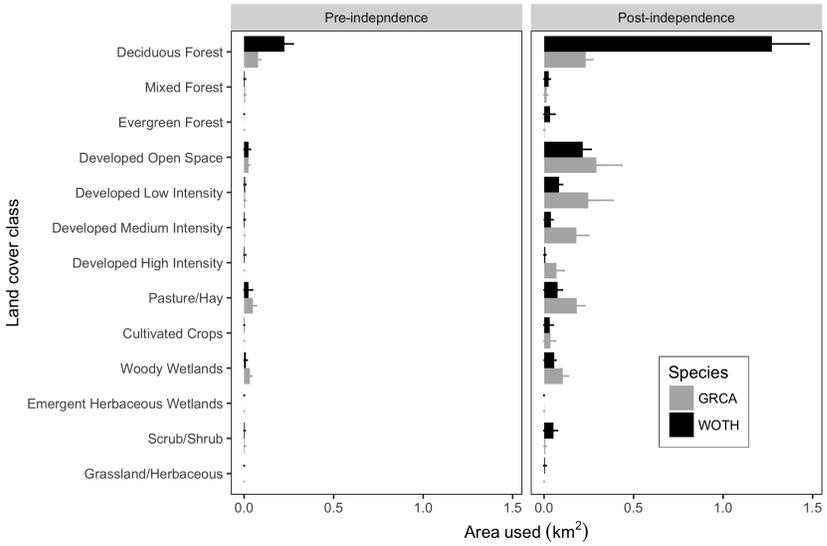


Figure 6.4: Comparison of area used (km²) within 15 National Land Cover Dataset land cover classes (weighted mean ± standard error) between Gray Catbirds (gray bars) and Wood Thrushes (black bars) during pre-independence and post-independence periods for post-fledglings during the breeding seasons from 2012–2014 in and around Newark, Delaware, USA.

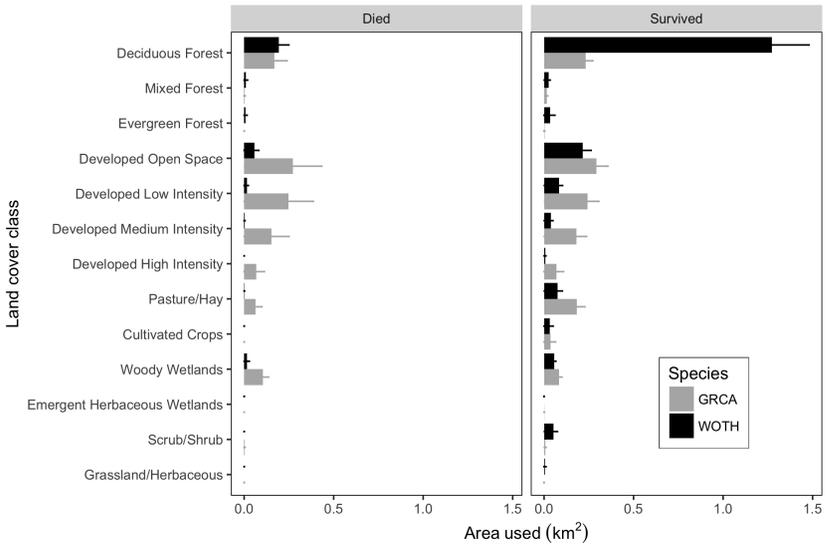


Figure 6.5: Comparison of area used (km²) within 15 National Land Cover Dataset land cover classes (weighted mean ± standard error) between Gray Catbirds (gray bars) and Wood Thrushes (black bars) that died or survived during the breeding seasons from 2012–2014 in and around Newark, Delaware, USA.

parisons indicated that catbirds and Wood Thrushes used areas similarly between fates within pre- and post-fledging periods (t.ratios ranged between $-2.21 - 1.40$, $P > 0.35$, in all cases). Moreover, we found no interactive effects of (post-fledging period x fate x land cover class x species) (PBtest = 16.8, nsim = 1000, $P = 0.16$).

Furthermore, we found that the catbirds used areas with a greater amount of roads ($0.12 \pm 0.03 \text{ km}^2$) compared to Wood Thrushes ($0.08 \pm 0.02 \text{ km}^2$; PBtest = 3.82, nsim = 1000, $P < 0.05$). We also detected a significant (species x post-fledging-period) interaction (PBtest = 5.51, nsim = 1000, $P < 0.04$), where catbirds used areas with 1.8 times the road area compared to Wood Thrushes during the post-independence period (Fig. 6.6).

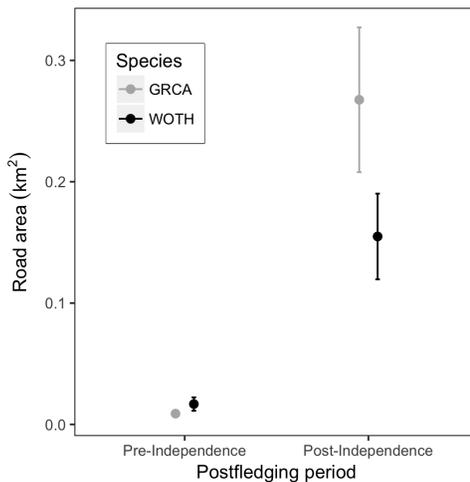


Figure 6.6: Comparison of space use associated with road density for Gray Catbirds and Wood Thrushes during pre-independence and post-independence periods for post-fledging during the breeding seasons from 2012–2014 in and around Newark, Delaware, USA.

6

As can be seen in Table 6.3, the analysis of the movement features underlying space use patterns while accounting for all potential interaction terms, revealed that movement features including home range area, path length, total displacement, and net displacement were all similar between catbirds and Wood Thrushes (PBtest < 1.73, nsim=1000, $P > 0.19$ in all cases). However, despite a lack of differences between species, we do uncover more F patterns between post-fledging periods. For example, the home range area was 6.9 times greater during post-independence ($1.57 \pm 0.20 \text{ km}^2$) compared to pre-independence ($0.23 \pm 0.04 \text{ km}^2$) periods (PBtest = 42.3, nsim=1000, $P < 0.001$; Fig. 6.7A). Path length differed between pre- and post-fledging periods (PBtest = 109.4, nsim = 1000, $P < 0.001$), where post-independence path length ($132.3 \pm 4.5 \text{ m}$) was double the path length during pre-independence ($63.5 \pm 2.5 \text{ m}$); Fig. 6.7B). Total displacement moved was greater during post-independence ($3,989 \pm 331 \text{ m}$) than

pre-independence (945.4 ± 59 m) periods (PBtest = 86.3, nsim = 1000, $P < 0.001$; Fig. 6.7C). Additionally, the net displacement was also greater during post-independence (720.7 ± 80.4 m) than during pre-independence (174.5 ± 25.5 m) periods (PBtest = 44.2, nsim = 1000, $P < 0.001$; Fig. 6.7D).

Table 6.3: Mean (and standard error) of movement metrics of post-fledging Gray Catbirds and Wood Thrushes during pre- and post-independence periods from 2012–2014 in and around Newark, Delaware, USA

| Movement feature | Gray Catbird | | Wood Thrush | |
|---|-----------------------------------|------------------------------------|-----------------------------------|------------------------------------|
| | Pre-independence (44 individuals) | Post-independence (32 individuals) | Pre-independence (56 individuals) | Post-independence (49 individuals) |
| Brownian bridge area (km ²) | 0.17 (0.03) | 1.30 (0.22) | 0.27 (0.07) | 1.77 (0.29) |
| Path length (m) | 57.6 (3.9) | 130.6 (6.6) | 67.7 (3.3) | 133.5 (6.8) |
| Total displacement (m) | 780.6 (93.2) | 4,244 (534) | 1,084 (70.9) | 3,829 (425) |
| Net displacement (m) | 173.3 (31) | 759.4 (105) | 178.6 (9.8) | 696.4 (114) |

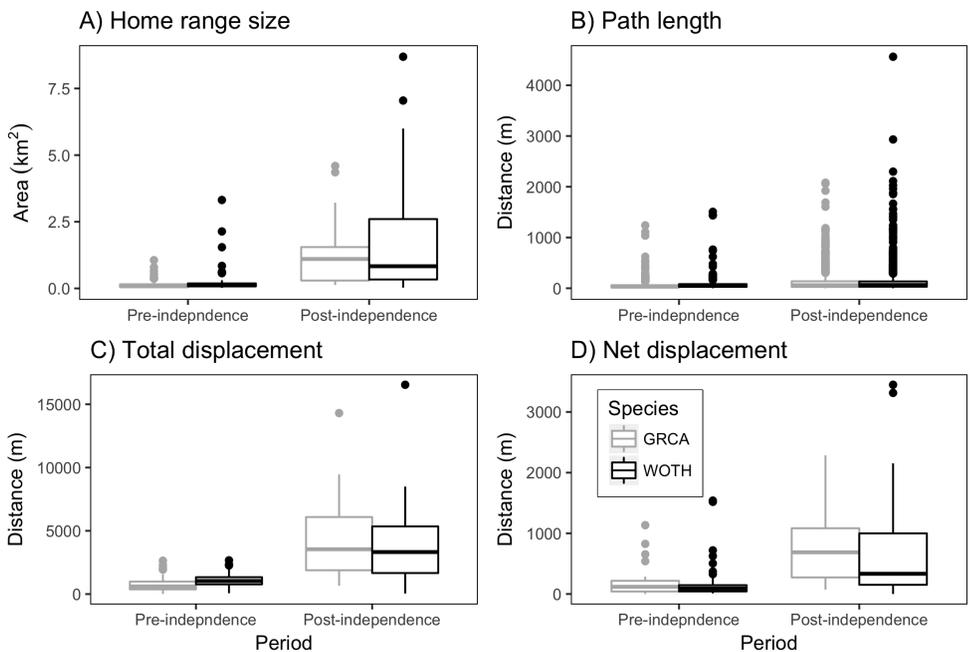


Figure 6.7: Comparison of pre- and post-independence periods for A) Brownian bridge area (km²), B) path length (m), C) total displacement moved (m), and D) net displacement (m). Note differing units and scales on y-axis.

Patterns in movement metrics predicted by the fate of fledglings were found for total displacement moved (PBtest = 16.2, nsim = 1000, $P < 0.001$) and net displacement (PBtest = 4.32, nsim = 1000, $P < 0.04$). After controlling for post-fledging

period effects, we find that fledglings that survived, had a greater total displacement (2594 ± 210 m) and a greater net displacement (460.5 ± 49.3 m) compared to the total displacement (744 ± 213 m) and net displacement (188.6 ± 61.4 m) of birds that died.

Finally, we find a significant interaction for the total displacement moved between post-fledging period and fate (PBtest = 5.25, nsim = 1000, $P < 0.03$), where surviving birds moved a greater total displacement ($4,296 \pm 355$ m) than birds that died ($1,741 \pm 541$ m) during the post-independence period.

6.4.3 Discussion

We demonstrated how Neotropical migratory species with similar evolutionary- and life-history characteristics can exhibit differential and contrasting behavioral responses in their space use within an urbanized landscape. Here, we compared two species (*i.e.*, the Gray Catbird and Wood Thrush) that were previously known to exhibit urban-adapting and urban-avoiding proclivities, respectively. Our expectations that catbirds would use areas with less forest and more roads compared to Wood Thrushes, were supported by our results of estimated space use of both species in relation to a representative urbanized landscape within our study area, located in the coastal mid-Atlantic United States, and lies in the heart of the megapolopolis extending from Boston, Massachusetts to Washington, DC, USA (Morrill, 2006).

We found that catbirds used areas with more developed land cover and more roads than Wood Thrushes, and in contrast, Wood Thrushes used much greater areas of deciduous forest during the entire post-fledging period. These patterns were further amplified during post-independence, when fledglings left natal patches, navigating through the available suitable and unsuitable habitat within the urbanized landscape of our study area. These results clearly show how avoidant or tolerant behavior of fledglings to areas of higher road density can have repercussions on differential fledgling survival found in other recent studies (Adalsteinsson et al., In prep.). Our results were not surprising and supported our expectations, yet are nonetheless important in helping to better understand how space use during the post-fledging period within the annual cycle, is related to the effects of urbanization on songbird populations (Lepczyk et al., 2017).

Patterns we found in basic movement features of Wood Thrushes such as mean net displacement (m) were similar to previous studies of post-fledging movement (Anders et al., 1998; Rivera et al., 1998; Vernasco et al., 2017). However, estimates of home range area differed considerably. For example, Anders et al. (1998) estimated that home range sizes of fledgling Wood Thrushes ranged between 0.026 and 0.25 km² whereas in our study estimated home range areas ranged between 0.019 and 8.7 km². In light of the development of improved analysis methods for

estimating space use (Walter et al., 2015; Demšar et al., 2015), observed differences are of course, understandable. Although these ranges do overlap towards the lower range, the discrepancy in differing upper ranges is likely due to differences in the landscapes and habitat composition between the two studies. The study of Anders et al. (1998) took place in a contiguous mature forest compared to our study area's fragmented urban landscape. Both simulation-based (Hand et al., 2014), and empirical studies have demonstrated support for how differences in landscape permeability within urban areas can influence bird movement (Tremblay and St Clair, 2009; Baguette et al., 2013; Evans et al., 2011).

Life within urbanized landscapes has proved difficult for breeding birds (Marzluff, 2001; Rodewald and Shustack, 2008; Aronson et al., 2014), particularly due to inherent losses of suitable breeding habitat (Hoekstra et al., 2005). However, not all species respond to direct and indirect effects of urbanization similarly, likely due to complex and interacting differences among evolutionary- and life-history traits, habitat requirements, and behavioral and phenotypic flexibility (Naef-Daenzer and Gruebler, 2016). In particular, achieving the habitat requirements (e.g., availability and arrangement) that are needed for post-fledging birds to forage, avoid mortality risks, and survive until migratory dispersal can be challenging in urban landscapes (Whittaker and Marzluff, 2012; Seress and Liker, 2015). Differential space use is also inherently linked to post-fledging survival probability, due to the differences in exposure to both natural and anthropogenic mortality factors (Ausprey and Rodewald, 2011; Vitz and Rodewald, 2011; Balogh et al., 2011). In particular, Adalsteinsson et al. (In prep.) found that post-fledging survival was higher for Wood Thrushes than Gray Catbirds. However, both species experienced an increased risk of mortality due to anthropogenic factors during the post-independence period. This makes sense, as it is the period when fledglings are no longer requiring parental provisioning, and are leaving their natal forest patches to explore potentially hostile matrix environments for their first time.

As increasing rates of urbanization are predicted to continue to have adverse effects on ecosystems (Seto et al., 2012), species that depend on critical breeding habitat such as migratory Neotropical migratory birds are also predicted to experience concomitant declines (Rottenborn, 1999; Blair, 2004; Ladin et al., 2016). Given equivocal results as to the effects of urbanization on forest-breeding bird populations from previous studies (Rodewald et al., 2013), continued research in this area seems warranted. To manage increasingly urbanized ecosystems, and design landscapes to mitigate potential negative effects, studies ought to continue investigating how species respond, and adapt to both direct and indirect anthropogenic effects of urbanization including loss of habitat, habitat fragmentation, and increased road densities. Furthermore, models that predict how populations respond to increasing rates of urbanization may be valuable for selecting the most successful management action.

6.5 Modeling land use

In this section, the groundwork is laid for the construction of a spatially explicit data-driven model to predict how populations of Gray Catbird and Wood Thrush fledglings respond to increasing rates of habitat loss and fragmentation during the post-independent period. The latter are likely to result from imminent increases in the urbanization rate over the next decades. More specifically, we incorporated some preliminary ideas into a spatially explicit movement model that simulates consecutive locations of individual fledglings on a daily basis (corresponding to time resolution of the tracking data), so that the land use of the *in silico* individuals is similar to the one of the real-world individuals. We started with a simple model that can easily be used, adapted, interpreted and communicated. Its predictive power was verified by means of a cross-validation.

6.5.1 Relevant land cover classes

Forest, Road and Urban canopy cover are the land cover classes that are known to steer movement of the post-fledglings through the landscape. These classes reflect the degree of urbanization and they were selected given the preferences of the species (Section 6.3) and the results of the data analysis (Section 6.4). As such, we constructed three 30 x 30 m resolution binary maps using the available geographical information (Section 6.2), one per considered land cover. Each pixel in such a map was assigned a value of one if the land cover was present, and zero otherwise. The spatial extent of the study area was delineated by a rectangular area with *y*-coordinates varying between 2,019,885 m and 2,045,235 m, and *x*-coordinates between 1,695,795 m and 1,720,905 m (Albers Equal Area projection).

More specifically, for constructing the binary Forest map, a pixel was assigned a value of one if it was labeled as Deciduous Forest, Evergreen Forest, Mixed Forest or Woody Wetlands in the NLCD map. Subsequently, the binary maps for Road and Urban canopy cover were constructed on the basis of the high-resolution land cover map produced by the Chesapeake Conservancy. For what concerns Road, the pixels of the high-resolution map of size 1 x 1 m were aggregated into pixels of size 30 x 30 m (as such, each pixel contains 30 x 30 pixels of the high-resolution map) by calculating, for each of the latter, the proportion of high-resolution pixels that were either labeled as Impervious Roads or as Tree canopy over Roads. Then, a pixel was assigned a value of one if this proportion was greater than or equal to 0.1, and zero otherwise. As such, a binary Road map was obtained. Also for Urban canopy cover, the pixels of the high-resolution map were first aggregated in 30 x 30 m pixels by calculating the proportion of 1 x 1 m pixels that were either labeled as Tree Canopy or Scrub/Shrub. Then, a pixel was assigned a value of one

if the proportion in a pixel was greater than or equal to 0.5, and if the pixel was not labeled as Forest, and zero otherwise (arbitrary choice).

The study area together with the resulting VHF tracking data are shown in Figs. 6.8 and 6.9, respectively. The tracking data set corresponding to the post-independent period contains 32 individuals and 1040 unique locations for Gray Catbird and 49 individuals and 1463 unique locations for Wood Thrush. The NLCD map, the high-resolution land cover map produced by the Chesapeake Conservancy and, finally, the binary Forest, Road and Urban canopy cover maps are used as base layers. Considering the binary maps, 26 % of the study area is covered by Forest, 15 % by Road and 18 % by Urban canopy.

6.5.2 Spatially explicit movement model

6.5.2.1 Overview

Our approach involves modeling subsequent angles of movement and step (path) lengths, and is, as such, similar to the models of Tracey et al. (2005) and Hubbard et al. (2004). A simulation generally consists of the following steps.

First, a starting location $\mathbf{x}(1)$, *i.e.* the location \mathbf{x} at day 1, and the number of locations n_d , corresponding to the number of days in the simulated period, are defined. In this dissertation, we choose to define an *in silico* individual for each real-world individual, so that the starting locations and lengths of the simulated periods of the *in silico* individuals correspond to the ones of the real-world individuals.

Then, for every *in silico* bird (Gray Catbird (\mathcal{G}) or Wood Thrush (\mathcal{W})), the step length $v(2)$ (the distance traveled per day) is sampled from an empirical distribution function (EDF) that was derived for each species separately on the basis of all available tracking data. The value obtained from sampling is a positive real number. To construct the EDF, a discretization step of 1 m was chosen. Given the sampled step length, an angle of movement $\alpha(2)$ is chosen to obtain the next location $\mathbf{x}(2)$. For that purpose, n_α angles ranging from 0° to 360° in steps of $360/n_\alpha$ are considered. We consider n_α to be either 4, 8, 16 or 36. In order not to confine movement to a few fixed angles, a random angle θ is sampled from a uniform distribution function $U(0, 360/n_\alpha)$ and added to each considered movement angle. As a convention, 0° points to the east, while 90° points to the north.

From these n_α angles, an appropriate one has to be chosen. This choice is based on the presence/absence of Forest F , Road R and Urban canopy cover U . More specifically, we account for the presence of these land cover classes near the current and next location, and in the area in between. Therefore, the TIBB is calculated for every angle α_j with $j = 1, \dots, n_\alpha$ on the basis of the current location and the possible next location. The mobility factors used during these calculations

are species dependent and correspond to the median values obtained from the tracking data, being $\sigma_m^2 = 303.6 \text{ m}^2 \text{ h}^{-1}$ for Gray Catbird and $\sigma_m^2 = 550.6 \text{ m}^2 \text{ h}^{-1}$ for Wood Thrush. For every TIBB, we delineate an area where the individual is expected to be located during 95 % of the studied time interval (one day) on the basis of the resulting PDF. This area is denoted in the remainder as the 95 % TIBB area. Therefore, we calculate a threshold so that the subset of pixels having a PDF value in their center greater than or equal to this threshold represents 95 % of the volume under the PDF (Kie et al., 2010).

In each such 95 % TIBB area j , the proportions of Forest P_j^F , Road P_j^R and Urban canopy cover P_j^U are determined, which are subsequently transformed into suitability index (SI) values. The SIs linking proportions to SI values are different for the Gray Catbird and the Wood Thrush, and incorporate the species behavior. Here, a completely unsuitable, respectively completely suitable, environment corresponds to SI value zero, respectively one, whereas an environment with a low, respectively high, suitability is assigned an SI value between zero and 0.5, respectively between 0.5 and one. Like in Section 5.4, we consider 0.5 as a neutral value and we use the uninorm aggregation function to combine the three obtained SI values per angle of movement into one overall habitat suitability index (HSI) value \mathcal{H}_j , $j = 1, \dots, n_\alpha$ (De Baets and Fodor, 1999). It holds that the proportions Forest, Road and Urban canopy cover, the SI values and the HSI values are confined to the unit interval. In this work, two possible SIs are considered, *i.e.* $S^{e,s}(P_j^e)$ and $\tilde{S}^{e,s}(P_j^e)$, where e represents the type of land cover and s the species under consideration (See Section 6.5.2.2).

After transforming the proportions of Forest (P_j^F), Road (P_j^R) and Urban canopy cover (P_j^U) into one overall HSI value (\mathcal{H}_j) for each TIBB, n_α HSI values are obtained to which a decision rule is applied to determine the angle of movement. This is done in such a way that angles of movement corresponding to the highest available HSI values are more likely to be chosen. Here, we consider three different decision rules, *i.e.* \mathcal{D}_1 , \mathcal{D}_2 and \mathcal{D}_3 (see Section 6.5.2.3). After choosing a movement angle and resulting next location $\mathbf{x}(2)$, the consecutive locations are simulated by repeatedly applying this procedure.

The algorithm to simulate post-fledging movement of a Gray Catbird or a Wood Thrush individual during the post-independent period can be found in Algorithm 2. Since four options for the number of considered angles of movement, two SIs, and three decision rules are considered, all together 24 models were evaluated in terms of their performance (Table 6.4).

6

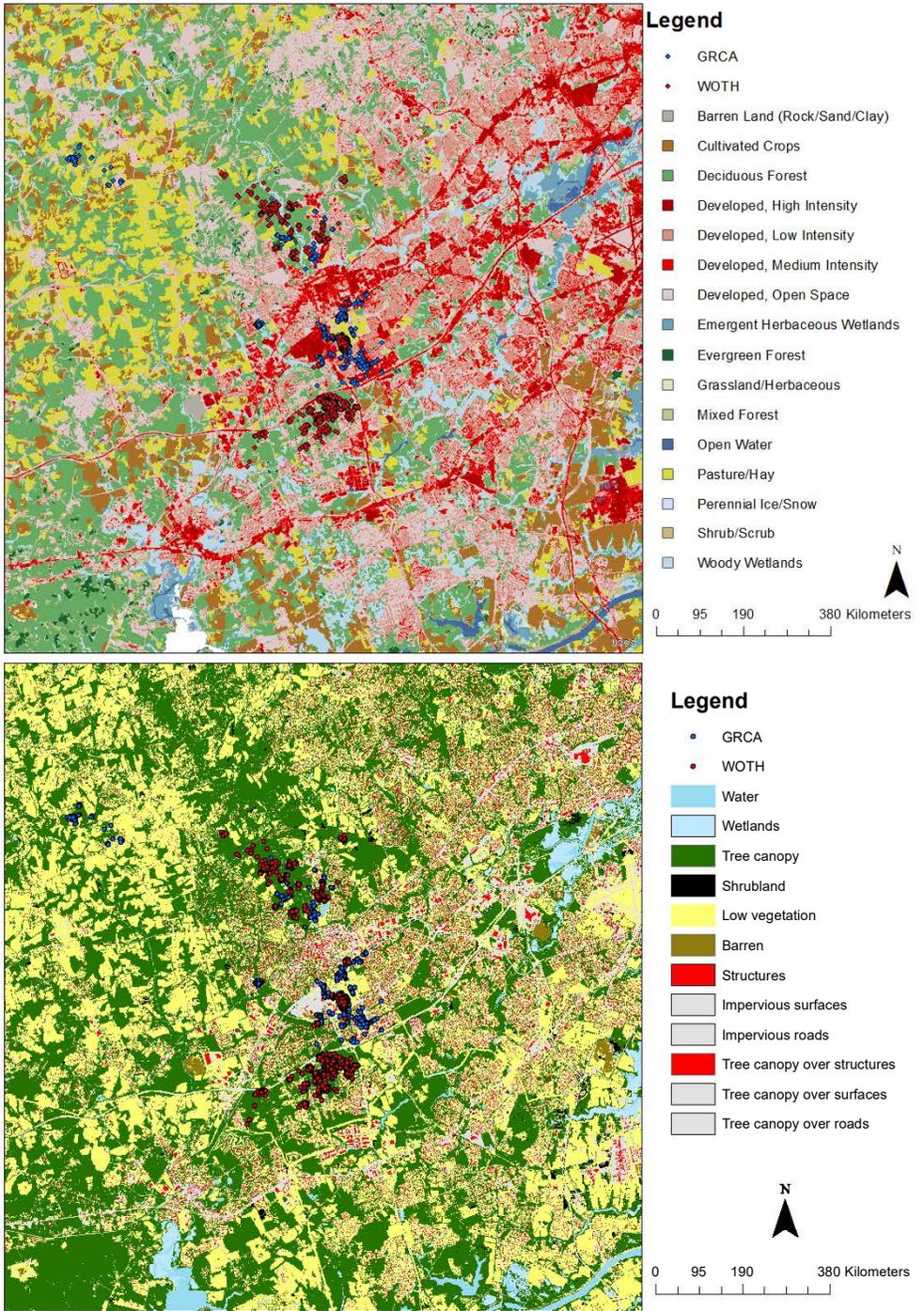


Figure 6.8: Study area together with the VHF tracking data of the post-fledglings during the post-independent period. NLCD map as base layer (top) – the high-resolution land cover map produced by the Chesapeake Conservancy as base layer (bottom).

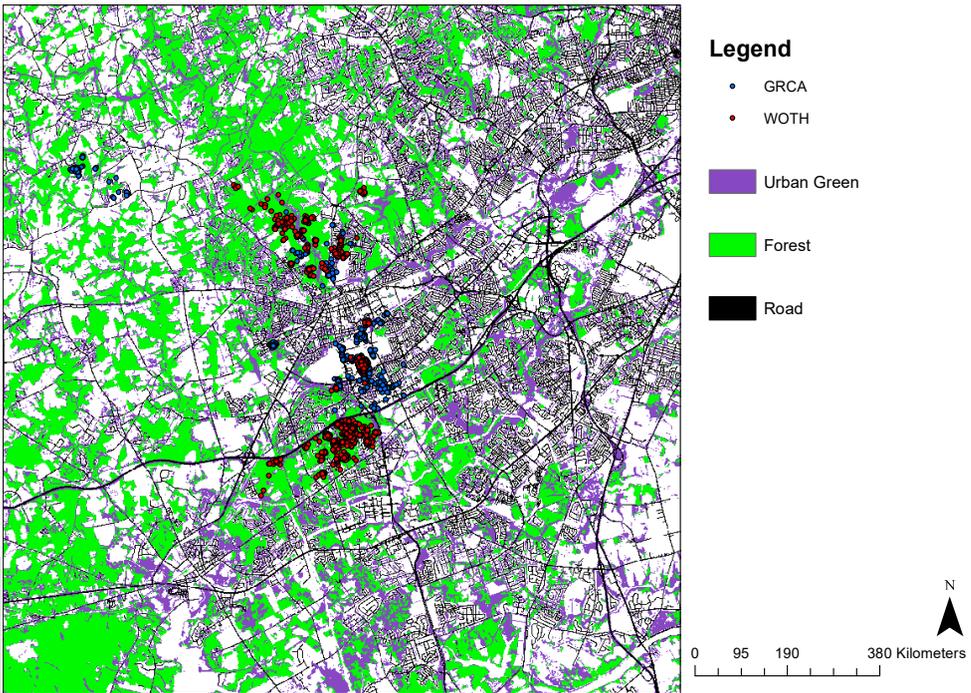


Figure 6.9: Study area together with the VHF tracking data of the post-fledglings during the post-independent period. The three binary maps of Forest, Road and Urban canopy cover, that are used in the spatially explicit movement model are shown as base layer.

Algorithm 2: Pseudo code of the spatially explicit simulation model of post-fledgling movement of a Gray Catbird or Wood Thrush individual during the post-independent period.

Define the species s : Gray Catbird (\mathcal{G}) or Wood Thrush (\mathcal{W})

Determine the EDF of the step length v for the species

Define the starting location $\mathbf{x}(1)$

Define the length of the simulated period n_d (days)

for $i = 2$ **to** $i = n_d$ **do**

 Sample step length $v(i)$

 Sample a random angle $\theta \sim U(0, 360/n_d)$

for $j = 1$ **to** $j = n_a$ **do**

 (n_a is either 4, 8, 16 or 36)

 Consider movement angle $a_j = 360(j - 1)/n_a + \theta$

 Delineate 95 % TIBB area j

 Determine the proportions of Forest P_j^F , Road P_j^R and Urban canopy cover

P_j^U in the 95 % TIBB area

 Transform each proportion to an SI value (Two SIs, *i.e.* $S^{e,s}(P_j^e)$ and $\tilde{S}^{e,s}(P_j^e)$, are considered (cfr. Subsection 6.5.2.2))

 Use the uninorm aggregation function to combine the three obtained SI values into one overall HSI value \mathcal{H}_j

end for

 Apply a decision rule to the HSI values \mathcal{H}_j with $j = 1, \dots, n_a$ to obtain the movement angle $\alpha(i)$ and resulting next location (Three decision rules, *i.e.* \mathcal{D}_1 ,

\mathcal{D}_2 and \mathcal{D}_3 are considered (cfr. Subsection 6.5.2.3))

end for

return Consecutive visited locations, together with the corresponding step lengths, angles of movement, proportions of Forest, proportions of Road and proportions of Urban canopy cover in the 95 % TIBB area

(Since four cases for number of considered movement angles, two SIs, and three decision rules are considered, all together 24 models are obtained.)

Table 6.4: Overview of the constructed models considered for Gray Catbird and Wood Thrush.

| Model ID | Suitability index | Number of angles of movement n_a | Decision rule |
|----------|-------------------|------------------------------------|---------------|
| 1 | $S^{e,s}$ | 4 | D_1 |
| 2 | $S^{e,s}$ | 4 | D_2 |
| 3 | $S^{e,s}$ | 4 | D_3 |
| 4 | $S^{e,s}$ | 8 | D_1 |
| 5 | $S^{e,s}$ | 8 | D_2 |
| 6 | $S^{e,s}$ | 8 | D_3 |
| 7 | $S^{e,s}$ | 16 | D_1 |
| 8 | $S^{e,s}$ | 16 | D_2 |
| 9 | $S^{e,s}$ | 16 | D_3 |
| 10 | $S^{e,s}$ | 36 | D_1 |
| 11 | $S^{e,s}$ | 36 | D_2 |
| 12 | $S^{e,s}$ | 36 | D_3 |
| 13 | $\tilde{S}^{e,s}$ | 4 | D_1 |
| 14 | $\tilde{S}^{e,s}$ | 4 | D_2 |
| 15 | $\tilde{S}^{e,s}$ | 4 | D_3 |
| 16 | $\tilde{S}^{e,s}$ | 8 | D_1 |
| 17 | $\tilde{S}^{e,s}$ | 8 | D_2 |
| 18 | $\tilde{S}^{e,s}$ | 8 | D_3 |
| 19 | $\tilde{S}^{e,s}$ | 16 | D_1 |
| 20 | $\tilde{S}^{e,s}$ | 16 | D_2 |
| 21 | $\tilde{S}^{e,s}$ | 16 | D_3 |
| 22 | $\tilde{S}^{e,s}$ | 36 | D_1 |
| 23 | $\tilde{S}^{e,s}$ | 36 | D_2 |
| 24 | $\tilde{S}^{e,s}$ | 36 | D_3 |

6.5.2.2 Suitability index

Two SIs are considered that reflect the preferences of the Gray Catbird and the Wood Thrush fledglings during the post-independent period towards the land covers Forest, Road and Urban canopy cover. As a first step, the 95 % TIBB areas are determined in between each pair of consecutive locations in the tracking data set, after which the proportions of each land cover class found in these 95 % TIBB areas are calculated. Then, for each species and land cover class, the average proportions and standard deviations (Table 6.5), and relative frequency histograms (bin width 0.05, Fig. 6.10) of the proportions across the 95 % TIBB areas determined on the basis of each pair of consecutive locations in the entire tracking data set, are determined. Comparing the values in Table 6.5 with the land cover available in the study area (*i.e.* 26% Forest, 15% Road and 18% Urban canopy cover), it appears that Gray Catbird often visited areas that are labeled as Urban canopy cover, while Wood Thrush preferred areas with a lot of forest and avoided roads.

The first SI uses these average proportions \bar{P}_e of the land cover e observed across the 95 % TIBB areas defined in between each pair of consecutive locations in the tracking data set. More specifically, it is assumed that an average proportion corresponds to the neutral SI value of 0.5, and that the value of an arbitrary proportion is transformed into an SI value by calculating the deviation from the average. On the basis of expert knowledge and the data analysis (Section 6.4), we arrive at:

$$S^{e,s}(P_j^e) = \begin{cases} \frac{(P_j^F - \bar{P}_F) + 1}{2} & , \text{ if } e = F \text{ and } s = \mathcal{W}, \\ \frac{((1 - P_j^R) - (1 - \bar{P}_R)) + 1}{2} & , \text{ if } e = R \text{ and } (s = \mathcal{W} \text{ or } s = \mathcal{G}), \\ \frac{(P_j^U - \bar{P}_U) + 1}{2} & , \text{ if } e = U \text{ and } s = \mathcal{G}, \\ 0.5 & , \text{ otherwise,} \end{cases} \quad (6.1)$$

which is applied to every 95 % TIBB area j . For Gray Catbird, Eq. (6.1) expresses that the higher the proportion of Urban canopy cover in 95 % TIBB area j is compared to the average proportion across the 95 % TIBB areas defined in between the

Table 6.5: The average proportions of Forest, Road and Urban canopy cover, together with the standard deviations (stdv), that are found across the 95 % TIBB areas defined in between each pair of consecutive locations in the entire tracking data set.

| Land use class | Gray Catbird | | Wood Thrush | |
|--------------------|--------------|----------|-------------|---------|
| | Average (%) | stdv (%) | Average (%) | stdv(%) |
| Forest | 26.69 | 26.83 | 70.87 | 22.80 |
| Road | 14.61 | 18.70 | 5.26 | 8.02 |
| Urban canopy cover | 26.53 | 17.83 | 15.51 | 12.96 |

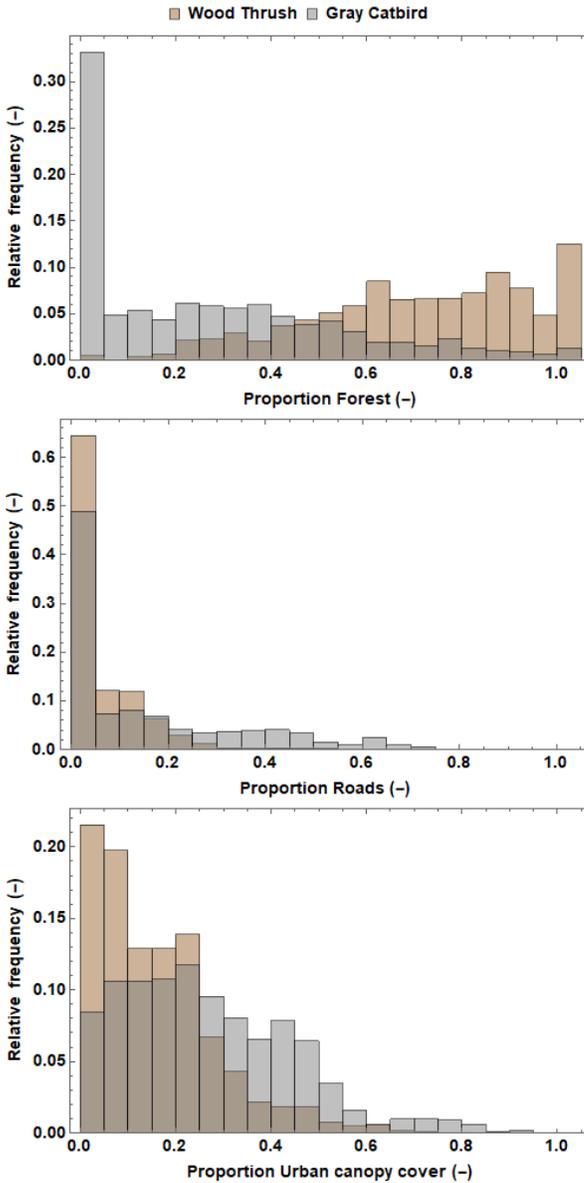


Figure 6.10: Relative frequency histograms (bin width 0.05) of proportions Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between each pair of consecutive locations in the entire tracking data set.

pairs of consecutive locations in the tracking data set, the more suitable the considered region, and hence the higher its SI value. On the contrary, the lower the proportion of Urban canopy cover in 95 % TIBB area j is compared to the average proportion across the 95 % TIBB areas defined in between the pairs of consecu-

tive locations in the tracking data set, the less suitable the considered region and, hence, the lower its SI value is. For Wood Thrush, however, Eq. (6.1) expresses that the higher the proportion of Forest in 95 % TIBB area j is compared to the average proportion across the 95 % TIBB areas defined in between the pairs of consecutive locations in the tracking data set, the more suitable and the higher its SI value is and, the other way around. Finally, it dictates for both species that the higher the proportion of Road in 95 % TIBB area j is as compared to the average proportion across the 95 % TIBB areas defined in between the pairs of consecutive locations in the tracking data set, the less suitable and the lower its SI value is, and the other way around. In other words, Road and Urban canopy cover for Gray Catbird, and, Forest and Road for Wood Thrush may lead to SI values different from 0.5 and influence the overall HSI value. By adding 1 to the deviation from the average and by dividing the right hand side of Eq. (6.1) by 2, $S^{e,s}(P_j^e)$ is confined to the unit interval, and $S^{e,s}(\bar{P}^e)$ corresponds to the neutral SI value of 0.5.

Although the first SI is intuitive, it might not correctly reflect the behavior of the species because it assumes linear relationships between the proportions of land cover and SI values. According to U.S. Fish and Wildlife Service (1981), all kinds of continuous functions can be used as SI. In some circumstances, it is even more appropriate to define a histogram and as such, use classes. For the second SI, we used the information contained in the relative frequency histograms of the proportions Forest, Road and Urban canopy cover across the 95 % TIBB areas defined in between each pair of consecutive locations in the tracking data set (bin width 0.05, Fig. 6.10).

First, we fitted known distribution functions to the relative frequency histograms, such as a Beta, Birnbaum-Saunders, Extreme value, Gamma, Generalized extreme value, Generalized Pareto, Inverse Gaussian, Logistic, Log-logistic, Lognormal, Nakagami, Normal, Rayleigh, Rician, t location-scale and Weibull distributions. These attempts, however, did not lead to satisfactory results in the sense that none of the distributions fit the histograms accurately. As a consequence, we defined the SIs ourselves. We chose functions that are determined by two proportions α_1 and α_2 , which are species and land cover dependent and $\alpha_1 \leq \alpha_2$. Proportions lower than α_1 are assumed to be completely suitable, respectively unsuitable (depending on the land cover), while proportions higher than α_2 are assumed to be completely unsuitable, respectively suitable. In between α_1 and α_2 , a linear relationship is assumed. As such, the neutral SI value of 0.5 corresponds to $(\alpha_1 + \alpha_2)/2$. In order to avoid a denominator equal to zero when using the uninorm aggregation function, completely suitable, respectively unsuitable correspond to SI value 0.99, respectively 0.01. For Forest and Urban canopy cover, the second SI can then be

defined as

$$\tilde{S}_2^{e,s}(P_j^e) = \begin{cases} 0.01 & , \text{ if } 0 \leq P_j^e \leq \alpha_1, \\ 0.01 + (0.99 - 0.01) \frac{P_j^e - \alpha_1}{\alpha_2 - \alpha_1} & , \text{ if } \alpha_1 < P_j^e < \alpha_2, \\ 0.99 & , \text{ if } \alpha_2 \leq P_j^e \leq 1, \end{cases} \quad (6.2)$$

while for Road the second SI can then be defined as

$$\tilde{S}_2^{e,s}(P_j^e) = \begin{cases} 0.99 & , \text{ if } 0 \leq P_j^e \leq \alpha_1, \\ 0.99 - (0.99 - 0.01) \frac{P_j^e - \alpha_1}{\alpha_2 - \alpha_1} & , \text{ if } \alpha_1 < P_j^e < \alpha_2, \\ 0.01 & , \text{ if } \alpha_2 \leq P_j^e \leq 1. \end{cases} \quad (6.3)$$

The two proportions α_1 and α_2 can be determined automatically by varying them in small steps in the unit interval, and selecting the parameter set that gives rise to the best fitting piecewise function. However, in order to reduce computing time, we extracted, for each land cover class and species, two proportions from the relative frequency histograms in Fig. 6.10. More specifically, α_1 and α_2 are equal to the proportions where there is a change in slope of the line connecting the points formed by the relative frequency of proportions lying in each bin and the average proportion in the considered bin. As such, we assume that the proportions where this slope changes indicate a behavioral change of the species towards the land cover class under consideration. Here, Mathematica (Version 11.0, Wolfram Research Inc., Champaign, US) was used to fit the following piecewise function $f^{e,s}(P_j^e)$

$$f^{e,s}(P_j^e) = \begin{cases} S^{e,s}(0) + \frac{S^{e,s}(\alpha_1) - S^{e,s}(0)}{\alpha_1} P_e & , \text{ if } 0 \leq P_j^e \leq \alpha_1 \\ S^{e,s}(\alpha_1) + \frac{S^{e,s}(\alpha_2) - S^{e,s}(\alpha_1)}{\alpha_2 - \alpha_1} (P_e - \alpha_1) & , \text{ if } \alpha_1 < P_j^e < \alpha_2 \\ S^{e,s}(\alpha_2) + \frac{S^{e,s}(1) - S^{e,s}(\alpha_2)}{1 - \alpha_2} (P_e - \alpha_2) & , \text{ if } \alpha_2 \leq P_j^e \leq 1. \end{cases} \quad (6.4)$$

For that purpose, it minimizes the sum of squared errors by applying the optimization algorithms (conjugate) gradient, (quasi) Newton, Nelder-Mead, simulated annealing, differential evolution or by performing a random search. The function automatically picks the method to use based on the type of problem (Wolfram Research Inc., 2018a,b). The fitted functions can be found in Fig. 6.11, while α_1 and α_2 are shown in Table 6.6 per land cover class and species. Analogously to the first SI, only the proportions of Forest and Road are taken into account for Wood Thrush, while this is Road and Urban canopy cover for Gray Catbird. The resulting SIs per land cover class can be found in Fig. 6.11.

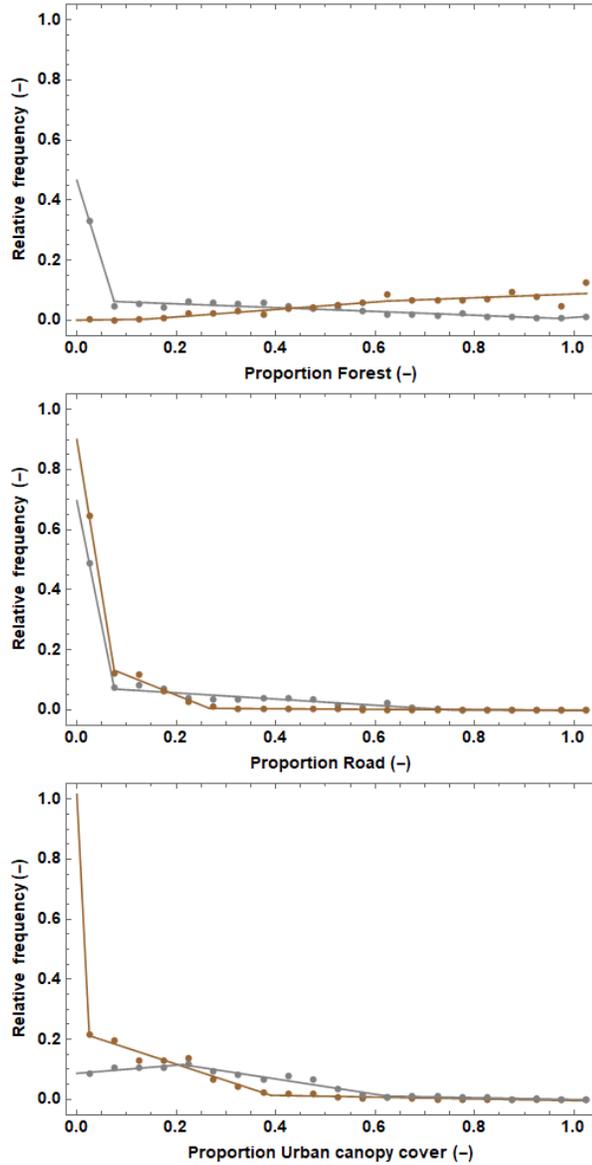


Figure 6.11: Average proportions Forest, Road and Urban canopy cover in each bin (bin width 0.05) of the histograms shown in Fig 6.10 in function of the relative frequency of proportions lying in the bin. The solid lines are the graphs of a piecewise function described by Eq. (6.4). Gray corresponds to Gray Catbird and brown to Wood Thrush.

Table 6.6: The values of parameters α_1 and α_2 per land cover class and species.

| Land use class | Gray Catbird | | Wood Thrush | |
|--------------------|--------------|------------|-------------|------------|
| | α_1 | α_2 | α_1 | α_2 |
| Forest | 0.076 | 0.260 | 0.139 | 0.625 |
| Road | 0.075 | 0.713 | 0.075 | 0.267 |
| Urban canopy cover | 0.173 | 0.625 | 0.025 | 0.390 |

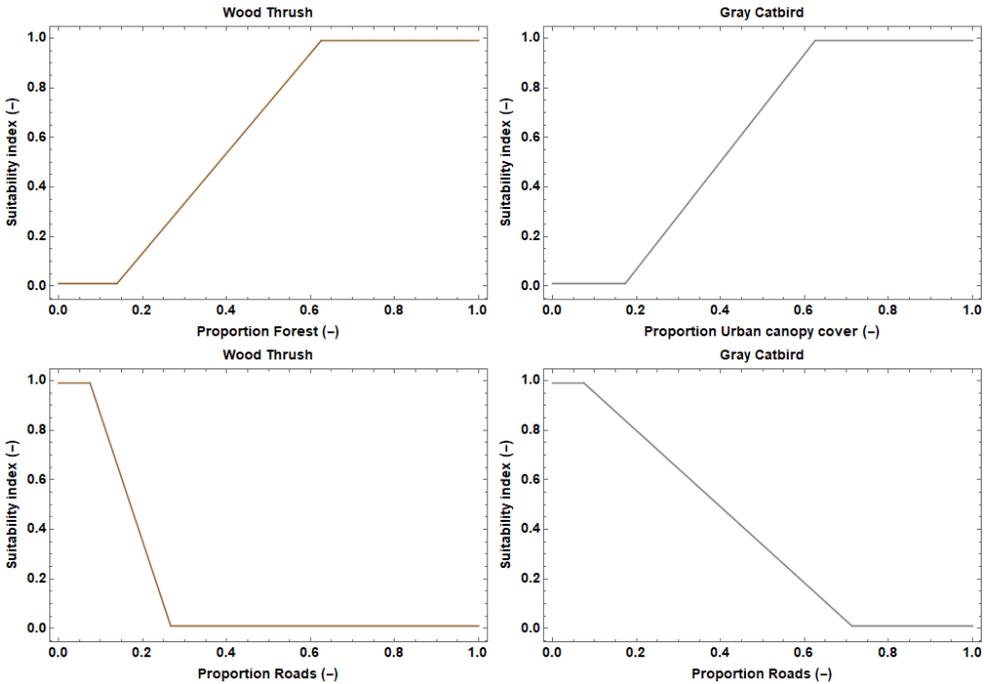


Figure 6.12: Suitability index values as a function of the proportions of Forest and Road for Wood Thrush (left), and as a function of the proportions of Urban canopy cover and Road for Gray Catbird (right).

6.5.2.3 Decision rule

After transforming the proportions of Forest, Road and Urban canopy into SI values and aggregating them into one HSI value \mathcal{H}_j for each angle of movement j , n_a HSI values are obtained. On the basis of these values, one angle of movement needs to be chosen. Here, it should be more likely to choose angles corresponding to higher HSI values. In the most simple case, the angle having the maximum HSI value is chosen. This decision rule is referred to as \mathcal{D}_1 .

It is, however, unlikely that an individual always chooses the best option. Animals in general, and fledglings in particular, explore new regions and might avoid predation or other individuals, which results in a movement angle different from the one with the highest HSI value. Therefore, a weighted choice might be more appropriate, where the weight w_j , *i.e.* the probability of choosing angle of movement α_j , equals:

$$w_j = \frac{\mathcal{H}_j}{\sum_{j=1}^{n_a} \mathcal{H}_j}. \tag{6.5}$$

The weights resulting from the indices \mathcal{H}_j , $j = 1, \dots, n_a$, however, cannot be used directly since they are often very similar, especially when many angles of move-

ment are considered, which results in an almost random choice of the angle of movement. Therefore, a scaling function is used to convert these indices so to obtain more meaningful weights. In literature on genetic algorithms, several scaling functions can be found (Sadjadi, 2004).

The first one selected is called top scaling where the top HSI values are set to the same value, while the others are set to zero (Sadjadi, 2004). Here, however, the $n_a/4$ top HSI values are not set to the same value, but the original index values are used. More specifically, we determine the HSI value at position $n_a/4$ in the list of HSI values sorted in descending order. All HSI values lower than this value are set to zero. If no HSI values with the same value occur, this corresponds to one maintained HSI value if $n_a = 4$, two if $n_a = 8$, four if $n_a = 16$ and nine if $n_a = 36$. Then, a weighted choice is made using these scaled HSI values and Eq. (6.5). The decision rule, using top scaling, is referred to as \mathcal{D}_2 . By only considering the weights of the top 25 %, the number of considered angles of movement decreases so that the resulting probabilities w_j lead to a non-random choice of the angle of movement. Still, for $n_a = 36$, this does not hold and an almost random choice is made out of the top 25 %.

Alternatively, the decision rule can use exponential scaling (\mathcal{D}_3). This involves ranking the n_a HSI values, where each \mathcal{H}_j gets a rank r . Here, we choose for dense ranking where the ranks are consecutive integers starting at 1 and ending at the number of unique HSI values. Rank values are not skipped in the event of ties. Therefore, there will be no gaps in the sequential rank numbering and, as such, equal HSI values receive the same rank (Kellenberger and Groom, 2015). The scaled HSI values are then calculated from

$$\mathcal{H}_j^r = m^{(1-r)}. \quad (6.6)$$

We choose $m = 1.5$, meaning that the value of the scaled HSI equals 1.5 times the value of the scaled HSI of the previous best choice. The scaled HSI values based on the rank of \mathcal{H}_j , $j = 1, \dots, n_a$ can be found in Fig. 6.13, in case that no HSI values with the same value occur. Depending on n_a , only the top n_a dots need to be considered. A lower m implies more similar HSI values and a higher probability of choosing a less suitable option, while a higher m implies a higher probability of choosing the best option. Hence, we argue that 1.5 is an appropriate value for m as it enables new habitat exploration, without ignoring its suitability. In Fig. 6.14, the probabilities of choosing an angle of movement corresponding to rank r , are shown for every considered value of n_a . These probabilities are computed using the scaled HSI values and Eq. (6.5). It can be seen that the probabilities are approximately the same for an arbitrary rank r , irrespective of n_a . As such, the more possible angles of movement, the more options that are considered for ranking, as such it is likely that more suitable options are visited in the end.

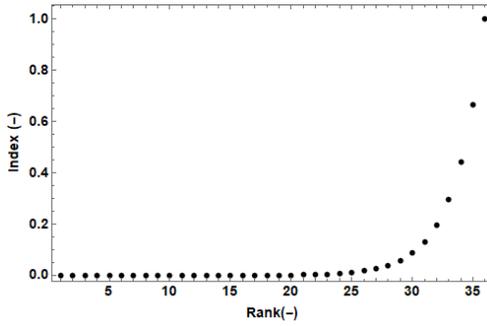


Figure 6.13: HSI values calculated from Eq. (6.6) based on the rank of the original HSI values.

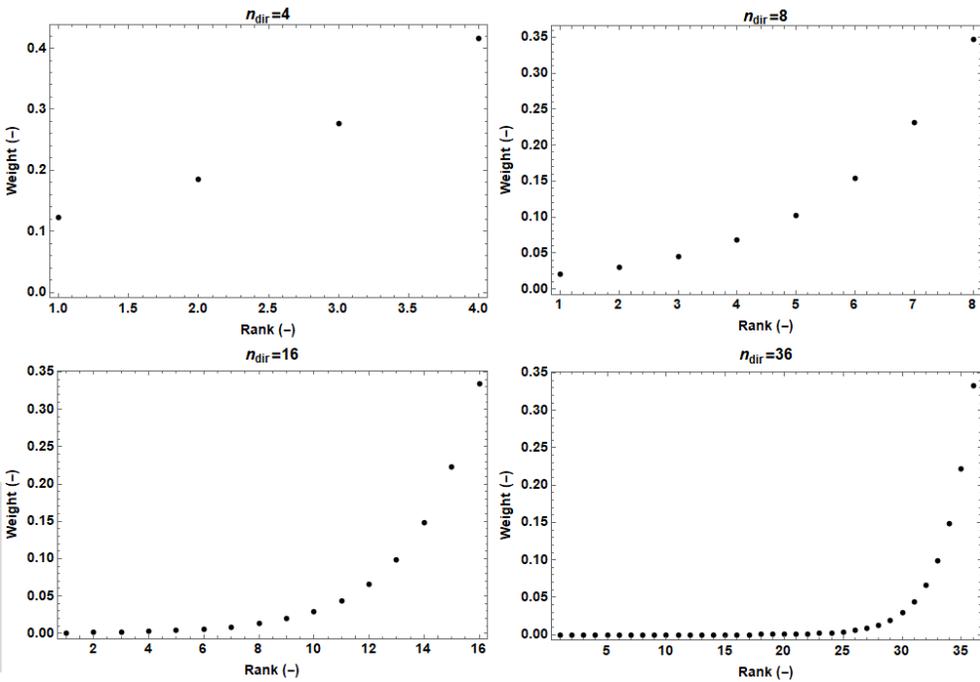


Figure 6.14: The probabilities of choosing the angle of movement as a function of its rank when there are 4, 8, 16 and 36 possible angles of movement.

6.5.3 Model selection

6.5.3.1 Introduction

The performance of the 24 models was evaluated objectively using the Match and the Kolmogorov-Smirnov distances (Rubner et al., 2000; Werman et al., 1985;

Frank and Massey, 1951). For each model, these two distances were calculated for each land cover class, as such, quantifying the discrepancy between the observed and the simulated proportions of the land cover class found in the 95 % TIBB areas defined in between the pairs of consecutive locations (Rubner et al., 2000). The Match distance quantifies the total difference between the EDFs of the observed and simulated proportions (Rubner et al. (2000) and Eq. (3.4)), while the Kolmogorov-Smirnov distance gives the maximum difference between the corresponding EDFs (Rubner et al. (2000) and Eq. (3.5)).

Model selection was performed on the basis of 50 simulated paths per individual and the discretization step of the proportions to construct the EDF was 0.05. Since all data was used to build and validate the models, we investigated the fit of the simulated data to the real-world data, hence the descriptive power of the model. The lower the distances, the more accurate the fit is. Predictive power was tested by means of a cross-validation in Subsection 6.5.4.

6.5.3.2 Gray Catbird

The Match and Kolmogorov-Smirnov distances for each land cover class and each model are shown in Fig. 6.15. The sum of the Match and the sum of the Kolmogorov-Smirnov distances for Forest, Road and Urban canopy cover per model are shown in Fig. 6.16. It can be inferred that Model 9 yields the lowest sums and clearly outperforms the other models, while Model 10 leads to the highest sums. Consequently, the discrepancies between the land use of the in silico and the real-world individuals are the smallest for Model 9 and the largest for Model 10.

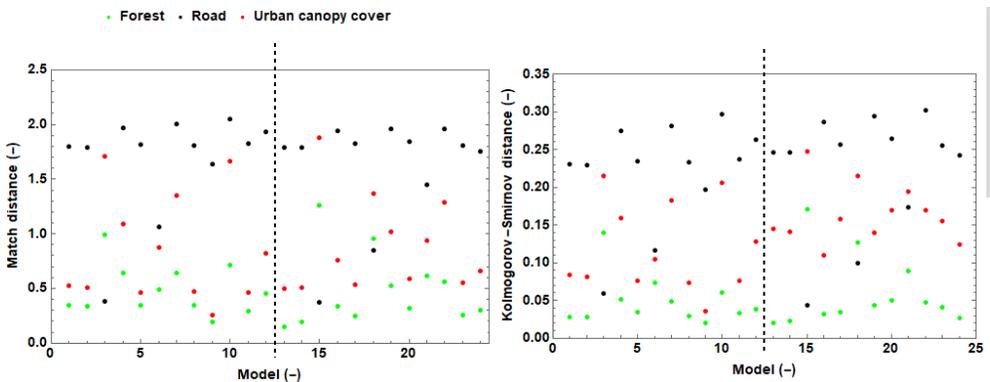


Figure 6.15: The Match and Kolmogorov-Smirnov distances for each land cover class and each model for Gray Catbird.

Since Gray Catbird movement is only governed by the proportions of Road and Urban canopy cover, the sum of the Match and the sum of the Kolmogorov-Smirnov distances for these covers are shown per model in Fig. 6.17. Based on these

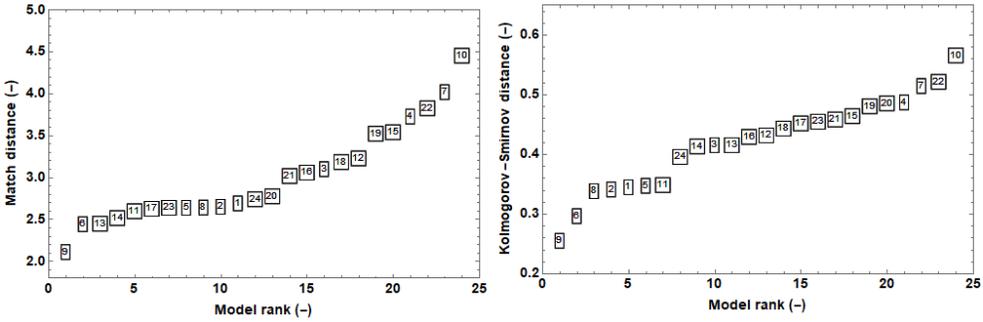


Figure 6.16: The sum of the Match and the sum of the Kolmogorov-Smirnov distances for Forest, Road and Urban canopy cover for Gray Catbird. The sums are sorted in ascending order and the plot markers show the model IDs as listed in Table 6.4.

sums, Models 3, 6 and 9 can be identified as best performing and Model 10 as performing the worst, which corroborates the findings when all land cover classes are considered.

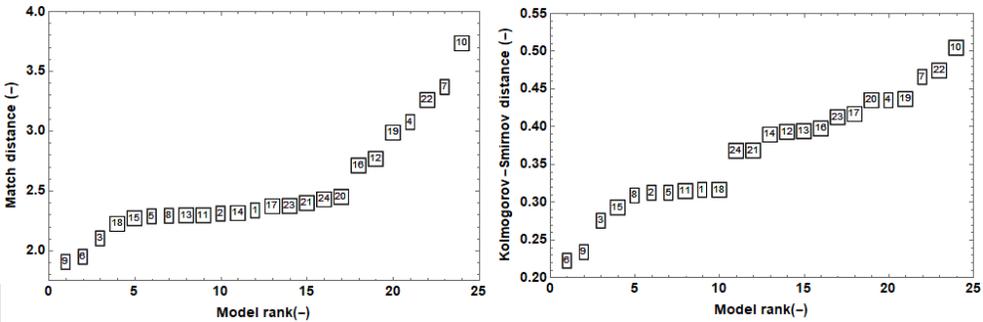


Figure 6.17: The sum of the Match and the sum of the Kolmogorov-Smirnov distances for Road and Urban canopy cover for Gray Catbird. The sums are sorted in ascending order and the plot markers show the model IDs as listed in Table 6.4.

The relative frequency histograms (bin width 0.05) of proportions Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between the pairs of consecutive locations are shown for the Models 3, 6 and 9, *i.e.* the three best models, in Fig. 6.18. For Model 3, the proportions of Road are simulated correctly, while the relative frequencies of the simulated proportions of Forest and Urban canopy cover in bin $[0, 0.05]$ are too high. These decrease for Model 6, which leads to a higher accuracy. For Model 9, the proportions of Forest and Urban canopy cover are simulated very accurately, but the proportions of Road are shifted towards low proportions.

Considering the results in Figs. 6.15–6.18, Model 9, which is based on suitability index $S^{e,s}(P_f^e)$, decision rule D_3 and $n_a = 16$, may be considered as the best model

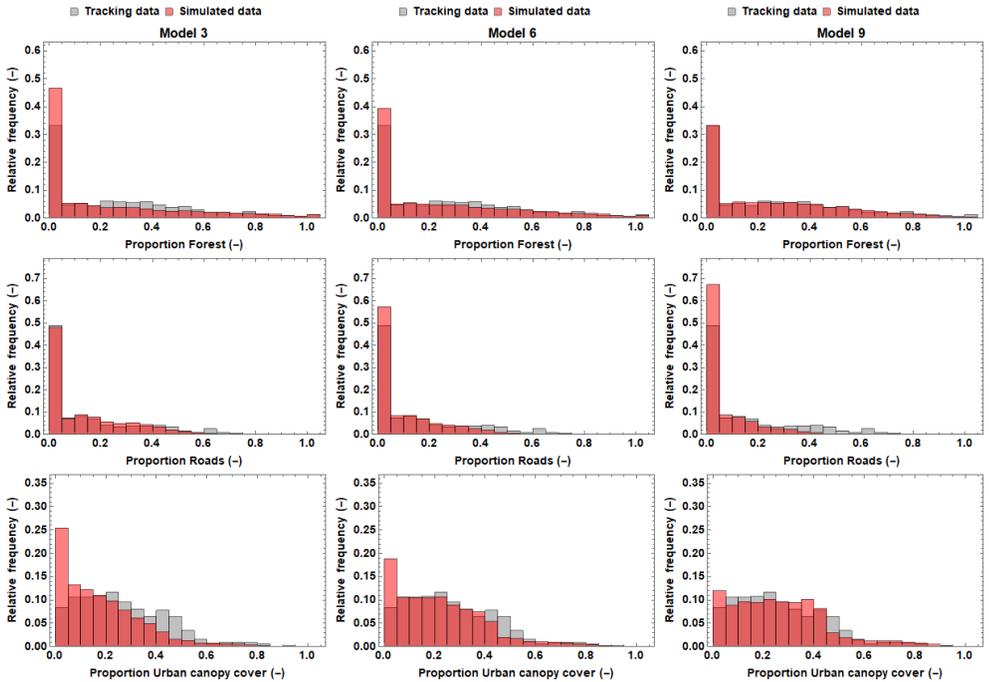


Figure 6.18: The relative frequency histograms (bin width 0.05) of the proportions Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between the pairs of consecutive locations for Gray Catbird. Both the histograms based on the observed and on the simulated proportions by Models 3, 6 and 9, are shown.

to simulate daily locations for Gray Catbird, so that the land use of the *in silico* individuals is similar to the one of the real-world individuals.

6.5.3.3 Wood Thrush

The Match and Kolmogorov-Smirnov distances for each land cover class and each model are shown in Fig. 6.19. The sum of the Match and the sum of the Kolmogorov-Smirnov distances for Forest, Road and Urban canopy cover per model are shown in Fig. 6.20. It can be inferred that Models 21 and 24 have the lowest sums and hence outperform the other models. Model 15 has the highest Match distance sum, while Models 7 and 10 have the highest Kolmogorov-Smirnov distance sums. As such, the discrepancies between the land use of the *in silico* and the real-world individuals are the least pronounced for Models 21 and 24, and the most pronounced for Models 7, 10 and 15.

Since Wood Thrush movement is only governed by proportions of Forest and Road, the sum of the Match and the sum of the Kolmogorov-Smirnov distances for these land cover classes are shown per model in Fig. 6.17. Based on these sums, Mod-

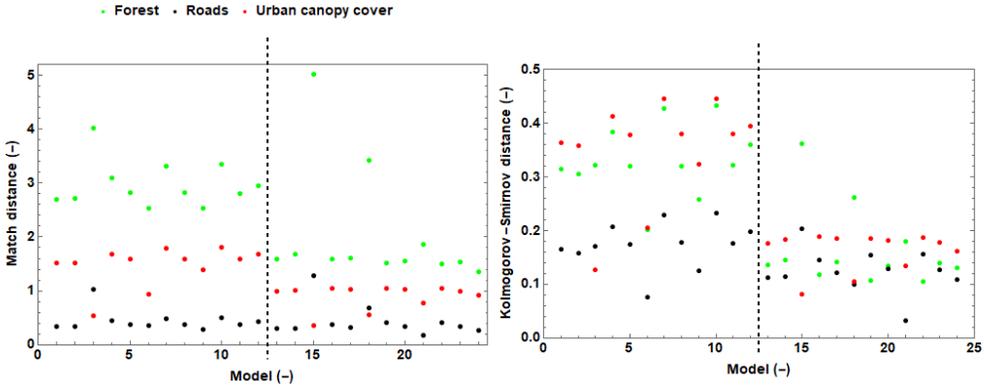


Figure 6.19: The Match and Kolmogorov-Smirnov distances for each land cover class and each model for Wood Thrush.

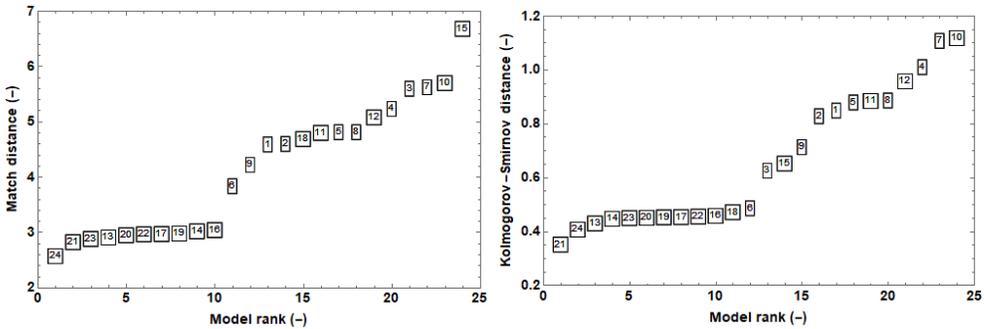


Figure 6.20: The sum of the Match and the sum of the Kolmogorov-Smirnov distances for Forest, Road and Urban canopy cover for Wood Thrush. The sums are sorted in ascending order and the plot markers show the model IDs as listed in Table 6.4.

6

els 21 and 24 are again identified as the best and Models 7, 10 and 15 as the worst.

The relative frequency histogram (bin width 0.05) of the proportions Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between the pairs of consecutive locations are shown for Models 21 and 24 in Fig. 6.22. For both models, the proportions of Road are simulated accurately, while the relative frequency histograms of Forest and Urban canopy cover have some bins indicating considerable differences between the simulated and observed relative frequencies. Especially bin [0, 0.05] has a higher simulated relative frequency, as compared to the observed one. This indicates that there are too few areas visited by the in silico individuals that are labeled as Forest or Urban canopy cover. As such, the use of more natural habitat might be underestimated.

Considering the results in Figs. 6.19–6.22, both Models 21 and 24 may be consid-

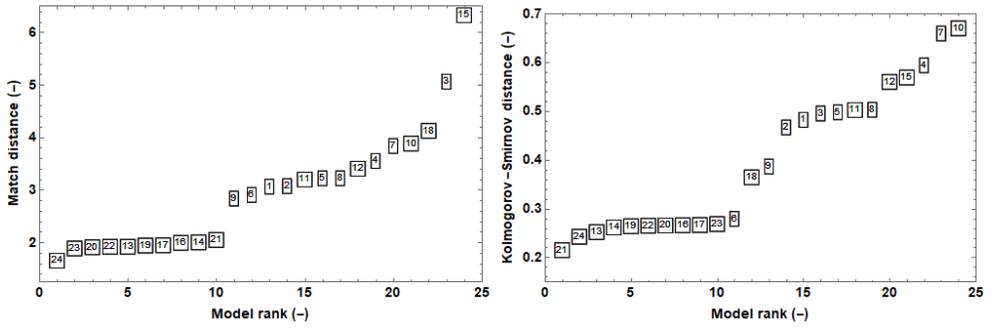


Figure 6.21: The sum of the Match and the sum of the Kolmogorov-Smirnov distances for Forest and Road for Wood Thrush. The sums are sorted in ascending order and the plot markers show the model IDs as listed in Table 6.4.

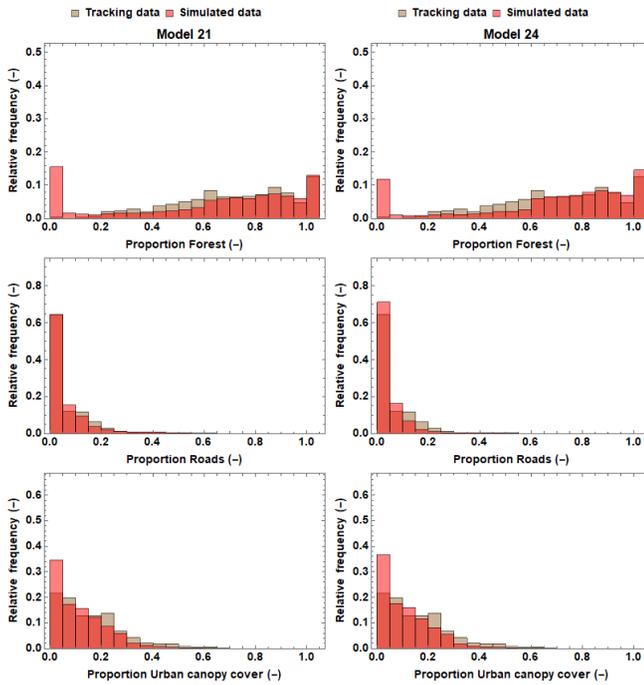


Figure 6.22: The relative frequency histograms (bin width 0.05) of the proportions Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between the pairs of consecutive locations for Wood Thrush. Both the histograms based on the observed and on the simulated proportions by Models 21 and 24, are shown.

ered as the best models for Wood Thrush. They both use suitability index $\tilde{S}^{e,s}(P_j^e)$ and decision rule \mathcal{D}_3 , and n_a is 16 for Model 21 and n_a is 36 for Model 24.

6.5.3.4 Variability among simulations

Using the Match and Kolmogorov-Smirnov distances, it was concluded that Model 9 for Gray Catbird, and Models 21 and 24 for Wood Thrush simulated proportions Forest, Road and Urban canopy cover with the highest accuracy. Thus far, these distances were calculated on the basis of the proportions of Forest, Road and Urban canopy cover obtained from 50 simulated paths per individual. To assess variability in descriptive power among the 50 repetitions of the simulation of one path per individual, the Match and Kolmogorov-Smirnov distances were calculated for each repetition. As such, 50 values of both distances were obtained per land cover class, for which box plots are shown in Fig. 6.23. Here, it can be seen that Model 9 for Gray Catbird leads to lower distances than Models 21 and 24 for Wood Thrush for the proportions of Forest and Urban canopy cover, while it yields higher distances for Road. Furthermore, it can be seen that Forest leads to a higher variability in Match distances as compared to the other land use classes. Considering the Kolmogorov-Smirnov distances, there are no remarkable differences in variability noticeable.

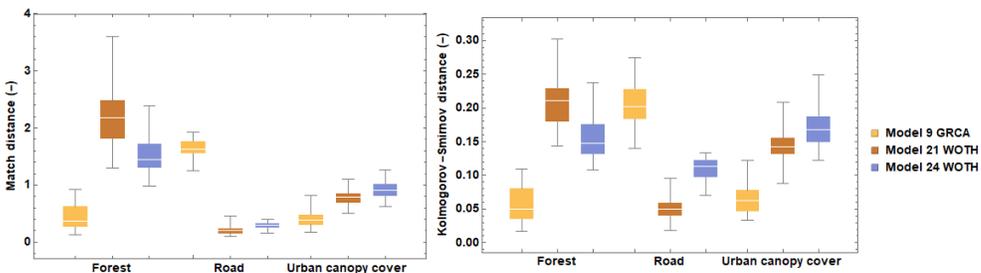


Figure 6.23: A box plot for the Match (left) and Kolmogorov-Smirnov (right) distances for Forest, Road and Urban canopy cover that were obtained from simulating 50 times one path per individual.

6

6.5.4 Model performance and cross-validation

6.5.4.1 Introduction

The predictive power of Model 9 for Gray Catbird and Models 21 and 24 for Wood Thrush was verified by means of a cross-validation (Elith and Leathwick, 2009). Therefore, the tracking data of each species was split in five folds and each of them was considered once as test set to evaluate the predictive power of the model calibrated using the remaining data (training set). For Gray Catbird, tracking data of 32 individuals were available. As such, each test set contains tracking data of six or seven individuals. For Wood Thrush, tracking data of 49 Wood Thrush

individuals were available and each test set contains tracking data of nine or ten individuals.

One round of cross-validation involved the simulation of consecutive locations for the individuals present in a test set. The starting locations and numbers of simulated days of the *in silico* individuals corresponded to the ones of the real-world individuals in the considered test set. The SI was calibrated for each round of cross-validation from the training set. More specifically, for Gray Catbird, \bar{P}_R and \bar{P}_U , and for Wood Thrush, α_1 and α_2 for Forest and Road were tuned using the training set. The Match and Kolmogorov-Smirnov distances were calculated between the relative frequency histograms, with bin width 0.05, of the simulated and real-world proportions Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between each pair of consecutive locations.

A five-fold cross-validation was performed. The model parameters for each fold can be found in Table 6.7. The values of \bar{P}_R , \bar{P}_U , α_1 and α_2 for Road are similar across the test sets and similar to the values obtained when using all data as training set (cfr. Tables 6.6 and 6.5). For Wood Thrush, however, α_1 and α_2 for Forest are sometimes mutually different and differ as well from α_1 and α_2 obtained when using all data as training set, which might indicate that there is quite some variability among the Wood Thrush individuals.

Table 6.7: Calibrated model parameters using the training set for each fold of cross-validation.

| Model parameter | Gray Catbird | | Wood Thrush | | | |
|-----------------|--------------|-------------|----------------------|----------------------|--------------------|--------------------|
| | \bar{P}_R | \bar{P}_U | α_1 Forest | α_2 Forest | α_1 Road | α_2 Road |
| Test set 1 | 0.125 | 0.269 | 0.025 | 0.770 | 0.075 | 0.267 |
| Test set 2 | 0.160 | 0.257 | 0.269 | 0.625 | 0.075 | 0.288 |
| Test set 3 | 0.148 | 0.265 | 0.217 | 0.625 | 0.075 | 0.268 |
| Test set 4 | 0.137 | 0.277 | 0.375 | 0.608 | 0.075 | 0.211 |
| Test set 5 | 0.160 | 0.260 | 0.259 | 0.602 | 0.075 | 0.282 |

The Match and Kolmogorov-Smirnov distances for each of the five folds, together with the average distances, can be found in Tables 6.8– 6.10. As a reference, the distances obtained when using all tracking data in the calibration and validation stage, are also shown.

To verify whether the observed and simulated proportions of each land cover class might be drawn from the same distribution function, a two sample Kolmogorov-Smirnov test (2-sided) was applied (Press, 2007). Here, we can reject, to a certain required level of significance, the null hypothesis that the observed and simulated proportions of a land cover class might be drawn from the same distribution function. Rejecting the null hypothesis indicates that the proportions are from dif-

ferent distributions. Failing to reject the null hypothesis, only indicates that the proportions might be consistent with a single distribution function (Press, 2007). If the Kolmogorov-Smirnov distance is larger than a critical value determined by the level of significance, here, 0.05, and by the number of observed and simulated proportions, the null hypothesis is rejected (Pearson and Hartley, 1972; Université de Montréal, 2018).

6.5.4.2 Gray Catbird

The reference distances for Gray Catbird (Table 6.8), obtained when using all data as training set, are the lowest for Forest and the highest for Road. As such, the proportions of Forest were simulated with the highest accuracy by Model 9, followed by the proportions of Urban canopy cover and finally, the proportions of Road. This is also shown by the two-sample Kolmogorov-Smirnov test, *i.e.* the observed and simulated proportions of Forest and Urban canopy cover might be drawn from the same distribution function, while this is not the case for Road. Alternatively, this might as well be inferred from the relative frequency histograms (bin width 0.05) of proportions Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between the pairs of consecutive locations (Fig. 6.18). Here, the maximum and the sum of the absolute differences between the observed and simulated relative frequencies in each bin, are 0.0180 and 0.0828 for Forest, 0.1828 and 0.3932 for Road and 0.0358 and 0.2155 for Urban canopy cover.

For what concerns the predictive power of Model 9 for Gray Catbird, it can be concluded from the average distances that the proportions of Forest and Urban canopy cover were simulated with the highest accuracy, whereas the proportions of Road were simulated with the lowest accuracy. Furthermore, the two-sample Kolmogorov-Smirnov tests showed that the proportions Forest and Urban canopy cover were accurately simulated for two test sets, while the proportions of Road were accurately simulated for one test set. The null hypothesis was rejected for every land cover class for two test sets. As such, we cannot substantiate the predictive power of Model 9. Still, for some of the land cover classes, especially Forest and Urban canopy cover, the proportions were accurately simulated.

Table 6.8: The Match and Kolmogorov-Smirnov (K-S) distances for a five-fold cross-validation, together with the average distances for Gray Catbird. As a reference, the distances obtained when using all data is used in the calibration and validation stage, are shown. If the observed and simulated proportions can be consistent with a single distribution function according to a two-sample Kolmogorov-Smirnov test, an asterisk (*) is shown.

| Land use class | Forest | | Road | | Urban canopy cover | |
|----------------|--------|--------|-------|--------|--------------------|--------|
| | Match | K-S | Match | K-S | Match | K-S |
| Test set 1 | 0.368 | 0.089* | 3.358 | 0.332 | 0.468 | 0.071* |
| Test set 2 | 1.301 | 0.14 | 0.447 | 0.066* | 0.788 | 0.148 |
| Test set 3 | 0.415 | 0.089* | 1.2 | 0.176 | 0.414 | 0.061* |
| Test set 4 | 0.54 | 0.084 | 2.09 | 0.342 | 1.276 | 0.188 |
| Test set 5 | 2.671 | 0.243 | 0.642 | 0.104 | 1.24 | 0.181 |
| Mean | 1.059 | 0.129 | 1.547 | 0.204 | 0.837 | 0.13 |
| Reference | 0.199 | 0.021* | 1.639 | 0.197 | 0.258 | 0.036* |

6.5.4.3 Wood Thrush

Let us now look at the distances for Wood Thrush when Model 21 is considered (Table 6.9). Here, the reference distances are the lowest for Road and the highest for Forest. As such, the proportions of Road were simulated with the highest accuracy, followed by the proportions of Urban canopy cover and finally, the proportions of Forest. The two-sample Kolmogorov-Smirnov tests showed that the simulated proportions of Road may be drawn from the same distribution function, while the null hypothesis was rejected for Forest and Urban canopy cover. Also a visual inspection of the relative frequency histograms (bin width 0.05) of proportions Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between the pairs of consecutive locations, showed that Model 21 was not able to accurately simulate proportions Forest and Urban canopy cover. Especially the bin [0, 0.05] has a high simulated relative frequency compared to the observed one, which is not desired (Fig. 6.22). The maximum and the sum of the absolute differences between the observed and simulated relative frequencies in each bin, are 0.1760 and 1.0232 for Forest, 0.1548 and 0.4963 for Road and 0.2617 and 0.7883 for Urban canopy cover.

For what concerns the predictive power of Model 21 for Wood Thrush, it can be concluded on the basis of the average distances that the proportions of Road were simulated with the highest accuracy, followed by the proportions of Urban canopy cover and Forest. Furthermore, the two-sample Kolmogorov-Smirnov tests showed that the proportions Road were accurately simulated for three test sets, and the null hypothesis was rejected for every land cover class for two test sets. As such, we cannot substantiate the predictive power of Model 21 for every test set and land cover class. Still, proportions of Road were accurately simulated in three out

of five test set sets. Proportions Forest and Urban canopy cover were not simulated accurately.

Table 6.9: The Match and Kolmogorov-Smirnov (K-S) distances for a five-fold cross-validation, together with the average distances. Here, Wood Thrush and Model 21 are under consideration. As a reference, the distances obtained when using all data is used in the calibration and validation stage, are shown. If the observed and simulated proportions can be consistent with a single distribution function according to a two-sample Kolmogorov-Smirnov test, an asterisk (*) is shown.

| Land use class | Forest | | Road | | Urban canopy cover | |
|----------------|--------|-------|-------|--------|--------------------|-------|
| | Match | K-S | Match | K-S | Match | K-S |
| Test set 1 | 2.225 | 0.199 | 0.256 | 0.047* | 1.087 | 0.218 |
| Test set 2 | 1.136 | 0.098 | 0.133 | 0.038* | 0.452 | 0.135 |
| Test set 3 | 1.135 | 0.104 | 0.203 | 0.045* | 0.889 | 0.184 |
| Test set 4 | 4.697 | 0.395 | 0.648 | 0.120 | 0.920 | 0.176 |
| Test set 5 | 2.707 | 0.251 | 0.405 | 0.135 | 0.753 | 0.261 |
| Mean | 2.38 | 0.209 | 0.329 | 0.077 | 0.820 | 0.195 |
| Reference | 1.859 | 0.180 | 0.167 | 0.033* | 0.774 | 0.134 |

Finally, Model 24 for Wood Thrush is considered (Table 6.9). Similar to Model 21, the reference distances are the lowest for Road and the highest for Forest and Urban canopy cover. The null hypotheses of the two-sample Kolmogorov-Smirnov tests were rejected for all land cover classes. Also a visual inspection of the relative frequency histograms (bin width 0.05) of proportions Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between the pairs of consecutive locations, showed that Model 24, like Model 21, was not able to accurately simulate proportions Forest and Urban canopy cover (Fig. 6.22). The maximum and sum of absolute differences between the observed and simulated relative frequencies in each bin, are 0.2133 and 1.1863 for Forest, 0.2233 and 0.6251 for Road and 0.2830 and 0.8326 for Urban canopy cover.

For what concerns the predictive power of Model 24 for Wood Thrush, it can be inferred on the basis of the average distances that the proportions of Roads were simulated with the highest accuracy. Furthermore, the two-sample Kolmogorov-Smirnov tests showed that the proportions of Road were accurately simulated for one test set and the null hypothesis was rejected for every land cover class for four test sets. As such, we fail to prove the predictive power of Model 24 for every test set and land cover class.

6

Table 6.10: The Match and Kolmogorov-Smirnov (K-S) distances for a five-fold cross-validation, together with the average distances. Here, Wood Thrush and Model 24 are under consideration. As a reference, the distances obtained when using all data is used in the calibration and validation stage, are shown. If the observed and simulated proportions can be consistent with a single distribution function according to a two-sample Kolmogorov-Smirnov test, an asterisk (*) is shown.

| Land use class | Forest | | Road | | Urban canopy cover | |
|----------------|--------|-------|-------|--------|--------------------|-------|
| | Match | K-S | Match | K-S | Match | K-S |
| Test set 1 | 2.024 | 0.179 | 0.314 | 0.118 | 1.282 | 0.260 |
| Test set 2 | 1.260 | 0.122 | 0.351 | 0.107 | 0.665 | 0.165 |
| Test set 3 | 1.091 | 0.117 | 0.431 | 0.116 | 1.070 | 0.215 |
| Test set 4 | 3.452 | 0.322 | 0.267 | 0.067* | 1.043 | 0.193 |
| Test set 5 | 2.111 | 0.178 | 0.677 | 0.193 | 0.733 | 0.239 |
| Mean | 1.988 | 0.184 | 0.408 | 0.120 | 0.959 | 0.214 |
| Reference | 1.358 | 0.130 | 0.271 | 0.110 | 0.919 | 0.162 |

Finally, it can be seen that Model 9 for Gray Catbird leads to lower distances than Models 21 and 24 for Wood Thrush for the proportions of Forest and Urban canopy cover, while it yields higher distances for Road. Both Models 21 and 24 for Wood Thrush have a similar performance, so we were not able to indicate a single model of preference.

6.5.5 Discussion

In this section, some preliminary ideas are given concerning the construction of a model to simulate consecutive locations of individual Gray Catbird and Wood Thrush fledglings during the post-independent period on a daily basis in such a way that the land use of the *in silico* individuals is similar to the one of the real-world individuals. This allows to establish the foundations for a model that helps predicting how populations of these fledglings respond to increasing rates of land cover changes. Ultimately, the responses of both species under consideration towards (locally) increased or decreased urbanization can be studied in detail, as such creating a tool for making informed decisions that will benefit a sustainable resource management.

Our modeling approach involves simulating subsequent movement angles and step lengths. Here, the step lengths are sampled from an EDF constructed on the basis of the tracking data. Each angle of movement is chosen among n_a options on the basis of the prevailing proportions of Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between the current location and the n_a possible next locations. Several alternatives for n_a , SI and the decision rule are considered. The models that simulated proportions Forest, Road and Urban

canopy cover with the highest accuracy are Model 9 for Gray Catbird and Model 21 and 24 for Wood Thrush.

Considering the obtained distances (Tables 6.8–6.10) and the relative frequency histograms (bin width 0.05) of the observed and simulated proportions Forest, Road and Urban canopy cover found in the 95 % TIBB areas defined in between the pairs of consecutive locations (Figs. 6.18 and 6.22), it is clear that these models did not accurately simulate the observed proportions. Hence, we failed to prove their predictive power. Nevertheless, some of the land cover classes were accurately simulated for some test sets and when using all data as training set.

To improve the model performance, we suggest to widen the data analysis so that more insight into the behavior of the Gray Catbird and Wood Thrush fledglings during the post-independent period are obtained. Here, it might be valuable to investigate the variability in behavior among the individuals, other relevant (environmental) characteristics that might influence movement behavior, the correlations between step length, movement angle and land cover classes, and so on.

Furthermore, it might be necessary to consider different model structures. For example, a diffusion coefficient could be introduced, or other SIs and decision rules might be explored. For instance, a sigmoid function, or a triangular or trapezoidal function could be chosen as SI. Moreover, it might not be appropriate to determine α_1 and α_2 by the proposed piecewise linear function. Alternatively, we could introduce some (a priori) restrictions on the values of the model parameters, or they could be chosen on the basis of quantiles or could be determined automatically by varying them in small steps.

As illustrated, there are many options to improve the spatially explicit movement model. First, however, we need to gain a more detailed insight into the behavior before making any changes to the evaluated models. Furthermore, adaptations should be made in the light of the ultimate goal of the model, that is, estimating the effect of further urbanization on the post fledglings. If, for example, habitat fragmentation needs to be considered, next to habitat loss, this environmental characteristic should be quantified and included. At the species level, survival rates might be taken into account to quantify the effect of urbanization. All together, it is clear that more work is needed before land use management can be based on the simulations obtained from them.

Since the definition of the SIs is highly adaptable, the presented modeling methodology – including model construction, model selection and (cross-) validation – can be adopted for modeling the movement of other species. Therefore, the SIs should be adapted to account for preferences, a time step should be chosen and information on typical step lengths should be included. In this section, we assumed that movement was possible in any direction. However, movement of species might be limited by barriers. Therefore, it might be necessary to define restrictions on the step lengths and turning angles. This can, for example, be done

by combining SIs by taking the minimum instead of using the uninorm aggregation function. Given the limited amount of environmental and behavioral information that is taken into account in our methodology, it is a simple and easy way to make a first attempt in modeling the movement behavior of a species under consideration. If this attempt would already give some indication that the model approximates the species behavior, it might serve as a good foundation that can be extended and fine tuned as desired.

7

Conclusions and perspectives

As a consequence of technological advances, people and a wide range of animals can be automatically tracked in high resolution over a long period of time. This results in a huge amount of data containing information on the behavior of the tracked individuals and the characteristics of the environment they are moving through. These detailed movement trajectories call for appropriate analysis and modeling methods. As illustrated in Chapter 2, there are many methods available. The challenge lies in selecting the most appropriate ones that, whilst accounting for the available tracking data and the research question at stake, extract efficiently the desired information. Therefore it is important to get familiar with the unique vocabulary in movement science, the particular characteristics of tracking data, the prevailing analysis tools and the available types of movement models. As such, fundamental knowledge is obtained to choose the most appropriate methods. It might, however, be necessary to adapt or to combine methods to fit the research question or the available tracking data at stake. It might even be necessary to develop new, custom-made methods.

In this research, we are joining this effort and selected, adapted and developed mathematical methods and models to help answering research questions. To reach this goal, available tracking data were used to gain knowledge on how, when,

where and why individuals move, and to parameterize, verify and validate models. The data analysis and obtained models should eventually help a well-founded decision making. For this research, we relied on existing tracking datasets. The specific research objectives were:

1. Build a spatially explicit marching model of the Belgian major endurance event '100 km Dodentocht' to predict the effect of a modified starting procedure;
2. Derive the analytical description of the time-integrated Brownian bridge;
3. Unravel possible sources of poisoning for the Eurasian Eagle Owl (*Bubo bubo*) in Limburg, the Netherlands;
4. Estimate the effects of urbanization on post-fledging space use of two bird species, *i.e.* the Gray Catbird (*Dumetella caroliniensi*) and the Wood Thrush (*Hylocichla mustelina*), which are phylogenetically closely related though exhibit differential responses to urbanization.

Several methods for analyzing and modeling movement phenomena were used. An overview can be found in Table 7.1. Due to the fact that each data set contained data of a different type (VHF radio, satellite, GPS and so on), and considers a different species, region and research question, all analysis and modeling approaches were specifically chosen. There was no one-fits-all approach. Some of the applied methods already existed, some were adapted and some were newly developed. In the following, we discuss each applied methodology per research objective and highlight the obtained generic added value. More specifically, insight is given in the limitations, advantages and perspectives.

Table 7.1: An overview of the studied tracking data sets, the type of tracking data, spatial dimension and the analysis and modeling approaches used to solve the research question.

| Tracking data set | Tracking data type | Space dimension | Analysis and modeling approaches |
|--------------------------|--------------------|-----------------|--|
| 100 km Dodentocht | passive RFID | 1D | calculation of general movement metrics, correlation analysis, IBM, spatially-explicit marching model |
| Eurasian Eagle Owl | GPS | 2D | expert-based HSI map, TIBB, movement metric-based approach for determining distinct movement behavior, convex hull, hotspot analysis, ENFA |
| Gray Catbird/Wood Thrush | VHF radio | 2D | calculation of movement metrics, TIBB, calculation of the weighted relative proportions and area of each land cover class in the home range, generalized linear mixed-effects models, IBM, spatially-explicit movement model |

The modeling methodology – including model construction, selection and validation – presented for the Belgian major endurance event '100 km Dodentocht'

(Chapter 3) is, to the best of our knowledge, the first of its kind aiming at building a spatially explicit model that mimics the dynamics of such an endurance event. We concluded that our model is not able to predict the exact number of participants passing by at every time interval, but it is nevertheless useful to predict when it will be busy/calm at the checkpoints. Furthermore, the model was used to predict the effect of a modified starting procedure. For this, three scenarios were tested to reduce the crowdedness at the start of the event. It was concluded that either the start should be moved to a location outside the town center, where streets are at least 25% wider, or that the marchers should start in two groups at two different locations, and that these groups should ideally merge at about 20 km after the start.

Our modeling methodology completely relied on an RFID dataset. In our study, the data analysis uncovered the effects of age, gender and previous marching speeds on the marching speeds along each section and on retiring. The obtained patterns were encoded in terms of conditional distribution functions and copulas, which were derived from the data. As a consequence, our presented methodology needs a large amount of data to reliably define the conditional experimental distribution functions (EDFs). If this data is available, the methodology is relatively easy to implement. Otherwise, our methodology cannot be used. Since most endurance (walking, running, cycling) events have a predefined route with several checkpoints and make use of RFID tracking systems to follow their participants, our methodology can easily be used to model similar events. Next to RFID data, other types of tracking data are suitable for defining the conditional EDFs. Bluetooth tracking data, for example, could also be used.

We believe that we could have obtained a better model performance if more influencing factors would have been taken into account. It would, for example, be interesting to quantify the effect of weather conditions. As such, organizers could better prepare for extremely hot or rainy events. Also an inclusion of the effect of trail layout would be valuable to indicate possible bottlenecks, to define optimal checkpoint locations and so on. To be able to include more influencing factors, sufficient data should be available to quantify their effect. On the basis of the available information for the '100 km Dodentocht', we were not able to include other influencing factors next to age, gender and previous marching speeds. For other events, other relevant information concerning factors that might influence the event dynamics might be available. Therefore, it is essential to perform a data analysis to determine the factors that need to be included in the marching model prior to building the model.

Furthermore, we believe that our model performance could be improved if resting periods would be taken into account. Therefore, we had a look at the available GPS tracking data. However, since not all participants take a rest at the same time and since the resting periods vary, we could not draw general conclusions about resting that could have been used to correct the marching speeds obtained from

the RFID tracking data. If more GPS tracking data would have been available and if this data would have been better structured so that the time step would have been smaller and more similar over the registrations, this dataset would have been more useful for both data analysis and modeling. Marching speeds could then, for example, be defined as an average of the instantaneous marching speeds along a section, as such, accounting for breaks. Alternatively, one might think about a more detailed model where sections are not considered as a whole and movement along sections is simulated in detail. With high resolution GPS tracking data it might even be possible to build a two-dimensional model. Again, if sufficient data is available to uncover overall patterns and to define conditional EDFs, our model methodology can be applied. Otherwise, it is not suitable.

If a different setting is considered, where people are not moving from a starting to an ending location along a predefined trail, our modeling methodology, where the information contained in a large amount of tracking data is used to steer movement, will probably not be suitable. While moving, individuals want to reach some internal goal and are influenced by the (constraints of the) environment he/she is moving through. Both the goal and the environment can change whilst moving. As such, the setting to be modeled is more dynamic and a lot of influencing factors need to be taken into account. Even if a lot of data would be available, it would still be difficult to extract the desired information and to find enough individuals with the same characteristics to decide on general behavior or to build experimental distribution functions. Some intermediate setting, where people, for example, want to move from a starting to an ending location and where a limited number of routes are available, might however be doable.

The presented spatially explicit model may be used to assess the possible effects of changes to the trail, the organization and the characteristics of the marchers on the marching speeds and the resulting passing times. It is crucial to know the impact of such modifications on the event dynamics before implementing them, but aside from relying on a mathematical model, there is no way to accomplish this. The only condition is that the considered scenario can be translated into changing model parameters. The organizing committee of an event can also make a first estimation of the effect of a scenario based on their experience, which is also valuable. For them, a scenario analysis with a movement model might confirm that what they are planning to do, is valuable or not. Furthermore, by making use of a model, quantification of the impact is possible.

The added value of our and other models mimicking the dynamics of mass events can be further improved by the inclusion of information on medical interventions, evacuation and the supply chain along the trail. Furthermore, a dynamic model, where model results are updated with information gathered during the ongoing event, would be highly informative. Ultimately, a dynamic model integrating the information of all involved parties should be constructed in order to guarantee a smooth and safe organization. Ideally, this model should be highly adaptable to

other events and a tool should be available to ensure an easy model implementation, automatic calculation and visualization of the desired information.

In Chapter 4, we derived the analytical description of the time-integrated Brownian bridge (TIBB). The presented expression can be applied when considering basic Brownian bridges (Horne et al., 2007), dynamic Brownian bridges (Kranstauber et al., 2012), biased random bridges (Benhamou, 2011) and bivariate Gaussian bridges (Kranstauber et al., 2014). As a consequence, the results will be more accurate, the calculations less time consuming and the selection of an appropriate discretization step in time unnecessary. In order to make a difference for ecologists, the analytical description should be included in the R BBMM- and move-packages.

To unravel possible sources of poisoning for the Eurasian Eagle Owl (*Bubo bubo*) in Limburg, the Netherlands, two approaches were implemented to highlight environmental characteristics (ECs) and areas important for the species in the region (Chapter 5). Both approaches demonstrated that the former quarries in the region are very important for the Eurasian Eagle Owls since they are used as nesting sites. Other ECs that attract the species are pastures, differences in altitude, forest edges and water bodies. Major roads and town centers will most likely be avoided. Highlighted areas and ECs can be used to help prioritizing areas that need protection. Furthermore, they can help in setting up a sampling strategy to further investigate possible sources of pollution since samples are best taken in areas that attract the species. In this chapter, we mainly selected and combined already existing methods. As such, the overall added value of this chapter concerning methodology is limited.

In the first approach, literature and expert knowledge was used to determine the ECs that attract or repel the species in the region. The importance of each EC was quantified by suitability indices (SIs). Here, we chose 0.5 as a neutral value so that it is more easy to weigh the chosen SIs and motivate them when making use of qualitative information coming from literature and expert knowledge. To keep a neutral value of 0.5 when aggregating the SI values per pixel, we relied on a uninorm aggregation function (De Baets and Fodor, 1999). This has, to the best of our knowledge, never been done before in this context. In fact, all kind of functions can be chosen for aggregating SIs but the aggregation should lead to habitat suitability index (HSI) values in the unit interval and it should reflect the relationship between the environmental characteristics (U.S. Fish and Wildlife Service, 1981; Van Horne and Wiens, 1991). However, none of the often used ones kept the neutral value of 0.5 so that the uninorm was chosen.

As a result of the first approach, a habitat suitability (HS) map was obtained which is useful for representing the spatial distribution of the ECs that have the highest influence on the occurrence and abundance of a species in a simple and understandable way. HS maps are widely used for representing the overall suitability of study areas. They quantify every preference that is known from the species (U.S.

Fish and Wildlife Service, 1981; Van Horne and Wiens, 1991). Eventually the obtained HS map could be used to guide field studies for locating possible sources of poisoning in the study area. Here, field studies are best conducted in areas that are highly suitable and found to attract the owls. These highly suitable areas might be more valuable as a guideline for field studies compared to the locations of the GPS tracking data themselves since they take into account general preferences of the species and do not rely on discrete observations made on a limited number of individuals.

GPS tracking data were used to evaluate the resulting HS map. This type of data is not commonly used for validation. Typically, validation is based on an independent set of data on the sites where the species is present and where it is absent. Unfortunately, this type of data was not available for our species in the region. Therefore, we combined the methods from Hirzel et al. (2006) and Reynolds-Hogland and Mitchell (2007) to perform validation. The obtained Ivlev's electivity and Manly's habitat selection indices are biased towards high HSI values since many GPS fixes in our dataset are located close to each other in a limited number of pixels, which consequently all have a very high probability of occurrence. Furthermore, they are located in highly suitable pixels, close to a possible nesting site. Moreover, since we compared the pixels that were probably visited with those available in the landscape, the values of the indices depended on the definition of the study area. If detailed information on habitat suitability is needed in more detail, the study area might be limited to the area of interest, for example the home range or the neighborhood of the nesting site.

In contrast to the first approach to determine the most important ECs for the Eagle Owls in Limburg, the second approach is data-driven. First, we removed the fixes corresponding to locations where owls were probably not hunting on the basis of an analysis of five relevant movement features. Next, we studied the use of space by means of a hotspot analysis and a convex hull, and the ecological niche factor analysis (ENFA) was performed to pinpoint the most important ECs for the species.

The hotspot analysis and convex hull were chosen because they are a straightforward way to get an overview of the frequently visited regions and use of space. Alternatively, a probability density function (PDF) describing the probability of occurrence, or a home range could have been defined by, for example, the TIBB.

Compared to other frequently used analyses, like logistic regression, Gaussian logistic regression, generalized linear models, discriminant analysis, Mahalanobis distances and artificial neural networks, ENFA does not rely on a form of presence/absence data of the focal species in a set of sampled locations. It requires only presence data (Hirzel et al., 2001, 2002). ENFA reduces, like PCA, the number of variables by combining them in so-called latent, uncorrelated factors (Hirzel et al., 2002). It can be indicated which of these factors explain most of the variance (Jolliffe, 1990; Yong and Pearce, 2013) and the relationships between the ECs

and the factors are expressed by so-called loadings (Hirzel et al., 2002). By analyzing these loadings, the influence of each EC on the species can be determined. ENFA builds upon Hutchinson's concept of an ecological niche (Hirzel et al., 2002; Basille et al., 2008) and, as such, factors have an ecological meaning. We find the latter interesting when wildlife is studied.

Finally, the differential space use of post-fledgling Gray Catbird (*Dumetella carolinensis*) and Wood Thrush (*Hylocichla mustelina*) individuals was investigated. As a first step, a tracking data analysis was performed. Here, the PDF describing the probability of occurrence and the home range were determined for each individual by making use of the TIBB. From the PDF and home range, the weighted relative proportion and area of each land cover class were calculated. To statistically test whether the relative space use differs between species, a generalized linear mixed-effects model (GLMM) was used which included all interactions between species, land cover class, fate and post-fledgling period. These tests indicate which (combination of) explanatory variables are influencing the weighted relative area but they do not give insight into the type of difference among the variables themselves, *i.e.* among the species, land cover class, fate and period. Therefore, the mean relative proportions and area per land cover class were inspected and compared between species, periods and fate. The main conclusion is that the tracked individuals of Gray Catbird used less forests and more roads than the Wood Thrush, which matches the expectations that Wood Thrush avoids urban areas, whereas the Gray Catbird is adapted to life in a landscape dominated by humans. The GLMM could also have been used to assess the effect of, or relationship between, explanatory variables on the weighted relative area. The interpretation is, however, not this straightforward since the link function and random effects need to be taken into account when interpreting the model parameters (UCLA: Statistical Consulting Group, 2018).

Taking into account the results of the data analysis, some preliminary ideas are given concerning the construction of a model to simulate consecutive locations of individual Gray Catbird and Wood Thrush fledglings during the post-independent period on a daily basis, in such a way that the land use of the *in silico* individuals is similar to that of the real-world individuals. Here, it was concluded that our models did not accurately simulate the observed proportions. Hence, we failed to prove their predictive power. Nevertheless, some of the land uses were accurately simulated for some test sets and when using all data as training set. To improve the model performance, we suggest to widen the data analysis so that more insights into the behavior of the fledglings are obtained. It is clear that more environmental and behavioral information needs to be included in the model. Furthermore, adaptations should be made in the light of the ultimate goal of the model, that is, estimating the effect of further urbanization on the post fledglings. Here, not only habitat loss but also fragmentation and the spatial distribution of the land uses might be important and need to be included. Survival rates may be used to quan-

tify the effect of urbanization. As can be seen, more work is needed before land use management can be based on the obtained simulations. Nevertheless, given the limited amount of information that is now taken into account in this stage, and given that there is already some indication that the model approximates the species behavior, it might serve as a good starting point.

Since the definition of the SIs is highly adaptable, the presented modeling methodology – including model construction, model selection and (cross-) validation – can be adopted for modeling the movement of other species. Therefore, the SIs should be adapted to account for preferences, a time step should be chosen and information on typical step lengths should be included. Furthermore, other aggregation functions and decision rules can be used. Our methodology offers a simple and easy way to make a first attempt in modeling the movement behavior of a species under consideration. It might serve as a good starting point that can be extended and fine tuned as desired.

To finish this chapter, we would like to share some general findings concerning data exploration, tracking data analysis and modeling of movement phenomena, which could serve as an inspiration to future research.

Prior to any data analysis, we found it useful to first look at the metadata of the tracking data to determine the type of data, time resolution, number of observations, coordinate system, measurement error, study period and so on. Also a concise literature study of the species was indispensable. Furthermore, information on the landscape where the species is moving through has to be considered. Here, remote-sensing-based data layers on topography and land cover provide such information at spatial scales from several kilometers to less than a meter, and with a possibility of a daily frequency (Kays et al., 2015). A visual inspection of the fixes depicted on, for example, a land use map, gives an idea about the spatial extent of the data and the type of landscape that is under consideration. The characteristics of landscapes do not only change spatially, they also change over time. Just think of the seasonal cycles observed within one calendar year and the natural day cycle (van Moorter et al., 2013). Furthermore, human-caused environmental changes influence movement. In this dissertation, we assumed static environmental conditions since the time frames of our studies were limited and we wanted to keep the analysis and modeling complexity relatively low. For other studies, it might be desirable to take a dynamic environment into account. Here, remote-sensing-based data layers with a high temporal resolution (MODIS for example (Kays et al., 2015)) can be used and real-time environmental information can be obtained from sensors logging environmental parameters (temperature, wind or water depth) that are combined with tracking devices (Demšar et al., 2015). The integrating of such detailed animal and environmental information is, however, very challenging (Kays et al., 2015). Finally, we noticed that it is important to clearly understand the research question at stake. Therefore, an in-depth discussion with the owners of the data and the stakeholders should be conducted at the beginning of each study.

This might also give preliminary insights into the species behavior and highlight some peculiarities concerning the data retrieval and study area.

After this data exploration, computational movement analysis is required to quantify the movement phenomenon under consideration. Here, several movement metrics, such as the covered distance per time interval (average speed in between two fixes), the direction of movement compared to a fixed location and the turning angle, are relevant. Moreover, it is essential to determine the relationship between the visited locations and the ECs. Ideally, a large dataset, containing many individuals, should be available. Otherwise, caution should be exercised when drawing conclusions concerning the species on the basis of the tracking data.

After extracting knowledge on the species behavior, movement models can be built for analyzing, describing, simulating and predicting movement dynamics. We found it most pragmatic to work with discrete time-continuous space models. In this way, movement is conceptualized by means of a series of turning angles and step lengths (Section 2.6.3). Hence, the mathematics involved is similar across all of such models.

In summary, technological advances in data collection and increased computer power are revolutionizing the field of movement science. To extract knowledge from the huge amount of tracking data that is nowadays available, convenient analysis and modeling methods need to be developed and applied to support a well-founded management decision making. My research is joining this effort. In this dissertation, one theoretical chapter and three chapters of a more applied nature have been presented. For what concerns the latter, several analysis methods and models helped to answer the research questions at stake. In this way, this research offers an added value to the involved parties, serves as a source of inspiration for future research with similar research questions and, foremost, it contributes to an improved management and conservation of the environment we live in.

Bibliography

- 100 km Dodentocht Kadee Bornem, 2015. 100 km Dodentocht Kadee Bornem.
URL www.dodentocht.be
- Abo Wind, 2017. Bestandsentwicklung des Uhus (*Bubo bubo*) in Deutschland.
URL https://www.abo-wind.com/de/pdf/Presse2015/2015-09_Uhu_Bestandsentwicklung_Deutschland.pdf
- Abramowitz, M., Stegun, I. A., 1972. Error Function and Fresnel Integrals. Ch. 7 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications, New York, USA.
- Actionchallenge, 2015. Walk it, jog it, or run it!
URL <http://www.actionchallenge.com/activities/treks-walks/uk-treks/100-km-challenge>
- Adalsteinsson, S. A., Buler, J. J., Bowman, J. L., D'Amico, V., Ladin, Z. S., Shriver, W. G., In prep. For fledgling songbirds in an urban landscape, first predators then anthropogenic hazards drive mortality risk.
- Alerstam, T., Hedenström, A., Åkesson, S., 2003. Long-distance migration: evolution and determinants. *Oikos* 103 (2), 247–260.
- Allen, A. M., Singh, N. J., 2016. Linking movement ecology with wildlife management and conservation. *Frontiers in Ecology and Evolution* 3, 155.
- Anders, A. D., Faaborg, J., Thompson III, F. R., 1998. Postfledging dispersal, habitat use, and home-range size of juvenile Wood Thrushes. *The Auk* 115 (2), 349–358.

- Aronson, M. F., La Sorte, F. A., Nilon, C. H., Katti, M., Goddard, M. A., Lepczyk, C. A., Warren, P. S., Williams, N. S., Cilliers, S., Clarkson, B., et al., 2014. A global analysis of the impacts of urbanization on bird and plant diversity reveals key anthropogenic drivers. *Proceedings of the Royal Society B* 281 (1780), 20133330.
- Ausprey, I. J., Rodewald, A. D., 2011. Postfledging survivorship and habitat selection across a rural-to-urban landscape gradient. *The Auk* 128 (2), 293–302.
- Baca, A., Dabnichki, P., Heller, M., Kornfeind, P., 2009. Ubiquitous computing in sports: A review and analysis. *Journal of Sports Sciences* 27 (12), 1335–1346.
- Baetens, J., Van Nieuland, S., Pauwels, I., De Baets, B., Mouton, A., Goethals, P., 2013. An individual-based model for the migration of pike (*Esox lucius*) in the river Yser, Belgium. *Ecological modelling* 258, 40–52.
- Baguette, M., Blanchet, S., Legrand, D., Stevens, V. M., Turlure, C., 2013. Individual dispersal, landscape connectivity and ecological networks. *Biological Reviews* 88 (2), 310–326.
- Balogh, A. L., Ryder, T. B., Marra, P. P., 2011. Population demography of Gray Catbirds in the suburban matrix: sources, sinks and domestic cats. *Journal of Ornithology* 152 (3), 717–726.
- Banks, J., 1998. *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. John Wiley & Sons, New York, USA.
- Basille, M., Calenge, C., Marboutin, E., Andersen, R., Gaillard, J.-M., 2008. Assessing habitat selection using multivariate statistics: Some refinements of the ecological-niche factor analysis. *Ecological Modelling* 211 (1), 233–240.
- Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- Batschelet, E., 1972. Recent statistical methods for orientation data. In: *Proceedings of Conference on Animal Orientation and Navigation capabilities*, Virginia, USA, September 9-13, 1970 . pp. 61–91.
- Belgian LifeWatch Infrastructure, 2018. Fish acoustic receiver network. <http://www.lifewatch.be/nl/node/123>.
- Benhamou, S., 2011. Dynamic approach to space and habitat use based on biased random bridges. *PLoS ONE* 6, e14592.
- Benítez-López, A., Alkemade, R., Verweij, P. A., 2010. The impacts of roads and other infrastructure on mammal and bird populations: a meta-analysis. *Biological Conservation* 143 (6), 1307–1316.

- Berger-Wolf, T. Y., Rubenstein, D. I., Stewart, C. V., Holmberg, J. A., Parham, J., Menon, S., Crall, J., Van Oast, J., Kiciman, E., Joppa, L., 2017. Wildbook: Crowdsourcing, computer vision, and data science for conservation. arXiv preprint arXiv:1710.08880.
- Berthold, P., 2001. Bird migration: a general survey. Oxford University Press on Demand, Oxford, United Kingdom.
- Bewick, V., Cheek, L., Bal, J., 2004. Statistics review 12: Survival analysis. Critical Care 8 (5), 389–394.
- Birdlife international, 2018a. (*Bubo bubo*). (amended version published in 2016) The IUCN Red List of Threatened Species 2017.
URL <http://datazone.birdlife.org/species/factsheet/22688927>
- Birdlife international, 2018b. IUCN Red List for Birds. <http://www.birdlife.org>.
- Blair, R., 2004. The effects of urban sprawl on birds at multiple levels of biological organization. Ecology and Society 9 (5), 2.
- Boelman, N. T., Gough, L., Wingfield, J., Goetz, S., Asmus, A., Chmura, H. E., Krause, J. S., Perez, J. H., Sweet, S. K., Guay, K. C., 2015. Greater shrub dominance alters breeding habitat and food resources for migratory songbirds in Alaskan arctic tundra. Global Change Biology 21 (4), 1508–1520.
- Boggie, M. A., Mannan, R. W., 2014. Examining seasonal patterns of space use to gauge how an accipiter responds to urbanization. Landscape and Urban Planning 124, 34–42.
- Borodin, A., 1989. Brownian local time. Russian Mathematical Surveys 44 (2), 1–51.
- Boyce, M. S., Vernier, P. R., Nielsen, S. E., Schmiegelow, F. K., 2002. Evaluating resource selection functions. Ecological Modelling 157 (2), 281–300.
- Brady, S. P., Richardson, J. L., 2017. Road ecology: shifting gears toward evolutionary perspectives. Frontiers in Ecology and the Environment 15 (2), 91–98.
- Breed, G. A., Costa, D. P., Jonsen, I. D., Robinson, P. W., Mills-Flemming, J., 2012. State-space methods for more completely capturing behavioral dynamics from animal tracks. Ecological Modelling 235, 49–58.
- Breslow, N. E., 1975. Analysis of survival data under the proportional hazards model. International Statistical Review 43 (1), 45–57.
- Bridge, E. S., Thorup, K., Bowlin, M. S., Chilson, P. B., Diehl, R. H., Fléron, R. W., Hartl, P., Kays, R., Kelly, J. F., Robinson, W. D., Wikelski, M., 2011. Technology on the move: recent and forthcoming innovations for tracking migratory birds. BioScience 61 (9), 689–698.

- Brillinger, D. R., Preisler, H. K., Ager, A. A., Kie, J. G., 2004. An exploratory data analysis (eda) of the paths of moving animals. *Journal of Statistical Planning and Inference* 122 (1-2), 43–63.
- Brum-Bastos, V. S., Long, J. A., Demšar, U., 2018. Weather effects on human mobility: a study using multi-channel sequence analysis. *Computers, Environment and Urban Systems*.
- Buchin, K., Sijben, S., Arseneau, T., Willems, E., 2012. Detecting movement patterns using Brownian bridges. In: *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*. Redondo Beach, CA, USA, pp. 119–128.
- Bullard, F., 1991. Estimating the home range of an Animal: a Brownian Bridge Approach. Master thesis, University of North Carolina, Chapel Hill, USA.
- Burghardt, T., Čalić, J., 2006. Analysing animal behaviour in wildlife videos using face detection and tracking. *IEE Proceedings-Vision, Image and Signal Processing* 153 (3), 305–312.
- Burt, W. H., 1943. Territoriality and home range concepts as applied to mammals. *Journal of Mammalogy* 24 (3), 346–352.
- Byrne, M., Clint, M., Hinton, J., Chamberlain, M., Collier, B., 2014. Using dynamic Brownian bridge movement modelling to measure temporal patterns of habitat selection. *Journal of Animal Ecology* 83 (5), 1234–1243.
- Cagnacci, F., Boitani, L., Powell, R. A., Boyce, M. S., 2010. Animal ecology meets gps-based radiotelemetry: a perfect storm of opportunities and challenges. *The Royal Society* 365 (1550), 2157–2162.
- Cain, M. L., 1989. The analysis of angular data in ecological field studies. *Ecology* 70 (5), 1540–1543.
- Calabrese, F., Ferrari, L., Blondel, V. D., 2015. Urban sensing using mobile phone network data: a survey of research. *Acm computing surveys (csur)* 47 (2), 25.
- Calenge, C., 2006. The package adehabitat for the R software: a tool for the analysis of space and habitat use by animals. *Ecological Modelling* 197 (3), 516–519.
- Candolin, U., Wong, B. B., 2012. *Behavioural Responses to a Changing World: Mechanisms and Consequences*. Oxford University Press, Oxford, United Kingdom.
- CBS, Den Haag; Planbureau voor de Leefomgeving, Den Haag/Bilthoven en Wageningen UR, Wageningen, 2015. *Flora van wegbermen 1999-2014 (indicator 1433, versie 04, 25 november 2015)*.

- Chapman, J. W., Klaassen, R. H., Drake, V. A., Fossette, S., Hays, G. C., Metcalfe, J. D., Reynolds, A. M., Reynolds, D. R., Alerstam, T., 2011. Animal orientation strategies for movement in flows. *Current Biology* 21 (20), R861–R870.
- Chatfield, C., 2004. *The Analysis of Time Series: An Introduction, Sixth Edition*. Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hall/CRC, New York, USA.
- Chesapeake Conservancy, 2017. High - Resolution Land Cover Dataset 101. Annapolis, Maryland.
URL <http://chesapeakeconservancy.org/conservation-innovation-center/high-resolution-data/land-cover-data-project/>
- Chi, Z., Pozdnyakov, V., Yan, J., 2015. On expected occupation time of brownian bridge. *Statistics and Probability Letters* 97, 83–87.
- Cimprich, D., Moore, F., 1995. *The Birds of North America* No. 167. Washington, D.C.: The Academy of Natural Sciences and Philadelphia, PA: The American Ornithologists' Union., Ch. Cimprich, D., F. Moore. 1995. Gray Catbird (*Dumetella carolinensis*). A Poole, F Gill, eds. Washington, D.C.: The Academy of Natural Sciences and Philadelphia, PA: The American Ornithologists' Union.
- Cioc, M., 2009. *The Game of Conservation-International Treaties to Protect the World's Migratory Animals*. Ohio University Press, Ohio (USA).
- Clayton, D. G., 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65 (1), 141–151.
- Clobert, J., Galliard, L., Cote, J., Meylan, S., Massot, M., 2009. Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. *Ecology letters* 12 (3), 197–209.
- Cochran, W. W., Lord Jr, R. D., 1963. A radio-tracking system for wild animals. *The Journal of Wildlife Management* 27 (1), 9–24.
- Codling, E. A., Plank, M. J., Benhamou, S., 2008. Random walk models in biology. *Journal of the Royal Society Interface* 5 (25), 813–834.
- Cohen, E. B., Barrow Jr, W. C., Buler, J. J., Deppe, J. L., Farnsworth, A., Marra, P. P., McWilliams, S. R., Mehlman, D. W., Wilson, R. R., Woodrey, M. S., Moore, F. R., 2017. How do en route events around the Gulf of Mexico influence migratory landbird populations? *The Condor* 119 (2), 327–343.
- Colorado, Z. G., Rodewald, A. D., 2017. Patterns of change in body condition in wintering Neotropical-Nearctic migratory birds in shaded plantations in the Andes. *Agroforestry Systems* 91 (6), 1129–1137.

- Cooper, N. W., Hallworth, M. T., Marra, P. P., 2017. Light-level geolocation reveals wintering distribution, migration routes, and primary stopover locations of an endangered long-distance migratory songbird. *Journal of Avian Biology* 48 (2), 209–219.
- Cooper, N. W., Sherry, T. W., Marra, P. P., 2014. Modeling three-dimensional space use and overlap in birds. *The Auk* 131 (4), 681–693.
- Cox, D., Oakes, D., 1984. *Analysis of Survival Data*. Chapman & Hall, New York, USA.
- Cox, W. A., Thompson, F. R., Cox, A. S., Faaborg, J., 2014. Post-fledging survival in passerine birds and the value of post-fledging studies to conservation. *The Journal of Wildlife Management* 78 (2), 183–193.
- Daamen, W., Hoogendoorn, S. P., 2006. Free speed distributions for pedestrian traffic. In: 85th Annual Meeting Transportation Research Board. Washington, USA, pp. 1–13.
- Dalbeck, L., Heg, D., 2006. Reproductive success of a reintroduced population of Eagle-owls *Bubo bubo* in relation to habitat characteristics in the Eifel, Germany. *Ardea* 94 (1), 3–21.
- Daly, A. J., Baetens, J. M., De Baets, B., 2016. The impact of resource dependence of the mechanisms of life on the spatial population dynamics of an in silico microbial community. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 26 (12), 123121.
- De Baets, B., Fodor, J., 1999. Van Melle's combining function in MYCIN is a representable uninorm: an alternative proof. *Fuzzy Sets and Systems* 104 (1), 133–136.
- de Ulzurrun, G. V.-D., Baetens, J., Van den Bulcke, J., De Baets, B., 2017. Modelling three-dimensional fungal growth in response to environmental stimuli. *Journal of Theoretical Biology* 414, 35–49.
- Delgado, M. D. M., Penteriani, V., 2007. Vocal behaviour and neighbour spatial arrangement during vocal displays in Eagle-owls (*Bubo bubo*). *Journal of Zoology* 271 (1), 3–10.
- Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., Weiskopf, D., Weibel, R., 2015. Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology* 3 (1), 5.
- Desbiez, A. L. J., Bodmer, R. E., Santos, S. A., 2009. Wildlife habitat selection and sustainable resources management in a neotropical wetland. *International Journal of Biodiversity and Conservation* 1 (1), 011–020.

- Dewhurst, O. P., Evans, H. K., Roskilly, K., Harvey, R. J., Hubel, T. Y., Wilson, A. M., 2016. Improving the accuracy of estimates of animal path and travel distance using gps drift-corrected dead reckoning. *Ecology and evolution* 6 (17), 6210–6222.
- Dingle, H. D., Drake, V. A., 2007. What is migration? *BioScience* 57 (2), 113–121.
- Dodge, S., Weibel, R., Ahearn, S. C., Buchin, M., Miller, J. A., 2016. Analysis of movement data. *International Journal of Geographical Information Science* 30 (5), 825–834.
- Donázar, J. A., 1989. Variaciones geográficas y estacionales en la alimentación del Búho Real (*Bubo bubo*) en Navarra. *Ardeola* 36 (1), 25–39.
- Drake, V., 1985. Radar observations of moths migrating in a nocturnal low-level jet. *Ecological Entomology* 10 (3), 259–265.
- Dunne, P., 2006. *Pete dunne's essential field guide companion*. Houghton Mifflin Harcourt, New York, USA.
- Elith, J., Leathwick, J. R., 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics* 40, 677–697.
- Eng, M. L., Stutchbury, B. J., Burke, D. M., Elliott, K. A., 2011. Influence of forest management on pre-and post-fledging productivity of a neotropical migratory songbird in a highly fragmented landscape. *Canadian Journal of Forest Research* 41 (10), 2009–2019.
- Environmental Systems Research Institute (ESRI), 2016. *ArcGIS Desktop Help 10.2 Geostatistical Analyst*.
URL <http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/optimized-hot-spot-analysis.htm>
- Erni, B., Liechti, F., Bruderer, B., 2005. The role of wind in passerine autumn migration between europe and africa. *Behavioral Ecology* 16 (4), 732–740.
- Erritzoe, J., Mazgajski, T. D., Rejt, Ł., 2003. Bird casualties on European roads—a review. *Acta Ornithologica* 38 (2), 77–93.
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *2nd Int. Conf. Knowledge Discovery and Data Mining (KDD96)*. Association for the Advancement of Artificial Intelligence Press, pp. 226–231.
- Evans, M., Gow, E., Roth, R. R., Johnson, M. S., Underwood, T. J., 2011. Wood Thrush (*Hylocichla mustelina*). In *The Birds of North America* (A. F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA.

- Fieberg, J., Matthiopoulos, J., Hebblewhite, M., Boyce, M. S., Frair, J. L., 2010. Correlation and studies of habitat selection: problem, red herring or opportunity? *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1550), 2233–2244.
- Frank, J., Massey, J., 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association* 46, 68–78.
- Frank, M. J., 1979. On the simultaneous associativity of $f(x, y)$ and $x + y - f(x, y)$. *Aequationes Mathematicae* 19 (1), 194–226.
- Gibbs, J. P., Shriver, W. G., 2002. Estimating the effects of road mortality on turtle populations. *Conservation Biology* 16 (6), 1647–1652.
- Gilruth, A., 2016. Eagle owls - are they making a comeback in Britain? Game and Wildlife Conservation Trust.
URL <https://www.gwct.org.uk/blogs/news/2016/april/eagle-owls-%E2%80%93-are-they-making-a-comeback-in-britain/>
- Gómez, C., Bayly, N. J., Norris, D. R., Mackenzie, S. A., Rosenberg, K. V., Taylor, P. D., Hobson, K. A., Cadena, C. D., 2017. Fuel loads acquired at a stopover site influence the pace of intercontinental migration in a boreal songbird. *Scientific reports* 7 (1), 3405.
- Goodenough, A. E., Coker, D. G., Wood, M. J., Rogers, S. L., 2017. Overwintering habitat links to summer reproductive success: intercontinental carry-over effects in a declining migratory bird revealed using stable isotope analysis. *Bird Study* 64 (4), 433–444.
- Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E., 2009. Aggregation Functions. *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, UK.
- Gradshteyn, I., Ryzhik, I., 2007. *Table of Integrals, Series, and Products*, 7th Edition. Academic Press, New York, USA.
- Grillmayer, L., 2013. Radio-frequency identification-overview. *Innovative Internet Technologies and Mobile* 25, 25–33.
- Grüebler, M. U., Naef-Daenzer, B., 2010. Survival benefits of post-fledging care: experimental approach to a critical part of avian reproductive strategies. *Journal of Animal Ecology* 79 (2), 334–341.
- Grzybowski, J. A., 1983. Patterns of space use in grassland bird communities during winter. *The Wilson Bulletin* 95 (4), 591–602.
- Gudmundsson, J., Laube, P., Wolle, T., 2008. Movement patterns in spatio-temporal data. In: *Encyclopedia of GIS*. Springer, New York, USA, pp. 726–732.

- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecology letters* 8 (9), 993–1009.
- Guisan, A., Zimmermann, N. E., 2000. Predictive habitat distribution models in ecology. *Ecological modelling* 135 (2-3), 147–186.
- Gurarie, E., Andrews, R. D., Laidre, K. L., 2009. A novel method for identifying behavioural changes in animal movement data. *Ecology Letters* 12 (5), 395–408.
- Gurarie, E., Bracis, C., Delgado, M., Meckley, T. D., Kojola, I., Wagner, C. M., 2016. What is the animal doing? Tools for exploring behavioural structure in animal movements. *Journal of Animal Ecology* 85 (1), 69–84.
- Halekoh, U., Højsgaard, S., 2014. A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models—the r package pbkrtest. *Journal of Statistical Software* 59 (9), 1–30.
- Hand, B. K., Cushman, S. A., Landguth, E. L., Lucotch, J., 2014. Assessing multi-taxa sensitivity to the human footprint, habitat fragmentation and loss by exploring alternative scenarios of dispersal ability and population size: a simulation approach. *Biodiversity and Conservation* 23 (11), 2761–2779.
- Hanks, E. M., Hooten, M. B., Alldredge, M. W., 2015. Continuous-time discrete-space models for animal movement. *The Annals of Applied Statistics* 9 (1), 145–165.
- Heintzenberg, F., 2008. Roofvogels en uilen: alle soorten van Europa, 1st Edition. Tirion Uitgevers, Baarn, The Netherlands.
- Herrlinger, E., 1973. Die Wiedereinbürgerung des Uhus *Bubo bubo* in der Bundesrepublik Deutschland. In: *Bonner Zoologische Monographien* No. 4. Zoologisches Forschungsinstitut und Museum Alexander Koenig, Bonn, Germany.
- Het Kadaster, 2016. Basisregistratie topografie: Catalogus en productspecificaties versie 2.2.1. Tech. rep., Het Kadaster, Apeldoorn, The Netherlands.
- Hewson, C. M., Thorup, K., Pearce-Higgins, J. W., Atkinson, P. W., 2016. Population decline is linked to migration route in the common cuckoo. *Nature Communications* 7, 12296.
- Hirzel, A. H., Hausser, J., Chessel, D., Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83 (7), 2027–2036.
- Hirzel, A. H., Helfer, V., Metral, F., 2001. Assessing habitat-suitability models with a virtual species. *Ecological modelling* 145 (2-3), 111–121.

- Hirzel, A. H., Le Lay, G., 2008. Habitat suitability modelling and niche theory. *Journal of Applied Ecology* 45 (5), 1372–1381.
- Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., Guisan, A., 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling* 199 (2), 142–152.
- Hobson, K. A., Kardynal, K. J., 2015. Western veeries use an eastern shortest-distance pathway: New insights to migration routes and phenology using light-level geolocators. *The Auk* 132 (3), 540–550.
- Hoekstra, J. M., Boucher, T. M., Ricketts, T. H., Roberts, C., 2005. Confronting a biome crisis: global disparities of habitat loss and protection. *Ecology letters* 8 (1), 23–29.
- Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., Megown, K., 2015. Completion of the 2011 national land cover database for the conterminous united states—representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing* 81 (5), 345–354.
- Hooghiemstra, G., 2002. On explicit occupation time distributions for Brownian processes. *Statistics & Probability Letters* 56 (4), 405–417.
- Hooten, M. B., Johnson, D. S., McClintock, B. T., Morales, J. M., 2017. *Animal Movement: Statistical Models for Telemetry Data*. CRC Press, Taylor & Francis, Boca Raton, Florida, United States.
- Horne, J., Garton, E., Krone, S., Lewis, J., 2007. Analyzing animal movements using Brownian bridges. *Ecology* 88 (9), 2354–2363.
- Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50 (3), 346–363.
- Hubbard, S., Babak, P., Sigurdsson, S. T., Magnússon, K. G., 2004. A model of the formation of fish schools and migrations of fish. *Ecological Modelling* 174 (4), 359–374.
- Huston, M., DeAngelis, D., Post, W., 1988. New computer models unify ecological theory. *BioScience* 38 (10), 682–691.
- Hutchinson, M., 1957. Concluding remarks. In: *Cold Spring Harbour Symposia on quantitative biology: population studies: animal ecology and demography*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA. pp. 415–427.
- Hutto, R. L., 1985. Habitat selection by nonbreeding, migratory land birds. In: *Cody, M. L. (Ed.), Habitat Selection in Birds*. Academic Press, Orlando, USA, pp. 455–476.

- International Union for Conservation of Nature and Natural Resources, 2018. The IUCN Red List of Threatened Species. <http://www.iucnredlist.org/>.
URL <http://www.iucnredlist.org/>
- Ivlev, V. S., 1960. Experimental ecology of The Feeding of Fishes. Yale University Press, New Haven, Connecticut, USA.
- Janssen, R., 2011. Vondst kraamkolonie Bechsteins vleermuis. Nieuweling op rode lijst? Zoogdier 22, 13–16.
- Jenkins, J. M., Hart, M., Eggert, L. S., Faaborg, J., 2017. Sex-specific movements in postfledging juvenile ovenbirds (*Seiurus aurocapilla*). The Wilson Journal of Ornithology 129 (4), 846–850.
- Jolliffe, I. T., 1990. Principal component analysis: a beginner's guide-I. Introduction and application. Weather 45 (10), 375–382.
- Karunaratne, F., 2012. Landscape and Wildlife Photography.
URL <https://www.flickr.com/photos/madhawak/>
- Kays, R., Crofoot, M. C., Jetz, W., Wikelski, M., 2015. Terrestrial animal tracking as an eye on life and planet. Science 348 (6240), aaa2478.
- Kays, S., Tilak, M., Crofoot, T., Fountain, D., Obando, A., Ortega, F., Kuemmeth, J., Mandel, G., Swenson, T., Lambert, B., Hirsch, Wilkelski, M., 2011. Tracking animal location and activity with an automated radio telemetry system in tropical rainforest. The Computer Journal 52, 255–271.
- Kellenberger, K., Groom, C., 2015. Discovering ranking functions. In: Expert T-SQL Window Functions in SQL Server. Springer Science+Business Media New York, USA, pp. 17–31.
- Kennedy-Mars Sittard, 2015. Overzicht aller tijden.
URL <http://www.kennedymars.org/wp-content/uploads/2015/12/KMS-2015-Overzicht-aller-tijden.pdf>
- Kie, J. G., Matthiopoulos, J., Fieberg, J., Powell, R. A., Cagnacci, F., Mitchell, M. S., Gaillard, J.-M., Moorcroft, P. R., 2010. The home-range concept: are traditional estimators still relevant with modern telemetry technology? Philosophical Transactions of the Royal Society B: Biological Sciences 365 (1550), 2221–2231.
- King, D., Degraaf, R., Smith, M.-L., Buonaccorsi, J., 2006. Habitat selection and habitat-specific survival of fledgling ovenbirds (*Seiurus aurocapilla*). Journal of Zoology 269 (4), 414–421.
- Kolmogorov, A., 1956. Grundbegriffe der wahrscheinlichkeitrechnung. ergebnisse der mathematik. 1933. translated by N. Morrison, Foundations of probability. Chelsea Publishing Company 2, 31–32.

- König, C., Weick, F., Becking, J. H., 2010. *Owls of the World*, 2nd Edition. Helm Identification Guides. Bloomsbury Publishing, London, UK.
- Kraak, M.-J., 2003. The space-time cube revisited from a geovisualization perspective. In: Proc. 21st International Cartographic Conference. pp. 1988–1996.
- Kranstauber, B., Kays, R., LaPoint, S., Wilkelski, M., Safi, K., 2012. A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *Animal Ecology* 81 (4), 738–746.
- Kranstauber, B., Safi, K., Bartumeus, F., 2014. Bivariate Gaussian bridges: directional factorization of diffusion in Brownian bridge models. *Movement Ecology* 2 (1), 5.
- Kranstauber, B., Smolla, M., 2014. Visualizing and analyzing animal track data. R package version 1.2.475.
URL <http://computational-ecology.com/main-move.html><http://cran.r-project.org/web/packages/move/index.html>
- Ladin, Z. S., D'Amico, V., Baetens, J. M., Roth, R. R., Shriver, W. G., 2016. Predicting metapopulation responses to conservation in human-dominated landscapes. *Frontiers in Ecology and Evolution* 4, 122.
- Ladin, Z. S., Van Nieuland, S., Adalsteinsson, S. A., D'Amico, V., Bowman, J. L., Buler, J. J., Baetens, J. M., De Baets, B., Shriver, W. G., 2018. Differential post-fledging habitat use of Nearctic-Neotropical migratory birds within an urbanized landscape. *Movement Ecology* 6 (17).
- Langley, R. B., 1998. The UTM Grid System. *GPS WORLD*, 46–50.
- Latombe, G., Parrott, L., Basille, M., Fortin, D., 2014. Uniting statistical and individual-based approaches for animal movement modelling. *PloS one* 9 (6), e99938.
- Lens, L., Dhondt, A. A., 1994. Effects of habitat fragmentation on the timing of crested tit *parus cristatus* natal dispersal. *Ibis* 136 (2), 147–152.
- Lepczyk, C. A., La Sorte, F. A., Aronson, M. F., Goddard, M. A., MacGregor-Fors, I., Nilon, C. H., Warren, P. S., 2017. Global patterns and drivers of urban bird diversity. In: *Ecology and Conservation of Birds in Urban Environments*. Springer, Heidelberg, Germany, pp. 13–33.
- Liao, J. C., 2007. A review of fish swimming mechanics and behaviour in altered flows. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 362 (1487), 1973–1993.
- Lombardi, L., Fernández, N., Moreno, S., Villafuerte, R., 2003. Habitat-related differences in rabbit (*Oryctolagus cuniculus*) abundance, distribution, and activity. *American Society of Mammalogists* 84 (1), 26–36.

- Lourenço, R., Tavares, P. C., del Mar Delgado, M., Rabaça, J. E., Penteriani, V., 2011. Superpredation increases mercury levels in a generalist top predator, the Eagle Owl. *Ecotoxicology* 20 (4), 635–642.
- Malkinson, T., 2009. Current and emerging technologies in endurance athletic training and race monitoring. In: *Science and Technology for Humanity (TIC-STH)*, 2009 IEEE Toronto International Conference. IEEE, Toronto, ON, USA, pp. 581–586.
- Manly, B., McDonald, L., Thomas, D., McDonald, T., Erickson, W., 2002. *Resource Selection by Animals: Statistical Analysis and Design for Field Studies*. Springer The Netherlands.
- Marchesi, L., Sergio, F., Pedrini, P., 2002. Costs and benefits of breeding in human-altered landscapes for the Eagle Owl (*Bubo bubo*). *Ibis* 144 (4), 164–177.
- Mardia, K. V., 1972. A multi-sample uniform scores test on a circle and its parametric competitor. *Journal of the Royal Statistical Society. Series B (Methodological)* 34 (1), 102–113.
- Markham, A., 2008. *On a Wildlife Tracking and Telemetry System: A Wireless Network Approach*. Ph.D. thesis, Department of Electrical Engineering, University of Cape Town, South Africa.
- Marković, N., Sekuła, P., Vander Laan, Z., Andrienko, G., Andrienko, N., 2018. Applications of trajectory data from the perspective of a road transportation agency: Literature review and maryland case study. *IEEE Transactions on Intelligent Transportation Systems* (99), 1–12.
- Martin, T. E., Geupel, G. R., 1993. Nest-monitoring plots: Methods for locating nests and monitoring success (métodos para localizar nidos y monitorear el éxito de estos). *Journal of Field Ornithology* 64 (4), 507–519.
- Martínez, J. A., Serrano, D., Zuberogoitia, I., 2003. Predictive models of habitat preferences for the Eurasian eagle owl (*Bubo bubo*): a multiscale approach. *Ecography* 26 (1), 21–28.
- Martínez, J. E., Calvo, J. F., 2000. Selección de habitat de nidificación por el Búho Real (*Bubo bubo*) en ambientes mediterráneos semiáridos. *Ardeola* 47 (2), 215–220.
- März, R., 1958. *Der Uhu: Bubo bubo*, 2nd Edition. Ziemsen Verlag, Wittenberg Lutherstadt, Deutsche Demokratische Republik.
- Marzluff, J. M., 2001. Worldwide urbanization and its effects on birds. In: *Avian ecology and conservation in an urbanizing world*. Springer, Boston, MA, USA, pp. 19–47.

- Mascanzoni, D., Wallin, H., 1986. The harmonic radar: a new method of tracing insects in the field. *Ecological Entomology* 11 (4), 387–390.
- Masden, E. A., Haydon, D. T., Fox, A. D., Furness, R. W., Bullman, R., Desholm, M., 2009. Barriers to movement: impacts of wind farms on migrating birds. *ICES Journal of Marine Science* 66 (4), 746–753.
- McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M., Morales, J. M., 2014. When to be discrete: the importance of time formulation in understanding animal movement. *Movement Ecology* 2 (1), 21.
- McClintock, B. T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B. J., Morales, J. M., 2012. A general discrete-time modeling framework for animal movement using multistate random walks. *Ecological Monographs* 82 (3), 335–349.
- McLane, A. J., Semeniuk, C., McDermid, G. J., Marceau, D. J., 2011. The role of agent-based models in wildlife ecology and management. *Ecological Modelling* 222 (8), 1544–1556.
- Mikkola, H., 2014. *Owls of the World - A Photographic Guide*, 2nd Edition. Bloomsbury Publishing, London, UK.
- Miller, H., 2010. The data avalanche is here. Shouldn't we be digging? *Journal of Regional Science* 50, 181–201.
- Millspaugh, J., Marzluff, J. M., 2001. *Radio Tracking and Animal Populations*. Academic Press, San Diego, CA, USA.
- Miosga, O., Gerdes, S., Krämer, D., Vohwinkel, R., 2015. Besonderes Uhu-Höhenflugmonitoring im Tiefland. *Natur in NRW* 3 (15), 35–39.
- Mitchell, M. S., Zimmerman, J. W., Powell, R. A., 2002. Test of a habitat suitability index for black bears in the southern appalachians. *Wildlife Society Bulletin* 30 (3), 794–808.
- Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., Fryxell, J. M., 2004. Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology* 85 (9), 2436–2445.
- Morrill, R., 2006. Classic map revisited: The growth of megalopolis. *The Professional Geographer* 58 (2), 155–160.
- Morris, P. A., 2006. *The New Hedgehog Book*. Whittet Books, London, UK.
- Mortenson, J. A., 1971. Location and capture of prey by the Great Horned Owl (*Bubo virginianus*). PhD Thesis, The University of Montana, USA.
- Naef-Daenzer, B., Früh, D., Stalder, M., Wetli, P., Weise, E., 2005. Miniaturization (0.2 g) and evaluation of attachment techniques of telemetry transmitters. *Journal of Experimental Biology* 208 (21), 4063–4068.

- Naef-Daenzer, B., Gruebler, M. U., 2016. Post-fledging survival of altricial birds: Ecological determinants and adaptation. *Journal of Field Ornithology* 87 (3), 227–250.
- Nathan, R., Getz, W., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., Smouse, P., 2008. A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences* 105, 19052–19059.
- Nationaal Geografisch Instituut, 2010. Productspecificatie Top10Vector v1.1. Tech. rep., Nationaal Geografisch Instituut, Brussels (Belgium).
- Natuurpunt, 2016. Oehoe maakt comeback in Vlaanderen.
URL <https://www.natuurpunt.be/nieuws/oehoe-maakt-comeback-vlaanderen-20160606#.WcEU-9FpGM8>
- Nelsen, R. B., 1998. *An Introduction to Copulas*. Springer, New York, USA.
- Nielson, M., R., Sawyer, H., McDonald, T. L., 2012. BBMM: Brownian bridge movement model. R package version 2.3.
URL <http://CRAN.R-project.org/package=BBMM>
- Noog, E., Geller, M., 1969. A table of integrals of the error functions. *Journal of Research of the National Bureau of Standards-B. Mathematical Sciences* 73 (1), 1–20.
- Oldham, R. S., Keeble, J., Swan, M. J. S., Jeffcote, M., 2000. Evaluating the suitability of habitat for the great crested newt (*Triturus cristatus*). *Herpetological Journal* 10 (4), 143–155.
- Ortigosa, G. R., De Leo, G. A., Gatto, M., 2000. VVF: integrating modelling and GIS in a software tool for habitat suitability assessment. *Environmental Modelling & Software* 15 (1), 1–12.
- Ottaviani, D., Lasinio, G. J., Boitani, L., 2004. Two statistical methods to validate habitat suitability models using presence-only data. *Ecological Modelling* 179 (4), 417–443.
- Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., King, R., 2017. Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *ASTA Advances in Statistical Analysis* 101 (4), 399–438.
- Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., Matthiopoulos, J., 2008. State-space models of individual animal movement. *Trends in Ecology & Evolution* 23 (2), 87–94.
- Paxton, K. L., Moore, F. R., 2017. Connecting the dots: Stopover strategies of an intercontinental migratory songbird in the context of the annual cycle. *Ecology and Evolution* 7 (17), 6716–6728.

- Pearson, E. S., Hartley, H. O., 1972. *Biometrika tables for statisticians*. Cambridge University Press, Cambridge, UK.
- Penteriani, V., 2002. Variation in the function of Eagle Owl vocal behaviour: territorial defence and intra-pair communication? *Ethology Ecology & Evolution* 14 (3), 275–281.
- Penteriani, V., Gallardo, M., Roche, P., 2002. Landscape structure and food supply affect Eagle Owl (*Bubo bubo*) density and breeding performance: a case of intra-population heterogeneity. *Journal of Zoology* 257 (3), 365–372.
- Pitman, J., 1998. The distribution of local times of a Brownian bridge. Tech. Rep. 539, Department of Statistics, University of California, USA.
- Polansky, L., Wittemyer, G., Cross, P. C., Tambling, C. J., Getz, W. M., 2010. From moonlight to movement and synchronized randomness: Fourier and wavelet analyses of animal location time series data. *Ecology* 91 (5), 1506–1518.
- Pop, M., Losif, R., Miu, I., Rozyłowicz, L., Popescu, V., 2018. Combining resource selection functions and home-range data to identify habitat conservation priorities for Brown Bears. *Animal Conservation*, 1–11.
- Pozdnyakov, v., Meyer, T., Wang, Y., Yan, J., 2014. On modeling animal movements using brownian motion with measurement error. *Ecology* 95, 247–253.
- Preisler, H. K., Ager, A. A., Johnson, B. K., Kie, J. G., 2004. Modeling animal movements using stochastic differential equations. *Environmetrics* 15 (7), 643–657.
- Press, W. H., 2007. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge University Press, Cambridge, UK.
- Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., Lemmon, A. R., 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526 (7574), 569.
- Pàges, J. F., Bartumeus, F., Hereu, B., López-Sanz, À., Romero, J., Alcoverro, T., 2013. Evaluating a key herbivorous fish as a mobile link: a Brownian bridge approach. *Marine Ecology Progress Series* 492, 199–210.
- Rappole, J. H., Tipton, A. R., 1991. New harness design for attachment of radio transmitters to small passerines (nuevo diseño de arnés para atar transmisores a passeriformes pequeños). *Journal of Field Ornithology* 62 (3), 335–337.
- Ray, N., Burgman, M. A., 2006. Subjective uncertainties in habitat suitability maps. *Ecological Modelling* 195 (3-4), 172–186.
- Reijnen, R., Foppen, R., 2006. *The ecology of transportation: managing mobility for the environment*. Springer, Dordrecht, The Netherlands, Ch. Impact of road traffic on breeding bird populations, pp. 255–274.

- Reynolds-Hogland, M. J., Mitchell, M. S., 2007. Effects of roads on habitat quality for bears in the southern Appalachians: a long-term study. *Journal of Mammalogy* 88 (4), 1050–1061.
- Rivera, J. V., Rappole, J., McShea, W., Haas, C., 1998. Wood Thrush postfledging movements and habitat use in northern Virginia. *Condor* 100 (1), 69–78.
- Rockwell, S. M., Wunderle, J. M., Sillett, T. S., Bocetti, C. I., Ewert, D. N., Currie, D., White, J. D., Marra, P. P., 2017. Seasonal survival estimation for a long-distance migratory bird and the influence of winter precipitation. *Oecologia* 183 (3), 715–726.
- Rodewald, A. D., Kearns, L. J., Shustack, D. P., 2013. Consequences of urbanizing landscapes to reproductive performance of birds in remnant forests. *Biological Conservation* 160, 32–39.
- Rodewald, A. D., Shustack, D. P., 2008. Urban flight: understanding individual and population-level responses of nearctic–neotropical migratory birds to urbanization. *Journal of Animal Ecology* 77 (1), 83–91.
- Roloff, G. J., Kernohan, B. J., 1999. Evaluating reliability of habitat suitability index models. *Wildlife Society Bulletin* 27 (4), 973–985.
- Ross, S., 1996. *Stochastic Processes*, Second Edition. John Wiley and Sons, New York, USA.
- Roth, R. R., Johnson, M. S., Underwood, T. J., 1996. *Wood Thrush: *Hylocichla mustelina**. American Ornithologists' Union, Washington, DC, USA.
- Rottenborn, S. C., 1999. Predicting the impacts of urbanization on riparian bird communities. *Biological Conservation* 88 (3), 289–299.
- Rubner, Y., Tomasi, C., Guibas, L. J., 2000. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision* 40 (2), 99–121.
- Ruiz-Gutierrez, V., Kendall, W. L., Saracco, J. F., White, G. C., 2016. Overwintering strategies of migratory birds: a novel approach for estimating seasonal movement patterns of residents and transients. *Journal of Applied Ecology* 53 (4), 1035–1045.
- Runge, C. A., Watson, J. E., Butchart, S. H., Hanson, J. O., Possingham, H. P., Fuller, R. A., 2015. Protected areas and global conservation of migratory birds. *Science* 350 (6265), 1255–1258.
- Running USA with Athlinks (2009-13) and Active.com (2005-08), March 2014. Running USA 's Annual Marathon Report. 2015.
URL <http://www.runningusa.org/index.cfm?fuseaction=news.details&ArticleId=332>

- Rusak, H., 2003. Forest Fragmentation. Federation of Ontario Naturalists (FON), Ontario, Canada.
- Sadjadi, F., 2004. Comparison of fitness scaling functions in genetic algorithms with applications to optical processing. In: Optical information systems II. Vol. 5557. International Society for Optics and Photonics, pp. 356–365.
- Salvadori, G., De Michele, C., Kottegod, N. T., Rosso, R., 2007. Extremes in Nature: an Approach using Copulas. Vol. 56. Springer Science & Business Media, Dordrecht, The Netherlands.
- Sargent, R. G., 2011. Verification and validation of simulation models. In: Jain, S., Creasey, R., Himmelspach, J., White, K. P., Fu, M. C. (Eds.), WSC'11 Winter Simulation Conference 2011. Phoenix, USA, pp. 183–198.
- Sauer, J., Niven, D. K., Hines, J. E., Ziolkowski, D. J., Pardieck, K. L., Fallon, J. E., Link, W. A., 2017. The North American Breeding Bird Survey, Results and Analysis 1966 – 2015. Version 2.07.2017. Tech. rep., USGS Patuxent Wildlife Research Center, Laurel, MD.
- Sawyer, H., Kauffman, M., Nielson, R., Horne, J., 2009. Identifying and prioritizing ungulate migration routes for landscape-level conservation. *Ecological Applications* 19 (8), 2016–2025.
- Seress, G., Liker, A., 2015. Habitat urbanization and its effects on birds. *Acta Zoologica Academiae Scientiarum Hungaricae* 61 (4), 373–408.
- Seto, K. C., Güneralp, B., Hutyra, L. R., 2012. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences* 109 (40), 16083–16088.
- Shamoun-Baranes, J., Alves, J. A., Bauer, S., Dokter, A. M., Hüppop, O., Koistinen, J., Leijnse, H., Liechti, F., van Gasteren, H., Chapman, J. W., 2014. Continental-scale radar monitoring of the aerial movements of animals. *Movement Ecology* 2 (1), 9.
- Shamoun-Baranes, J., Farnsworth, A., Aelterman, B., Alves, J. A., Azijn, K., Bernstein, G., Branco, S., Desmet, P., Dokter, A. M., Horton, K., et al., 2016. Innovative visualizations shed light on avian nocturnal migration. *PloS one* 11 (8), e0160106.
- Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., Fotheringham, A. S., 2016. Analysis of human mobility patterns from gps trajectories and contextual information. *International Journal of Geographical Information Science* 30 (5), 881–906.
- Sklar, M., 1959. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 8, 229–231.

- Smouse, P. E., Focardi, S., Moorcroft, P. R., Kie, J. G., Forester, J. D., Morales, J. M., 2010. Stochastic modelling of animal movement. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1550), 2201–2211.
- Solla, D., Shane, R., Bonduriansky, R., Brooks, R. J., 1999. Eliminating autocorrelation reduces biological relevance of home range estimates. *Journal of Animal Ecology* 68 (2), 221–234.
- Sorte, F. A. L., Fink, D., Hochachka, W. M., Aycrigg, J. L., Rosenberg, K. V., Rodewald, A. D., Bruns, N. E., Farnsworth, A., Sullivan, B. L., Wood, C., et al., 2015. Documenting stewardship responsibilities across the annual cycle for birds on us public lands. *Ecological Applications* 25 (1), 39–51.
- SOVON, 2012. Vogelatlas, 5th Edition. Nationaal Natuurhistorisch Museum Naturalis, KNNV Uitgeverij & EIS, Leiden, The Netherlands.
- SOVON, consulted in 2017. Oehoe: Verspreiding en aantalsontwikkeling. <http://s1.sovon.nl/soorten.asp?euring=7440>.
- Strachan, R., Moorhouse, T., 2006. *Water Vole Conservation Handbook*, 2nd Edition. Wildlife Conservation Research Unit, University of Oxford, Oxford, UK.
- Student, 1908. The probable error of a mean. *Biometrika* 6 (1), 1–25.
- Tang, J., Song, Y., Miller, H. J., Zhou, X., 2016. Estimating the most likely space–time paths, dwell times and path uncertainties from vehicle trajectory data: A time geographic method. *Transportation Research Part C: Emerging Technologies* 66, 176–194.
- Tang, W., Bennett, D. A., 2010. Agent-based modeling of animal movement: A review. *Geography Compass* 4 (7), 682–700.
- Taylor, C. M., Stutchbury, B. J., 2016. Effects of breeding versus winter habitat loss and fragmentation on the population dynamics of a migratory songbird. *Ecological Applications* 26 (2), 424–437.
- The Cornell Lab of Ornithology, 2018. All about birds. URL www.allaboutbirds.org
- Tomkiewicz, S., Ruller, M., Kie, J., Bates, K., 2010. Global positioning system and associates technologies in animal behavior and ecological research. *Philosophical transactions of the royal society B: Biological Science* 365 (1150), 2163–2176.
- Tracey, J., Zhu, J., Crooks, K., 2005. A set of nonlinear regression models for animal movement in response to a single landscape feature. *Journal of Agricultural, Biological, and Environmental Statistics* 10 (1), 1.

- Tremblay, M. A., St Clair, C. C., 2009. Factors affecting the permeability of transportation and riparian corridors to the movements of songbirds in an urban landscape. *Journal of Applied Ecology* 46 (6), 1314–1322.
- Trierweiler, C., Klaassen, R. H., Drent, R. H., Exo, K.-M., Komdeur, J., Bairlein, F., Koks, B. J., 2014. Migratory connectivity and population-specific migration routes in a long-distance migratory bird. In: *Proceedings of the Royal Society of London B: Biological Sciences*. Vol. 281. The Royal Society.
- Tucker, M. A., Böhning-Gaese, K., Fagan, W. F., Fryxell, J. M., Van Moorter, B., Albers, S. C., Ali, A. H., Allen, A. M., Attias, N., Avgar, T., et al., 2018. Moving in the anthropocene: Global reductions in terrestrial mammalian movements. *Science* 359 (6374), 466–469.
- UCLA: Statistical Consulting Group, 2018. Introduction to generalized linear mixed models.
URL <https://stats.idre.ucla.edu/other/mult-pkg/introduction-to-generalized-linear-mixed-models/>
- Uhlenbroek, C., 2008. *Animal Life*, 1st Edition. Tirion Uitgevers BV, Baarn, The Netherlands.
- UNEP and FAO, 1991. Operation of the prior informed consent procedure for banned or severely restricted chemicals in international trade. Decision guidance documents DDT. Tech. rep., United Nations Environment Programme and Food and Agriculture Organization of the United Nations, Rome - Geneva.
- Université de Montréal, 2018. Critical values for the two-sample kolmogorov-smirnov test (2-sided).
URL https://www.webdepot.umontreal.ca/Usagers/angers/MonDepotPublic/STT3500H10/Critical_KS.pdf
- Urbano, F., Cagnacci, F., Calenge, C., Dettki, H., Cameron, A., Neteler, M., 2010. Wildlife tracking data management: a new vision. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1550), 2177–2185.
- U.S. Fish and Wildlife Service, 1981. Standards for the development of habitat suitability index models. Tech. rep., U.S. Fish and Wildlife Service, Washington D.C., USA.
- van de Poel, J. L., Dekker, J., Van Langevelde, F., 2015. Dutch hedgehogs (*Eri-naceus europaeus*) are nowadays mainly found in urban areas, possibly due to the negative effects of badgers (*Meles meles*). *Wildlife Biology* 21 (1), 51–55.
- van den Brink, N., Jansman, H., 2005. Verontreinigingen in oehoes (*Bubo bubo*) uit Limburg en Twente; onverwacht verhoogde concentraties van PCBs in oehoes uit Limburg. Tech. rep., Alterra, Wageningen, The Netherlands, Alterra-Rapport 1317.

- Van der Spek, S., Van Schaick, J., De Bois, P., De Haan, R., 2009. Sensing human activity: Gps tracking. *Sensors* 9 (4), 3033–3055.
- Van Dyck, H., Baguette, M., 2005. Dispersal behaviour in fragmented landscapes: routine or special movements? *Basic and Applied Ecology* 6 (6), 535–545.
- Van Horne, B., Wiens, J. A., 1991. Forest bird habitat suitability models and the development of general habitat models. Tech. rep., COLORADO STATE UNIV FORT COLLINS.
- van Lierop, S., Janssen, R., 2014. Inventarisatie van oehoe-territoria en hun broedsucces in Limburg in 2014. Tech. rep., Bionet Stein en van Lierop Natuuradvies & Onderzoek, Nijmegen, The Netherlands.
- van Moorter, B., Bunnefeld, N., Panzacchi, M., Rolandsen, C. M., Solberg, E. J., Sæther, B.-E., 2013. Understanding scales of movement: animals ride waves and ripples of environmental change. *Journal of Animal Ecology* 82 (4), 770–780.
- Van Nieuland, S., Baetens, J. M., De Baets, B., 2016. Spatially Explicit Modelling of the Belgian Major Endurance Event ‘The 100 km Dodentocht’. *PloS one* 11 (10), e0164981.
- Van Nieuland, S., Baetens, J. M., De Meyer, H., De Baets, B., 2015. An analytical description of the time-integrated Brownian bridge. *Computational and Applied Mathematics* 36 (1), 1–19.
- Van Nieuland, S., Baetens, J. M., Janssen, R., De Baets, B., 2018a. A validated expert-based habitat suitability map for Eagle Owls in Limburg, the Netherlands, submitted to *European Journal of Wildlife Research*.
- Van Nieuland, S., Baetens, J. M., Janssen, R., De Baets, B., 2018b. Environmental characteristics of possible hunting grounds of the Eurasian Eagle Owl in Limburg, the Netherlands. *Ardea*.
- Vananderoye, S., May 2014. Antwerp 10 Miles is big business.
URL <http://www.stampmedia.be/2014/05/antwerp-10-miles-is-big-business/>
- Veerarajan, T., 2008. *Probability, Statistics and Random Processes*. Tata McGraw-Hill, New Delhi, India.
- Venuti, F., Bruno, L., 2007. An interpretative model of the pedestrian fundamental relation. *Comptes Rendus Mécanique* 335 (4), 194–200.
- Vernasco, B. J., Sillett, T. S., Marra, P. P., Ryder, T. B., 2017. Environmental predictors of nestling condition, postfledging movement, and postfledging survival in a migratory songbird, the Wood Thrush (*Hylocichla mustelina*). *The Auk* 135 (1), 15–24.

- Versichele, M., Neutens, T., Delafontaine, M., Van de Weghe, N., 2012. The use of bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the ghent festivities. *Applied Geography* 32 (2), 208–220.
- Vitz, A. C., Rodewald, A. D., 2011. Influence of condition and habitat use on survival of post-fledging songbirds. *The Condor* 113 (2), 400–411.
- Voous, K. H., Cameron, A., 1990. *Owls of the Northern Hemisphere*. Collins, London, UK.
- Voskamp, P., 2004. Opmars van oehoes in Zuid-Limburg. *Limburgse Vogels* 14, 1–8.
- Wakeley, J. S., 1988. A method to create simplified versions of existing habitat suitability index (HSI) models. *Environmental Management* 12 (1), 79–83.
- Walck, C., 1996. *Handbook on Statistical Distributions for Experimentalists*. Particle Physics Group Fysikum - University of Stockholm.
- Walter, W. D., Onorato, D. P., Fischer, J. W., 2015. Is there a single best estimator? selection of home range estimators using area-under-the-curve. *Movement Ecology* 3 (1), 10.
- Wassink, G., 2007. De oehoe (*Bubo bubo*) in Nederland en het Duitse laagland in 2007. Tech. rep., Lievelede, The Netherlands.
- Wassink, G., 2010a. GPS-telemetrieonderzoek aan vijf Nederlandse oehoes (*Bubo bubo*) in 2008 en 2009. Tech. rep., Vogelwerkgroep Zuidoost-Achterhoek, Winterswijk, The Netherlands.
- Wassink, G., 2010b. Wat is er aan de hand met de oehoe in limburg? *Limburgse Vogels* 20, 52–58.
- Wassink, G., 2011a. GPS-onderzoek aan de oehoe in 2010. Tech. rep., Oehoewerkgroep Nederland, Lievelede, The Netherlands.
- Wassink, G., 2011b. Nestplaatskeus van oehoes in het grensgebied van Nederland en Duitsland. *Uilen* 2 (11), 49–55.
- Wassink, G., 2012. GPS-onderzoek aan de oehoe in 2011/2012. Tech. rep., Oehoewerkgroep Nederland, Lievelede, The Netherlands.
- Wassink, G., 2014a. Dispersie van jonge oehoes in beeld gebracht met satellietzenders en GPS -loggers. *Limosa* 87, 91–98.
- Wassink, G., 2014b. GPS-onderzoek aan de oehoe in 2013/2014. Tech. rep., Oehoewerkgroep Nederland, Lievelede, The Netherlands.

- Wassink, G., 2014c. Populatieontwikkeling van de oehoe (*Bubo bubo*) in Nederland en West-Duitsland met een blik op de toekomst. Tech. rep., Wassink Natuurprojecten, Lievelede, The Netherlands.
- Wassink, G., 2016. Oehoewerkgroep Nederland.
URL <http://www.oehoewerkgroep.nl>
- Weidmann, U., 1993. Transporttechnik der Fussgänger. Tech. Rep. Ivt Report no. 90., ETH Zürich, Germany.
- Werman, M., Peleg, S., Rosenfeld, A., 1985. A distance metric for multidimensional histograms. *Computer Vision, Graphics, and Image Processing* 32 (3), 328–336.
- Westfall, P. H., 2014. Kurtosis as peakedness, 1905–2014. RIP. *The American Statistician* 68 (3), 191–195.
- Whittaker, K., Marzluff, J. M., 2012. Post-fledging mobility in an urban landscape. *Studies in Avian Biology* 45, 183–198.
- Wikelski, M., Kays, R., 2014. Movebank: archive, analysis and sharing of animal movement data. World Wide Web electronic publication. <http://www.movebank.org>, accessed on 25/09/2014.
- Wilcove, D., 2008. No way home: the decline of the world's great animal migrations. Island press, Washington, USA.
- Wilcove, D., Wikelski, M., 2008. Going, going, gone: Is animal migration disappearing? *Plos Biology* 6 (7), e188.
- Wilson, R. P., Ducamp, J., Rees, G., Culik, B., Niekamp, K., 1992. Wildlife telemetry: remote monitoring and tracking of animals. Ellis Horwood, London, UK, Ch. Estimation of location: global coverage using light intensity, pp. 131–134.
- Wilson, R. P., Grémillet, D., Syder, J., Kierspel, M. A., Garthe, S., Weimerskirch, H., Schäfer-Neth, C., Scolaro, J. A., Bost, C.-A., Plötz, J., Nel, D., 2002. Remote-sensing systems and seabirds: their use, abuse and potential for measuring marine environmental variables. *Marine Ecology Progress Series* 228, 241–261.
- Wilson, R. P., Liebsch, N., Davies, I. M., Quintana, F., Weimerskirch, H., Storch, S., Lucke, K., Siebert, U., Zankl, S., Müller, G., Zimmer, I., Scolaro, A., Campagna, C., Plötz, J., Bornemann, H., Teilmann, J., McMahon, C. R., 2007. All at sea with animal tracks; methodological and analytical solutions for the resolution of movement. *Deep Sea Research Part II: Topical Studies in Oceanography* 54 (3-4), 193–210.
- Wolfram Research Inc., 2018a. Introduction to Solving Nonlinear Equations.
URL <http://reference.wolfram.com/language/tutorial/UnconstrainedOptimizationIntroductionNonlinearEquations.html#771134545>

- Wolfram Research Inc., 2018b. Numerical Nonlinear Global Optimization.
URL <http://reference.wolfram.com/language/tutorial/ConstrainedOptimizationGlobalNumerical.html>
- Wolfram Research Inc., 2018c. Partitioning Data into Clusters.
URL <https://reference.wolfram.com/language/tutorial/PartitioningDataIntoClusters.html>
- Yackel Adams, A. A., Skagen, S. K., Savidge, J. A., 2006. Modeling post-fledging survival of lark buntings in response to ecological and biological factors. *Ecology* 87 (1), 178–188.
- Yager, R. R., Rybalov, A., 1996. Uninorm aggregation operators. *Fuzzy Sets and Systems* 80 (1), 111–120.
- Yan, J., Chen, Y., Lawrence-Apfel, K., Ortega, I., Pozdnyakov, V., Williams, S., Meyer, T., 2014. A moving-resting process with an embedded Brownian motion for animal movements. *Population Ecology* 56 (2), 401–415.
- Yong, A. G., Pearce, S., 2013. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology* 9 (2), 79–94.
- Young, R. P., Davison, J., Trewby, I. D., Wilson, G. J., Delahay, R. J., Doncaster, C. P., 2006. Abundance of hedgehogs (*Erinaceus europaeus*) in relation to the density and distribution of badgers (*Meles meles*). *Journal of Zoology* 269 (3), 349–356.
- Zajac, Z., Stith, B., Bowling, A. C., Langtimm, C. A., Swain, E. D., 2015. Evaluation of habitat suitability index models by global sensitivity and uncertainty analyses: A case study for submerged aquatic vegetation. *Ecology and Evolution* 5 (13), 2503–2517.

Curriculum vitae

Personalia

| | |
|----------------|--|
| Name | Steffie Van Nieuland |
| Date of birth | 11/08/1988 |
| Place of birth | Ghent, Belgium |
| Nationality | Belgian |
| E-mail | Steffie.VanNieuland@UGent.be, steffie_vn@hotmail.com |

Education

2009 - 2011 : M.Sc. bioscience engineering: land- and water management, Ghent University, Ghent, Belgium

2006 - 2009 : B.Sc. bioscience engineering: land- and forest management, Ghent University, Ghent, Belgium

2002 - 2006 : Secondary school Science Mathematics, Sint Janscollege, Ghent, Belgium

Employment

September 2011 - September 2018 : PhD candidate/teaching assistant at the Research Unit Knowledge-based Systems, Department of Data Analysis and Mathematical Modelling, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium

Tutorship of master and bachelor theses

Een Habitatgeschiktheidskaart Voor De Oehoe: Kennisgebaseerde Versus Datagedreven Modelling, academic year 2016 (Sofie Vriens).

Datagedreven modellering van dierenmigratie aan de hand van Brownse bruggen, academic year 2013 (Niels Vanheuverbeke).

Migratie van dieren: een fascinerend fenomeen, academic year 2012 (Sofie Balemans, Sofie Baeyens, Brecht D'hont, Anton Gistelincx)

Scientific output

Publications in international journals

Zachary S. Ladin, Steffie Van Nieuland, Solny A. Adalsteinsson, Vincent D'Amico, Jacob L. Bowman, Jeffrey J. Buler, Jan M. Baetens, Bernard De Baets, W. Gregory Shriver (2018) Differential post-fledging habitat use of Nearctic-Neotropical migratory birds within an urbanized landscape. *Movement Ecology* 6(17).

Steffie Van Nieuland, Jan M. Baetens and Bernard De Baets (2016) Spatially Explicit Modelling of the Belgian Major Endurance Event 'The 100 km Dodentocht' *PLOS one* 11(10).

Steffie Van Nieuland, Jan M. Baetens, Hans De Meyer, Bernard De Baets (2015) An analytical description of the time-integrated Brownian bridge. *Journal of Computational and Applied Mathematics* (36) 627-645.

Jan M. Baetens, Steffie Van Nieuland, Ine Pauwels, Bernard De Baets, Ans Mouton and Peter Goethals (2013) An Individual-based Model for the Migration of Pike (*Esox Lucius*) in the River Yser, Belgium. *Ecological modelling* (258) 40-52.

Ine Pauwels, Ans Mouton, Jan M. Baetens, Steffie Van Nieuland, Bernard De Baets and Peter Goethals (2013) Modelling a Pike (*Esox Lucius*) Population in a Lowland River Using a Cellular Automaton. *Ecological informatics* (17) 46-57.

Submitted manuscripts

Steffie Van Nieuland, Jan M. Baetens, René Janssen, Bernard De Baets (2018)
A validated expert-based habitat suitability map for Eurasian Eagle Owls in Limburg, the Netherlands. *European Journal of Wildlife Research* (submitted)

Steffie Van Nieuland, Jan M. Baetens, René Janssen, Bernard De Baets (2017)
Environmental characteristics of possible hunting grounds of the Eurasian Eagle Owl in Limburg, the Netherlands. *Ardea* (accepted)

Attended conferences

Steffie Van Nieuland, Jan M. Baetens, René Janssen, Bernard De Baets (2016)
Spatio-temporal behavior of the Eurasian Eagle Owl : An individual-based approach
Oral presentation at ECMTB 2016: European conference on Mathematical and Theoretical Biology in Nottingham (United Kingdom): 11 July - 15 July 2016

Steffie Van Nieuland, Jan M. Baetens, Bernard De Baets (2014) Animal migration modelling: an analytical description of the two-dimensional time-integrated Brownian bridge
Oral presentation at ISEC2014, 4th International Statistical Ecology Conference in Montpellier (France): 1-4 July 2014

Steffie Van Nieuland, Jan M. Baetens, Bernard De Baets (2014) An analytical description of the time-integrated Brownian bridge
Oral presentation at NSAB2014: 19th national symposium on applied biological science in Gembloux (Belgium): 7 February 2014

Steffie Van Nieuland, Jan M. Baetens, Bernard De Baets (2012) A spatially explicit migration model for pike (*Esox lucius* L.)
Poster presentation at ACRI 2012 Cellular automata for research and industry in Santorini (Greece): 24-27 September 2012

Active participation at international scientific events

NSAB2014: 18th national symposium on applied biological science (2013): Member of the organizing committee