A wildlife journey in space and time

Methodological advancements in the assessment and analysis of spatio-temporal patterns of animal movement across European landscapes

By Msc. Johannes De Groeve

PhD in Geography

Ghent University, 2018

Department of Geography
Ghent University, Belgium

Department of Biodiversity and Molecular Ecology
Research and Innovation Centre
Edmund Mach Foundation, Italy

Supervisors:
Prof. dr. Nico Van de Weghe (UGent)
Dr. Francesca Cagnacci (FEM)

Supervisors

Prof. dr. Nico Van de Weghe, Ghent University

Dr. Francesca Cagnacci, Research and Innovation Center, Edmund Mach Foundation

Members of the Examination Committee

Dr. Emiel van Loon, University of Amsterdam

Dr. Nicolas Morellet, University of Toulouse, INRA

Prof. dr. Jan Baetens, Ghent University

Prof. dr. Luc Lens, Ghent University

Prof. dr. Rudi Goossens, Ghent University

Prof. dr. Veerle Van Eetvelde, Ghent University

Dean: Prof. dr. Isabel Van Driessche

Rector: Prof. dr. Rik Van de Walle

TABLE OF CONTENT

I. INDEX OF TABLES

INTRODUCTION

DATA

Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods

Individual Movement - Sequence Analysis Methods (IM-SAM): characterising spatio-temporal patterns of animal trajectories across scales and landscapes.

Functional role of seasonal habitat determines spatio-temporal patterns of habitat use in a large herbivore

DISCUSSION

II. INDEX OF FIGURES

INTRODUCTION

DATA

Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods

Individual Movement - Sequence Analysis Methods (IM-SAM): characterising spatio-temporal patterns of animal trajectories across scales and landscapes.

Functional role of seasonal habitat determines spatio-temporal patterns of habitat use in a large herbivore

DISCUSSION

III. ABSTRACT

English:

Movement is one of the most fundamental processes for living entities on earth at the core of scientific disciplines such as ecology and geography. In animal ecology, ongoing progress in tracking and remote sensing technologies has spurred an explosion of movement and environmental data collected at high spatial and temporal resolution, at a large scale, so that the interaction between animal movement and habitat features can now be investigated in much more detail. As a result, in recent years the field of animal ecology has produced a growing body of studies on movement-based patterns leading to habitat use and selection. In this regard, GIScience has contributed with several visual analytical approaches to study animals in relation to their environment and habitat. However, the patterns behind the sequential use of different habitat classes have remained largely unexplored. Sequential habitat use is defined as the consecutive use of habitat features along the trajectory of an animal, extracted from the context of its spatial movement. By accounting for the sequence of use, it is possible to distinguish fundamentally different behavioural habitat use strategies that are important for the survival and fitness of an animal, such as habitat alternation versus random sequential use. Such distinctions would remain undetected by only considering the proportion of use. Sequential habitat use patterns occur in a spatial context, meaning sequential patterns are affected by what is actually available to the animal.

In this dissertation we merge knowledge from different fields to present an innovative method to study the relation between animals and their environment by accounting for the sequential use of habitats, and animal movement rules. We developed a visually effective method to analyse and visualise sequential habitat use patterns of animals at multiple spatio-temporal scales by combining real and simulated sequences of habitat use. To study sequential habitat use patterns we use Sequence Analysis Methods (SAM), an approach widely applied in molecular biology, as well as many applications in different fields, to measure dissimilarity between sequences of characters. In brief, we use dissimilarity algorithms to measure the distance between all pairs of sequences, and then apply a clustering algorithm to investigate how these sequences group together, which are visualised as dissimilarity trees. We propose a procedure consisting of three steps, including explo-

ration, simulation and classification. In the exploration phase, we build exploratory trees, which visualise *real* sequential habitat use patterns. Second, by applying animal movement models we simulate expected sequential habitat use patterns, and assess how spatial context, and especially habitat availability, affects the clustering of sequential patterns. Third, we combine real and simulated sequences to identify which simulated pattern is most parsimonious with the real sequences. The research progress has been presented in three main chapters. In Chapter 3 we present seminal methodological development where SAM was applied to animal movement data. In Chapter 4 we introduce further methodological advancements to extend the applicability of SAM to animal ecology. In Chapter 5 we present a large-scale multi-population ecological application. All research was performed using GPS movement data of roe deer and environmental data provided by the Euroungulates database project.

Chapter 3 presents the first application of SAM to identify ecologically relevant sequential patterns in animal habitat use. We exemplify the method using ecological data consisting of simulated and real trajectories from a roe deer population (*Capreolus capreolus*) in the Italian Alps, expressed as ordered sequences of four habitat use classes, i.e. high/open, high/closed, low/open, low/closed. In essence, the SAM framework identifies relevant sequential patterns in real trajectories by measuring their similarity to spatially-explicit simulated trajectories with known sequential patterns. Simulation trajectories were generated in arenas resembling the landscape structure of the roe deer population. Chapter 4 extends SAM to an individual-based approach (i.e. IM-SAM, Individual Movement – Sequence Analysis Methods), that is applicable over multiple populations. Specifically, instead of performing simulations in landscape-like arenas, we use real individual home ranges, thus accounting for individual spatial context, and landscape composition and structure. To assess usability of our advanced framework we investigate the sequential use of open and forest habitats for nine roe deer populations ranging in landscapes with different geographic contexts and anthropogenic disturbance. We also discuss implications for conservation and management. Chapter 5 addresses the functional role of landscapes throughout seasons by identifying both population level and individual level variability in the sequential habitat use patterns of roe deer, identified in the former nine roe deer populations. We show how identified sequential habitat use patterns can be treated as variables, and analysed with standard and well-accepted statistical methods.

While the (IM-)SAM framework was developed for studying sequential habitat use in specific, we highlight that its methodological steps and study design can easily be generalised. Indeed, its dissimilarity and clustering algorithms, temporal resolution, sampling units, and number of classes for which sequential patterns are investigated can all be customised for the specific research questions in mind. (IM-)SAM is easily applicable to different types of sequential data that describe aspects of an animal's internal (e.g. heart rate) or external state (e.g. temperature). Through improvements in technology, including the growing number of information that can be collected through sensors (GPS trackers, biologgers and satellites), improving database infrastructures and the instant availability of advanced R packages dedicated to animal movement, (IM-)SAM could be easily integrated in a wide range of both local and broad-scaled behavioural spatio-temporal studies.

Nederlands:

Beweging is een van de meest fundamentele processen voor levende wezens op aarde dat aan de basis ligt van heel wat wetenschappelijke disciplines zoals ecologie en geografie. In dierenecologie heeft technologische vooruitgang in de ontwikkeling van tracking- en remote sensing technologieën recent geleid tot een explosie aan informatie over beweging en de omgevingscontext. Zowel de resolutie als de schaal waarin deze informatie nu beschikbaar is blijft toenemen en laten beter dan ooit toe te bestuderen hoe de bewegingen van dieren samenhangen met habitatkenmerken. De wetenschappelijke interesse in dierenecologie is de laatste decennia sterk ontwikkeld omwille van zulke technologische vooruitgang, resulterend in een groeiend aantal studies die gedetailleerde veranderingen in habitatgebruik en -selectie toelaten. Ook in geografische informatiewetenschap en gerelateerde onderzoeksgebieden zijn verschillende visueel-analytische benaderingen ontwikkeld om dieren te bestuderen in relatie tot hun habitat. Wat echter zelden in rekening wordt genomen, is de volgorde of sequentie waarin verschillende habitatklassen worden gebruikt. Sequentieel habitatgebruik wordt gedefinieerd als het opeenvolgende gebruik van habitatkenmerken langs het bewegingstraject van een dier. Door rekening te houden met de sequentie in habitatgebruik, is het mogelijk om fundamenteel verschillende strategieën te onderscheiden die een belangrijke invloed hebben op de overleving en *fitness* van organismen, zoals een afwisselend patroon in habitatgebruik naargelang dag/nacht versus een willekeurig patroon in habitatgebruik. Dergelijke patronen kunnen niet ontdekt worden in de meer klassieke habitatanalyses, waarbij enkel wordt gekeken naar proporties in habitatgebruik. De volgorde waarin habitats langs een bewegingstraject worden gebruikt hangt ook af van de ruimtelijke context. Bijgevolg zijn geobserveerde habitatgebruikspatronen afhankelijk van wat beschikbaar is voor de diersoort.

In deze dissertatie brengen we kennis uit verschillende wetenschappelijke disciplines samen om tot een nieuwe methode te komen voor het bestuderen van habitats die zowel de volgorde van het gebruik, als de ruimtelijke bewegingscontext in rekening brengen. We stellen een visueel aantrekkelijke methode voor die het analyseren en visualiseren van sequentiële habitatgebruikspatronen van dieren toelaat op basis van een combinatie van reële en gesimuleerde sequenties in habitatgebruik. De methode laat toe om habitatgebruik op meerdere tijdruimtelijke schalen te verkennen en te analyseren. Om sequenties in habitatgebruik te bestuderen, gebruiken we *Sequence Analysis Methods* (SAM), een analysetechniek die standaard wordt toegepast in de moleculaire biologie, en verschillende andere vakgebieden. De basistechniek bestaat erin om de ongelijkheid of dissimilariteit tussen verschillende karaktersequenties (e.g. COOCOO) te berekenen, waarbij elk karakter (e.g. O, C) een status (e.g. open/gesloten habitat) voorstelt op een bepaald tijdstip. Samengevat maken we gebruik van dissimilariteitsalgoritmen om de afstand tussen alle sequenties te berekenen, en gebruiken vervolgens een clusteringalgoritme wat toelaat om de dissimilariteit tussen sequenties visueel voor te stellen in een dendrogram. De algemene procedure die we voorstellen bestaat uit drie stappen: exploratie, simulatie en classificatie. In de exploratiefase genereren we dendrogrammen die reële habitatgebruikspatronen visualiseren. In de simulatiefase genereren we de verwachte habitatgebruikspatronen en analyseren we hoe ruimtelijke context de gegenereerde trajecten beinvloedt. In de classificatiefase combineren we de reële en gesimuleerde sequenties om te identificeren welk gesimuleerd patroon het best overeenkomt met de reële sequenties. De voortgang van het onderzoek wordt beschreven in drie hoofdstukken. In hoofdstuk 3 presenteren we de rudimentaire methodologische ontwikkeling, in hoofdstuk 4 beschrijven we methodologische innovaties en in hoofdstuk 5 presenteren we een ecologische toepassing. Al het onderzoek werd uitgevoerd met behulp van GPS-localisatie van reeën en habitatinformatie geleverd door het *Euroungulates database* project.

Hoofdstuk 3 presenteert de eerste toepassing van SAM om ecologisch relevante habitatgebruikspatronen te identificeren. We illustreren de methode met behulp van ecologische gegevens die bestaan uit gesimuleerde en reële trajecten van een reeënpopulatie (*Capreolus capreolus*) in de Italiaanse Alpen, uitgedrukt als opeenvolgende karaktersequenties van vier habitatgebruiksklassen (status), namelijk hoog/open habitat, hoog/gesloten habitat, laag/open habitat, laag/gesloten habitat. In essentie identificeert het SAM-framework relevante sequentiële patronen in reële trajecten door hun gelijkenis te meten met ruimtelijk expliciet gesimuleerde trajecten met bekende sequentiële patronen. Gesimuleerde trajecten werden gegenereerd in gegenereerde arena's die lijken op de landschapsstructuur van de reeënpopulatie. Hoofdstuk 4 breidt SAM uit naar een individuele benadering (IM-SAM, *Individual Movement – Sequence Analysis Methods*). In plaats van het uitvoeren van simulaties in gegenereerde arena's gebruiken we hier de eigenlijke *home range* van elk dier. Door deze uitbreiding nemen we de individuele ruimtelijke bewegingscontext in rekening en dus de landschapsstructuur waarin een dier beweegt. Om de bruikbaarheid van ons geavanceerde framework te beoordelen, onderzoeken we het sequentieel habitatgebruik van open- en gesloten habitats voor negen reeënpopulaties, die zich bevinden in diverse landschappen in Europa. We bespreken ook de implicaties voor behoud en beheer. In hoofdstuk 5 focussen we meer op de functionele rol van landschappen doorheen de seizoenen door het identificeren van populatie- en individuele variabiliteit in de sequentiële habitatgebruikspatronen van reeën, geïdentificeerd in de voormalige negen reeënpopulaties. We tonen aan hoe de geidentificeerde sequentiële habitatgebruikspatronen verder kunnen worden gebruikt als variabelen in statistische analyses.

Hoewel (IM-)SAM werd ontwikkeld voor het bestuderen van sequentieel habitatgebruik in het bijzonder, benadrukken we dat deze methodologie generalizeerbaar is om een breed spectrum aan onderzoeksvragen te behandelen. Bijvoorbeeld, dissimilariteits- en clusteringalgoritmen, temporele resoluties, studie eenheden en het aantal klassen (aantal karakters) waarvoor patronen worden onderzocht, kunnen worden aangepast aan de specifieke onderzoeksvragen. Bovendien is (IM-)SAM gemakkelijk toepasbaar op verschillende informatiebronnen die de interne status (e.g. hartslag) of omgevingswereld (e.g. temperatuur) van een dier kunnen beschrijven. Door de continue verbeteringen in sensor-technologie (GPS trackers, biologgers en satellieten), de verbetering van database-infrastructuren en de onmiddellijke beschikbaarheid van geavanceerde R software met specifieke codes speciaal ontwikkeld voor de studie van bewegende dieren, kan (IM-)SAM gemakkelijk geïntegreerd worden zowel in lokale als continentale/globale tijdruimtelijke gedragsstudies.

IV. PUBLICATIONS

A1 Publications

De Groeve, J., Van de Weghe, N., Ranc, N., Neutens, T., Ometto, L., Rota-Stabelli, O. and Cagnacci, F., 2016. Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods. *Methods in Ecology and Evolution*, **7** (3),369-379. IF: 6.554

De Groeve, J., Cagnacci, F., Ranc, N., Bonnot, N., Gehr, B., Heurich, M., Hewison, A.J.M., Morellet, N., Mysterud, A., Kroeschel, M., Linnell, J.D.C., Sandfort, R. and Van de Weghe, N., *(submitted)* Individual Movement - Sequence Analysis Methods (IM-SAM): characterising spatio-temporal patterns of animal trajectories across scales and landscapes. *International Journal of Geographical Information Science*. IF: 2.502

De Groeve, J., Van de Weghe, N., Ranc, N., Bonnot, N., Gehr, B., Heurich, M., Hewison, A.J.M., Morellet, N., Mysterud, A., Kroeschel, M., Linnell, J.D.C., Sandfort, R. and Cagnacci, F., *(to be submitted)* Functional role of seasonal habitat determines spatio-temporal patterns of habitat use in a large herbivore. *Ecography*. IF: 4.902

Through additional material of this research and thesis not included in the aforementioned publications, I am also contributing to the preparation of two further manuscripts:

Salvatori, M., De Groeve, J., Van de Weghe, N., De Baets, B., Moorcroft, P., Focardi, S., Bonnot, N., Gehr, B., Heurich, M., Hewison, A.J.M., Morellet, N., Kroeschel, M., Licoppe, A., Pedrotti, L., Signer, J. and Cagnacci F., (*in preparation*) Habitat use seen from the satellites: what can we catch? A cross-population study on two cervid species. Target journal: *Journal of Applied Ecology.* IF: 5.30

Cagnacci, F., De Groeve, J., the Euroungulates consortium and Urbano F., (*in preparation*) Euroungulates: a bottom-up collaborative science initiative for movement ecology, *Movement Ecology*. IF-score from June 2018.

Conference publications

De Groeve, J., Van de Weghe, N. and Cagnacci, F., 2017. Studying a species at the distribution range scale – spatiotemporal habitat use patterns in roe deer, Invited speaker, *Open Science Tools, Data & Technologies for Efficient Ecological & Evolutionary Research Symposium*, 7-8 Dec, Amsterdam, The Netherlands.

De Groeve, J., Van de Weghe, N. and Cagnacci, F., 2017. Poster: Sequential Habitat Use by Animals: A Methodological Workflow. *Lorentz Center Workshop “Movement: New Sensors, New Data, New Challenges”*, 21-25 Aug., Leiden, The Netherlands.

De Groeve, J., Van de Weghe, N., Ranc, N., Moorcroft, P. and Cagnacci, F., 2017. Pan-European Geographic Diversity in Spatiotemporal Sequential Habitat Use by a Large Herbivore. *NSF Mobility Workshop*, Ohio State University, 11-12 May, Columbus, Ohio, USA.

Cagnacci, F. and De Groeve, J., 2016. Landscape of fear in a human-dominated environment: An approach to incorporate spatio-temporal components in diel cycles. *Canadian Society for Ecology and Evolution*, CSEE 2016, 7-11 July, St. John’s, Newfoundland.

De Groeve, J., Van de Weghe, N., Ranc, N., Neutens, T., Ometto, L., Rota-Stabelli, O. and Cagnacci, F., 2016. Extracting similar movement patterns in animal trajectories using sequence analysis methods. *Association of American Geographers Annual Meeting,* AAG 2016, 29 Mar.-2 Apr., San Francisco, California, USA.

V. PREFACE

A doctoral thesis is rightly considered as the most essential part in the PhD degree. It gives a student the opportunity to demonstrate his or her full academic capacities, which have been established over their years of education. Four academic years are devoted to collecting information, processing knowledge and acquiring insights into a specific topic. Therefore, it is of utmost importance to study a subject which is dear to the heart. My search has brought me to this research, a combination of my major interests in geography and animal behaviour with my fascination for technological advancements.

First, I would like to thank the members of my Examination Committee. With their extremely relevant feedback, stemming from a wide range of study fields, I could critically evaluate and improve the first draft of my dissertation.

This research could by no means have been carried out without the help of several people. In my case, there can be no doubt that I received great quality guidance, and would therefore like to express my gratitude to a number of persons. Firstly I would like to thank my supervisors Dr. Francesca Cagnacci and Prof. dr. Nico Van de Weghe.

I would like to express my gratitude to Francesca Cagnacci, from the Research and Innovation Centre Fondazione Edmund Mach (FEM), for this intense collaboration. Francesca was one person that was by my side during my entire research project. She is not only a great scientist, but also big-hearted. She is truly proud of my accomplishments as well as her contribution to it. She not only cares about professional research development, but also personal and skill development. Next to scientific support, she helped me to find my interests, provided motivation, helped me to build more confidence in writing and presenting, and gave personal support and encouragement. She is much more than a supervisor, as she is also a friend.

I would also like to thank my supervisor at Ghent University Nico Van de Weghe for the opportunities he has afforded me, which have accumulated in this exciting and interesting endeavour enabling my study of animal behaviour in a geographical context. During my master thesis, he opened a discussion at “Summer School and young researchers forum: Moving objects and knowledge discovery (August 2012)” with Francesca Cagnacci, which initially opened my eyes to the research possibilities out there. After my masters, it was he who suggested I go to Italy for an internship to further develop my research skills. He always gave me freedom in my research, and in personal life choices. He knew I built an exciting life in Italy, and did not interfere with that. Moreover, when I was in Belgium he always made time for me for meetings, and showed his excitement for the research I was performing. I would like to thank him for his general great support throughout the PhD.

One of the most fundamental aspects of research is teamwork. At FEM, I have been fortunate to work in an amazing lab with a great atmosphere. Besides working together we did plenty of activities such as dining, hiking, and climbing. In particular, I would like to thank PhD student Nathan Ranc for his enthusiasm, the mutual love of listening to music in the 'death metal dungeon' (i.e. our office), and also for his scientific support. It was he who developed mechanistic movement models for the purpose of my research. Without his contributions, reaching the point that I am now would have been intensely more complicated. Secondly, I would like to thank Dr. Federico Ossi. Fede was already in our lab when I arrived for the first time in Italy. Having a good knowledge of models, he has helped me a great deal with the interpretations. I would also like to thank Marco Salvatori, whom I supported in his master thesis. Over a course of one year he helped me with a specific topic that supports my PhD research. Throughout this year, he reached a level of scientific maturity beyond expectation of a master student. I'd be remiss excluding fellow PhD students, Steffen Mumme and Andrea Corradini, from my list of thank-you's; despite joining our research group in November 2017, it feels like we have known each other for much longer. I would like to thank Whitney Hansen for bringing joy to the lab, for proof reading the final version of this thesis, and for being awesome. I would also like to thank the many other people that joined our lab over the years, in particular Tobia Sforna, Ginevra Lilli, Marrit Leenstra, Priscilla Bonanni, Maria Beatriz al Meida, Julius Ross, Paola Semenzato, Wibke Peters, Daniele Falcinelli, Filippo Bianchini, Sara Mancinelli, Simona Picardi, Valentina Zini and Maria Valent. I am sure I will keep in touch with many of them.

Due to my absence at Ghent University for large parts of my PhD, I have less close relationships with colleagues. Nevertheless, when I was there I immediately felt at home in the research lab. For this reason, I would like to thank my colleagues Jasper Beernaert, Tim Baert, Pengdong Zhang and Alexander Duytschaever. I am also grateful to Dr. Tijs Neutens who, together with Nico Van de Weghe, supported me not only during the master thesis, but also throughout the first year of the PhD, before his career took another turn. I would also like to note my appreciation for Helga Vermeulen's administrative help.

I would also like to thank the people behind EuroDEER, primarily the database developer Dr. Ferdinando Urbano, for teaching me PostgreSQL, technical support, and for always giving me such immediate feedback. He helped a lot in developing my current database skills. Also I would like to thank all co-authors and collaborators within the EuroDEER network. In particular, I would like to thank Dr. Nicolas Morellet, not only for his support in my research, but also for his immediate, accurate and detailed responses on EuroDEER requests. He understands the importance of data curatorship and respects it, which makes data collection and its validation procedure truly a pleasant experience. Furthermore, I would like to thank collaborators and co-authors from FEM. In particular, Dr. Lino Ometto and Dr. Omar Stabelli-Rota, who provided great support in molecular biology and phylogeny, especially in the first years of my PhD. I would be grateful to continue working with them. During my PhD, I also had the opportunity to go for an exchange period to Harvard University. There I joined the research lab of Prof. dr. Paul Moorcroft for two months, whom I would like to thank for the irreplaceable opportunity to join a lab at one of the best universities of the world. I would also like to thank Dr. Emiel Van Loon, in particular for his interest in my research, his personal invitation to a workshop in Amsterdam, the Netherlands, and the very important opportunities that he is giving me.

Special thanks also to wildlife photographer Bruno Bressan, who provided the wonderful roe deer photographs on the front and back cover of the dissertation.

I would also like to thank all funding organs that supported my PhD, and related activities such as courses and meetings. Specifically, Special Research Fund (BOF, 13 months), research project of Prof. dr. Nico Van de Weghe and Prof. dr. Bernard De Baets (5 months), EuroDEER and Fondazione Edmund Mach (30 months), International Research School in Applied Ecology (IRSAE), Research Foundation – Flanders (FWO) and National Science Foundation (NSF), Netherlands Institute for Ecology (NIOO-KNAW) and the Lorentz Center for providing grants to pursue my scientific goals.

In conclusion, I would like to thank my family, friends, and my ex-band Celestial Wolves for their encouragement, trust, and the requisite distraction. Last but not least, I would like to thank my girlfriend, Cleopatra Leontidou, for her love, support, and for the wonderful times we have together.

CHAPTER 1

INTRODUCTION

1.1. BACKGROUND

Movement, defined as the change in spatial location of an entity over time (i.e. the trajectory), is one of the most fundamental processes for entities on earth. Detailed insights in movement patterns are crucial for many scientific disciplines such as ecology, sociology, geography and meteorology. For instance, studying movement has enhanced scientists' understanding of animal behaviour (Nathan 2008), traffic and crowd management (Batty *et al.* 2003, Verhein & Chawla 2006, Versichele *et al.* 2012, Çolak *et al.* 2016), patient monitoring (Shoval *et al.* 2008) and environmental hazard predictions (Dodge 2011).

Movement ecology is one of these exciting fields, focusing on why, how, when and where animals move, and in what way movement is linked to external factors (Nathan 2008). Knowledge in animal movement allows us to answer research questions that are essential to understanding important ecological and evolutionary processes that lay at the basis of survival and fitness of a species, i.e. the ability to increase or maintain population abundance in the next generation. With movement data we can better understand how animals use resources, habitats, and land use types within specific environmental contexts, or how animals interact with conspecifics or with other species (Cagnacci *et al*. 2010). Animal movement has always fascinated mankind: first scientific contributions to the field date back to the philosophers and naturalists of the Greek Classical Antiquity. Bird migrations were recorded by Hesiod, Homer, Herodotos and the founding father of zoology, Aristotle. The latter wrote the first comprehensive work *Περὶ ζῴων κινήσεως* (*On the Motion of Animals*) describing common features of various animal movements (Nathan 2008). While trail tracking and bird ringing (Jespersen 1950) have been around for centuries, only with the recent introduction of telemetry technologies field studies has animal ecology taken a big leap (LeMunyan *et al*. 1959).

Technological innovation in the last decades has led to a tremendous increase in the availability of movement data (Rutz & Hays 2009). In particular, remote tracking (e.g. GPS) and sensing (e.g. satellites) technologies have revolutionised movement ecology, among other fields of study. Advances in sensor technology have now allowed their implementation at unforeseen rates. Longer battery life has resulted in higher temporal resolution and longer monitoring periods, the accuracy of both tracking sensors and satellites have increased, a wider variety of species can be studied due to the miniaturisation of devices and development of new technologies (e.g. ICARUS, International Cooperation for Animal Research Using Space, https://icarusinitiative.org/), and multiple sensors can be installed in one device (e.g. gyroscopes, accelerometers, magnetometers, proximity loggers, physiological sensors, neurologgers) (Demšar *et al.* 2015a, Kays *et al.* 2015). Hence, current technologies not only allow the study of explicit movement of animals, but also multiple aspects of animals' internal states (e.g. heart-rate, body temperature) and their external environment (e.g. temperature, pressure, interactions, resources). The growing amount of multi-sensor tracking data is turning movement ecology into a field of big data (Kays *et al.* 2015), and hence collaborative initiatives for storage and management of such data have been developed. Some examples are Movebank (Wikelski & Kays 2014), ZoaTrack (Dwyer *et al.* 2015), Euroungulates (www.euroungulates.org), the Seabird Tracking Database (http://www.seabirdtracking.org/) and Ocean Tracking Network (OTN) (http://oceantrackingnetwork.org/, see Campbell *et al.* 2016 for a detailed list). Availability of such big data repositories provides new opportunities and challenges for the analysis of movement data. Such e-infrastructures allow easier access to more multi-species movement datasets and simultaneous comparison of various movement patterns across taxa (e.g. Tucker *et al.* 2017).

Among the many challenges, one of the hardest is the development of robust methods to explore, visualise and analyse bulks of movement data. In this context, interdisciplinary work between movement ecologists and GIScientists offers huge potential for establishing permanent networks of research (Shamoun-Baranes *et al.* 2011, Demšar & van Loon 2013, Demšar *et al.* 2015a, Kays *et al.* 2015). Compounds of movement data in different fields are very similar: spatial features (points, lines and polygons), movement parameters (e.g. speed, step length, turning angle), environmental context information and pattern similarity measures can typically be computed for any kind of moving object (animals, humans, hurricanes, icebergs, vehicles; Laube *et al.* 2007, Andrienko *et al.* 2008, Dodge *et al.* 2009).

Through this dissertation I promote innovation in animal movement ecology by presenting a novel methodology to study spatio-temporal sequential habitat use by animals within their geographic context, using interdisciplinary techniques. The seeds of this project trace back to 2013, through participation in the European Commission Funded COST Action – Knowledge Discovery from Moving Objects (MOVE), a European collaborative network focusing on knowledge extraction from large amounts of any kind of moving objects data. This dissertation intends to merge knowledge from movement ecology and GIScience to give us a better understanding of animal behaviour in space and time.

1.2. MOVEMENT AND ENVIRONMENTAL CONTEXT

1.2.1. Habitat use and selection

Understanding how animals move within their environmental context is one of the keystones of movement ecology (Manly *et al.* 2002, Nathan 2008). For example, habitat loss induced by climate change (e.g. desertification, acidification) may result in altitudinal or latitudinal range shifts of animal species (Parmesan *et al.* 2006, Wilson *et al.* 2007, Linares *et al.* 2015). Land use modifications, such as deforestation and urban development, may limit the amount of suitable habitats, lead to habitat fragmentation, hamper animal movements (see Tucker *et al.* 2018) and potentially result in population decline, or even local extinction of a species (e.g. Zemanova *et al.* 2017). Hence, to identify the impact of environmental changes on a species, analytical methods must identify and describe which specific habitats and resources are used, as well as where and when.

The habitat of a species is comprised by those places required by an animal to fulfill its resource needs, such as food, shelter, protection and mates for reproduction. Habitats can be described by a combination of biotic (e.g. predator presence, forest cover) and abiotic (e.g. temperature, soil, snow cover) characteristics of an environment. To understand habitat requirements of a species, ecologists perform habitat use and habitat selection studies (Lele *et al.* 2011). Habitat use can be defined as the actual use, whatever the availability, and habitat selection is the use disproportionate to the availability (Gaillard *et al.* 2010). Both habitat use and selection are recognised to occur at multiple scales in space and time (Morris *et al.* 1987, Orians *et al.* 1991). At the spatial scale, four orders of habitat selection are usually recognised (see the seminal paper: Johnson 1980). The first-order, corresponding to the distribution range (i.e. the geographical region in which the species can be found). The second-order, at which animals select their home range (i.e. the area an animal crosses during normal feeding and reproductive behaviour, excluding excursions (Burt, 1943)). The third-order, at which resources within the home range are selected, and the fourth-order, or the selection of specific items of a resource type (Johnson 1980). For example, considering our study species the European roe deer (*Capreolus capreolus*), the distribution of the species covers most of Europe and parts of Asia (level 1), they usually select home ranges that include a mosaic of forest and open habitats (level 2), they are a generally ecotonal species with a preference for edge habitats (level 3) and they are concentrate selectors, preferring specific nutritious plant types often found in edge habitat (level 4).

Moreover, at each spatial scale, habitats can reflect daily, seasonal or annual changing patterns, related to variations in trade-offs that lead to different behavioural decisions over time (Mysterud *et al.* 1999a, Godvik *et al.* 2009, Bremset Hansen *et al.* 2009, Bonnot *et al.* 2013). A herbivore in temperate regions may select more forest habitat during winter, while showing a daily alternation between forest and open habitats during summer, because of seasonal variation in trade-offs between food, cover and weather conditions. More precisely, in winter such open habitats are often avoided because of a thicker snow layer, no availability of food and cover and colder temperatures than found in the forest (Mysterud *et al.* 1997, Mysterud *et al.* 1999a, Mysterud & Østbye 2006, Ratikainen *et al.* 2007, Ewald *et al.* 2014, Ossi *et al.* 2015). Instead, in the summer, open habitats are attractive due to the availability of highly nutritious plants, but are mainly accessed at night in order to avoid encounters with humans during daytime and to avoid predation at night (Mysterud *et al.* 1999b, Hewison *et al.* 2001, Bjørneraas *et al.* 2011, Meisingset *et al.* 2013, Padié *et al.* 2015, Dupke *et al.* 2017, Coppes *et al.* 2017, Gehr *et al.* 2018). Note however that in habitat use, the fact that an individual allocates time to find both food and cover does not necessarily imply they trade cover for food acquisition. If food is limited, a positive correlation between food quality or abundance and risk might exist over the landscape because resources at the best feeding sites (secure and food rich) should be depleted first (Mysterud & Ims 1998, Bonnot *et al.* 2013, Gaudry *et al.* 2018). This situation may lead to a trade-off between food availability and risk avoidance as food-rich sites may be associated to higher mortality risk, as is often hypothesised for populations ranging in the European fragmented landscape (Mysterud & Ims 1998, Mysterud 1999).

1.2.2. Methods in ecology

In the field of ecology many exploratory and inferential methods have been developed to investigate the relationship between animal movement and their environment. Exploratory methods use multivariate analysis techniques (e.g. General Niche-Environment System Factor Analysis, GNESFA) or decision tree learning methods (e.g. random forest), to identify relevant variables describing the habitat, or the realised niche[1], used by a population (see Chapter 3). Inferential methods mainly use logistic regression models testing the disproportion between used and available habitat features (i.e. habitat selection), and are generally referred to as Resource Selection Functions, RSFs (Boyce *et al.* 2002). Depending on what area is defined as used and as available, different spatial scales of selection can be accounted for (Boyce *et al.* 2002, Forester *et al.* 2009). For instance, second-order selection (i.e. selection of the home range) defines *habitat availability* at the landscape scale, assuming that all individuals in a population have access to the same area, while *habitat use* is defined at the individual home range scale (e.g. individual Minimum Convex Polygon, MCP, Forester *et al.* 2009). Step Selection Functions (SSFs) are an extension of RSFs, comparing habitat variables used at an observed step, with those at another random step, both generated from the same starting point (Fortin *et al.* 2005, Thurfjell *et al.* 2014). By using this approach, habitat availability is measured dynamically along the trajectory of an animal.

While different spatial scales of selection have been studied extensively for many species (e.g. Forester *et al.* 2007, Herfindal *et al.* 2009), analytical methods linking this to the time scale of selection have only been developed more recently (Mysterud *et al.* 1999b, Godvik *et al.* 2009, Rivrud *et al.* 2010, Dupke *et al.* 2017). Many studies on habitat use/selection account for the temporal component by using regression modelling approaches (e.g. Generalised Linear Models). With the increasing temporal resolution of

---

[1] A concept that is often confused with, but also related to habitat, is the niche (Whittaker *et al.* 1973; Rodriguez-Cabal *et al.* 2013). The niche concept is defined in several ways (Grinnell, 1917; Elton, 1927; Hutchinson, 1957; Bruno *et al.* 2003; Rodriguez-Cabal *et al.* 2013), but its most common definition is as a multi-dimensional space defined by a range of physical and biological variables' states in which a species can persist (Hutchinson, 1957). The fundamental niche are the abiotic factors of an environment a species would occupy in case of no competition, while the realised niche is the part of the fundamental niche that is occupied in case of competition (Hutchinson, 1957). The inferential multivariate analysis techniques that describe the relation between animal movement and environment (see in-text citations) are mainly based on this latter definition.

tracking data, the temporal sampling units of such models have a higher frequency. For example, while in the past, annual changes in ecological covariates were averaged or aggregated at the seasonal scale, now many studies have incorporated measures at monthly (Morellet *et al.* 2013, Dupke *et al.* 2017), weekly (Morellet *et al.* 2013, Mancinelli *et al.* 2015, Kämmerle *et al.* 2017, Ossi *et al.* 2017) and even daily scales (Rivrud *et al.* 2010; Pagon *et al.* 2013). Circadian or seasonal patterns can be investigated by including the time of the day or day length, respectively, in the models (Pagon *et al.* 2013; Kämmerle *et al.* 2017), for example by using splines (Generalised Additive Models, GAM or GAMM; Zúñiga *et al.* 2016). For an RSF/SSF, accounting for the temporal component would mean that instead of computing one selection function (SFs), multiple SFs are computed for each time unit, which are then used as input in a later model.

While traditional ecological habitat studies mainly relied on statistical methods, less attention was given to geovisualisation, which refers to the visualisation of movement data in relation to the environmental context, disregarding its strength for data exploration to build ecological hypothesis and to investigate visually complex spatio-temporal dynamic dependencies intrinsic to animal movement data (Demšar *et al.* 2015a). While statistical models are very well suited to test hypothesis and quantify results, combining that with visual outputs, could result in better conceptual understanding of spatio-temporal dynamics underlying habitat use/selection. We are gradually approaching a shift now, with more and more available tools for dynamic exploration and sound statistical frameworks for studying animal movement in relation with their environment, which will be discussed in more detail in the next paragraph.

1.2.3. Methods in GIScience

Recent growth in data availability has spurred geo-informaticians' interest in animal movement data. Since habitats are used and selected at multiple spatio-temporal scales, robust exploration of environmental features at multiple scales are of fundamental importance (Demšar & van Loon 2013). Developing these exploratory tools requires high computational skills and a long data preparation process that are often outside the range of interest of ecologists (Slingsby & van Loon 2016). Instead, ecologists are especially interested in answering ecological questions through hypothesis testing, using the best available tools (Xavier & Dodge 2014). Joint research between GIScientists and ecologists has lead to several fruitful geovisualisation tools and methodological frameworks to study movement in relation to the environmental context (Demšar *et al.* 2015, see figure 1.1 for some examples which are discussed more in detail below).

Several exciting interactive tools have been developed for exploratory visual analysis of animal movement trajectories in relation with the environment and other miscellaneous data (Andrienko *et al.* 2011, Spretke *et al.* 2011, Xavier & Dodge 2014, Slingsby & Van Loon 2016[2], Konzack *et al.* 2018). In general, these tools combine spatial aggregation (Andrienko *et al.* 2011, Slingsby & Van Loon 2016) or trajectory segmentation techniques (Spretke *et al.* 2011, Konzack *et al.* 2018), and use those as the units for plotting aggregated statistics about the movement or environmental context on maps. For example, for migratory birds, habitat features of stopovers are important, since they require the availability of enough resources to replenish energy deposits to continue migration (Wang & Moore 2005). By trajectory segmentation, stopovers and migrations can be differentiated, and statistics on habitat characteristics could be summarised on the map at each stopover location. A good example of the latter is the tool developed by Konzach *et al.* (2018, Fig.1.1a). While the current tool only allows implicit visual exploration of identified stopovers in relation to environmental context, a future extension will integrate explicit visualisation methods to summarise environmental variables as described above. Another interesting example is the roe deer case study of Andrienko *et al.* (2011, Fig.1.1b), which plots histograms of the aggregated hourly use of open habitats over the monitoring period for each spatial cluster corresponding to an individual roe deer. Such a geovisualisation gives a prompt insight in the individual variability in the use of open habitats between day and night. Dynamic Multivariate Visualization of Movement (DYNAMO, Fig.1.1c) instead is a tool for animating trajectories, which can be annotated by a variety of environmental variables, useful for hypothesis building, but also for outreach (Xavier & Dodge 2014). DYNAMO is exemplified using two case studies, one where tiger movement is animated on a elevation map, highlighting the relation between topography and movement speed, and one where albatross movement is animated and annotated using movement speed and wind speed (Fig.1.1c; see also https://somayehdodge.info/projects/).

---

[2] Link to program: http://gicentre.org/birdGPS/

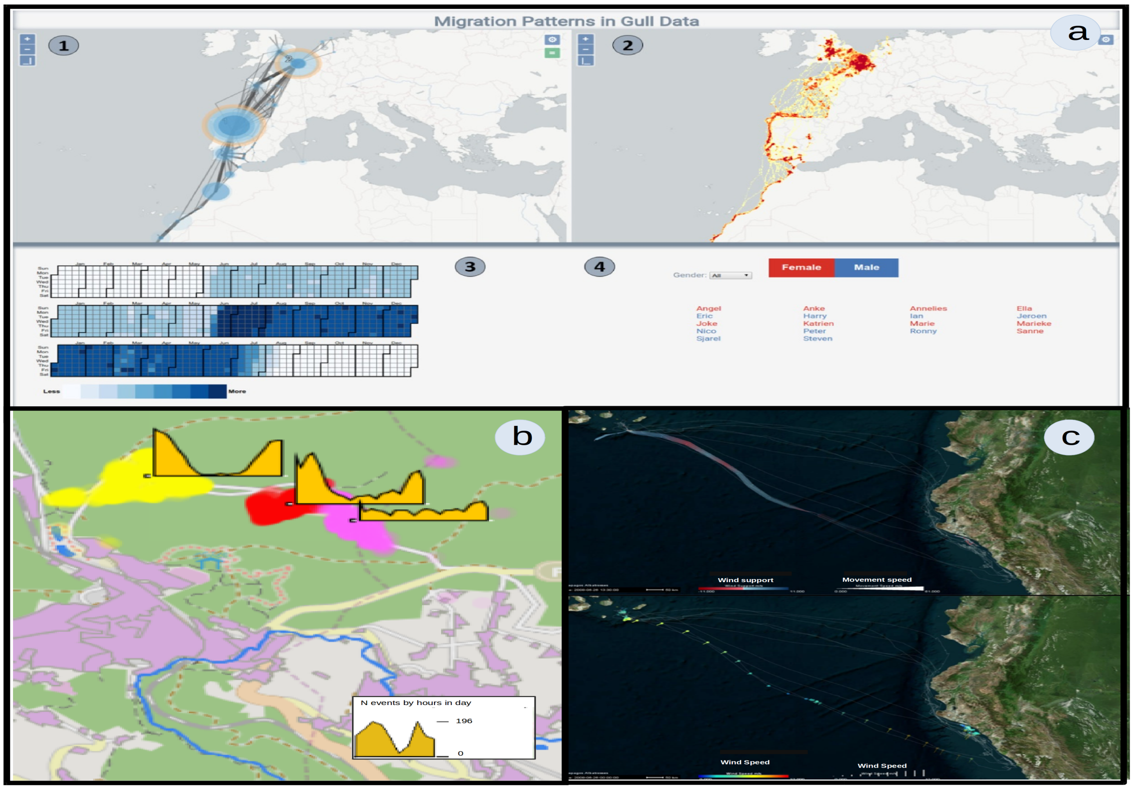


Figure 1.1. Extracts of interactive mapping tools to explore animal movement. (a) Stopover visualisation tool of Konzach *et al.* (2018) including stopover aggregation visualisation (1, upper-left) and density map of gull GPS trajectories (2, upper-right), and a calender view (3, lower-left) showing the temporal distribution of stopovers for the selected list of animals (4, lower-right). (b) Daily use of open habitat for three spatial clusters of roe deer GPS locations using the event-based interactive tool V-Analytics developed by Andriendo *et al.* (2011). (c) Extract of animated Albatros track annotated by wind support and movement speed (upper) and wind speed (lower) using the tool DYNAMO (Xavier & Dodge 2014).

Several methods have been developed that take into account the environment for computation of home ranges and utility distributions (UD). For example, Demšar *et al.* (2015b) proposed a 3D-visualisation of a home range, where x,y-plane is space, and the z-axis is time, and aggregated information of used habitats characterise the space-time cube. Long (2016) presents a model that combined classical time geography and least-cost path analysis through landscapes with specific environmental characteristics (e.g. dense forest, topography) to generate home ranges and UDs (see also R package wildlifeTG, http://jedalong.github.io/wildlifeTG).

Also for measuring trajectory similarity and trajectory segmentation several methods have been developed that account for environmental context. Toor *et al.* (2016) developed a trajectory segmentation algorithm based on time changes in habitat use using random forest models. Buchin *et al.* (2012, 2014) developed a trajectory similarity algorithm that takes into account both movement and environmental context, but has been developed using hurricane movement data. Another original contribution is the procedure developed by De Groeve *et al.* (2016), discussed in Chapter 3 of this thesis, which measures similarity directly based on the sequence of used habitats along the trajectory of an animal.

GIScientists, together with ecologists, further contributed in the development of automatic trajectory annotation software such as the R package RNCEP (Kemp *et al.* 2012) and Environmental-Data Automated Track Annotation (Env-DATA) integrated in the data-sharing database movebank (www.movebank.org; Dodge *et al.* 2013, Bohrer *et al.* 2015). Such developments have facilitated, and improved accessibility of large-scaled dynamic environmental datasets. Although the many examples described above, the link between movement and environmental context in GIScience is still considered as largely ignored (Dodge 2016, Holloway & Miller 2017).

1.2.4. Sequential habitat use

The trajectory is the basic analytical unit in movement studies. However, mostly it is not the trajectory *an sich*, but the underlying information which can be extracted from a trajectory that is of relevance, such as movement parameters, or any measure of the internal or external state of an animal (Dodge *et al.* 2009; Nathan 2008). For instance, by linking activity states inferred from accelerometer data with GPS locations we can understand where along a trajectory certain behaviours (i.e. internal state: feeding, resting, running, flying) are observed (Shamoun-Baranes *et al.* 2012, Chimienti *et al.* 2016), or, by relating meteorological data to bird trajectories it has been discovered that changing weather conditions (i.e. external state) may trigger long-distance migration movements (Shamoun-Baranes *et al.* 2010).

While fine-scaled spatio-temporal data availability has considerably improved our ability to study animal movement in relation to habitats, current methods rarely account for its temporally-dynamic nature, and in particular the *sequential use of habitats*. Sequential use is defined as the sequence of locations (the trajectory) measured against the underlying ordered pattern of habitat use. Not accounting for sequential use, only gives insights in the proportion of use, while the same proportion could correspond to various biologically important different behaviours.

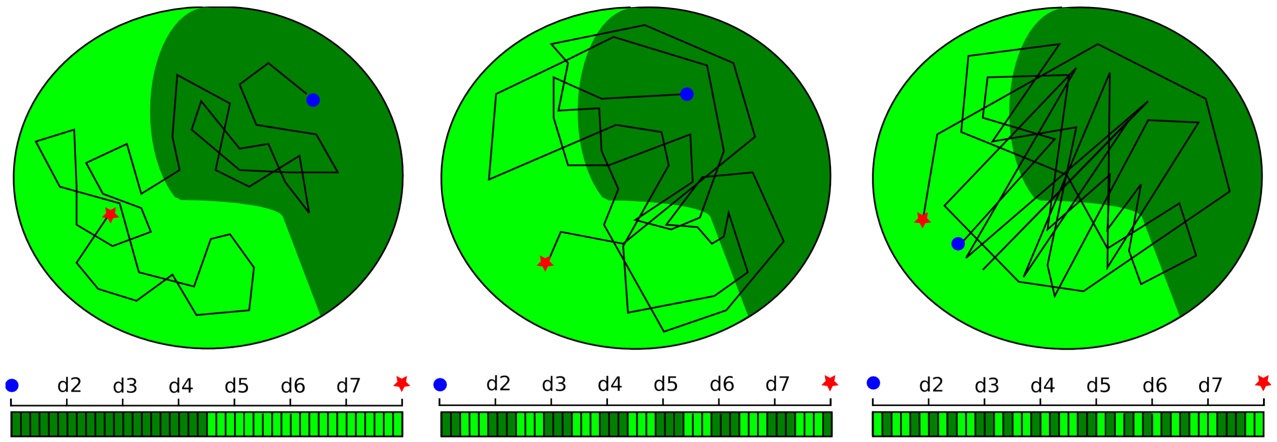


Figure 1.2. Concept of sequential habitat use. Blue dot and red star indicate the start and end of a weekly trajectory with a fix resolution of four hours (six locations a day) within a home range (circle; dark green, closed habitat; light green, open habitat). Below the home range, we represent the corresponding sequential habitat use patterns of the trajectory.

We present the concept of sequential habitat use in figure 1.2 for two distinct classes, open and closed habitats, and a non-exhaustive number of sequential patterns (i.e. temporal clustered use, day-night alternating use or random use). We exemplify with these two classes because those were the main habitat classes that were studied throughout this dissertation.

Temporal clustered use (Fig.1.2, left) in this example indicates a switch in habitat use over a longer time period. The change in the use could for example be the result of sub-seasonal functional home ranges (Couriot *et al.* 2018) or could reflect different summer and winter ranges (Cagnacci *et al.* 2011). Although closed and open habitats are available to the animal, it chooses in a certain period only to use open habitats, and in another, only closed. Such a change could be related to changes in resource availability in open and closed habitats or a change in the environment. For example, for temperate ungulates, sudden snow could result in a change from only use of open to only use of closed habitat (Mysterud *et al.* 1997, Mysterud *et al.* 1999a, Ossi *et al.* 2015).

Daily habitat alternation (Fig.1.2, middle), with mainly use of open habitats during night and closed habitats during day, reflect a clear switch in behaviour that can be related to trade-offs in availability of food and cover in open and closed habitats (see above). Many animals, including humans, indeed show clear repeated patterns in behaviour. For example, many herbivores increase their activity during dusk and dawn, while day vs night activity seems to vary throughout the seasons (Cederlund 1989, Krop-Benesch *et al.*

2012, Pagon *et al.* 2013, Ensing *et al.* 2014, Kämmerle *et al.* 2017). Further, herbivores show rumination cycles, where the use of cover is higher during rumination than during feeding bouts (Cederlund, 1981). Animals may also repeatedly visit the same resource patches (Fagan *et al.* 2013, Bracis *et al.* 2018). Such repeated patterns in the use of certain patches may moreover be related to memory components, and thus spatial and temporal awareness of an animal in the availability of specific resources (e.g. Fagan *et al.* 2013). Indeed, since recently many studies focus on quantification and modelling of recursive use of dynamic resource patches (Berger-Tal & Bar-David 2015, Bracis *et al.* 2018, see discussion Chapter 5).

Finally, random sequential patterns (Fig.1.2, right) instead, is a pattern that reflects the spatial structure of an area, and show a similar importance for different habitat types at specific periods of the day or the season. In such cases, it might be that the habitat types, in our case open and closed habitats, contain similar resources.

In addition, the emerging sequential habitat use patterns also depend from the spatial context in which animals are moving. For example, the strength of alternation might be stronger in a home range with equal availability of open and forest habitats, but could as well be observed in home ranges where habitats show disproportional availability. Here we attempt to provide a framework for *exploration*, *simulation* and *classification* of sequential habitat use patterns which also takes into account the individuals' geographic context. To do this we use string matching algorithms (see 1.2.5. Sequence dissimilarity analysis) in combination with a spatially-explicit movement modelling approach (see 1.2.6. Movement model) to identify specific sequential patterns of use (Chapter 3 and 4).

1.2.5. Sequence dissimilarity analysis

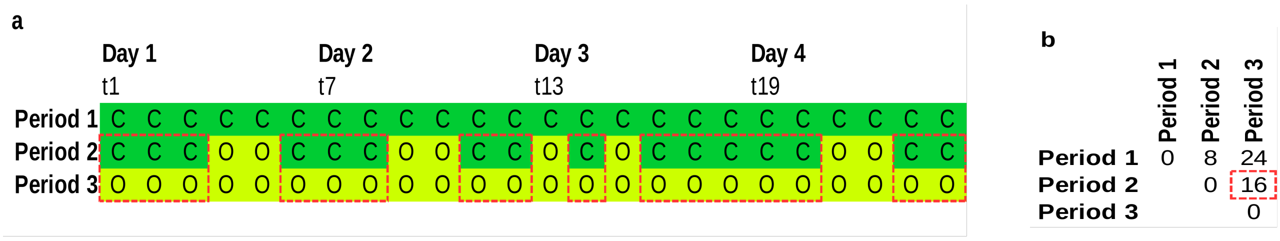
One way to extract knowledge from movement trajectories is by measuring their (dis)similarity. Measuring dissimilarity between trajectories has a long tradition and many different algorithms are available. An overview on the complexity of available dissimilarity measures can be found in several publications (e.g. Dodge 2011, Magdy *et al.* 2015). Here, we are especially interested in string matching algorithms, which use as input *sequences of characters*. The most important fields with a tradition in the use of string matching algorithms are molecular biology (Sanger & Tuppy 1951, Needleman & Wunsch 1970), computer science (Hamming 1950, Wagner & Fischer 1974), linguistics (Prokić *et al.* 2009) and sociology (Abbott 1995), and in each field several different algorithms have been developed. Among them are the Hamming distance, Levenstein distance, bag distance (fast levenstein distance), dynamic time warping, N-grams (number of substrings of length n), longest common prefix and Jaro variants (Jaro algorithm, Jaro-Winkler, sorted Winkler, permuted Winkler) (see Recchia & Louwerse 2013, Studer & Ritschard 2016). There are many applications where string matching algorithms play a role including natural language processing, spam filters, search engines, plagiarism detection, bioinformatics, feature detection in digital images, forensics, cryptography, etc. (Soni *et al.* 2014).

A general characteristic of such dissimilarity measures is that they try to minimise the number of operations in order to equate those character sequences. Here we exemplify this with the Levenstein distance (or edit distance), also used to measure dissimilarity between DNA and protein sequences, which calculates sequence dissimilarity based on the number of deletions, insertions, or substitutions (i.e. operations) required to transform one sequence into another. Equating the words in the string “Dear Dr Deer” with the former distance metric will require a minimum of two operations (i.e. two insertions/deletions) to equate the word pairs “D**ea**r-Dr” and “Dr-D**ee**r”, and one operation (i.e. a substitution) to equate word pairs “De**e**r-De**a**r”, so the latter word pair is considered more similar. Importantly, algorithms are not always interchangeable between different fields of study. For instance, the Levenstein distance is under discussion in social sciences where time is often an important component (Studer & Ritschard 2016), since the operations insertion and deletion would warp time along the sequence. Also for sequences of animal habitat use, the temporal component is essential, and such time shifts should be avoided.

Following several applications in GIS-related fields (Shoval & Isaacson 2007, Wilson 2008, Delafontaine *et al.* 2012, Chavoshi *et al.* 2015) we decided to test the applicability for animal movement data. Since the procedure, referred to as Sequence Analysis Method (SAM), is described in detail in Chapter 3 and 4 we will only discuss the general principles in brief. In essence, SAM measures dissimilarity between each pair of character sequences using one of the many dissimilarity algorithms. Calculated dissimilarities are then stored in a dissimilarity matrix. While other time dependent algorithms are available, such as dynamic time warping (Giorgino 2009), we have developed our approach using the Hamming distance (HD), which was the first developed and the simplest distance measure available (Hamming 1950). Since this is the first application on ecological data we decided that it is better to use a robust and easy-to-follow distance measure. Specifically, the HD counts the number of mismatches between sequences of equal length (Gabadinho *et al.*

2011, Fig.1.3). The more mismatches, the more dissimilar two sequences are from each other. The Hamming distance can also include substitution weights for classes that are more similar to each other, especially relevant when input classes are more similar to each other than others (see Chapter 3). Next, using a cluster algorithm (e.g. Hierarchical cluster algorithm of Ward) computed dissimilarities can be visually represented by a dissimilarity tree. The leaves in such a tree are the sequences, and nodes represent clusters (groups of sequences) for which the branch lengths measure the degree of dissimilarity. Consequently, the longer a branch the larger the dissimilarity. In a next step the robustness of the classification can be calculated, which allows to identify the most relevant number of clusters in a tree. Finally, after identifying the optimal number of clusters, the cluster types need to be identified, based on their specific characteristics.

While the further chapters will show SAMs potential for analysing temporal habitat use sequences, it should be noted that there are also other temporal explicit analytical approaches. For instance, even though differing in many aspects, time-to-event and discrete event models (Haccou & Meelis, 1992) that are widely applied in survival analysis, often have similar underlying questions as those dealt with in this dissertation.



a

	Day 1						**Day 2**						**Day 3**						**Day 4**					
	t1						t7						t13						t19					
Period 1	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
Period 2	C	C	C	O	O	C	C	C	O	O	C	C	O	C	O	C	C	C	C	C	O	O	C	C
Period 3	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O

b

	Period 1	**Period 2**	**Period 3**
Period 1	0	8	24
Period 2		0	16
Period 3			0

Figure 1.3. Example of Hamming distance for three four-days open (O)-closed (C) habitat use sequences (a), corresponding to three different periods (period 1, 2, 3). The red boxes show the mismatches between period 2 and 3 (dotted line). Dissimilarities of each pair of sequences is represented in the dissimilarity matrix (b).

1.2.6. Movement models

Mechanistic Movement models have a long tradition in the field of ecology to interpret animal movement data (Gurarie 2008). The main purpose of such movement models is to provide mathematical formulations of movement based ecological processes (e.g. home ranging, migration, foraging). Movement of living beings is very complex, with trajectories being at random, linear, circular, tortuous, directional, or occurring in a combination of the previous (see Gurarie *et al.* 2017), and moreover, the general characteristics of movement parameters vary over spatial and temporal scales (Fryxell *et al.* 2008, Gurarie & Ovaskainen 2011). For instance, by subsampling a location dataset, Fryxell *et al.* (2008) showed that different ecological processes emerged at different spatial and temporal scales, with movement characterising home ranging vs. dispersal and migration at the largest scale, feeding and resting cycles at the intermediate scale, and fine responses to direct stimuli at the finest scale.

Hence, to capture such complexity a burgeoning number of deterministic models have been proposed and developed to evaluate and predict movement, including fractal analysis (Mandelbrot, 1967; Jander 1982, Turchin 1996, Webb *et al.* 2009) and a large number of random walk based models (e.g. Patlak *et al.* 1953, Kareiva & Shigesada 1983, Dunn & Brown 1987, Alt 1990, Marsh & Jones 1988, Klafter *et al.* 1996, Codling *et al.* 2008). Most mechanistic movement models are based on the random walk (RW, McClintock *et al.* 2014), which in most general terms simulates trajectories by drawing random values from empirical or theoretical distributions of movement parameters such as the step length and the turning angle (Hollaway & Miller 2018). Correlated random walks (CRW) are *correlated* because any given step will tend to be in the same direction as the previous. CRW are characterised by *forward persistence* in the movement achieved by drawing at random turning angles from a unimodal distribution peaking around zero (Patlak *et al.* 1953, Kareiva & Shigesada 1983). Biased random walks (BRW) are *biased* because any given step will orient towards a preferred direction. BRW are characterised by *directional bias* in the movement achieved by drawing at random step orientations from a distribution centered on a preferred direction (Marsh & Jones 1988, Benhamou 2006). Biased and correlated random walks (BCRW) combine both directional bias and forward persistence (Codling *et al.* 2008). Lévy walks (Klafter *et al.* 1996) are uncorrelated random walk where the step length follows a fat-tailed distribution, resulting in random walk-like movements with occasional long steps (Gurarie 2008). Further, mechanistic movement models based on random walk are generalisable into two main families, discrete and continuous models. In discrete models, distributions (of steps and turning angles) are generated for certain time units to model a trajectory, while in continuous-time or correlated velocity models (CVM), diffusion models that account for autocorrelation between locations are used, and this is referred to as the Ornstein-Uhlenbeck process (OU; Gurarie *et al.* 2017). The OU process is a continuous-time random walk with a tendency to drift towards a central location (McClintock *et al.* 2014, Gurarie *et al.* 2017). Since aforementioned CVM were not used here, we will not further discuss them.

In this dissertation we performed simulations of realistic movement trajectories accounting for spatial context to generate expected sequential habitat use patterns. The movement model needed to be defined in such a way that roe deer's movement behaviour was realistically simulated at two spatio-temporal scales. First, the model should capture the typical home ranging behaviour of roe deer (i.e. a spatially-constrained movement). Second, the model should include the species' tendency of using habitats in a specific manner on a daily basis (i.e. see *supra*: homogeneous, daily alternation, random). To achieve this goal we could rely on the literature of memory-based movement models (Van Moorter *et al.* 2009, Fagan *et al.* 2013). In specific, we used a stochastic BCRW model, where the bias is towards previously visited habitat classes (Börger *et al.* 2008, Van Moorter *et al.* 2009). For homogeneous closed and open patterns, the directional bias is kept constant for the respective habitat classes, both during day and night. For alternating patterns, directional bias varies through time, with main attraction towards closed habitats during day and open habitats during night. The random pattern is our null model, showing no attraction for open or closed habitats. Hence, the random trajectories will result in a sequential habitat use pattern that reflects the availability and distribution of habitats within the movement context of the animal, given the specific movement model. In the (IM-)SAM-procedure, such random patterns will be identified as a separate cluster, when habitats are more equally present, or group together with the most similar simulated sequential patterns (homogeneous closed/open, alternation). As such, it is possible to understand whether a simulated sequential pattern occurs at random or not within a given landscape, for instance the day-night alternation between open and closed habitats.

1.3. STUDY SPECIES

European roe deer (*Capreolus capreolus*) is one of the most common and best studied ungulate species in Europe. The number of mature animals is estimated to be 15,000,000, and at present their distribution range extends almost the whole Europe, until central Asia (Fig.1.4, Lovari *et al.* 2016). Due to its relatively small size (18-49 kg, 60-75 cm at the shoulder) roe deer is an important prey species and is part of the diet of many carnivores, including grey wolf (*Canis lupus lupus*), Eurasian lynx (*Lynx lynx*), red fox (*Vulpes vulpes*), Eurasian brown bear (*Ursus arctos arctos*) and free-ranging dogs (*Canis familiaris*) (Aanes *et al.* 1998, Lovari *et al.* 2016). Roe deer is also an important and highly managed game species in most parts of Europe (Cederlund *et al.* 1998).

Figure 1.4. Distribution of european roe deer.

This species, described as a generalist at species, population, and at individual level, can adapt to a wide range of natural and human habitats, including deciduous, mixed and coniferous forests, shrubland, moorland, pastures, arable land and suburban gardens (Linnell *et al.* 1998). They thrive in landscapes with a mix of forest and agricultural land (Aulak & Babińska-Werka 1990) including high amounts of ecotonal habitats (i.e. edges, Tufto *et al.* 1996). Roe deer are considered as native forest dwellers that adjusted to the European landscape by colonising human-dominated open habitats (Linnell *et al.* 1998). Such behavioural plasticity, and thus the species' ability to cope with changing environmental conditions, is manifested in different habitats throughout their distributional range (Hewison *et al.* 1998, Hewison *et al.* 2001). In essence, roe deer's expansion to open habitats is thus considered as a plastic response of the species to human induced landscape changes. Often a distinction is now made between two ecotypes, forest and field roe deer (Fruziński *et al.* 1982, San José *et al.* 1997), because of fundamental behavioural differences including diet composition (Abbas *et al.* 2011), group size (Jepsen & Topping 2004) and space use strategy (Tufto *et al.* 1996). These two ecotypes, may even exist next to each other within the same study area (e.g. Hewison *et al.* 2001).

Figure 1.5. European roe deer (*Capreolus capreolus*). Illustration from Spagnesi *et al.* (1999).

Specifically, roe deer maintain a concentrate selectors feeding strategy, typically for small ungulates, but their diet includes a wide range of species including fungi, lichens, mosses, ferns, coniferous and deciduous tree leaves, shrubs, forbs, cultivated plants, grasses and sedges (Duncan *et al.* 1998). Their diet mainly varies in relation to the available habitat types (Abbas *et al.* 2011). Specifically, with increasing availability of fields in the landscape, roe deer eat more seeds, roots and fruits, while leaves are the dominant food resource for forest roe deer (Duncan *et al.* 1998). Seasonal variations in diet composition also exist, but are less marked (Cornelis *et al.* 1999, Abbas *et al.* 2011). Instead, feeding behaviour, or budget time in animal activities, may vary extensively throughout the seasons, showing as much as double feeding-resting bouts in summer compared to winter period (resp. 16 vs 8 bouts per day, Cederlund *et al.* 1981). Roe deer is mainly crepuscular showing highest activities during dusk and dawn (Pagon *et al.* 2013, Krop-Benesch *et al.* 2013), and mainly access nutritious food resources often found in more open habitats during night, probably to avoid human disturbance and predation (Mysterud *et al.* 1999b). During winter roe deer of both sexes more often group together when they are foraging in open habitats. In open landscapes group sizes are usually larger than those in forest landscapes (Liberg *et al.* 1998, Hewison *et al.* 2001, Jepsen & Topping 2004). Instead, during summer (mid July–mid August) individuals of both sexes are more often solitary, when females, or does, are taking care of their offspring, and males, or bucks, are rutting and defending their territories from conspecifics for mating (Liberg *et al.* 1998).

In general, roe deer are not territorial, except males in summer (Bramley 1970), during which they defend their resources (Vanpé *et al.* 2009) for one (or more) females, thus showing a low degree of polygyny (Vanpé *et al.* 2007). During the rut territorial males exhibit aggression, marking and patrolling of the territory (Sempéré *et al.* 1998). Since both sexes are mainly sedentary over their lifetime, access to partners is in general limited (Hewison *et al.* 1998). Moreover, females are also monoestrous, having only one cycle a year (Hoffmann *et al.* 1978), with an oestrous during the rut of only circa 36 hours (Sempéré *et al.* 1998, Debeffe *et al.* 2014). As shown by Debeffe *et al.* (2014) in six roe deer populations, females also perform breeding dispersal to non-contiguous male home ranges, a tactic that improves their chances to find a mate and that may limit the risk of inbreeding. Although the rutting season ends in August, the implantation of the fertilised egg in the ovary only follows in January, and parturition dates, or fawn birth, is highly synchronised between populations, ranging between early-May and early-June, with females giving birth consistently earlier when aging (Plard *et al.* 2012).

The behavioural plasticity of the species is also reflected in the diverse space use strategies (i.e. common space use strategies are residency, nomadism, migration) and variation in home range size observed within and between populations. Specifically, annual home ranges can range from less than twenty to more than two hundred hectares (Mysterud 1999, Morellet *et al.* 2013), and partial migration is observed in several populations, where only a portion of the individuals migrate seasonally between distinct summer and winter ranges (Cagnacci *et al.* 2011). In some populations instead almost all animals are migratory or residents (Cagnacci *et al.* 2016). Hence, much variability has been observed in the occurrence, and characteristics of migratory behaviour across populations, and also across years within populations (Peters *et al.* 2017). Moreover, it has been recently shown that roe deer shift functional areas within their range (Couriot *et al.* 2018). High spatio-temporal variability in behaviour between and within populations makes roe deer a very interesting model species to study sequential habitat use.

1.4. PROBLEM STATEMENT

While concepts of habitat use are very well defined, the temporal and sequential component that is underlying the trajectory of animals is hardly accounted for. In general, the aim of this PhD project can be divided into two main components, a methodological, and ecological one, which are further discussed below in detail.

1.4.1. Methodological aim

The first aim of this PhD is to develop a new methodological framework, that merges knowledge from geography and ecology, to study sequential patterns in animal habitat use, and thus how animals use their habitats as an ordered temporal sequence along their trajectories. We aimed to develop a quantitative method that is also visually effective to study animal habitat use changes over multiple temporal scales. This method should be able to take into account issues typically encountered in movement data and habitat studies, such as missing data and spatial autocorrelation, availability and distribution of habitat features. Moreover, the method should be able to identify both population and individual level sequential patterns. The methodological part of this work consists out of two main parts. First, in Chapter 3 we developed the tree-based SAM approach using GPS movement data of a single roe deer population as a case study, based on sequential patterns of four discretised habitat variables (Open-High, Open-Low, Closed-High, Closed-Low). Second, in Chapter 4 we adjusted the methodological framework to an individual-based approach so that it can be applied over multiple populations in a variety of geographic contexts, and only focused on two habitat classes (open and closed habitats).

1.4.2. Ecological aim

The second aim of this PhD is to assess spatio-temporal patterns of habitat use of multiple roe deer populations, and to evaluate the ecological relevance of sequential habitat use in understanding the relation between roe deer movement, and the environment. In particular, using the developed methodology we aimed to further test the applicability and described sequential use of open and closed habitats within and between different roe deer populations ranging in different environmental contexts in Europe. In general, given the high adaptability of roe deer, we evaluated whether sequential patterns in the use of open and closed habitats vary within and between populations and throughout the year, and whether sequential patterns depend on the habitat structure and composition.

Specifically, because roe deer is a generalist species, that can occupy many different habitats (Linnell *et al.* 1998), we expected that sequential habitat use would differ between populations, reflecting the habitat composition and structure of an area. We expected the sequential habitat use patterns to deviate from random, since roe deer show a high adaptation to spatio-temporal distribution of resources (Peters *et al.* 2017). Given that roe deer show a preference for more open habitats during night time (Mysterud *et al.* 1999b), we expected day-night alternation between closed and open habitats to emerge as a common space use technique, and to be different from a random sequential pattern. Moreover, given seasonal variations in roe deer physiology (Liberg *et al.* 1998), rumination cycles (Cederlund *et al.* 1981) and in resource distribution in a given space (Mysterud *et al.* 1999, Dupke *et al.* 2017) we also expected that sequential patterns of use may vary across seasons. Results of the ecological study are presented in more detail in Chapter 5.

1.5. DISSERTATION STRUCTURE

The following chapters represent the research activities performed during this PhD. Chapter 2 is dedicated to my data resource, the Euroungulates database, to which I also contributed to considerably by importing, standardising, validating and managing datasets. Chapter 3 and 4 focus on the methodological aim (Chapter 3: published; Chapter 4: submitted), while Chapter 5 focuses on the ecological aim (close to submission). Each chapter is accompanied by several appendices (Chapter 3, five appendices; Chapter 4, seven appendices; Chapter 5, one appendix and Chapter 6, two appendices), which further clarify important data preparation and methodological steps, or represent figures supporting the results. These appendices were not integrated in the main body of the chapters to avoid the reader to be distracted from the main story-line.

A strength of this PhD is its collaborative nature, starting from the joint supervisorship, and the needed expertise drawing from diverse disciplines as GIScience and animal and quantitative ecology. Therefore, while (IM-)SAM methodology has been mainly developed by my own contribution, here I would like to stress that input of co-authors was fundamental to obtain the research outputs as here presented, as expected in a truly interdisciplinary scientific work. Movement models used in Chapter 3 and 4, in particular, were not developed by myself, but by experts in the field of mechanistic movement modelling, and specifically the PhD student Nathan Ranc, under guidance of his supervisors Prof. dr. Paul Moorcroft and Dr. Francesca Cagnacci. Indeed, during my PhD I obtained a scholarship to spend some visiting time at Prof. Moorcroft's lab in order to integrate the aforementioned models into my research work. Hence, I jointly wrote the movement model portions (Appendix 3B and Appendix 4C) with my collaborators.

REFERENCES

Aanes, R., Linnell, J.D.C., Perzanowski, K., Karlsen, J. and Odden, J., 1998. Roe deer as prey. *In:* Andersen, R., Duncan, P and Linnell, J.D.C., eds. *The European roe deer: the biology of success.* Oslo: Scandinavian University Press, 337-372.

Abbas, F., Morellet, N., Hewison, A.J.M., Merlet, J., Cargnelutti, B., Lourtet, B., Angibault, J.M., Daufresne, T., Aulagnier, S. and Verheyden, H., 2011. Landscape fragmentation generates spatial variation of diet composition and quality in a generalist herbivore. *Oecologia*, **167** (2), 401-411.

Abbott, A., 1995. Sequence analysis: new methods for old ideas. *Annual Review of Sociology*, **21**, 93-113.

Alt, W., 1990. Correlation analysis of two-dimensional locomotion paths. *In:* Alt, W. and Hoffman, G., eds. *Biological Motion: Proceedings of a workshop helf in Königswinter Germany. Lecture Notes in Biomathematics*, Berlin, Germany: Springer-Verlag. 254-268.

Andrienko, N., Andrienko, G., Pelekis, N., Spaccapietra, S., 2008. Basic Concepts of Movement Data. *In:* Giannotti F. and Pedreschi D., eds. *Mobility, Data Mining and Privacy.* Berlin: Springer.

Andrienko, G., Andrienko, N. and Heurich, M., 2011. An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, **25** (9),1347-1370.

Andrienko, N. and Andrienko, G., 2013. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, **12** (1), 3-24.

Aulak, W. and Babińska-Werka, J., 1990. Use of agricultural habitats by roe deer inhabiting a small forest area. *Acta Theriologica*, **35** (1-2), 121-127.

Batty, M., DeSyllas, J. and Duxbury, E., 2003. The discrete dynamics of small-scale spatial events: agent-based models of mobility in carnivals and street parades. *International Journal of Geographical Information Science*, **17** (7), 673-697.

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

Berger-Tal, O. and Bar-David, S., 2015. Recursive movement patterns: review and synthesis across species. *Ecosphere*, **6** (9), 1-12.

Bjørneraas, K., Solberg, E.J., Herfindal, I., Moorter, B.V., Rolandsen, C.M., Tremblay, J.P., Skarpe, C., Sæther, B.E., Eriksen, R. and Astrup, R., 2011. Moose Alces alces habitat use at multiple temporal scales in a human-altered landscape. *Wildlife Biology*, **17** (1), 44-54.

Bohrer, G., Kays, R., Davidson, S.C., Wienzierl, R., Dodge, S., Mcclain, K.M. and Wikelski, M., 2015. Remote sensing in support of endangered species management and animal movement research–the Env-DATA tool pack. In *Proceedings of the 66th International Astronautical Congress, International Astronautical Federation*.

Bonnot, N., Morellet, N., Verheyden, H., Cargnelutti, B., Lourtet, B., Klein, F. and Hewison, A.J.M., 2013. Habitat use under predation risk: hunting, roads and human dwellings influence the spatial behaviour of roe deer. *European journal of wildlife research*, **59** (2), 185-193.

Börger, L., Dalziel, B.D. and Fryxell, J.M., 2008. Are there general mechanisms of animal home range behaviour? A review and prospects for future research. *Ecology letters*, **11** (6), 637-650.

Boyce, M.S., Vernier, P.R., Nielsen, S.E. and Schmiegelow, F.K., 2002. Evaluating resource selection functions. *Ecological modelling*, **157** (2-3), 281-300.

Bracis, C., Bildstein, K.L. and Mueller, T., 2018. Revisitation analysis uncovers spatio-temporal patterns in animal movement data. *Ecography*. **41**, 1-11.

Bramley, P.S., 1970. Territoriality and reproductive behaviour of roe deer. *Journal of reproduction and fertility. Supplement*, **11**.

Bremset Hansen, B., Herfindal, I., Aanes, R., Sæther, B.E. and Henriksen, S., 2009. Functional response in habitat selection and the tradeoffs between foraging niche components in a large herbivore. *Oikos*, **118** (6), 859-872.

Bruno, J.F., Stachowicz, J.J. and Bertness, M.D., 2003. Inclusion of facilitation into ecological theory. *Trends in Ecology & Evolution*, **18** (3), 119-125.

Buchin, M., Dodge, S. and Speckmann, B., 2012. Context-Aware Similarity of Trajectories. *In: Proceedings of the 6th International Conference on Geographic Information Science.* Berlin: Springer, 43-56.

Buchin, M., Dodge, S. and Speckmann, B., 2014. Similarity of trajectories taking into account geographic context. *Journal of Spatial Information Science*, **2014** (9), 101-124.

Burt, W.H., 1943. Territoriality and home range concepts as applied to mammals. *Journal of mammalogy*, **24** (3), 346-352.

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

Cagnacci, F., Focardi, S., Heurich, M., Stache, A., Hewison, A.J.M., Morellet, N., Kjellander, P., Linnell, J.D., Mysterud, A., Neteler, M. and Delucchi, L., 2011. Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors. *Oikos*, **120** (12), 1790-1802.

Cagnacci, F., Focardi, S., Ghisla, A., Van Moorter, B., Merrill, E.H., Gurarie, E., Heurich, M., Mysterud, A., Linnell, J., Panzacchi, M. and May, R., 2016. How many routes lead to migration? Comparison of methods to assess and characterize migratory movements. *Journal of Animal Ecology*, **85** (1), 54-68.

Campbell, H. A., Urbano, F., Davidson, S., Dettki, H., and Cagnacci, F. 2016. A plea for standards in reporting data collected by animal-borne electronic devices. *Animal Biotelemetry* **4 (**1). 1-4.

Cederlund, G., 1981. Daily and seasonal activity pattern of roe deer in a boreal habitat [Capreolus capreolus, Sweden]. *Swedish Wildlife Research,* **11** (8), 315-353.

Cederlund, G., 1989. Activity patterns in moose and roe deer in a north boreal forest. *Ecography*, **12** (1), 39-45.

Cederlund, G., Bergqvist, J., Kjellander, P., Gill, R., Gaillard, J.M., Duncan, P., Ballon, P. and Boisaubert, B., 1998. Managing roe deer and their impact on the environment: maximising benefits and minimising costs. *In:* Andersen, R., Duncan, P and Linnell, J.D.C., eds. *The European roe deer: the biology of success.* Oslo: Scandinavian University Press, 337-372.

Chapron, G., Kaczensky, P., Linnell, J.D., von Arx, M., Huber, D., Andrén, H., López-Bao, J.V., Adamec, M., Álvares, F., Anders, O. and Balčiauskas, L., 2014. Recovery of large carnivores in Europe's modern human-dominated landscapes. *Science*, **346** (6216), 1517-1519.

Chavoshi, S.H., De Baets, B., Neutens, T., De Tré, G. and Van de Weghe, N., 2015. Exploring dance movement data using sequence alignment methods. *PloS one*, **10** (7), p.e0132452.

Chimienti, M., Cornulier, T., Owen, E., Bolton, M., Davies, I.M., Travis, J.M. and Scott, B.E., 2016. The use of an unsupervised learning approach for characterizing latent behaviors in accelerometer data. *Ecology and evolution*, **6** (3), 727-741.

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

Çolak, S., Lima, A. and González, M.C., 2016. Understanding congested travel in urban areas. *Nature communications*, **7**, 10793.

Coppes, J., Burghardt, F., Hagen, R., Suchant, R. and Braunisch, V., 2017. Human recreation affects spatio-temporal habitat use patterns in red deer (Cervus elaphus). *PloS one*, **12** (5), p. e0175134.

Cornelis, J., Casaer, J. and Hermy, M., 1999. Impact of season, habitat and research techniques on diet composition of roe deer (Capreolus capreolus): a review. *Journal of Zoology*, **248** (2), 195-207.

Couriot, O., Hewison, A.J.M., Saïd, S., Cagnacci, F., Chamaillé-Jammes, S., Linnell, J.D., Mysterud, A., Peters, W., Urbano, F., Heurich, M. and Kjellander, P., Nicoloso, S., Berger, A., Sustr, P., Kroeschel, M., Soennichsen, L., Sandfort, R., Gehr, B. and Morellet, N., 2018. Truly sedentary? The multi-range tactic as a response to resource heterogeneity and unpredictability in a large herbivore. *Oecologia*, 1-14.

Debeffe, L., Focardi, S., Bonenfant, C., Hewison, A.M., Morellet, N., Vanpé, C., Heurich, M., Kjellander, P., Linnell, J.D., Mysterud, A. and Pellerin, M., 2014. A one night stand? Reproductive excursions of female roe deer as a breeding dispersal tactic. *Oecologia*, **176** (2), 431-443.

De Groeve, J., Van de Weghe, N., Ranc, N., Neutens, T., Ometto, L., Rota-Stabelli, O. and Cagnacci, F., 2016. Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods. *Methods in Ecology and Evolution*, **7** (3), 369-379.

Delafontaine, M., Versichele, M., Neutens, T. and Van de Weghe, N., 2012. Analysing spatiotemporal sequences in Bluetooth tracking data. *Applied Geography*, **34**, 659-668.

Demšar, U. and van Loon, E., 2013. Visualising movement: The seagull. *Significance*, **10** (5), 40-42.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., Weiskopf, D. and Weibel, R., 2015a. Analysis and visualisation of movement: an interdisciplinary review. *Movement ecology*, **3** (1), 5.

Demšar, U., Buchin, K., van Loon, E.E. and Shamoun-Baranes, J., 2015b. Stacked space-time densities: a geovisualisation approach to explore dynamics of space use over time. *GeoInformatica*, **19** (1), 85-115.

Dodge, S., Weibel, R. and Forootan, E., 2009. Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, **33** (6), 419-434.

Dodge, S., 2011. *Exploring Movement Using Similarity Analysis*. PhD thesis. University of Zurich.

Dodge, S., Bohrer, G., Weinzierl, R., Davidson, S.C., Kays, R., Douglas, D., Cruz, S., Han, J., Brandes, D. and Wikelski, M., 2013. The environmental-data automated track annotation (Env-DATA) system: linking animal tracks with environmental data. *Movement Ecology*, **1** (3), 1-14.

Dodge, S., 2016. From observation to prediction: the trajectory of movement research in GIScience. *In:* Onsrud, H., Kuhn, W., eds. *Advancing geographic information science: the past and next twenty years*. Needham: GSDI Association Press, 123-135.

Duncan, P., Tixier, H., Hoffman, R.R. and Lechner-Doll, M., 1998. The European roe deer: a portrait of a successful species. *In:* Andersen, R., Duncan, P and Linnell, J.D.C., eds. *The European roe deer: the biology of success.* Oslo: Scandinavian University Press, 91-116.

Dunn, G.A. and Brown, A.F., 1987. A unified approach to analysing cell motility. *Journal of Cell Science*, **1987** (8), 81-102.

Dupke, C., Bonenfant, C., Reineking, B., Hable, R., Zeppenfeld, T., Ewald, M. and Heurich, M., 2017. Habitat selection by a large herbivore at multiple spatial and temporal scales is primarily governed by food resources. *Ecography*, **40** (8), 1014-1027.

Dwyer, R.G., Brooking, C., Brimblecombe, W., Campbell, H.A., Hunter, J., Watts, M., and Franklin, C.E., 2015. An open web-based system for the analysis and sharing and analysis of animal tracking data. *Animal Biotelemetry* **3** (1), 1-11.

Elton, C.S., 1927. *Animal ecology.* London, UK: Sidgwick and Jackson.

Ensing, E.P., Ciuti, S., de Wijs, F.A., Lentferink, D.H., ten Hoedt, A., Boyce, M.S. and Hut, R.A., 2014. GPS based daily activity patterns in European red deer and North American elk (Cervus elaphus): indication for a weak circadian clock in ungulates. *PLoS One*, **9** (9), e106997.

Ewald, M., Dupke, C., Heurich, M., Müller, J. and Reineking, B. 2014. The application of LiDAR remote sensing and activity sensors to analyze winter habitat selection of European roe deer. *Forestry,* **2014** (5), 1374-1390.

Fagan, W.F., Lewis, M.A., Auger-Méthé, M., Avgar, T., Benhamou, S., Breed, G., LaDage, L., Schlägel, U.E., Tang, W.W., Papastamatiou, Y.P. and Forester, J., 2013. Spatial memory and animal movement. *Ecology letters*, **16** (10), 1316-1329.

Forester, J.D., Ives, A.R., Turner, M.G., Anderson, D.P., Fortin, D., Beyer, H.L., Smith, D.W. and Boyce, M.S., 2007. State–space models link elk movement patterns to landscape characteristics in Yellowstone National Park. *Ecological Monographs*, **77** (2), 285-299.

Forester, J.D., Im, H.K. and Rathouz, P.J., 2009. Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. *Ecology*, **90** (12), 3554-3565.

Fortin, D., Hawthorne, L.B., Boyce, M.S., Smith, D.W., Duchesne, T. and Mao, J.S. 2005. Wolves influence elk movements: behaviour shapes a trophic cascade in Yellowstone national park. *Ecology,* **86**, 1320-1330.

Fruziński, B., Kałuziński, J. and Baksalary, J., 1982. Weight and body measurements of forest and field roe deer. *Acta theriologica*, **27** (33), 479-488.

Fryxell, J.M., Hazell, M., Börger, L., Dalziel, B.D., Haydon, D.T., Morales, J.M., McIntosh, T. and Rosatte, R.C., 2008. Multiple movement modes by large herbivores at multiple spatiotemporal scales. *Proceedings of the National Academy of Sciences*, 0801737105.

Gabadinho, A., Ritschard, G., Mueller, N.S. and Studer, M., 2011. Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, **40** (4), 1-37.

Gaillard, J.M., Hebblewhite, M., Loison, A., Fuller, M., Powell, R., Basille, M. and Van Moorter, B., 2010. Habitat–performance relationships: finding the right metric at a given spatial scale. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365** (1550), 2255-2265.

Gaudry, W., Gaillard, J.-M., Said, S., Bonenfant, C., Mysterud, A., Morellet, N., Pellerin, M. and Calenge, C., 2018. Same habitat types but different use: evidence of context-dependent habitat selection in roe deer across populations. *Scientific reports*, **8** (1), 5102.

Giorgino, T., 2009. Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software*, **31** (7), 1-24.

Gehr, B., Hofer, E.J., Ryser, A., Vimercati, E., Vogt, K. and Keller, L.F., 2018. Evidence for nonconsumptive effects from a large predator in an ungulate prey? *Behavioral Ecology,* 1-12

Godvik, I.M.R., Loe, L.E., Vik, J.O., Veiberg, V., Langvatn, R. and Mysterud, A., 2009. Temporal scales, trade-offs, and functional responses in red deer habitat selection. *Ecology*, **90** (3), 699-710.

Grinnell, J., 1917. The niche-relationships of the California Thrasher. *The Auk*, **34** (4), 427-433.

Gurarie, E. 2008. Models and analysis of animal movements: From individual tracks to mass dispersal. University of Washington.

Gurarie, E. and Ovaskainen, O., 2011. Characteristic spatial and temporal scales unify models of animal movement. *The American Naturalist*, **178** (1), 113-123.

Gurarie, E., Fleming, C.H., Fagan, W.F., Laidre, K.L., Hernández-Pliego, J. and Ovaskainen, O., 2017. Correlated velocity models as a fundamental unit of animal movement: synthesis and applications. *Movement ecology*, **5** (1), 13.

Haccou, P. and Meelis, E., 1992. *Statistical analysis of behavioural data: an approach based on time-structured models*. Oxford, UK: Oxford University Press.

Hamming, R.W., 1950. Error detecting and error correcting codes. *Bell Labs Technical Journal*, **29** (2), 147-160.

Herfindal, I., Tremblay, J.P., Hansen, B.B., Solberg, E.J., Heim, M. and Sæther, B.E., 2009. Scale dependency and functional response in moose habitat selection. *Ecography*, **32** (5), 849-859.

Hewison, A.J.M., Vincent, J.P. and Reby, D., 1998. Social organisation of European roe deer. *In:* Andersen, R., Duncan, P. and Linnell, J.D.C., eds. *The European roe deer: the biology of success.* Oslo: Scandinavian University Press, 189-219.

Hewison, A.J.M., Vincent, J.P., Joachim, J., Angibault, J.M., Cargnelutti, B. and Cibien, C., 2001. The effects of woodland fragmentation and human activity on roe deer distribution in agricultural landscapes. *Canadian journal of zoology*, **79** (4), 679-689.

Hoffmann, B., Barth, D. and Karg, H., 1978. Progesterone and estrogen levels in peripheral plasma of the pregnant and nonpregnant roe deer (Capreolus capreolus). *Biology of reproduction*, **19** (5), 931-935.

Holloway, P. and Miller, J.A., 2017. Analysis and Modeling of Movement, *In*: Huang, B., *Comprehensive Geographic Information Systems.* Amsterdam: Elsevier, 177.

Hutchinson, G.E., 1957. Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*. **22**: 415-427.

Jander, R., 1982. Random and systematic search in foraging insects. *In:* Breed, M.D., Michener, C.D. and Evans, H.E. eds. *The biology of social insects*. Boulder, Colorado, USA: Westview Press, 28-31.

Jepsen, J.U. and Topping, C.J., 2004. Modelling roe deer (Capreolus capreolus) in a gradient of forest fragmentation: behavioural plasticity and choice of cover. *Canadian journal of zoology*, **82** (9), 1528-1541.

Jespersen, P. and Tåning, Å., 1950. Studies in bird migration, being the collected papers of H. Chr. C. Mortensen 1856–1921. Copenhagen: Munksgaard.

Johnson, D.H., 1980. The Comparison of usage and availability measurements for evaluating resource preference. *Ecology,* **61**, 65-71.

Kämmerle, J.L., Brieger, F., Kroeschel, M., Hagen, R., Storch, I. and Suchant, R., 2017. Temporal patterns in road crossing behaviour in roe deer (Capreolus capreolus) at sites with wildlife warning reflectors. *PloS one*, **12** (9), e0184761.

Kareiva, P.M. and Shigesada, N., 1983. Analyzing insect movement as a correlated random walk. *Oecologia*, **56** (2-3), 234-238.

Kays, R., Crofoot, M.C., Jetz, W. and Wikelski, M., 2015. Terrestrial animal tracking as an eye on life and planet. *Science*, **348** (6240), aaa2478.

Kemp, M.U., Emiel van Loon, E., Shamoun-Baranes, J. and Bouten, W., 2012. RNCEP: global weather and climate data at your fingertips. *Methods in Ecology and Evolution*, **3** (1), 65-70.

Klafter, J., Shlesinger, M.F. and Zumofen, G. 1996. Beyond Brownian motion. *Physics Today,* **49**, 33-39.

Konzack, M., Gijsbers, P., Timmers, F., van Loon, E., Westenberg, M.A. and Buchin, K., 2018. Visual exploration of migration patterns in gull data. *Information Visualization*, 1473871617751245.

Krop-Benesch, A., Berger, A., Hofer, H. and Heurich, M., 2013. Long-term measurement of roe deer (Capreolus capreolus)(Mammalia: Cervidae) activity using two-axis accelerometers in GPS-collars. *Italian Journal of Zoology*, **80** (1), 69-81.

Laube, P., Imfeld, S. and Weibel, R. 2005. Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science,* **19** (6), 639-668.

Lele, S.R., Merrill, E.H., Keim, J. and Boyce, M.S., 2013. Selection, use, choice and occupancy: clarifying concepts in resource selection studies. *Journal of Animal Ecology*, **82** (6), 1183-1191.

LeMunyan, C.D., White, W., Nyberg, E. and Christian, J.J., 1959. Design of a miniature radio transmitter for use in animal studies. *The Journal of Wildlife Management*, **23** (1), 107-110.

Liberg, O., Johansson, A., Andersen, R. and Linnell, J.D.C., 1998. Mating system, mating tactics and the function of male territoriality in roe deer. *In:* Andersen, R., Duncan, P and Linnell, J.D.C., eds. *The European roe deer: the biology of success.* Oslo: Scandinavian University Press, 221-256.

Linares, C., Vidal, M., Canals, M., Kersting, D.K., Amblas, D., Aspillaga, E., Cebrián, E., Delgado-Huertas, A., Díaz, D., Garrabou, J. and Hereu, B., 2015. Persistent natural acidification drives major distribution shifts in marine benthic ecosystems. *Proc. R. Soc. B,* **282** (1818), 20150587.

Linnell, J.D.C., Duncan, P. and Andersen, R., 1998. The European roe deer: a portrait of a successful species. *In:* Andersen, R., Duncan, P and Linnell, J.D.C., eds. *The European roe deer: the biology of success.* Oslo: Scandinavian University Press, 11-22.

Long, J.A., 2016. A field-based time geography for wildlife movement analysis. *International Conference on GIScience Short Paper Proceedings*, 1 (1).

Magdy, N., Sakr, M.A., Mostafa, T. and El-Bahnasy, K., 2015. Review on trajectory similarity measures. *IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS),* **2015**, 613-619.

Mancinelli, S., Peters, W., Boitani, L., Hebblewhite, M. and Cagnacci, F., 2015. Roe deer summer habitat selection at multiple spatio-temporal scales in an Alpine environment. *Hystrix, the Italian Journal of Mammalogy*. **26** (2), 132-140.

Mandelbrot, B., 1967. How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science*, **156** (3775), 636-638.

Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L. and Erickson, W.P., 2002. Introduction to Resource Selection Studies. *In: Resource selection by animals: statistical design and analysis for field studies*. Dordrecht: Kluwer Academic Publishers, 1-15.

Marsh, L. and R. Jones, 1988. The form and consequence of random walk movement models. *Journal of Theoretical Biology,* **133**, 113-131.

McClintock, B.T., Johnson, D.S., Hooten, M.B., Ver Hoef, J.M. and Morales, J.M., 2014. When to be discrete: the importance of time formulation in understanding animal movement. *Movement Ecology*, **2** (1), 21.

Meisingset, E.L., Loe, L.E., Brekkum, Ø., Van Moorter, B. and Mysterud, A., 2013. Red deer habitat selection and movements in relation to roads. *The Journal of Wildlife Management*, **77** (1), 181-191.

Morellet, N., Bonenfant, C., Börger, L., Ossi, F., Cagnacci, F., Heurich, M., Kjellander, P., Linnell, J.D., Nicoloso, S., Sustr, P. and Urbano, F., 2013. Seasonality, weather and climate affect home range size in roe deer across a wide latitudinal gradient within Europe. *Journal of Animal Ecology,* **82** (6), 1326-1339.

Morris, D.W., 1987. Ecological scale and habitat use. *Ecology,* **68** (2), 362-369.

Mysterud, A., Bjørnsen, B.H. and Østbye, E., 1997. Effects of snow depth on food and habitat selection by roe deer Capreolus capreolus along an altitudinal gradient in south-central. *Wildlife Biology*, **3** (1). 27-33.

Mysterud, A. and Ims, R.A., 1998. Functional responses in habitat use: availability influences relative use in trade-off situations. *Ecology*, **79** (4), 1435-1441.

Mysterud, A., 1999. Seasonal migration pattern and home range of roe deer (Capreolus capreolus) in an altitudinal gradient in southern Norway. *Journal of zoology*, **247** (4), 479-486.

Mysterud, A., Lian, L.B. and Hjermann, D.Ø., 1999a. Scale-dependent trade-offs in foraging by European roe deer (Capreolus capreolus) during winter. *Canadian journal of zoology,* **77** (9), 1486-1493.

Mysterud, A., Larsen, P.K., Ims, R.A. and Østbye, E., 1999b. Habitat selection by roe deer and sheep: does habitat ranking reflect resource availability?. *Canadian Journal of Zoology*, **77** (5),776-783.

Mysterud, A. and Østbye, E., 1999. Cover as a habitat element for temperate ungulates: effects on habitat selection and demography. *Wildlife Society Bulletin (1973-2006)*, **27** (2), 385-394.

Nathan, R., 2008. An emerging movement ecology paradigm. *Proceedings of the National Academy of Sciences*, 105 (49), 19050–19051.

Needleman, S.B. and Wunsch, C.D., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, **48** (3), 443-453.

Orians, G.H. and Wittenberger, J.F., 1991. Spatial and temporal scales in habitat selection. *The American Naturalist,* **137**, S29-S49.

Ossi, F., Gaillard, J.M., Hebblewhite, M. and Cagnacci, F., 2015. Snow sinking depth and forest canopy drive winter resource selection more than supplemental feeding in an alpine population of roe deer. *European journal of wildlife research*, **61** (1), 111-124.

Ossi, F., Gaillard, J.M., Hebblewhite, M., Morellet, N., Ranc, N., Sandfort, R., Kroeschel, M., Kjellander, P., Mysterud, A., Linnell, J.D.C., Heurich, M., Soennichsen, L., Sustr, P., Berger, A., Rocca, M., Urbano, F. and Cagnacci, F. 2017. Plastic response by a small cervid to supplemental feeding in winter across a wide environmental gradient. *Ecosphere*, **8** (1).

Padié, S., Morellet, N., Hewison, A.J.M., Martin, J.L., Bonnot, N., Cargnelutti, B. and Chamaillé-Jammes, S., 2015. Roe deer at risk: teasing apart habitat selection and landscape constraints in risk exposure at multiple scales. *Oikos*, **124** (11), 1536-1546.

Pagon, N., Grignolio, S., Pipia, A., Bongi, P., Bertolucci, C. and Apollonio, M., 2013. Seasonal variation of activity patterns in roe deer in a temperate forested area. *Chronobiology international,* **30** (6), 772-785.

Parmesan, C., 2006. Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 637-669.

Patlak, C.S., 1953. Random walk with persistence and external bias. *The bulletin of mathematical biophysics*, ***15*** (3), 311-338.

Peters, W., Hebblewhite, M., Mysterud, A., Spitz, D., Focardi, S., Urbano, F., Morellet, N., Heurich, M., Kjellander, P., Linnell, J.D.C. and Cagnacci, F., 2017. Migration in geographic and ecological space by a large herbivore. *Ecological Monographs*, **87** (2), 297-320.

Plard, F., Gaillard, J.M., Bonenfant, C., Hewison, A.M., Delorme, D., Cargnelutti, B., Kjellander, P., Nilsen, E.B. and Coulson, T., 2013. Parturition date for a given female is highly repeatable within five roe deer populations. *Biology letters*, **9** (1), 20120841.

Prokić, J., Wieling, M. and Nerbonne, J., 2009. Multiple sequence alignments in linguistics. *Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education, Association for Computational Linguistics,* 18-25.

Ratikainen, I.I., Panzacchi, M., Mysterud, A., Odden, J., Linnell, J.D.C. and Andersen, R., 2007. Use of winter habitat by roe deer at a northern latitude where Eurasian lynx are present. *Journal of Zoology*, **273** (2),192-199.

Recchia, G. and Louwerse, M.M., 2013. A Comparison of String Similarity Measures for Toponym Matching. In *COMP@ SIGSPATIAL,* 54-61.

Rivrud, I.M., Loe, L.E. and Mysterud, A., 2010. How does local weather predict red deer home range size at different temporal scales? *Journal of Animal Ecology,* **79** (6), 1280-1295.

Rodriguez-Cabal, M.A., Barrios-Garcia, M.N. and Nuñez, M.A., 2012. Positive interactions in ecology: filling the fundamental niche. *Ideas in Ecology and Evolution*, **5**, 37-41.

Rutz, C. and Hays, G.C., 2009. New frontiers in biologging science. *Biology Letters,* **5**, 289-292.

San José, C., Lovari, S. and Ferrari, N., 1997. Grouping in roe deer: an effect of habitat openness or cover distribution? *Acta Theriologica*, **42** (2), 235-239.

Sanger, F. and Tuppy, H., 1951. The amino-acid sequence in the phenylalanyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates. *Biochemical Journal*, **49** (4), 463.

Sempéré, A.J., Mauget, R. and Mauget, C., 1998. Reproductive physiology of roe deer. *In:* Andersen, R., Duncan, P and Linnell, J.D.C., eds. *The European roe deer: the biology of success.* Oslo: Scandinavian University Press, 161-188.

Shamoun-Baranes, J., Bouten, W. and van Loon, E.E., 2010. Integrating meteorology into research on migration. *Integrative and Comparative Biology,* **50** (3), 280-292.

Shamoun-Baranes, J., van Loon, E.E., Purves, R.S., Speckmann, B., Weiskopf, D. and Camphuysen, C.J., 2011. Analysis and visualization of animal movement. *Biology Letters.* **8** (1), 6-9.

Shamoun-Baranes, J., Bom, R., van Loon, E.E., Ens, B.J., Oosterbeek, K. and Bouten, W., 2012. From sensor data to animal behaviour: an oystercatcher example. *PloS one*, **7** (5), e37997.

Shoval, N. and Isaacson, M., 2007. Sequence alignment as a method for human activity analysis in space and time. *Annals of the Association of American geographers*, **97** (2), 282-297.

Shoval, N., Auslander, G.K., Freytag, T., Landau, R., Oswald, F., Seidl, U., Wahl, H.W., Werner, S. and Heinik, J., 2008. The use of advanced tracking technologies for the analysis of mobility in Alzheimer's disease and related cognitive diseases. *BMC geriatrics*, **8** (7).

Slingsby, A. and van Loon, E., 2016. Exploratory visual analysis for animal movement ecology. *Computer Graphics Forum,* **35** (3), 471-480.

Soni, K.K., Vyas, R. and Sinhal, A., 2014. Importance of string matching in real world problems. *International Journal Of Engineering And Computer Science*, **3** (6), 6371-6375.

Spagnesi, M., Toso, S. and Catalano, U., 1999. Iconografia dei mammiferi d'Italia.: Ministero dell'ambiente, Servizio conservazione natura. Tavole di Umberto Catalano. Ghigi, A., Ed. Instituto Nazionale per la Fauna Selvatica.

Spretke, D., Bak, P., Janetzko, H., Kranstauber, B., Mansmann, F. and Davidson, S., 2011. Exploration through enrichment: a visual analytics approach for animal movement. *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems,* 421-424.

Studer, M. and Ritschard, G., 2016. What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **179** (2), 481-511.

Toor, M.L., Newman, S.H., Takekawa, J.Y., Wegmann, M. and Safi, K., 2016. Temporal segmentation of animal trajectories informed by habitat use. *Ecosphere*, **7** (10), e01498.

Thurfjell, H., Ciuti, S. and Boyce, M.S., 2014. Applications of step-selection functions in ecology and conservation. *Movement ecology*, **2** (4).

Tucker, M.A., Böhning-Gaese, K., Fagan, W.F., Fryxell, J.M., Van Moorter, B., Alberts, S.C., Ali, A.H., Allen, A.M., Attias, N., Avgar, T. and Bartlam-Brooks, H., 2018. Moving in the Anthropocene: Global reductions in terrestrial mammalian movements. *Science*, **359** (6374), 466-469.

Tufto, J., Andersen, R. and Linnell, J.D.C., 1996. Habitat use and ecological correlates of home range size in a small cervid: the roe deer. *Journal of Animal Ecology*, **65** (6), 715-724.

Turchin, P., 1996. Fractal analyses of animal movement: a critique. *Ecology*, **77** (7), 2086-2090.

Van Moorter, B., Visscher, D., Benhamou, S., Börger, L., Boyce, M.S. and Gaillard, J.M., 2009. Memory keeps you at home: a mechanistic model for home range emergence. *Oikos*, **118** (5), 641-652.

Vanpé, C., Kjellander, P., Galan, M., Cosson, J.F., Aulagnier, S., Liberg, O. and Hewison, A.M., 2007. Mating system, sexual dimorphism, and the opportunity for sexual selection in a territorial ungulate. *Behavioral Ecology*, **19** (2), 309-316.

Vanpé, C., Morellet, N., Kjellander, P., Goulard, M., Liberg, O. and Hewison, A.J., 2009. Access to mates in a territorial ungulate is determined by the size of a male's territory, but not by its habitat quality. *Journal of Animal Ecology*, **78** (1), 42-51.

Verhein, F. and Chawla, S., 2006. Mining spatio-temporal association rules, sources, sinks, stationary regions and thoroughfares in object mobility databases. *In:* Li Lee, M., Tan, K.L. and Wuwongse, V., eds. *Database Systems for Advanced Applications. DASFAA 2006. Lecture Notes in Computer Science*, **3882**. Berlin: Springer, 187-201.

Versichele, M., Neutens, T., Delafontaine, M. and 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.

Wagner, R.A. and Fischer, M.J., 1974. The string-to-string correction problem. *Journal of the Association for Computing Machinery (JACM)*, **21** (1), 168-173.

Wang, Y. and Moore, R.F., 2005. Long-distance bird migrants adjust their foraging behavior in relation to energy stores. *Acta zoologica Sinica*, **51** (1), 12-23.

Webb, S.L., Riffell, S.K., Gee, K.L. and Demarais, S., 2009. Using fractal analyses to characterize movement paths of white-tailed deer and response to spatial scale. *Journal of Mammalogy*, **90** (5), 1210-1217.

Wikelski, M. and Kays, R., 2014. Movebank: archive, analysis and sharing of animal movement data. *World Wide Web electronic publication*.

Wilson, R.J., Gutierrez, D., Gutierrez, J. and Monserrat, V.J., 2007. An elevational shift in butterfly species richness and composition accompanying recent climate change. *Global Change Biology*, **13** (9), 1873-1887.

Wilson, C., 2008. Activity patterns in space and time: calculating representative Hagerstrand trajectories. *Transportation*, **35** (4), 485-499.

Whittaker, R.H., Levin, S.A. and Root, R.B., 1973. Niche, habitat, and ecotope. *The American Naturalist*, **107** (955), 321-338.

Xavier, G. and Dodge, S., 2014. An exploratory visualization tool for mapping the relationships between animal movement and the environment. *In: MapInteract 2014 - Proceedings of the 2nd ACM SIGSPATIAL International Workshop on MapInteraction,* 36-42.

Zemanova, M.A., Perotto-Baldivieso, H.L., Dickins, E.L., Gill, A.B., Leonard, J.P. and Wester, D.B., 2017. Impact of deforestation on habitat connectivity thresholds for large carnivores in tropical forests. *Ecological Processes*, **6** (21).

Zúñiga, D., Falconer, J., Fudickar, A.M., Jensen, W., Schmidt, A., Wikelski, M. and Partecke, J., 2016. Abrupt switch to migratory night flight in a wild migratory songbird. *Scientific reports*, **6**, 34207.

CHAPTER 2

DATA

2.1. THE EUROUNGULATES PROJECT

This PhD project relies on GPS movement and environmental data provided by the Euroungulates project. Euroungulates is the umbrella project of three species-specific networks (EuroDEER, EurREDDEER and EuroBOAR) and provide database infrastructure respectively for roe deer, red deer (*Cervus elaphus*) and wild boar (*Sus scrofa*). The seeds of this data sharing project were planted in 2008, which quickly resulted in the initiation of the EuroDEER collaborative project. Due to its success, the network was extended in 2014 to red deer and in 2016 to wild boar, which is how Euroungulates came to life. The main aim of Euroungulates is to produce collaborative science, supported by a community of specialists in the species, providing study area specific insights. Indeed, data sharing and collaborative science, can lead to scientific breakthroughs, and facilitates large-scaled, long-term and multi-species studies (Open Science Symposium, 7-8 Dec 2017; EuroDEER meeting VI,VII, VIII, IX, X). For instance, with such broad studies the effect of climate and habitat change on animal behaviour and distributions can be properly quantified (Carpenter *et al.* 2009, Cagnacci *et al.* 2011, Morellet *et al.* 2013). Scientific outputs of the collaborative network EuroDEER evidence the recent progress that has been made in clarifying such broad-scaled continental patterns in roe deer movement ecology, including identifying partial-migration behaviour, and the resident to migration movement plasticity (Cagnacci *et al.* 2011, 2016, Peters *et al.* 2017), home range variability along a latitudinal gradient (Morellet *et al.* 2013), breeding responses (Debeffe *et al.* 2014), feeding sites use (Ossi *et al.* 2017) and sub-seasonal functional home ranges (Couriot *et al.* 2018).

Anno 2018, the three databases combined contain more than 8,340,000 GPS locations, 70,800 VHF locations and 53,550,000 activity records, corresponding in total to 2490 animals, collected for more than 60 study areas from 18 countries, with the oldest records dating back to 1982. In table 2.1 statistics are summarised for each database separately.

Table 2.1. Statistics of the Euroungulates database

	EuroDEER	EuREDDEER	EuroBOAR
Animals	1,714	516	260
Study areas	28	20	16
Countries	15	15	18
Partners	33	33	26
GPS locations	4,741,364	2,791,066	811,129
VHF locations	46,859	19,449	4,517
Activity record	53,546,056	0	0

2.1.1. Database software platform

Euroungulates datasets are stored in a Spatial Relational DataBase Management System (SRDBMS) developed on PostgreSQL/PostGIS, hosted by a server-client architecture developed at the Research and Innovation Center the Edmund Mach Foundation (FEM-CRI, Information System for Analysis and Management of Ungulate Data, ISAMUD, 2005, Cagnacci & Urbano 2008). An SRDBMS is a database designed to store, manage, query, and manipulate large amounts of spatial data. Like a GIS (QGIS, GRASS, OpenJUMP) a spatial database allows to perform spatial operations, with the advantage that these can be efficiently performed over large datasets. In a relational database, data is stored in multiple indexed tables that are linked to each other by common fields. Moreover, with a set of database connection parameters the database can be connected with GIS and statistical software, and allows to perform computations with the interconnected software of choice. Also, several procedural languages are implemented directly within the base distribution of PostgreSQL, including PL/pgSQL, PL/Tcl, PL/Perl, PL/Python, and several external procedural languages are maintained (e.g. PL/Java, PL/R) among which PL/R is also available for use in the Euroungulates database. Using PL/R it is possible to call R functionality within the database. The database is also available online via phpPgAdmin interface (http://eurodeer.org/) and a practical guide is available allowing other movement ecologists to replicate the standardised database structure for their own animal movement data (Urbano & Cagnacci 2014).

2.1.2. Database structure

The core of the Euroungulates database are the GPS, VHF and accelerometer data. These tables in relation with sensor deployment information (i.e. start and end time of deployment of the collar to an animal) allow to define the relation between movement data, an animal and a sensor. While these datasets are extremely valuable, it is especially the link with ancillary data about the individuals, populations and environmental context that allow to thoroughly understand the movement of animals. See figure 2.1 for a simplified graphical representation of the database structure using GPS data. Because of the major importance of such ancillary data, Euroungulates maintains a database structure that can also accommodate this information. Over the last couple of years the data processing protocol for ancillary data has been standardised, which includes meticulous validation procedures and quality checks. Ancillary data includes standardised information on individual animal characteristics (e.g. age, sex), sensor characteristics (e.g. brand), population level information (e.g. management regimes, predator presence, competitors, human disturbance, density estimation), capture data (e.g. capture time, manipulation time, injuries), individual contacts data (e.g. last contact time, type of contact, cause and date of mortality) and many validated static and dynamic environmental data layers. See table 2.2 and figure 2.2 for an overview of environmental layers available in the Euroungulates database.

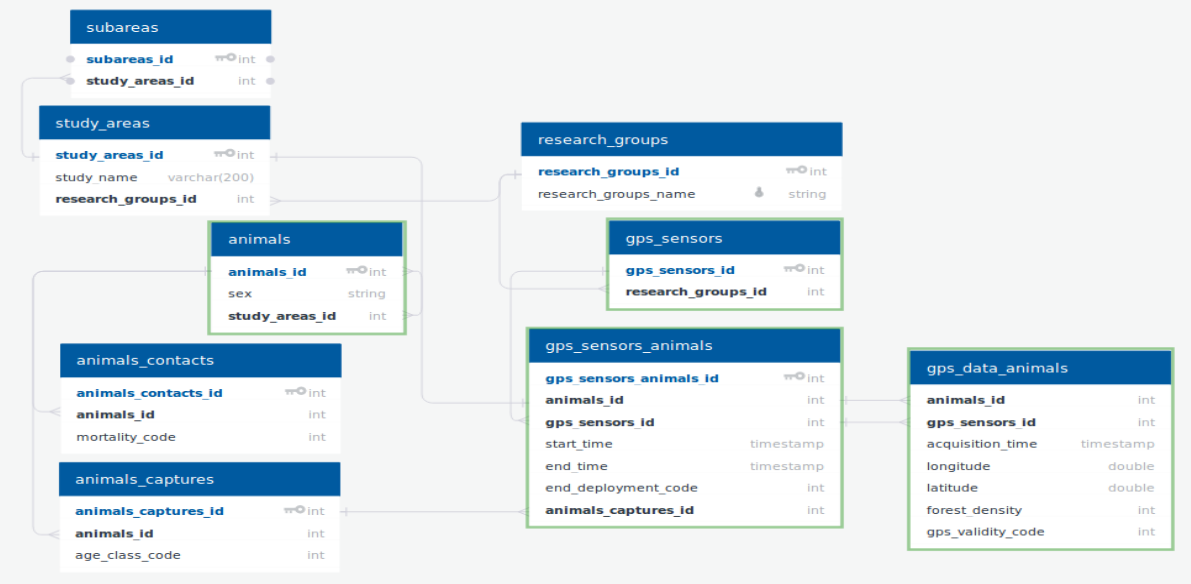


Figure 2.1. Simplified graphical representation of the database for GPS data. The core data tables are shown in green boxes. Ancillary data are stored in different tables such as in animals_captures, animals_contacts, subareas (i.e. population data) and environmental data in the GPS data animals table (e.g. forest density).

The database further includes various commonly used functions to facilitate data preparation and analytical procedures, including a function to assign a fix to daylight or night hours based on the timestamp and location, a regularisation function to create a regular time series of locations using a defined time interval and time buffer, a simple linear interpolation function to calculate the geometry to remove unwanted time gaps between two subsequent locations, a function to calculate the age class corresponding to each GPS location and one to calculate geometric parameters such as step length, turning angle and speed between consecutive relocations (see section documentation at the github page[3] for examples of these functions). Hence, Euroungulates is a unique platform, offering big data analysis opportunities, but with very detailed and verified information and several analytical tools, thus allowing to answer very specific ecological questions at a broad scale. Since the database structure is very well documented in the user guide available at the EuroDEER github page, section documentation, we will further mainly focus on the protocols for data collection and data quality.

Table 2.2. Environmental layers available in the Euroungulates database, see for details on the EuroDEER website (http://eurodeer.org/environmental-covariates/). In light grey, the datasets that were used in Chapters 3, 4 and 5.

Dataset	**Spatial and temporal resolution**
NDVI MODIS	250 m 16-days
NDVI MODIS smoothed	250 m 10-days
NDVI MODIS smoothed – BOKU University	250 m 7-days
SNOW MODIS	500 m
Tree Cover Density (TCD, 2012)	20 m
Corine Land Cover (CLC, 1990,2000,2006,2012)	100 m (& vector 2012)
ASTER Digital Elevation Model (ASTER DEM)	30 m
SRTM Digital Elevation Model (SRTM DEM)	90 m
Digital Elevation Model over Europe (EU-DEM)	30 m
NDVI-based temporal variability: 1. Constancy, 2. Contingency, 3. Predictability	250 m
NDVI-based average vegetation phenology	250 m
SNOW MODIS-based yearly winter severity	500 m

---

[3] https://github.com/eurodeer/eurodeer_db

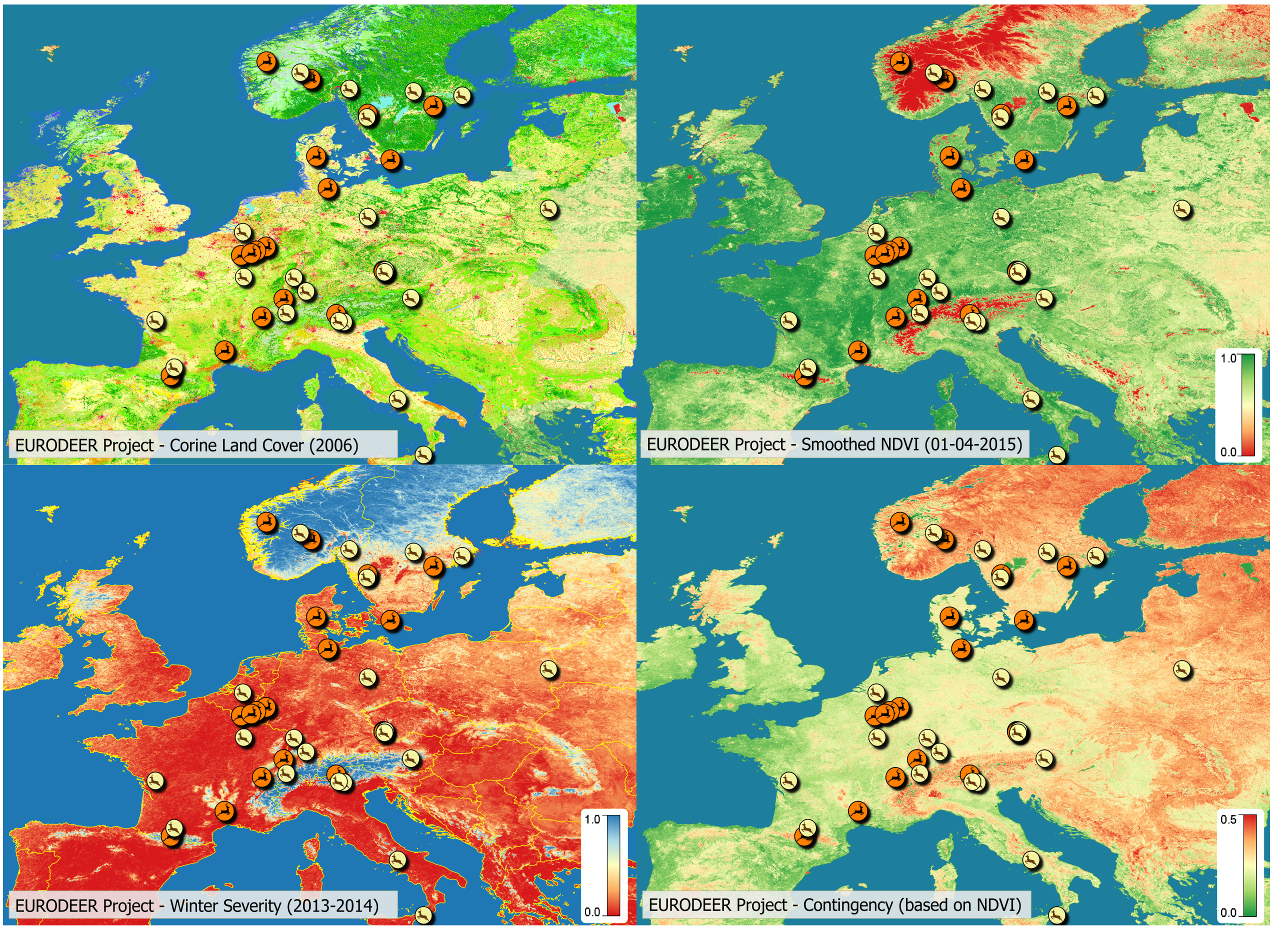


Figure 2.2. Four environmental layers available in the Euroungulates database. From upper-left to bottom-right Corine Land Cover (2006), Smoothed NDVI (01-04-2015), Contigency (based on NDVI) and Winter Severity (2013-2014). Figures were extracted from the EuroDEER website (http://eurodeer.org/environmental-covariates/).

2.1.3. Data collection protocol

Extending a movement database to multiple populations calls for data harmonisation practices (i.e. to combine data from different sources in a standard way) (Urbano et al. 2010). Currently there are no well developed standards available for biologging datasets, unlike in other biological and geographical fields, such as organisations Biodiversity Information Standards (TDWG; Darwin core, ABCD) and the Open Geospatial Consortium (OGC; Keyhole Markup Language, KML; Geography Markup Language, GLM), respectively (Percivall 2009). Because of the lack of standards it may occur that even sensors of different generations of the same brand may structure and format data in very different ways. Available formats are becoming increasingly complex, especially with the implementation of multiple sensors on one device, which calls for a biologging standard for easy retrieval, storage and processing of data collected through biologging devices. Several organisations such as movebank and the biologging society are making efforts to develop a biologging standard.

To facilitate data harmonisation in the Euroungulates database several standardised protocols for data collection of tracking (i.e. GPS, VHF, accelerometer) and ancillary data (e.g. capture, contacts) have been implemented. Such protocols were developed through expert-based knowledge by collaborators of the Euroungulates network together with the main database developers (Cagnacci, F. & Urbano F.). Modifications and improvements to protocols are reported on the website, as well as presented and discussed during the annual meetings organised at one of the collaborators' institutes. Once the protocols are established, they are disseminated over different research groups, which can be freely adopted for future projects as minimum data collection requirement. A concise overview will be given on some of the protocols on which I have worked most intensively for their correct implementation, concerning GPS, VHF and accelerometer collection protocols and captures and contacts protocols.

To harmonise different GPS datasets minimum requirements for providing data include the coordinates in WGS84, timestamps in UTC and the corresponding sensor and animal identifier, together with animals, sensors and animals-sensors tables (see above). Additional information concerns altitude, temperature and any information that allows to assess the quality of the GPS locations, such as the number of satellites and the Dilution Of Precision (DOP). In general the preference is given to the raw input files directly retrieved from the sensors, so that introduced bias in the data is minimised. Further, the protocol for GPS data collection suggested initially a three hour fix schedule, but there is no clear consensus. Also rolling schedules have been suggested to cover all hours or higher fix rates during captures. Since storage capacity is becoming better, schedules are now often set to a temporal resolution of one hour. Several functions have been implemented to deal with irregularities in the sampling schedules, such as a regularisation and interpolation tool (see above).

VHF datasets include similar minimum requirements (i.e. coordinates, timestamps, animal and sensor identifier), but since they are often old and pre-processed they need to be carefully dealt with. Often locations are only collected in local coordinate reference systems and timestamps are provided in local time formats usually accounting for time saving. Therefore data providers need to ensure to provide the correct metadata on the spatial reference system and time format in which original coordinates and timestamps are stored. With this metadata coordinates and timestamps can be transformed into corresponding formats in which spatial and temporal locations are stored in the database (i.e. WGS84 and UTC, respectively).

The harmonisation of accelerometer data is more difficult since even for the same sensor measurements can be stored in several modes (Krop-Benesch *et al.* 2010). For instance, the company Vectronic Aerospace GmbH distinguishes seven modes. Three of them measure acceleration on two axes (x, forward/backward motion; y, side-ward and rotary motion) by averaging over a certain interval (300s, 152s, 63s) and four modes give measurements using a different set of parameters (e.g. mode 5: head angle/acceleration, mode 6: Acceleration Peak/Acceleration Threshold) and the sampling interval can be set manually in steps of 8s (i.e. 8, 16, 24, etc.). The standard setting of these collars, and hence for most sensors in the database, provide acceleration averaged over 300s measured on two axes (i.e. mode 1). Since accelerometer data is often not directly comparable it is stored into five different tables, each corresponding to a specific sensor and mode. Accelerometer data further provides challenges in data storage and processing, since many more records are collected than is usual the case for GPS. To link accelerometer data to GPS data, it is further important to always provide the associated GPS sensor of the accelerometer. The protocol on acceleration data further provides suggestions for error checks and potential correction but these are mainly to be performed by the analyst.

The protocols for captures consist out of two main parts, a behavioural and a physiological assessment. The measurements are non-invasive and should minimise the costs and handling time during the captures. The behavioural protocol is based on two subjective scales assessing how resistant the animal was during handling and how the animal reacts directly after release. Physiological protocols include measures such as hind foot length, body mass, heart rate (beats/min) and measurement time and rectal body temperature (°C) and measurement time. Moreover also the time of capture, start and end of handling and release time are expected to be recorded, as well as individual traits such as sex and estimated age. In the eurodeer database age is simplified into three relevant classes, fawn (less than one year), yearling (less than two years) and adults, where the change data between classes is set to the first of April. Note that 'fawns' are usually caught only a month or two before the 'age-change date', The capture protocols have been developed by Kjellander, P., Nicoloso, S., Hewison, M. and Cagnacci, F. while the age classification was developed by Morellet, N. in discussion with the Euroungulates community.

The contacts table reports only the last contact with the animal *after* the last deployment, including mortality. This is an important source of information for survival studies. Until recent, contacts were provided together with the animals table, but since a couple of months there is a new standard sheet to collect these data.

2.1.4. Data cleaning protocol

One of the biggest added values of the Euroungulates database, is the time investment to standardisation and validation of the data. Data sources can be very heterogeneous due to a variety of data collection and processing protocols followed by research groups. Although that standardised protocols have been developed, often new research groups were not aware of those, data might be collected several years back or protocols might not be implemented. The path from data source to a structured database can be treacherous, where in each step errors can be introduced. For instance, mistakes can be introduced during field work by recording wrong information, during transcription of collected field data, due to conversion issues between different software platforms, or by incorrect automated data handling scripts. As a result, data provided are rarely consistent and without errors. For these reasons, before data can be shared among the community data quality check is a fundamental step, since erroneous data can lead to misleading scientific outputs, also known in computer science as “Garbage In, Garbage Out” (GIGO). The general philosophy of Euroungulates is not to remove data, instead GPS locations are tagged using validation codes, which allows collaborators to assess the reliability of the data. A part of the validation procedure has been based on a published standardised protocol (Bjørneraas *et al.* 2010), which has been further developed in Urbano *et al.* (2014). A summary of the validity codes is given in table 2.3. In addition, numerous data consistency checks are performed so that relational tables are harmonised. For instance, if start and end of deployment are consistent with actual GPS data, if there are multiple individuals marked with the same sensor at the same time, if capture dates are consistent with the start of deployment, if there are GPS locations, deployments, capture or contact dates after the death of an animal, if the age classes are consistent, etc. Queries used to hunt for errors are stored and openly available at the eurodeer github page, in the data management section (see footnote 3 for the link).

Table 2.3. GPS validity codes in the Euroungulates database

Code	Description
0	Record with no coordinates
1	Valid record (valid position)
2	Record with a low degree of reliability
3	Suspicious location: to be checked
11	Record wrong: out of the study area
12	Record wrong: impossible spike
13	Record wrong: impossible place (e.g. lake or sea)
14	Record wrong: not associated to any animal (out of deployment time range)
15	Record wrong: impossible movement
21	Record wrong: duplicated timestamp
22	Record wrong: impossible timestamp
23	Record with no timestamp
31	Record wrong: belonging to a dataset with very suspicious spatial patterns

2.1.5. Data collection perspective

ASIMUD, the predecessor of Euroungulates, was specifically developed for a single population and could easily be enhanced by many features such as automated data entry (Cagnacci & Urbano 2008). Currently the data import procedure, starting from data collection to quality checks to final import are mainly performed through intense data curatorship, an essential task when dealing with heterogeneous datasets. Experience as data base curator over the last four years, privileged me to gain a deep insight into the database structure and which potential errors to encounter. Since most protocols are standardised and available the next step is to further streamline the data collection protocol. While currently data can be provided into any kind of format (e.g. CSV, DBF, Microsoft Excel or Access formats) we aim to facilitate this through a web application, or R shiny app. Such a web application should allow data providers to upload the necessary data to a web-interface (e.g. drag and drop, navigate to file, copy/paste from file), identify automatically evident errors and inconsistencies (e.g. time of death before last capture) and finally provide direct access to the data curator through a temporary folder in the database. Data entry could be made more interactive, involved and even pleasant, by showing some simple statistics, figures and maps of the providers' own dataset. For example, when uploading captures we could show a histogram with the hind-food length and by matching the size of the capture location to the hind-food size we could show on a map if there is a spatial association, or, when uploading contact information we could show the main reason of mortality through a barchart and color-coded locations of the associated event on a map. Based on my experience as data curator, well-documented data entry templates have been developed, which highlights potential errors and inconsistencies that have been identified observed in the data throughout the years[4]. These tables could be used as further template for developing such a web application, or Shiny app. To design such an application and assure that the data entry is user-friendly a questionnaire could also be distributed to understand better how data providers prepare the data files and to match the procedure to their general interests.

2.2. DATASETS

From a total of twenty-eight populations available in the EuroDEER database we extracted GPS movement data of nine populations which retained a large enough sample size. The subsampling was constrained to GPS trajectories with a fixed 4-hour relocation interval at fixed time stamps (0, 4, 8, 12, 16, 20 h). We chose an interval of 4 hours as a trade-off between number of sequences and temporal resolution. When the temporal resolution decreases the sample size increases, but instead sequential patterns become less detailed. Populations concern Southcentral Norway, NK5; Southwest France, FR8; Switzerland or Swiss Alps, CH25; Southern Germany, DE15, DE31; Southeast Germany or Bavarian Forest, DE2; Northern Italy or Italian Alps, IT1, IT24; Eastern Austria or Austian Alps, AU17 (Fig.2.3). The population of the Italian Alps (IT1) has been used to develop SAM as described in Chapter 3, while all nine populations were used for the development of IM-SAM in Chapter 4 and 5. All characteristics of these two datasets are described in table 2.4, highlighting the differences and similarities of the datasets. While a detailed description is given in each chapter we will here describe the general characteristics of the movement and environmental data of each dataset.

---

[4] https://docs.google.com/spreadsheets/d/1uTIsVvQPQ7cu190wBvrwg1qHcZP5dwwaVBqEr4KUrog/edit?usp=sharing

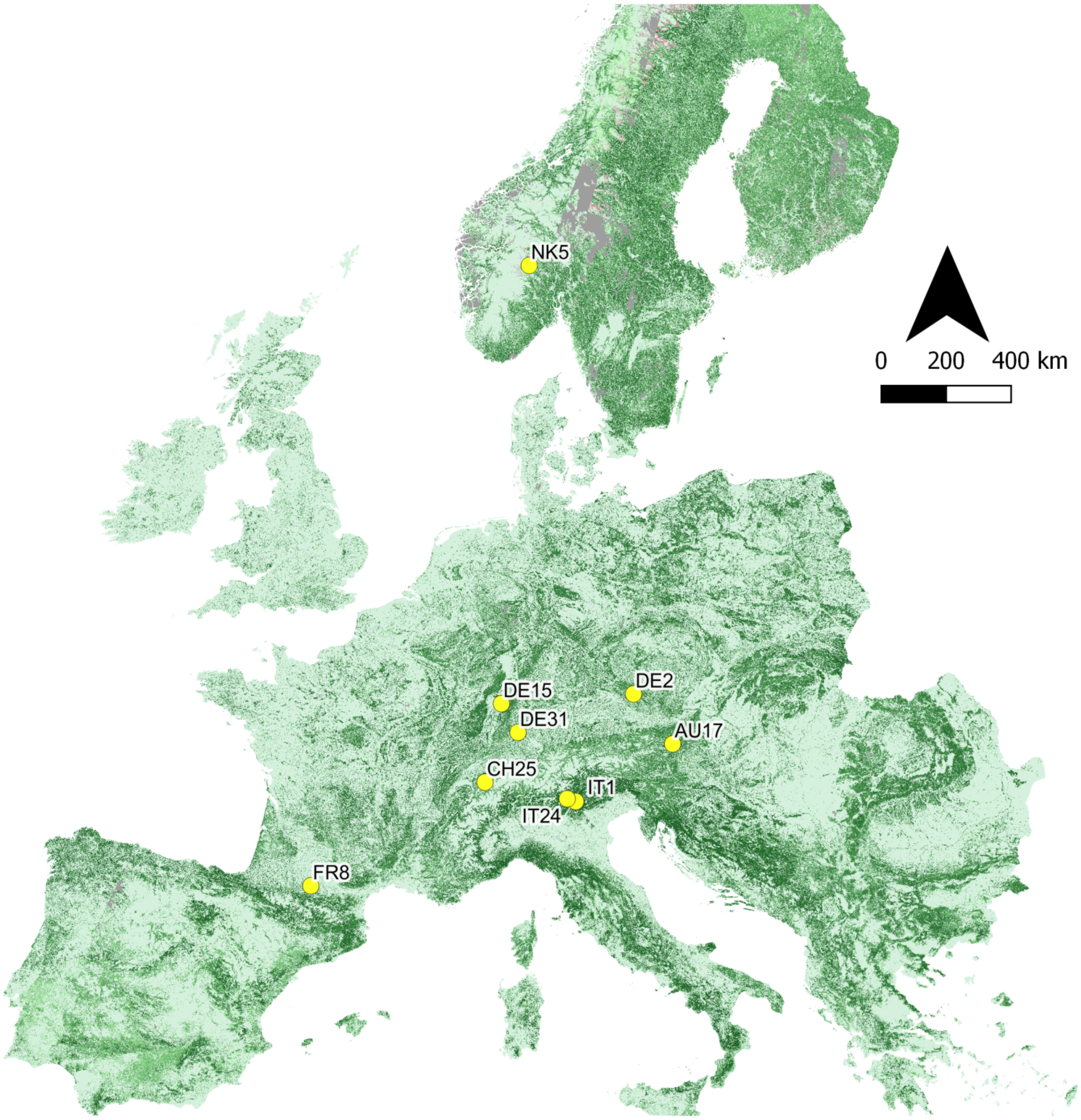


Figure 2.3. Study areas of the nine roe deer populations, mapped on the High Resolution Layer – Tree Cover Density 2012 (TCD, EEA 2012). Study areas are located in Southcentral Norway (NK5, 60°46'07"N 8°23'02"E), Southwest France (FR8, 43°20'01"N 0°43'57"E), Switzerland or Swiss Alps (CH25, 46°33'38"N 7°31'54"E), Southern Germany (DE15, 48°39'10"N 7°59'42"E; DE31, 47°50'24"N 8°43'47"E), Southeast Germany or Bavarian Forest (DE2, 48°57'31.9"N 13°23'32.7"E), Northern Italy or Italian Alps (IT1, 46°01'57"N 11°01'16"E; IT24, 46°04'53"N 10°43'12"E) and Eastern Austria or Austrian Alps (AU17, 47°26'39"N 15°06'07"E).

The population in the Italian Alps as used in Chapter 3 consists out of 111 bimonthly trajectories post-processing, corresponding to 26 animals and 40,626 relocations. Each trajectory consists out of 366 consecutive relocations collected at a fixed 4-hour interval (i.e. 6 relocations a day) for 61 days (i.e. bimonth) over a fixed yearly schedule starting on January 1st. We annotated bimonthly trajectories by combining two geographical parameters, habitat openness and elevation. Habitat openness was derived from the Corine Landcover 2006 (CLC, EEA, 2010; 100 m spatial resolution) and elevation from the Global Digital Elevation Model of ASTER (NASA-GDEM-ASTER, 2012; 30 m). We combined both raster datasets and reclassified the environment into four habitat classes distinguishing high-closed (HC), high-open (HO), low-closed (LC), low-open (LO). Each sequence thus represents the habitat use of these four classes over a period of 61 days. To each bimonthly sequence we further linked several individual traits including the age class (fawn, less than 1 year; yearling, between 1 and 2 years; adult, more than 2 years), sex (male/female) and migration status (migrant yes/no).

The nine populations as used in Chapter 4 and 5 consist out 5,402 biweekly sequences post-processing, corresponding to 404 animals and 518,592 relocations. Trajectories represented 96 consecutive relocations at the same fixed 4-hour interval corresponding to 16 days (i.e. biweek) over a fixed yearly schedule starting on January 1st . Biweekly trajectories were annotated with the percentage of tree cover (0-100%) extracted from the High Resolution Layer- Tree Cover Density 2012 (TCD, EEA 2012, 20 m). TCD raster values were reclassified to two classes of habitat openness using a cut-off point of pixel-level tree cover density at 50%, distinguishing closed (C, >50%) and open (O, <50%). Hence, each sequence here represents the habitat use of two classes over a period of 16 days. Each biweekly sequence was further associated with the biweekly period (1 to 23), at individual level by seven landscape metrics measured for each individual MCP (Chapter 5), the sex (male/female) and the population.

Table 2.4. Number of populations, animals, relocations and sequences, sequence characteristics, environmental variables and covariates used in Chapter 3 compared to Chapter 4 and 5.

	Chapter 3	**Chapter 4 & 5**
Populations in numbers		
Populations	1	9
Animals	26	404
Relocations+NA's	40,626	518,592
Sequences	111	5,402
Sequence Characteristics		
Temporal interval	4 hours	
Timestamps	0, 4, 8, 12, 16, 20 h	
Length	61 days (366 relocations)	16 days(96 relocations)
Number of classes	4	2
Environmental data		
Datasets	CLC (100 m) & ASTER-DEM (30 m)	TCD (20 m)
Classes	High-Open; High-Closed; Low-Open; Low-Closed	Open, Closed
Codes	HO, HC, LO, LC	O, C
Covariates		
	sex, age, migration	landscape metrics, biweek, population, sex

REFERENCES

Bjørneraas, K., Van Moorter, B., Rolandsen, C.M. and Herfindal, I., 2010. Screening global positioning system location data for errors using animal movement characteristics. *The Journal of Wildlife Management*, **74** (6), 1361-1366.

Cagnacci, F. and Urbano, F., 2008. Managing wildlife: a spatial information system for GPS collars data. *Environmental Modelling & Software*, **23** (7), 957-959.

Cagnacci, F., Focardi, S., Heurich, M., Stache, A., Hewison, A.J.M., Morellet, N., Kjellander, P., Linnell, J.D., Mysterud, A., Neteler, M. and Delucchi, L., 2011. Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors. *Oikos*, **120** (12), 1790-1802.

Cagnacci, F., Focardi, S., Ghisla, A., Moorter, B., Merrill, E.H., Gurarie, E., Heurich, M., Mysterud, A., Linnell, J., Panzacchi, M. and May, R., 2016. How many routes lead to migration? Comparison of methods to assess and characterize migratory movements. *Journal of Animal Ecology*, **85** (1), 54-68.

Carpenter, S.R., Armbrust, E.V., Arzberger, P.W., Stuart Chapin III, F., Elser, J.J., Hackett, E.J., Ives, A.R., Kareiva, P.M., Leibold, M.A., Lundberg, P. and Mangel, M., 2009. Accelerate synthesis in ecology and environmental sciences. *BioScience*, **59** (8), 699-701.

Couriot, O., Hewison, A.J.M., Saïd, S., Cagnacci, F., Chamaillé-Jammes, S., Linnell, J.D., Mysterud, A., Peters, W., Urbano, F., Heurich, M. and Kjellander, P., Nicoloso, S., Berger, A., Sustr, P., Kroeschel, M., Soennichsen, L., Sandfort, R., Gehr, B. and Morellet, N., 2018. Truly sedentary? The multi-range tactic as a response to resource heterogeneity and unpredictability in a large herbivore. *Oecologia*, 1-14.

Debeffe, L., Focardi, S., Bonenfant, C., Hewison, A.J.M., Morellet, N., Vanpé, C., Heurich, M., Kjellander, P., Linnell, J.D., Mysterud, A. and Pellerin, M., Sustr, P., Urbano, F. and Cagnacci, F., 2014. A one night stand? Reproductive excursions of female roe deer as a breeding dispersal tactic. *Oecologia*, **176** (2), 431-443.

Krop-Benesch, A., Berger, A., Streich, J., Scheibe, K. 2010. User's manual activity pattern, VECTRONICS Aerospace.

Morellet, N., Bonenfant, C., Börger, L., Ossi, F., Cagnacci, F., Heurich, M., Kjellander, P., Linnell, J.D., Nicoloso, S., Sustr, P. and Urbano, F., 2013. Seasonality, weather and climate affect home range size in roe deer across a wide latitudinal gradient within Europe. *Journal of Animal Ecology,* **82** (6), 1326-1339.

Ossi, F., Gaillard, J.M., Hebblewhite, M., Morellet, N., Ranc, N., Sandfort, R., Kroeschel, M., Kjellander, P., Mysterud, A., Linnell, J.D.C., Heurich, M., Soennichsen, L., Sustr, P., Berger, A., Rocca, M., Urbano, F. and Cagnacci, F. 2017. Plastic response by a small cervid to supplemental feeding in winter across a wide environmental gradient. *Ecosphere*, **8** (1).

Percivall, G., 2010. The application of open standards to enhance the interoperability of geoscience information. *International Journal of Digital Earth*, **3** (S1), 14-30.

Peters, W., Hebblewhite, M., Mysterud, A., Spitz, D., Focardi, S., Urbano, F., Morellet, N., Heurich, M., Kjellander, P., Linnell, J.D.C. and Cagnacci, F., 2017. Migration in geographic and ecological space by a large herbivore. *Ecological Monographs*, **87** (2), 297-320.

Urbano, F., Cagnacci, F., Calenge, C., Dettki, H., Cameron, A. and Neteler, M., 2010. Wildlife tracking data management: a new vision. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **365** (1550), 2177-2185.

Urbano, F. and Cagnacci, F., 2015. *Spatial database for GPS wildlife tracking data*. Italija: Springer.

CHAPTER 3

Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods

De Groeve, J., Van de Weghe, N., Ranc, N., Neutens, T., Ometto, L., Rota-Stabelli, O., Cagnacci, F.

In: Methods in Ecology and Evolution, 7 (3), 369-379.

ABSTRACT. Digital tracking technologies have considerably increased the amount and quality of animal trajectories, enabling the study of habitat use and habitat selection at a fine spatial and temporal scale. However, current approaches do not yet explicitly account for a key aspect of habitat use, namely the sequential variation in the use of different habitat features.To overcome this limitation, we propose a tree-based approach that makes use of Sequence Analysis Methods, derived from molecular biology, to explore and identify ecologically relevant sequential patterns in habitat use by animals. We applied this approach to ecological data consisting of simulated and real trajectories from a roe deer population (*Capreolus capreolus*), expressed as ordered sequences of habitat use. We show that our approach effectively captured spatio-temporal patterns of sequential habitat use by roe deer. In our case study, individual sequences were clustered according to the sequential use of the elevation gradient (first-order), and of open/closed habitats (second-order). We provided evidence for several behavioural processes, such as migration and daily alternating habitat use. Some unexpected patterns, such as homogeneous sequences of use of open habitat, could also be identified. Our findings advocate the importance of dealing with the sequential nature of movement data. Approaches based on Sequence Analysis Methods are particularly useful and effective since they allow exploring temporal patterns of habitat use in a synthetic and visually captive manner. The proposed approach represents a useful and effective way to classify individual movement behaviour across populations and species. Ultimately, this method can be applied to explore the temporal scale of ecological processes based on movement.

3.1. INTRODUCTION

Recent advances in digital tracking technology and increased availability of high-resolution environmental data by remote sensing have facilitated the collection of spatio-temporal series of animal-borne data (Cagnacci *et al.* 2010). Application of satellite navigation technology (e.g. Global Positioning System, GPS) to individual animals allows recording temporal sequences of animal locations at an unprecedented spatio-temporal resolution. Moreover, by projecting these locations onto spatial layers, including satellite images, it is possible to obtain robust and standardised information about the habitat of these animals (Urbano *et al.* 2010).

At present, an array of both exploratory and inferential methods is available to the analyst to investigate the relation between animal movement and the use of habitat. Exploratory methods apply multivariate analysis techniques (e.g. General Niche-Environment System Factor Analysis, GNESFA, The K-select analysis, (Canonical) Outlying Mean Index analysis; see R package AdehabitatHS of Calenge (2011) for an overview) to identify relevant variables describing the habitat (or the realised niche) of a population. Similarly, decision tree learning methods, such as random forest and CART modelling, are data mining techniques that present decision rules for classifying a set of data based on associated explanatory variables (see R package rpart of Therneau *et al.* (2014)). Conversely, inferential methods mainly consist of a variety of regression models testing the disproportion between used and available habitat units (i.e. habitat selection; Johnson 1980), such as Resource Selection Functions, RSF (Boyce *et al.* 2002) and Step Selection Functions, SSF (Fortin *et al.* 2005). In essence, exploratory methods offer a description of animals' habitat, whereas inferential methods allow to test specific hypotheses (Calenge & Basille 2008). In this sense, the first can be used to select explanatory variables which are relevant for the application of the latter (e.g. Calenge 2006, Wittemyer *et al.* 2008, Dray *et al.* 2010).

Despite the proliferation of exploratory methods, current approaches rarely evaluate the sequential use of habitats by animals, i.e. the sequence of locations (trajectory) vs. the underlying ordered pattern of habitat use. In movement ecology, temporal patterns have been addressed, for example, by exploration of temporal autocorrelation of movement parameters (Wittemyer *et al.* 2008, Dray *et al.* 2010). We wish to draw attention on the meaningfulness of temporal patterns when describing habitat choices. For example, the same proportion of habitat use in a certain time interval, may correspond to very different sequential patterns. An animal may continuously use a single habitat type, then switch to another, or, in contrast, alternate the use of both. Such spatio-temporal patterns may correspond to alternative space use tactics, and find a deep ecological significance. Currently, insights in the spatial patterns of use of multiple habitat-related variables are easy to obtain, for example using suitability maps (e.g. Calenge 2006); however, very few methods provide insights into spatio-temporal patterns combined. An interesting publication in that direction comes from Benhamou & Riotte-Lambert (2012) presenting a framework using movement-based kernel density estimation (utilisation distribution) and computation of residence time combined to explore areas of intensive use. Here, we are interested into methods to visually explore the sequential and thus temporal structure of habitat use.

In other research areas, such as geovisual-analytics, important progress has been made in terms of visually exploring sequential data at variable spatio-temporal scales (Andrienko *et al.* 2011). Buchin *et al.* (2012), for example, developed a geometric algorithm for trajectory clustering that takes into account environmental context parameters such as temperature and habitat type. In sociology, on the other hand, the link between sequential order of human behaviour and space use has been investigated using Sequence Analysis Methods (SAM) (Abbott 1995). This technique is principally used in the field of bioinformatics to evaluate the degree of similarity among DNA or protein sequences, but has also been applied successfully in transportation science (Wilson 2008), tourism research (Shoval & Isaacson 2007) and indoor navigation (Delafontaine *et al.* 2012).

Sequence Analysis Methods to our knowledge has never been applied to explore spatio-temporal patterns in sequential habitat use by animals. The essence of this approach is the possibility to ‘extract’ ordered sequences of habitat classes occupied along trajectories by means of clusters, which can be conveniently visualised in trees and validated by measures of statistical reliability. Moreover, SAM allows to deal with two common issues of GPS-based location datasets: missing data points (i.e. acquisition failures by GPS sensors, Frair *et al.* 2010), and spatial correlation (Dray *et al.* 2010).

In this study, we aimed to evaluate the applicability of SAM to movement ecology data for exploratory purposes by analysing both simulated trajectories and time-stamped locations of individually tracked roe deer (*Capreolus capreolus*) from a partially migratory alpine population. The analysis consisted of several steps (Fig.3.1). We first produced a classification tree based on bimonthly sequences of habitat use by individual roe deer. We used this first exploratory classification to hypothesise potential patterns of sequential habitat use. Then, we produced simulated trajectories with those patterns, at different proportion of habitat availability, and classified them in trees. We then re-added the real trajectories to simulation trees, while accounting for their relative proportion of habitat availability. Finally, we evaluated the biological relevance of such classification on the basis of ecological predictions. This way, we explored spatio-temporal patterns of real trajectories and evaluated them in a simulated experimental setting.

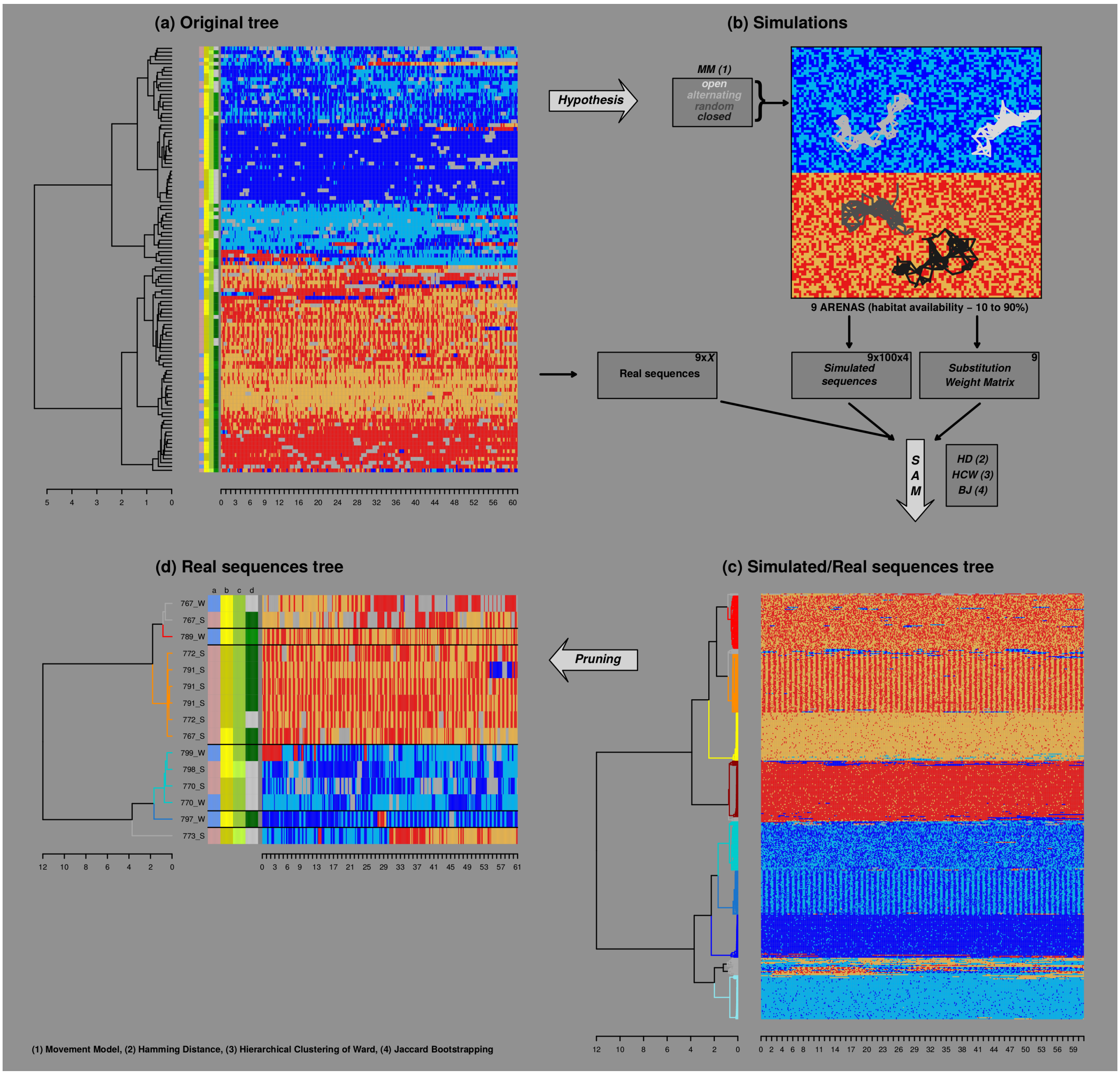


Figure 3.1. Flowchart of the complete process to explore spatio-temporal habitat use patterns of real trajectories by sequence similarity analysis. In essence, based on simple classification trees of real trajectories (a), we simulated trajectories with a-priori defined patterns of sequential habitat use (b), that in turn were used to produce classification trees accounting for spatial correlation, at different proportion of habitat availability (b/c). We then introduced the real trajectories (c), accounting for the relative proportion of habitat availability, and assessed their classification into tree clusters resulting from the simulation exercise (c/d).

3.2. MATERIAL AND METHODS

3.2.1. Study area and real trajectories

The studied animal population consisted of 26 European roe deer equipped with a GPS collar (GPS-Plus D, Vectronic Aerospace GmbH), of which 16 were females and 10 males. Six of them were collared as fawns (i.e. less than 1 year old; 2 females and 4 males), one as yearling (i.e. one female between 1 and 2 year old), while all others were collared as adults. Figure 3.2 pictures the cumulative sum of the 90% fixed-kernel home ranges (KDE) with reference smoothing parameter (href; Worton 1995) of individual roe deer using Home Range Extension (Rodgers & Carr 2002). This area extends across the Monte Bondone-Monte Stivo range, west of Trento and Adige valley and east of Valle dei Laghi, in North-east Italy (46°4'N, 11°7'E). Elevation ranges between 200 and 2300 m above sea level (m a.s.l.). Along this altitudinal gradient, climate is extremely varied, ranging from semi-Mediterranean and temperate (<1000m a.s.l. defined as 'low elevation') to semi-Alpine and Alpine (>1000m a.s.l. defined as 'high elevation') conditions. The study area is mainly covered by broad-leaved, coniferous and mixed forest (defined as 'closed habitat', representing 50% and 75% of high and low elevation, respectively), alternated by pastures (defined as 'open habitat', representing 50% and 25% of high and low elevation, respectively). Relevantly, the high-elevation range is mainly constituted by protected land, whereas the low elevation is not and is characterised by more anthropic land use.

The sampling period spanned from 2005 to 2008, during which GPS collars yielded a total of 54 845 time-stamped locations. The interval and duration of sampling were different among individuals, depending on date of capture and battery exhaustion. Locations were acquired at a pre-determined temporal interval Δt of 4 hours at fixed time stamps (0, 4, 8, 12, 16 and 20h) except for cold winter months (January and February), when Δt was 6h (0, 6, 12 and 18h) in order to save battery. We linked GPS locations to two geographical parameters retrieved from remote sensing sources: habitat openness [EEA-Corine Landcover (CLC) 2006 IV, European Environment Agency, EEA (2010), resolution = 100m] and elevation [NASA-ASTER GDEM, Ministry of Economy, Trade and Industry of Japan, METI, National Aeronautics and Space Administration of the United States, NASA (2012) = = 30m]. We combined and reclassified environmental parameters into four classes: high-closed (HC), high-open (HO), low-closed (LC), low-open (LO) (Table 3.1 and Fig.3.2). In this application, we used four classes to reduce the complexity of data analysis for demonstration. On the one side, the classes correspond to well-defined vegetation successional types in the alpine habitat of this population (Cagnacci *et al.* 2011); on the other side, they are meaningful for roe deer habitat use traits, since they are known to prefer ecotonal habitats and forest edges (Tufto *et al.* 1996).

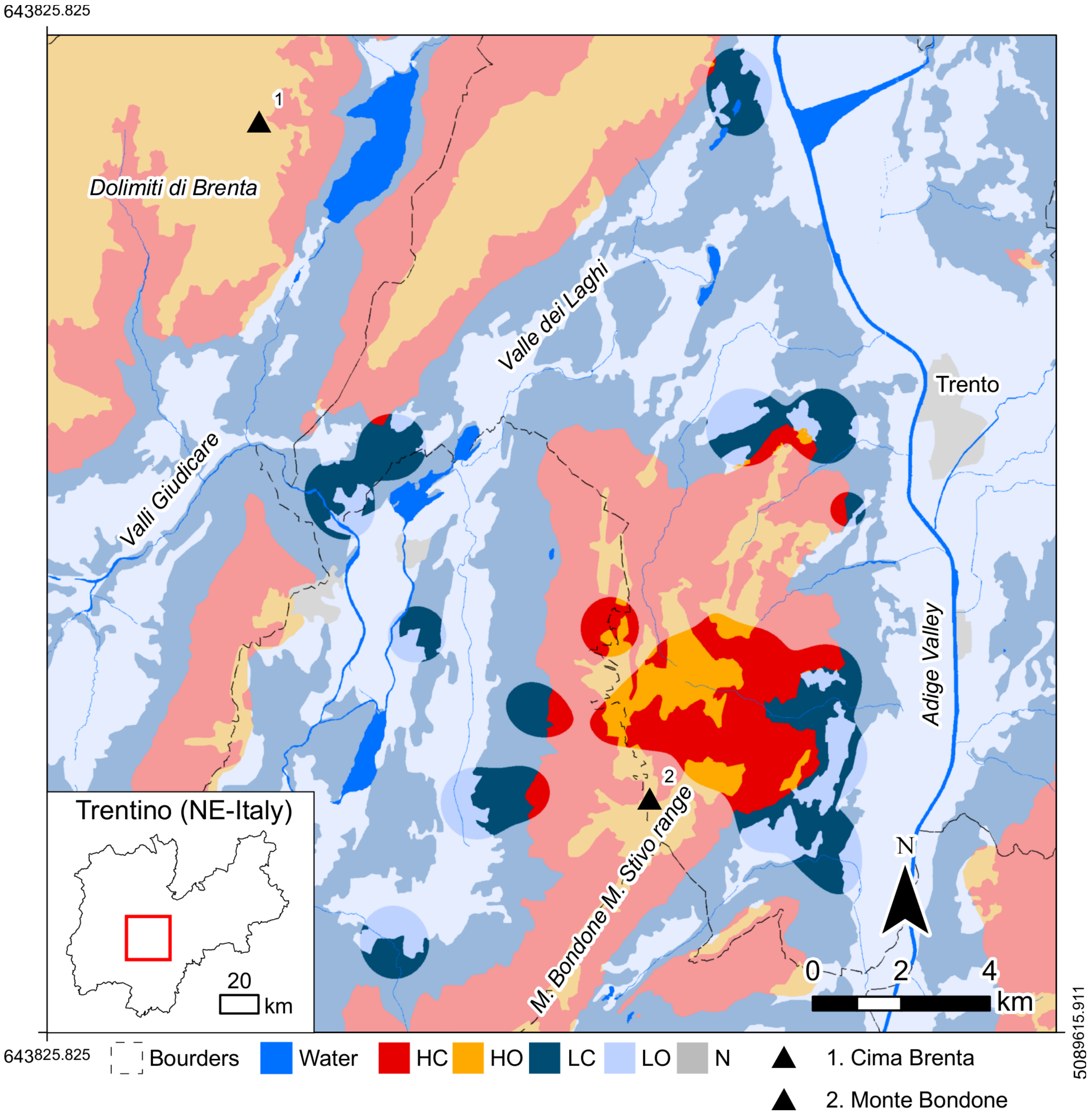


Figure 3.2. The study area, indicated by the darker irregular shaped polygons, was computed as the cumulative sum of home ranges of 26 individually marked roe deer (kernel density estimator, 90% polygons). For codes of habitat classes, see Table 1. N: Urban areas.

Then, we recoded sequences of locations into sequences of the environmental classes above, adding asterisks to account for missing locations due to acquisition failure. More precisely, the input sequences for SAM describe the habitat use by individual animals at regular time stamps (0, 4, 8, 12, 16 and 20h) over a period of two months. While SAM is able to deal with missing data, if they are too frequent they can over-fragment the sequence and thus bias the similarity measurement. To avoid a bias in the downstream analyses, we therefore excluded bimonthly sequences with more than 40% of missing data. After removal, the input file consisted of a total of 111 sequences (min 21, max 24 per bimonthly period), which were reclassified as belonging to summer (May-October) or winter (November-April) season, based on snowfalls and typical alpine climate (Ramanzin *et al.* 2007; Cagnacci *et al.* 2011). Only seven sequences could be retained for January-February, therefore we excluded the whole period from further analysis. Finally, sequences were associated with sex, age (fawn, yearling or adult) and migration occurrence (migrant, non-migrant, no data), of each individual. We obtained a first visualisation of the habitat use patterns represented in our sample by creating a simple classification tree of the 111 sequences (Fig.3.1a). This classification tree was based on Hamming distance (Gabadinho *et al.* 2011; see below for more details), without accounting for spatial correlation, nor habitat availability. Thus, this first classification tree does not allow to objectively classify the sequential use of habitat types, but can be used to build hypotheses for expected models of habitat use.

Table 3.1. Reclassification of the environmental parameters elevation and habitat openness resulting in combined classes (coded as HC, HO, LC and LO).

Elevation	Habitat openness	Habitat use classes
Low (<1000m)	Closed[1]	LC
Low	Open[2]	LO
High (>1000m)	Closed	HC
High	Open	HO

[1] Forest: Corine Landcover-classes 311, 312, 313, 323, 324

[2] No Forest: All other Corine Landcover-classes, except inland water

3.2.2. Simulation arenas and simulated trajectories

We simulated nine squared gridded arenas, each composed of 10 000 squared cells of 100 x 100 m mirroring the availability of environmental classes present in the real settings (Fig.1b, Appendix 3A). Since elevation classes are highly clustered in our study area (see Fig.3.2), we split each arena in two equal parts, corresponding to high and low elevations. We obtained the final four categories by randomly assigning all cells to either open or closed habitats, with varying proportions from 10 to 90% in each arena. Thus, simulated landscapes covered all possible habitat prevalence, which may occur within individual home ranges in this specific study area.

Within each arena, we generated simulated sequential habitat use using a simple spatially-explicit stochastic movement model (Fig.3.1b, grey box ‘MM’; Appendix 3B). For the simulated trajectories, we chose 4 different patterns of habitat selection, based on preliminary observations on the classification tree of real trajectories (Fig.3.1a), and previous knowledge on roe deer ecology. We thus distinguished homogeneous use of closed or open habitats, random and alternating use (i.e. day-night patterns) of open and closed habitats. The random pattern represented the ‘control’ in our simulated experimental settings. The homogeneous closed and alternation between closed and open were the expected patterns according to roe deer ecology, and specifically the known preference for forest and ecotonal habitats (Tufto *et al.* 1996). Homogeneous open represented the alternative hypothesis. We simulated 100 trajectory replicates for each of the nine arenas and four behaviours of habitat selection (total of 3600 sequences, Fig.3.1b, grey box ‘simulated sequences’). Because sequential use of elevation was strongly dependent on the release location of the simulated agents, each of the 100th set of simulations had the same seed random locations across arenas and behaviours (i.e. trajectories had the same seed in groups of 36). We then trimmed the simulated trajectories to match the length of the real bimonthly roe deer trajectories i.e. 366 steps. Finally, simulated sequences of habitat use were extracted from these calibrated trajectories.

3.2.3. General procedure of sequence similarity analysis for calculation of dissimilarity trees

Sequence Analysis Methods are based on sequence similarity measures that are used to identify groups of sequences showing similar behaviour. The input of such analysis always relies on a *dissimilarity matrix*, which provides the dissimilarity, or ‘distance’ among all possible pairs of sequences. Among the available distance measures, we chose Hamming distance (HD) to ascertain the dissimilarity matrix, as it is considered the most suitable for sequences with a temporal dimension. HD relies solely on two operations: identity and substitution, and in fact, it computes the minimum number of substitutions to equate a number of sequences of equal length (Gabadinho *et al.* 2011). In a more optimised HD, also weights can be assigned to substitutions: that is, HD computation can be based on a *substitution weight matrix*. Figure 3.3 gives a conceptual example of two alternative HD of the character strings 'Kapreolo' and 'Capriolo', respectively the word for roe deer in Esperanto and Italian. Both distinguish six identities and two substitutions, but differ from each other in weights assigned to substitutions. The substitution between the letters K and C gets a lower weight (i.e. probability) in 'a' (substitution score = 0.4, Fig.3.3a) than in 'b' (substitution score = 1, Fig.3.3b), since the former HD takes into account the phonetic similarity. Consequently, the total dissimilarity in 'a' will be lower than in 'b'.

	K	*A*	*P*	*R*	**E**	*O*	*L*	*O*	
	C	*A*	*P*	*R*	**I**	*O*	*L*	*O*	
a	0.4	0	0	0	1	0	0	0	**1.4**
b	1	0	0	0	1	0	0	0	**2**

Figure 3.3. Two alternative computations of the Hamming distance, based on substitution weights given to each operation performed to equate the sequences (reported under each character). Identities are in italic and substitutions bolt.

In our case, the dissimilarity matrix computed by the HD algorithm was based on substitution weights for all combinations of habitat classes, and constrained to their availability and distribution, which together determine patterns of spatial correlation. Therefore, *we derived the substitution weights from spatial correlation* of habitat classes in the simulated arenas (Fig.3.1b, grey box ‘substitution weight matrix’, see Appendix 3C for computation and simulations of the substitution weights).

The successive step in data analysis is using the HD dissimilarity matrix to calculate a dissimilarity tree. Here, we used Ward's method (Fig.3.1b/c, HCW), the most common hierarchical agglomerative cluster procedure (Gabadinho *et al.* 2011). The resulting trees are a representation of the dissimilarity among the habitat use sequences of animals. The distance, or ‘branch length’, between leaves (which stay for individuals) and nodes (which stay for groups of individuals) can be used as a proxy for the dissimilarity between the portions of the tree: the longer the distance, the higher the dissimilarity.

The following step is the identification of ‘clusters’, or portions of the tree indicating association between sequences. This is the most important step to use trees for exploratory purposes and the subsequent interpretation of results. We determined the number of clusters present in a tree by a cut-off distance, based on a repeatable bootstrapping procedure (Fig.3.1b/c, BJ). For this application we performed 1000 iterations and calculated the Jaccard bootstrapping index (i.e. bootmean, Hennig 2007) for a cut-off distance that separated from two to twenty clusters. This index compares the similarities of the original clusters to the most similar clusters in bootstrapped datasets. The cut-off distance was determined as the maximum number of clusters where the median bootmean is higher than 0.9, a value corresponding to highly stable clusters. This threshold is conservative in a sense that normally values above 0.75 correspond to robust classification.

3.2.4. Determining classification trees of simulated trajectories

Using this procedure, we computed a dissimilarity tree and identified clusters for each simulated landscape by using the 400 simulated sequences (100 replicates x 4 patterns of sequential habitat use, Fig.3.1c). To account for spatial correlation of the four habitat classes, we recomputed the substitution matrix for each landscape. In fact, although all grid cells were assigned to habitat classes (open/closed) at random, this still corresponded to different spatial correlation structure across arenas, an effect of the different proportion of classes. Then, we investigated the sequence composition of individual clusters distinguished by the classification trees to assess the liability of the method to *correctly group sequences of different sequential behaviour*. This also allowed us to identify the name of each cluster.

3.2.5. Determining spatio-temporal patterns of real trajectories

In the next step we introduced the real trajectories into one of the classification trees (Fig.3.1c), according to the relative proportion of habitat availability as follows. For each real bimonthly sequence, we measured the availability of open/closed habitat within the corresponding bimonthly home range. We then associated them to the simulated trajectories referring to the arena with the same habitat proportion. For example, if an individual bimonthly home range showed an open habitat availability between 45 and 55%, the corresponding sequences were associated to the simulated sequences originated from the 50% arena. Then, each tree was recalculated for the combined set of sequences (simulated and corresponding real sequences). Using this approach, simulation sequences could be used as a guide for classification of real sequences to their most similar sequential behavioural group (Fig.3.1c/d).

3.2.6. Expected spatio-temporal patterns of real trajectories

Based on previous knowledge on roe deer ecology, and the individual descriptive variables, we formulated predictions of tree clustering. This was the core of our study, to assess the meaningfulness of SAM for exploration of spatio-temporal sequences of ecological data. In roe deer populations of northern and alpine environments, some individuals reach higher elevations in summer, when habitat suitability increases, and return at lower elevation in rigid winter conditions. Other individuals, instead, occupy the same low-elevation range all year round (Ramanzin *et al.* 2007; Cagnacci *et al.* 2011). This phenomenon is known as partial migration. On these premises, we expected individuals to classify into two main clusters according to the use of elevation: animals with a constant use of the same elevation range (winter and summer sequences), and animals with a seasonal shift in elevation range associated to migration (winter sequences separated from summer sequences) (P1a). Roe deer reproductive season is concentrated in summer, when both males and females exploit the best environmental conditions to meet the high energetic demand of mating and giving birth (Hewison *et al.* 1998). If migration is linked to habitat quality, we expect both sexes to show similar patterns of migration and thus of elevation range use (P1b, but see Ramanzin *et al.* 2007). Likewise, since fawns are not yet engaged in reproduction, we may expect a lower rate of migrating individuals (P1c). Alternatively, they may follow the mother in the migrating movements. Predictions on sequential use of open and closed habitats are less straightforward. Roe deer are known to prefer habitats providing cover and protection, especially intermediate stages of forest succession and ecotonal habitats (e.g. Tufto *et al.* 1996). Therefore, we predict that animals would show a sequential use of habitats different from random (P2a). In particular, we predict a separation between animals using only closed habitat and those showing a combined use of closed and open habitats (P2b). When animals use both habitats, we expect open habitats to be used mainly during night, due to the anti-disturbance and anti-predatory behaviour of roe deer, translating in an alternating sequential use of open and closed (P2c) (Saïd & Servanty 2005). For similar reasons a constant use of open habitat is instead less likely (P2d).

3.3. RESULTS

3.3.1. Classification trees of simulated trajectories

The application of HD algorithm to simulated bimonthly sequences generated 9 trees, each corresponding to landscapes with different habitat availability (i.e. T_{10} to T_{90}; see Fig.3.1c and Appendix 3D for all trees). In all trees the topological relationships between tree branches indicated two main orders of classification: first-order clusters, splitting the sequences in two groups according to the preferential use of different elevations (C_1 and C_2; e.g. Fig.3.1c), and several second-order clusters separating animals with different sequential use of open and closed habitats (C, R, A, O, U; see text below, Fig.3.1c and Appendix 3D: second-order clusters are distinguished by coloured branches).

At first-order, all trees showed a significant separation (Jaccard bootstrapping index $B_J>0.95$) between high and low elevation trajectories, thus defining a 'high-elevation' cluster (C_1) and a 'low-elevation' cluster (C_2). At second-order, our defined bootstrap threshold (i.e. $B_{J,median} > 0.9$) identified 6-8 main clusters, (T_{10}-T_{30} & T_{70}-T_{90}= 6 clusters, with 40-80 sequences in each cluster; T_{40}-T_{60}= 8 clusters, with about 40 sequences in each), and 3-7 small clusters (with less than 10 sequences each). Separation of the main second-order clusters was significant (most clusters $B_J > 0.95$), whereas for smaller clusters, it was not (most clusters $B_J < 0.75$). For both high and low elevation, the main second-order clusters corresponded to different patterns of sequential open/closed habitat use. The classification of sequences in such clusters was highly dependent on habitat availability in arenas. For trees deriving from arenas with similar open/closed habitat proportion (40-60%: T_{40}, $T_{50,}$ T_{60}, Fig.3.1c, Figs 3D.4-3D.6 in Appendix 3D), we distinguished all four simulated sequential behaviours: homogeneous closed (C; brown and dark blue branches), random (R; red and cyan branches), alternating (A; orange and blue branches) and homogeneous open (O; yellow and light blue branches). In this case, 99% of sequences in each cluster was of the same sequential behaviour. For example, all sequences in the blue cluster of figure 3.1c are homogeneous closed. Conversely, for landscapes where habitats were disproportionally available (Open < 30% or Open > 70%: $T_{10,}$ T_{20}, T_{30} and $T_{70,}$ T_{80}, T_{90}, Figs. 3D.1-3D.3, 3D.7-3D.9 in Appendix 3D), random sequences clustered with homogeneous sequences of the dominant habitat (C with R: brown and dark blue branches; O with R: light blue and yellow branches). The random sequential use of habitat according to availability (i.e. the control case) was therefore effectively represented by a separated random cluster only when sequences were not trivial (e.g. T_{10} and T_{90} obviously led to homogeneous sequences ‘at random’).

Finally, small clusters, corresponded to trajectories indicating a mixed use of high and low elevation, for a specific pattern of sequential use of open and closed habitats (e.g. mixed-alternating). Alternatively, small clusters were undefined (U) due to a too small number of sequences (<5 sequences), and can be considered as outliers. Both mixed and undefined clusters were coloured grey in trees (Fig.3.1c, Appendix 3D).

3.3.2. Spatio-temporal patterns of real trajectories and discussion of the study case

a. Sequential habitat use patterns

We used the trees based on mixed simulated and real trajectories to assess the classification of real trajectories according to sequential habitat use. To visualise how real trajectories were classified, we extracted (i.e. pruning) the real sequences from nine different trees (Fig.3.1d and 3.4, see also Appendix 3E for a different visualisation). The first-order clusters C_1 and C_2 distinguish very well high- from low-elevation sequences (Fig.3.4, Appendix 3E right *vs* left panels).

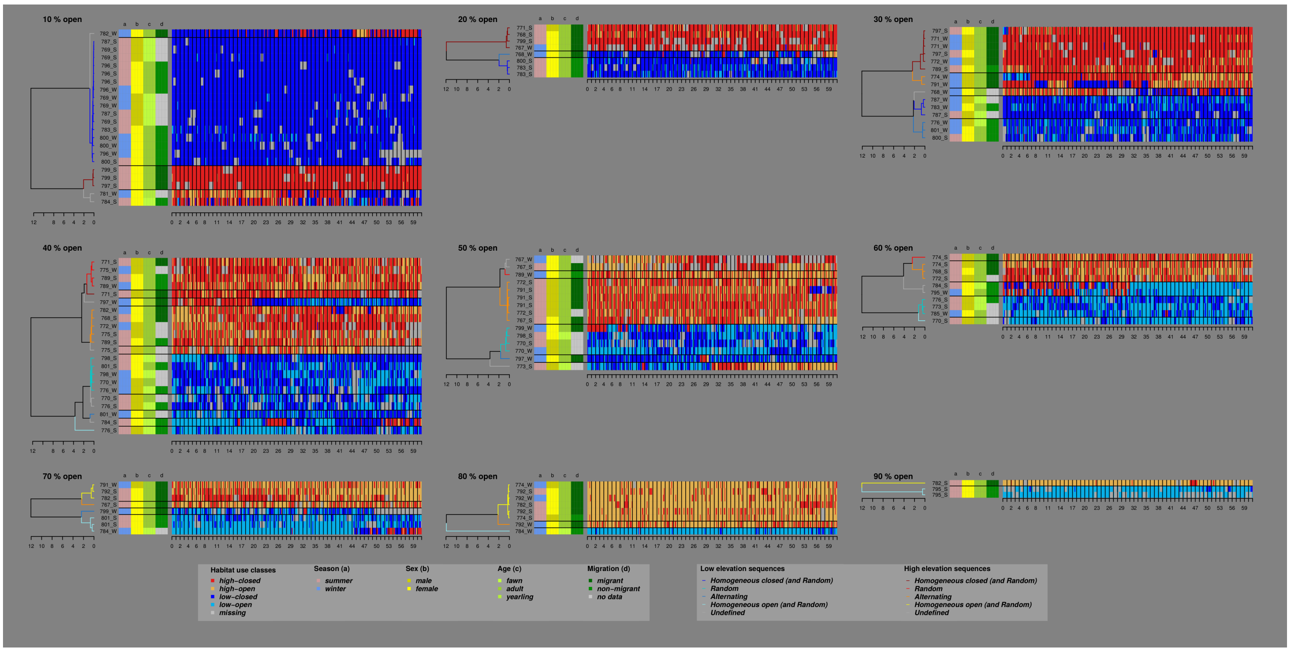


Figure 3.4. Roe deer bimonthly sequences of habitat use classes (right portion of the nine panels; daily scale reported bottom-right), and resulting dissimilarity trees, based on the Hamming distance (left portion of the panels). The real sequences were extracted (pruned) from their corresponding simulated tree from which sequences were derived from a trajectory running over nine different arenas with varying habitat proportions (10-90 %). Tree 'leaves' represent a real sequence, while 'nodes' indicate their clustering. Branch lengths (distance between leaves, and first common node: bottom-left for the scale) indicate the dissimilarity between individual sequences. The split into two main clusters define first-order clusters separation, based on use of elevation classes (high, C1; low, C2). Second-order clusters are based on use of open/closed habitats and distinguish homogeneous closed (C, brown and dark blue branches), random (R, red and cyan branches), alternating (A, orange and blue branches) and homogeneous open (O, yellow and light blue branches). Grey branches (U) are sequences with undefined classification or clusters with mixed sequences of high and low. The id gives the animal code and season (Summer, S; Winter, W). Variables season, age, sex and migration are represented as colour-coded bars between trees and sequences (see legend for meaning of the colours).

More interesting is the classification of sequences in different sequential patterns of open/closed habitat use (second-order). Within our roe deer population, we found evidence of the four different patterns of open/closed sequential use that have been addressed in our simulations. Specifically, 98 sequences out of 111 were classified according to those patterns (88.3%). Moreover, some interesting differences emerged in sequential use of open/closed habitats for different elevations (Fig.3.4, Appendix 3E). At high elevation, half of the sequences have an alternating (C_1A: 32%, 18 sequences) or homogeneous closed pattern (C_1C: 25%, 14 sequences), whereas only 11% of the sequences are random (C_1R: 6 sequences). At low elevation, 63% of the sequences are homogeneous closed (C_2C: 41%, 22 sequences) or random (C_1R: 22%, 12 sequences), and conversely alternating sequences are uncommon (C_1A: 13%, 7 sequences). Surprisingly, both at high and low elevation, there are also homogeneous open sequences (C_1O: 18%, 10 sequences; C_2O: 14%, 8 sequences). At the study area scale, different sequential patterns of open/closed habitat use according to elevation are possibly linked to those habitats availability in the study area (Figure 3.2: open more available at high elevation). Indeed, individual habitat use sequences are obviously related to the availability of habitat classes within the home range (Fig.3.4). In particular, sequences classified as homogeneous sequences (C and O) derived from home ranges with a large proportion of one specific habitat type (Fig.3.4.1-3.4.3, <30% open; Fig.3.4.7-3.4.9, >70% open). Instead, heterogeneous sequences showed an alternating or random pattern (A and R), and derived from home ranges where both habitat types were available (Fig.3.4.4-3.4.6, >40% open and <60% open). Interestingly, though, both homogeneous and heterogeneous trajectories lay next to each other in the study area (Fig.3.5). Also, several sequences with the same home range availability were assigned to different sequential patterns. For example, 776_W was classified in C_2A and 787_S in C_2C, when they both have 30% open in home ranges (Fig.3.4.3). Or, 771_S was classified in C_1C and 789_S in C_1R, when they both have 40% open in home ranges (Fig.3.4.4).

Finally, sequences switching between high and low were classified in the mixed_alternating cluster (only one sequence) or in the undefined clusters (i.e. as an outlier) (Fig.3.4, C_1U: 14%, 8 sequences; C_2U: 9%, 5 sequences). Notably, C_1U mainly consists of sequences changing from high to low elevation (e.g. Fig.3.4.4, 797_W; Fig.3.4.6, 784_S, 795_W), whereas C_2U shows the opposite pattern (e.g. Fig.3.4.5, 773_S), as a result of the migration between winter and summer ranges (Fig.3.5). Alternatively, undefined sequences may correspond to sequential behaviours we did not simulate, such as the shift from homogeneous closed to homogeneous open sequential habitat use within the bimonthly period (e.g. Fig.3.4.5, 767_W).

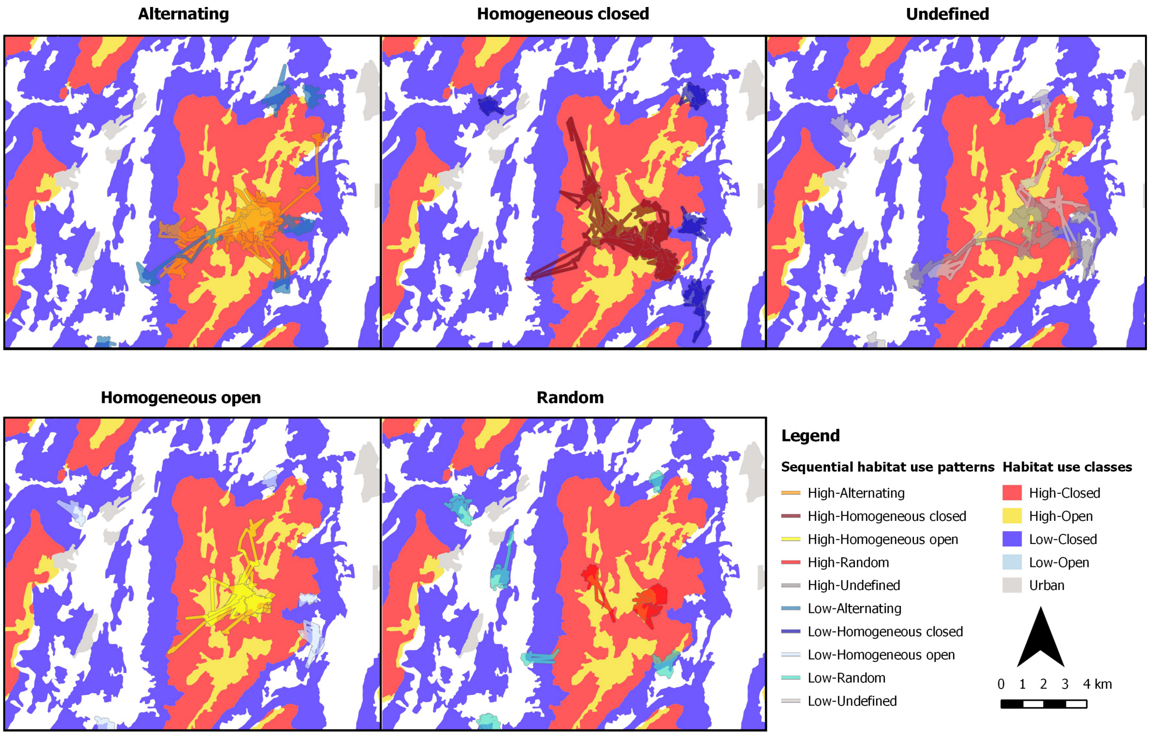


Figure 3.5. Clusters of real bimonthly trajectories with similar patterns of sequential habitat use in their spatial context. Legend in bottom-right panel.

b. Sequential habitat use vs descriptive variables

Descriptive variables of individual sequences, in particular migration occurrence, sex and age enabled to further interpret the classification of trajectories in first- and second-order clusters. In winter, individual trajectories were classified both at high and at low elevation, regardless the space use strategy (i.e. both migrants and non-migrants; see Fig.3.6a and 3.4). Conversely, in summer migrants' trajectories were always classified at high elevation and non-migrants' mainly at low elevation. Sequences that were classified as alternating mainly belonged to migrants, whereas homogenous open sequences equally corresponded to migrant and non-migrant individuals (Fig.3.6c and 3.4). Finally, more than 60% of the non-migrant sequences were either classified as homogeneous closed or random.

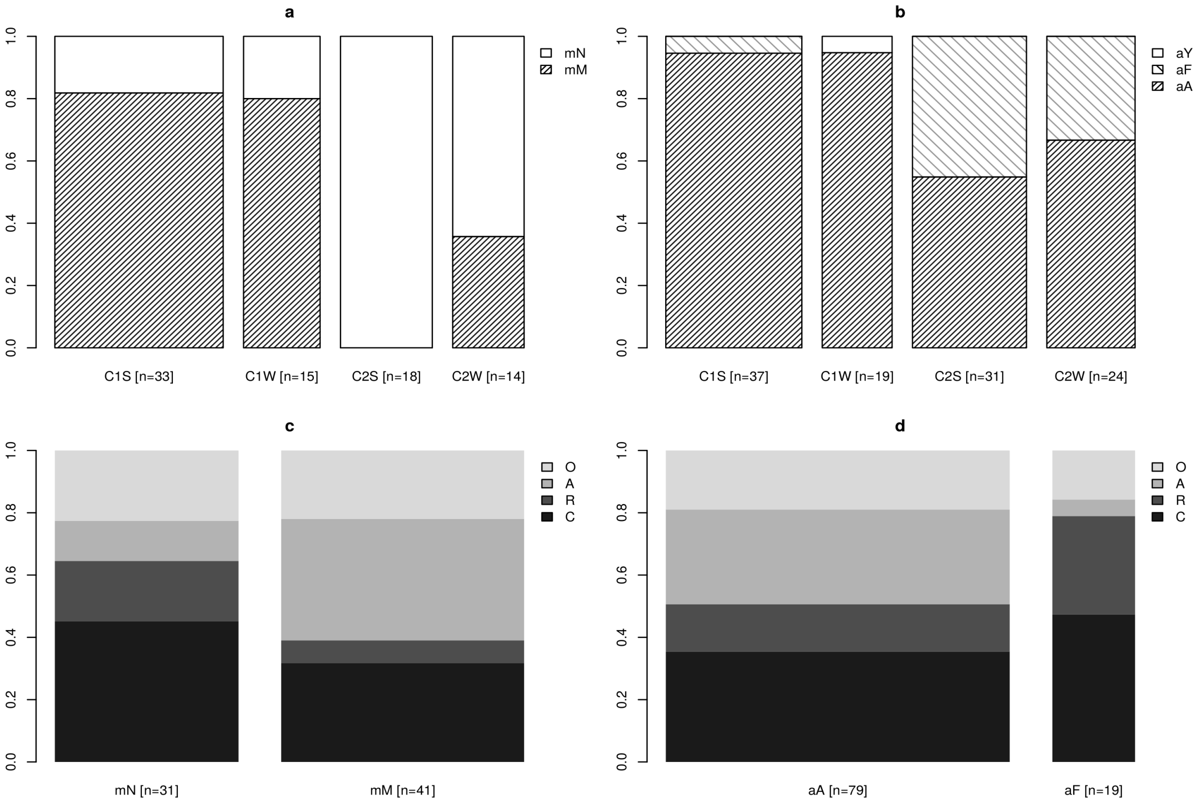


Figure 3.6. Individual descriptive characteristics are summarised by the clusters to which real individual sequences were assigned. Migration (%; panel a; migrant-mM, non-migrant-mN, no data-mX) and age (%; panel b; adult-aA, yearling-aY, fawn-aF) in relation to first-order clusters (high, C1; low, C2), both in Summer and Winter. Second-order clusters (%; C, R, A, O; see legend, figure 3.4 for meaning of codes) in relation to migration (panel c) and age (panel d).

Interestingly, trajectories of fawns were classified at low elevation, both in summer and winter, except one sequence (Fig.3.6b and 3.4). In terms of second-order clusters, trajectories of fawns were classified in the great majority as homogeneous closed or random (Fig.3.6d and 3.4). Adults' sequences were more equally distributed between different patterns of habitat use.

c. Assessment of predictions and case study discussion

The first clear spatio-temporal pattern that emerged in the study of roe deer population through exploratory analysis with SAM is a differential use of altitude. This can be linked to *migration* for two pieces of evidence: first, some individuals with sequences classified at high elevation in some trees, also showed sequences classified at low elevation in other trees (Fig.3.4, e.g. individual 799). Second, some sequences from the same individuals included both high and low altitude in the migration periods, when animals move between seasonal ranges (Fig.3.4, e.g. individual 797, Fig.3.5-undefined). Moreover, the classification of sequences in high and low elevation clusters can be attributed to *partial migration* for a further evidence; that is, sequences of some other individuals were always included in the low-elevation cluster (Fig.3.4, e.g. individual 783), or always across both (Fig.3.4, e.g. individual 784), that is, they were resident. As such, prediction P1a is supported.

The property of SAM as a valid exploratory tool to identify spatio-temporal patterns of individual movements were also highlighted when looking at descriptive variables. Sequential use of high and low elevation was same among sexes (Fig.3.4: sequences of both sexes were included in both clusters; P1b), but not across age classes (Fig.3.6 and 3.4; P1c). This is a novel result for roe deer, that opens up new directions of research. For example, age class could be included as an explanatory variable to assess partial migratory behaviour of roe deer.

The most innovative outcome of our analysis was the spatio-temporal analysis of open/closed habitat use by individual roe deer. The most obvious, and yet relevant, result was that roe deer did not always use open/closed habitat at random (P2a). As predicted, sequences of individuals were clearly distinguished between homogeneous and heterogeneous use or shifted between the two categories. We intend to underline that by SAM exploratory analysis, we always looked at sequential *use* of habitat; that is, habitat selection was not considered here. The term ‘random’ use, therefore, does not refer to habitat use according to total habitat availability, as in classic habitat selection analysis (Manly *et al.* 2002), but to sequential use of available habitats as it would happen by chance. Indeed, we found that sequential use of habitats was heavily influenced by availability in the home range. For example, home ranges including a high percentage of closed habitat likely resulted in a ‘random’ homogenous use of closed (Fig.3.4.1-3.4.3, Fig.3.5-homogeneous closed). Conversely, home ranges including both open and closed habitat may lead to ‘random’ use of both (Fig.3.4.4-3.4.6; Fig.5-random). Here, we have two observations. On the one side, the fact that trajectories with a different spatio-temporal pattern laid next one to each other in the study area (Fig.3.5) indicates that sequential habitat use reflects the selection operated by animals to define their home ranges within the study area (i.e. second-order habitat selection, *sensu* Johnson 1980). On the other side, though, we had clear evidence of *alternative* tactics of temporal use of habitats available within the home range (several examples of sequences with same total habitat availability, but different sequential patterns, see above and Fig.3.4). When both open and closed were available in home ranges, one pattern stuck out as particularly well represented, i.e. the alternating use of both habitats on a daily basis (P2c; Fig.3.4 and 3.5). To our knowledge, this is among the first systematic exploratory analyses of a behavioural pattern previously empirically described, and linked to individual personalities (Bonnot *et al.* 2015). According to our results, the propensity of roe deer to expose to open habitats can result in a systematic spatio-temporal pattern of habitat use, which is evident only in some individuals (Fig.3.4). Even more, our explorations indicated that migrant individuals used this pattern much more than resident, at equal habitat availability in the home range (Fig.3.4.4-3.4.6 and 5-alternating). Migration attitude in partial migratory populations has been previously attributed to a 'boldness' syndrome (Chapman *et al.* 2011). Investigating the effect of personalities at both seasonal and daily habitat use scale is a very exciting research direction that our exploratory results suggest.

A further result that supports the individual differences in sequential habitat use is the presence of homogenous open sequences, an unexpected and yet observed pattern both at high and low elevation (P2d; Fig.3.4).

3.4. DISCUSSION: APPLICATION OF SAM TO ANIMAL HABITAT USE

In this paper, we showed that SAM is a useful and powerful tool to explore and compare sequences of habitat use by animals, and extract common spatio-temporal patterns. Importantly, we took into consideration an aspect of animal ecology largely overlooked in literature. Indeed, one of the most interesting and captivating outcomes of SAM is that different topological levels of trees are associated to hierarchical similarities between the individuals' sequential use of environmental features. Repetitive patterns of sequential habitat use are informative with respect to the interaction between animals and their environment. In fact, well-known phenomena such as migration, or feeding-resting cycles emerge as repetitive patterns of habitat use at different temporal scales, as shown by our study case.

To understand the potential usefulness of SAM for movement ecology, we shall first clarify what SAM *is not*. SAM *is not* a spatial explicit method and *does not* provide predictive models of habitat use, nor of animal distribution. Instead, SAM embeds the temporal component of habitat use, in the form of real ordered sequences of used habitat classes. SAM can provide information at the *population and individual level*, by clustering individuals in trees through robust algorithms that search for dissimilarities in spatio-temporal patterns of habitat use. Based on all these considerations, we see SAM as an approach to *explore* temporal patterns in habitat use across an animal population. As such, it differentiates from and it complements current exploratory methods in habitat analysis. The R package AdehabitatHS provides a suite of niche-based methods that are primarily used to express the realised niche (Calenge & Basille 2008). However, niche-based methods are also meaningful to identify the most relevant explanatory variables, and their relationships, for predictive habitat selection models, such as RSF (Boyce *et al.* 2002). RSF model the probability of disproportional habitat use in a hypothetical-deductive framework, by means of selection of competing models. A preliminary investigation of most relevant variables thus allow to better express hypotheses, and focuses the analysis. SAM may complement the aforementioned approaches by also exploring the temporal component of habitat use patterns. A big advantage of niche-based methods or recursive partitioning trees is their ability to provide a graphical representation of the importance and relation between variables. Similarly, SAM summarises common patterns of sequential habitat use across the population, by clustering. Importantly, we suggest that the way to handle the length of the sequences (i.e. time resolution and total duration) should depend on the research questions. In the study case, we decided to split the individuals' sequences in bimonthly periods, since we were interested in intra-annual patterns. This design implies that each individual is present more than once in the trees, and caution must be taken when interpreting the results (i.e. pseudo-replication Tukey, 1977).

In our exercise, we considered a simple combination of habitat classes, although the R package TraMineR allows to define more complex combinations of environmental parameters. All the same, sequences based on many habitat classes, would lead to a very articulated dissimilarity tree, thus likely difficult to interpret. We therefore suggest a rationale in the combined and complementary use of the aforementioned methods, when exploring spatio-temporal patterns in animal movement data. First, niche-based or recursive partitioning methods or other simpler multivariate approaches (e.g. PCA) can be applied to identify the most important and least related environmental parameters. Then, ordered sequences of locations can be matched to those variables by spatial join with relevant geographic layers, and sequences of habitat use thus obtained. Sequence Analysis can then be performed, and provide a representation of existing spatio-temporal patterns, or be used to formulate new hypotheses evaluated through a classic model selection approach.

Technically, we suggest to simulate expected sequential behaviours (e.g. based on preliminary classifications or previous knowledge, Fig.3.1a), and use them as a guide to extract the sequential habitat use pattern of real sequences, while accounting for spatial correlation (through the substitution matrix) and habitat availability. As our study case shows, real-world sequences can exhibit more complex behaviours than those represented by simulations. We suggest that if the great majority of real trajectories are classified in cluster types derived from simulations, then the real sequential habitat use is well represented. Otherwise, one may want to change the simulation rules, which can be easily modified in the movement model.

Sequence Analysis Methods is a well-suited method for data acquired by animal-borne tracking technologies, since the method can account for two main limitations related to animal movement data: missing locations (Frair *et al.* 2010) and spatial correlation (Dray *et al.* 2010). We explicitly remarked for the first time that the weight matrix used by SAM can be used to deal with spatial correlation. Specifically, we offer a quantifiable and repeatable assessment of the spatial correlation between different habitat classes (i.e. substitutions matrix). Notwithstanding the novelty, future studies may assess the sensitivity of SAM output to changes in the substitution matrix.

While the framework presented in this paper is relevant and innovative in ecological studies, in practice there are also some limitations to SAM. First, temporal correlation is not directly accounted, whereas there is a clear temporal association between consecutive observations. To account for temporal correlation, a variant of the Hamming distance known as the Fuzzy Hamming distance (Bookstein *et al.* 2001) could be used, but is only developed for binary data. Secondly, several researchers in evolutionary biology (e.g. Morrison 2010) and social sciences (e.g. Wu 2000) have argued that many steps in SAM are based on subjective decisions (e.g. definition of classes, parameter settings, interpretation of results).

These limitations notwithstanding, we believe that SAM offers great advantages and new insights in movement ecology studies. For instance, the method can be promptly extended to multiple or other species than roe deer as well as to other spatial (e.g. home range) and temporal resolutions (e.g. hours) and reveal yet underappreciated or overlooked patterns.

DECLARATIONS

Acknowledgements

We are grateful to MOVE (COST Action IC0903) to provide funding for two Short-Term Scientific Missions (STSMs) to conduct this research. Also we thank Fondazione Edmund Mach and Special Research Fund (BOF) of Ghent University for their financial support to JDG. Finally, we wish to thank the collaborative EuroDEER project to provide geospatial and environmental data. We are particularly grateful to Ferdinando Urbano for his support with data management, to Matteo Marcantonio for precious suggestions in R programming and to Matthias Studer for advice on TraMineR. The movement model was developed by Nathan Ranc under guidance of Prof. dr. Paul Moorcroft and Dr. Francesca Cagnacci.

Data accessibility

We provide the R script (mee312453-sup-0006-AppendixS6.r[5]) to repeat the Sequence Analysis as here presented. The full dataset is made available in the Dryad digital repository (De Groeve *et al.* 2015).

---

[5] https://besjournals.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2F2041-210X.12453&attachmentId=116121544

REFERENCES

Abbott, A., 1995. Sequence analysis: new methods for old ideas. *Annual Review of Sociology*, **21**, 93-113.

Andrienko, G., Andrienko, N. and Heurich, M., 2011. An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science,* **25**, 1347-1370.

Benhamou, S. and Riotte-Lambert, L., 2012. Beyond the Utilization Distribution: Identifying home range areas that are intensively exploited or repeatedly visited. *Ecological Modelling*, **227**, 112-116.

Bonnot, N., Verheyden, H., Blanchard, P., Cote, J., Debeffe, L., Cargnelutti, B., Klein, F., Hewison, A.J.M. and Morellet, N., 2014. Interindividual variability in habitat use: evidence for a risk management syndrome in roe deer? *Behavioral Ecology*, **26** (1), 105-114.

Bookstein, A., Klein, S.T. and Raita, T., 2001. Fuzzy Hamming distance: a new dissimilarity measure. *Combinatorial pattern matching*, **2089**, 86-97.

Boyce, M.S., Vernier, P.R., Nielsen, S.E. and Schmiegelow, F.K.A., 2002. Evaluating resource selection functions. *Ecological Modelling*, **157**, 281-300.

Buchin, M., Dodge, S. and Speckmann, B., 2012. Context-Aware Similarity of Trajectories. *Proceedings of the 6th International Conference on Geographic Information Science (GIScience).* Berlin: Springer.

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

Cagnacci, F., Focardi, S., Heurich, M., Stache, A., Hewison, A.J.M., Morellet, N., Kjellander, P., Linnell, J.D.C., Mysterud, A., Neteler, M., Delucchi, L., Ossi, F. and Urbano, F., 2011. Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors. *Oikos,* **120**, 1790-1802.

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** (3-4), 516-519.

Calenge, C. and Basille, M., 2008. A general framework for the statistical exploration of the ecological niche. *Journal of Theoretical Biology*, **252**, 674– 685.

Chapman, B.B., Hulthén, K., Blomqvist, D.R., Hansson, L.-A., Nilsson, J.-Å., Brodersen, J., Nilsson, P.A., Skov, C. and Brönmark C., 2011. To boldly go: individual differences in boldness influence migratory tendency. *Ecology Letters*, **14**, 871-876.

De Groeve, J., Van de Weghe, N., Ranc, N., Ometto, L., Rota-Stabelli, O. & Cagnacci, F. 2015. Data from: Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods (SAM). *Dryad Digital Repository*. doi:10.5061/dryad.h4f7p.

Delafontaine, M., Versichele, M., Neutens, T. and Van de Weghe, N., 2012. Analysing spatio-temporal sequences in Bluetooth tracking data. *Applied Geography*, **34**, 659-668.

Dray, S., Royer-Carenzi, M. and Calenge, C., 2010. The exploratory analysis of autocorrelation in animal-movement studies. *Ecological Research*, **4**, 34-41.

European Environment Agency, EEA (2010) CORINE Land Cover 2006. URL http://www.eea.europa.eu/dataand-maps/data/corine-land-cover-2000-2006 (accessed 20 August 2015).

Fortin, D., Hawthorne, L.B., Boyce, M.S., Smith, D.W., Duchesne, T. and Mao, J.S., 2005. Wolves influence elk movements: behaviour shapes a trophic cascade in Yellowstone national park. *Ecology,* **86**, 1320-1330.

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

Gabadinho, A., Ritschard, G., Müller, N.S. and Studer, M., 2011. Analyzing and visualizing state sequences in R with TraMineR, *Journal of Statistical Software*, **40**, 1-37.

Hennig, C., 2007. Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis*, **52**, 258-271.

Hewison, A.J.M., Vincent, J.P. and Reby, D., 1998. Social organisation of European roe deer. *In:* Andersen R., Duncan P., and Linnell, J.D.C., eds. *The European roe deer: the biology of success.* Oslo: Scandinavian University Press, 189-219.

Johnson, D.H., 1980. The Comparison of usage and availability measurements for evaluating resource preference. *Ecology,* **61**, 65-71.

Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L. and Erickson, W.P., 2002. Introduction to Resource Selection Studies. *In: Resource selection by animals: statistical design and analysis for field studies*. Dordrecht: Kluwer Academic Publishers, 1-15.

Ministery of Economy, Trade, and Industry of Japan, METI, National Aeronautics and Space Administration of the United States, NASA (2012) Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 1 (GDEM V1). URL http://asterweb.jpl.nasa.gov/gdem.asp (accessed 20 August 2015).

Morrison, D.A., 2010. Sequence Alignment: Methods, Models, Concepts, and Strategies. *Systematic Biology*, **59**, 363-365.

Ramanzin, M., Sturaro, E. and Zanon, D. (2007) Seasonal migration and home range of roe deer (*Capreolus capreolus*) in the Italian eastern Alps. *Canadian Journal of Zoology*, **85**, 280-289.

Rodgers, A.R. and Carr, A.P., 1998. *HRE: The Home Range Extension for ArcView. Ontario Ministry of Natural Resources*, Centre for Northern Forest Ecosystem Research, Thunder Bay, Ontario, Canada. URL http://flash.lakeheadu.ca/~arodgers/hre/ [accessed 4 March 2014]

Saïd, S. and Servanty, S., 2005. The influence of landscape structure on female roe deer home-range size. *Landscape Ecology*, **20**, 1003-1012.

Shoval, N. and Isaacson, M., 2007. Sequence alignment as a method for human activity analysis in space and time. *Annals of the Association of American Geographers,* **97**, 282-297.

Therneau, T., Atkinson, B. and Ripley, B., 2014. Recursive Partitioning and Regression Trees. URL http://cran.r-project.org/web/packages/rpart/rpart.pdf [accessed 10 October 2014]

Tufto, J., Andersen, R. and Linnell, J.D.C., 1996. Habitat use and ecological correlates of home range size in a small cervid: the roe deer. *Journal of Animal Ecology*, **65**, 715-724.

Tukey, J.W., 1977. Exploratory data analysis, Reading: Addison-Wesley.

Urbano, F., Cagnacci, F., Calenge, C., Dettki, H., Cameron, A. and Neteler, M., 2010. Wildlife tracking data management: a new vision. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, **365**, 2177-2185.

Wilson, C., 2008. Activity patterns in space and time: calculation representative Hägerstrand ClustalG software. *Transportation,* **35**, 485-499.

Wittemyer, G., Polansky, L., Douglas-Hamilton, I. and Getz, W.M., 2008. Disentangling the effects of forage, social rank, and risk on movement autocorrelation of elephants using Fourier and wavelet analyses. *Proceedings of the National Academy of Sciences of the United States of America,* **105**, 19108-19113.

Worton, B.J., 1995. Using Monte-Carlo Simulation to Evaluate Kernel-Based Home-Range Estimators. *Journal of Wildlife Management*, **59**, 794-800.

Wu, L.L., 2000. Some Comments on Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect. *Sociological Methods and Research*, **29**, 41-64.

APPENDIX 3A – SIMULATED ARENAS

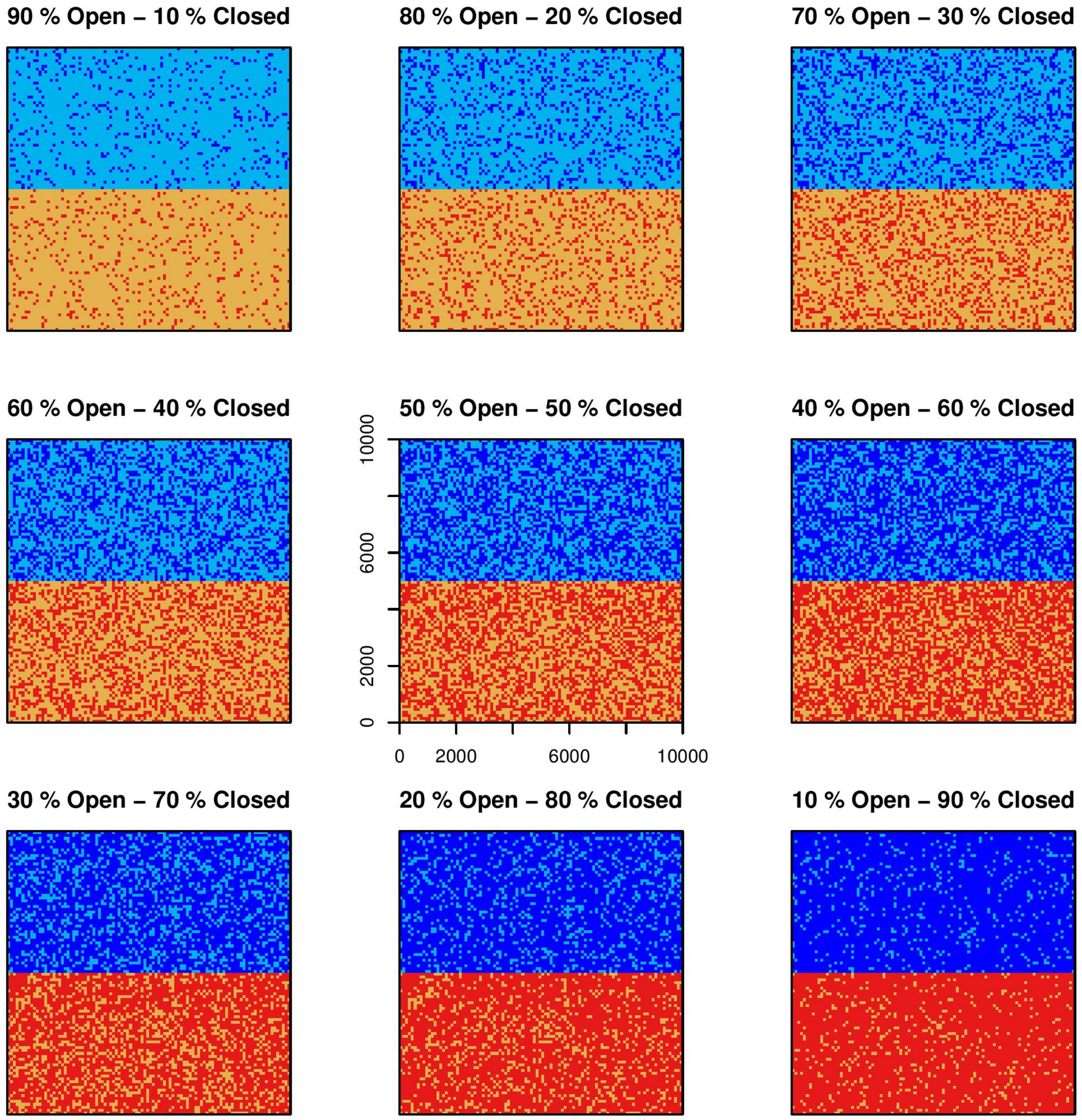


Figure 3A.1. Nine square gridded arenas, each composed of 10,000 squared cells of 100 x 100m, with varying proportions of open (light colours) and closed (dark colours) habitats (10 to 90%) at high (blue) and low elevation (red/orange). Simulated landscapes mirror the availability of environmental classes present in individual home ranges in the study area.

APPENDIX 3B – THE MOVEMENT MODEL

We simulated movement as a stochastic and spatially as well as temporally discrete process. The movement takes place on a grid consisting of N square cells (N=10,000). The probability of an animal to move from its position on cell i at time $t-1$ (i_{t-1} but for convenience abbreviated to $i \in \{1,2,..,N\}$) to any given cell $j \in \{1,2,..,N\}$ at time *t* is given by:

$$P_{i,j,t} = \frac{A_{i,j,t}}{\sum_{j=1}^{N} A_{i,j,t}}$$

with $A_{i,j,t}$ the attractiveness of cell *j* for an animal which is currently at cell *i*. $A_{i,j,t}$ is given by

$$A_{i,j,t} = \begin{cases} \dfrac{\beta_{S,h_j,t} * m_{j,t}}{d_{i,j}}, & d_{i,j} \leq dmax \\ 0, & otherwise; \end{cases}$$

where $\beta_{S,h_j,t}$ is the selection coefficient for habitat type *h* on cell *j*, h_j , at time *t* based on selection behaviour *S*, $m_{j,t}$ is the number of steps since the last visit from cell *i* to *j* at time *t*, and $d_{i,j,t}$ the squared Euclidean distance between cells *i* and *j*. We imposed a maximum distance, *dmax,* that an animal could travel in a single time step.

The selection coefficient $\beta_{S,h_j,t}$ was calculated for two habitat types *h* (open, *o*; or closed, *c*) and four selection behaviours *S* (random, *R*; homogeneous open, *O*; homogeneous closed, *C;* or alternating, *A*). We generated a random selection behaviour through a spatio-temporally constant selection coefficient (i.e. $\beta_{R,c,t} = \beta_{R,o,t} = 1.0$). Homogeneous closed and homogeneous open selection behaviours were defined as temporally constant but conditional on habitat types. Specifically, homogeneous closed selection was characterised by a high selection for closed habitat and a low selection for open habitat (i.e. $\beta_{C,c,t} = 1.0$ and $\beta_{C,o,t} = 0.1$), and *vice versa* for homogeneous open selection (i.e. $\beta_{O,c,t} = 0.1$ and $\beta_{O,o,t} = 1.0$). Unlike the other three selection behaviours, the alternating behaviour was modelled by a selection coefficient dependent on both habitat and time. Specifically, the coefficient values for closed and open habitat types were permuted every three simulation steps to capture circadian habitat selection patterns (6 iterations = one day):

$$\beta_{A,c,t} = \begin{cases} 0.1, & 6l+1 \leq t \leq 6l+3 \\ 1.0, & 6l+4 \leq t \leq 6(l+1) \end{cases}$$

$$\beta_{A,o,t} = \begin{cases} 1.0, & 6l+1 \leq t \leq 6l+3 \\ 0.1, & 6l+4 \leq t \leq 6(l+1) \end{cases}$$

with l the daily cycle iterations.

The coefficient $m_{j,t}$ to *j* at time *t* represents a memory component to the cell attraction value and captures the spatially-localised nature of movement behaviour that is common in roe deer (Hewison *et al.* 1998). Specifically, following the approach of van Moorter *et al.* (2009), we assumed the the attraction to a given cell increases with the time since last visit $m_{j,t}$. The memory value on any cell that has not yet been visited by the simulated animal was initialised to 1. When the animal visits a cell, the memory coefficient drops to 0 and then increments with every time step:

$$m_{j,t} = \begin{cases} 0, & j = i_{t-1} \\ m_{j,t-1} + 1, & j \neq i_{t-1} \; and \; j \in \{i_1, i_2, i_3, .., i_{t-2}\} \end{cases}$$

Furthermore, we assumed that a cell's attraction declined inversely proportional to the squared distance between cells *i* and *j*, $d_{i,j}$. Beyond this threshold, the cell attraction value was set to 0. The squared distance is calculated by

$$d_{i,j} = (x_i - x_j)^2 + (y_i - y_j)^2$$

With (x_i, y_i) and (x_j, y_j) the Cartesian coordinates of the centers for cells *i* and *j* respectively.

The simulations ran independently for 732 steps, and chose *N* = 10,000 and dmax = 500 meters. The first half was treated as a burning period, necessary for the simulated animals to build a memory of the landscape and establish a home range (van Moorter *et al.* 2009).

REFERENCES

Hewison, A.J.M., Vincent, J.P. and Reby, D., 1998. Social organisation of European roe deer. In: Andersen R., Duncan P., and Linnell, J.D.C., eds. *The European roe deer: the biology of success.* Oslo: Scandinavian University Press, 189-219.

van Moorter, B., Visscher, D., Benhamou, S., Börger, L., Boyce, M. and Gaillard, J.-M., 2009. Memory keeps you at home: a mechanistic model for home range emergence. *Oikos*, **118**, 641–652.

APPENDIX 3C – SUBSTITUTION MATRIX TO ACCOUNT FOR SPATIAL CORRELATION

The essence of Sequence Analysis Methods is to compare sequences and cluster them according to similarity. This is done according to a dissimilarity matrix, where substitutions between characters are weighted (see Fig.3.3 for a generic example of dissimilarity computation). So, a substitution matrix provides the weights for all possible substitutions. In our specific case, weights assigned to substitutions between habitat classes affect the dissimilarity measure between temporal sequences of habitat use. For example, two locations with different habitat openness and equal elevation, are 'less dissimilar' than two locations with different habitat openness and different elevation. However, there is a further issue to take into account when expressing the substitution matrix, i.e. spatial correlation of environmental variables. Here we explain how to measure spatial correlation and we assess the influence of spatial distribution on substitution weights. Then, we show how spatial correlation is accounted for in our study.

3C.1. From spatial correlation matrix to substitution matrix

Spatial correlation of environmental variables is defined as the spatial dependence between habitat values that can be more similar (positive correlation) or less similar (negative correlation) than would be expected by random (Dray *et al.* 2010). Accounting for spatial correlation is fundamental when examining temporal sequences of habitat use, since patterns of use are obviously ***constrained*** to the ***distribution and availability*** of the different habitat patches. Thus, when defining the dissimilarity weight of habitat classes, we took into account their actual spatial correlation. To measure spatial correlation, we used a spatial resolution of 100 m. The resolution of the spatial units corresponds to the median step length between two consecutive GPS data points (Δt=4 h) among the trajectories of all individuals.

Substitution weights were derived with a common method to compute spatial correlation from categorical data, namely Join-Count Statistics Rook Case (JCS), using spatial analysis software PaSSaGE 2 (Rosenberg & Anderson 2011). Join-Counts express how often polygons sharing an edge (i.e. neighbours) have the same or a different value. Usually JCS is based on dichotomic maps (i.e. presence/absence of one habitat type in spatial units), but Join-Counts can also be calculated for more than two classes (Rosenberg & Anderson 2011). In our study we computed the JCS for each of the four habitat classes (HC, HO, LC, LO, see legend in main text) and presented the Join-Counts as the proportion of the total number of compared pairs (i.e. neighbours) in a matrix of this type:

$$M_a = \begin{array}{ccccc} & HC & HO & LC & LO \\ HC & JC_{HC,HC} & JC_{HC,HO} & JC_{HC,LC} & JC_{HC,LO} \\ HO & & JC_{HO,HO} & JC_{HO,LC} & JC_{HO,LO} \\ LC & & & JC_{LC,LC} & JC_{LC,LO} \\ LO & & & & JC_{LO,LO} \end{array}$$

where

M_a: Similarity matrix based on Join-Counts (spatial correlation matrix)

$JC_{i,j}$: Proportion of Join-Counts of habitat classes (i,j) over all possible habitat pairs.

Whereas M_a gives similarities among habitat classes (partially showed in table 3C.1), a substitution matrix (M_s) relies on dissimilarities. Therefore we converted M_a as follows:

$M_s=1 – M_a/100$

Thus, Join-Counts, expressed as the proportion of the total number of compared pairs where a specific habitat neighbour relationship is *not* observed, are used to assess the dissimilarity values of the substitution matrix (Table 3C.2). Therefore, values of M_s range from 0 to 1, where 0 stands for similar and 1 for dissimilar. Whereas M_a will give the proportion of pairs with a specific habitat neighbour relationship (e.g. LC-LO), M_s will give the proportion of pairs where this relationship is not observed. For example for LC-LO the value in M_a will be 0.09, whereas in M_s the value will be 0.91 (Table 3C.2c).

3C.2. Effects of spatial correlation on substitution matrices

To assess how different levels of patchiness influence the substitution matrix, we created three artificial layers with different levels of stereotyped spatial autocorrelation of habitat classes (completely dispersed, completely random and completely clustered), while ***keeping the availability constant across classes***. We used a polygon having the shape of the study area, and transformed it into a grid with the specified resolution (100m). Each cell was then assigned one of the four habitat classes, according to the simulated distribution of habitat patches (Fig.3C.1). In all cases, levels of spatial correlation were measured by Join-Count Statistics Rook Case and quantified by Z-scores, expressed as:

$$Z=((O_{JC} - E_{JC})/stdev)$$

where:
O_{JC}: observed Join-Counts, i.e. number of adjacent polygons having the same value.
E_{JC}: expected Join-Counts under random distribution.
O_{JC} and E_{JC} were computed for each habitat class separately. The Z-score thus expressed the level of spatial correlation with respect to the variation in the area. Specifically, it indicates how many standard deviations the O_{JC} is above or below E_{JC}. Positive values correspond to a clustered pattern, values around zero to a random pattern and negative values to a dispersed pattern.

JCS and Z-scores of simulated scenarios (Table 3C.1) resulted in very different substitution matrices (Table 3C.2). In particular, the negative spatial correlation (Z_{HC}=-29.31; Z_{HO}=-29.35; Z_{LC}=-29.95; Z_{LO}=-29.21, Table 3C.1a) of the dispersed pattern translated in a substitution matrix without identities (dissimilarity of identities=1.00, Table 3C.2a), since there are no adjacent units with same values. For a random distribution (Z_{HC}=-0.02; Z_{HO} =0.55; Z_{LC}=-1.11; Z_{LO}=1.53, Table 3C.1b), dissimilarities between habitat use classes were more equally distributed and the resulting substitution matrix showed similar values across all combinations of habitat classes, both for identical and non-identical combinations (Table 3C.2b). Finally, the positive spatial autocorrelation (Z_{HC}=86.78; Z_{HO}=87.87; Z_{LC}=84.83; Z_{LO}=85.92, Table 3C.1c) of the clustered pattern corresponded to lower dissimilarities for identical than for non-identical combinations of habitat classes (Table 3C.2c).

Our simulation exercise therefore showed the need to measure spatial correlation and to account for that in a substitution matrix. Next, we show how we performed this in our study.

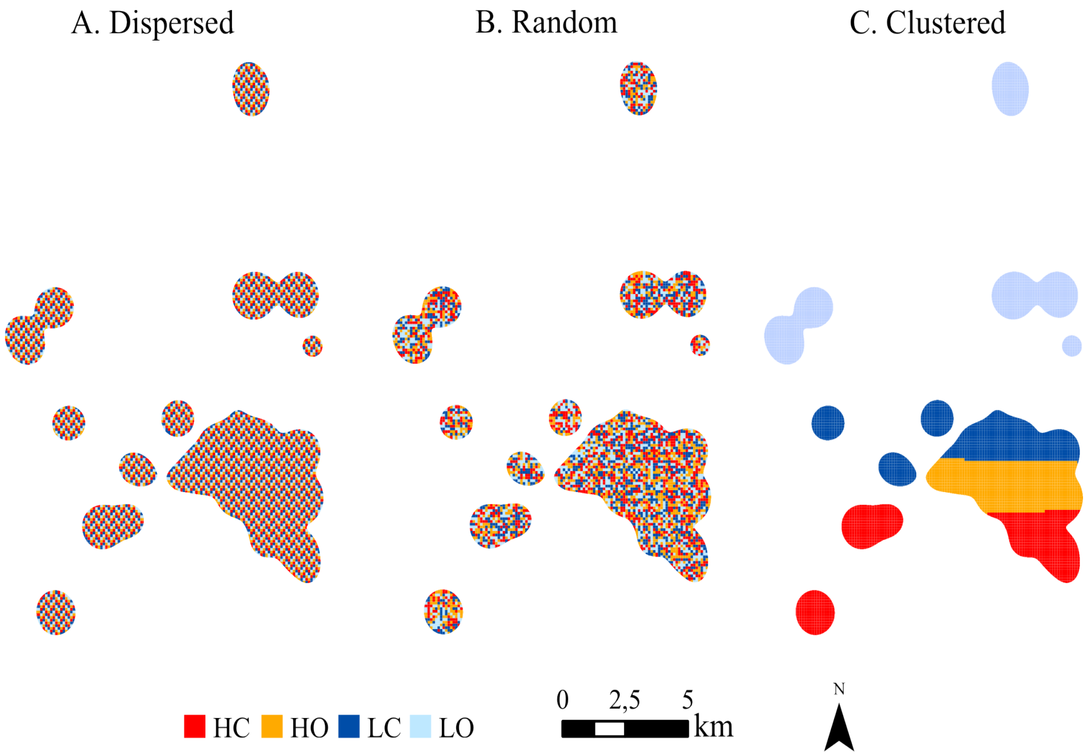


Figure 3C.1. Simulated dispersed (a), random (b) and clustered (c) spatial distribution of habitat patches based on the shape of the study area.

Table 3C.1. Join-Count Statistics for simulated scenarios (a, b, c). EJC=the expected number of adjacent spatial units with the same habitat, considering the null hypothesis of random distribution; OJC=the observed number of adjacent spatial units with the same habitat; SdE=the Standard deviation, Z=the Z-score (how many standard deviations the OJC is above or below EJC) and P=P-value of Z scores.

Simulation	**Spatial distribution**	**T-T**	**# Pairs**	E_{JC}	O_{JC}	Sd_E	Z	P
A	**Dispersed**	**HC-HC**	8220	509.61	0.00	17.39	-29.31	< 0.001
		HO-HO	8220	510.56	0.00	17.40	-29.35	< 0.001
		LC-LC	8220	526.75	0.00	17.59	-29.95	< 0.001
		LO-LO	8220	506.78	0.00	17.35	-29.21	< 0.001
B	**Random**	**HC-HC**	8220	513.39	513.00	17.43	-0.02	0.98
		HO-HO	8220	513.39	523.00	17.43	0.55	0.58
		LC-LC	8220	513.39	494.00	17.43	-1.11	0.27
		LO-LO	8220	513.39	540.00	17.43	1.53	0.13
C	**Clustered**	**HC-HC**	8220	513.39	2026.00	17.43	86.78	< 0.001
		HO-HO	8220	513.39	2045.00	17.43	87.87	< 0.001
		LC-LC	8220	513.39	1992.00	17.43	84.83	< 0.001
		LO-LO	8220	513.39	2011.00	17.43	85.92	< 0.001

Table 3C.2. Substitution matrices for simulated scenarios (a, b, c), expressed as dissimilarity measured using Join-Counts, i.e. the proportion of the total number of adjacent polygons where the specified habitat neighbour relationship is not observed. Dissimilarity values for identities and non-identical combinations are summarised for each substitution matrix where 'Id' is the dissimilarity for identities and 'Non-id μ (σ)" is the mean and standard deviation of the dissimilarity for non-identical combinations of habitats. Low values along the diagonal give an indication that habitat classes are highly correlated whereas high values, close to 1, are obtained for habitat pairs which are rarely observed.

a. Dispersed

	HC	**HO**	**LC**	**LO**	**Id**	**Non-id. μ (σ)**
HC	**1.00**	0.91	0.91	0.94	1.00	0.92(0.01)
HO	0.91	**1.00**	0.94	0.91	1.00	0.92(0.01)
LC	0.91	0.94	**1.00**	0.91	1.00	0.92(0.01)
LO	0.94	0.91	0.91	**1.00**	1.00	0.92(0.01)

b. Random

	HC	**HO**	**LC**	**LO**	**Id**	**Non-id μ (σ)**
HC	**0.94**	0.94	0.94	0.94	0.94	0.94(0.00)
HO	0.94	**0.94**	0.94	0.94	0.94	0.94(0.00)
LC	0.94	0.94	**0.94**	0.94	0.94	0.94(0.00)
LO	0.94	0.94	0.94	**0.93**	0.94	0.94(0.00)

c. Clustered

	HC	**HO**	**LC**	**LO**	**Id**	**Non-id. μ (σ)**
HC	**0.75**	1.00	1.00	1.00	0.75	1.00(0.00)
HO	1.00	**0.75**	1.00	1.00	0.75	1.00(0.00)
LC	1.00	1.00	**0.76**	1.00	0.76	1.00(0.00)
LO	1.00	1.00	1.00	**0.76**	0.76	1.00(0.00)

3C.3. Substitution matrices for simulated arenas

In this study, spatial correlation was assessed for our 9 simulated arenas each corresponding to different availabilities of open/closed habitats, with a random distribution (see *Material and Methods, Simulation arenas and simulated trajectories* paragraph). As an effect of the varying proportions of habitats, the spatial correlation structure was different across arenas and could be matched with those from real bimonthly home ranges. Contrary to our simulation exercise described above, the spatial correlation structure was a consequence of different habitat availability in the arenas, as typical in the real case. Indeed, the arenas showed positive relation between spatial correlation (i.e. increasing Z-scores) and availability of open or closed habitats (Fig.3C.2). Our simulated landscapes in arenas therefore gave a good assessment of the real habitat prevalence and distribution (i.e. in home ranges).

Here we provide the JCS and Z-scores for the 9 simulated arenas (table 3C.3), together with the corresponding substitution matrices (table 3C.4).

Table 3C.3. Join-Count Statistics for 9 arenas with variable proportions of open and closed habitats (expressed as the percentage of open habitat). See table 3C.1 for explanation of codes.

% open	**T-T**	**# Pairs**	E_{JC}	O_{JC}	Sd_E	Z	P
10	**HC x HC**	19800	4009.01	7982	35.1	113.191	< 0.001
	HO x HO	19800	49.41	107	6.68	**8.62126**	< 0.001
	LC x LC	19800	4009.01	7985	35.1	113.276	< 0.001
	LO x LO	19800	49.41	101	6.68	**7.72305**	< 0.001
20	**HC x HC**	19800	3167.52	6312	33.99	92.5119	< 0.001
	HO x HO	19800	197.82	405	12.67	**16.352**	< 0.001
	LC x LC	19800	3167.52	6312	33.99	92.5119	< 0.001
	LO x LO	19800	197.82	402	12.67	**16.1152**	< 0.001
30	**HC x HC**	19800	2425.05	4805	32.18	73.9574	< 0.001
	HO x HO	19800	445.25	872	17.97	**23.7479**	< 0.001
	LC x LC	19800	2425.05	4858	32.18	75.6044	< 0.001
	LO x LO	19800	445.25	916	17.97	**26.1964**	< 0.001
40	**HC x HC**	19800	1781.58	3514	29.67	58.3896	< 0.001
	HO x HO	19800	791.68	1552	22.56	**33.7021**	< 0.001
	LC x LC	19800	1781.58	3565	29.67	60.1085	< 0.001
	LO x LO	19800	791.68	1588	22.56	**35.2979**	< 0.001
50	**HC x HC**	19800	1237.13	2438	26.47	45.3672	< 0.001
	HO x HO	19800	1237.13	2438	26.47	**45.3672**	< 0.001
	LC x LC	19800	1237.13	2473	26.47	46.6895	< 0.001
	LO x LO	19800	1237.13	2467	26.47	**46.4628**	< 0.001
60	**HC x HC**	19800	791.68	1554	22.56	33.7908	< 0.001
	HO x HO	19800	1781.58	3530	29.67	**58.9289**	< 0.001
	LC x LC	19800	791.68	1575	22.56	34.7216	< 0.001
	LO x LO	19800	1781.58	3539	29.67	**59.2322**	< 0.001
70	**HC x HC**	19800	445.25	862	17.97	23.19143	< 0.001
	HO x HO	19800	2425.05	4811	32.18	**74.1439**	< 0.001
	LC x LC	19800	445.25	886	17.97	24.52699	< 0.001
	LO x LO	19800	2425.05	4825	32.18	**74.5789**	< 0.001
80	**HC x HC**	19800	197.82	373	12.67	13.8264	< 0.001
	HO x HO	19800	3167.52	6285	33.99	**91.7176**	< 0.001
	LC x LC	19800	197.82	383	12.67	14.6156	< 0.001
	LO x LO	19800	3167.52	6281	33.99	**91.5999**	< 0.001
90	**HC x HC**	19800.00	49.41	95.00	6.68	6.82485	< 0.001
	HO x HO	19800.00	4009.01	7970.00	35.10	**112.849**	< 0.001
	LC x LC	19800.00	49.41	97.00	6.68	7.12425	< 0.001
	LO x LO	19800.00	4009.01	7971.00	35.10	**112.877**	< 0.001

Table 3C.4. Substitution matrices for the nine arenas. Dissimilarities of substitutions vary according to availability, thus presenting different correlation structures. For example, substitutions between open/closed habitats for same elevation classes (e.g. HCxHO, LCxLO) are less dissimilar when open/closed habitats are more proportional (40% to 60%) (see values in bolt for an example of HcxHO).

10%	HC	HO	LC	LO
HC	0.597	**0.911**	0.996	0.999
HO	0.911	0.995	0.999	1
LC	0.996	0.999	0.597	0.911
LO	0.999	1	0.911	0.995

20%	HC	HO	LC	LO
HC	0.681	**0.842**	0.997	0.999
HO	0.842	0.98	0.999	1
LC	0.997	0.999	0.681	0.842
LO	0.999	1	0.842	0.98

30%	HC	HO	LC	LO
HC	0.757	**0.789**	0.998	0.999
HO	0.789	0.956	0.999	0.999
LC	0.998	0.999	0.755	0.794
LO	0.999	0.999	0.794	0.954

40%	HC	HO	LC	LO
HC	0.823	**0.758**	0.998	0.999
HO	0.758	0.922	0.999	0.999
LC	0.998	0.999	0.82	0.763
LO	0.999	0.999	0.763	0.92

50%	HC	HO	LC	LO
HC	0.877	**0.749**	0.999	0.999
HO	0.749	0.877	0.999	0.998
LC	0.999	0.999	0.875	0.752
LO	0.999	0.998	0.752	0.875

60%	HC	HO	LC	LO
HC	0.922	**0.759**	0.999	0.999
HO	0.759	0.822	0.999	0.998
LC	0.999	0.999	0.92	0.761
LO	0.999	0.998	0.761	0.821

70%	HC	HO	LC	LO
HC	0.956	**0.789**	1	0.999
HO	0.789	0.757	0.999	0.997
LC	1	0.999	0.955	0.791
LO	0.999	0.997	0.791	0.756

80%	HC	HO	LC	LO
HC	0.981	**0.839**	1	0.999
HO	0.839	0.683	0.999	0.997
LC	1	0.999	0.981	0.839
LO	0.999	0.997	0.839	0.683

90%	HC	HO	LC	LO
HC	0.995	**0.910**	1.000	1.000
HO	0.910	0.597	0.999	0.996
LC	1.000	0.999	0.995	0.910
LO	1.000	0.996	0.910	0.597

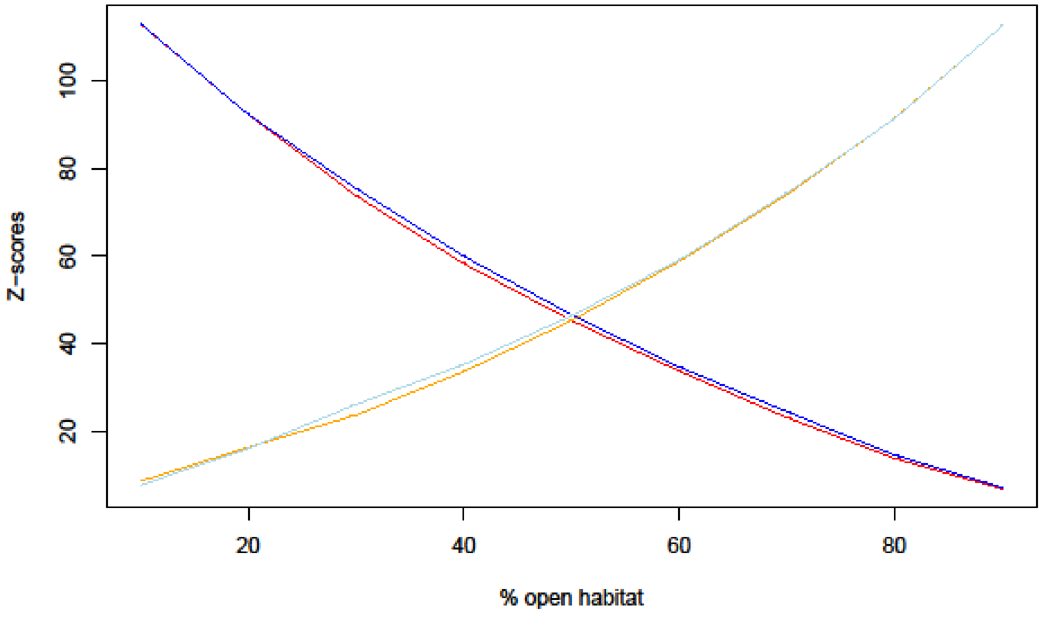


Figure 3C.2. Spatial correlation (Z-scores) in relation to habitat availability (x-axis, % open habitat) for HCx-HC (red), HOxHO (yellow), LCxLC (blue), LOxLO (light blue).

REFERENCES

Dray, S., Royer-Carenzi, M. and Calenge, C., 2010. The exploratory analysis of autocorrelation in animal-movement studies. *Ecological Research*, **4**, 34-41.

Rosenberg, M.S., and Anderson, C.D., 2011. PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2. *Methods in Ecology and Evolution*, **2**, 229-232.

APPENDIX 3D – TREES

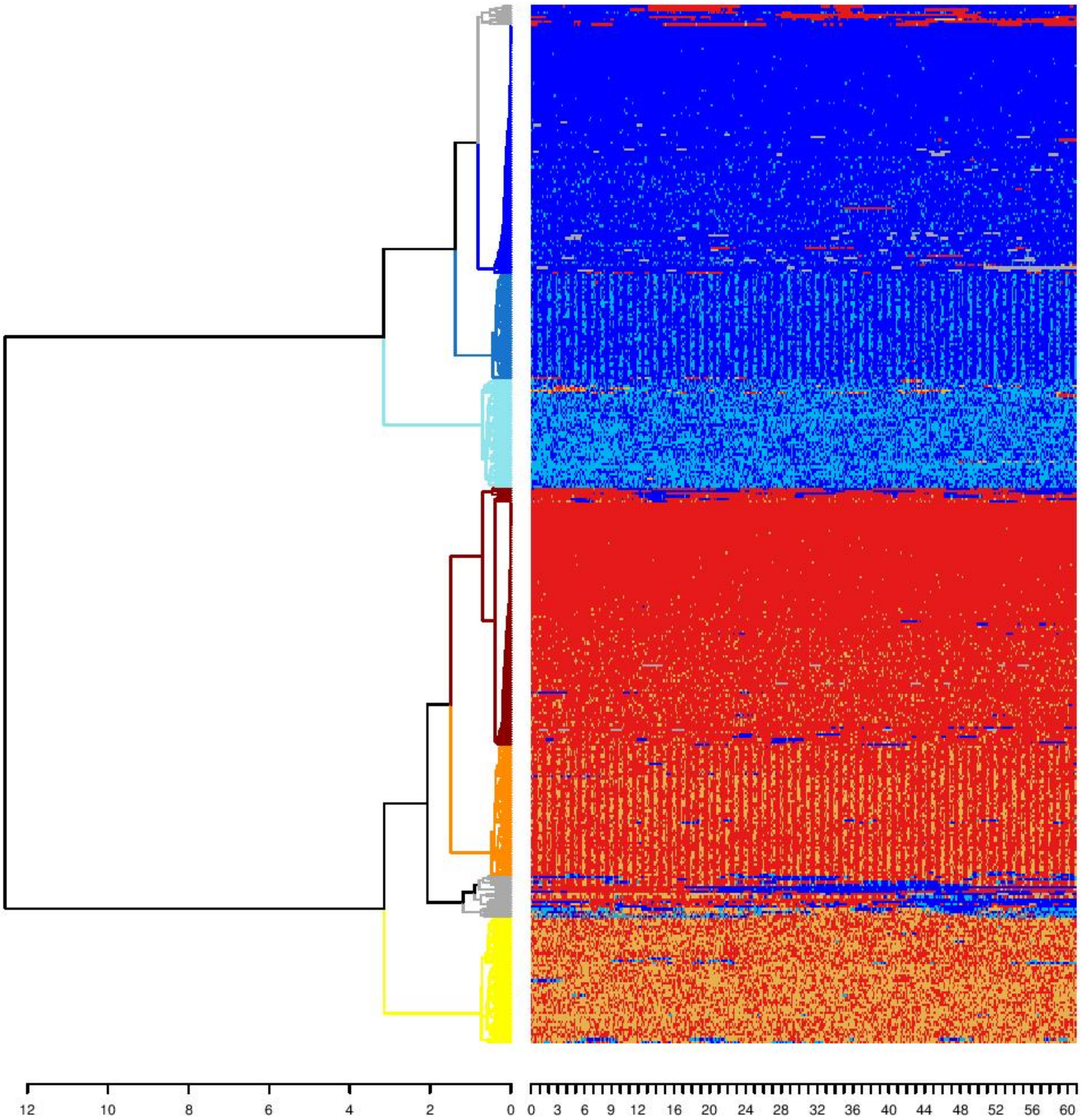


Figure 3D.1. Simulated and roe deer bimonthly sequences of habitat use classes (right portion of the panel; daily scale reported bottom-right), and resulting dissimilarity tree, based on the Hamming distance algorithm (left portion of the panel). The simulated sequences are derived from a trajectory running over an arena with 10% open habitat (T_{10}). Real sequences were included with simulated ones if referred to home ranges with the same open habitat proportion. Tree 'leaves' represent a real or simulated sequence, while 'nodes' indicate their clustering. Branch lengths (distance between leaves, and first common node: bottom-left for the scale) indicate the dissimilarity between individual sequences. The split into two main clusters define first-order clusters separation, based on use of elevation classes (high, low). Second-order clusters are based on use of open/closed habitats and distinguish homogeneous closed and random (brown and dark blue branches), alternating (orange and blue branches) and homogeneous open (yellow and light blue branches). Grey branches are sequences with undefined classification or clusters with mixed sequences of high and low.

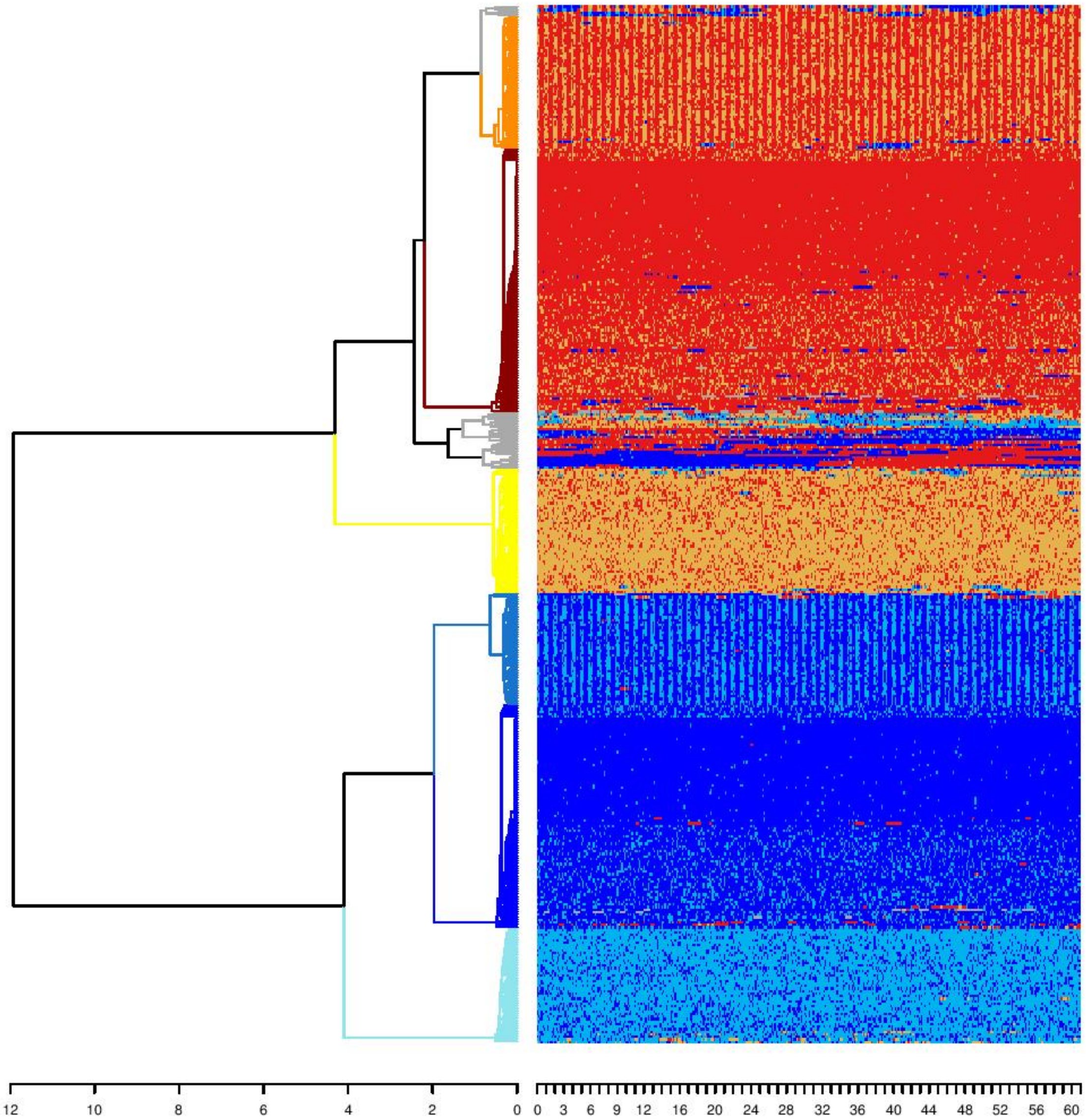


Figure 3D.2. Simulated and roe deer bimonthly sequences of habitat use classes (right portion of the panel; daily scale reported bottom-right), and resulting dissimilarity tree, based on the Hamming distance algorithm (left portion of the panel). The simulated sequences are derived from a trajectory running over an arena with 20% open habitat (T_{20}). Real sequences were included with simulated ones if referred to home ranges with the same open habitat proportion. Tree 'leaves' represent a real or simulated sequence, while 'nodes' indicate their clustering. Branch lengths (distance between leaves, and first common node: bottom-left for the scale) indicate the dissimilarity between individual sequences. The split into two main clusters define first-order clusters separation, based on use of elevation classes (high, low). Second-order clusters are based on use of open/closed habitats and distinguish homogeneous closed and random (brown and dark blue branches), alternating (orange and blue branches) and homogeneous open (yellow and light blue branches). Grey branches are sequences with undefined classification or clusters with mixed sequences of high and low.

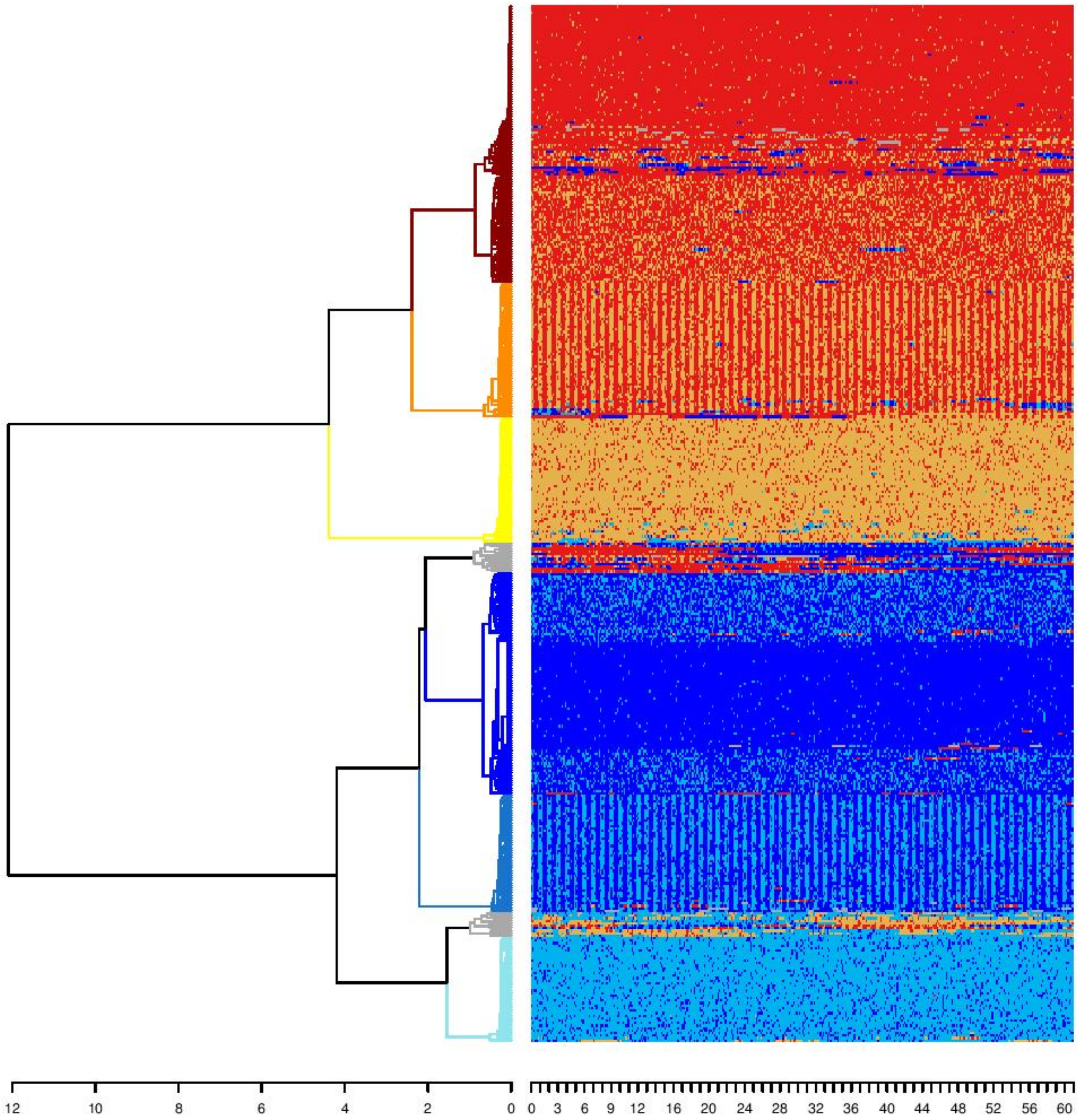


Figure 3D.3. Simulated and roe deer bimonthly sequences of habitat use classes (right portion of the panel; daily scale reported bottom-right), and resulting dissimilarity tree, based on the Hamming distance algorithm (left portion of the panel). The simulated sequences are derived from a trajectory running over an arena with 30% open habitat (T_{30}). Real sequences were included with simulated ones if referred to home ranges with the same open habitat proportion. Tree 'leaves' represent a real or simulated sequence, while 'nodes' indicate their clustering. Branch lengths (distance between leaves, and first common node: bottom-left for the scale) indicate the dissimilarity between individual sequences. The split into two main clusters define first-order clusters separation, based on use of elevation classes (high, low). Second-order clusters are based on use of open/closed habitats and distinguish homogeneous closed and random (brown and dark blue branches), alternating (orange and blue branches) and homogeneous open (yellow and light blue branches). Grey branches are sequences with undefined classification or clusters with mixed sequences of high and low.

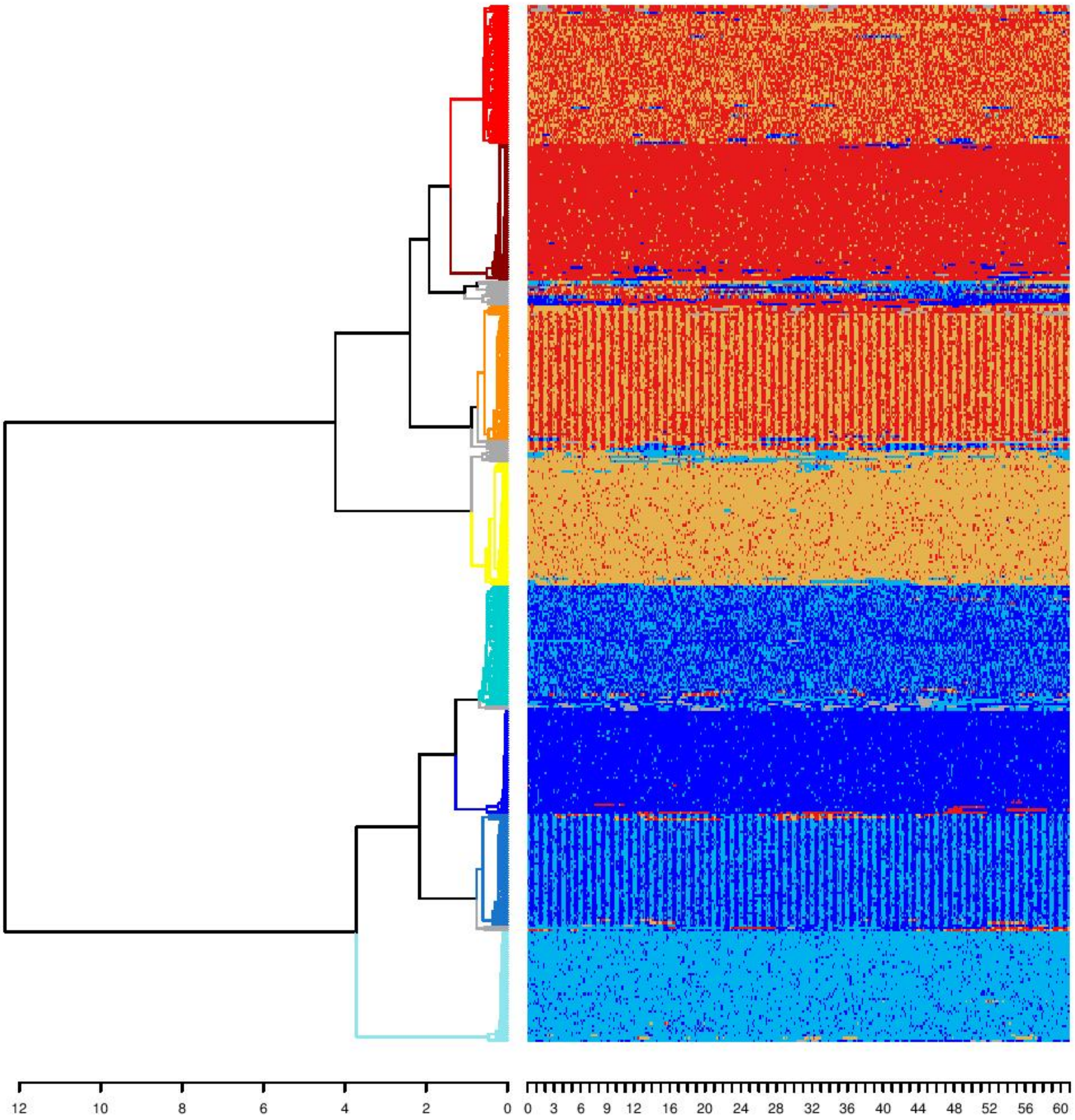


Figure 3D.4. Simulated and roe deer bimonthly sequences of habitat use classes (right portion of the panel; daily scale reported bottom-right), and resulting dissimilarity tree, based on the Hamming distance algorithm (left portion of the panel). The simulated sequences are derived from a trajectory running over an arena with 40% open habitat (T_{40}). Real sequences were included with simulated ones if referred to home ranges with the same open habitat proportion. Tree ‘leaves’ represent a real or simulated sequence, while ‘nodes’ indicate their clustering. Branch lengths (distance between leaves, and first common node: bottom-left for the scale) indicate the dissimilarity between individual sequences. The split into two main clusters define first-order clusters separation, based on use of elevation classes (high, low). Second-order clusters are based on use of open/closed habitats and distinguish homogeneous closed (brown and dark blue branches), random (red and cyan branches), alternating (orange and blue branches) and homogeneous open (yellow and light blue branches). Grey branches are sequences with undefined classification or clusters with mixed sequences of high and low.

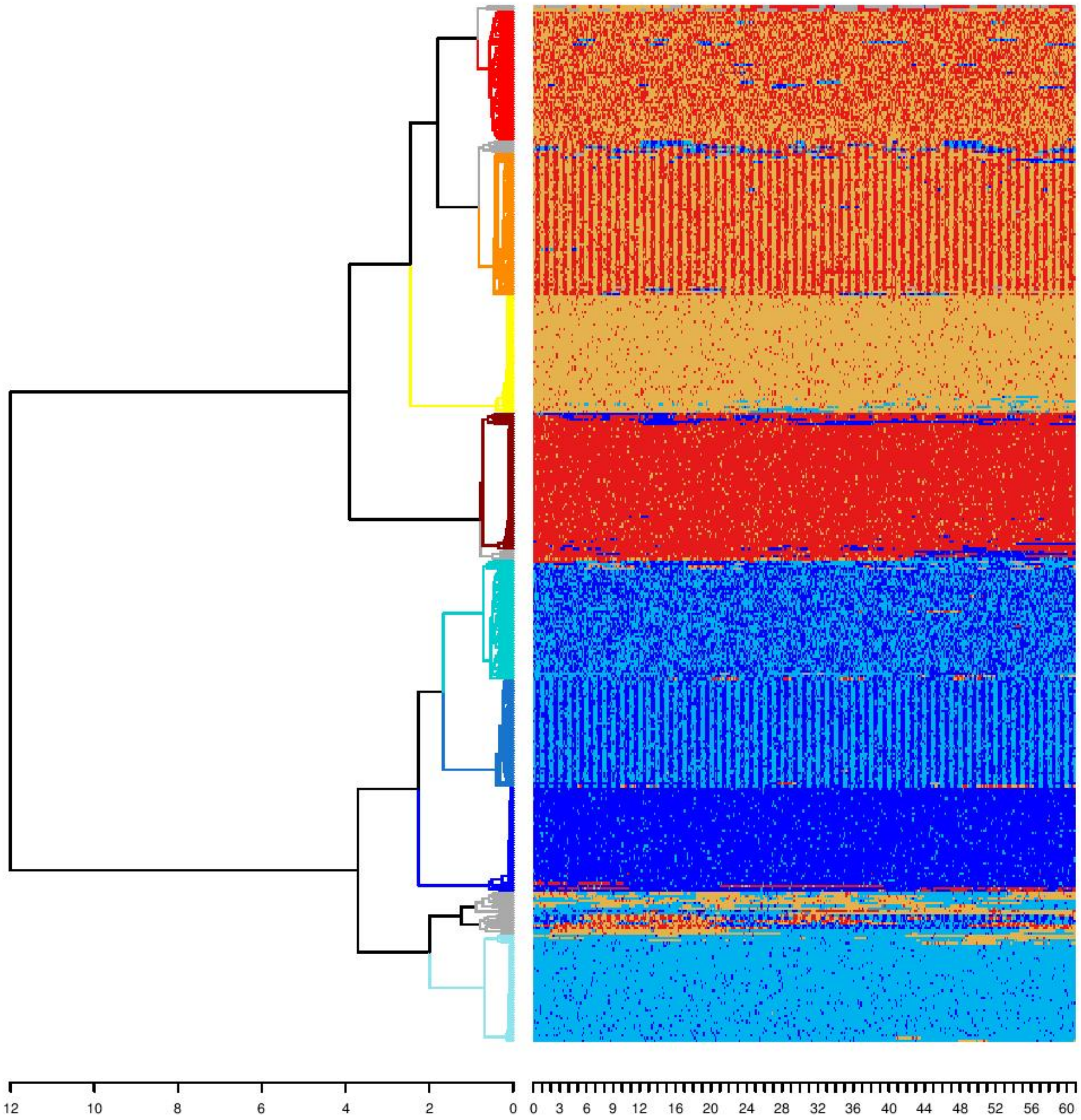


Figure 3D.5. Simulated and roe deer bimonthly sequences of habitat use classes (right portion of the panel; daily scale reported bottom-right), and resulting dissimilarity tree, based on the Hamming distance algorithm (left portion of the panel). The simulated sequences are derived from a trajectory running over an arena with 50% open habitat (T_{50}). Real sequences were included with simulated ones if referred to home ranges with the same open habitat proportion. Tree 'leaves' represent a real or simulated sequence, while 'nodes' indicate their clustering. Branch lengths (distance between leaves, and first common node: bottom-left for the scale) indicate the dissimilarity between individual sequences. The split into two main clusters define first-order clusters separation, based on use of elevation classes (high, low). Second-order clusters are based on use of open/closed habitats and distinguish homogeneous closed (brown and dark blue branches), random (red and cyan branches), alternating (orange and blue branches) and homogeneous open (yellow and light blue branches). Grey branches are sequences with undefined classification or clusters with mixed sequences of high and low.

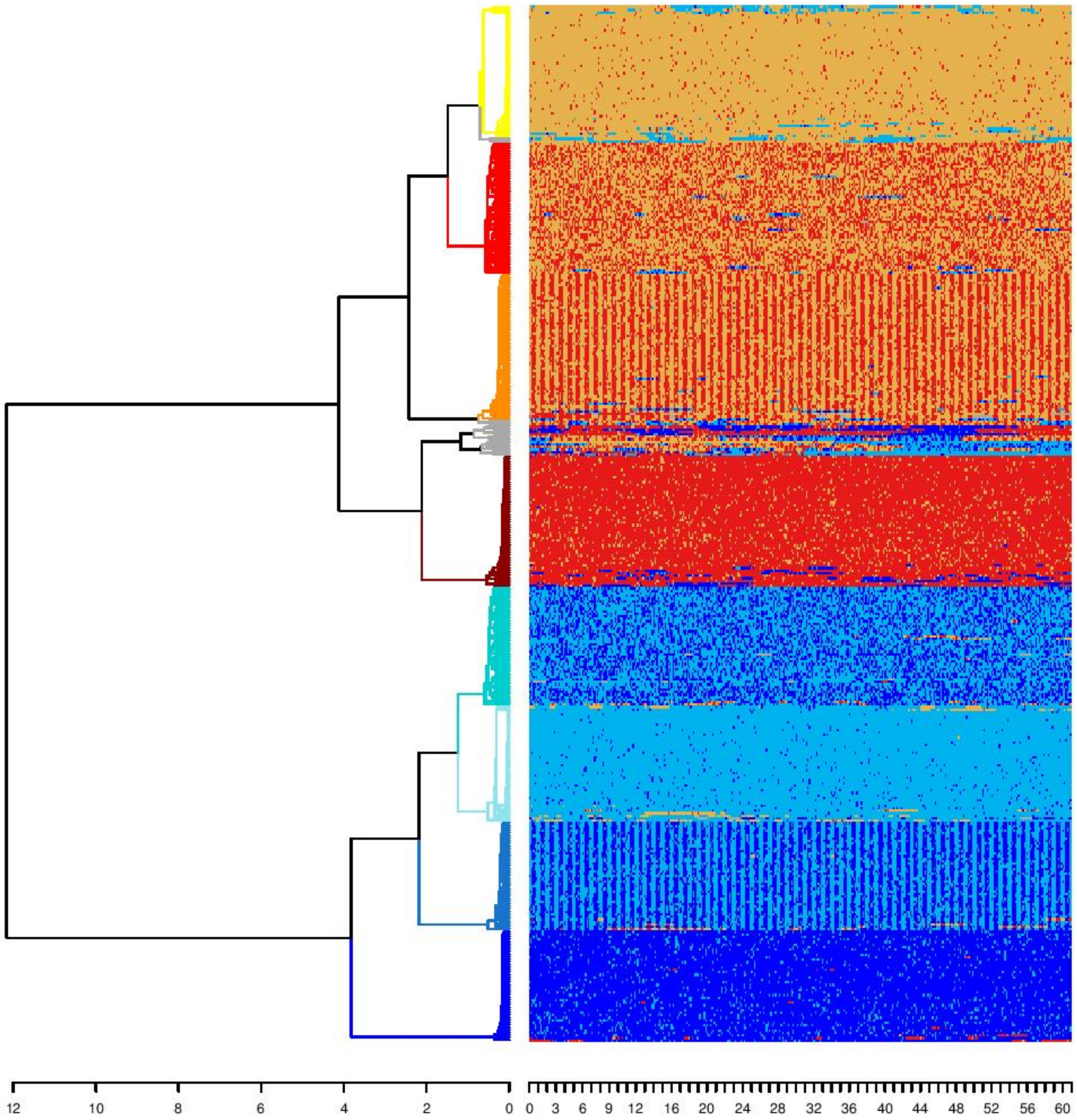


Figure 3D.6. Simulated and roe deer bimonthly sequences of habitat use classes (right portion of the panel; daily scale reported bottom-right), and resulting dissimilarity tree, based on the Hamming distance algorithm (left portion of the panel). The simulated sequences are derived from a trajectory running over an arena with 60% open habitat (T_{60}). Real sequences were included with simulated ones if referred to home ranges with the same open habitat proportion. Tree 'leaves' represent a real or simulated sequence, while 'nodes' indicate their clustering. Branch lengths (distance between leaves, and first common node: bottom-left for the scale) indicate the dissimilarity between individual sequences. The split into two main clusters define first-order clusters separation, based on use of elevation classes (high, low). Second-order clusters are based on use of open/closed habitats and distinguish homogeneous closed (brown and dark blue branches), random (red and cyan branches), alternating (orange and blue branches) and homogeneous open (yellow and light blue branches). Grey branches are sequences with undefined classification or clusters with mixed sequences of high and low.

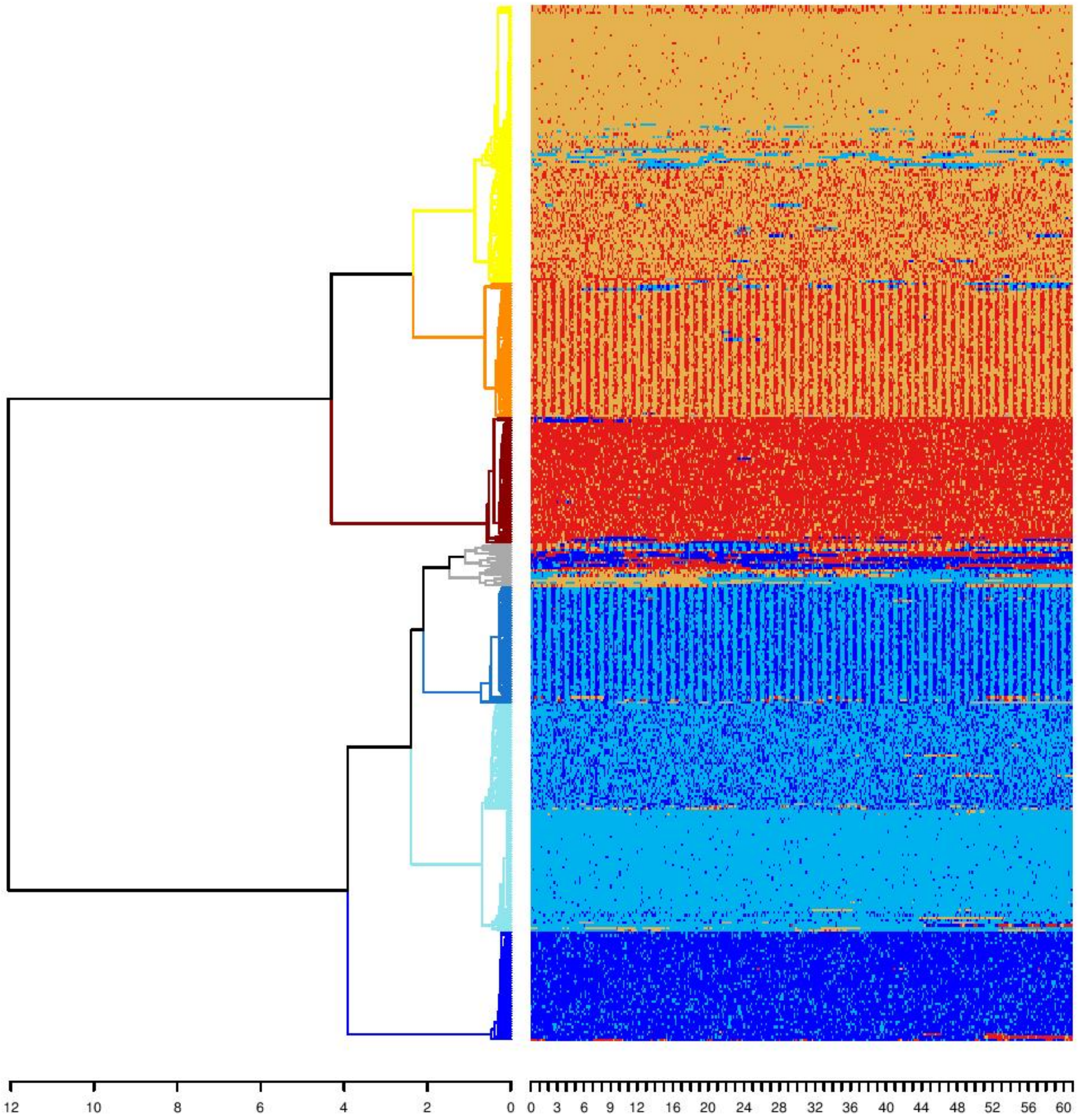


Figure 3D.7. Simulated and roe deer bimonthly sequences of habitat use classes (right portion of the panel; daily scale reported bottom-right), and resulting dissimilarity tree, based on the Hamming distance algorithm (left portion of the panel). The simulated sequences are derived from a trajectory running over an arena with 70% open habitat (T_{70}). Real sequences were included with simulated ones if referred to home ranges with the same open habitat proportion. Tree 'leaves' represent a real or simulated sequence, while 'nodes' indicate their clustering. Branch lengths (distance between leaves, and first common node: bottom-left for the scale) indicate the dissimilarity between individual sequences. The split into two main clusters define first-order clusters separation, based on use of elevation classes (high, low). Second-order clusters are based on use of open/closed habitats and distinguish homogeneous closed (brown and dark blue branches), alternating (orange and blue branches) and homogeneous open and random (yellow and light blue branches). Grey branches are sequences with undefined classification or clusters with mixed sequences of high and low.

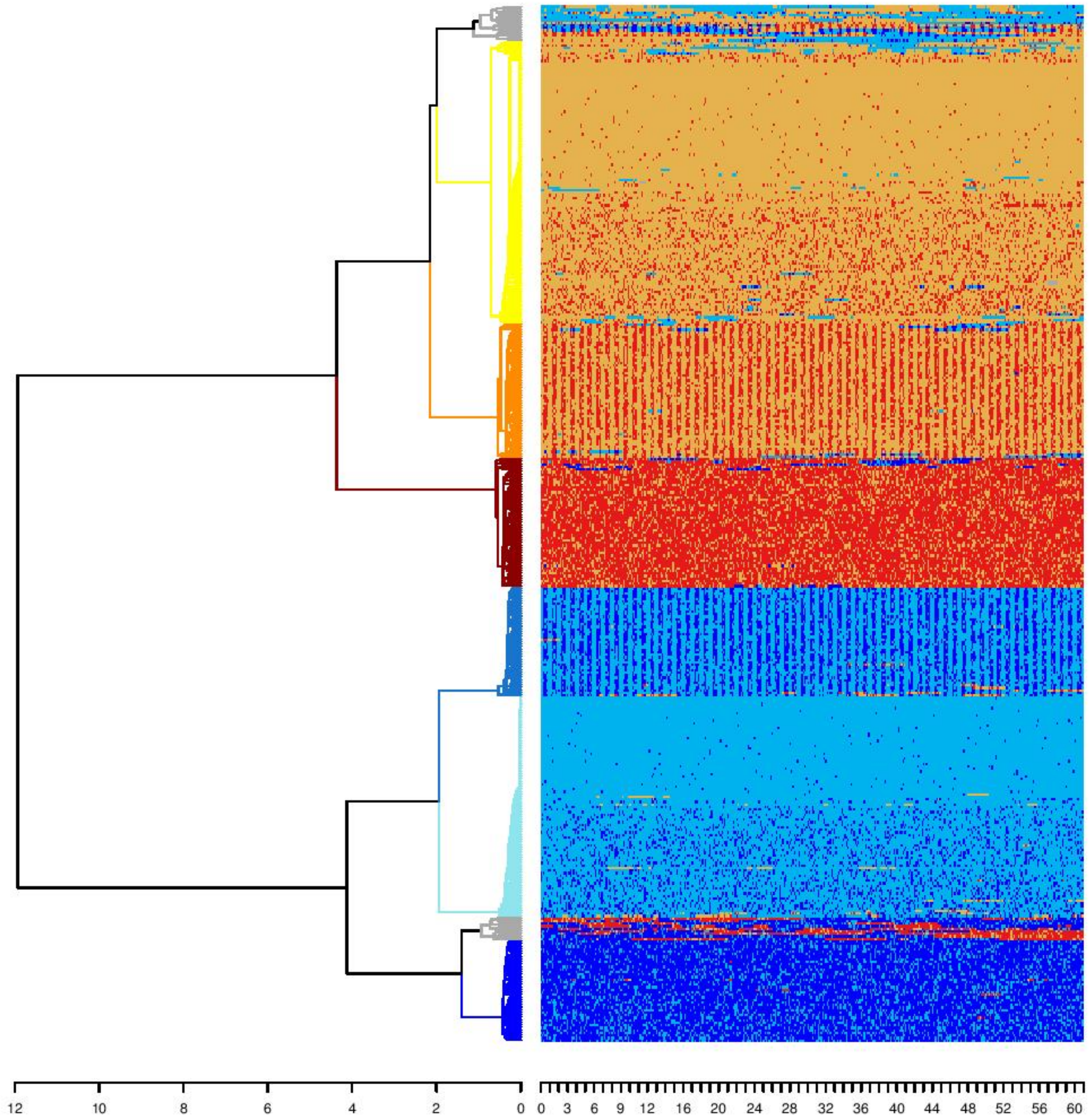


Figure 3D.8. Simulated and roe deer bimonthly sequences of habitat use classes (right portion of the panel; daily scale reported bottom-right), and resulting dissimilarity tree, based on the Hamming distance algorithm (left portion of the panel). The simulated sequences are derived from a trajectory running over an arena with 80% open habitat (T_{80}). Real sequences were included with simulated ones if referred to home ranges with the same open habitat proportion. Tree 'leaves' represent a real or simulated sequence, while 'nodes' indicate their clustering. Branch lengths (distance between leaves, and first common node: bottom-left for the scale) indicate the dissimilarity between individual sequences. The split into two main clusters define first-order clusters separation, based on use of elevation classes (high, low). Second-order clusters are based on use of open/closed habitats and distinguish homogeneous closed (brown and dark blue branches), alternating (orange and blue branches) and homogeneous open and random (yellow and light blue branches). Grey branches are sequences with undefined classification or clusters with mixed sequences of high and low.

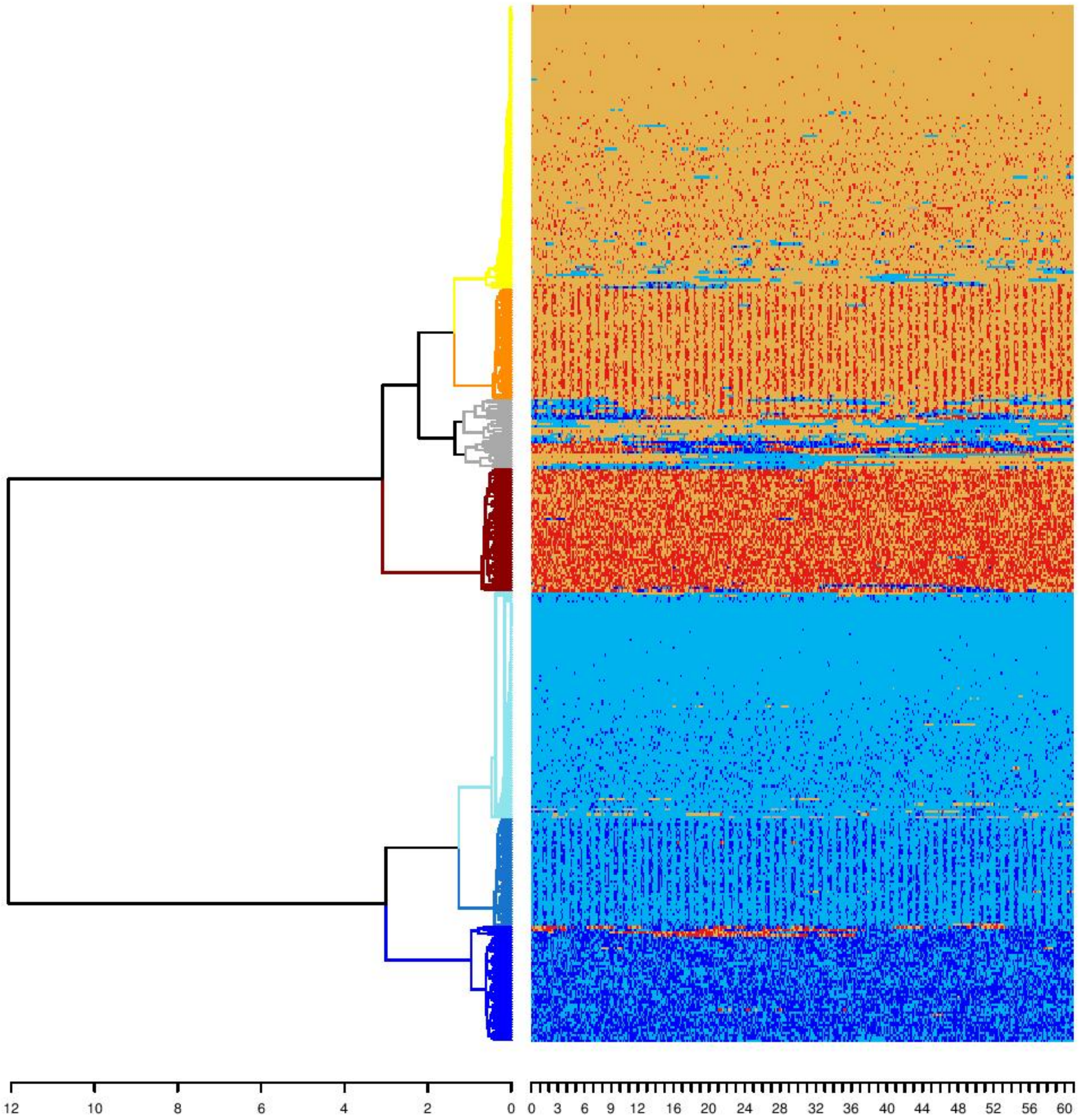


Figure 3D.9. Simulated and roe deer bimonthly sequences of habitat use classes (right portion of the panel; daily scale reported bottom-right), and resulting dissimilarity tree, based on the Hamming distance algorithm (left portion of the panel). The simulated sequences are derived from a trajectory running over an arena with 90% open habitat (T_{90}). Real sequences were included with simulated ones if referred to home ranges with the same open habitat proportion. Tree 'leaves' represent a real or simulated sequence, while 'nodes' indicate their clustering. Branch lengths (distance between leaves, and first common node: bottom-left for the scale) indicate the dissimilarity between individual sequences. The split into two main clusters define first-order clusters separation, based on use of elevation classes (high, low). Second-order clusters are based on use of open/closed habitats and distinguish homogeneous closed (brown and dark blue branches), alternating (orange and blue branches) and homogeneous open and random (yellow and light blue branches). Grey branches are sequences with undefined classification or clusters with mixed sequences of high and low.

APPENDIX 3E – CLUSTERS OF REAL BIMONTHLY SEQUENCES

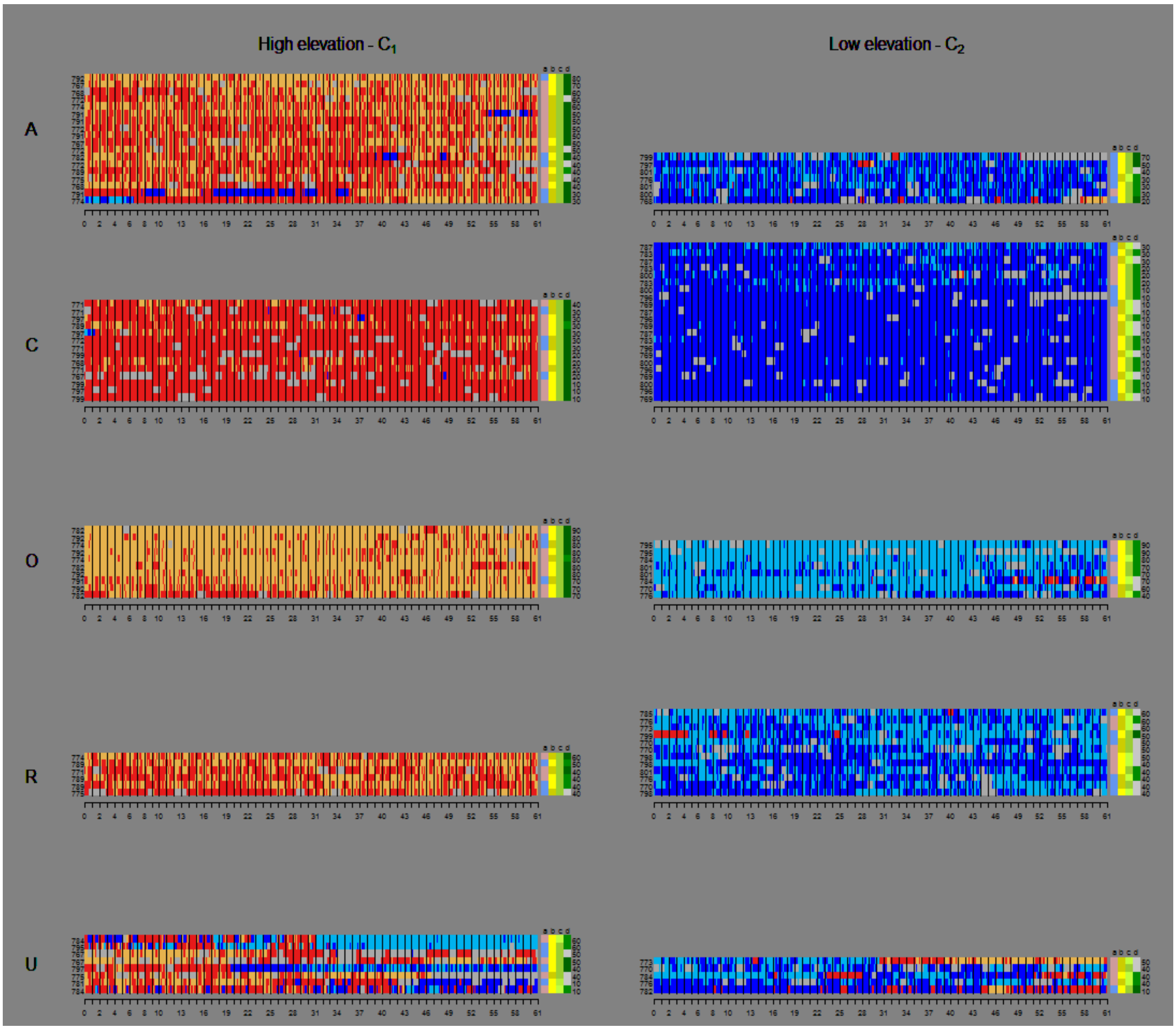


Figure 3E.1. Clusters of real bimonthly sequences with similar patterns of sequential habitat use, extracted from the corresponding dissimilarity trees. Left and right ‘columns’ refer to high and low elevation clusters. Alternating (A), homogeneous closed (C), homogeneous open (O), random when available (R) and undefined (U) clusters at both elevations are represented by row. Midnight of consecutive days is indicated by vertical black lines. The id (left y-axis, e.g. W_sF_aA_mM) gives the individual code. Variables season, age, sex and migration are respectively represented as colour-coded bars at the right side of the sequences (see Fig.3.4 for the legend). The y-axis on the right side gives the proportion (%) of open habitats within the roe deer's bimonthly home range.

CHAPTER 4

Individual Movement - Sequence Analysis Methods (IM-SAM): characterising spatio-temporal patterns of animal trajectories across scales and landscapes

De Groeve, J., Cagnacci, F., Ranc., N., Bonnot., N., Benedikt, G., Heurich, M., Hewison, A.J.M., Kroeschel, M., Linnell, J.D.C., Morellet, N., Mysterud, A., Sandfort, R., Van de Weghe, N.

Submitted in: International Journal of Geographical Information Science

ABSTRACT - We present methodological advances to a recently developed framework to study sequential habitat use by animals using a visually-explicit and tree-based Sequence Analysis Method (SAM), derived from molecular biology and more recently used in time geography. Habitat use sequences are expressed as annotations obtained by intersecting GPS movement trajectories with environmental layers. Here, we develop IM-SAM, where we use individual home ranges as the reference spatial context. To assess the applicability of our new framework, we investigated the sequential use of open and closed habitats across multiple European roe deer populations ranging in landscapes with markedly contrasting structure. Starting from simulated sequences based on a mechanistic movement model, we found that different sequential patterns of habitat use were better distinguished as separate, robust clusters, with less variable cluster size, when habitats were proportional within the home range. Application on real roe deer sequences showed that our approach effectively captured variation in spatio-temporal patterns of sequential habitat use, and provided evidence for several behavioural processes, such as day-night habitat alternation. By characterising sequential habitat use patterns of animals, we may better evaluate the temporal trade-offs in animal habitat use and how they are affected by changes in landscapes. This improved framework could also be used by geographers to study human movement patterns, or the relationships between animal and human spatio-temporal movements.

4.1. INTRODUCTION

Understanding which habitats are used by animals through space and time is important to establish cost-effective and time-varying policies essential for species conservation and wildlife management purposes. In Europe, for example, landscapes are strongly affected both by human intensification of land use (e.g. urbanisation, deforestation and road construction), leading to habitat loss and fragmentation on the one side, and by the abandonment of traditional agro-systems due to decreased economic viability on the other (Jaeger *et al.* 2011). The latter, in combination with conservation policies (e.g. Bern Convention (Council of Europe, 1979); Habitat and Bird directive (European Commission, 1992, 2009)), has favoured the re-colonisation of native wildlife species across a variety of landscapes. For example, Europe currently has twice as many wolves than the USA, with more than twice the human population density and half of the land surface area (Chapron *et al.* 2014). Although many species with specific ecological requirements are decreasing in distribution and abundance due to fragmentation and habitat impoverishment (e.g. Hanski *et al.* 2005, Krauss *et al.* 2010), several others are gradually adjusting to the European anthropogenic landscape by also using human-dominated land use types (i.e. agricultural land, pastures, urban areas). In this context, spatio-temporal dynamics of animal movement behaviour and habitat use may reveal patterns of important applied value. For example, several ungulate species show higher activity and intensified movement at dusk and dawn, resulting in more road-crossings (Kämmerle *et al.* 2017) and hence vehicle collisions during twilight. Similarly, the fragmented patchwork of forest habitat within a matrix of agricultural areas has favoured the use of crops or crop remains as a forage resource by ungulates. For example, roe deer (Aulak & Babinska-Werka 1990; Cibien *et al.* 1989; Hewison *et al.* 2001), and especially wild boar (Podgórski *et al.* 2013) have adapted very well to agro-ecosystems or even urbanised areas (i.e. phenomenon of ‘urban wildlife’; Magle *et al.* 2012), and can reach local high abundances that may result in conflicts. Consequently, we need robust methodological approaches to understand the temporal patterns in the use of complementary anthropogenic resources (such as food and cover) in order to take appropriate management actions.

Habitats provide the resources (e.g. food, cover, thermal protection) that species need for survival and reproduction (Manly *et al.* 2002). European-level mapping products (e.g. Corine Landcover, Copernicus; see eea.europa.eu) which provide a human representation of the land use are often used by movement ecologists to quantify such resources (e.g. forest cover) by linking these maps with GPS locations obtained from animal tracking projects. Improved spatio-temporal resolution and range of both remote sensing products and animal tracking datasets are allowing ecologists to derive ever more detailed animal trajectories annotated with habitat information, and hence facilitate the study of the animal-habitat relationship over time (Cagnacci *et al.* 2010, Demšar *et al.* 2015, Kays *et al.* 2015). Using concepts from the discipline of time geography (e.g. space-time cubes, space-time prisms), ecologists and geo-informaticians have together developed several temporally-explicit methods to visualise and analyse movement in relation to habitat type and structure (see Demšar *et al.* 2015 for overview). One way of considering time dependency is to investigate sequentiality, which takes into account the temporal order in which behavioural, ecological or movement components occur. Several promising methods have been proposed often relying on approaches developed in different research fields or by combining several concepts. Fourier and wavelet transforms have been used to simultaneously detect repetitive, periodical patterns in ecological variables, such as habitat features, at multiple temporal scales (Wittemyer *et al.* 2008, Bar-David *et al.* 2009, Polansky *et al.* 2010, Riotte-Lambert *et al.* 2013, 2016). Minimal conditional entropy was used to identify a scale of repetitiveness in resource patch visitation (i.e, traplining, Riotte-Lambert *et al.* 2016). Utilisation distribution in combination with residence time was also proposed to identify recursive patterns in space use from movement data (Benhamou and Riotte-Lambert 2012). Finally, the Sequence Analysis Methods (SAM) approach was recently suggested to measure similarity between temporally ordered sequences of habitat use for individuals or populations (De Groeve *et al.* 2016). SAM is a tree-based approach developed in computer science to measure dissimilarity between multiple strings of characters, and has subsequently been used in different fields of study. It was first adopted to measure dissimilarity between DNA and protein sequences. The popularity of the technique in molecular biology resulted in several applications in other fields such as in sociology to study life courses (e.g. Abbott 1995, Wilson 2006, Gabadinho *et al.* 2011), in time geography for transportation science (Wilson 2008), in tourism research (Shoval and Isaacson 2007), in indoor navigation (Delafontaine *et al.* 2012) and in choreography research (Chavoshi *et al.* 2015); and, recently, in the field of animal movement ecology (De Groeve *et al.* 2016).

De Groeve *et al.* (2016) showed that for a given proportion of habitat used, animals can show very different sequential patterns. For example, while animals may equally use open and closed habitats over a given time-window, their sequential use patterns were markedly different (from random to day-night alternating patterns). Based on De Groeve *et* *al.* (2016), we described here an improved methodological framework accounting for individual variability in landscape heterogeneity (i.e. the distribution of resources in a landscape), IM-SAM (Individual Movement - Sequence Analysis Method). In essence, the SAM-framework of De Groeve *et al.* (2016) used simulated sequential habitat use patterns to determine the classification of real habitat use sequences by using movement simulations in artificially generated landscapes that mimic habitat composition and structure of an animal's movement context. Instead, in IM-SAM we use the real landscape context for an individual animal (i.e. its home range) and hence account for the true variation in habitat composition and structure among individuals. Since simulations of sequential patterns generate more complex outputs in real landscapes, the clustering of sequences into trees requires a robust procedure to define cluster cut-offs (i.e. the number of clusters in a tree). With these improvements (i.e. individual based simulations, improved cut-off definition), we could expand the applicability of IM-SAM from animal trajectories derived from a single local context (i.e. single population) to a continental scale (i.e. multiple populations at the European level).

We evaluated this adjusted framework in a case study using GPS movement data of nine European roe deer (*Capreolus capreolus*) populations with contrasting landscape structure from the EuroDEER collaborative initiative (euroungulates.org). As roe deer are generally described as a forest species, but often feed on rich resources available in more open habitats (e.g. meadow, crop), we described sequential use of two simple habitat classes, open and closed, across 404 individual animals and expressed regularised animal trajectories as character sequences, representing the used habitats.

4.2. MATERIAL AND METHODS – HABITAT USE SEQUENTIAL ANALYSIS

To describe sequential use of open and closed habitats for individual animals, we followed a workflow modified from De Groeve *et al.* (2016) that can be summarised in four steps (Fig.4.1). First, for each individual roe deer, we produced an exploratory tree based on biweekly trajectories annotated with habitat categories and used to formulate hypotheses of expected patterns of sequential habitat use (Fig.4.1a). Next, we generated stochastic movement rules for such expected patterns of sequential habitat use, and ran the movement models within each individual's home range in order to produce individual specific simulated trajectories (Fig.4.1b), and individual level simulation trees (Fig.4.1c). Simulation trees were therefore based on the true landscape context where each individual ranged. Finally, we combined real and simulated trajectories to produce trees where real and simulated habitat use sequences with a high degree of similarity were grouped together, so to assign a specific sequential habitat use type to real trajectories (Fig.4.1c/d). Finally, we pruned the output tree and visualised only the classified real trajectories, to facilitate interpretation (Fig.4.1d). We now describe each step in detail.

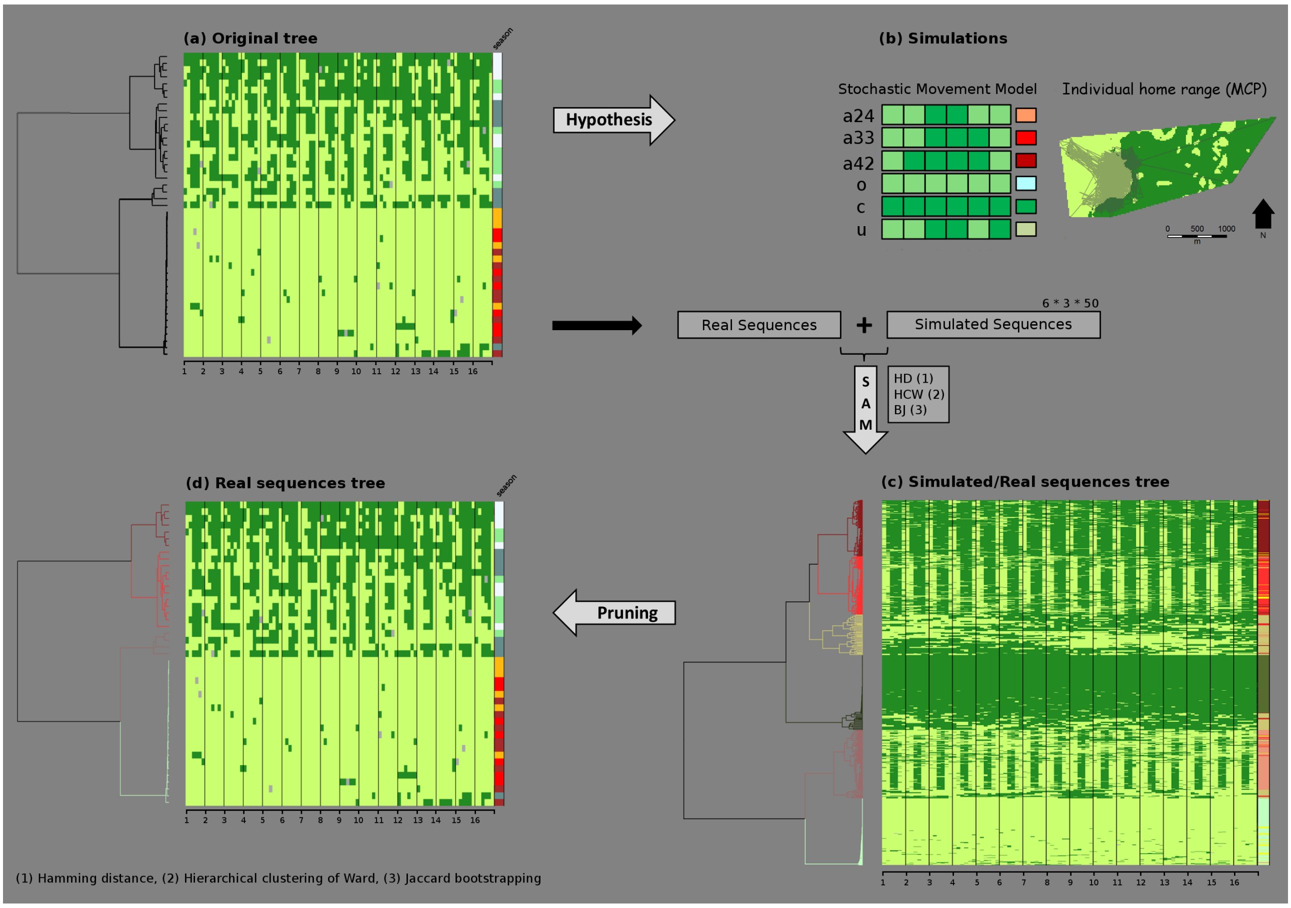


Figure 4.1. Workflow chart of the procedure to classify spatio-temporal habitat use patterns of individual animal trajectories using Individual Movement-Sequence Analysis Method (IM-SAM). Trees represent sequence dissimilarities between habitat use sequences, indicated by the branch lengths. Each tree leave corresponds to one biweekly sequence, which were visualised beside the tree, together with a colour-coded bar representing a variable related to each sequence. See main text for details.

4.2.1. Real trajectories - exploratory trees (Fig.4.1a)

We extracted roe deer trajectories from the EuroDEER database (Cagnacci *et al.* 2011, euroungulates.org) and subsampled them into 16-day GPS trajectories with a fixed 4-hour relocation interval (0, 4, 8, 12, 16 and 20h) over a fixed yearly schedule starting on January 1st (e.g. 01/01-16/01, 17/01-01/02, etc.). To increase the sample size we also included relocations within 1 hour from the above 4-hour intervals, after verifying that this did not affect exploratory tree clustering. If multiple GPS locations were within this time window we selected the closest one in time to the reference time stamps. Instead, if no fix was obtained for a reference time stamp we filled the gap with a missing value (NA). We annotated each 16-day trajectory with the percentage tree cover (0-100%) extracted from the High Resolution Layer-Tree Cover Density 2012 (TCD, EEA 2012, 20m spatial resolution), thus obtaining the *biweekly sequences* of habitat use. We reclassified TCD into two distinct classes, using a cut-off point for pixel-level tree cover density of 50%, distinguishing closed (C, ≥50%) and open (O, <50%) habitat classes. The final dataset consisted of 5,402 biweekly habitat use sequences of 404 animals (236 females and 168 males) of nine European roe deer populations characterised by varying forest composition (see map in Chapter 2, Fig. 2.3: Southcentral Norway, NK5; Southwest France, FR8; Switzerland or Swiss Alps, CH25; Southern Germany, DE15, DE31; Southeast Germany or Bavarian Forest, DE2; Northern Italy or Italian Alps, IT1, IT24; Eastern Austria or Austian Alps, AU17). After processing, the dataset consisted of 14,607 missing values (2.82%) and 503,985 true GPS locations (97.18%), of which 273,230 (52.69%) were classified as open and 230,755 (44.50%) as closed habitat. See Appendix 4A for the complete data preparation procedure and Appendix 4B for the TCD-raster validation and details about the threshold definition.

We obtained an initial visualisation of the habitat use patterns by creating exploratory trees separately for each individual (Fig.4.1a). These trees were used to describe sequential patterns and helped to build hypotheses for expected models of sequential habitat use (see below). Sequence Analysis Methods use a dissimilarity algorithm to compute the distance (dissimilarity) between all possible pairs of sequences. All these pairwise distances are written into a dissimilarity matrix, for example through the Hamming distance algorithm (HD), which calculates the minimum number of character substitutions (i.e. O and C) required to match a number of sequences of equal length (Gabadinho *et al.* 2011). From the HD dissimilarity matrix, we subsequently calculated dissimilarity trees using a hierarchical clustering algorithm (Ward's method, Gabadinho *et al.* 2011, De Groeve *et al.* 2016).

4.2.2. From SAM to IM-SAM

a. Individual-specific simulated trajectories (Fig.4.1b)

We computed individual home ranges as 100% Minimum Convex Polygons (MCP) for each roe deer separately as a simple representation of the available space in which movement of that individual may occur. We then intersected each home range with the TCD raster re-classified as open/closed habitat as described above, and characterised all home ranges by their relative habitat proportion. We obtained both home ranges with a marked prevalence of a given habitat, for example 10% of open and 90% of closed, and home ranges where the proportion of the two habitat categories was similar (for example 40% open and 60% closed). Hence, in the following analyses, we classified home ranges into 5 classes (0.1 to 0.5) of ‘relative proportion of open/closed habitat’ for illustrative purposes (e.g. Fig.4.2). So, for example, the 0.1 class indicates both 10% open - 90% closed, and 10% closed- 90% open habitat, unless otherwise specified.

Within each home range we simulated sequential habitat use patterns using a simple spatially-explicit stochastic movement model to express expected sequential habitat use patterns (see Appendix 4C for the full description). In particular, according to the exploratory trees obtained from real sequences with six locations per day (Fig.4.1a), we identified six characteristic patterns of sequential habitat use (Fig.4.1b): homogeneous use of closed habitat, or pattern 'c', the resulting day-night sequence (DS) being: DS=CCCCCC; homogeneous use of open habitats, or pattern 'o' (DS=OOOOOO); and three patterns of day-night alternating use of both open and closed habitats, or patterns 'a'. The alternating patterns were generated on the assumption that roe deer use open and closed habitats in relation to the daylight cycle, with use of open habitat more prevalent at night (Mysterud *et al.* 1999; Bonnot *et al.* 2013). In addition, we accounted for variation in day length over the different locations included in the study according to the ephemeris. We distinguished the following patterns: dominant use of open habitat from 16:00 to 08:00 (pattern 'a24', DS=OOCCOO), equal use of both habitats - open from 20:00 to 08:00, closed from 08:00 to 20:00 (pattern 'a33', DS=OOCCCO), and dominant use of closed habitat from 04:00 to 20:00 (pattern 'a42', DS=OCCCCO). Finally, we defined a pattern of random use of the landscape (hence reflecting its structure), or pattern 'u', as a control. The seeds of stochastic simulations were random release locations within each individual home range, whereas successive steps were based on six sets of habitat selection rules matching the aforementioned sequential behaviours. For completeness, we ran the movement simulations with three selection coefficients for each selection pattern to account for behavioural variability (selection coefficient ratios: low, 1:0.2; intermediate, 1:0.1; high, 1:0.01). Each movement simulation was repeated 50 times for each parameter set (i.e. 6 habitat selection rules x 3 selection coefficients), varying release location between repetitions but holding release location constant across parameter sets for any given repetition. We thus obtained 900 simulated sequences per individual home range (6 habitat selection rules x 3 selection coefficients x 50 repetitions).

b. Individual-specific simulation trees (Fig.4.1c)

By running the HD dissimilarity algorithm (see above) on the individual-specific simulated sequences, we obtained 404 individual simulated trees, which illustrate the dissimilarity among expected habitat use sequences for each individual separately. Note that HD can be customised by assigning weights to character substitutions when comparing sequences. For example, in De Groeve *et al.* (2016) substitution weights were derived from spatial autocorrelation of habitat classes within simulated landscapes. However, in IMSAM we considered substitution weights to be redundant because simulated sequences were modelled within each individual home range and habitat was classified as a binary category. Hence, spatial structure and thus spatial autocorrelation, were directly accounted for by using this individual-specific modelling approach.

c. Robust classification of individual-specific simulation trees: defining the cut-off distance

In the obtained trees, the leaves are the sequences, and while remaining nodes represent clusters (groups of sequences) for which the branch lengths measure the extent of dissimilarity. In other words, the longer the branch length, the higher the dissimilarity between clusters. Hence, the number of clusters that are identified in a tree depends on a cut-off value that should be selected to obtain the most robust dissimilarity tree (Hennig *et al.* 2007). To assess robustness, an iterative procedure of sequence re-clustering such as bootstrapping is generally used. Bootstrapping metrics express the consistency in reclassification of sequences in the same clusters. The same procedure can be repeated for different cut-off values (and therefore number of clusters), using the optimisation of bootstrapping metrics as a criterion to decide upon the best cut-off point. In IM-SAM, we propose a standardised procedure to identify the most robust and informative tree.

For the trees based on real trajectories only (Fig.4.1a) we did not include any cut-off, since we used them for exploratory purposes only. For the clustering of simulated trajectories only, and both simulated and real trajectories in the final classification (Fig.4.1c), instead, we allowed all cut-off values (that correspond to distances from the last common node) that generated from a minimum of 2 to a maximum of 20 clusters per tree. For each cut-off value, we computed the Jaccard bootstrapping index for each cluster (BJ, or boot-mean; Hennig 2007) using 1000 iterations (De Groeve *et al.* 2016), and then we computed the median BJ of all clusters $(BJ_{MD} \in [0,1])$, and the BJ interquartile range $(BJ_{IQ} \in [0,1])$. We finally defined a combined bootstrapping index $\mathrm{BJ}_{\mathrm{IQMD}}$ $(BJ_{IQMD} \in [0,1])$ that we computed for each cut-off value:

$$\mathrm{BJ}_{\mathrm{IQMD}} = \frac{\mathrm{BJ}_{\mathrm{MD}} + (1 - \mathrm{BJ}_{\mathrm{IQ}})}{2}$$

To evaluate the optimised cut-off value, we used a semi-automated selection procedure based on the optimisation of the $\mathrm{BJ}_{\mathrm{IQMD}}$ index. Specifically, we plotted $\mathrm{BJ}_{\mathrm{IQMD}}$ in relation to the number of clusters for each individual (Fig.4.2, top panel and Appendix 4D, Fig.4D.1). In most cases, the plot showed two peaks: a primary peak, corresponding to a cut-off value that generated trees with two to three clusters, that separate sequences based on general dissimilarity (for example: homogeneous sequences from all others); and a secondary peak, corresponding to a cut-off value that generated trees with five to eight clusters, catching the complexity of the sequences, i.e. the diversity of sequential habitat use patterns generated by the simulations (see also De Groeve *et al.* 2016). Hence, we excluded the primary peak and defined the cut-off value based on maximum values of $\mathrm{BJ}_{\mathrm{IQMD}}$ within the secondary peak (Fig.4.2, upper panel, light blue region). Then, we did a visual check of all trees derived from the cut-off values selected as above to identify potential inconsistencies.

d. Identification of cluster types ('cluster tagging')

Once we obtained the 'optimal' tree, we classified each cluster on the basis of the sequential patterns that composed the cluster ('cluster tagging'). First, we calculated the proportion of each sequential habitat use pattern in each cluster. Then, to identify cluster types, we appended each pattern representing at least 10% of the cluster to a tag, ordered by abundance (first criterion) and giving priority to homogeneous patterns (second criterion). For example, if a cluster consisted of 40% a24 sequences, 30% a33 sequences, and 20%, a42 sequences, the resulting tag was a24_a33_a42 (first criterion); or, a composition of 30% of O sequences, 30% of a24, 20% of a33, and 15% of a42, was tagged as O_a24_a33_a42 (second criterion). We also used a simplified classification by retaining the most abundant sequential habitat use pattern only as the cluster tag (e.g. a24 and O in the two examples above, respectively).

4.2.3 Simulated and observed trajectories - classification trees (Fig.4.1c/1d)

In a final step, we obtained the classification tree for each individual by re-running the HD algorithm on both real and simulated sequences, and by using the cut-off distance as defined above. This way, simulated sequences were used as a *guide for classification of real sequences* to their most similar cluster type (real habitat use sequence tagging, Fig.4.1c). For visualisation purposes, we pruned the classification tree by only retaining real sequences (Fig.4.1d).

4.3. RESULTS

4.3.1. Classification of individual-specific simulation trees

The application of the HD algorithm to simulated biweekly sequences generated 404 trees, one per individual home range. According to the simulation procedure (par.4.2.2a, Fig.4.1b)), we expected trees to be composed of 6 clusters with 150 sequences each. Instead, we found a substantial deviation from this expectation (Appendix 4E) that we attributed to individual differences in the environmental context, and notably to the relative proportion of open vs. closed habitat within the individual’s home range.

Using the IM-SAM cut-off criterion (i.e. the second peak in the plot of BJ_{IQMD} for all cut-off values; Fig.4.2, top panel), we automatically identified 394 simulation trees, mainly composed of five (164 trees, 40.59%), six (192 trees, 47.52%) or seven (38 trees, 9.41%) clusters, whereas trees with three and eight clusters were rare (7 and 3 respectively, or 1.73% and 0.74%). The overall average BJ_{IQMD} of these trees was 0.749±0.111 which corresponded to a high BJ_{MD} (0.850±0.066), indicating high inter-cluster dissimilarity, and a low BJ_{IQ} (0.123±0.078), indicating low variability in inter-cluster dissimilarity, thus confirming the robustness of the cluster classification. For 32 trees, the number of clusters was manually adjusted to distinguish clear and obvious clusters (from five or six to six, seven, or eight clusters), but this resulted in a negligible change in average BJ_{IQMD} (decrease of 0.007, 0.742±0.115). Furthermore, we noted that the relative proportion of open vs. closed habitat in the home range affected cluster quality (i.e. clustering robustness). Specifically, BJ_{IQMD} increases when the proportion of closed and open habitats is more or less balanced (Fig.4.2, top panel: higher to lower values from dark to light trend lines). The corresponding BJ_{MD} values also increase when the proportions of open and closed are similar (Fig.4.2, top panel: green shade on the trend lines), whereas the corresponding BJ_{IQ} values decrease (Fig.4.2, top panel: orange to green points along the trend lines). Despite these differences between classes of habitat proportion, the trend in BJ_{IQMD} was consistent, with a second peak for values of between five and seven clusters, and a rapid drop after that.

The proportion of open vs. closed habitat in the home range also affected the cluster size within simulation trees. When the proportion of open and closed habitat in the home range was similar, the trees for simulated sequences were more evenly classified between different clusters, whereas when a given habitat type was preponderant in the home range, some clusters were composed of a larger number of sequences than others (Fig.4.2, lower panel, e.g. median lower than 150, and more outliers for class 0.1).

The above indicates that it is easier to distinguish different sequential habitat use patterns when the relative proportion of habitats available to the individual is similar.

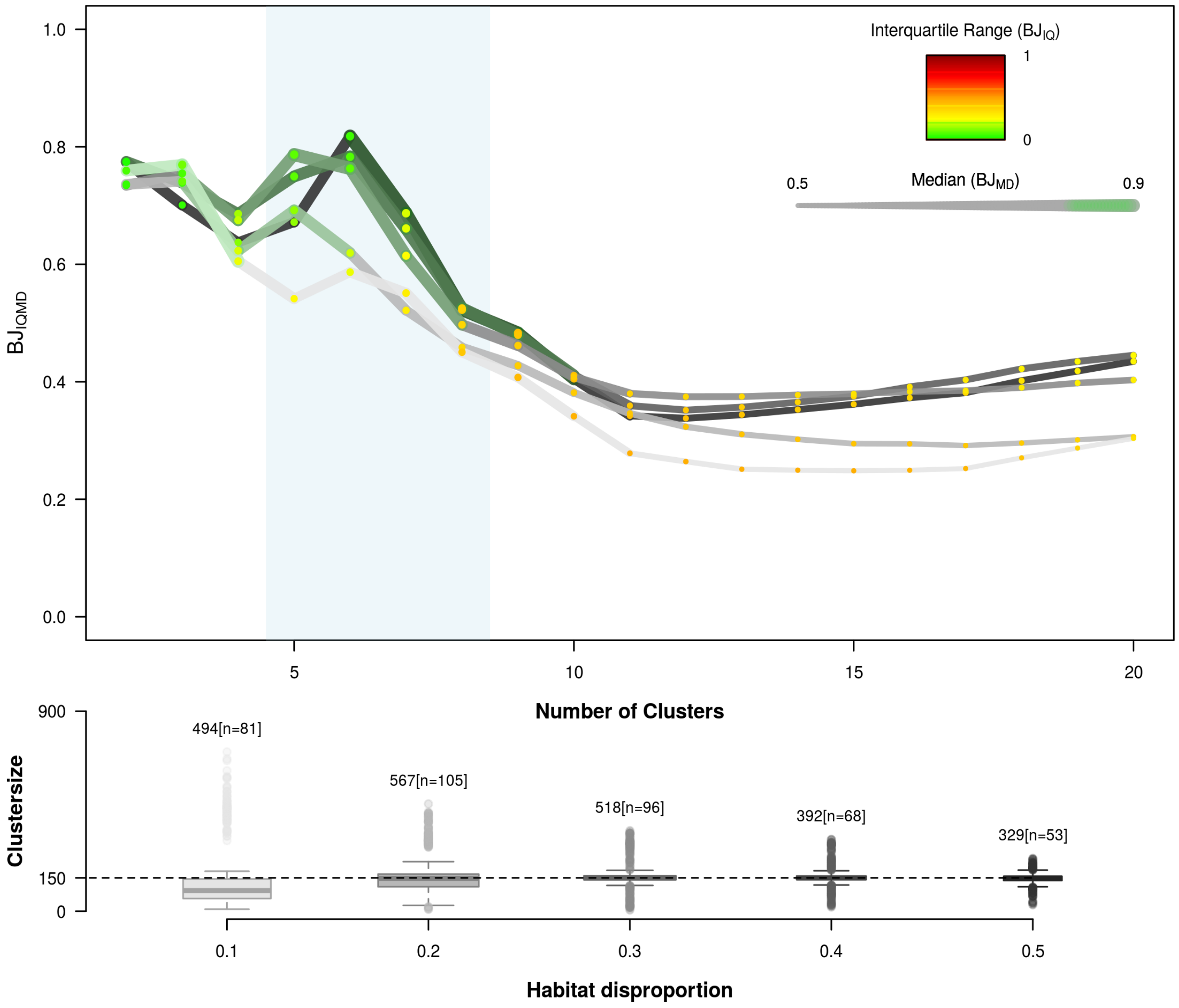


Figure 4.2. Top plot: Trend lines representing the combined bootstrap index BJ_{IQMD} (combining the bootstrap median BJ_{MD} and the bootstrap interquartile range BJ_{IQ}), for different cut-offs (2-20 clusters) in all 404 individual-based simulation trees. The output BJ_{IQMD} values are classified by the relative habitat proportion in the individual home ranges (gradient from light to dark grey, from 0.1 to 0.5). The plot also represents the contribution of BJ_{MD} and BJ_{IQ} to the combined index BJ_{IQMD}. Specifically, the thickness of the lines corresponds to BJ_{MD}; when BJ_{MD} is larger than 0.8, a green transparency is added to the grey BJ_{IQMD} trend lines. The colour of the dots along the trend lines represents BJ_{IQ}, with values decreasing from orange to green (bright green: $BJ_{IQ} < 0.2$). The transparent light blue region is the window corresponding to the second peak in BJ_{IQMD} that was chosen as the cut-off criterion for final simulation trees. **Bottom plot:** Boxplots visualise how the total 900 sequences simulated for each 404 home ranges are distributed between clusters, when the cut-off based on BJ_{IQMD} is used to define the corresponding simulated trees. The trees are classified by the relative habitat proportion in the home ranges (from 0.1, to 0.5 - gradient of grey as in the top panel). When there is no habitat preponderance (i.e. relative habitat proportion of 0.4 or 0.5), the sequences are almost equally distributed between clusters (i.e. median cluster size close to 150, with very few outliers).

4.3.2. Identification of cluster types ('cluster tagging')

We identified 16 main cluster types among the 404 simulated trees, each representing more than 0.5% of all clusters (Fig.4.3, circles and bars in the lower panel). The 2300 clusters consisted of one or more sequential habitat use patterns, in differing proportions. Specifically, 1613 clusters consisted of one sequential habitat use pattern (70% of total number of clusters, Fig.4.3, top-left panel; a24: 378 clusters; a33: 272; a42: 386; c: 290; o: 170; u: 125). 513 (22%) and 139 (6%) clusters consisted of two or three sequential habitat use patterns, respectively (o_u: 170 clusters; a33_a42: 125; c_u: 89; a33_a24: 52; a24_a33: 29; a42_a33: 21; u_a33: 13; o_u_a24: 84; c_u_a42, 27). The remaining 33 cluster types represented less than 3% of all clusters in total and were also used to classify real trajectories (these rare cluster types were omitted from Fig.4.3; see Appendix 4F for the full set of classified cluster types).

The number of sequential habitat use patterns occurring in clusters was affected by the relative proportion of open vs. closed habitat in the home range, with clear-cut cluster identification (i.e. one pattern per cluster) for trees derived from home ranges with similar relative proportions of open and closed habitat, and more 'unclear' cluster identification (i.e. with 3 to 4 sequential habitat use patterns) for trees derived from home ranges with a preponderance of one habitat type only (Fig.4.3, top-left panel).

Similarly, the occurrence and relative importance of cluster types within trees also depended on the relative proportion of open vs. closed habitat in the home range (Fig.4.3, top-right panel). For example, cluster types with one alternating sequential habitat use pattern (tags: a24, a33, a42) occurred in trees corresponding to home ranges irrespective of habitat composition, whereas the combination of homogeneous closed, random and alternation a42 (tag: c_u_a42) was only recorded in trees where the home range contained less than 40% open habitat. Indeed, sequences with random and alternating habitat use patterns grouped together with homogeneous sequences when one habitat was prevalent in the home range (Fig.4.3, top-right panel; the patterns are mirrored for high proportion of open or closed habitat). Importantly, only a small proportion of alternating patterns clustered together with a random pattern of use, indicating that alternating patterns rarely occurred at random.

To sum up, we have shown that sequences of a-priori defined habitat use patterns generated within individual home ranges mostly clustered amongst themselves, as expected, but there was some variation in the pattern. Clusters were characterised by one or more sequential habitat use pattern, and this ‘cross-pattern’ clustering was dependent on the relative composition of open and closed habitats in the home range.

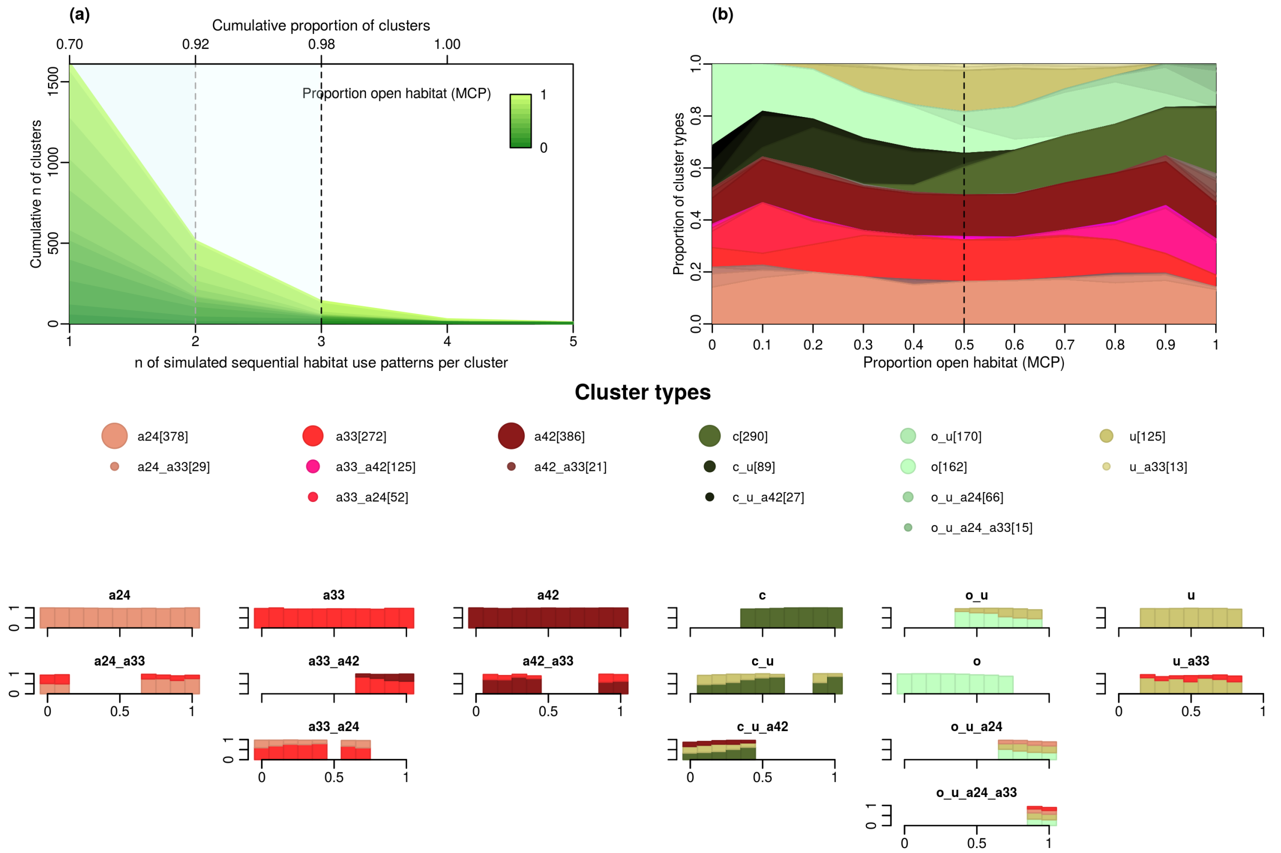
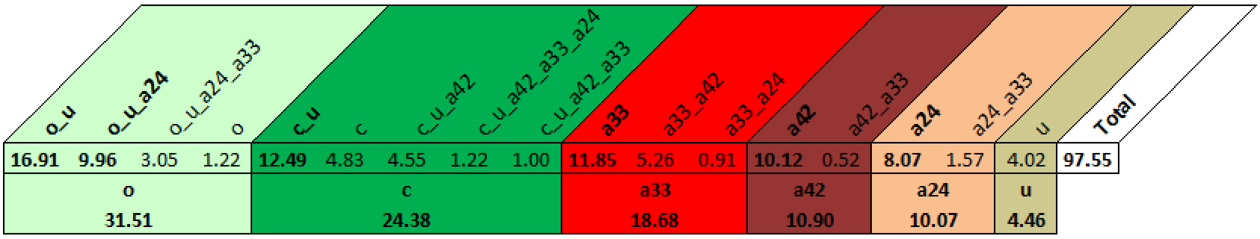


Figure 4.3. Bottom panel - Cluster types: coloured circles represent all sixteen main cluster types identified in simulation trees, scaled by the number of clusters of that type (actual number of clusters in brackets). The colour is the legend for the top-right figure (see below). The horizontal bars indicate the relative proportion of each sequential habitat use pattern in each cluster type (vertical axis) and the occurrence of each cluster type in dependence on open habitat proportion in home ranges (MCP, from 0 to 1). **Top panel left – Cluster composition (a)**: frequency of occurrence of cluster types composed by up to five sequential habitat use patterns. The coloured gradient indicates the proportion of open habitat in home ranges on which simulation trees were based. **Top panel right – Cluster composition (b)**: relative proportion of cluster types in simulated trees (main y-axis), plotted against the proportion of open habitat in home ranges (main x-axis).

4.3.3. Classification of real animal trajectories into cluster types

After classification of simulated sequences, we re-ran SAM also including real trajectories so that these were grouped with the most similar cluster types. After pruning (i.e. filtering out of the simulated sequences), we obtained the classification tree of the real sequences for each individual (See Fig.4.1d for an example, and Appendix 4G for a sample of 35 classification trees). In total, 69.40% of the real habitat use sequences were classified into only 6 cluster types (i.e. o_u, o_u_a24, c_u, a33, a42, a24, Table 4.1, in bolt), and 97.55% of all real sequences into 17 cluster types (Table 4.1). The remaining 2.45% of all real habitat use sequences matched another 21 cluster types. All six a-priori sequential habitat use simulated patterns were represented amongst real sequences. Specifically, 40.11% of the sequences were classified as one of the 6 cluster types including a single sequential habitat use pattern (for example, 4.83% of the real sequences were classified as homogeneous closed, and 11.85% were classified as alternation a33, Table 4.1). Another 38.20% were classified into cluster types that included a combination of two patterns, especially a combination of homogeneous open/closed with random patterns (i.e. 29.40%, o_u, c_u, Table 4.1), or a combination of two alternating patterns (i.e. 8.26%, a33_a42, a33_a24, a42_a33, a24_a33). Finally, a remaining 21.69% of the real sequences were classified into cluster types that included a combination of three or more sequential habitat use patterns. When these results were considered with the most simplified classification (i.e. retaining the most abundant pattern only), the majority of the sequences were classified as homogeneous open (o, 31.51%), and homogeneous closed (c, 24.38%). More than 40% of the sequences were classified as the three types of habitat alternation (a33, 18.68%; a42, 10.90%; a24, 10.07), while the smallest proportion of habitat patterns corresponded to random sequential use of habitat (u, 4.46%).

Table 4.1. Percentage of real sequences classified in the 17 most common cluster types (different from the most common cluster types for simulated sequences), and in the simplified classification.



o_u	**o_u_a24**	o_u_a24_a33	o	**c_u**	c	c_u_a42	c_u_a42_a33_a24	c_u_a42_a33	**a33**	a33_a42	a33_a24	**a42**	a42_a33	**a24**	a24_a33	u	**Total**
16.91	**9.96**	3.05	1.22	**12.49**	4.83	4.55	1.22	1.00	**11.85**	5.26	0.91	**10.12**	0.52	**8.07**	1.57	4.02	**97.55**
o **31.51**				**c** **24.38**					**a33** **18.68**			**a42** **10.90**		**a24** **10.07**		**u** **4.46**	

4.4. DISCUSSION

In this paper, we propose a generalisable ecological application of Sequence Analysis Methods, IM-SAM to describe sequential habitat use of GPS-tracked animals. Below, we first discuss the methodological advances of IM-SAM. Second, we consider in further detail the ecological relevance of variation in the observed spatio-temporal patterns of roe deer sequential habitat use across Europe. Finally, we discuss how IM-SAM can be used for other ecological and geographical applications.

4.4.1. IM-SAM-procedure

IM-SAM provides a suitable method to detect similarity in sequential patterns in movement data of animal species. The IM-SAM framework involves three methodological steps. First, exploration trees are built using real sequences only (Fig.4.1a). Then simulation trees are generated taking into account the individual spatial context using simulated sequences only (Fig.4.1b/c). Finally, classification trees are produced based on real and simulated sequences combined (Fig.4.1c/d). One of the most important advances of IM-SAM, compared to other previous ecological applications of the SAM framework (De Groeve *et al.* 2016), is to account for the spatial context in which an animal moves, by generating individual-specific, spatially-explicit simulated sequences. In this way, individual sequential patterns of habitat use can be extracted in a standardised and comparable manner across a diversity of landscapes, as done here, facilitating multi-population comparisons. In this study, we generated simulations based on day-night habitat use patterns. This might not be optimal for other studies. The simulation rules must be based on the question addressed, on the behavioural traits of the species, and the spatio-temporal resolution of the study. For example, when studying migration-timing and the use of summer vs winter ranges, simulations might be better based on a weekly timeframe. The IM-SAM procedure only detects sequential patterns that are coded within the simulation rules. While this insures the robustness of the approach, it requires an attentive evaluation of the exploration trees to define the expected sequential habitat use patterns. In this sense, exploratory trees represent the empirical observations on which to build the set of hypotheses. Moreover, exploratory trees allow the identification of other repetitive patterns in a dataset, such as the occurrence of missing data that are often associated with particular habitats (e.g. forest, topographic complexity) and/or behaviours (e.g. resting; Frair *et al.* 2010). They also allow to determine associations between given sequential patterns and other environmental variables, such as the season (like in our example Fig.4.1a).

In comparison with the simplified model used in De Groeve *et al.* (2016), we here simulated sequences within the true landscape context giving more complex and variable patterns, and hence resulting in more complex dissimilarity trees requiring methodological refinements to obtain robust classifications. Specifically, we needed a better definition for the cut-off value for cluster definition. Indeed, the optimal number of clusters can be derived using many different approaches. Several SAM-applications (e.g. Shoval and Isaacson 2007) define the cut-off value by visual exploration of clusters in dissimilarity trees without considering the robustness of the clusters. More objective methods use within- and between-cluster quality assessments, such as silhouette plots (Rousseeuw *et al.* 1987), or, as often used in DNA-analysis, cluster stability procedures based on bootstrapping (e.g. Jaccard bootstrapping, BJ). Due to a large sample size, we not only needed an improved method, but also a semi-automated standardised approach. Here, we extended the approach of De Groeve *et al.* (2016), based on median values in BJ, by also accounting for the interquartile range in BJ, and defining a combined bootstrapping index. In our case, the sequences corresponded to simulated behaviours (i.e. discrete trajectories) that were obtained through a set of stochastic rules, applied to real and highly heterogeneous environments. Hence, some variability in the output sequences, and so in the clustering, can be expected, especially when the proportion of alternative habitats is highly unequal within individual home ranges (see Fig.4.3). For this reason, when evaluating the threshold for the ‘optimal’ classification into trees, we combined a central tendency (BJ_{MD}), and a dispersion measure (BJ_{IQ}) into a unique index. We think that this procedure could be appropriate for other SAM applications.

Another potential application of IM-SAM is to identify the spatio-temporal patterns of habitat use present in clusters with tags that are assigned by a standard set of rules. Such ‘tags’ can be used as levels of a categorical variable in established statistical modelling frameworks, such as multivariate statistics (e.g. Jongman *et al.* 1995) or Generalised Linear Modelling (e.g. Pinheiro and Bates 2000; Zuur *et al.* 2009). SAM was originally applied to the ecological context as a spatio-temporal exploratory tool (De Groeve *et al.* 2016). IM-SAM takes this a step forward, opening the potential to use spatio-temporal patterns as a variable in spatial ecological modelling. For example, these patterns can be studied in relation to environmental drivers, such as temperature, season, presence of predators, anthropogenic factors, such as hunting regime or agricultural practices, population level characteristics, such as density, and individual life-history traits, e.g. sex, age, or reproductive success. Life-history traits is any trait characterising upcoming events throughout the lifetime of an animal. For instance, being male or female will result in different choices and behaviours throughout the whole life cycle.

One of the most important advantages of IM-SAM is the possibility to express sequences as multi-level habitat categories, although here we used a simple case of two alternative habitat types (open vs. closed). Note that for continuous or discretised habitat variables, which are expected to be spatially correlated, substitution weights are essential to correct for classes that are more similar to each other. For example, in case of four habitat classes with different forest cover density (<25%, 25-50%, 50-75%, >75%) a forest cover density of <25% is more similar to the 25-50% category than to the >75% one. In the R package TraMineR such substitution weights can be user-defined or assessed automatically based on sequence characteristics (Gabadinho *et al.* 2011). While automatic computation of substitution weights is sufficient for exploration trees, we recommend assessing them directly by measuring spatial correlation at the relevant scale (i.e. median step length; see De Groeve *et al.* 2016) for simulation and classification trees. However, while the exploration phase can handle a large number of classes, we recommend minimising the number of levels of habitat categories to the most essential for simulation and classification. Indeed, a higher number of levels generate more complex simulated sequential habitat use patterns. For example, in our simplified case study with two habitat types, we identified six expected sequential habitat use patterns.

4.4.2. Ecological insights and geographical applications

Insights derived from IM-SAM can be used to better understand the impact of human land use on animal habitat use, since both animals and humans move in the same landscape (Demšar *et al.* 2015). For example, in our study, a large proportion of real habitat use sequences of roe deer that were classified as homogeneous open, meaning that a non-negligible number of deer intensively used human-exploited agricultural lands (i.e. crops: Southwest France, Southern Germany; husbandry: Switzerland; Aulak & Babinska-Werka 1990). Indeed, agricultural areas may simultaneously provide both high-quality food and cover resources for roe deer, at least during certain parts of the year (Hewison *et al.* 2001, Cimino & Lovari 2006, Bjørneraas *et al.* 2011). Homogeneous closed sequences, on the other hand, were more common in forest landscapes but also occurred in agricultural landscapes, suggesting that some individuals are strictly associated to a given habitat type. Interestingly, 40% of the real sequences from all roe deer populations were classified as alternating patterns between open and closed habitats. This suggests that not only the proportion, but also the sequential order in which open and closed habitats are used, is an important metric for characterising the space use strategy of individual roe deer.

Activity and physiological circadian cycles, such as feeding-rumination, may explain the observed alternation between open and closed habitats. Indeed, roe deer are known mainly to select for forest and cover habitats during rumination and resting (Cederlund 1981), and to favour edges and open areas at peak of foraging activity at twilight and during night (Pagon *et al*. 2013). This pattern may be less pronounced in areas with less human disturbance, such as for a Canadian Elk population (Ensing *et al*. 2014). Indeed, in human-dominated European landscapes, habitat alternation is likely a behavioural response of ungulates to landscape heterogeneity and human activities. Because rich open landscapes are often associated with higher risk of predation or disturbance, in such human-dominated environments, prey species must generally trade their acquisition of high-quality resources against risk avoidance. By alternating between rich open areas and more closed forest habitats, with less forage but a higher degree of shelter, prey may hence resolve the risk-resource trade-off (Fraser and Huntingford 1986). In particular, wild ungulate species, including roe deer, generally use closed refuge habitats during daytime, when human disturbance is greater, and rich open habitats during nighttime (e.g. Bonnot *et al*. 2013; Padié *et al*. 2015). Finally, our results indicate that most of the time, the sequential use of open and closed habitats by roe deer were not random, since only a small proportion of real sequences indicated a random pattern of habitat use.

Human-driven landscape modifications have become an integral part of ecosystems, especially in Europe. Results of this paper strongly suggest that urban planning and landscape management plans should account for the spatio-temporal habitat requirements of key animal species, and not only the ‘time-compounded’ habitat suitability (e.g. through occurrence-based Species Distribution Models). This shall be especially important in the evaluation of habitat connectivity, at different spatial scales. For example, infrastructure developments should take into account sequential patterns of habitat use to ensure that barriers have limited effects on animal movements (Kämmerle *et al.* 2017).

While IM-SAM was applied here on animal habitat use sequences, human geography may also benefit from this novel framework. From a technical point of view, while several

Sequence Analysis studies exist in time geography, to the best of our knowledge, robust classifications such as those obtained by bootstrapping and the use of BJ_{IQMD} have rarely been explored. Moreover, the routine applicability of IM-SAM could be supported by the use of a popular data analysis software (R version 3.4.1., R Core Team 2017; package TraMineR, Gabadinho *et al*. 2011). Conceptually, with the ongoing advances in human and animal tracking techniques, IM-SAM could ultimately be used as a tool to simultaneously compare patterns of space use in animals and humans. For example, mapping sequential animal and human space use in the same area could help understand if and how they differ or conflict. Alternatively, potential effects of traffic, recreation, hiking, cycling and other human activities could be assessed by modelling them as environmental drivers of sequential habitat use. Furthermore, after characterising animal and human sequential space use, one could explore the sequential pattern of non-movement related metrics obtained through biologging, such as activity, heart rate, body temperature, or food intake (see Ropert-Coudert *et al*. 2005).

DECLARATIONS

Acknowledgements

We thank Fondazione Edmund Mach and Special Research Fund (BOF) of Ghent University for their financial support to JDG under grant number 01sf2313. We thank Research Foundation – Flanders (FWO) for the travel grant for study abroad to Harvard University under grant number V417616N to JDG. FC was supported by the Sarah and Daniel Hrdy Fellowship during part of the development of this ms. We wish to thank the EuroDEER collaborative project to provide geospatial and environmental data (paper no. 008 of the EuroDEER series; www.euroungulates.org). The EuroDEER spatial database is hosted by Fondazione Edmund Mach. We are grateful to the Copernicus project for the use of free data (European Union, EEA - European Environment Agency). Here we confirm that all conditions are met required for free, full and open access to this data set, established by the Copernicus data and information policy Regulation (EU) No 1159/2013 of 12 July 2013.. We are particularly grateful to Ferdinando Urbano for his support with data management, and to Lino Ometto for precious advice in data exploration. The movement model was developed by Nathan Ranc under guidance of Prof. dr. Paul Moorcroft and Dr. Francesca Cagnacci. The GPS data collection of the Fondazione Edmund Mach was supported by the Autonomous Province of Trento under grant number 3479 to FC (BECO-CERWI—behavioural Ecology of Cervids in Relation to Wildlife Infections), and by the help of the Wildlife and Forest Service of the Autonomous Province of Trento and the Hunting Association of Trento Province (ACT). The Norwegian participation was funded by the Research Council of Norway (grant number 251112). The GPS data collection of the CEFS-INRA was supported by the "Move-It" ANR grant ANR-16-CE02-0010-02.

Data availability statement

Data availability statement, The data that support the findings of this study will be openly available in Zenodo at https://doi.org/10.5281/zenodo.1254230

REFERENCES

Abbott, A., 1995. Sequence analysis: new methods for old ideas. *Annual review of sociology*, **21** (1), 93-113.

Andersen, R., Duncan, P., and Linnell, J.D.C., 1998. *The European roe deer: the biology of success.* Oslo: Scandinavian University Press.

Aulak, W. and Babińska-Werka, J., 1990. Use of agricultural habitats by roe deer inhabiting a small forest area. *Acta Theriologica*, **35** (1-2), 121-127.

Bar-David, S., Bar-David, I., Cross, P.C., Ryan, S.J., Knechtel, C.U. and Getz, W.M., 2009. Methods for assessing movement path recursion with application to African buffalo in South Africa. *Ecology*, **90** (9), 2467-2479.

Benhamou, S. and Riotte-Lambert, L., 2012. Beyond the Utilization Distribution: Identifying home range areas that are intensively exploited or repeatedly visited. *Ecological Modelling*, **227**, 112-116.

Berger-Tal, O. and Bar-David, S., 2015. Recursive movement patterns: review and synthesis across species. *Ecosphere*, **6** (9), 1-12.

Bissonette, J. A. and Storch, I. 2003. Part 1: Conceptual and Quantitative Linkages. *In:* Bissonette, J. A. and Storch, I., eds. *Landscape Ecology and Resource Management: Linking Theory with Practice*. Washington DC: Island Press, 11-14.

Bjørneraas, K., Solberg, E.J., Herfindal, I., Moorter, B.V., Rolandsen, C.M., Tremblay, J.P., Skarpe, C., Sæther, B.E., Eriksen, R. and Astrup, R., 2011. Moose Alces alces habitat use at multiple temporal scales in a human-altered landscape. *Wildlife Biology*, **17** (1), 44-54.

Bonnot, N., Morellet, N., Verheyden, H., Cargnelutti, B., Lourtet, B., Klein, F. and Hewison, A.J.M., 2013. Habitat use under predation risk, hunting, roads and human dwellings influence the spatial behaviour of roe deer. European Journal of Wildlife Research, **59**, 185-193.

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

Cagnacci, F., Focardi, S., Heurich, M., Stache, A., Hewison, A.J.M., Morellet, N., Kjellander, P., Linnell, J.D., Mysterud, A., Neteler, M. and Delucchi, L., 2011. Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors. *Oikos*, **120** (12), 1790-1802.

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** (3-4), 516-519.

Cederlund, G., 1981. Daily and seasonal activity pattern of roe deer in a boreal habitat [*Capreolus capreolus*, Sweden]. *Swedish Wildlife Research,* **11** (8), 315-353.

Chavoshi, S.H., De Baets, B., Neutens, T., De Tré, G. and Van de Weghe, N., 2015. Exploring dance movement data using sequence alignment methods. *PloS one*, **10** (7), e0132452.

Chapron, G., Kaczensky, P., Linnell, J.D., von Arx, M., Huber, D., Andrén, H., López-Bao, J.V., Adamec, M., Álvares, F., Anders, O. and Balčiauskas, L., 2014. Recovery of large carnivores in Europe’s modern human-dominated landscapes. *Science*, **346** (6216), 1517-1519.

Cibien, C., Bideau, E., Boisaubert, B. and Maublanc, M.L., 1989. Influence of habitat characteristics on winter social organisation in field roe deer. *Acta Theriologica*, **34** (14), 219-226.

Cimino, L. and Lovari, S., 2003. The effects of food or cover removal on spacing patterns and habitat use in roe deer (Capreolus capreolus). Journal of Zoology, **261** (3), 299-305.

Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora (The Habitat Directive).

De Groeve, J., Van de Weghe, N., Ranc, N., Neutens, T., Ometto, L., Rota-Stabelli, O. and Cagnacci, F., 2016. Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods. *Methods in Ecology and Evolution*, **7** (3), 369–379.

De Groeve, J., Cagnacci, F., Ranc., N., Bonnot., N., Benedikt, G., Marco, H., Kroeschel, M., Linnell, J., Hewison, A.J.M., Morellet, N., Mysterud, A., Sandfort, R., Van de Weghe, N. 2018. Data from: *Individual Movement - Sequence Analysis Method (IM-SAM): characterising spatio-temporal patterns of animal trajectories across scales* *and landscapes* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.1254230

Delafontaine, M., Versichele, M., Neutens, T. and Van de Weghe, N., 2012. Analysing spatiotemporal sequences in Bluetooth tracking data. *Applied Geography*, **34**, 659-668.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., Weiskopf, D. and Weibel, R., 2015. Analysis and visualisation of movement: An interdisciplinary review. *Movement ecology*, **3** (5), 1-24.

Directive 2009/147/EC of the European Parliament and of the Council of 30 November 2009 on the conservation of wild birds (The Bird Directive).

Dupke, C., Bonenfant, C., Reineking, B., Hable, R., Zeppenfeld, T., Ewald, M. and Heurich, M., 2017. Habitat selection by a large herbivore at multiple spatial and temporal scales is primarily governed by food resources. *Ecography*, **40** (8), 1014-1027.

Ensing, E.P., Ciuti, S., de Wijs, F.A., Lentferink, D.H., ten Hoedt, A., Boyce, M.S. and Hut, R.A., 2014. GPS based daily activity patterns in European red deer and North American elk (Cervus elaphus): indication for a weak circadian clock in ungulates. *PLoS One*, **9** (9), e106997.

Forester, J.D., Ives, A.R., Turner, M.G., Anderson, D.P., Fortin, D., Beyer, H.L., Smith, D.W. and Boyce, M.S., 2007. State–Space models link elk movement patterns to landscape characteristics in yellowstone national park. *Ecological Monographs*, **77** (2), 285-299.

Frair, J.L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N.J. and Pedrotti, L., 2010. Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **365** (1550), 2187-2200.

Fraser, D.F. and Huntingford, F., 1986. Feeding and avoiding predation hazard: The behavioral response of the prey. *Ethology*, **73**, 56e68.

Jongman, R.H..G., Ter Braak, C.J.F. and Van Tongeren, O.F.R., 1995. *Data analysis in community and landscape ecology*. Cambridge: Cambridge University Press.

Gabadinho, A., Ritschard, G., Mueller, N.S. and Studer, M., 2011. Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, **40** (4), 1-37.

Gaillard, J.M., Hebblewhite, M., Loison, A., Fuller, M., Powell, R., Basille, M. and Van Moorter, B., 2010. Habitat-performance relationships: finding the right metric at a given spatial scale. *Philosophical Transactions of the Royal Society B: Biological Sciences,* **365** (1550), 2255-2265.

Godvik, I.M.R., Loe, L.E., Vik, J.O., Veiberg, V., Langvatn, R. and Mysterud, A., 2009. Temporal scales, trade-offs, and functional responses in red deer habitat selection. *Ecology* **90** (3), 699-710.

Hanski, I., 2005. Landscape fragmentation, biodiversity loss and the societal response – The longterm consequences of our use of natural resources may be surprising and unpleasant. *EMBO reports*, **6** (5), 388-392.

Hennig, C., 2007. Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis*, **52**, 258-271.

Herfindal, I., Tremblay, J.P., Hansen, B.B., Solberg, E.J., Heim, M. and Sæther, B.E., 2009. Scale dependency and functional response in moose habitat selection. *Ecography,* **32** (5), 849-859.

Hewison, A.J.M., Vincent, J.P., Joachim, J., Angibault, J.M., Cargnelutti, B. and Cibien, C., 2001. The effects of woodland fragmentation and human activity on roe deer distribution in agricultural landscapes. *Canadian journal of zoology*, **79** (4), 679-689.

Jaeger, J.A., Soukup, T., Schwick, C., Madriñán, L.F. and Kienast, F., 2011. Landscape fragmentation in Europe, *European Environment Agency, Federal Office for the Environment, EEA-FOEN.*

Johnson, D.H., 1980. The Comparison of usage and availability measurements for evaluating resource preference. *Ecology,* **61**, 65-71.

Kämmerle, J.L., Brieger, F., Kroeschel, M., Hagen, R., Storch, I. and Suchant, R., 2017. Temporal patterns in road crossing behaviour in roe deer (Capreolus capreolus) at sites with wildlife warning reflectors. *PLoS One,* **12** (9), 1-18.

Kays, R., Crofoot, M.C., Jetz, W. and Wikelski, M., 2015. Terrestrial animal tracking as an eye on life and planet. *Science* 348, aaa2478.

Krauss, J., Bommarco, R., Guardiola, M., Heikkinen, R.K., Helm, A., Kuussaari, M., Lindborg, R., Öckinger, E., Pärtel, M., Pino, J. and Pöyry, J., 2010. Habitat fragmentation causes immediate and time-delayed biodiversity loss at different trophic levels. *Ecology letters*, **13** (5), 597-605.

Magle, S.B., Hunt, V.M., Vernon, M. and Crooks, K.R., 2012. Urban wildlife research: past, present, and future. *Biological Conservation*, **155**, 23-32.

Mancinelli, S., Peters, W., Boitani, L., Hebblewhite, M. and Cagnacci, F., 2015. Roe deer summer habitat selection at multiple spatio-temporal scales in an Alpine environment. *Hystrix, the Italian Journal of Mammalogy*. **26** (2), 132-140.

Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L. and Erickson, W.P., 2002. Introduction to Resource Selection Studies. *In: Resource selection by animals: statistical design and analysis for field studies*. Dordrecht: Kluwer Academic Publishers, 1-15.

Morellet, N., Bonenfant, C., Börger, L., Ossi, F., Cagnacci, F., Heurich, M., Kjellander, P., Linnell, J.D.C., Nicoloso, S., Sustr, P. and Urbano, F., 2013. Seasonality, weather and climate affect home range size in roe deer across a wide latitudinal gradient within Europe. *Journal of Animal Ecology*, **82** (6), 1326-1339.

Mysterud, A., Lian, L.B. and Hjermann, D.Ø., 1999a. Scale-dependent trade-offs in foraging by European roe deer (Capreolus capreolus) during winter. *Canadian journal of zoology*, **77** (9), 1486-1493.

Mysterud, A., Larsen, P.K., Ims, R.A. and Østbye, E., 1999b. Habitat selection by roe deer and sheep: does habitat ranking reflect resource availability? *Canadian Journal of Zoology,* **77** (5), 776-783.

Nathan, R., 2008. An emerging movement ecology paradigm. *Proceedings of the National Academy of Sciences*, **105** (49), 19050-19051.

Padié, S., Morellet, N., Hewison, A.J.M., Martin, J.L., Bonnot, N., Cargnelutti, B. and Chamaillé-Jammes, S., 2015. Roe deer at risk: teasing apart habitat selection and landscape constraints in risk exposure at multiple scales. *Oikos*, **124** (11), 1536-1546.

Pagon, N., Grignolio, S., Pipia, A., Bongi, P., Bertolucci, C. and Apollonio, M., 2013. Seasonal variation of activity patterns in roe deer in a temperate forested area. *Chronobiology international*, **30** (6), 772-785.

Pinheiro, J.C. and Bates, D.M., 2000. *Mixed-effects models in S and S-PLUS.* New York: Springer.

Podgórski, T., Baś, G., Jędrzejewska, B., Sönnichsen, L., Śnieżko, S., Jędrzejewski, W. and Okarma, H., 2013. Spatiotemporal behavioral plasticity of wild boar (Sus scrofa) under contrasting conditions of human pressure: primeval forest and metropolitan area. *Journal of Mammalogy*, **94** (1), 109-119.

Polansky, L., Wittemyer, G., Cross, P.C., Tambling, C.J. and Getz, W.M., 2010. From moonlight to movement and synchronized randomness: Fourier and wavelet analyses of animal location time series data. *Ecology,* **91**, 1506-1518.

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Riotte-Lambert, L., Benhamou, S. and Chamaillé-Jammes, S., 2013. Periodicity analysis of movement recursions. *Journal of theoretical biology*, **317**, 238-243.

Riotte-Lambert, L., Benhamou, S. and Chamaillé-Jammes, S., 2016. From randomness to traplining: a framework for the study of routine movement behavior. *Behavioral Ecology*, **28** (1), 280-287.

Ripple, W.J., Chapron, G., López-Bao, J.V., Durant, S.M., Macdonald, D.W., Lindsey, P.A., Bennett, E.L., Beschta, R.L., Bruskotter, J.T., Campos-Arceiz, A. and Corlett, R.T., 2017. Conserving the World’s Megafauna and biodiversity: The fierce urgency of now, *Bioscience*, **67** (3), 197-200.

Rivrud, I.M., Loe, L.E., Mysterud, A., 2010. How does local weather predict red deer home range size at different temporal scales? *Journal of Animal Ecology,* **79** (6), 1280-1295.

Ropert-Coudert, Y. and Wilson, R.P., 2005. Trends and perspectives in animal-attached remote sensing. *Frontiers in Ecology and the Environment*, **3** (8), 437-444.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, **20**, 53-65.

Shoval, N. and Isaacson, M., 2007. Sequence alignment as a method for human activity analysis in space and time. *Annals of the Association of American geographers*, **97** (2), 282-297.

Smulders, M., Nelson, T.A., Jelinski, D.E., Nielsen, S.E., Stenhouse, G.B. and Laberee, K., 2012. Quantifying spatial–temporal patterns in wildlife ranges using STAMP: a grizzly bear example. *Applied Geography*, **35** (1-2), 124-131.

Treaty No. 104 of the Council of Europe of 19 September 1979 on the conservation of European Wildlife and Natural Habitats. Bern, Switzerland.

Wilson, C., 2006. Reliability of sequence-alignment analysis of social processes: Monte Carlo tests of ClustalG software. *Environment and Planning A*, **38** (1), 187-204.

Wilson, C., 2008. Activity patterns in space and time: calculating representative Hagerstrand trajectories. *Transportation*, **35** (4), 485-499.

Wittemyer, G., Polansky, L., Douglas-Hamilton, I. and Getz, W.M., 2008. Disentangling the effects of forage, social rank, and risk on movement autocorrelation of elephants using Fourier and wavelet analyses. *Proceedings of the National Academy of Sciences*, **105** (49), 19108-19113.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M., 2009. *Mixed Effects Models and Extensions in Ecology with R*. Berlin: Springer.

APPENDIX 4A – DATA PREPARATION

For the SAM analysis we extracted from the EuroDEER database (Cagnacci *et al.* 2011, euroungulates.org) 16-day regularised sequences (i.e. *biweekly sequences*) at a fixed four hour relocation interval (0, 4, 8, 12, 16 and 20h) on a fixed yearly schedule starting on January 1st (e.g. 01/01-16/01, 17/01-01/02, etc.). The regularisation consisted in trajectory sub-sampling or gap-filling with NA values. To increase the sample size we also included relocations within 1 hour from the reference time stamps, after verifying that this would not affect exploratory tree clustering (i.e, sensitivity analysis). Below, we discuss in detail the sensitivity analysis (4A.1) and missing data treatment (4A.2).

4A.1. Sensitivity analysis

To increase the number of habitat use sequences we allowed relocations to deviate up to 1h from the reference time stamps (i.e. 0, 4, 8, 12, 16, 20h). More precisely, we increased our sample by extracting three hour temporal resolution sequences (8 fixes per day - i.e. 0, 3, 6, 9, 12, 15, 18, 21h *or* 2, 5, 8, 11, 14, 17, 20, 23h) which were then subsampled to two alternative sequences at four hour temporal resolution (6 fixes per day – i.e, 0, 3, 9, 12, 15, 21h *or* 23, 5, 8, 11, 17, 20h). For example, we subsampled the sequences including locations for time stamps 0, 3, 6, 9, 12, 15, 18, 21h by including the locations at timestamps 0, 3, 9, 12, 15, 21h only. The locations corresponding to the timestamps 9, 15 and 21h were then shifted by one hour, forward or backward, to match the reference time stamps 8, 16 and 20h. This rule increased the sample of 736 sequences for 41 animals from three populations (i.e. IT24, AU17, NK5). To investigate whether this subsampling-regularisation rule would affect the results of SAM we performed a sensitivity analysis using hourly resolution GPS movement data of population CH25[6].

First, we investigated the temporal autocorrelation between habitat use classes of hourly consecutive fixes (Fig.4A.1). As expected, we observed high autocorrelation between hourly consecutive fixes (i.e. a high proportion of consecutive fixes with the same habitat class), but differences occurred between night, day, dusk and dawn. Temporal autocorrelation was highest for night locations, with approximately 85% of the fixes having

---

[6] We performed the sensitivity analysis solely for study area CH25, because of the availability of one hour temporal resolution GPS movement data, and its heterogeneous landscape, where potential bias from sub-sampling-regularisation are expected to be particularly relevant.

the same habitat class than the previous fix. For day locations, autocorrelation slightly decreased to approximately 82% and dropped to a minimum of approximately 80% during dusk (7-8h) and dawn (18-20h). Thus, a relatively high percentage (between 15 and 20%) of the “shifted” fixes may acquire a different habitat use class due to subsampling than the one with the reference time stamp, with the highest effects on dusk and dawn locations.

We further investigated the effects of subsampling on the sequence clustering in SAM. Specifically, using the same hourly dataset (CH25) we extracted 16 days regularised sequences of habitat use using the sampling designs described above (the reference timestamps and the two subsampled sequences at 4 hrs resolution). A basic SAM procedure, was then run for the three sequence datasets combined. Next, we cut the tree by distinguishing from two up to hundred clusters, and calculated the proportion of identical sequences obtained with different sampling designs, which grouped in a different cluster. Results of the sensitivity analysis showed that subsampling had a negligible effect on sequence clustering. For a cut-off from 0 to 20 clusters less than 1% of the triplets of subsampled sequences were not classified in the same cluster. We thus expect a misclassification for 7 out of 736 subsampled sequences, corresponding to 0.12% of the total 5727 sequences that were processed in the overall analyses.

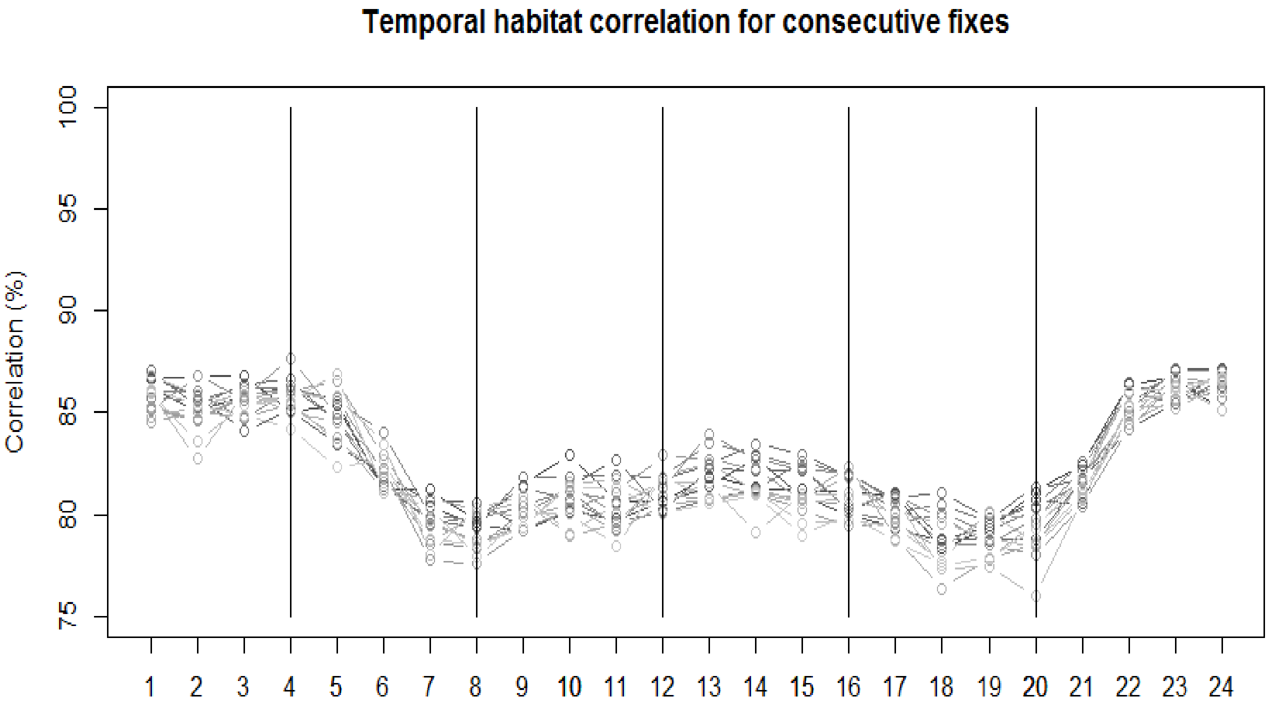


Figure 4A.1. Lineplot giving the temporal autocorrelation (%) for habitat use classes for consecutive hourly timestamps. Temporal autocorrelation is expressed as the percentage of fixes with the same habitat use class at timestamp t and t-1.

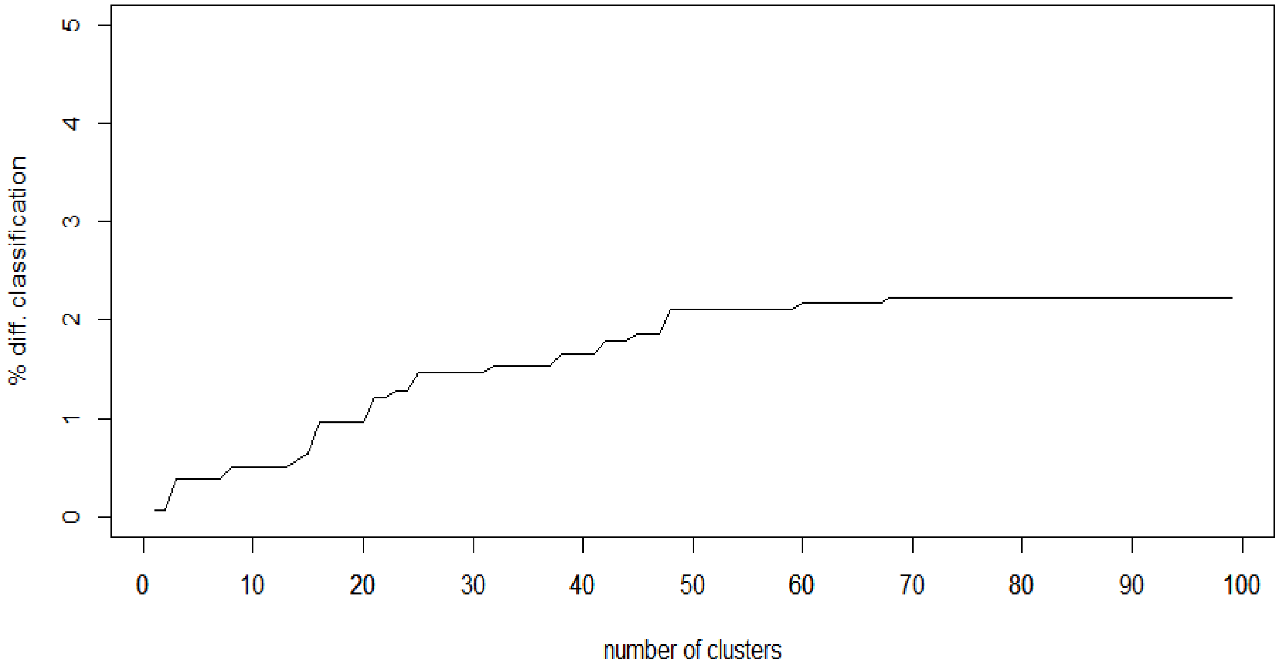


Figure 4A.2. Percentage of sequences that is classified in a different cluster after subsampling from three to four hour resolution, for a cut-off from 2 to 100 clusters.

4A.2. Missing data treatment

For the extraction of biweekly sequences we maintained the NAs distinguishing two types: NA-1, due to the regularisation of trajectories (gap-filling), and thus not corresponding to any GPS acquisition attempt; and NA-2, that are GPS acquisition failures (e.g. due to lack of satellite coverage). While NA-1 are an artifact of the regularisation process, NA-2 are instead potentially ecologically meaningful. Indeed, NA-2 may be correlated with the habitat type where the GPS acquisition was attempted (Frair *et al.* 2010). Therefore, we *a priori* removed any biweekly sequence containing more than 20% of NA-1. Using this procedure we obtained 549792 GPS locations for 432 individual roe deer (250 females and 182 males). These corresponded to 5,727 biweekly sequences of 9 populations characterised by varying forest prevalence and forest structure. To obtain a better insight in the NA-structure (frequency and distribution), we first used the exploratory SAM trees (e.g. Fig.4A.3). NA-1s showed to have a random distribution, while NA-2s were more common during daytime and correlated to forest or closed habitat use. For example, habitat use sequences associated to the exploratory trees in figure 4A.3 show a clear daily alternation in NA-2s with main occurrence during daytime (light-grey bars). Exploration of trees also showed that NA-2s were especially common in three populations (i.e. IT1, 30.86%; DE2, 27.20%; FR8, 25.19%), where data is derived from old sensors with lower performance, and hence a high proportion of NA-2 (Fig.4A.4). Indeed, forest habitat use and NA-2s followed the same daily trend, again only for these three populations (Fig.4A.4). Based on this NA exploration and supporting information from literature (e.g. Cargnelutti *et al.* 2007) we developed a two-step procedure where both NA-1s and NA-2s were subsequently imputed a new value, described below in more detail.

First, *all* NA-1s were imputed as C, O or NA-2 by random-sampling each timestamp of each sequence separately. After this stage, all NA-1s were therefore reassigned to new values. Then, when NA-2 were present, we reclassified them to forest, for a maximum of 10% of the entire sequence, i.e. 9 fixes, and randomly across timestamps. We only performed this reclassification for the three populations mentioned earlier. The reclassification criterion is based on Cargnelutti *et al.* (2007) *w*ho ran a GPS performance analysis in FR8. They found that missing fixes (i.e. NA-2 type) were 10% of the total number of expected fixes, and that in *all* cases NA-2 occurred in the forest and *not* in open habitat. Similar figures were found from an unpublished study (Cagnacci F. pers.comm), that run tests in an Alpine area with a forest density comparable to that of IT1, IT24, DE2, and using the same collar brand and model of such studies. Hence, the expected number of missing fixes was 10% of the total according to the aforementioned assessments, and they were all expected to correspond to true locations in the forest. In this study, we found higher proportion of NA-2s in the same populations (IT1, DE2, FR8) possibly due to other environmental components that may affect GPS acquisition, such as topography, or behaviour, and importantly collar performance (Frair *et al.* 2010, for a review). Because it was not possible to account for all these variables, across the different populations (i.e. complex models of GPS performance were not available), we decided to account for the proportion of missed locations due to forest cover, only, according to the cited assessments. Hence, we applied this conservative proportion for the reclassification of NA-2s, and left as were any NA-2 exceeding such proportion. Finally, if NA-2s still represented more than 30% of all fixes in a sequence, the sequence was removed from the analysis (n=169, or 3% of all original sequences and n=7, or 1.5% of all individuals).

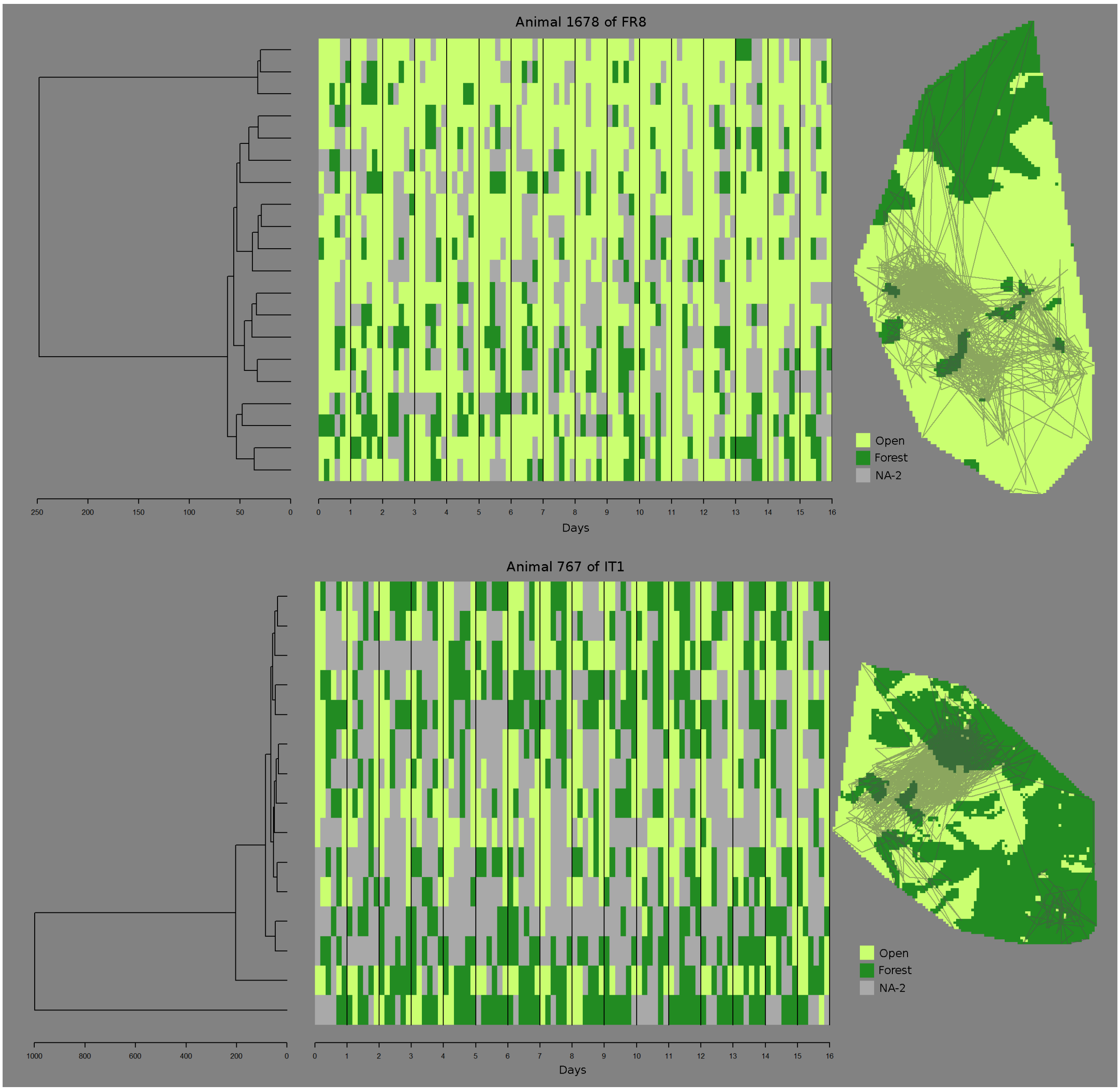


Figure 4A.3. Two examples of exploratory dissimilarity trees (left) with corresponding habitat use sequences (middle, open, lightgreen; closed, darkgreen; NA-2, lightgrey) and spatial movement context (right) highlighting the regular pattern of NA-2s, mainly present during daytime. Vertical lines on top of the habitat use sequences mark a new day. The upper tree is animal 1678 of population FR8, with mainly open habitat within the home range, and the lower tree is animal 767 of population IT1, with mainly forest habitat within the home range.

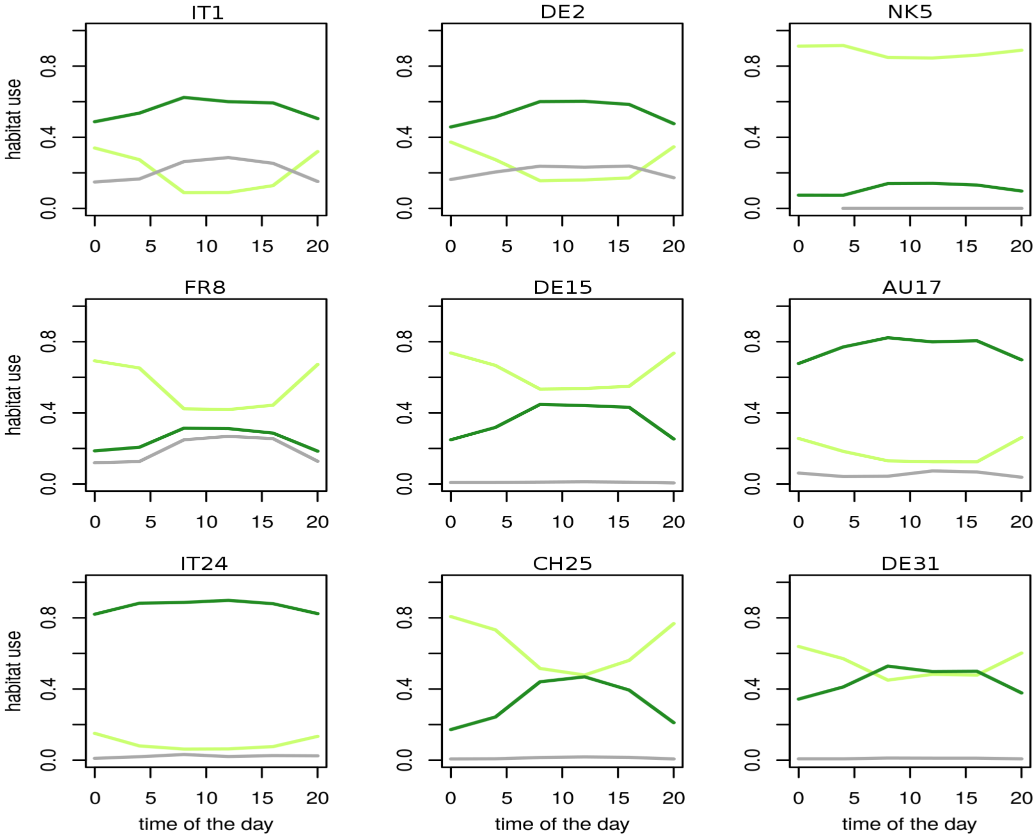


Figure 4A.4. Daily proportional use of open (light green) and closed habitats (dark green) and NA-2s for each population. Three populations (IT1, FR8 and DE2) showed a high proportion of NA-2s, with main prevalence during daytime. Note the similar trend between closed habitat use and NA-2s for the latter populations.

4A.3. Final dataset

As a result of the abovementioned resampling and missing data treatment, we obtained 5,558 sequences and 425 animals. Moreover, all sequences from another twenty animals were removed due to the extremely homogeneous composition of the home range (habitat ratio of 1:99 or more) – five individuals with open habitat only and fifteen with forest only. Finally, one animal with an extremely large home range was removed, because it slowed down computations. As a result of this procedure, the final dataset used for the analyses consisted of 5,402 sequences from 404 animals (236 females and 168 males). The relocations were partitioned as follows: 14,607 NAs (2.82%) and 503,985 actual GPS locations (97.18%), from which 273,230 (52.69%) were classified as open and 230,755 (44.50%) as forest. See table 4A.1 for a summary of the number of animals, sequences and relocations pre-and post-processing.

Table 4A.1. Number of animals, sequences, relocations pre- and post-standardisation of sequences per population. Note that some populations have a lower number of relocations in forest post-processing due to removal of sequences with more than 30% of NAs after the missing data treatment and removal of animals with extreme habitat disproportion (1:99).

Population	**Animals**		**Sequences**		**Relocations & NAs**				**Forest relocations**	
	pre	post	pre	post	pre		post		pre	post
IT1	26	26	561	521	41,157	12,699	46,440	3,576	30,031	35,636
DE2	45	40	409	375	30,868	8,396	33,142	2,858	21,180	23,608
NK5	16	12	264	203	25,063	281	19,488	0	3,573	2,513
FR8	198	182	1,254	1,107	96,161	24,223	102,416	3,856	29,901	37,473
DE15	33	33	866	865	81,749	1,387	82,261	779	29,661	29,818
AU17	8	7	142	141	12,848	784	12,788	748	10,392	10,328
IT24	23	21	399	358	36,759	1,545	33,681	687	33,143	30,018
CH25	71	71	1,559	1,559	144,439	5,225	147,811	1,853	48,104	49,560
DE31	12	12	273	273	25,691	517	25958	250	11,616	11,801
Total (n)	432	**404**	5,727	**5,402**	494,735	55,057	**503,985**	**14,607**	217,601	**230,755**
				%	89.99	10.01	**97.18**	**2.82**	39.58	**44.50**

REFERENCES

Cagnacci, F., Focardi, S., Heurich, M., Stache, A., Hewison, A.J.M., Morellet, N., Kjellander, P., Linnell, J.D., Mysterud, A., Neteler, M. and Delucchi, L., 2011. Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors. Oikos, *120* (12), 1790-1802.

Cargnelutti, B., Coulon, A., Hewison, A.J.M., Goulard, M., Angibault, J.M. and Morellet, N., 2007. Testing Global Positioning System performance for wildlife monitoring using mobile collars and known reference points. *Journal of Wildlife Management*, *71* (4), 1380-1387.

Frair, J.L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N.J. and Pedrotti, L., 2010. Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 365 (1550), 2187-2200.

APPENDIX 4B – RASTER VALIDATION AND THRESHOLD DEFINITION

4B.1. Introduction

To define open and closed habitats at European level we could use two readily available datasets: Corine Land Cover[7] (CLC, EEA 2006, 2012, 100 m spatial resolution) and the High-Resolution Layer-Tree Cover Density 2012[8] (TCD, EEA 2012, 20 m spatial resolution). While CLC is the most commonly used dataset among ecologists for land use and habitat identification, there are several limitations, especially when studying roe deer habitat use patterns. As pointed out by Pekkarinen *et al.* (2009) CLC fails in mapping forest or open patches smaller than the Minimum Mapping Unit (MMU) of 25 Ha, due to a too low spatial resolution (100m) and due to generalisations for defining land cover classes. For example, the CLC-class 'heterogeneous agricultural areas' (CLC-class 21) may include small patches of forest, agricultural fields and urban context, such as farms or dispersed houses. Roe deer's small average home range size (30 Ha) and its preference for ecotonal habitat (Tufto *et al.* 1996), make that CLC potentially misses a significant portion of important open or closed habitat. Therefore, we used instead the High Resolution Layer-Tree Cover Density 2012 (TCD, EEA 2012), which provides the proportion of tree cover (0-100%) per pixel available at a spatial resolution of 20 m or 100 m. This dataset is currently the most detailed validated forest cover dataset that is directly available for use and is provided by the same source as CLC (i.e. eea.europa.eu). Validation of the product was done by an independent consortium, using two validation units: at lot level and country level. The validation procedure was performed using three forest cover density thresholds (10, 20, 30%) by measuring commission (wrong inclusion, or overestimation) and omission (wrong exclusion, or underestimation) errors for the aggregated TCD layer with a resolution of 100 m. In this report, commission error is described by the Producer's Accuracy (PA, i.e. correctly classified points divided by the total number of points) and omission error by the Users Accuracy (UA, i.e. correctly classified points of a class, devided by the total number of points of the class). General results of the validation show that the TCD layer exceeds the target accuracy at pan-European level for commission error and meets the minimum accuracy for omission error. Furthermore, there is country level variability, with

---

[7] http://www.eea.europa.eu/publications/COR0-landcover

[8] https://land.copernicus.eu/pan-european/high-resolution-layers/forests/tree-cover-density/status-maps/2012?tab=metadata

high accuracies for most countries of interest, and with the highest accuracies in general for the 30% forest threshold (see table 4B.1).

Because of country level variability, we performed a local validation analysis which included all except one population (i.e. Southcentral Norway) and helped to define the most appropriate forest cover threshold to identify forest and open habitat. Note that also some other areas are included (i.e. Belgium and Denmark) due to relevance of the study for the EuroDEER collaborative project. To decide which threshold to use we investigated whether TCD showed a gain in accuracy compared to CLC. In total we used two different validation datasets. Norway was excluded due to difficulties in identifying validation points as forest or open habitats.

Table 4B.1. Producers (PA) and Users (UA) accuracy and confidence interval for countries of interest using the 30% forest cover threshold (see full validation report here).

Country	PA (%)	95% CI	UA (%)	95% CI
Switzerland	90.1	2.5	95.5	2.8l
Italy	93.6	1.4	79.1	3.7
Germany	84.7	9.6	92.4	2.4
France	80.2	8.1	90.7	2.0
Norway	72.8	3.1	94.4	3.8

4B.2. Validation data

The validation data consists out of two independently collected datasets, covering a representative but small subset of regions in West-Europe (mainly parts of Belgium, Denmark, France, Germany, Italy and Switzerland). The two datasets have a decreasing spatial coverage. Figure 4B.1 marks the spatial extents of the different regions for each validation dataset, coloured in green and red.

Respectively, datasets concern field observations collected through the Land Use/Cover Area frame Statistical Survey 2012 (LUCAS, Fig.4B.1, green, 5 regions, 4859 points and 0.046 points=km2) and visual interpretation of high resolution satellite imagery (VISAL, Fig.4B.1, red, 9 regions, 1402 points and 0.034 points=km2). LUCAS is a free and open-access field survey monitoring and validation dataset, with project coordination by the Statistical Office of the European Communities (Eurostat, Tóth *et al.* 2013). Since 2006 LUCAS collected every three years land cover/land use, agro-environmental and soil data for georeferenced point locations. LUCAS 2012, which is used in this study, comprises in total around 271.000 surveyed points for 27 countries of the European Union (Eurostat, 2015). The dataset uses eight main land cover classes (A00 to H00) distinguishing up to 84 subclasses, specifying up to species level for the woodland (C00, 21 subclasses) and cropland (B11, 40 subclasses) class. Since we are only interested in forest, the woodland class (C00) was identified as forest and all other classes (A00-B00 & D00 to H00) as non-forest. To ensure reliability only survey points with a distance below four meters between observation location and the LUCAS point were used. In total the validation dataset consisted out of 4859 points (0.046 points=km2) over five regions (Fig.4B.1, green polygons).

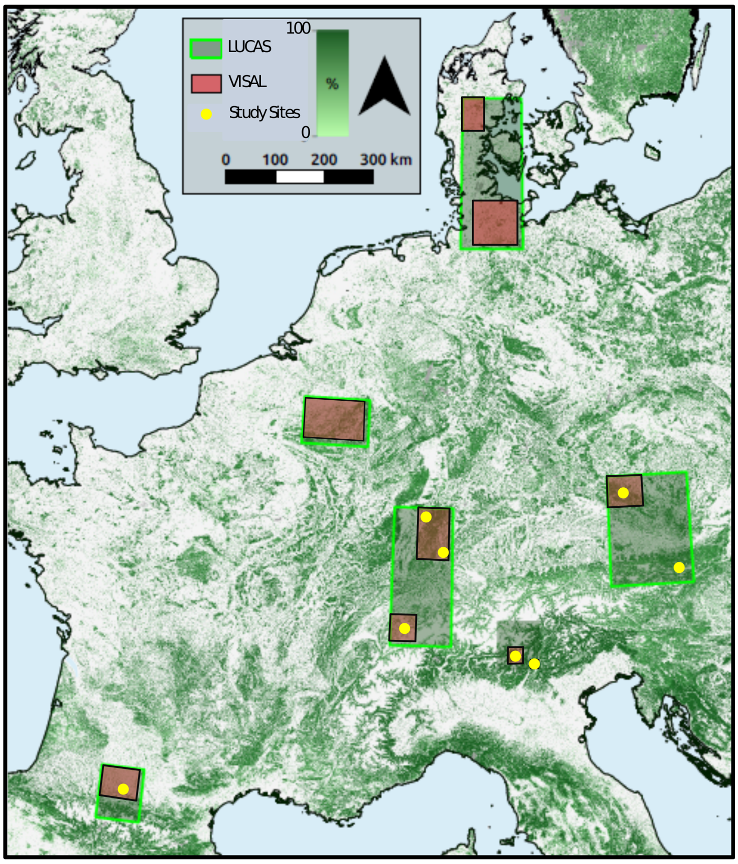


Figure 4B.1. LUCAS and VISAL validation areas and the location of the study sites. Note that Norway is not included in the map.

Over the five LUCAS regions nine different sub-regions were identified, marking the boundaries of the VISAL dataset (Fig.4B.1, red polygons). In each sub-region a random set of 150 points was generated which were classified as forest or non-forest using high resolution GoogleEarth and BingMaps imagery. Although those two sources have their limitations, such as unknown geometric precision, the absence of metadata and variable spatial resolution, they were used since they provide easy access and free high-resolution imagery covering the whole extent of this study (see also Pekkarinen *et al.* 2009). Points were defined as forest if the location was within a forest patch larger than 0.5 Ha. If points were at the exact boarder between forest and non-forest they were excluded for further analysis. In total VISAL consists out of 1402 points (0.034 points=km2).

4B.3. Methods

To assess thematic accuracy of TCD and CLC we performed a cross-validation analysis using the two validation datasets LUCAS and VISAL. First, the forest cover percentage of each validation point was extracted from the corresponding raster cell of TCD and the land use class for CLC. For CLC we considered the three forest classes CLC 311, 312 and 313 as closed habitat and all others as open habitat. Then we used classic methods to measure raster classification accuracy. Specifically, confusion matrices were built and accompanied statistics computed for classification thresholds ranging from 1 to 60 % forest cover. The related statistics were used to assess the accuracy of each classification threshold and consist out of two per class accuracy measures, aka producer's accuracy (Sensitivity, Sf and Specificity, Sn), and the Overall Accuracy (OA) including Confidence Intervals (CI), which are computed based on the assumption of a binomial distribution (but see Kuhn *et al.* 2008; Pekkarinen *et al.* 2009). For each validation dataset statistics were computed over all regions and for each region separately. Respectively, S_f and S_n express the proportion of expected forest and non-forest validation points that were correctly classified by the raster map:

$$S_f = TP/(TP+FN)$$
$$S_n = TN/(TN+FP)$$

Where
TP = True Positive or correctly classified as forest.
FP = False Positive or wrongly classified as forest.
TN = True Negative or correctly classified as non-forest.
FN = False Negative or wrongly classified as non-forest.

OA expresses the proportion of all validation points (forest and non-forest) that were classified correctly by the raster map.

$$OA=(TP+TN)/n$$

After examining the change in accuracy over classification thresholds the described statistical procedure was repeated for CLC. Then we computed the gain/loss in accuracy compared to CLC. Instead of computing statistics for each classification threshold, an increment of 10% was used up to a cut-off of 60% (t1, t2, t3, t4, t5, t6).

4B.4. Results

Measuring accuracy over all validation points of different sites we found a very high Overall Accuracy (OA), Sensitivity (S_f) and Specificity (S_n), with a low standard deviation below a forest cover threshold of 60% (Table 4B.2). In that sense, the classification of the TCD raster dataset was considered more accurate by this local validation study than by the official validation. The histogram of the distribution of forest cover classes shows a low proportion of validation points with a forest cover threshold between 10 and 60%, explaining the low standard deviation over all thresholds (Fig.4B.2). Similar to the official validation we found that higher thresholds result in higher overall accuracy. Specifically, a maximum accuracy was found at a forest cover threshold of 49% for VISAL and at 40% for LUCAS. Moreover, TCD showed on average an improved Overall Accuracy compared to CLC (Fig.4B.3). Average Sensitivity (S_f) and Specificity (S_n), are almost constant up to the 40% threshold, and respectively decrease and increase at the 50% threshold. This means that a larger proportion of open is correctly classified (S_n), but also that a larger proportion of closed is wrongly classified as open (S_f). Note though, that Sn and Sf show larger regional differences than the Overall Accuracy. Due to limited variation in Overall Accuracy below the 60% threshold, a maximum Overall Accuracy between 40% (LUCAS) and 49% (VISAL) and a more balanced Producer’s accuracy (Specificity and Sensitivity) we selected 50% as a threshold to distinguish closed from open habitats. Note that this threshold is specifically defined for our analysis. Due to local differences in accuracy we certainly suggest always to perform a validation analysis before usage. Validation could be performed by visual exploration of satellite images, or better, by building a field (and/or satellite) point validation dataset.

Table 4B.2. Summary statistics for accuracy measures (Overall Accuracy, OA; Sensitivity, S_f; Specificity, S_n). Statistics include the accuracy averaged over all forest cover thresholds below a cut-off of 60% (μ(σ)) and the forest cover threshold (FT%) at which the maximum accuracy (max) is reached.

		μ(σ)	max (FT%)
VISAL	OA	0.919 (0.003)	0.923 (49%)
	S_f	0.872 (0.016)	0.881
	S_n	0.914 (0.013)	0.94
LUCAS	OA	0.892 (0.003)	0.895 (40%)
	S_f	0.773 (0.02)	0.786
	S_n	0.926 (0.006)	0.94

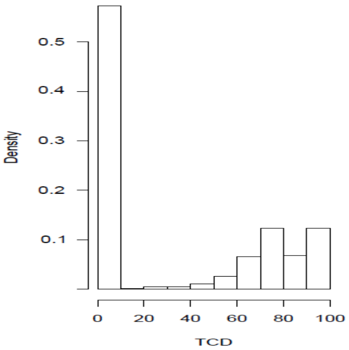
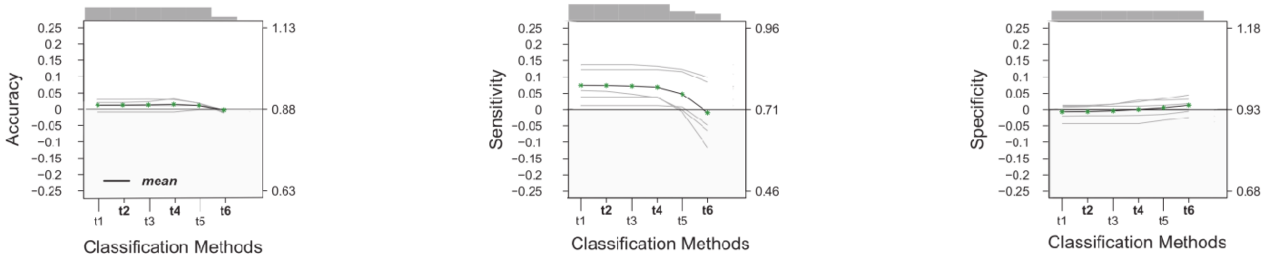


Figure 4B.2. Distribution of forest cover density for validation points.

(a) lucas



(b) visal

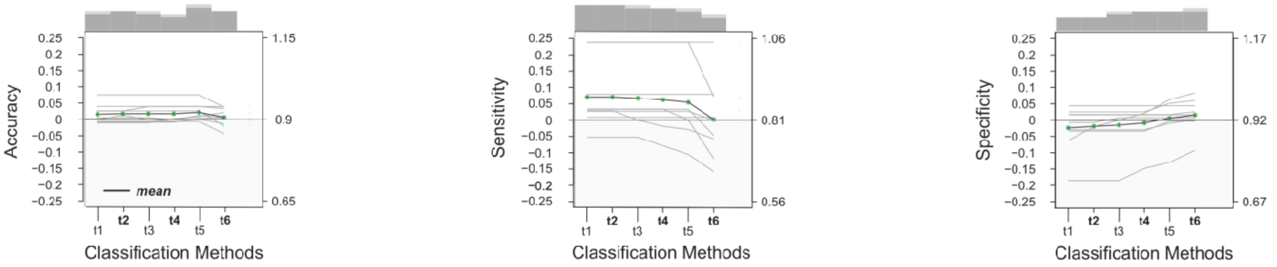


Figure 4B.3. Gain (y > 0)/Loss (y < 0) in overall (OA target > 0:85) and producer's accuracy (Sensitivity, S_f; Specificity, S_n; target > 0:70) for TCD (green points, t1-t6 (10-60%)) compared to CLC over different regions using the visal and lucas validation datasets. The second y-axis (right) is the reference accuracy of CLC. When target accuracy is reached points are indicated by asterisks, else by points. Black lines correspond to the means and grey lines to study specific values (per region). Histograms on top of the plots give the total number of datasets that have a better accuracy than CLC.

REFERENCES

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. *Journal Of Statistical Software*, **28** (5), 1-26.

Pekkarinen, A., Reithmaier, L. and Strobl., P., 2009. Pan-European forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **64** (2), 171-183.

Tóth, G., Jones, A. and Montanarella, L., 2013. LUCAS Topsoil Survey Methodology, data and results. JRC Technical Reports. Luxembourg. Publications Office of the European Union, EUR26102 – Scientific and Technical Research series.

Tufto, J., Andersen, R. and Linnell, J.D.C., 1996. Habitat use and ecological correlates of home range size in a small cervid: the roe deer. *Journal of Animal Ecology*, **65** (6), 715-724.

APPENDIX 4C – THE MOVEMENT MODEL

The movement simulations took place within rasterised home ranges of individual roe deer, as opposed to the artificial arenas used in De Groeve *et al.* (2016). The raster grid resolution was defined to match the minimum mapping width of the habitat layer – reclassified High Resolution Layer-Tree Cover Density 2012 dataset (TCD, EEA 2012) – i.e. 20 meters. We simulated movement as a stochastic and spatially as well as temporally discrete process. The movement takes place on a grid consisting of *N* square cells, with N the total number of 20 x 20 m grid cells encompassing the individual roe deer home range. Following De Groeve *et al.* (2016), the probability of an animal to move from its position on cell i at time *t-1* (i_{t-1} but for convenience abbreviated to $i \in \{1,2,..,N\}$) to any given cell j $\in \{1,2,..,N\}$ at time *t* is given by

$$P_{i,j,t} = \frac{A_{i,j,t}}{\sum_{j=1}^{N} A_{i,j,t}}$$

with $A_{i,j,t}$ the attractiveness of cell *j* for an animal which is currently at cell *i*. $A_{i,j,t}$ is given by

$$A_{i,j,t} = \frac{\beta_{s,H_j,r,t} * m_{j,t}}{d_{i,j}}$$

where $\beta_{s,H_j,t}$ is the selection coefficient for habitat type *H* on cell *j*, H_j , at time *t* based on selection behaviour *s*, $m_{j,t}$ is the number of time steps since the last visit from cell *i* to *j* at time *t*, and $d_{i,j}$ the squared Euclidean distance between cells *i* and *j*. As opposed to De Groeve *et al.* (2016), we did not introduce a maximum travelled distance, *dmax*, to the calculation of the cell attractiveness $A_{i,j,t}$ since the movement of the simulated animal was constrained within a realistic spatial region (i.e. the observed individual roe deer home range).

The selection coefficient $\beta_{s,H_j,r,t}$ was calculated for two habitat types *H* (open, *O;* or closed, *C*), six selection behaviours *s* (random, *u*; homogeneous open, *o*; homogeneous closed, *c;* and three types of daily alternation, *a24*, *a33* and *a42*) and three selection intensities by changing selection coefficient ratios *r*: 1:0.2 (low, r_1), 1:0.1 (intermediate, r_2) and 1:0.01 (high, r_3). The alternating selection behaviour ‘*a33’* is comparable to the behaviour ‘*A*’ in De Groeve *et al.* (2016). We detail below the values of the selection coefficient for the low ratio i.e. $\beta_{s,H_j,r_1,t}$. We generated a random selection behaviour through a spatiotemporally constant selection coefficient (i.e. $\beta_{u,C,r_1,t}=\beta_{u,O,r_1,t}=1.0$). Homogeneous closed and homogeneous open selection behaviours were defined as temporally constant but conditional on habitat types. Specifically, homogeneous closed selection was characterized by a high selection for closed habitat and a low selection for open habitat (i.e. $\beta_{c,C,r_1,t}=1.0$ and $\beta_{c,O,r_1,t}=0.2$), and *vice versa* for homogeneous open selection (i.e. $\beta_{o,C,r_1,t}=0.2$ and $\beta_{o,O,r_1,t}=1.0$). Unlike the previous behaviours, the three alternating selection behaviours were modelled by a selection coefficient dependent on both habitat and time. Specifically, the coefficient values for closed and open habitat types were permuted every three simulation steps to capture circadian habitat selection patterns (6 iterations = one day).

$$\beta_{a24,C,r_1,t} = \begin{cases} 0.2, & 6l+1 \le t \le 6l+2 \\ 1.0, & 6l+3 \le t \le 6(l+1) \end{cases}$$

$$\beta_{a24,O,r_1,t} = \begin{cases} 1.0, & 6l+1 \le t \le 6l+2 \\ 0.2, & 6l+3 \le t \le 6(l+1) \end{cases}$$

$$\beta_{a33,C,r_1,t} = \begin{cases} 0.2, & 6l+1 \le t \le 6l+3 \\ 1.0, & 6l+4 \le t \le 6(l+1) \end{cases}$$

$$\beta_{a33,O,r_1,t} = \begin{cases} 1.0, & 6l+1 \le t \le 6l+3 \\ 0.2, & 6l+4 \le t \le 6(l+1) \end{cases}$$

$$\beta_{a42,C,r_1,t} = \begin{cases} 0.2, & 6l+1 \le t \le 6l+3 \\ 1.0, & 6l+4 \le t \le 6(l+1) \end{cases}$$

$$\beta_{a42,O,r_1,t} = \begin{cases} 1.0, & 6l+1 \le t \le 6l+4 \\ 0.2, & 6l+5 \le t \le 6(l+1) \end{cases}$$

with $l \in \{0,1,2,..,n\}$ the daily cycle iterations (see table 4C.1). Compared to the movement simulations used in De Groeve *et al.* (2016), two patterns of habitat selections (i.e. ‘a24’ and ‘a42’) and two selection intensities (i.e. r_1 = 1:0.2 and r_3 = 1:0.01) have been added.

Table 4C.1. Daily variations in selection coefficients (β) for open and forest habitats used to simulate the six habitat use patterns. Numerical values described below correspond to the low selection intensity treatment (r_1).

Habitat use patterns	β **for open habitat**	β **for closed habitat**
Random ‘u’	1.0, 1.0, 1.0, 1.0, 1.0, 1.0	1.0, 1.0, 1.0, 1.0, 1.0, 1.0
Homogeneous closed ‘c’	0.2, 0.2, 0.2, 0.2, 0.2, 0.2	1.0, 1.0, 1.0, 1.0, 1.0, 1.0
Homogeneous Open ‘o’	1.0, 1.0, 1.0, 1.0, 1.0, 1.0	0.2, 0.2, 0.2, 0.2, 0.2, 0.2
Alternating ‘a24’	1.0, 1.0, 0.2, 0.2, 1.0, 1.0	0.2, 0.2, 1.0, 1.0, 0.2, 0.2
Alternating ‘a33’	1.0, 1.0, 0.2, 0.2, 0.2, 1.0	0.2, 0.2, 1.0, 1.0, 1.0, 0.2
Alternating ‘a42’	1.0, 0.2, 0.2, 0.2, 0.2, 1.0	0.2, 1.0, 1.0, 1.0, 1.0, 0.2

Second, we introduced a memory component, $m_{j,t}$, to the cell attraction value in order to simulate the emergence of a home range from the simulated movement trajectory (van Moorter *et al.* 2009; Fagan *et al.* 2013), which is the usual pattern of space use observed in roe deer. More specifically, we assumed that the attraction to a given cell increases with the time since last visit from cell *i* to *j* at time t, $m_{j,t}$, used here as a proxy for memory (Schlägel & Lewis 2014). The memory value on any cell that has not yet been visited by the simulated animal was initialized to 1. When the animal visits a cell, the memory coefficient drops to 0 and then increments with every time step:

$$m_{j,t} = \begin{cases} 0, & j = i_{t-1} \\ m_{j,t-1} + 1, & j \neq i_{t-1} \text{ and } j \in \{i_1, i_2, i_3, \ldots, i_{t-2}\} \end{cases}$$

Furthermore, we assumed that a cell’s attraction declined inversely proportional to the squared distance between cells *i* and *j,* given by

$$d_{i,j} = (x_i - x_j)^2 + (y_i - y_j)^2$$

With (x_i, y_i) and (x_j, y_j) the Cartesian coordinates of the centers for cells *i* and *j* respectively.

The simulations ran independently for 293 iterations. The first 197 iterations were treated as a burning period, necessary for the simulated animals to initialise a memory of the landscape (van Moorter *et al.* 2009). We used a shorter burning period than De Groeve *et al.* (2016) since the movement was, regardless of memory development, constrained within the observed individual home range. The last 96 iterations (i.e., biweekly trajectory) were used in the IM-SAM analyses.

REFERENCES

De Groeve, J., Van de Weghe, N., Ranc, N., Neutens, T., Ometto, L., Rota-Stabelli, O. and Cagnacci, F., 2016. Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods. *Methods in Ecology and Evolution*, **7** (3), 369–379.

Fagan, W.F., Lewis, M.A., Auger-Méthé, M., Avgar, T., Benhamou, S., Breed, G., LaDage, L., Schlägel, U.E., Tang, W.W., Papastamatiou, Y.P. and Forester, J., 2013. Spatial memory and animal movement. *Ecology letters*, **16** (10), 1316-1329.

van Moorter, B., Visscher, D., Benhamou, S., Börger, L., Boyce, M. and Gaillard, J.-M., 2009. Memory keeps you at home: a mechanistic model for home range emergence. *Oikos*, **118**, 641–652.

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

APPENDIX 4D – BOOTSTRAPPING

To identify the optimal number of clusters in each of the 404 trees, we used a repeatable bootstrapping procedure, based on Jaccard bootstrapping (BJ). We defined our bootstrapping index BJ_{IQMD} as a combination of BJ median (BJ_{MD}, i.e. overall cluster quality of a tree) and BJ interquartile range (BJ_{IQ}, i.e. variance of cluster quality of a tree). This procedure is described in more detail in the main body.

Here, we include three figures giving more detailed information on figure 4.2 in the main body. Specifically, the figures give the trend lines at different cut-offs and for different habitat proportions for BJ_{MD} (Fig.4D.1), BJ_{IQ} (Fig.4D.2) and the combined index BJ_{IQMD} (Fig.4D.3) separately. The figures give support to our decision to identify the cut-off based on a combined index (BJ_{IQMD}). When habitats are more proportional within a home range (MCP) the number of clusters at which BJ_{MD} maximises increases (Fig.4D.1, from red to green dots). On the other hand, BJ_{IQ} has two dips, a primary dip at two or three clusters and a secondary dip at five or six clusters which becomes more pronounced when habitats are more proportional (Fig.4D.2, from red to green dots). Combining both measures (BJ_{MD}, BJ_{IQ}), described by the formula in the main body, increases how well the secondary peak is pronounced (Fig.4D.3).

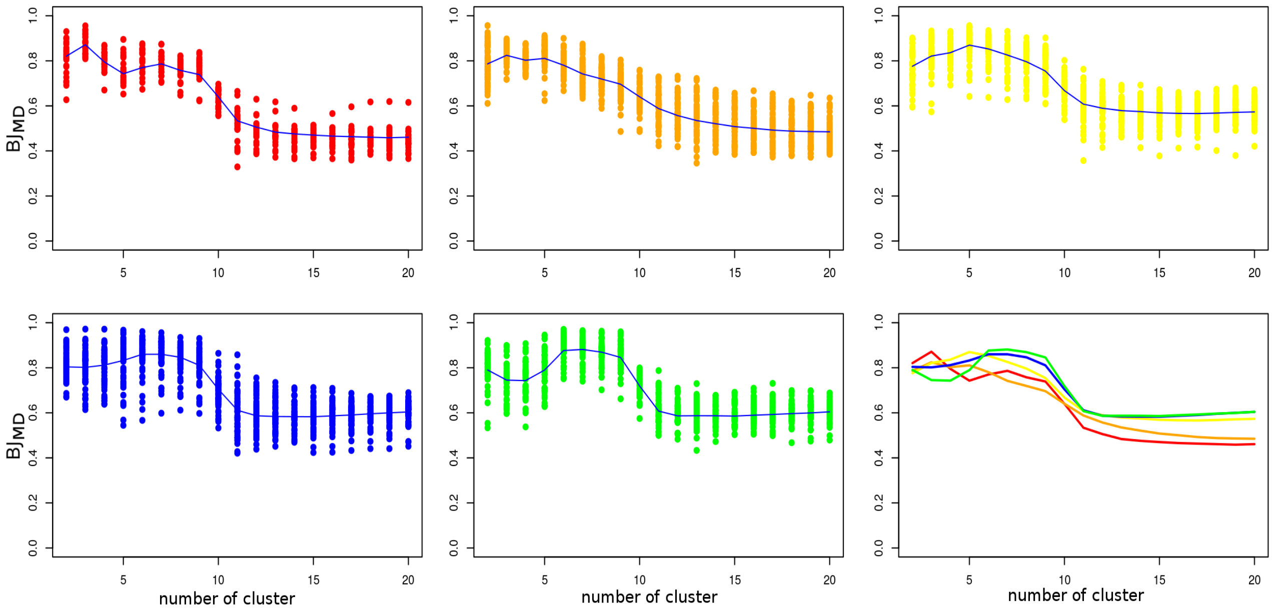


Figure 4D.1. Trend lines and points representing the bootstrap median BJ_{MD} for different cut-offs (2-20 clusters) in all 404 individual-based simulated trees. The higher BJ_{MD} the higher the cluster's quality of a tree. The output BJ_{MD} values are classified into five different relative habitat proportions derived from the individual home ranges, each visualised in a different plot (red, 0.1; orange, 0.2; yellow, 0.3; blue, 0.4; green, 0.5). The lower-right figure only includes the trend lines of each relative habitat proportion class. When habitats in the home ranges are more proportional the BJ_{MD} maximises at a higher number of clusters.

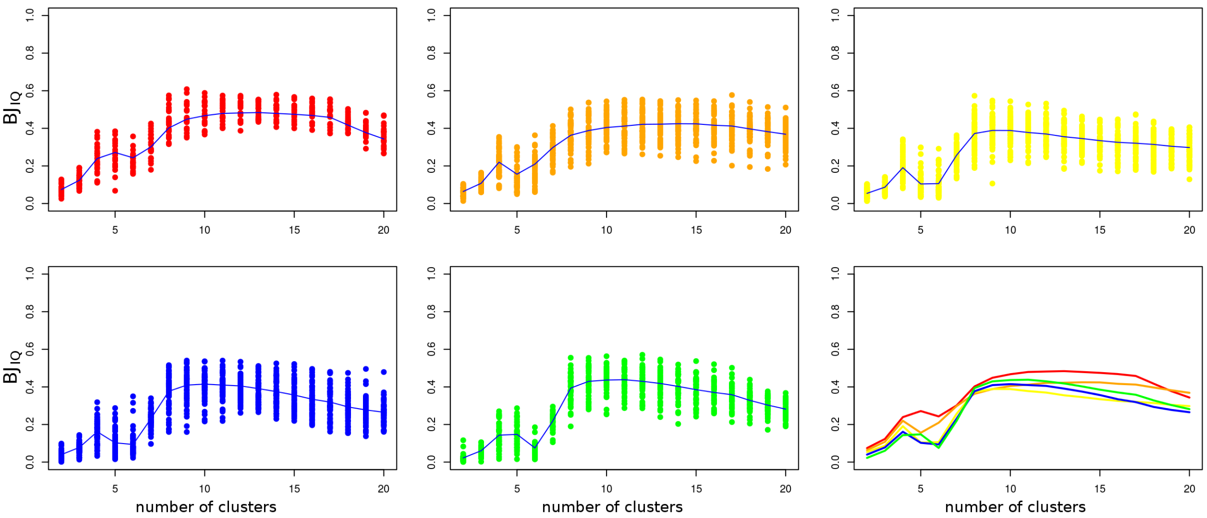


Figure 4D.2. Trend lines and points representing the bootstrap interquartile range BJ_{IQ} (0-1) for different cut-offs (2-20 clusters) in all 404 individual based simulated trees. The lower BJ_{IQ} the lower variability of cluster quality in a tree. The output BJ_{IQ} values are classified into five different relative habitat proportions derived from the individual home ranges, each visualised in a different plot (red, 0.1; orange, 0.2; yellow, 0.3; blue, 0.4; green, 0.5). The lower-right figure only includes the trend lines of each habitat proportion class. Trend lines show two dips, a primary dip, corresponding to a cut- off of two or three clusters; and a secondary dip, corresponding to a cut-off of five to six clusters. When habitats in the home ranges are more proportional the second peak (5-6 clusters) BJ_{IQ} becomes more pronounced.

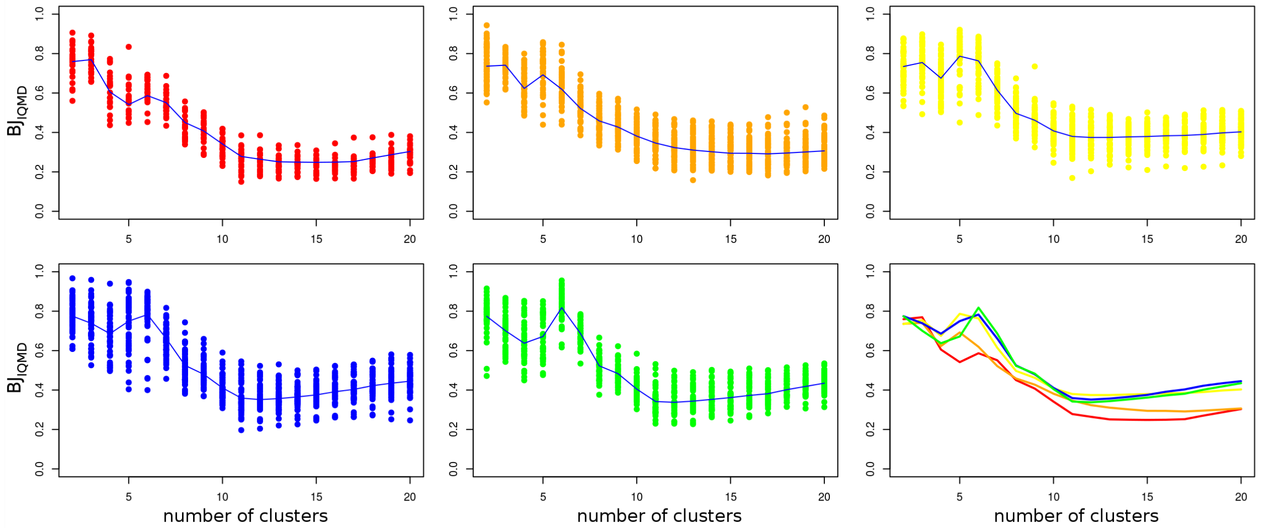


Figure 4D.3. Trend lines and points representing the combined bootstrapping index BJ IQMD (combining bootstrap interquartile range BJ_{IQ} and bootstrap median BJ_{MD}), for different cut-offs (2-20 clusters) in all 404 individual-based simulated trees. The output BJ_{IQMD} values are classified into five different relative habitat proportions derived from the individual home ranges, each visualised in a different plot (red, 0.1; orange, 0.2; yellow, 0.3; blue, 0.4; green, 0.5). The lower-right figure only includes the splines of each habitat proportion class. When habitats are more proportional the secondary peak is more pronounced at a cut-off of six clusters, while the primary peak is more pronounced when habitats are disproportional.

APPENDIX 4F – CLUSTER TYPES

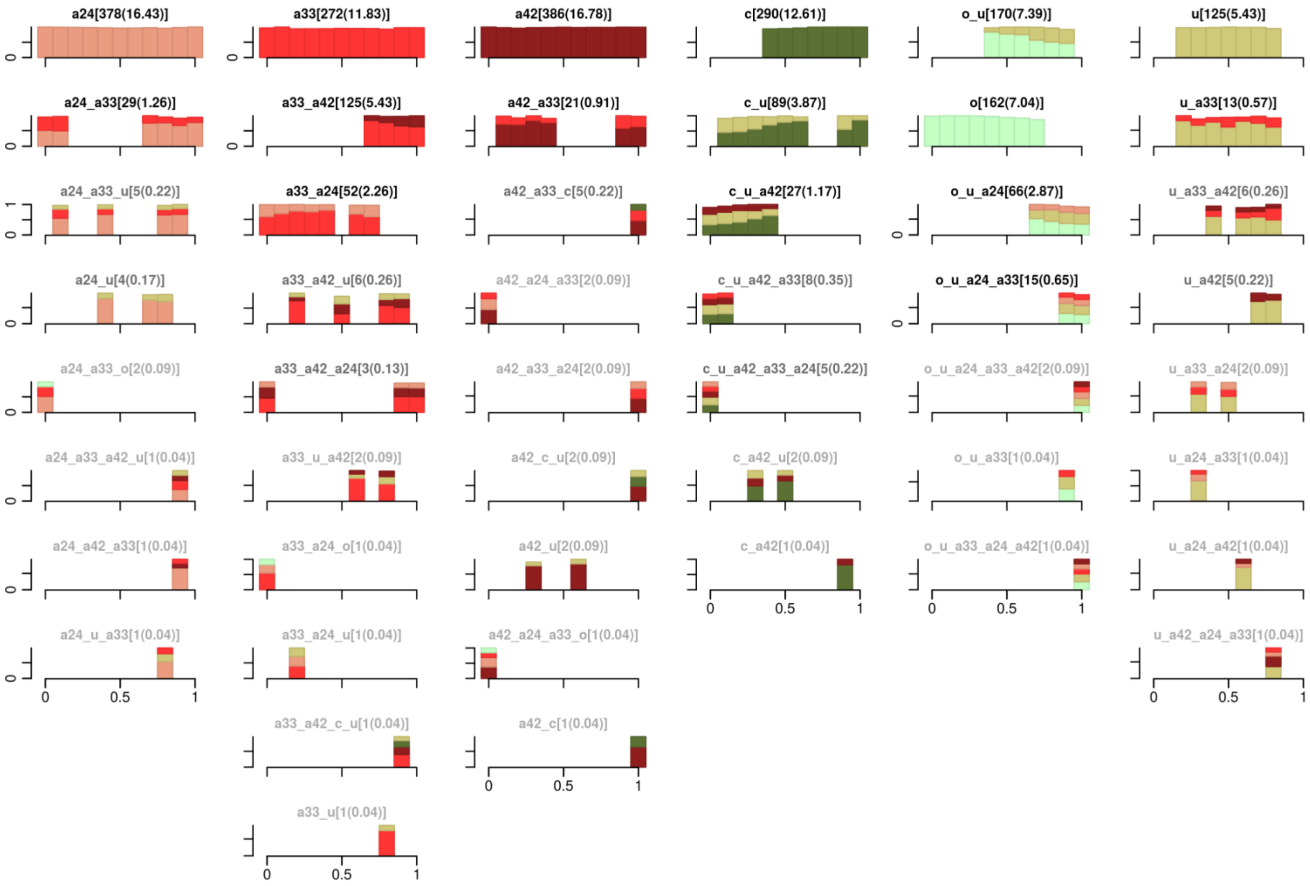


Figure 4F.1. Summary of relative proportions of each sequential habitat use pattern in each cluster type (vertical axis) and the occurrence of each cluster type in dependence on open habitat proportion in home ranges (from 0 to 1). Between square brackets we provide the absolute and proportion of clusters of each identified cluster type. Cluster types titles are in light grey when corresponding to less than 0.5 percent of all clusters.

APPENDIX 4E/4G

Appendices 4E and 4G are available via the following download link:
https://drive.google.com/file/d/1_gYwWIzdfWddZk7J_TrvfcKYDONvSWb3/view?usp=sharing
https://drive.google.com/file/d/11mW34EdMkHt6lOlv7RSyhxxifYA5Ag1v/view?usp=sharing

Through this link we provide respectively a subset of 35 simulation-real sequences trees (Appendix 4E) and pruned classification trees (Appendix 4G), five for each population.

In the submitted manuscript these appendices are included as supplementary material.

CHAPTER 5

Functional role of seasonal habitat determines spatio-temporal patterns of habitat use in a large herbivore

De Groeve, J., Van de Weghe, N., Ranc., N., Bonnot, N., Hewison, A.J.M., Heurich, M., Kroeschel, M., Linnell, J.D.C., Morellet, N., Mysterud, A., Gehr, B., Sandfort, R., Cagnacci, F.

To be submitted in: Ecography

ABSTRACT - Animal habitat use is a particularly important concept in movement ecology, for example, for understanding species' range shifts in response to global change. Fine-scaled spatio-temporal data retrieved from satellite and digital tracking technologies have considerably improved our ability to study animal movement across a range of geographic contexts. However, current approaches rarely account for its temporally-dynamic nature, and in particular the sequential use of habitats with different functional roles. Specifically, sequential use is defined as the sequence of locations (the trajectory) measured against the underlying ordered pattern of habitat use, within an individual specific geographic context (i.e. an individual's home range). Not accounting for sequential use, as in most habitat analysis, only gives insights in the proportion of use, while same proportion could correspond to various biologically important behaviours, such as random, day-night or temporal clustered use. To overcome this limitation, we applied a visually-informative and robust tree-based approach using Sequence Analysis Methods (SAM), derived from molecular biology and more recently used in time geography, to identify ecologically relevant sequential patterns in animal habitat use. This approach was applied to ecological data, for the first time, consisting of simulated and real trajectories of roe deer (*Capreolus capreolus*) across Europe. Our aim was to quantify variation in sequential use of open and closed habitats in nine populations of deer ranging in areas with markedly different landscape structure. The SAM results provide consistent evidence for several behavioural processes, notably day-night habitat alternation, where forest is mainly used during daytime, while there is more of open habitat at night. Interestingly, animals with similar proportions of habitat types in their home range could adopt very different sequential habitat use strategies (e.g. alternating vs random), in particular, local landscape structure in the home range often affected sequential habitat use patterns. A disproportion of habitat availability within the home range resulted in less pronounced daily habitat alternation than when habitat types were proportional. Further, other landscape attributes such as habitat heterogeneity influenced the sequential use of habitats. Lastly, roe deer showed a seasonal shift in the sequential use of different habitats, with more diurnal alternation in habitat use during early spring and early autumn in most populations. Our SAM approach highlights how habitat types with different functions lead to context dependent use and shed new light on how trade-offs shape animal space use.

5.1. INTRODUCTION

Understanding how animals move in relation to their environmental context is one of the keystones of movement ecology, and allows identification of which, where and when resources are used (Manly *et al.* 2002, Nathan 2008). Combined advances in remote monitoring (e.g. GPS) and sensing (e.g. Satellite imagery) technologies has facilitated the juxtaposition between habitat layers and temporal sequences of animal locations, collected at higher temporal resolution over longer monitoring periods and with higher spatial accuracy (Cagnacci *et al.* 2010, Urbano *et al.* 2010, Kays *et al.* 2015). Accordingly, more and more analytical methods have been developed to deal with such large, high-resolution and multidimensional datasets (Demšar *et al.* 2015a).

Habitat use and selection are recognised to occur at multiple scales in space and time (Morris *et al.* 1987, Orians *et al.* 1991). Habitat selection is defined as a hierarchical and multi-scale process by which animals *choos*e resources in a habitat, while habitat use is related to how much animals *us*e these resources to fulfil their requirements (Johnson 1980), within certain tolerance limits (i.e. fundamental niche, Hutchinson 1965). In large herbivores, fitness and survival are affected by the distribution of multiple resources (Law *et al.* 1998), including availability of forage plants, resting sites, as well as refuge habitat for protection from threats, such as predators and extreme weather (e.g. Fryxell and Sinclair 1988; Post and Stenseth 1999; Hebblewhite and Merrill 2009). Depending on the spatio-temporal heterogeneity of such resources, these requirements may result in the use of a multifunctional space (Peters *et al.* 2017). For example, an area of high forest coverage may provide food, but also cover and thermal protection. Hence, human-defined landcover typologies are often associated with multiple environmental and ecological variables important for a species. In this sense, the geographical and ecological definitions of habitat practically correspond (see Bissonette and Storch 2003, p.13). In general, the functional roles of a space (characterised as a *habitat* in the *geographical sense*, from now on) and its use by animals are also dependent on many exogenous factors such as landscape structure and heterogeneity, population density, human density, hunting, inter- and intra-specific interactions (e.g. competition, mutualism) and predator-prey interactions. For example, the level of predation risk of a habitat depends on the nature of risk, i.e. the type of predator (chasing, ambush), and its spatio-temporal distribution in the prey's landscape (e.g. intense, but pulsing; continuous, but homogeneous and predictable), but also by the heterogeneity level and ready availability of refuges. In addition, habitat roles may change over *time* at various temporal scales (seasonal, daily). For example, at the seasonal scale, pastures may have a limited role in winter for some herbivore species due to exposure to cold temperatures and winds, limited food resources and deeper snow, but may be attractive in summer due to the availability of highly nutritious resources (Mysterud *et al.* 1997, Bjørneraas *et al.* 2011, Dupke *et al.* 2017). Agricultural land may be used more during summer when there are crops providing both cover and food, but not during winter when they are more exposed to predation and human disturbance. At the daily scale, animals may trade-off food for cover when using pastures and forest habitat, respectively, which leads to an alternating circadian pattern in the use of habitats (Owen-Smith *et al.* 2010, Dupke *et al.* 2017).

In this paper, we argue that the description of spatio-temporal patterns in sequential habitat use by individuals in contrasting, heterogeneous landscapes may help to provide insights into the complex relationship between habitats, and the resources they provide to meet animals' needs. Sequences of habitat use indicate not only *how much* animals use habitats, but also *the way* they use them. For example, ruminants alternate feeding bouts with resting period. At the daily scale, these activities may be distinguished by the alternation between habitats (i.e. *the way*), more than by the proportion of used habitats (i.e. *how much*). Our general hypothesis is that because open and closed habitats represent multiple and different resources throughout the year in temperate ecosystems, large herbivores will show distinct daily patterns of use (i.e. habitat sequences), and that these patterns will change seasonally. Moreover, we expect that patterns will depend on the *structure* of the landscape (i.e. proportion and heterogeneity), as well as on individual characteristics.

To explore this, we use an enhanced methodological framework developed as described in Chapter 4 which makes use of Individual Movement-Sequence Analysis Methods (IM-SAM), a visually-informative and robust tree-based method to explore and classify ecologically relevant sequential patterns in animal habitat use. De Groeve *et al.* (2016) suggests taking the sequential temporal component of habitat use into account when exploring the relation between habitat, and its function for animals. Indeed, the sequential order with which habitats are used, and not only the proportion of their use, contributes to define the spatial behaviour of an individual. Here, we applied IM-SAM on sequences of use of open and closed habitats by individual European roe deer (*Capreolus capreolus*) for nine populations across Europe for which the availability and distribution of forest habitat varied markedly. .

The European roe deer is one of the most adaptable and widespread ungulate species in Europe. This generalist has adapted to the heterogeneous landscape of Europe by colonising open habitats, such as agricultural fields and grasslands, which are often more exposed to human-related disturbance (Andersen *et al.* 1998). Roe deer show a preference for ecotonal habitats (i.e. edges) where highly digestible forage resources are available (Tufto *et al.* 1996) and adjust feeding strategy and diet with changes in landscape heterogeneity (Abbas *et al.* 2011). Behavioural plasticity is observed with very distinct space use and feeding strategies between and within populations (Abbas *et al.* 2011). In addition, seasonal variability in space use is also frequently observed, corresponding to a gradient of migration strategies within and between populations (Cagnacci *et al.* 2011). At a daily temporal scale roe deer are typically characterised by a bimodal crepuscular activity pattern, and a general increased activity during the night (Pagon *et al.* 2013, Krop-Benesch *et al.* 2013). High spatio-temporal variability in behaviour between and within populations makes roe deer a very interesting model species to study sequential habitat use. While there are many detailed local studies on roe deer habitat use and selection (e.g. De Groeve *et al.* 2016, Mancinelli *et al.* 2015, Dupke *et al.* 2017), multi-population studies are rare, and there has been no attempt to explore the actual sequence of habitat use in such context.

We investigated individual trajectories of roe deer from nine European populations as sequences in the use of open and closed habitat, with a biweekly resolution. Then, we categorised these sequences as spatio-temporal patterns of use: homogeneous use of either closed or open; alternation between closed and open (with different alternation schedules); and random use. We compared the spatio-temporal patterns of use across populations, and individuals, throughout the year to assess the hypothesis that roe deer respond to landscape composition and fragmentation, and to resource seasonality, by adjusting habitat use both as relative proportion and sequential use.

Because roe deer can adapt to different environments (generalist behaviour), we expected that the sequential patterns of habitat use would differ between populations, reflecting gradients of habitat composition and structure. We expected the sequential habitat use to deviate from random, thereby reflecting roe deer's ecological plasticity, i.e. adaptation to spatio-temporal distribution of resources (i.e. Peters *et al.* 2017), and environmental constraints such as weather, human disturbance and predation risk. Since movement and habitat selection in roe deer can be the result of individual responses (i.e. personality, life history traits) and other environmental constraints (e.g. social fences; red deer in Mysterud *et al.* 2011) we expect that similar habitat composition and structure of home ranges may result into different patterns of sequential habitat use at the individual level. Finally, because open and closed habitats are composites of resources that vary in time, we expect the sequential patterns of their use to vary across seasons.

5.2. MATERIAL AND METHODS

5.2.1. Trajectory data

GPS-based trajectories for roe deer were extracted from the EuroDEER database (Cagnacci *et al.* 2011, euroungulates.org) following the procedure as described in Chapter 4 (par. 4.2.1, Appendix 4A & 4B). To summarise, we extracted 16-day regularised sequences (i.e. *biweekly sequences*) with a fixed four hour relocation interval (0, 4, 8, 12, 16 and 20h) over a fixed yearly schedule starting on January 1st (e.g. 01/01-16/01, 17/01-01/02, etc.) and linked the GPS locations to the reclassified High Resolution Layer-Tree Cover Density 2012 (TCD, EEA 2012), distinguishing closed (C, >50%) and open (O, <50%) habitats. In table 5.1 we summarise the number of individuals and sequences per population. For transparency population codes refer to the official country codes (i.e. NK, FR, DE, IT, AU and CH) and the EuroDEER study site identifier (i.e. 1, 2, 5, 8, 15, 17, 24, 25, 31).

Table 5.1. Populations, number of animals and sequences.

Country	Country code	EuroDEER ID	Animals	Sequences
Norway	NK	5	12	203
France	FR	8	182	1107
Switzerland	CH	25	71	1559
Germany	DE	15	33	865
		31	12	273
		2	40	375
Italy	IT	1	26	521
		24	21	358
Austria	AU	17	7	141
		Total	**404**	**5402**

5.2.2. Sequential habitat use analysis

To describe the variability in sequential use of open and closed habitats for individual animals, we applied (IM-)SAM as proposed in Chapter 4 (par. 4.2.2., Appendix 4D), i.e. we classified the biweekly sequences of used open/closed habitats using a tree-based method based on sequence dissimilarity. In essence, within each individual’s home range described by a Minimum Convex Polygon (MCP), simulated trajectories, were generated following six hypothesis-based sequential habitat use types. Specifically, homogeneous closed, 'c'; homogeneous open, 'o', random, 'u', alternation mainly open, 'a24', alternation equal open-closed, 'a33' and alternation mainly closed, 'a42'. Subsequently, we identified the most similar simulated sequential habitat use type for observed trajectories by quantifying sequence dissimilarity between observed and simulated trajectories. Observed trajectories were hence classified into six characteristic types of sequential habitat use as generated by simulated trajectories. Because clusters were often described by multiple sequential patterns, dependent from habitat availability, we only considered the main sequential pattern in a cluster (Chapter 4 par. 4.3.3 & Table 4.1). The sequential habitat use type, associated to real sequences are further referred to by the six codes (c, o, u, a24, a33, a42).

5.2.3. Statistical Analysis

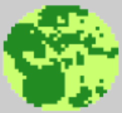
a. Environmental and individual-based covariates

After identification of sequential habitat use patterns, we investigated their spatial and temporal variability in relation to landscape metrics (i.e. structure and composition of the landscape), and individual-based ecological variables. For this purpose, we calculated landscape metrics with the R package SDMTools, which includes 37 measures of the reference program for spatial pattern analysis FRAGSTATS (VanDerWal *et al.* 2014). Because we observed that many landscape metrics were highly correlated, we performed a covariate selection procedure. In particular, we first computed the area of reference for each *population* by merging all *biweekly* MCPs (i.e. MCP computed per sequence, not per individual) and generated a random sample of 1000 equal-sized circles of 310m diameter within this area, which corresponds approximately to the median home range size of 30 Ha. Next, we intersected each circle with the reclassified TCD raster layer and derived all 37 landscape metrics, using both habitat classes as references (O and C). We thus obtained a total of 9000 values (1000 buffers x 9 populations) for each metric. We plotted these values against the proportion of open habitat in each buffer, and we fitted a spline to the point distribution. We grouped the metrics into similar distributional types, and chose one from each group, thus limiting correlation between covariates. Specifically, we selected Proportion (P), Splitting Index (SPLIT), Patch Cohesion Index (COHESION), Landscape Shape Index (LSI), Patch Density (PD) and Edge Density (ED), which we calculated for both habitat types (Fig.5.1 right panel; Table 5.2 for definitions with an illustration for one study site). Abbreviations are the official terms as defined in the FRAGSTATS documentation[9]. Next, these metrics were obtained for each individual as the average of values computed for 200 random equal-sized circles (310 m diameter) within each individual MCP, which we used as our main covariates (Fig.5.1, left panel for each population). With this approach, measures covered the complete surface of larger MCPs. We also included sex (male, female) and biweek of each sequence (1 to 23, see trajectory analysis) as additional individual and population level covariates.

---

[9] https://www.umass.edu/landeco/research/fragstats/documents/fragstats.help.4.2.pdf

Table 5.2. Description of the six fragstat landscape metrics used as covariates for the sequential habitat use pattern variability. To the right, an example of the mean and maximum values of each covariate calculated for open habitats in the study site CH25.

Fragstat Measures	Description	Mean	Max
Proportion $0 \leq \mathbf{P} \leq 1$	P gives the proportion of a class.		
Patch Cohesion Index $0 < \mathbf{COHESION} < 100$	COHESION is a measure of connectivity between patches of a class. Patch cohesion increases if the patch types become more clumpy.		
Landscape Shape Index **LSI** ≥ 1, without limit	LSI is a measure of class aggregation or clumpiness and equals the total edge length of a class, given in number of raster cells, divided by the minimum edge length of the class possible for a maximally aggregated class.		
Patch Density **PD** > 0, constrained by cell size	PD gives the number of patches of a class divided by total landscape area (m^2), multiplied by 10 000 and 100 (100 Ha).		
Edge Density **ED** ≥ 0, without limit	ED sums the edge segments (m) of a class, divided by the total landscape area (m^2), multiplied by 10 000 (Ha).		
Splitting Index $1 \leq \mathbf{SPLIT} \leq$ number of cells in the landscape squared	SPLIT gives the cumulative patch area distribution and is interpreted as the number of patches with a constant patch size when the landscape is divided in S patches, where S is the splitting index.		

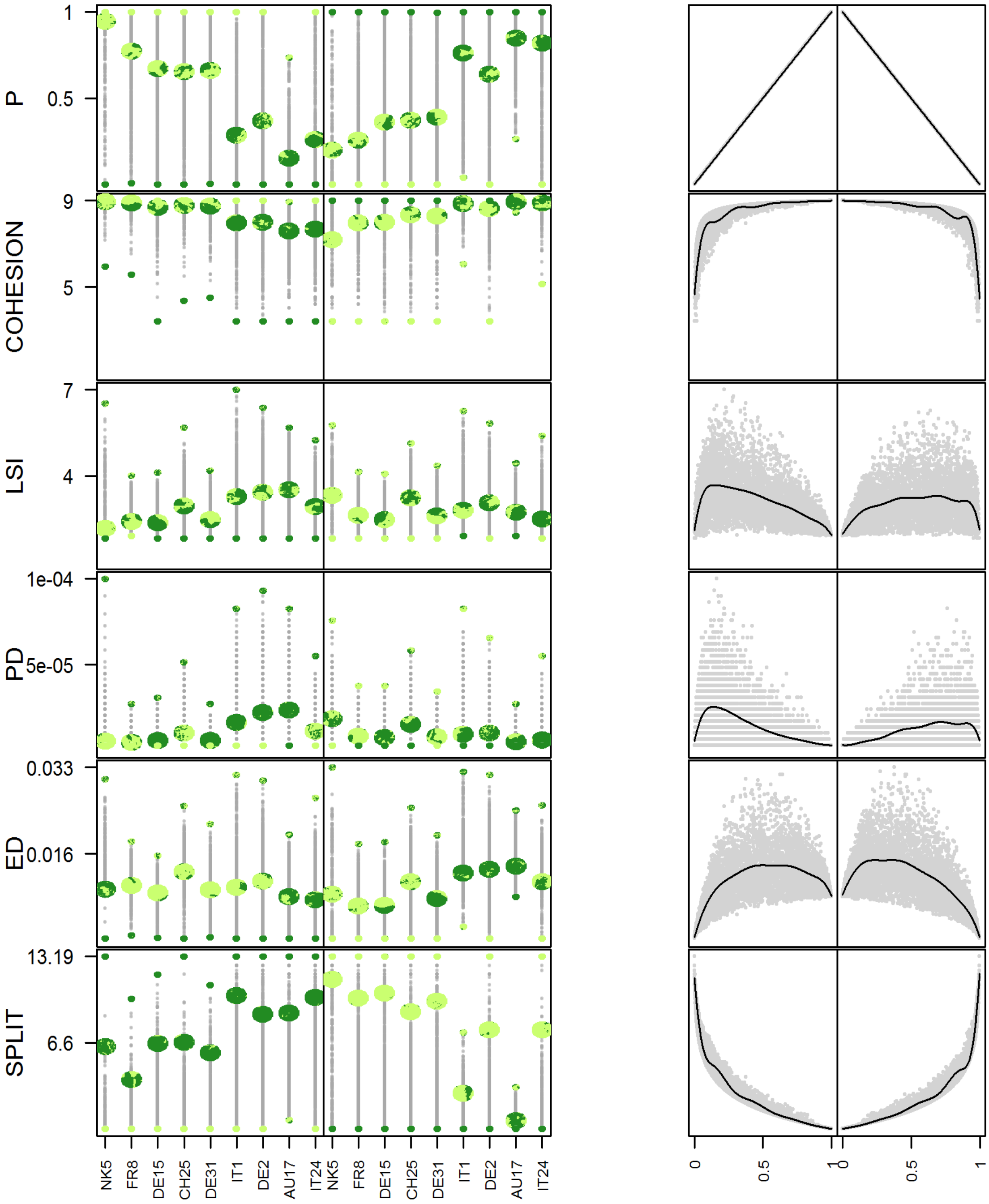


Figure 5.1. Point distribution plots (right): distribution of landscape metric values in relation to the proportion of open habitats (x-axis, 0-1) plotted for all populations, for open (left) and forest habitats (right). **Buffer plots (left)**: mean, minimum and maximum values of landscape metrics, for each population. Population codes as in the main text.

b. Multivariate statistics of habitat use sequential patterns in relation to environmental variables

To answer our main research questions, we used a stepwise-analysis procedure. First, we performed a Detrended Correspondence Analysis (DCA; Hill and Gauch 1980) to explore the relationship between sequential habitat use patterns and population identity. The input for this analysis consisted of an abundance table of sequential habitat use patterns (the 'species' in multivariate statistics terminology) per population (the 'sites'). DCA is a multivariate statistics technique where new 'variables' (the ordination axes) are generated that maximise the separation among groups of 'species' sampled at different 'sites'. In other words, the axes should separate the sequential habitat use patterns according to their relative abundance across populations. DCA rescales and detrends the variables along the axis (ter Braak and Verdonschot 1995).

Then, to identify *which* landscape variables affected sequential habitat use, we performed a Principal Component Analysis followed by a Stepwise Redundancy Analysis (PCA/RDA; ter Braak and Verdonschot 1995) at the individual-level. Here, the input is an abundance table of sequential habitat use patterns (the 'species') per individual (the 'sites'), excluding any individual for which we had less than 10 biweekly sequences, together with a table of the described landscape metrics calculated for each individual MCP (the 'environment'). The goal of RDA is to determine the pattern of variation among 'species' assemblages (i.e. abundance of multiple 'species' at different 'sites') dependent on a set of environmental variables. In RDA, the ordination axes (or Principal Components of a PCA) for displaying the abundance of species in different sites are constrained to be linear combinations of the environmental variables measured at the same sites (Rao 1964, ter Braak and Verdonschot 1995).

In both DCA and RDA, the results are represented in 'biplots'. The prefix 'bi' refers to joint representation of two datasets (Jongman *et al.* 1995), i.e. the ordination of sequential habitat use patterns (and environmental axes in the case of RDA) along two generated gradients, the ordination axes. In RDA, the association between sequential habitat use patterns and environmental variables and the two ordination axes is therefore represented as vectors over the joint space between the two axes. Vectors are a standardised representation of RDA canonical coefficients (regression coefficients) of environmental variables on both axes (ter Braak 1986, McGarigal *et al.* 2000, Legendre and Gallagher 2001). For all multivariate statistics, we used the R package vegan.

Finally, to quantify temporal variability in sequential habitat use patterns across seasons, we summarised and visualised the abundance of sequential habitat use patterns per biweekly period, for each population separately.

5.3. RESULTS

5.3.1. Sequential habitat use patterns vary among populations.

The sequential habitat use patterns (i.e. 'species' in multivariate analysis terminology) separated mainly along the first DCA axis (Eigenvalue=0.34, Axis length rescaled=2.12, Fig.5.2), which was mainly associated with homogeneous sequential patterns. Furthermore, some populations (i.e. ‘sites’) were clearly associated with these sequential patterns. Specifically, the minimum and maximum site scores for the first axis were respectively -0.95 for Southcentral Norway (NK5, homogeneous open) and 1.17 for the Italian Alps (IT24, homogeneous closed). Italian and Austrian Alps (IT1, AU17) and Bavarian Forest (DE2) were also associated with the latter. The second DCA axis was much less discriminatory for separating sequential habitat use patterns (Eigenvalue=0.04, Axis length rescaled=0.59), and mainly distinguished the alternating habitat use sequences alternation24/alternation33 (a24 score=1.55; a33 score=1.01) from alternation42 (a42 score=-1.45). Interestingly, the sequential habitat use types alternation24 and alternation33 were more closely related to homogeneous open than to homogeneous closed on the first axis (a24 score=-0.33 and a33 score=-0.09), whereas the opposite was true of alternation42 (a42 score=0.40). The minimum and maximum sample score for the second axis were -0.35 for Southcentral Norway (NK5, negatively associated with alternation) and 0.24 for Southwest France (FR8, positively associated with alternation a33 and a24), respectively. A gradient length of 0.59 SD units is very low and shows that the opposite ends of the second axis are rather similar. Notably, although, some sites had similar proportion of closed habitat (Fig.5.2, top-right infra panel), they did not overlap in the DCA diagram, for example Southern Germany (DE15) and Southwest France (FR8) occurred at opposite ends of the second axis, with Southwest France (FR8) associating with the habitat use sequences alternation24/alternation33 (site score=0.24), but not Southern Germany (DE15, site score=-0.25).

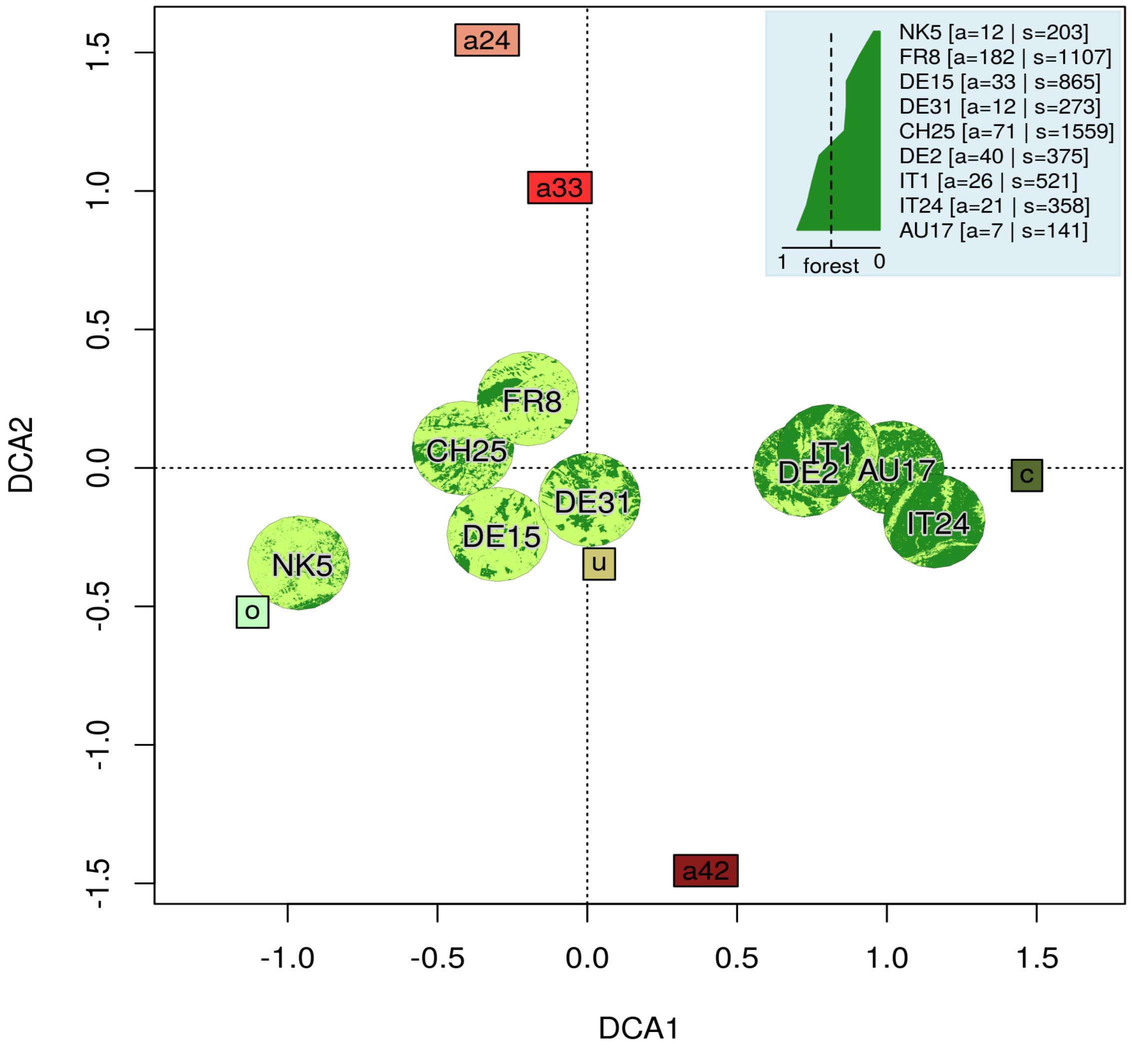


Figure 5.2. Detrended Correspondence Analysis (DCA) using abundances of sequential habitat use patterns ('species' in multivariate statistics terminology; coloured tags: c, o, u, a24, a33, a42) across populations ('sites', here represented as circular extracts of population ranges) as input data. The graph in the upper-right corner indicates the availability of closed habitat per population including the number of animals (a) and sequences (s).

5.3.2. Sequential habitat use patterns vary among individuals

The occurrence of sequential habitat use patterns was not constrained along canonical axes using individual abundances only (i.e. number of sequences for each sequential habitat use pattern per individual – individuals were scattered across the four quadrants of the biplot, Fig.5.3), so that the Principal Component (PC) axes were unconstrained, with eigenvalues of the same order of magnitude (PC1=1.69, PC2=1.54, PC3=0.88, PC4=0.85). Instead, sequential habitat use patterns were highly correlated with the individual level habitat structure metrics within the home range (Fig.5.3, Appendix 5A).

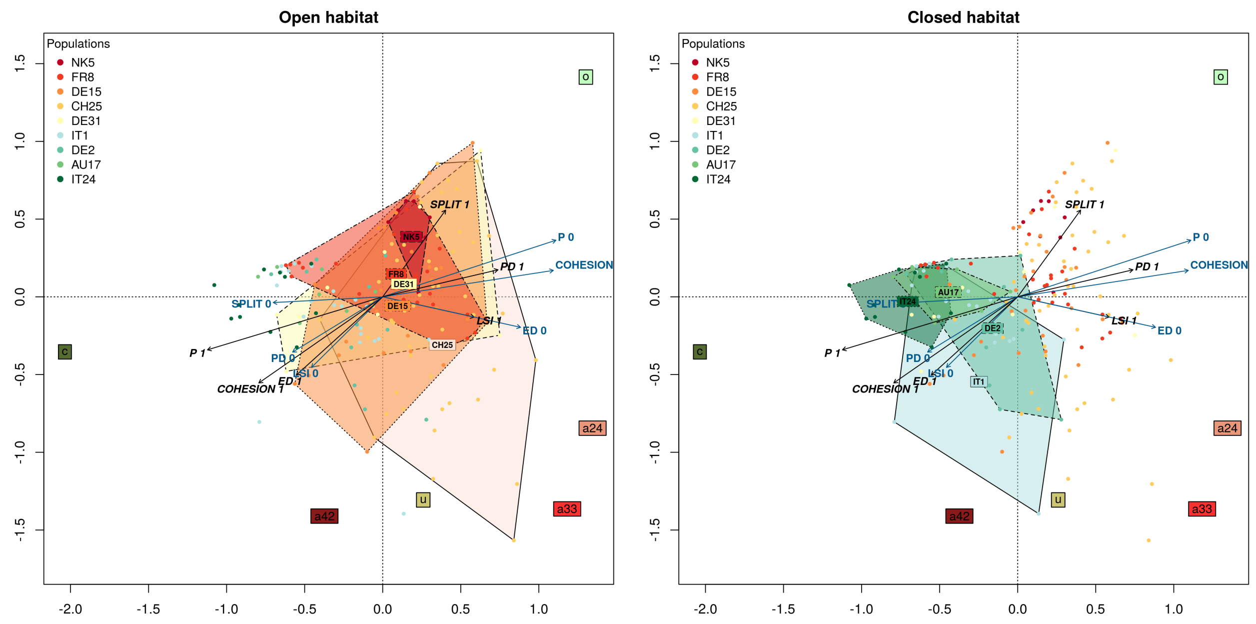


Figure 5.3. Principal Component Analysis/Redundancy Analysis biplots, highlighting individual variability in sequential habitat use patterns, in relation to the landscape metrics. In multivariate terminology, sequential habitat use patterns represent the ‘species’ (the rectangular labels in the biplot), individuals represent the ‘sites’ (the dots in the biplots), and landscape metrics the ‘environmental variables’ (the vectors in the biplots). Points are distinguished by population with different colours, and wrapped in minimum convex polygon of the same colour, to show the variability of sequential habitat use between individuals (the larger the polygon, the more variable the sequential habitat use among individuals). For readability, we plotted the *same* PCA twice, visualising all individuals in both plots, but polygons from different populations, i.e. with prevalence of open habitat on the left and of closed habitat on the right. Landscape metrics are: proportion, P; Splitting Index, SPLIT; Patch Cohesion Index, COHESION; Landscape Shape Index, LSI; Patch Density, PD; Edge Density, ED, computed both for the open habitat class (dark blue vectors) and the closed habitat class (black vectors).

Specifically, homogenous open ‘o’ was obviously positively associated with the proportion of open habitat (P0, PC1=0.95), and also with the landscape shape index of closed (LSI1, PC1=0.98), open patch cohesion (COHESION0, PC1=0.99), closed patch density (PD1, PC1=0.97) and edge density of open patches (ED0, PC1=0.98), on the first axis, and the splitting index of closed (SPLIT1, PC2=0.56), on the second axis. Homogeneous closed ‘c’ was obviously associated with the proportion of closed (P1, PC1=-0.96), and also with the landscape shape index of open (LSI0, PC1=-0.71), closed patch cohesion (COHESION1, PC1=-0.82), open patch density (PD0, PC1=-0.85), edge density of closed patches (ED1, PC1=-0.74), and the splitting index of open (SPLIT0, PC1=-1), on the first axis. Alternation24 'a24' was associated to landscape shape of closed (LSI1, PC1=0.98) and edge density of open patches (ED0, PC1=0.98) on the first axis, whereas alternation42 'a42' was associated to the landscape shape index of open (LSI0, PC1=-0.71) and edge density on closed patches (ED1, PC1=0.98) on the first axis and the landscape shape index of open (LSI0, PC2=-0.71), closed patch cohesion (COHESION1, PC2=-0.57) and edge density of closed patches (ED1, PC2=-0.67) on the second axis.

The spread of individual animal data (coloured points) within a population (corresponding coloured polygons) on the biplot (Fig.5.3) highlighted substantial individual variability in sequential habitat use patterns. In general, with few exceptions, many individuals aggregated close to the center, indicating that they did not adopt a single type of sequential habitat use. Instead, several, but not all, aligned along the environmental covariate vectors, indicating a stronger correlation between the metrics describing habitat structure within the individual’s home range and the type of sequential habitat use. Specifically, populations with a high habitat preponderance of one habitat type, such as Southcentral Norway (NK5) and Italian Alps (IT24) (average proportion of open habitat in MCPs greater than 0.9 and smaller than 0.1, respectively, see top-right panel in Fig.5.2) had small, asymmetrical polygons, indicating little variability among individuals, with a prevalence of a homogeneous sequential habitat use pattern. Interestingly, with habitat proportion of open or closed greater than 0.1, individual variability increased. For example, the polygon for Southwest France (FR8, average proportion of open habitat in MCPs between 0.8 and 0.9) was much larger, spanning the four quadrants of the biplot, and indicating that individuals with all types of sequential habitat use patterns occurred on this study site. For populations where the proportion of open and closed habitat were more or less equal (average proportion of open/closed habitat in MCPs between 0.3 and 0.5), such as Southern Germany (DE15, DE31) and Italian Alps (IT1), the size of the polygons in the biplot was especially large.

5.3.3. Sequential habitat use patterns vary through time.

The relative proportions of the types of sequential habitat use patterns across time (biweeks) showed clear seasonal variability within populations (Fig.5.4). Some temporal trends in the abundance of sequential habitat use patterns were common across multiple populations. First, a bimodal seasonal pattern in habitat alternation (a24, a33, a42) was observed (i.e. Southwest France (FR8) and Austrian (AU17), Italian (IT1) and Swiss Alps (CH25); Fig.5.4), with a peak of habitat alternation in early spring (around 5th and 6th biweek) and a second peak in the fall (between 16th and 20th biweek, depending on the population). This trend was therefore observed both for populations in landscapes where open habitat was prevalent (Swiss Alps (CH25) and Southwest France (FR8)), in which case the variability concerned homogeneous open (o) and habitat alternation (a24, a33, a42), whereas the use of homogeneous closed was consistent throughout the year; and for populations in landscapes where closed habitat was prevalent (Italian (IT1) and Austrian Alps (AU17)), in which case the variability concerned the use of homogeneous closed (c) and mainly habitat alternation (a24, a33, a42). Interestingly, in certain cases more than half of the sequences of a biweek were alternating ones. Second, homogeneous closed sequences were prevalent during winter in several populations (Italian and Austrian Alps (IT24, IT1, AU17), Southern Germany (DE15, DE31) and Bavarian Forest (DE2)). Third, even in very closed landscapes, some alternation occurred in spring-summer (Italian and Austrian Alps (IT24, AU17), Bavarian Forest (DE2; see Fig.5.2, top-right panel for the proportion of closed). Finally, further peculiarities in sequential habitat use trends were noticed for the populations in Southern Germany (DE15 and DE31). In the former (DE15) we observed alternating and homogeneous closed sequences during winter and mainly homogeneous open sequences from late spring to early autumn; in the latter (DE31) we observed a more irregular pattern.

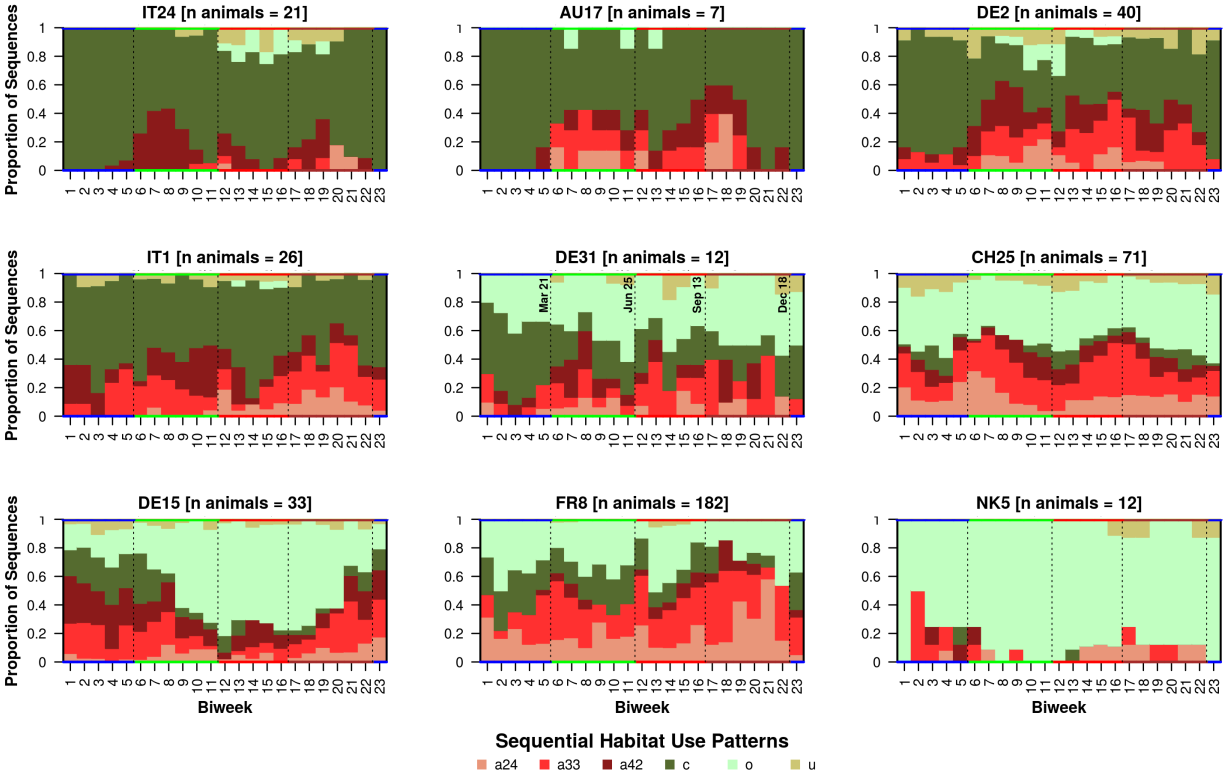


Figure 5.4. Temporal variability in sequential habitat use patterns for each population, reported as biweekly proportions (1-23). The coloured segments of the x axis indicate the seasons (blue-winter, green-spring, red-summer, brown-autumn), and the vertical dotted lines the season transitions.

5.4. DISCUSSION

The sequential habitat use pattern identified for each sequence was found to be dependent upon the environmental context (Fig.5.2 and Fig.5.3) and seasons (Fig.5.4). At the population level (Fig.5.2), the sequential habitat use patterns were influenced by habitat proportion (mainly first DCA axis) but less clearly by structure (first & second DCA axis). However, the low maximum values for both first and especially second DCA axis potentially indicated a high individual variability within populations in the use of different sequential patterns. Results were further confirmed by the individual level analysis (Fig.5.3), which showed that even for populations with a high preponderance of one habitat type, variability in sequential habitat use patterns were observed among individuals. At the individual level we could better explain this variability by habitat composition and structure of the animal's range. Specifically, homogeneous sequential habitat use, i.e. homogeneous open and homogeneous closed, were associated both with high proportion and connectivity of the respective habitat classes and also with heterogeneous landscapes, where patches of both habitat classes were structured so as to include a lot of ecotonal habitat (i.e. edges). On the other hand, alternation, and especially alternation with mainly use of open habitat (i.e. a24), was mainly present in heterogeneous landscapes, as expected. Due to seasonal variability in sequential patterns (Fig.5.4), most individuals cannot be related to one specific sequential habitat use pattern.

5.4.1. Functional habitat types in different contexts

Habitat classifications in animal ecology studies are human simplifications of the animals' perception of the landscape. For animals, habitats represent an ensemble of resources that respond to their needs (e.g. feeding, hiding, sheltering), or that represent constraints to their distribution and fitness (e.g. exposure to bad weather conditions, predation). Our analysis took advantage of the high-resolution pan-European TCD raster dataset, which we used, for the first time, to perform a standardised multi-population analysis on sequential use of two simplified, functionally different habitat classes, open and closed, at high spatial resolution. While one habitat type, such as forest, may provide multiple resources, the same need can be satisfied by resources present in different habitats. While TCD generally identifies open/closed habitats with precision, it may overlook small features that can be used by roe deer which provide similar functions as forest or open habitats. Specifically, roe deer may alternate between dense and sparse areas in the forest, use small open patches (e.g. clearcuts) or edges (e.g. forest roads). In mainly open or heterogeneous landscapes, roe deer may find cover in hedgerows or field margins, as observed in Southwest France (FR8; see Morellet *et al.* 2011). A more detailed classification of different habitats that represent open and closed (e.g. old/young forest, clearcuts, shrubland, type of forest,etc.), such as those derived from LiDAR (Lone *et al.* 2014, Lone *et al.* 2017, Ciuti *et al.* 2018), might be needed to capture more fine-scaled patterns. Nevertheless, such detail is currently available at the local or regional scale only. Images from most recent satellites, such as Sentinel 2 (ESA, https://sentinel.esa.int) could reach a resolution down to 1m, so that fine scale responses to habitat composition and function will possibly be detectable in future studies.

5.4.2. Individual variability

As expected, the occurrence of sequential habitat use patterns within populations helped to characterise the spatio-temporal use of landscapes, i.e. each population can be described mainly by one homogeneous and one/two alternation patterns. Instead, sequential habitat use patterns do not characterise individuals, as in individuals generally cannot be tagged with one sequential habitat use pattern only. This finding means that individuals often use their habitat with an ensemble of sequential patterns which vary across seasons. Although individual sequential habitat use was better classified considering abundance and structure of open and closed habitats, we observed differences among individuals with similar values of habitat structure metrics. For example, Figure 5.5 shows two different individuals with MCPs similar for habitat composition, structure, and size, that nevertheless use their local environment in very different ways. Roe deer A (59% open habitat) showed a homogeneous use of open habitat during summer/autumn and alternating or random use of both open and closed during winter/early-spring. Roe deer B (52% open habitat) showed instead an alternating use of open and closed during late spring/summer, and homogeneous use of closed habitat during autumn/winter.

Indeed, we observed inter-individual variability even in populations with a strong prevalence of one habitat type over the other (e.g. Southwest France - FR8). Inter-individual variability in the sequential habitat use may be related to external factors (e.g. social fences, competition, risk of predation/hunting), or to changes in the internal state of animals (e.g. physiological phases, life-history traits), and/or to their personality. An increasing body of work relates individual behavioural variability to animal personalities (Dall *et al.* 2004, Bonnot *et al.* 2015). Indeed, for many species it has been shown that some individuals are bolder, more active and take more risks than others (Chapman *et al.* 2011, Bonnot *et al.* 2015). For example, some individuals will minimise risk by using safer habitats but obtain less access to high quality food resources, while others will expose themselves to a more risky context in order to obtain high quality food (Bonnot *et al.* 2015; Padié *et al.* 2015). Hence, while the same habitats are available at the population level, different individuals can exhibit very different space and habitat patterns of use. We propose that the sequence with which habitats are used may be part of the ‘behavioural syndrome’, defined as an ensemble of correlated behavioural traits occurring within or across behavioural contexts (Sih, et al. 2004), hence an important movement tactic to study.

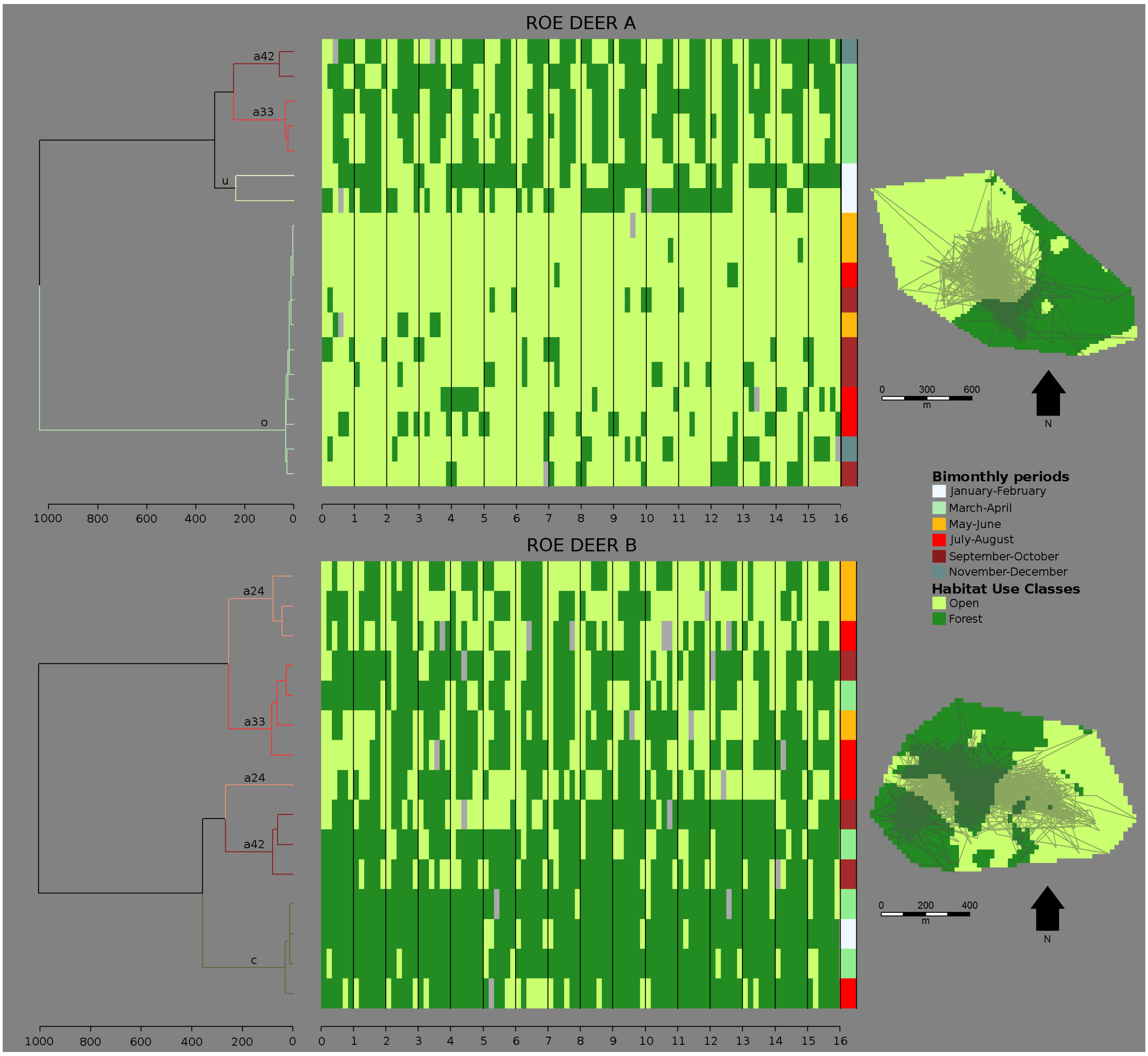


Figure 5.5. Classification trees (left) for 16-day habitat use sequences (middle) and the corresponding bimonthly period (bar at the right of the sequences) for two roe deer with similar MCP composition and structure. Branches are coloured based on the identified cluster type by IM-SAM (see Chapter 4 for more details on the SAM-procedure).

5.4.3. Temporal variability in sequential use of habitat types

Habitats can be considered *dynamic resource composites*. On one side, resources change seasonally, on the other, physiological and behavioural phases of animals affect their resource needs. In the European landscape, herbivores' essential resources (i.e. food and cover) may change due to human interference (i.e. agriculture and forest management practices) or naturally (i.e. phenological cycle of vegetation, Pettorelli *et al.* 2005). Hence, habitat use by animals should adapt to such changes. We showed that habitats can be used by animals in a plastic manner by adjusting their sequential use. Although relatively overlooked so far, the order in which habitats are used can be considered a tactic to address different resource needs and constraints, such as avoidance of predators, or hunting, at different temporal scales. We found that seasonal resource dynamics corresponded to temporal variability in sequential habitat use patterns by roe deer. In some populations, seasonal variability are associated with partially migratory behaviour, where some individuals stay at lower elevations in mainly closed habitats while others migrate between low and high elevations and alternate between open and closed habitats (Cagnacci *et al.* 2011, De Groeve *et al.* 2016). For example, migrant individuals in the Italian Alps (IT1) showed alternation between open and closed habitats during summer and a relatively homogeneous use of closed habitats during winter. Elsewhere, seasonal changes in sequential patterns do not correspond to a range shift (i.e. in space), but possibly to a different functional use of habitats. For example, in Southern Germany (DE15), a minimally productive area in winter has abundant crops in summer, which can also offer cover. As a consequence, alternation or homogeneous use of closed habitats during winter abruptly changes into homogeneous use of open in summer. Interestingly, many populations show a bimodal pattern with more alternation between closed and open habitats in spring, that might be explained by green up in open habitats such as pastures, and in fall, perhaps due to availability of plants with delayed maturation or crop left-overs. For instance, the Swiss Alpine population (CH25), which ranges in a landscape characterised by its partly-managed Alpine habitats (e.g. cattle pastures), showed a very pronounced bimodal seasonal change in habitat alternation. Similarly, closed environments may provide a protective service from extreme weather (Mysterud *et al.* 1997, Mysterud and Østbye 2006, Ratikainen *et al.* 2007). This is especially true for Alpine/mountainous environments such as the Italian Alps (IT1, IT24), the Austrian Alps (AU17), and the Bavarian Forest (DE2): roe deer used closed habitats homogeneously in winter, most likely due to the reduced snow depth with respect to open habitats, thermal protection, and increased food availability compared to open habitats (Ewald *et al.* 2014).

Temporal patterns in sequential habitat use can be expected to follow behavioural and physiological phases, further to environmental seasonality. In some populations we see a drop in the alternating use of habitats and a slight increase in the use of open habitats during the 12th-15th biweek, corresponding to the rutting period (e.g. Italian Alps - IT1, IT24; Bavarian Forest - DE2). Other patterns may emerge at a finer temporal resolution than the one considered in this paper. For example, roe deer doe hide their fawns during early life-stages (first weeks of life), only returning to feed them (Van Moorter *et al.* 2009). This behaviour should correspond to an alternation between different habitats that however we did not observe, but might become evident at a higher resolution (temporal and habitat classes) as we suggest below.

5.4.4. Methodological considerations

Periodicity and sequentiality have recently received more attention in the animal movement literature (Berger-Tal and Bar-David 2015). Fourier and wavelet transformations allow the identification of multi-scaled temporal resolutions of repetitive behaviours in ecological or movement variables (Wittemyer *et al.* 2008, Bar-David *et al.* 2009, Polansky *et al.* 2010, Riotte-Lambert *et al.* 2013, 2016). Recently, Riotte-Lambert *et al.* (2016) investigated the degree of repetitiveness in resource patch visitation (i.e. the sequence with which patches are revisited). Benhamou and Riotte-Lambert (2012) combined Utilisation Distribution and residence time to gain insights into recursive use. Demšar *et al.* (2015b) suggested a 3D-geovisualisation to characterise whether re-visitations are random in time or follow a sequential pattern. Analysing the repetitiveness and sequentiality of resource and habitat use is essential to evaluate the influence of resource predictability and dynamics on animal movement. In addition, movement recursions provide insights into memory based processes governing animal movement and space use patterns (Fagan *et al.* 2013, Berger-Tal and Bar-David 2015, Riotte-Lambert *et al.* 2016). Here we propose IM-SAM as a complementary methodological framework to investigate spatio-temporal patterns of habitat use. IM-SAM offers a complete and robust approach to explore, visualise, identify and classify complex and multi-scaled sequential patterns over multiple populations. Importantly, the computation of habitat sequence similarity accounts for individual variability in habitat availability and distribution (see Chapter 4).

While this study applied IM-SAM to habitat use sequences, other important components of a species internal or external state (Nathan 2008) can be studied with this approach. For example, instead of habitats one might look into sequentiality in movement parameters (e.g. step length, speed, turning angle) and derived metrics such as temporal autocorrelation (Wittemyer *et al.* 2008, Boyce *et al.* 2010), or movement and behavioural states (e.g. foraging/resting, exploratory/encamped (Morales *et al.* 2004); Gurarie 2009, 2016), or areas of intensive use (Benhamou and Riotte-Lambert 2012, Berger-Tal and Bar-David 2015). Finally, other non-movement parameters obtained through biologging, such as behaviours, activity, heart rate, body temperature, or food intake (see Ropert-Coudert *et al.* 2005) could also be analysed as sequence of events.

IM-SAM is not only applicable to binary datasets (open-closed) but could be extended to more complex classifications, and higher temporal resolutions (e.g. hours). Increasing granularity of sensing and tracking data offers several avenues of application of the IM-SAM approach to explore and identify sequential patterns, within and between populations at multiple spatio-temporal scales.

DECLARATIONS

Acknowledgements

We thank Fondazione Edmund Mach and Special Research Fund (BOF) of Ghent University for their financial support to JDG under grant number 01sf2313. We thank Research Foundation – Flanders (FWO) for the travel grant for study abroad to Harvard University under grant number V417616N to JDG. FC was supported by the Sarah and Daniel Hrdy Fellowship during part of the development of this ms. We wish to thank the EuroDEER collaborative project to provide geospatial and environmental data (paper no. 009 of the EuroDEER series; www.euroungulates.org). The EuroDEER spatial database is hosted by Fondazione Edmund Mach. We are grateful to the Copernicus project for the use of free data (European Union, EEA - European Environment Agency). Here we confirm that all conditions are met required for free, full and open access to this data set, established by the Copernicus data and information policy Regulation (EU) No 1159/2013 of 12 July 2013.. We are particularly grateful to Ferdinando Urbano for his support with data management, and to Lino Ometto for precious advice in data exploration. The GPS data collection of the Fondazione Edmund Mach was supported by the Autonomous Province of Trento under grant number 3479 to FC (BECOCERWI—behavioural Ecology of Cervids in Relation to Wildlife Infections), and by the help of the Wildlife and Forest Service of the Autonomous Province of Trento and the Hunting Association of Trento Province (ACT). The Norwegian participation was funded by the Research Council of Norway (grant number 251112). The GPS data collection of the CEFS-INRA was supported by the "Move-It" ANR grant ANR-16-CE02-0010-02.

REFERENCES

Abbas, F., Morellet, N., Hewison, A.J.M., Merlet, J., Cargnelutti, B., Lourtet, B., Angibault, J.M., Daufresne, T., Aulagnier, S. and Verheyden, H., 2011. Landscape fragmentation generates spatial variation of diet composition and quality in a generalist herbivore. *Oecologia*, **167** (2), 401-411.

Andersen, R., Duncan, P., and Linnell, J.D.C., 1998. *The European roe deer: the biology of success.* Oslo: Scandinavian University Press.

Bar-David, S., Bar-David, I., Cross, P.C., Ryan, S.J., Knechtel, C.U. and Getz, W.M., 2009. Methods for assessing movement path recursion with application to African buffalo in South Africa. *Ecology*, **90** (9), 2467-2479.

Benhamou, S. and Riotte-Lambert, L., 2012. Beyond the Utilization Distribution: Identifying home range areas that are intensively exploited or repeatedly visited. *Ecological Modelling*, **227**, 112-116.

Berger-Tal, O. and Bar-David, S., 2015. Recursive movement patterns: review and synthesis across species. *Ecosphere*, **6** (9), 1-12.

Bissonette, J. A. and Storch, I. 2003. Part 1: Conceptual and Quantitative Linkages. *In:* Bissonette, J. A. and Storch, I., eds. *Landscape Ecology and Resource Management: Linking Theory with Practice*. Washington DC: Island Press, 11-14.

Bjørneraas, K., Solberg, E.J., Herfindal, I., Moorter, B.V., Rolandsen, C.M., Tremblay, J.P., Skarpe, C., Sæther, B.E., Eriksen, R. and Astrup, R., 2011. Moose Alces alces habitat use at multiple temporal scales in a human-altered landscape. *Wildlife Biology*, **17** (1), 44-54.

Bonnot, N., Verheyden, H., Blanchard, P., Cote, J., Debeffe, L., Cargnelutti, B., Klein, F., Hewison, A.J.M. and Morellet, N., 2015. Interindividual variability in habitat use: evidence for a risk management syndrome in roe deer? *Behavioral Ecology*, **26** (1), 105-114.

Boyce, M.S., Pitt, J., Northrup, J.M., Morehouse, A.T., Knopff, K.H., Cristescu, B. and Stenhouse, G.B., 2010. Temporal autocorrelation functions for movement rates from global positioning system radiotelemetry data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365** (1550), 2213-2219.

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

Cagnacci, F., Focardi, S., Heurich, M., Stache, A., Hewison, A.J.M., Morellet, N., Kjellander, P., Linnell, J.D., Mysterud, A., Neteler, M. and Delucchi, L., 2011. Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors. *Oikos*, **120** (12), 1790-1802.

Chapman, B.B., Hulthén, K., Blomqvist, D.R., Hansson, L.A., Nilsson, J.Å., Brodersen, J., Anders Nilsson, P., Skov, C. and Brönmark, C., 2011. To boldly go: individual differences in boldness influence migratory tendency. *Ecology Letters*, **14** (9), 871-876.

Ciuti, S., Tripke, H., Antkowiak, P., Gonzalez, R.S., Dormann, C.F. and Heurich, M., 2018. An efficient method to exploit LiDAR data in animal ecology. *Methods in Ecology and Evolution*. **9** (4): 893-904.

Dall, S.R., Houston, A.I. and McNamara, J.M., 2004. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology letters*, **7** (8), 734-739.

De Groeve, J., Van de Weghe, N., Ranc, N., Neutens, T., Ometto, L., Rota-Stabelli, O. and Cagnacci, F., 2016. Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods. *Methods in Ecology and Evolution*, **7** (3), 369–379.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., Weiskopf, D. and Weibel, R., 2015a. Analysis and visualisation of movement: an interdisciplinary review. *Movement ecology*, **3** (1), 5.

Demšar, U., Buchin, K., van Loon, E.E. and Shamoun-Baranes, J., 2015b. Stacked space-time densities: a geovisualisation approach to explore dynamics of space use over time. *GeoInformatica*, **19** (1), 85-115.

Dupke, C., Bonenfant, C., Reineking, B., Hable, R., Zeppenfeld, T., Ewald, M. and Heurich, M., 2017. Habitat selection by a large herbivore at multiple spatial and temporal scales is primarily governed by food resources. *Ecography*, **40** (8), 1014-1027.

Ewald, M., Dupke, C., Heurich, M., Müller, J. and Reineking, B. 2014. The application of LiDAR remote sensing and activity sensors to analyze winter habitat selection of European roe deer. *Forestry,* **2014** (5), 1374-1390.

Fagan, W.F., Lewis, M.A., Auger-Méthé, M., Avgar, T., Benhamou, S., Breed, G., LaDage, L., Schlägel, U.E., Tang, W.W., Papastamatiou, Y.P. and Forester, J., 2013. Spatial memory and animal movement. *Ecology letters*, **16** (10), 1316-1329.

Fryxell, J.M. and Sinclair, A.R.E., 1988. Causes and consequences of migration by large herbivores. *Trends in ecology & evolution*, **3** (9), 237-241.

Gurarie, E., Andrews, R.D. and Laidre, K.L., 2009. A novel method for identifying behavioural changes in animal movement data. *Ecology letters*, **12** (5), 395-408.

Gurarie, E., Bracis, C., Delgado, M., Meckley, T.D., Kojola, I. and 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.

Hebblewhite, M. and Merrill, E.H., 2009. Trade-offs between predation risk and forage differ between migrant strategies in a migratory ungulate. *Ecology*, **90** (12), 3445-3454.

Hill, M.O. and Gauch, H.G., 1980. Detrended correspondence analysis: an improved ordination technique. *In:* van der Maarel, E., ed. *Classification and ordination. Advances in vegetation science*, Dordrecht: Springer, 47-58.

Hutchinson, G.E., 1965. *The Ecological Theater and the Evolutionary Play*. Yale: University Press.

Johnson, D.H., 1980. The Comparison of usage and availability measurements for evaluating resource preference. *Ecology,* **61**, 65-71.

Jongman, R.H..G., Ter Braak, C.J.F. and Van Tongeren, O.F.R., 1995. *Data analysis in community and landscape ecology*. Cambridge: Cambridge University Press.

Kays, R., Crofoot, M.C., Jetz, W. and Wikelski, M., 2015. Terrestrial animal tracking as an eye on life and planet. *Science*, **348** (6240), aaa2478.

Krop-Benesch, A., Berger, A., Hofer, H. and Heurich, M., 2013. Long-term measurement of roe deer (Capreolus capreolus)(Mammalia: Cervidae) activity using two-axis accelerometers in GPS-collars. *Italian Journal of Zoology*, **80** (1), 69-81.

Law, B.S. and Dickman, C.R., 1998. The use of habitat mosaics by terrestrial vertebrate fauna: implications for conservation and management. *Biodiversity & Conservation*, **7** (3), 323-333.

Legendre, P. and Gallagher, E.D. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia,* **129** (2), 271–280.

Lone, K., Loe, L.E., Gobakken, T., Linnell, J.D., Odden, J., Remmen, J. and Mysterud, A., 2014. Living and dying in a multi-predator landscape of fear: roe deer are squeezed by contrasting pattern of predation risk imposed by lynx and humans. *Oikos*, **123** (6), 641-651.

Lone, K., Mysterud, A., Gobakken, T., Odden, J., Linnell, J. and Loe, L.E., 2017. Temporal variation in habitat selection breaks the catch-22 of spatially contrasting predation risk from multiple predators. *Oikos*, **126** (5), 624-632.

Mancinelli, S., Peters, W., Boitani, L., Hebblewhite, M. and Cagnacci, F., 2015. Roe deer summer habitat selection at multiple spatio-temporal scales in an Alpine environment. *Hystrix, the Italian Journal of Mammalogy*. **26** (2), 132-140.

Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L. and Erickson, W.P., 2002. Introduction to Resource Selection Studies. *In: Resource selection by animals: statistical design and analysis for field studies*. Dordrecht: Kluwer Academic Publishers, 1-15.

McGarigal, K., Stafford, S. and Cushman, S. 2000. *Multivariate statistics for wildlife and ecology research*. New York: Springer.

Morales, J.M., Haydon, D.T., Frair, J., Holsinger, K.E. and Fryxell, J.M., 2004. Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, **85** (9), 2436-2445.

Morellet, N., Van Moorter, B., Cargnelutti, B., Angibault, J.M., Lourtet, B., Merlet, J., Ladet, S. and Hewison, A.J.M., 2011. Landscape composition influences roe deer habitat selection at both home range and landscape scales. *Landscape Ecology*, **26** (7), 999-1010.

Morris, D.W., 1987. Ecological scale and habitat use. *Ecology,* **68** (2), 362-369.

Mysterud, A., Bjørnsen, B.H. and Østbye, E., 1997. Effects of snow depth on food and habitat selection by roe deer Capreolus capreolus along an altitudinal gradient in south-central. *Wildlife Biology*, **3** (1). 27-33.

Mysterud, A. and Østbye, E., 2006. Effect of climate and density on individual and population growth of roe deer Capreolus capreolus at northern latitudes: the Lier valley, Norway. *Wildlife Biology,* **12** (3), 321-329.

Mysterud, A., Loe, L.E., Zimmermann, B., Bischof, R., Veiberg, V. and Meisingset, E., 2011. Partial migration in expanding red deer populations at northern latitudes–a role for density dependence?. *Oikos*, **120** (12), 1817-1825.

Nathan, R., 2008. An emerging movement ecology paradigm. *Proceedings of the National Academy of Sciences*, 105 (49), 19050–19051.

Orians, G.H. and Wittenberger, J.F., 1991. Spatial and temporal scales in habitat selection. *The American Naturalist,* **137**, S29-S49.

Owen-Smith, N., Fryxell, J.M. and Merrill, E.H., 2010. Foraging theory upscaled: the behavioural ecology of herbivore movement. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365** (1550), 2267-2278.

Padié, S., Morellet, N., Hewison, A.J.M., Martin, J.L., Bonnot, N., Cargnelutti, B. and Chamaillé-Jammes, S., 2015. Roe deer at risk: teasing apart habitat selection and landscape constraints in risk exposure at multiple scales. *Oikos*, **124** (11), 1536-1546.

Pagon, N., Grignolio, S., Pipia, A., Bongi, P., Bertolucci, C. and Apollonio, M., 2013. Seasonal variation of activity patterns in roe deer in a temperate forested area. *Chronobiology international,* **30** (6), 772-785.

Peters, W., Hebblewhite, M., Mysterud, A., Spitz, D., Focardi, S., Urbano, F., Morellet, N., Heurich, M., Kjellander, P., Linnell, J.D.C. and Cagnacci, F., 2017. Migration in geographic and ecological space by a large herbivore. *Ecological Monographs*, **87** (2), 297-320.

Pettorelli, N., Gaillard, J.M., Mysterud, A., Duncan, P., Chr. Stenseth, N., Delorme, D., Van Laere, G., Toïgo, C. and Klein, F., 2006. Using a proxy of plant productivity (NDVI) to find key periods for animal performance: the case of roe deer. *Oikos*, **112** (3), 565-572.

Polansky, L., Wittemyer, G., Cross, P.C., Tambling, C.J. and Getz, W.M., 2010. From moonlight to movement and synchronized randomness: Fourier and wavelet analyses of animal location time series data. *Ecology,* **91**, 1506-1518.

Post, E. and Stenseth, N.C., 1999. Climatic variability, plant phenology, and northern ungulates. *Ecology*, **80** (4), 1322-1339.

Rao, C.R., 1964. The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A*, **26** (4), 329-358.

Ratikainen, I.I., Panzacchi, M., Mysterud, A., Odden, J., Linnell, J. and Andersen, R., 2007. Use of winter habitat by roe deer at a northern latitude where Eurasian lynx are present. *Journal of Zoology*, **273** (2), 192-199.

Riotte-Lambert, L., Benhamou, S. and Chamaillé-Jammes, S., 2013. Periodicity analysis of movement recursions. *Journal of theoretical biology*, **317**, 238-243.

Riotte-Lambert, L., Benhamou, S. and Chamaillé-Jammes, S., 2016. From randomness to traplining: a framework for the study of routine movement behavior. *Behavioral Ecology*, **28** (1), 280-287.

Ropert-Coudert, Y. and Wilson, R.P., 2005. Trends and perspectives in animal-attached remote sensing. *Frontiers in Ecology and the Environment*, **3** (8), 437-444.

Ter Braak, C.J., 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67** (5), 1167-1179.

Ter Braak, C.J. and Verdonschot, P.F., 1995. Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic sciences*, **57** (3), 255-289.

Tufto, J., Andersen, R. and Linnell, J.D.C., 1996. Habitat use and ecological correlates of home range size in a small cervid: the roe deer. *Journal of Animal Ecology*, **65** (6), 715-724.

Urbano, F., Cagnacci, F., Calenge, C., Dettki, H., Cameron, A. and Neteler, M., 2010. Wildlife tracking data management: a new vision. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, **365**, 2177-2185.

Van Moorter, B., Gaillard, J.M., McLoughlin, P.D., Delorme, D., Klein, F. and Boyce, M.S., 2009. Maternal and individual effects in selection of bed sites and their consequences for fawn survival at different spatial scales. *Oecologia*, **159** (3), 669-678.

VanDerWal, J., Falconi, L., Januchowski, S., Shoo, L. and Storlie, C., 2014. SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises. *R package version 1*, 1-221. https://CRAN.R-project.org/package=SDMTools

Wittemyer, G., Polansky, L., Douglas-Hamilton, I. and Getz, W.M., 2008. Disentangling the effects of forage, social rank, and risk on movement autocorrelation of elephants using Fourier and wavelet analyses. *Proceedings of the National Academy of Sciences*, **105** (49), 19108-19113.

APPENDIX 5A – MULTIVARIATE STATISTICS

a. Sequential habitat use patterns varies across geographic space.

Table 5A.1. Eigenvalues, decorona values and axis lengths for all DCA axes.

	DCA1	DCA2	DCA3	DCA4
Eigenvalues	0.339	0.04155	0.041949	0.043878
Decorona values	0.3472	0.02674	0.001259	0.0004566
Axis lengths	2.1217	0.59292	0.559777	0.652769

Table 5A.2. Species scores of sequential habitat use patterns for all DCA axes.

Species	DCA1	DCA2	DCA3	DCA4	Totals
a24	-0.3343	1.5452	1.5579	-1.1969	544
a33	-0.0913	1.0125	0.9826	-1.1107	1009
a42	0.3958	-1.4542	-1.4992	0.7093	589
c	1.4677	-0.0273	-0.0644	0.2853	1317
o	-1.1173	-0.5206	-0.4831	0.2198	1702
u	0.0398	-0.3469	-0.2026	2.5071	241

Table 5A.3. Site scores of different populations for all DCA axes.

Sites	DCA1	DCA2	DCA3	DCA4	Totals
AU17	1.03402	-0.00729	-0.04012	0.10931	141
CH25	-0.40222	0.06439	0.0789	-0.06239	1559
DE15	-0.2861	-0.24716	-0.2406	0.10988	865
DE2	0.73635	-0.01462	-0.03209	0.19276	375
DE31	0.02045	-0.12257	-0.12266	0.08971	273
FR8	-0.18494	0.24236	0.24257	-0.27759	1107
IT1	0.81677	0.05285	0.02669	0.06284	521
IT24	1.17098	-0.1991	-0.22833	0.37517	358
NK5	-0.95068	-0.35056	-0.31721	0.14912	203

b. Sequential habitat use patterns varies among individuals

Table 5A.4. Inertia and Rank

	Inertia	Rank
Total	6	
Unconstrained	6	6

Inertia is correlations

PC1	PC2	PC3	PC4	PC5	PC6
1.6932	1.5372	0.879	0.8464	0.6685	0.3757

Table 5A.5. Species scores for sequential habitat use patterns for all PCA-axes

Species	PC1	PC2	PC3	PC4	PC5	PC6
a24	0.4420608	-0.2915859	0.30207016	-0.52696625	-0.55881103	0.1958325
a33	0.38860249	-0.4709225	0.18494347	0.0390161	0.71532156	0.2825046
a42	-0.12275717	-0.4867856	0.36961936	0.69389601	-0.33975326	-0.1201167
c	-0.66936854	-0.1222922	0.11051972	-0.14883178	-0.01723742	0.7087508
o	0.42797457	0.487813	-0.03129066	0.45786415	-0.15947641	0.5855124
u	0.08534961	-0.4509241	-0.85131507	0.08661502	-0.18677035	0.1491985

Table 5A.6. Site scores for individual animals for all PCA-axes

Sites	PC1	PC2	PC3	PC4	PC5	PC6
1	-0.028863109	-0.0312128914	0.0631903952	0.1288660973	-0.04344534	-0.158337326
3	-0.114992505	0.027301069	0.0057794067	-0.0488219378	0.007721166	0.0431837336
4	-0.031501331	-0.0995115645	-0.0576489537	0.1051555724	-0.1135378	-0.1026477751
5	0.005310877	-0.0370235782	0.0552886505	-0.0691199119	0.03767123	0.0036602976
11	-0.098566593	0.0373398162	-0.0048383141	-0.0580751064	0.01596633	0.0137648782
12	-0.124786256	0.0254231704	0.0080236292	-0.0519017889	0.007319801	0.0651995201
14	-0.019907758	-0.1264072468	0.1204590376	-0.0413222877	0.1399913	0.2033123456
19	0.048796821	-0.1379906053	-0.0941462639	0.0640127338	0.09830387	-0.0401199814
20	-0.085143481	0.0283482892	0.0060745812	-0.0223759864	-2.42881E-005	-0.0200696234
768	-0.024498849	-0.0351310431	-0.0122037643	-0.0149229693	-0.02812846	-0.0463148704
769	-0.11325722	0.0345229683	-0.0014719803	-0.062694883	0.01536428	0.0467885579
770	-0.098566593	0.0373398162	-0.0048383141	-0.0580751064	0.01596633	0.0137648782
771	-0.118463075	0.0128572705	0.0202821806	-0.0210760473	-0.007565065	0.035974085
772	0.051697564	-0.0480094111	0.0436998341	-0.1595161969	-0.05926671	0.0201314644
773	-0.017847539	-0.0109656528	-0.1106708139	-0.0236116301	-0.01675303	-0.0910804242
774	0.023466873	-0.2436908144	-0.094306077	0.1664965294	-0.03135186	0.0099032547
776	-0.113789435	0.028032856	0.0025330089	-0.0636150265	0.02794943	0.0645328421
782	-0.054455485	-0.0092939767	0.0151925812	-0.0448196177	-0.05798859	0.0391067414
783	-0.079511305	0.0346055011	-0.00532177	-0.0528355476	0.02935421	-0.0125224108
784	-0.009059926	-0.0497111932	0.0486627679	0.0187133896	0.14063	-0.0188887682
785	-0.088496711	-0.0222307437	0.0515672146	0.05968794	-0.03159003	-0.0523688985
787	-0.080213861	-0.0048597973	0.0240788497	-0.0272303055	0.06460926	0.0402371838
789	-0.004367297	-0.0473008335	0.0455947059	-0.0227062414	0.1549205	0.0233284753
791	-0.024394274	-0.0417026518	0.0699928963	-0.1035662977	-0.0160656	0.0797536037
792	-0.138039013	-0.1405884931	0.0881614156	0.1475006206	-0.07112607	0.1095312632
796	-0.108360344	0.0354619176	-0.0025940916	-0.0611549575	0.01556496	0.0357806647
797	-0.045797091	-0.0105355899	0.0209876515	-0.0330842023	0.1114126	0.0200196077

798	-0.030489654	-0.0032168283	0.0254798081	0.0331679984	0.01871711	-0.1271526832
799	-0.055197267	0.0063250614	0.0144633041	-0.0186105342	0.05302752	-0.0325465666
800	-0.105198754	0.0291789676	0.0035351841	-0.0457420867	0.00812253	0.0211679471
801	-0.025638213	-0.0512703584	0.0182195022	0.0171853213	0.1078843	-0.0031281818
837	0.034818344	0.1074265853	-0.0250507465	0.0438528987	0.001915151	-0.0084520063
1333	0.025248363	0.0679793922	0.0041287498	0.0575907991	-0.02786223	-0.0450051959
1335	0.052652174	0.0892893921	-0.0022560892	-0.0082884746	-0.0617786	-0.002964381
1336	0.006220078	0.0839880884	-0.0254746319	0.006565136	0.003500065	-0.0855754922
1338	0.047672399	0.0667064223	0.0055259571	-0.0219226472	-0.0311819	-0.0040790133
1339	0.039970367	0.0049497342	-0.1458924487	0.0269542342	0.03225936	-0.0094068036
1341	0.026047574	0.1077069011	-0.0274865349	0.036566211	-0.008257818	-0.0279860508
1343	0.017235353	0.0971652066	-0.0265923558	0.0232323999	-0.003032092	-0.0535813581
1347	0.017235353	0.0971652066	-0.0265923558	0.0232323999	-0.003032092	-0.0535813581
1425	-0.074082214	0.0420345627	-0.0104488705	-0.0503754786	0.01696974	-0.0412745881
1453	0.001096655	-0.0632064078	-0.2600374907	0.0381374577	-0.06741992	0.0697752994
1454	0.015032298	0.0945297829	-0.026368811	0.0198989471	-0.001725661	-0.0599801849
1455	-0.07612648	0.0140859149	0.0174345662	0.006656228	-0.01340204	-0.0667017786
1456	0.007648862	0.0230338051	-0.0954845262	-0.0189888818	-0.008267342	-0.095722727
1457	-0.004813473	0.0773549695	-0.0485785535	0.0131971391	-0.007994646	-0.0151298572
1495	-0.071090964	0.0027764266	0.0210760356	-0.0084372221	0.03219721	-0.0208719952
1497	-0.118154095	0.033584019	-0.000349869	-0.0642348086	0.0151636	0.0577964512
1499	-0.008167976	0.0058604069	0.0329522393	-0.0760914091	-0.02253411	-0.0072203519
1500	-0.092842383	-0.0294448065	0.0633654463	0.0601875062	-0.06559986	-0.045446803
1501	-0.040349011	-0.0158717541	0.0305416606	-0.029504785	0.07780413	0.0049259167
1503	-0.139476883	0.0226063225	0.011389963	-0.0565215656	0.006717755	0.0982231999
1506	-0.067963279	0.0309400731	-0.0001362699	-0.0625092935	0.01617469	-0.0172748865
1606	-0.109424774	0.022481693	0.0054158869	-0.0629952444	0.04073526	0.0712692329
1607	-0.18844564	0.0132168296	0.0226110758	-0.0719208212	0.004710933	0.2083021325
1610	-0.132844722	0.0307671711	0.0030164648	-0.0688545853	0.01456155	0.090820131
1611	-0.169476097	-0.0244808702	0.05938673	0.0205564037	-0.03994367	0.1206258271
1614	-0.161108651	-0.0090981224	0.0437618448	-0.0056495613	-0.02445675	0.1168275824
1618	-0.074680473	-0.0184053867	0.0359558027	0.0614800077	-0.08896244	0.0448374214
1619	-0.125605111	-0.0394598093	-0.1503639761	-0.0364400307	-0.0352248	0.1479275158
1621	-0.159682346	-0.0226029716	0.0571425075	0.0236362548	-0.0395423	0.0986100405
1623	-0.09614991	-0.0566665824	-0.0290780464	0.0486491156	-0.03228188	0.0236566342
1670	0.063751954	0.0541125556	0.0208908916	-0.0613104235	-0.08129477	-0.0158546986
1672	0.052862634	0.0034034931	0.0429433528	-0.1416197767	-0.1031453	-0.0866720216
1674	-0.102037163	0.0228960176	0.0096644598	-0.0303292158	0.0006800968	0.0065552295
1675	0.017984397	-0.0066188994	0.0617223025	0.0333534411	-0.09909185	-0.0873831982
1676	0.103493592	-0.0401344073	0.0823646435	-0.1800420992	-0.09404338	-0.0215695707
1678	0.094889155	-0.0198987089	0.0769171983	-0.1904014461	-0.1598733	-0.0358608898
1679	0.017005645	0.0194679192	0.0080685591	-0.0119309928	0.06813608	-0.0762904483
1714	-0.103463468	0.0364008669	-0.0037162029	-0.0596150319	0.01576565	0.0247727714
1796	0.032728467	0.0133740264	0.0165432446	0.031682234	0.1056985	-0.0339473835
1798	0.094731969	-0.046819388	0.0720452737	-0.1003349809	0.02653078	-0.0099718744
1807	0.03723186	0.02417611	0.0199251508	-0.0861449506	-0.040758	-0.0857608202
1808	0.030247507	0.0122023803	0.0093326397	-0.0415248074	0.07799838	-0.0834993626
1817	-0.026286629	0.0032258687	0.0313019693	0.0727545342	-0.02565871	-0.1506544196
1823	-0.052356102	0.0373491955	-0.0041280289	-0.024596069	0.02010164	-0.0583690477
1877	-0.103463468	0.0364008669	-0.0037162029	-0.0596150319	0.01576565	0.0247727714
1894	-0.093669717	0.0382787655	-0.0059604254	-0.0565351808	0.01616701	0.0027569849
1895	0.037398729	0.0207988342	0.0272217072	-0.1164787631	-0.09047719	-0.0995204206
1911	-0.108360344	0.0354619176	-0.0025940916	-0.0611549575	0.01556496	0.0357806647
1920	0.056393485	-0.0085939496	0.0196516618	-0.0165920852	0.1641743	-0.0301399545
1921	-0.089305056	0.0327276025	-0.0030775474	-0.0559153987	0.02895284	0.0094933758

1922	0.071649607	-0.0567666836	0.0417149317	-0.0447075629	0.2225618	-0.0285955436
1923	0.023430023	-0.0031486251	0.0060366852	-0.0272384767	0.1470356	-0.0949975054
1924	0.11496749	-0.0245784682	0.0961583433	-0.2331622431	-0.2402722	-0.0179131748
1925	0.034859794	0.1182485956	-0.028380714	0.0499000222	-0.01348354	-0.0023907436
1926	0.049866683	0.0284764584	0.0033541478	0.0127629567	0.1313584	-0.0182115836
1928	0.026983113	0.0985339498	-0.0134308505	0.0709790071	-0.02615691	-0.0223980458
1929	0.038768509	-0.0063446802	0.0111318067	-0.0159985542	0.1686879	-0.0623282433
1937	0.012746344	0.0702503388	-0.0194853312	0.0044712474	0.03037816	-0.0785015373
1967	0.003051705	0.0463358822	-0.0709941948	-0.0139395781	-0.01840046	-0.0794723257
1974	-0.015276473	0.0112374212	-0.0161682826	-0.0559529395	-0.02339605	-0.0892144405
1976	-0.035237362	-0.0568052999	-0.2498031844	-0.0318472502	-0.0567626	-0.0030846169
2041	0.028157851	-0.0436917398	0.0873346293	-0.0108810172	-0.05397695	0.0110550548
2042	0.008776857	-0.015496188	-0.2094790671	0.0033901777	-0.0219779	-0.0731078185
2043	0.065413488	0.0327251378	0.0124158583	-0.0002838098	0.08789286	0.0067593409
2044	-0.098475687	-0.0978042673	0.0920322516	0.2036688085	-0.1271783	-0.1099517376
2045	0.052528361	0.033490703	0.0180420266	-0.0536909984	-0.009644727	-0.040150495
2046	0.032408041	0.0506811104	-0.0081773641	0.0102838284	0.08021506	-0.045157147
2047	-0.08294792	-0.0883450393	0.1469341526	0.2824509709	-0.135349	0.0016974181
2048	0.060162176	-0.0127823737	0.0598839604	0.0587406287	0.04737513	0.030141641
2049	-0.049027123	-0.0656001753	0.1056129721	0.2150710412	-0.1051041	-0.1518892226
2053	0.031302396	-0.0613262397	-0.014915975	-0.0241661753	-0.04281504	0.0098564836
2054	0.095084158	-0.0252424351	0.0828746105	0.0995929798	0.03684314	0.0590373659
2055	0.054450037	0.0190015141	0.0349398465	0.1081755231	0.06990552	0.0082185185
2056	0.116832106	-0.0293736639	0.0732773866	0.0271598245	0.1075909	0.1048678837
2058	0.002138466	0.050326302	0.0017137045	-0.0459496095	-0.0194317	-0.0416365831
2059	-0.061968068	0.0104126505	0.0180732216	0.0103558612	-0.0002148446	-0.0819811743
2060	0.007156689	0.0281498204	0.0091974791	-0.0135443744	-0.002523443	-0.1179921387
2061	-0.118633416	-0.0201179156	0.0456778111	0.0054493383	0.01510484	0.0709893954
2063	-0.002043873	0.0074861602	0.0290960678	-0.0357987636	-0.04619485	-0.1130032129
2065	-0.006873762	-0.016165222	-0.0479372701	0.0799567875	0.006184378	0.0351891647
2066	0.017530007	0.0286881921	-0.055943144	0.038419614	0.01433239	-0.064539632
2067	0.02539101	0.068750758	-0.0134916431	0.0175046971	0.05052342	-0.028425555
2068	-0.04090246	-0.0503523217	0.0851775403	-0.0032670361	-0.05530395	0.0265277326
2069	0.109597866	0.1645649578	-0.0226613665	0.1390489228	0.003692566	0.1909243171
2070	-0.09357343	-0.0222278223	-0.0609664724	-0.0074624475	-0.01449525	0.0313395406
2071	0.001813968	0.0787172412	-0.0250275423	-0.0001017696	0.006112928	-0.0983731458
2072	0.067998218	-0.0766164329	0.0886137958	0.2011405443	-0.1358716	0.0994885222
2073	-0.005653062	-0.0198598664	0.0069831253	-0.0263477028	-0.0003247818	-0.0762149166
2074	0.131091714	-0.0430975727	0.0596256167	-0.0152737678	0.2145625	0.1001068661
2075	0.041552929	0.1011335702	-0.0150948569	0.017472321	-0.03632464	-0.0090763891
2076	-0.017400923	-0.1740634685	0.114853259	0.0991203905	0.1802416	0.1290490176
2078	0.100651726	0.1730702779	-0.0584552775	0.1433143763	-0.04274405	0.1843315528
2079	0.039224454	0.1126974326	-0.0254978361	0.0505198042	-0.0006977121	0.0043456473
2081	0.052484234	0.1393319846	-0.0301690723	0.0765676444	-0.023935	0.048799871
2082	-0.108093596	-0.0837256179	-0.0649576375	0.0715715935	-0.1032803	0.0694861177
2239	0.085843476	0.0729549898	0.0321781267	0.1841445071	0.01186855	0.112718622
2240	0.134273144	-0.0187182837	0.1228876613	-0.0850018382	-0.2485658	0.0848125995
2241	0.12466715	-0.0819024667	-0.0028459679	-0.1360095203	-0.1319599	0.1488506915
2242	0.171375029	-0.0710595967	0.1413542439	-0.1050906303	-0.1204237	0.112441741
2244	0.073247804	0.1305451876	-0.0078170776	0.1137203623	-0.04765418	0.0929749916
2246	0.065828519	0.117612582	-0.0105756037	0.0329262718	-0.06015629	0.0483696429
2247	0.11897258	0.0685356088	0.0204698466	0.1323714185	0.119156	0.1675562644
2248	-0.070366989	-0.0255276121	-0.0165127313	0.0051853274	-0.006892822	0.0484339211
2249	0.150317766	-0.2103139392	-0.0586673309	-0.0808194673	0.06865228	0.1050277038

2250	0.146671639	-0.2736824298	-0.2669408583	-0.0294325512	0.06289117	0.1193778019
2251	0.034520768	-0.0101450746	-0.0071861134	-0.0717809436	0.03559876	-0.0897038175
2252	0.066692984	-0.0368329782	0.0537975008	0.0059328605	0.1131781	-0.0220447657
2253	0.092360974	0.0005877808	0.0443558355	-0.0188123147	0.06568082	0.0353935507
2254	0.024532432	-0.0156988521	0.027388926	-0.0231594931	0.07941727	-0.1031691008
2255	-0.012841682	-0.00521152	0.0262967993	0.0102410894	0.02596655	-0.1215639438
2256	0.023199364	-0.1258367295	0.0381193049	0.1022996404	0.1408151	-0.0395637363
2258	-0.010966596	-0.010537619	0.0320241945	0.0556139552	0.02315545	-0.1506459696
2259	0.065495316	0.1010344751	-0.0148605033	0.0663327436	0.04521989	0.056886518
2260	0.009652023	-0.0268295823	-0.1823699131	-0.0131570173	-0.08726945	-0.0014547278
2261	-0.00426451	0.016831845	-0.0893684806	-0.0405690919	-0.03222187	-0.0869855676
2263	0.016005125	0.040896941	-0.0052625072	-0.002796381	0.08373299	-0.042337841
2264	0.061296454	0.1498736791	-0.0310632514	0.0899014555	-0.02916072	0.0743951782
2265	0.05792363	-0.1502080041	0.0096462212	0.013034561	-0.00989187	-0.0046813902
2266	0.101808369	-0.0201669344	0.0680942683	-0.0458670713	0.00772415	0.0310209343
2267	0.101329845	0.0613828129	0.0380253676	-0.0682838155	-0.1318868	0.0541021683
2268	0.009513836	-0.0264772227	0.0367061999	0.0142601432	0.07731056	-0.1056262737
2272	0.011944198	-0.0291622903	0.0551303537	-0.0055737448	-0.01033776	-0.0932672981
2273	0.023845589	0.0584061607	-0.0066465636	-0.0212895483	0.004924199	-0.0723895292
2274	0.066594642	-0.1263341939	0.0997802434	-0.0325599116	0.003894065	0.0075258087
2279	0.074695319	-0.1195227337	0.0313956044	-0.0780896537	-0.07206812	0.0037279986
2281	0.04266471	0.1062841556	-0.0819962466	0.0672766017	-0.03245533	0.04864762
2282	-0.009449539	-0.1580380971	-0.0482237996	0.1260381234	-0.08524839	-0.0516945772
2283	0.015511095	-0.107708667	-0.2841666952	-0.0134265262	-0.03918455	-0.0140927636
2284	0.041263317	0.0020468401	0.0185231291	-0.0521186967	0.07740401	-0.0705075543
2285	0.031153042	-0.0658264138	0.0720709501	0.0513980321	0.07825276	-0.094585223
2287	0.082879294	0.1213838484	-0.0050290052	0.0526266271	-0.0426239	0.1112432788
2288	0.004229697	-0.1315746408	-0.3697538727	0.0187945674	-0.07933719	-0.0157730426
2289	0.062373311	0.0117889983	0.0392970802	-0.0286248904	0.02802771	0.0418092311
2290	0.061256611	0.0690536937	0.003191356	0.0020708723	0.004051351	0.0113269382
2291	-0.016430721	-0.045622333	0.0270801679	0.0966906852	-0.005240156	-0.1606416879
2292	0.059772811	0.0313156545	-0.0022493567	0.01139408	-0.02746847	-0.0016061919
2294	0.030091653	-0.0557087641	-0.2083030878	0.0076541818	0.02704864	0.067817126
2295	0.106709441	-0.115460304	-0.029747121	-0.0854771515	0.06788285	0.0203921171
2296	0.056435643	-0.2045849487	0.1632162789	0.0782136899	-0.2057093	-0.0271711349
2297	0.056310742	0.0120241274	0.0383149835	0.0579216419	0.04322814	-0.0079975204
2298	0.105525494	0.1525395591	-0.007621164	0.0717143921	-0.09313296	0.1506074626
2300	0.079090441	0.0509165005	0.0259860132	-0.050070501	-0.0596424	0.0168145636
2301	0.043672014	0.1287902901	-0.0292748932	0.0632338333	-0.01870927	0.0232045637
2302	0.053834044	-0.0894834961	-0.0412102877	-0.1002289245	-0.008529119	0.0286144115
2303	0.051752483	0.0398989711	-0.0826689628	-0.0007950092	-0.03791908	0.0003104205
2305	0.010094509	0.059436165	-0.0116085189	-0.0149492552	0.01941014	-0.0740358802
2447	-0.079918636	0.0372487611	-0.0009984701	-0.0628754739	0.001681064	0.0044033359
2448	-0.114992505	0.027301069	0.0057794067	-0.0488219378	0.007721166	0.0431837336

Table 5A.7. Environmental fit of the landscape metrics. The values in column PC1 and PC2 give the direction cosines corresponding to the coordinates of the heads of the vectors. Vectors in figure 6 are scaled by their respective correlations (square root of r2) so that landscape metrics with better support are longer than those with a weaker support.

Landscape Metrics	PC1	PC2	r2	Pr(>r)	
P0	0.95017	0.31172	0.5454	0.001	***
P1	-0.95682	-0.29069	0.5503	0.001	***
COHESION0	0.98813	0.15361	0.4894	0.001	***
COHESION1	-0.81995	-0.57244	0.3745	0.001	***
LSI0	-0.70856	-0.70565	0.1655	0.001	***
LSI1	0.97612	-0.21724	0.144	0.001	***
PD0	-0.85144	-0.52445	0.1836	0.001	***
PD1	0.97337	0.22925	0.2302	0.001	***
ED0	0.9762	-0.21685	0.3272	0.001	***
ED1	-0.73967	-0.67297	0.2238	0.001	***
SPLIT0	-0.99862	-0.05253	0.1972	0.001	***
SPLIT1	0.5857	0.81053	0.1895	0.001	***

Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Permutation: free
Number of permutations: 999

CHAPTER 6

DISCUSSION

This dissertation represents the progression made throughout this PhD-project, from the development of SAM (Chapter 3) to IM-SAM (Chapter 4) followed by an ecological application of the latter (Chapter 5). My co-authors and I provided a framework that allows to first *explore*, *simulate* and then *classify* sequential habitat use patterns in animal trajectories, that accounts for individual geographical distribution and availability of habitats. In Chapter 5 we further showed how identified sequential patterns can be used as dependent variable(s) for further statistical analysis. In this discussion I will first summarise the strengths and weaknesses of (IM-)SAM. Second, I comment on the ecological results, discussing the main challenges and perspectives. In a third part, I highlight the prospects, including enhancements in environmental data sources, general applicability of SAM opening up future directions of research (see also Appendix 6A) and technical developments.

6.1. SAM AND IM-SAM

To highlight the strengths of (IM-)SAM, here I summarise the most important features of the proposed framework. First off, we offer a powerful tool for discovering spatio-temporal multi-scaled patterns, which combines theory from movement ecology and GIScience into one robust framework. (IM-)SAM is not only visually informative and effective, but also reliable, robust and quantifiable. Indeed, (IM-)SAM offers a strong visual exploratory framework for presenting hierarchical relations between habitat variables, and linked behavioural mechanisms. For instance, in Chapter 3 we saw how visual exploration could reveal that migration was related to animals ranging at high elevations, and that both at high and low elevation various sequential patterns could be observed in the use of open and closed habitats (Chapter 3, Fig.3.1a). Furthermore, the approach allows one to match real observed sequences to the best fitting simulated patterns based on expectations of sequential behaviour (Chapter 3, 4 & 5). This 'tagging' procedure (see Chapter 3 & 4) translates *de facto* habitat use sequences into levels of a categorical variable that can be treated into classic statistical modelling framework (see Chapter 4 and application in Chapter 5). Hence, (IM-)SAM reconciles the exploration of the spatio-temporal structure of movement data with the possibility to test hypotheses' ecological significance, i.e. the knowledge gap identified early on in Chapter 3. As mentioned in Chapter 4, the IM-SAM procedure is hypothesis driven and only detects sequential patterns that are coded within the simulation rules. Here I would like to stress that the proposed framework is *not* meant to be used as a data mining technique, that detects any kind of pattern disregarding its biologically meaningfulness. Instead, exploratory tree building should be used to ensure that biologically important sequential patterns are identified. As showed in Chapter 3, sequential patterns that do not fit within the simulated set may separate as unidentifiable, and usually small clusters. However, in Chapter 4, although uncommon opposite day-night patterns, with main use of open during day and closed during night, were not detected as such unidentified separate clusters.

In addition, both in SAM and IM-SAM we account for the spatial structure and distribution of habitat features that affect sequences of habitat use. While Chapter 3 performs such simulations in artificial landscapes generated on the basis of study area characteristics, in Chapter 4/5 we use the actual individual spatial movement context, or home range. In Chapter 3 we show how spatial autocorrelation could be incorporated into tree clustering by assigning substitution weights to habitat classes that are more similar to each other, as determined through Join-Count Statistics. In Chapter 4/5, because we took into account real individual-based landscapes, we could study in more detail how habitat availability influences cluster separation, hence the robustness of the classification trees, and we did not need to explicitly address spatial autocorrelation. Indeed, the relevance of spatial autocorrelation also depends on the number of classes and similarity between classes. In Chapter 4/5, we took into account only two classes, hence a simplified context with respect to Chapter 3, but in turn we fine-tuned our evaluation *between* different environmental contexts. Note that in Chapter 4/5 we accounted for individual-based landscapes by constraining movement simulations within the Minimum Convex Polygon delimited by the locations of each individual. MCPs have been traditionally used in ecology as parameter-free and purely geometrical representations of an animals' home range (Nilsen *et al*. 2008), that we considered optimal for the purpose. All the same, a sensitivity analysis on the changes obtained using other estimators could be performed in the future.

Finally, we also accounted for an otherwise often neglected issue in movement ecology: the effect of missing data on spatial analysis. While missing data are accepted by the R package TraMineR, they cannot be too abundant. In fact, we found that if missing data are excessive they affect the sequence separation in clusters. Based on results from Chapter 4, we suggest only including sequences with less than 30% of missing observations.

IM-SAM procedure is generalisable and adjustable to specific methodological choices. For instance, while we used the Hamming distance as a dissimilarity measure, other metrics could be used for exploring sequential patterns. The most promising approach is Dynamic Time Warping (see R package dtw, Giorgino 2009), which has been used in sociology, linguistics and musicology to account for time (Müller 2007), and potentially also as a multi-class extension of the Fuzzy Hamming Distance (Bookstein *et al.* 2001, see Chapter 3). Alternatively, a new dissimilarity measure based on pairs or triplets of states, similar to codon models[10], could be evaluated to account for temporal autocorrelation of consecutive states. Another component we did not explore, but has been discussed with experts in phylogeny, is a simulation study generating dissimilarity trees based on maximum likelihood or Bayesian methods (Douady *et al.* 2003). In phylogeny such methods are considered as superior over the current approach. However, in transferring these approaches, caution and close collaboration with phylogenists will be needed, to assure that parameters are correctly assigned and no wrong assumptions are made. Lastly, a potential extension of the classification of sequences into sequential habitat use patterns would be to measure the degree of alternation with respect to reference sequences (e.g. a completely homogeneous or random sequence; see Studer *et al.* (2016) for a similar application in sociology). Indeed, animal behaviour can be hardly classified into stereotypical behaviours; resource patch revisitation (Riotte-Lambert *et al.* 2016) and migration (Cagnacci *et al.* 2011) are known to occur in a gradient, where stereotypical behaviours are only the end points of such gradient.

---

[10] Codons are used in molecular biology and are described by three DNA or RNA nucleotides, e.g. CAG represents the amino acid glumine, described by the respective nucleotides Citocine (C), Adenine (A) and Guanine (G). A habitat use sequence could be coded as a triplet of consecutive habitat classes to account for temporal autocorrelation, e.g. for a sequence representing open (O) and closed (C) habitats **O*OC***_O_OC, triplets would be **OOC**, *OCO*, COO, OOC.

6.2. DISCUSSION OF ECOLOGICAL RESULTS

One of the main goals of this PhD was to explore the usefulness of our framework for the movement ecology community. We accomplished this through several case studies (Chapter 3 and Chapter 5). The main ecological results that reflect the general ecological aim of this dissertation are discussed in detail in Chapter 5. To summarise, we discovered that sequential habitat use patterns in roe deer were dependent on the population, on the individual, and on the environmental context in which animals move. At the population level habitat availability affected sequential use patterns, while at the individual level also habitat heterogeneity was shown to be relevant when explaining variability in sequential habitat use. However, the environmental context alone could not explain the identified sequential patterns. Within populations, where habitat accessibility is presumed to be similar for all individuals, we found variation in sequential habitat use. We also found clear temporal variability in the way animals sequentially used the habitats available in their home range throughout seasons. Finally, our mixed approach for data exploration and pattern recognition allowed us to investigate the dependence of sequential habitat use on life-history traits even futher. Preliminary results showed significantly more alternation for males compared to females during summer months, while males are territorial and females raise fawns (Liberg *et al.* 1998). Although preliminary, this direction of research is extremely promising, especially when considering the potential to associate life-history traits with individual personalities.

We applied and developed (IM-)SAM using GPS movement data of the European roe deer. This relatively small ungulate species is a challenging modelling species for several reasons. Specifically, roe deer show a high degree of behavioural plasticity, which manifests in a high heterogeneity in behavioural patterns (e.g. migrants/residents/facultative migrants: Peters *et al.* 2017) and in habitat occupancy (e.g. field/forest roe deer, Cederlund *et al*. 1991). Moreover, roe deer is an ecotonal species, and thus selects for edge habitats, small forest patches and transitional woodland habitat (Aulak & Babińska-Werka 1990, Tufto *et al.* 1996). By consequence, the link between roe deer movement data and habitat layers is sensitive to misclassification (Frair *et al.* 2010), which is especially a problem in large-scaled multi-population studies, or when high-resolution local habitat maps are not available. For instance, in the first case study (Chapter 3) we used Corine Land Cover 2006 (CLC) to identify open and closed habitats. Several small patches used in the high-elevation pastures were not detected by CLC (see also Pekkarinen *et al.* 2009). Hence we identified a relatively high proportion of homogeneous open sequences less frequently for the same population (IT1) in Chapter 4 & 5, where instead we used High-Resolution-Layer Tree Cover Density 2012 (TCD, EEA 2012) as the spatial layer for forest cover. Additional uncertainties may arise if there is a mismatch between the spatial resolution of raster layers and GPS errors, which may result in commission and omission errors (Frair *et al.* 2004, Frair *et al.* 2010).

Throughout this work, thanks to the recent availability of the High-Resolution-Layer Tree Cover Density 2012, we could perform a detailed and standardised habitat use analysis for roe deer at the European-scale and thus minimise such misclassification issues. Since this analysis, to our knowledge, was the first application of its kind, we performed a local raster validation analyses that is summarised in Chapter 4 – Appendix 4B. While local accuracy was high, we suggest to always perform additional visual and field-based checks before use. For this dissertation we only looked into sequential patterns of two simplified habitat classes: open and closed habitats. Habitat classifications are the result of a human simplification of the landscape, and key resources that drive animal habitat use/selection often remain hidden from the observer (Getz & Saltz, 2008). We suggest that the variability of sequential patterns over the year and across populations is the result of different types of resources emerging in open and closed habitats along the European gradient of landscapes and climate.

While here we only looked into two habitat classes, in future studies I would also like to take into account more detailed classes, different thresholds for forest cover, or the continuous forest cover value (see for example the study case 2 in Appendix 6A). For instance, through our work we observed that daily habitat alternation does not only occur between open and forest habitats, but also between forests with different densities or different open habitat types such as pastures and agricultural fields. While we observed mainly homogeneous use of open habitats in a population in Southcentral Norway, we found that animals actually do show alternating patterns, but at lower forest cover densities than defined by our threshold. This suggests that roe deer compensate their need for cover within a given habitat with different types of 'functional cover' (Mysterud & Østbye 1999). For example, in the population in Southwest France, animals in open habitats searched for cover in micro-selection of hedgerows (i.e. habitat *complementation*, see Morellet *et al.* 2011). At this moment, for large-scaled studies, environmental layers do not disentangle such high-resolution patterns, but ongoing progress in remote sensing promise a bright future for movement ecologists (see below). In addition, habitat types with exactly the same resources, may not necessarily have the same value. For instance, in proximity of sources of disturbance, such as roads or houses, a habitat could be less attractive. By combining multiple habitat variables, extracted from high resolution raster layers or maps, future work could explore how to get to a more refined habitat classification as input for (IM-)SAM.

In general, though, while higher resolution and better classified input environmental data would be important, I also see (IM-)SAM as a class of tools that help produce a synthesis of the complex relation between animals and their environment. In this sense, the 'Big Data' era in animal ecology (see Technical Prospects) offers on the one side huge opportunities, but on the other side challenges or even an increase in the *apparent* complexity of reality. Statistical and mathematical models (van Moorter et al. 2016), and synthesising methodological approaches as (IM-)SAM represent a response to such challenges.

A last important issue that I will cover here concerns missing data. GPS sensor performance is known to be affected by the animals internal state (e.g. resting), habitat characteristics (e.g. topography, cover density) and sensor type (Frair *et al*. 2010). Through (IM-)SAM exploration we discovered a high synchronicity between the use of closed habitats and missing observations, especially for older sensor types (see Chapter 4 - Appendix 4A). Hence, excluding such missing observations would impose habitat bias by underestimating the use of closed habitats. We highlighted the importance of accounting for missing data in habitat analysis (see Frair *et al.* 2010, Dokter *et al.* 2017), also for what concerns sequential habitat use. Several data imputation procedures have been evaluated, such as the models of the R package Amelia (Honaker *et al.* 2015), which allow to incorporate temporal correlation patterns. However, because our data did not fulfil the assumptions, another procedure was used based on a literature study on sensor performance (See Chapter 4 - Appendix 4A).

6.3. TECHNICAL PROSPECTS

6.3.1. (IM-)SAM and environmental data sources

I expect that future data collection will allow to tease better apart different habitat types. Current progress in Airborne Laser Scanning (ALS) technology (Light Detection And Ranging, LiDAR) has showed improvements for habitat identification at the local scale, with an increased detail in differentiating between low stands, old forest, young forest and shrub vegetation (Lone *et al.* 2014, Lone *et al.* 2017, Dupke *et al.* 2017, Ciuti *et al.* 2018). LiDAR products are becoming cheaper, better, are often publicly available and data processing over multiple study sites could be harmonised. In long-term, LiDAR plots in combination with satellite data and field validation plots from different eco-geographical areas may result in very detailed large-scale forest cover models. Global efforts have been made, in database development including 400 000 forest validation plots (Liang *et al.* 2016) and networks such as Globals (Global ALS Data Providers Database) aiming to establish a worldwide database of potential Airborne Laser Scanner data providers. Also Satellite data, such as the Landsat satellite archives (www.landsat.usgs.gov) and derived products from Copernicus Sentinel satellites (e.g. Sentinel 1, 2, 3; NASA-ESA, Copernicus, www.esa.int/Our_Activities/Observing_the_Earth) have a lot of potential. For instance, Landsat habitat variables could be generated at lower spatial resolution then lidar plots, but can be generated for large geographic regions and dynamically over a long time period (Duro *et al.* 2007). Although often pricy, cloud-based geospatial databases such as Google Earth Engine (www.earthengine.google.com), Mapbox (www.mapbox.com), MangoMap (www.mangomap.com) and CartoDB (www.carto.com), can facilitate analysis of such huge spatio-temporal datasets. An example of a widely used end-product developed with Google Earth Engine is the global forest cover dataset (Hansen *et al.* 2013), providing several freely available global raster products at 30 m spatial resolution: a forest cover density map and several forest gain and loss maps. A practical example is the global biodiversity mapping initiative, map of life (mol, www.mol.org/), that are currently developing high-resolution habitat suitability maps for multiple species, combining high-resolution habitat information available through Google Earth Engine (www.species.mol.org). Also crowd-funding initiatives such as openstreetmaps (www.openstreetmap.org, © OpenStreetMap contributors) are a precious resource for mapping different habitat variables, such as road density, forest cover, urban context. Openstreetmaps layers can be directly downloaded in a postgresql data base, or into R using R packages osmar (Eugster & Schlesinger 2013) and OpenStreetMap (Fellows & Stotz 2013).

6.3.2. (IM-)SAM extended applications

SAM and IM-SAM were developed for the discovery of similar patterns in habitat use sequences but have a large potential for studying other aspects related to the internal and external state of animals (e.g. movement parameters, behavioural states, animal interactions, physiological parameters, see Chapter 5 – Discussion) and could also be related to human movement data (see Chapter 4 – Discussion; see also Brum-Bastos *et al.* 2018). Moreover it can be extended to more complex classifications, to different temporal resolutions, and it can be used to explore both within and between population variations at multiple spatio-temporal scales. To exemplify (IM-)SAM's flexibility I developed four case studies (Appendix 6A) using different input variables and different temporal resolutions. Specifically, in case study 1 I look how daily sequential movement rates (m/s) measured at hourly resolution of one female roe deer cluster throughout the year. In case study 2 I show the clustering result when using continuous forest cover values as input instead of discretised classes. In case study 3 I explore yearly sequences of the use of summer and winter ranges at a daily resolution, based on a literature based simulated study. Finally, in case study 4 I highlight the potential of using the output of analytical animal movement R packages as input of (IM-)SAM.

6.3.3. (IM-)SAM technical developments

6.3.3.1. R Package

To improve applicability and transferability of (IM-)SAM methodology and procedure I aim for the future to develop a dedicated R package. This will increase accessibility of our framework and provide guidelines and case studies for customisation to specific research questions. Once this package is available, it could be used in combination with a suite of R packages that are currently available for computing several components of an animals' internal or external state for which the sequential pattern could be of relevance. For instance, adehabitatLT (Calenge 2006) and move (Kranstauber *et al.* 2017) could be used for calculating movement parameters, BCPA (behavioural Change Point Analysis, Gurarie *et al.* 2009, 2016) to identify different behavioural states, move (Kranstauber *et al.* 2017), adehabitatHR (Calenge 2006), BBMM (Nielson *et al.* 2013), tlocoh (Lyons *et al.* 2013), wildlifeTG (Long & Nelson 2015, Nelson *et al.* 2015, Long 2016) and stampr (Robertson *et al.* 2007; Long *et al.* 2018) for computing spatial objects related to home ranges and utilisation distributions, WildlifeDI (Long 2014) for computing dynamic interactions between animals, adehabitatHS (Calenge 2006) and Hab (Basille 2014) for exploring habitat variables, recurse (Bracis *et al.* 2018) for computing revisitation metric, RNCEP (Kemp *et al.* 2012) for extraction of global environmental data, rpostgis (Basille & Bucklin 2017) for extraction and import of both raster and vector data from and to a spatial data base, RPostgreSQL (Conway *et al.* 2013) for import and export of PostgreSQL database tables and rgrass7 (Bivand 2015) for advanced spatial computations using functionality of GRASS7.

Currently, several R functions have been developed, and are available via github (https://github.com/jedgroev/PhD-project), but further standardisations and tests of the scripts are needed before users can apply the analytical framework appropriately. For instance, I would like to implement a function for automatic assessment of dissimilarity weights for spatially correlation of habitat variables.

6.3.3.2. R Shiny

Currently, I am developing a R Shiny application. R shiny is a R package that allows to translate static into dynamic products delivered via simple web-based graphic-user interface. R shiny is gaining in popularity, especially in the United States of America where apps are even developed by newspapers to present interactive maps, tables and graphs. I developed a prototype of an interactive tree-based application, to explore, identify and visualise ecologically relevant and similar sequential patterns in habitat use, or other ecological, behavioural or movement variables. In the prototype app, several parameters can be set interactively: temporal resolution and range of the sequences, number and classification of habitat use classes and the environmental variable to visualise. In the first version of the app I aim to publish the data exploration step of (IM-)SAM. However, in the future, I also aim to include all steps of the proposed (IM-)SAM procedure (i.e. exploration, simulation, classification) in a R shiny app, that should make (IM-)SAM procedure more known and accessible to the community.

6.3.3.3. (IM-)SAM effectiveness

As described above, the (IM-)SAM has a large potential for visualising and analysing any type of sequential pattern (i.e. case studies) for which functions and tools will be made available through the baseline statistical software for ecologists R (i.e. R package and R Shiny app). While here we have showed the effectiveness of (IM-)SAM in visualising sequential patterns through various case studies (Chapter 3 & 4, Appendix 6A), in future work I wish to evaluate the prospective toolbox in a more coherent way. Through experiments, expert users (Andrienko *et al.* 2011; Konzach *et al.* 2018), in-depth user-experience questionnaires (Konzach *et al.* 2018) and eye-tracking analysis (Fu 2016; Keskin *et al.* 2017) one could evaluate users' cognitive processes in learning, acquiring and remembering information presented in the trees. The visual effectiveness and informativeness of (IM-)SAM could be evaluated in two consecutive steps.

In a first phase static trees could be provided for which sequence length and temporal resolution are *a priori* set. After basic instructions about the tool we could evaluate with eye tracking, or mouse hoover technology, in which sequence and how much time is spend to explore different parts of the visualisation. Users might first and more intensively explore the actual tree and dissimilarity between the branches, or the sequences and associated covariates, or the optional map that includes the animal trajectory. Note that when designing such an experiment, instructions should minimise bias in the sequence with which different parts are observed by users. After a first exploration of the visualisation specific analytical questions could be asked on the exploratory trees (see table 6.1 for some example questions). Given the interdisciplinary nature of the tool, such a user experiment could be performed with both biology and geography students. In a second phase a similar stepwise experiment could be performed for the interactive R Shiny application. The experiment could be performed for users that participated to the previous experiment and inexperienced users.

Table 6.1. Potential questions for evaluation of tool effectiveness

Q1	How many clusters can you identify in the dissimilarity tree?
Q2	Is there a hierarchy in the clusters?
Q3	How do sequential patterns in different clusters differentiate from each other?
Q4	Is there a spatial association between observed patterns?
Q5	Is there a relation between the clusters and associated sequences?

Another important aspect on the effectiveness and usability of a tool is its computation time. In Appendix 6B I provide a detailed report on the computation time of two essential methodological steps (i.e. IM-SAM tree building, bootstrapping), using generated random sequences by varying sequence length, number of sequences and number of classes. Results showed that there is a linear trend with increasing sequence complexity. The median computation time using 20 replicates ranged between 1.6 seconds (sequence length, 25; number of classes, 2) and 67.7 seconds (sequence length, 1000; number of classes, 25). Bootstrapping running time also showed a linear increase and were mainly influenced by the calculation of sequence dissimilarity component. The time to perform the bootstrapping itself does not change with an increasing sequence complexity. Another part for which evaluation of computation time is important are the movement models, but this was not tested systematically. As a rule of thumb, computation time for 900 simulated sequences was approximately half an hour.

REFERENCES

Andrienko, G., Andrienko, N. and Heurich, M., 2011. An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, **25** (9), 1347-1370.

Aulak, W. and Babińska-Werka, J., 1990. Use of agricultural habitats by roe deer inhabiting a small forest area. *Acta Theriologica*, **35** (1-2), 121-127.

Basille, M., 2014. hab: habitat and movement functions. – *R package version 1.19.2.*

Basille, M. and Bucklin, D., 2017. rpostgis: R Interface to a 'PostGIS' Database. https://CRAN.R-project.org/package=rpostgis

Bivand, R., 2015. rgrass7: Interface Between GRASS 7 Geographical Information System and R. *R package version 0.1-3.* https://CRAN.R-project.org/package=rgrass7

Bookstein, A., Klein, S.T. and Raita, T., 2001. Fuzzy Hamming distance: a new dissimilarity measure. *In:* Amir, A., ed. *Combinatorial Pattern Matching. CPM 2001. Lecture Notes in Computer Science*, 2089, Berlin: Springer, 86-97.

Bracis, C., Bildstein, K.L. and Mueller, T., 2018. Revisitation analysis uncovers spatio-temporal patterns in animal movement data. *Ecography*.

Brum-Bastos, V.S., Long, J.A. and Demšar, U., 2018. Weather effects on human mobility: a study using multi-channel sequence analysis. *Computers, Environment and Urban Systems*.

Cagnacci, F., Focardi, S., Heurich, M., Stache, A., Hewison, A.J.M., Morellet, N., Kjellander, P., Linnell, J.D., Mysterud, A., Neteler, M. and Delucchi, L., 2011. Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors. *Oikos*, **120** (12), 1790-1802.

Ciuti, S., Tripke, H., Antkowiak, P., Gonzalez, R.S., Dormann, C.F. and Heurich, M., 2018. An efficient method to exploit LiDAR data in animal ecology. *Methods in Ecology and Evolution*. **9** (4): 893-904.

Conway, J., Eddelbuettel, D., Nishiyama, T., Prayaga, S.K. and Tiffin, N., 2013. RPostgreSQL: R interface to the PostgreSQL database system. *R package version 0.4.* https://CRAN.R-project.org/package=RPostgreSQL

Dokter, A.M., Loon, E.E., Fokkema, W., Lameris, T.K., Nolet, B.A. and Jeugd, H.P., 2017. Analyzing time-ordered event data with missed observations. *Ecology and evolution*, **7** (18), 7362-7369.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F. and Douzery, E.J., 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Molecular Biology and Evolution*, **20** (2), 248-254.

Dupke, C., Bonenfant, C., Reineking, B., Hable, R., Zeppenfeld, T., Ewald, M. and Heurich, M., 2017. Habitat selection by a large herbivore at multiple spatial and temporal scales is primarily governed by food resources. *Ecography*, **40** (8), 1014-1027.

Duro, D.C., Coops, N.C., Wulder, M.A. and Han, T., 2007. Development of a large area biodiversity monitoring system driven by remote sensing. *Progress in Physical Geography*, **31** (3), 235-260.

Eugster, M.J.A. and Schlesinger, T., 2013. osmar: OpenStreetMap and R. *The R Journal*, **5** (1), 53-63. https://cran.r-project.org/web/packages/osmar

Fellows, I. and Stotz, J.P., 2013. OpenStreetMap: access to open street map raster images. *R package version 0.3*, **1**. https://CRAN.R-project.org/package=OpenStreetMap

Fu, J., 2016. Usability evaluation of software store based on eye-tracking technology. In *Information Technology, Networking, Electronic and Automation Control Conference, IEEE,* 1116-1119.

Frair, J.L., Nielsen, S.E., Merrill, E.H., Lele, S.R., Boyce, M.S., Munro, R.H., Stenhouse, G.B. and Beyer, H.L., 2004. Removing GPS collar bias in habitat selection studies. *Journal of Applied Ecology*, **41** (2), 201-212.

Frair, J.L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N.J. and Pedrotti, L., 2010. Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **365** (1550), 2187-2200.

Getz, W.M. and Saltz, D., 2008. A framework for generating and analyzing movement paths on ecological landscapes. *Proceedings of the National Academy of Sciences*, **105** (49), 19066-19071.

Giorgino, T., 2009. Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software*, **31** (7), 1-24.

Gurarie, E., Andrews, R. and Laidre, K., 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. and 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.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R. and Kommareddy, A., 2013. High-resolution global maps of 21st-century forest cover change. *Science*, **342** (6160), 850-853.

Honaker, J., King, G. and Blackwell, M., 2015. Amelia: A Program for Missing Data. *R package version 1.7.4.* https://CRAN.R-project.org/package=Amelia

Kemp, M.U., Emiel van Loon, E., Shamoun-Baranes, J. and Bouten, W., 2012. RNCEP: global weather and climate data at your fingertips. *Methods in Ecology and Evolution*, **3** (1), 65-70.

Keskin, M., Ooms, K., Dogru, A.O. and De Maeyer, P., 2018. Digital sketch maps and eye tracking statistics as instruments to obtain insights into spatial cognition. *Journal of Eye Movement Research*, **11** (3), 4.

Konzack, M., Gijsbers, P., Timmers, F., van Loon, E., Westenberg, M.A. and Buchin, K., 2018. Visual exploration of migration patterns in gull data. *Information Visualization*, 1473871617751245.

Kranstauber, B., Smolla, M. and Scharf, A.K. 2017. move: Visualizing and Analyzing Animal Track Data. *R package version 3.0.2.* https://CRAN.R-project.org/package=move

Liang, J., Crowther, T.W., Picard, N., Wiser, S., Zhou, M., Alberti, G., Schulze, E.D., McGuire, A.D., Bozzato, F., Pretzsch, H. and De-Miguel, S., 2016. Positive biodiversity-productivity relationship predominant in global forests. *Science*, **354** (6309), aaf8957.

Liberg, O., Johansson, A., Andersen, R. and Linnell, J.D.C., 1998. Mating system, mating tactics and the function of male territoriality in roe deer. *In:* Andersen, R., Duncan, P and Linnell, J.D.C., eds. *The European roe deer: the biology of success.* Oslo: Scandinavian University Press, 221-256.

Lone, K., Loe, L.E., Gobakken, T., Linnell, J.D., Odden, J., Remmen, J. and Mysterud, A., 2014. Living and dying in a multi-predator landscape of fear: roe deer are squeezed by contrasting pattern of predation risk imposed by lynx and humans. *Oikos*, **123** (6), 641-651.

Lone, K., Mysterud, A., Gobakken, T., Odden, J., Linnell, J. and Loe, L.E., 2017. Temporal variation in habitat selection breaks the catch-22 of spatially contrasting predation risk from multiple predators. *Oikos*, **126** (5), 624-632.

Long, J.A., 2014. wildlifeDI: Calculate Indices of Dynamic Interaction for Wildlife Telemetry Data. *R package version 0.2.* https://CRAN.R-project.org/package=wildlifeDI

Long, J.A. and Nelson, T.A., 2015. Home range and habitat analysis using dynamic time geography. *Journal of Wildlife Management,* **79** (3), 481-490.

Long, J.A., 2016. A field-based time geography for wildlife movement analysis. *International Conference on GIScience Short Paper Proceedings*, 1 (1).

Long, J., Robertson, C. and Nelson, T. 2018. stampr: Spatial-Temporal Analysis of Moving Polygons in R. *Journal of Statistical Software*, **84** (1), 1-19.

Lyons, A.J., Turner, W.C. and Getz, W.M., 2013. Home range plus: a space-time characterization of movement over real landscapes. *Movement Ecology,* **1** (1), 2.

Morellet, N., Van Moorter, B., Cargnelutti, B., Angibault, J.M., Lourtet, B., Merlet, J., Ladet, S. and Hewison, A.J.M., 2011. Landscape composition influences roe deer habitat selection at both home range and landscape scales. *Landscape Ecology*, **26** (7), 999-1010.

Müller, M., 2007. Dynamic Time Warping. *In:* Müller, M.*, Information retrieval for music and motion.* Berlin: Springer, 69-84.

Mysterud, A. and Østbye, E., 1999. Cover as a habitat element for temperate ungulates: effects on habitat selection and demography. *Wildlife Society Bulletin (1973-2006)*, **27** (2), 385-394.

Nelson, T., Long, J., Laberee, K. and Stewart, B., 2015. A time geographic approach for delineating areas of sustained wildlife use. *Annals of GIS,* **21** (1), 81-90.

Nielson, R.M., Sawyer, H. and McDonald, T.L., 2013. BBMM: Brownian bridge movement model. *R package version 3.0.* https://CRAN.R-project.org/package=BBMM

Nilsen, E.B., Pedersen, S. and Linnell, J.D. 2008. Can minimum convex polygon home ranges be used to draw biologically meaningful conclusions?. *Ecological Research*, **23** (3), 635-639.

Pekkarinen, A., Reithmaier, L. and Strobl, P., 2009. Pan-European forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **64** (2), 171-183.

Peters, W., Hebblewhite, M., Mysterud, A., Spitz, D., Focardi, S., Urbano, F., Morellet, N., Heurich, M., Kjellander, P., Linnell, J.D.C. and Cagnacci, F., 2017. Migration in geographic and ecological space by a large herbivore. *Ecological Monographs*, **87** (2), 297-320.

Robertson, C., Nelson, T.A., Boots, B. and Wulder, M.A. 2007, STAMP: Spatial-temporal analysis of moving polygons. *Journal of Geographical Systems*, **9**, 207-227.

Riotte-Lambert, L., Benhamou, S. and Chamaillé-Jammes, S., 2016. From randomness to traplining: a framework for the study of routine movement behavior. *Behavioral Ecology*, **28** (1), 280-287.

Studer, M., Ritschard, G., Gabadinho, A. and Müller, N.S., 2010. Discrepancy Analysis of Complex Objects Using Dissimilarities. *In:* Guillet, F., Ritschard, G., Zighed, D.A. and Briand, H., eds. *Advances in Knowledge Discovery and Management. Studies in Computational Intelligence*, **292.** Berlin: Springer, 3-19.

Tufto, J., Andersen, R. and Linnell, J.D.C., 1996. Habitat use and ecological correlates of home range size in a small cervid: the roe deer. *Journal of Animal Ecology*, **65** (6), 715-724.

Van Moorter, B., Rolandsen, C.M., Basille, M., & Gaillard, J.-M. 2016. Movement is the glue connecting home ranges and habitat selection. *Journal of Animal Ecology*, **85** (1), 21-31.

APPENDIX 6A – CASE STUDIES

In all case studies I applied a simplified version of the SAM algorithm. Specifically, I only built exploratory trees based on hamming distance, with automatic assessment of the substitution weight matrix (i.e. TRATE-method), or with a user-defined weight matrix that imposes a linear relationship between class dissimilarity (case study 2 and case study 4).

6A.1. Case study 1 – sequential movement rate (m/h)

For this case study I extracted GPS-locations with a fixed one hour relocation interval for a monitoring period of one year (365 days) of one female roe deer (Animal 2054) from Southern Germany (DE15), and computed the hourly movement rate in meters per hour (m/h). Thus, each sequence represents the daily sequential hourly movement rates. In a first step movement rates were reclassified into seven classes, using the k-means discretisation technique, applied to the distribution of *all* movement rates (Fig.6A.1). These class boundaries were used to reclassify the movement rate sequences. Next exploratory dissimilarity trees were built including corresponding sequences, and the seasons were added as coloured bars (Fig.6A.2). Although a very rough version of the algorithm was used, two main clusters could still be distinguished, and showed a seasonal clustering, related to variations in sunrise and sunset throughout the year: an upper cluster, corresponding to spring, summer and autumn, showing highest movement rates at 5-6h UTC and 19-20h UTC, a lower cluster, corresponding to autumn and winter, showing highest movement rates from 7-8h UTC and 17-18h UTC. In both clusters movement rates were generally higher during dusk and dawn (41-306 m/s), and lowest during daytime (<41 m/s). During night (41-187 m/s) movement rates seem to be higher than during daytime. Interestingly, visually also a small cluster could be detected, including sequences from 11 to 17th of May, which show very little variation throughout the day, likely corresponding to the period of parturition (Plard *et al*. 2012).

While patterns are clearly distinguished, and seasonal patterns are observed, application of the bootstrap algorithm did not show a significant difference between the clusters. Specifically, BJ_{MD} for a cut-off at 2 clusters was 0.54, while values below 0.60 indicate a not-well identified clusters. Increasing the number of clusters results in a decrease in BJ_{MD} and increase in BJ_{IQMD}. This low cluster robustness is likely because daily sequences only consist out of 24 consecutive states. Hence, limited changes in variability could potentially affect the clustering of sequences. I expect the clustering to improve when movement rates are computed instead at half hour resolution, which would result in sequences consisting out of 48 consecutive states.

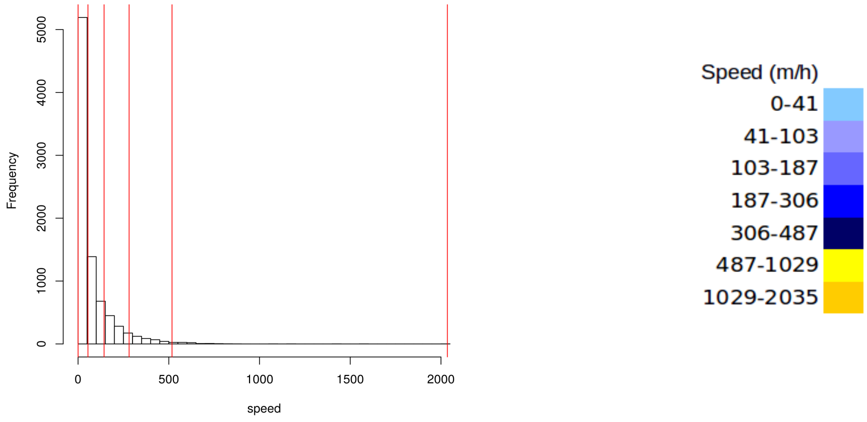


Figure 6A.1. Distribution of movement rates for animal 2054 and identified class-thresholds. colour-codes are used in the dissimilarity tree below.

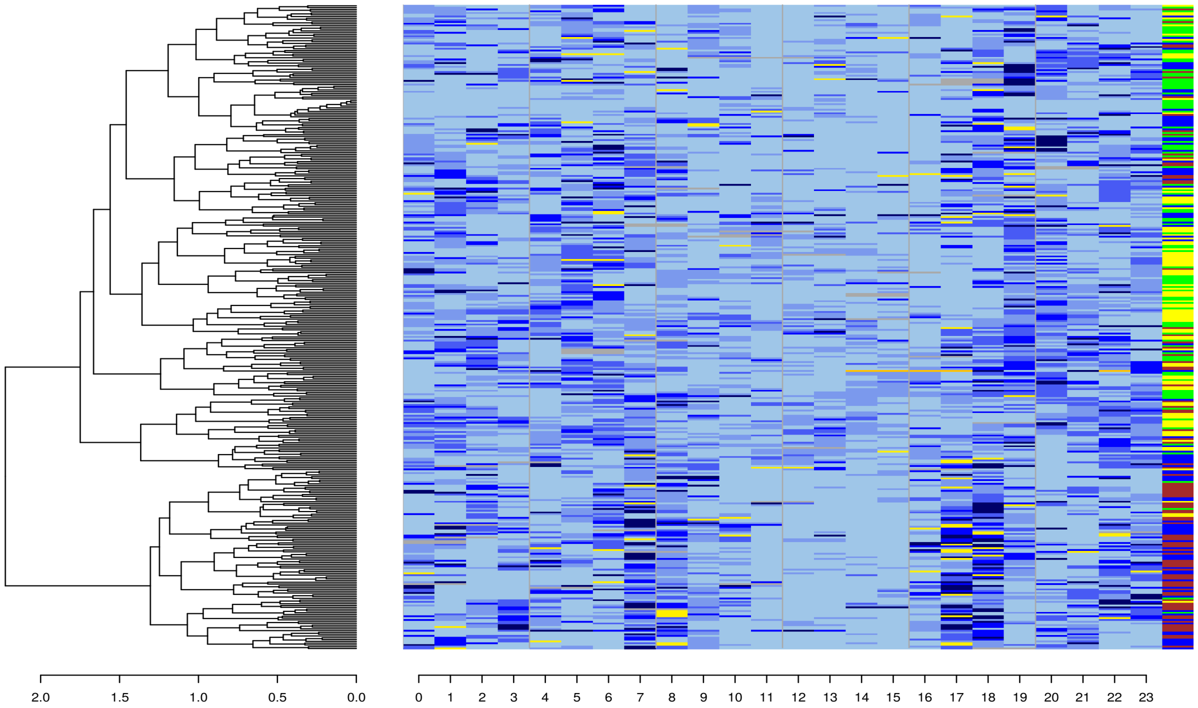


Figure 6A.2. Dissimilarity tree of daily movement rate (m/s) sequences, corresponding sequences (gradient of blue, give the five most common movement rate classes; yellow and orange, are very high movement rates). coloured bar on the right highlights the season of each sequence (winter, blue; spring, green; summer, yellow; autumn, brown).

6A.2. Case study 2 – Sequential use of forest cover density (0-100%)

For animal 2054 (see case study 1), the same daily GPS trajectories were intersected with the High Resolution Layer-Tree Cover Density 2012 (EEA 2012). Instead of reclassifying forest cover density into two distinct classes (as in Chapter 4 and 5) the original forest cover percentages were used. While these values were continuous, the simplified algorithm seemed to separate well different behaviours. The dissimilarity tree distinguished three clear clusters, that also show seasonal clustering in the sequential use of open and forest habitats (Fig.6A.3, coloured bars). The upper cluster, including sequences from late autumn to early spring, shows several daily habitat alternation patterns. Interestingly, the switch from forest to open habitat during the evening is sharper than the one during the morning. Specifically, in the morning this roe deer switched from open to forest habitat between 4h and 10h UTC, while the switch in the evening is more consistently around 16h or 17h UTC. The second cluster, also mainly during winter, shows a more continuous use of forest habitats from 0h to 16h or 17h UTC. The third cluster, includes most sequences from spring, summer and autumn and represents mainly sequences with a continuous open habitat use. Here the BJ algorithm showed a significant separation for two (BJ_{MD} = 0.98) and for three (BJ_{MD} = 0.84) clusters. However, if the cut-off is set to three clusters, the smallest one has a relatively low bootmean value (0.66).

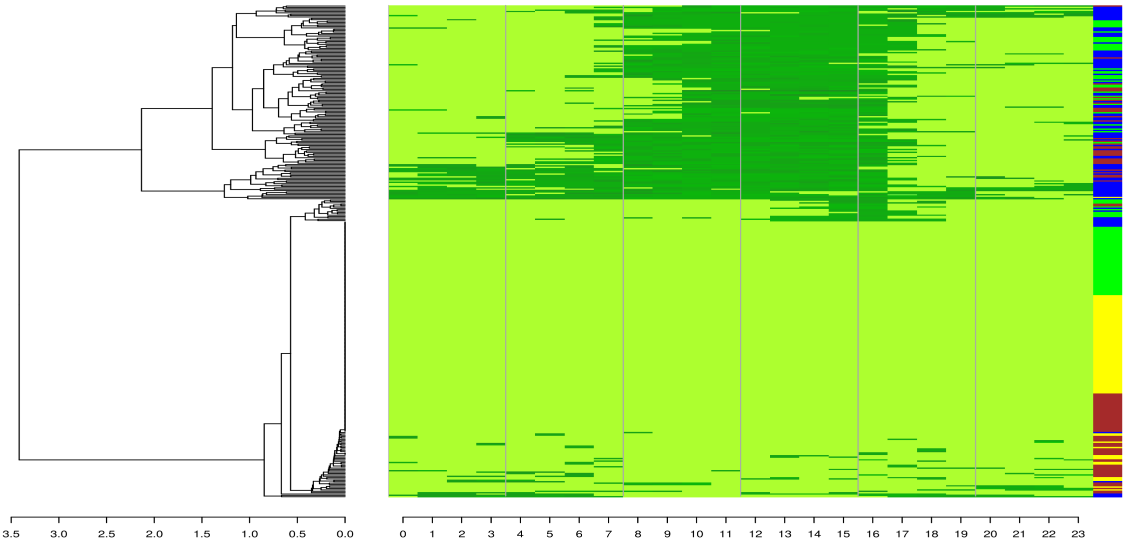


Figure 6A.3. Dissimilarity tree for forest cover density (0-100%) sequences, corresponding sequences (gradient of green, where darker greens indicate higher forest cover densities). coloured bar on the right highlights the season of each sequence (winter, blue; spring, green; summer, yellow; autumn, brown).

6A.3. Case study 3 – Simulated sequential daily use of summer and winter ranges

Partial Migration is typically observed for roe deer, meaning that migration occurs according to a gradient of behaviours (Cagnacci *et al.* 2011). Some individuals migrate in one instance, others switch several times between summer and winter range, and others remain resident in the same region over the whole year (i.e. residents). The onset of spring migration (winter to summer range) is more synchronous over different study sites in Europe (15 April - 15 May), most likely triggered by plant phenology and thus availability of highly nutritious plants during early spring. The onset of autumn migration (1 November – 15 February) is more variable, but one of the important explanatory variables is snow cover (Mysterud *et al.* 1999).

With SAM the variation on migration behaviour could be explored, within and between populations, by coding daily winter and summer range use over a year (i.e. sequences consisting of 365 characters). The winter range could be used as the reference range. Using the results from Cagnacci *et al.* (2011), a simplified simulation was developed to explore sequential use of summer and winter ranges. Specifically, summer and winter range use were simulated using the mean and standard deviation of the start of both summer and winter migration for the four populations included in Cagnacci *et al.* (2011): Norway, the Bavarian Forest, the Italian Alps and Southwest France (Table 6A.1). Per population we simulated 100 yearly sequences with a daily temporal resolution (365 states). Start dates of summer and winter migration were simulated assuming a normal distribution within each population. A dissimilarity tree shows the clustering of the simulated sequences, using the hamming distance and hierarchical clustering method of Ward (Fig.6A.4). Jaccard Bootstrapping procedure showed a significant cluster separation, mainly for a cut-off at two clusters. At this cut-off, the population of southwest France (yellow), having a late mean winter migration (see table 6A.1), was clearly distinguished as a separate cluster. Other populations, which show more similar start dates of winter/summer migration group together in one cluster. However, the sequences from Bavarian Forest and Norway are more mixed, then those of the Italian Alps.

Table 6A.1. Input for simulation data based on Cagnacci *et al.* (2011), giving the mean and standard deviation of the start day for summer and winter migration.

Population	**mean**		**Standard deviation (days)**		
	Summer M	**Winter M**	**Summer M**	**Winter M**	
Bavarian Forest	15 May	1 Dec	30	30	🟩
Italian Alps	1 May	1 Dec	10	15	▨
Southwest France	15 May	15 Feb	45	45	🟨
Norway	10 May	1 Nov	30	30	⬛

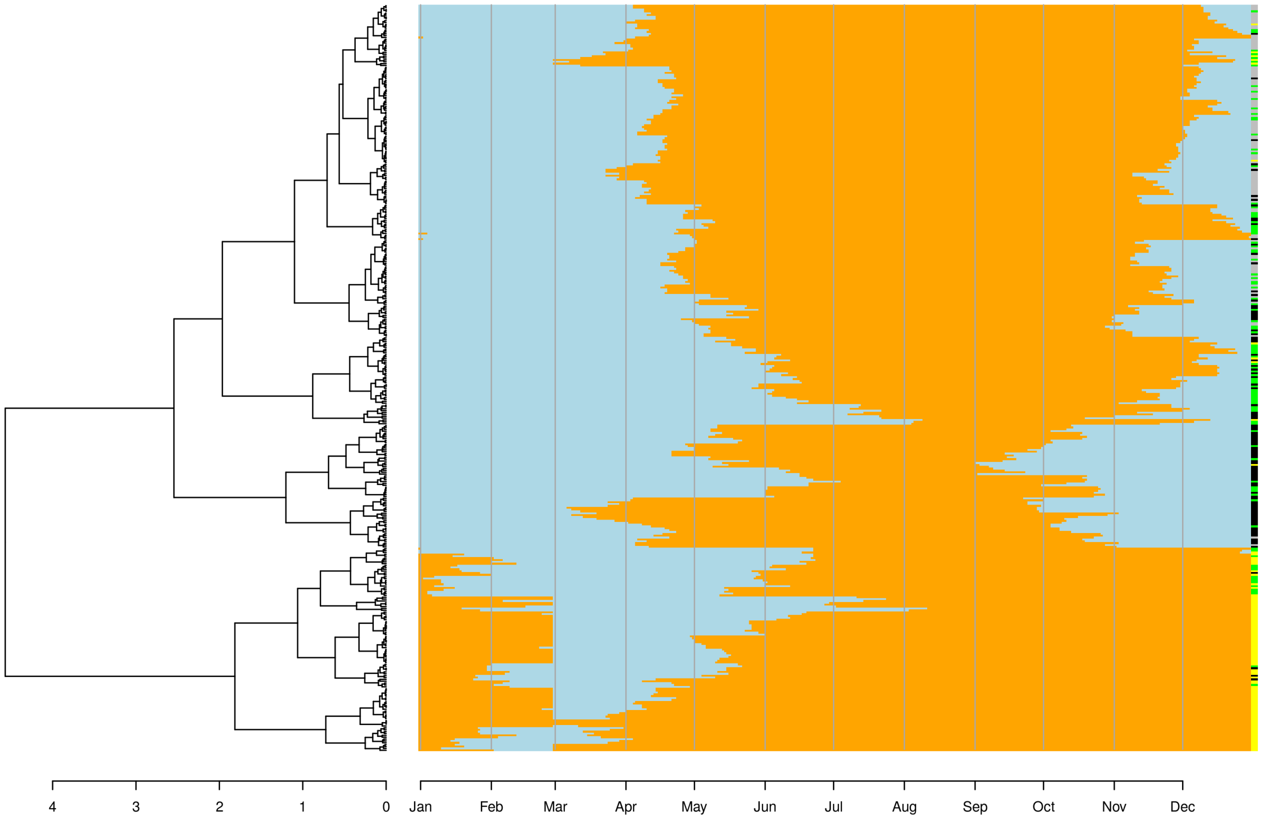


Figure 6A.4. Dissimilarity tree of the different simulated use of yearly summer and winter ranges (light blue and orange are respectively the winter and summer range). The corresponding population of each simulated sequence is highlighted by the coloured bar on the right (Bavarian Forest, green; the Italian Alps, grey; southwest France, yellow; Norway, black).

6A.4. Case study 4 – Sequential recursive use

This example shows that input data calculated with other R packages could be used to explore spatio-temporal patterns. Here recursive use of locations was computed using the R package recurse (Bracis *et al.* 2018) for the sample dataset 'Martin' (species not mentioned). The visualisation of a sequence dissimilarity tree can help to understand whether places of intensive use are repetitively used at similar timestamps or for consecutive days (Fig.6A.5). The sequential plot does not immediately show a clear daily pattern in the recursive use. The most significant cut-off is detected at two clusters, corresponding to a BJ_{MD} of 0.69.

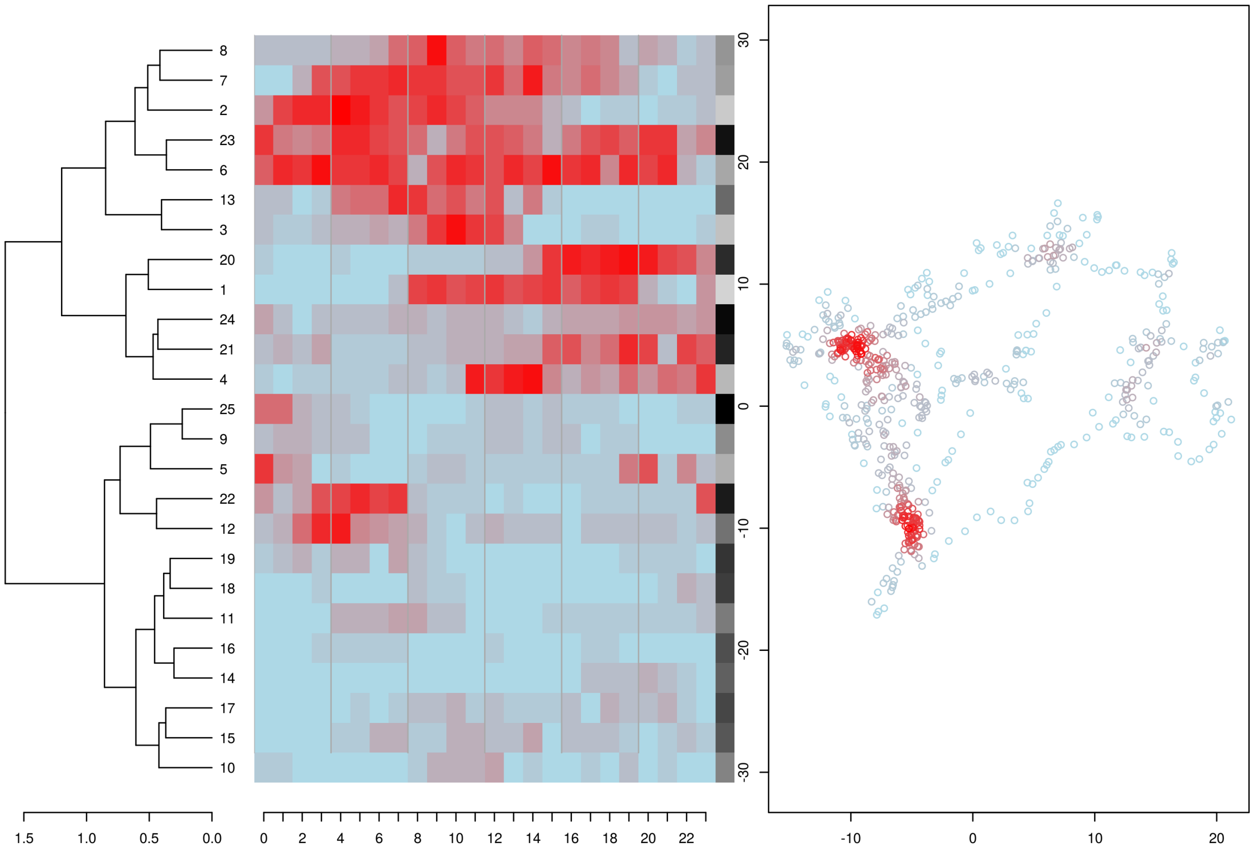


Figure 6A.5. Dissimilarity tree for recursive use sequences, corresponding sequences (from blue to red, 1 to 17 visits) and spatial points. coloured bar on the right highlights the day of each sequence (grey to black).

REFERENCES

Bracis, C., Bildstein, K.L. and Mueller, T., 2018. Revisitation analysis uncovers spatio-temporal patterns in animal movement data. *Ecography*. **41**, 1-11.

Cagnacci, F., Focardi, S., Heurich, M., Stache, A., Hewison, A.J.M., Morellet, N., Kjellander, P., Linnell, J.D.C., Mysterud, A., Neteler, M. and Delucchi, L., 2011. Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors. *Oikos*, **120** (12), 1790-1802.

Mysterud, A., 1999. Seasonal migration pattern and home range of roe deer (Capreolus capreolus) in an altitudinal gradient in southern Norway. *Journal of zoology*, **247** (4), 479-486.

Plard, F., Gaillard, J.M., Bonenfant, C., Hewison, A.J.M., Delorme, D., Cargnelutti, B., Kjellander, P., Nilsen, E.B. and Coulson, T., 2013. Parturition date for a given female is highly repeatable within five roe deer populations. *Biology letters*, **9** (1), 20120841.

APPENDIX 6B – RUNNING TIME

The (IM-)SAM procedure consists out of several different methodological steps for which computation time could be assessed. Using a set of generated random sequences (i.e. non-spatial explicit) I computed both the computation time for building (IM-)SAM trees and for the bootstrapping.

Considering (IM-)SAM trees (Fig.6B.1), the running time was computed for an increasing complexity of the dataset by varying the sequence length, the number of sequences and the number of classes. Specifically, the running time was calculated for seven different sequence lengths (i.e. 25, 50, 100, 250, 500, 750, 1000 states per sequence) using a constant sequence count (1000 sequences), and *vice versa* for seven different sequence counts (i.e. 25, 50, 100, 250, 500, 750, 1000 sequences) using a constant sequence length (1000 states per sequence), which was repeated for a different number of classes (i.e. 2 , 4 , 5, 10, 15, 20, 25 classes). For each parameter combination 20 replicates were run. Hence, in total computation time was calculated 980 times (7 sequence length x 7 number of classes x 20 replicates; 7 sequence counts x 7 number of classes x 20 replicates). Thus, the lowest and highest complexity of sequence length/sequence counts corresponds to 25 and 1000 states per sequence/sequence counts of 2 and 25 classes, respectively. To investigate how the computation time changes with complexity the median computation time was calculated for each parameter combination, which was used as input for linear regression models. A regression model was built using computation time as the response variable and sequence length/sequence count as a dependent variable, generated for each number of classes separately (Table 6B.1).

Computation time showed a strong linear trend with increasing complexity, where the median computation time ranges between 1.600 seconds (sequence length = 25; n classes = 2) and 67.656 seconds (sequence length = 1000; n classes = 25; Fig.6B.1, left), and between 1.209 seconds (n sequences = 25; n classes = 2) and 67.625 seconds (n sequences = 1000; n classes = 25; Fig.6B.1, middle). There is a linear increase in the computation time both when increasing the sequence length and sequence count (Fig.6B.1 & Table 6B.1 for summary of the regression models). The slope of the models is steeper when the number of classes included in generated sequences is higher. A regression model of the slopes indicates that the computation time also increases linearly with the number of classes (Fig.6B.1. right). All models had an R-square higher than 0.990.

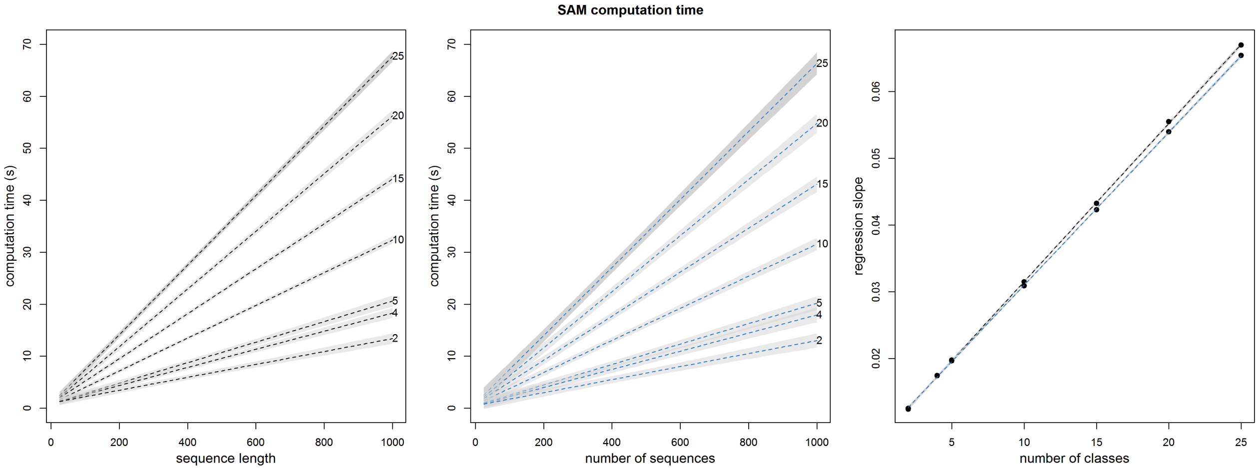


Figure 6B.1. Fitted lines of linear regression models of the median computation time of (IM-)SAM by varying the sequence length (left) and the number of sequences (middle) for a different number of habitat classes. To measure the effect of the number of classes on the running time, I extracted the slopes of each regression line shown in the former two plots.

Table 6B.1. Linear regression model outputs for (IM)-SAM trees

	classes	**Intercept**	**intercept.se**	**slope**	**slope.se**	**r.squared**	**p.value**
sequence length	2	0.968	0.286	0.012	0.001	0.990	0.000
	4	0.871	0.271	0.017	0.001	0.996	0.000
	5	0.851	0.298	0.020	0.001	0.996	0.000
	10	0.874	0.203	0.031	0.000	0.999	0.000
	15	0.878	0.232	0.043	0.000	0.999	0.000
	20	0.743	0.320	0.055	0.001	0.999	0.000
	25	0.729	0.300	0.067	0.001	1.000	0.000
sequence number	2	0.480	0.391	0.013	0.001	0.982	0.000
	4	0.493	0.405	0.017	0.001	0.990	0.000
	5	0.539	0.376	0.020	0.001	0.993	0.000
	10	0.692	0.349	0.031	0.001	0.998	0.000
	15	0.788	0.416	0.042	0.001	0.998	0.000
	20	0.833	0.521	0.054	0.001	0.998	0.000
	25	0.889	0.600	0.065	0.001	0.998	0.000

	Slope	**intercept**	**intercept.se**	**slope**	**slope.se**	**r.squared**	**p.value**
number of classes	seq. length	0.008	0.000	0.002	0.000	1.000	0.000
	seq. number	0.008	0.000	0.002	0.000	1.000	0.000

A similar analysis was performed for the bootstrapping procedure. Since similar trends were found for the sequence length and sequence counts, computations were limited by varying the sequence lengths only (i.e. 25, 100, 250, 500, 1000), computed for three classes (i.e. 2, 10, 15, 25). Also here 20 replicates were run for each parameter combination. In total computation time was thus calculated 400 times (5 sequence lengths x 4 number of classes x 20 replicates). For each parameter combination the median computation time was calculated and then used as input for linear regression models. As in Chapter 4 the number of bootstraps was set to 1000 and the cut-off from 2 to 15 clusters, meaning that the bootstrapping was repeated 14 times. Using these parameter settings, the median running time ranged between 6.84 minutes (sequence length = 25; n classes = 2) and 7.46 minutes (sequence length = 1000; n classes = 25), which corresponds to an absolute difference of maximum 37 seconds. Indeed, regression models show a negligible slope with increasing sequence length, and a slight increase in the intercept (Table 6B.2). To conclude, the complexity of the dataset has a minimal impact on the computation time of the bootstrapping. Instead, computation time is affected by the number of bootstraps and the maximum cut-off distance over which bootstraps are calculated.

Table 6B.2. Linear regression model outputs for bootstrapping

	number of classes	**intercept**	**intercept.se**	**Slope**	**slope.se**	**r.squared**	**p.value**
sequence length	25	7.056	0.035	0.000	0.000	0.931	0.008
	15	7.037	0.039	0.000	0.000	0.906	0.013
	10	7.035	0.037	0.000	0.000	0.908	0.012
	2	6.938	0.052	0.000	0.000	0.850	0.026

Curriculum Vitae Johannes De Groeve

Personal Information

Address	Terstraeten 6A, 9860 Balegem
Phone	+32 496 83 05 22, +39 327 075 67 04
E-mail	degroevejohannes@gmail.com, johannes.degroeve@ugent.be
Nationality	Belgian
Date and Place of Birth	5 September 1988, Ghent, Belgium
Gender	Male
Driver license	B

Education

2014 – 2018	**PhD in Geography** at Ghent University and Research and Innovation Centre, Edmund Mach Foundation (FEM/CRI)
2010 – 2012	**Master of Science in Geography** (magna cum laude) at Ghent University
2010 – 2012	**Preparatory Course to Master of Science in Geography** (magna cum laude) at Ghent University
2011 – 2012	**Erasmus Exchange Program** at University of Oslo (UiO)
2009 – 2010	**Master in Archaeology** (cum laude) at Ghent University
2006 – 2009	**Bachelor in Archaeology** (cum laude) at Ghent University

Internships

2016	**Internship** at Department of Organismic & Evolutionary Biology, Moorcroft Laboratory, Harvard University, MA, USA. (Apr.-June 2016)
2013	**STSM (Short-Term Scientific Mission) as researcher** at Research and Innovation Centre, Edmund Mach Foundation (FEM/CRI), San Michele all' Adige (TN), Italy (June-Oct. 2013)
	STSM (Short-Term Scientific Mission) as researcher at Research and Innovation Centre, Edmund Mach Foundation, San Michele all' Adige (TN), Italy (Feb.-Mar. 2013)
2006 – 2010	**Internship Archaeology** at multiple archaeological sites in Belgium

Work experience

2014-2018	**Database curator** at Euroungulate
2011	**GIS-specialist** at Vlaams Instituut voor Onroerend Erfgoed (V.I.O.E.)

Publications

Paper Publications

2016	De Groeve J, Van de Weghe N, Ranc N, Neutens T, Ometto L, Rota-Stabelli O and Cagnacci F (2016) *Extracting Spatio-Temporal Patterns In Animal Trajectories: An Ecological Application Of Sequence Analysis Methods (SAM)* Methods in Ecology and Evolution 7 (3): 369-379 DOI: 10.1111/2041-210X.12453 (IF: 6.344, Ranking: 8/149 (Ecology)).

Conference Publications

2017

De Groeve J, Van de Weghe N, Cagnacci F (2017) Studying a species at the distribution range scale – spatiotemporal habitat use patterns in roe deer, Invited speaker at **Open Science Tools, Data & Technologies for Efficient Ecological & Evolutionary Research Symposium**, 7-8 Dec, Amsterdam, The Netherlands.

De Groeve J, Van de Weghe N, Cagnacci F (2017) *Poster: Sequential Habitat Use by Animals: A Methodological Workflow.* **Lorentz Center Workshop "Movement: New Sensors, New Data, New Challenges"**, 21-25 Aug., Leiden, The Netherlands.

De Groeve J, Van de Weghe N, Ranc N, Moorcroft P, Cagnacci F (2017) *Pan-European Geographic Diversity in Spatiotemporal Sequential Habitat Use by a Large Herbivore*. **NSF Mobility Workshop**, Ohio State University, 11-12 May, Columbus, Ohio, USA.

2016

Cagnacci F & De Groeve J (2016) *Landscape of fear in a human-dominated environment: An approach to incorporate spatio-temporal components in diel cycles.* **Canadian Society for Ecology and Evolution, CSEE 2016**, 7-11 July, St. John's, Newfoundland.

De Groeve J, Van de Weghe N, Ranc N, Neutens T, Ometto L, Rota-Stabelli O and Cagnacci F (2016) *Extracting similar movement patterns in animal trajectories using sequence analysis methods.* **Association of American Geographers Annual Meeting, AAG 2016**, 29 Mar.-2 Apr., San Francisco, California, USA.

2013

De Groeve J, Van de Weghe N, Neutens T, Ometto L, Rota-Stabelli O and Cagnacci F (2013) *Roe deer (Capriolus capriolus) Spatio-Temporal Sequential Habitat Use. An application of Sequence Alignment Methods.* **MOVE Final Conference COST IC0903**, 30 Sept.-1 Oct., Technical University of Vienna, Austria.

De Groeve J, Van de Weghe N, Neutens T, Ometto L, Rota-Stabelli O and Cagnacci F (2013) *From Evolutionary Trees to Behavioral Ecology. Exploiting SAM to visualize spatio-temporal movement patterns.* **Evoluzione 2013 V congress of the Italian Society for Evolutionary Biology (SIBE)**, 28-31 Aug., Trento, Italy.

De Groeve J, Van de Weghe N and Cagnacci F (2013) *Extracting Temporal Patterns in Environmental and Movement Parameters of Roe Deer (Capreolus capreolus) A Methodological Study Using Sequence Alignment Methods.* **Workshop on Analysis & Visualization of MOVEment: Focus on Tangible Results, COST IC0903 MOVE meeting**, 4-5 Mar., Aristotle University of Thessaloniki, Greece.

Training

2017

Workshops at **Open Science Tools, Data & Technologies for Efficient Ecological & Evolutionary Research Symposium**, 7-8 Dec, Amsterdam, The Netherlands.

Interactive Data Analysis and Visualization with R Shiny. Course, 25-29 Sept., Barcelona, Spain.

Summer School in Applied Ecology. IRSAE Summer School 2017, 7-11 Aug., Evenstad, Norway.

2016

Negotiation in Conservation. Training Workshop, 8 Apr., Harvard University, MA, USA.

2015

Temporal and spatial analysis of ecological data in an open source environment. IRSAE course, 26-29 Oct., San Michele 'all Adige, Italy.

Getting started with High-Performance Computing. Course, 27 May-3 June, Ghent University, Belgium.

2014

Flames Summer School in Methodology and Statistics, 8-12 Sept., Ghent University, Belgium.

Methods to evaluate the effects of climate change on ecosystems and populations. IRSAE Summer School 2014, 4-8 Aug., Bo, Norway.

R for macroecological and global change studies. IRSAE course, 22-27 June, Aarhus, Denmark.

2013 **PostgreSQL and PostGIS workshop** organized by EURODEER, 11-12 July, Venice Mestre, Italy.

Introduction to R organized by FEM/CRI, 7,14 & 21 May, San Michele all' Adige (TN), Italy.

Grants

2017 **Lorentz Center grant** (€150) for Mobility Workshop in Leiden, The Netherlands. 21-25 Aug.

NSF Travel grant ($610) for Mobility Workshop at Ohio State University, Columbus, USA. 11-12 May.

2016 **FWO Travel grant for long stay abroad** (€3432), Harvard University, Boston, MA, USA. 26 Mar.-6 July.

2014-2017 **5 IRSAE Travel grants** (€189, €135, €166, €537, €717) for several meetings and courses.

2013 **2 MOVE COST Travel grants** (2 x €2500) for Short-Term Scientific Mission (STSM) at FEM-CRI, San Michele all' Adige (TN), Italy. Feb.-Oct.

Synergistic Activities

Teaching Teaching Assistant at Next Generation Data Management in Movement Ecology. IRSAE Summer School 2015, 1 – 10 July, FEM-CRI, San Michele all' Adige (TN), Italy.

Data Sharing & Scripts De Groeve J, Van de Weghe N, Ranc N, Neutens T, Ometto L, Rota-Stabelli O, Cagnacci F (2016) Data from: Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.h4f7p

Contributor to the Github account of EuroDEER by providing Python and PostgreSQL scripts. https://github.com/eurodeer/eurodeer_db

Personal Github account with R Functions produced during my PhD. https://github.com/jedgroev

Media MOVE COST project highlights, http://youtube.com/embed/HYs1ieMlB2I

'Big Hunter' website ('Studio su migrazione, due nuovi progetti'), 11 Aug. http://www.bighunter.it/

'Ansa' website ('Fauna selvatica, studio su migrazioni'), 08 Aug. http://www.ansa.it/trentino/

'La voce del Trentino' website ('le-migrazioni-della-fauna), 08 Aug. http://lavocedeltrentino.it/

Skills

Mother Tongue **Dutch**

Other languages

	Listening	Reading	Speaking	Writing
English	C2	C2	C2	C2
Italian	B2	B1	B1	B1
French	B1	B1	B1	A2
German	A1	A1	A1	A1

*A1/2: Basic user - B1/2: Independent user - C1/2 Proficient user

Computer skills

	Proficient User	Independent User
GIS	Quantum GIS, GRASS GIS, ArcGIS, GDAL	Mercator, OpenJump
Database	PgAdmin, PostgreSQL, SQL	Access
Analysis	R, PaSSage 2, ClustalTXY	SPSS
Programming		Python, Unix/Bash
Graphics	GIMP, Inkscape	
Other	Notepad++,Sublime Text, Office, LibreOffice	AutoCAD

Interests Guitar, bouzouki/baglamas, music, cooking, travelling, climbing/bouldering, jogging, swimming, GEO-ICT, R-scripting, data visualisation, animal ecology and ethology