



Norwegian University of Life Sciences
Faculty of Science and Technology

Philosophiae Doctor (PhD)
Thesis 2018:96

Extracting Movement Patterns from Geographical Trajectory Data

Uttrekk av bevegelsesmønstre fra
geografiske sporingsdata

Maryam Teimouri

Extracting Movement Patterns from Geographical Trajectory Data

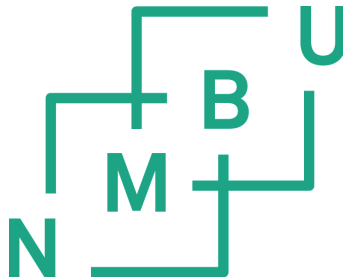
Uttrekk av bevegelsesmønstre fra geografiske sporingsdata

Philosophiae Doctor (PhD) Thesis

Maryam Teimouri

Norwegian University of Life Sciences
Faculty of Science and Technology

Ås 2018



Thesis number 2018:96
ISSN 1894-6402
ISBN 978-82-575-1564-5

A doctoral thesis at the Norwegian University of Life Sciences consists of either a monograph or a compendium of papers in published or manuscript form. In the latter case, the introductory part, from an overall perspective, summarizes and collates the research objectives, theoretical fundamentals, results, and conclusions presented in the separate papers, and documents the coherence of the thesis.

Evaluation Committee:

Dr. Kamran Safi

Department for migration and immuno-ecology

Max Planck Institute of Ornithology and University of Konstanz, Germany

Associate Professor Kevin Buchin

Department of Mathematics and Computer Science

Technical University Eindhoven, Germany

Associate Professor Ola Øvstedal (committee coordinator)

Faculty of Science and Technology

Norwegian University of Life Sciences, Norway

Supervisors:

Associate Professor Håvard Tveite

Faculty of Science and Technology

Norwegian University of Life Sciences, Norway

Associate Professor Ulf Geir Indahl

Faculty of Science and Technology

Norwegian University of Life Sciences, Norway

Acknowledgements

I am grateful to all those who have helped and inspired me during my doctoral study.

First of all, I would like to express my sincere gratitude to Prof. Håvard Tveite and Prof. Ulf Geir Indahl whose invaluable suggestions, guidance and encouragement helped me during the research period.

I wish to thank Prof. Øystein Holand, Dr. Nicolai Hermann Jørgensen and Dr. Geir Steinheim from department of 'Animal and Aquacultural Sciences' at Faculty of Biosciences who shared their expert knowledge with me and provided the dataset.

Additional thanks go to Dr. Hanne Sickel from 'Division of Survey and Statistics' at NIBIO and her husband Morten Sickel for facilitating data collection for the analyses in this thesis. I highly appreciate Dr. Christen H. Fleming from University of Maryland, one of the developers of the ctmr R package, who answered all my technical questions about the ctmr package patiently.

Beyond the academic context, special thanks to God for the gift of my little girl, Anita. She has changed my life from the way I look to the way I think to the way I interact with the world around me. I feel very grateful to my parents, and my siblings for their unconditional love and support when I was away from them studying. Finally, I would like to thank my spouse, Hossein, who has been always warming my heart over years. This thesis would never have been completed without his support.

This research was financially supported by Norwegian University of Life Sciences under project number 1301051701.

Summary

Thanks to the recent advent of inexpensive positioning technologies, data about movement of various mobile objects are collected in rapidly growing amounts. Massive repositories of movement data require developing new data analysis methods in order to reduce unnecessary complexity of the recorded data and extract meaningful information for decision making about the application in question.

In Norway, sheep are the most common large herbivores in most Norwegian mountain areas during the summer months. Remotely monitoring the position of sheep in the mountains is important for animal welfare and provides an increased understanding of grazing patterns. I bring together spatiotemporal concepts and techniques from GIScience, cluster analysis methods from knowledge discovery, statistical tests and reasoning from statistics, and animal movement analyses from the animal study literature in order to extract high-level information from low-level, raw movement data from sheep tracking.

During this research, different movement datasets with different temporal sampling frequencies became available for us. The first dataset is from an experimental site in Hol municipality in Buskerud County and contains the positions of many sheep during the whole summer season (2012) with a half hour sampling interval. The second dataset is from an experimental site in Valdres in Oppland County and contains four individual sheep tracked for some observation periods of 3-4 days (2016) with 10 second time interval.

It is obvious that some observable short time scale patterns will not be detectable for the coarser time scale. Two different segmentation methods tailored to two different sets of movement behaviors were developed for the available datasets. In addition to detection of the movement behavior of the animals for each temporal sampling frequency, a statistical method for estimating the measurement error was proposed. The developed methods for analyzing movement data are all based on the movement parameters of step length, speed and turning angles. I found that the derived movement parameters are influenced by different factors, including temporal sampling frequency. Hence, how to assess the impact of temporal scale on movement parameter estimation was investigated in the final stage of the research.

All in all, this study has been a three stage research process including the development of (I) a segmentation method according to Inactive and Active periods of animal movement paths from low frequency tracking data, (II) a segmentation method for classification animal movement paths from high frequency tracking data into foraging, resting and walking behavior, and (III) temporal sampling frequency assessments for movement parameter estimation.

The core of this thesis is presented in three scientific research papers corresponding to the three stages mentioned above.

Sammendrag

Takket være den stadig økende tilgangen på rimelig sporingsteknologi samles data om bevegelsene til ulike mobile objekter i stadig økende mengder. De store mengdene med sporingsdata krever at det utvikles nye metoder for å redusere unødvendig kompleksitet i de registrerte dataene og for å trekke ut meningsfull informasjon for beslutningstaking. I Norge er sau det vanligste husdyret i de fleste norske fjellområder i sommermånedene. Overvåking av hvor sauene er i fjellet er viktig for dyrevelferd og gir økt forståelse for beitemønstre. Jeg kombinerer rom-tid-konsepter og teknikker fra geografisk informasjonsvitenskap, cluster-analysemetoder fra datautforskning, statistiske tester og resonnement fra statistikk og bevegelsesanalyser fra dyrestudier for å trekke ut overordnet informasjon fra detaljerte sporingsdata fra sau. I løpet av denne forskningen fikk jeg tilgang til sporingsdatasett med ulik samplingsfrekvens. Det første datasettet, fra et felteksperiment i Hol kommune i Buskerud, inneholder posisjoner til mange sau fra hele sommersesongen (2012) med en halv times samplingsintervall. Det andre datasettet, fra et eksperiment i Valdres i Oppland, inneholder sporingsdata fra fire sau med observasjonsperioder på 3-4 dager (2016) og 10 sekunders samplingsintervall.

Det er åpenbart at noen observerbare høyfrekvente bevegelsesmønstre ikke kan avdekkes med lavfrekvensdata. To forskjellige segmenteringsmetoder skreddersydd for ulike sett med bevegelsesadferd ble utviklet for de tilgjengelige datasettene. I tillegg til gjenkjenning av bevegelsesadferden til dyrene, ble det foreslått en statistisk metode for estimering av målefeil. De utviklede metodene for å analysere bevegelsesdata er alle basert på bevegelsesparametrene trinnlengde, hastighet og svingvinkel. Vi fant at de avledede bevegelsesparametrene er påvirket av forskjellige faktorer, inkludert samplingsfrekvens. Derfor ble det i slutfasen av forskningen undersøkt hvordan man kan vurdere effekten av samplingsfrekvens på bevegelsesparameterestimering. Alt i alt har denne studien vært en tre-trinns forskningsprosess, inkludert utvikling av (I) en metode for å segmentere dyrs bevegelser i inaktive og aktive perioder fra lavfrekvente sporingsdata, (II) en segmenteringsmetode for klassifisering i beiting, hvile og gang fra høyfrekvente sporingsdata, og (III) samplingsfrekvensvurderinger for

bevegelsesparameterestimering.

Kjernen i denne avhandlingen er presentert i tre vitenskapelige forskningsartikler som svarer til de tre trinnene som er nevnt ovenfor.

List of Abbreviations

DOP: Dilution of Precision
PDOP: Position Dilution of Precision
HDOP: Horizontal Dilution of Precision
GPS: Global Positioning System
VHF: Very High Frequency
WLAN: Wireless Local Area Network
GIScience: Geographical Information Science
2drms: twice Distance Root Mean Squared
CEP: Circular Error Probability
IP: Inactive Period
POI: Point of Interest
ROI: Region of Interest
DBSCAN: Density Based Spatial Clustering of Application with Noise
DBCIP: Density Based Clustering method for Inactive Periods
OPTICS: Ordering points to identify the clustering structure
MinPts: Minimum Number of Points
SMoT: Stops and Moves of Trajectories
DTW: Dynamic Time Warping
LCSS: Longest Common Subsequence
BCPA: Behavioral Change Point Analysis
RW: Random Walk
CRW: Correlated Random Walk
BCRW: Biased Correlated Random Walk
MSD: Mean Squared Displacement
CTSP: Continuous-Time Stochastic Process
IID: Independent Identically Distributed
BM: Brownian Motion
OU: Ornstein-Uhlenbeck
IOU: Integrated OU
OUF: Ornstein-Uhlenbeck Foraging
SVF: Semi-Variance Function
RTSF: Representative Temporal Sampling Frequency

Contents

Acknowledgements.....	I
Summary.....	III
Sammendrag	V
List of Abbreviations.....	VII
1 Introduction.....	1
1.1 Introduction to Norwegian Sheep Farming.....	1
1.2 Motivation	3
1.3 Research Questions and Research Objectives.....	5
1.4 Research Process and Research Papers	6
1.5 Structure of the Thesis.....	7
2 Theoretical Background and State of the Art.....	9
2.1 Animal Tracking	9
2.2 Animal Movement Trajectories.....	12
2.3 Identifying Active and Inactive States from Animal Movement Trajectories ...	13
2.4 Identifying Movement Behavior States from Animal Movement Trajectories .	16
2.5 Stochastic Animal Movement Models	19
3 Methods and Results.....	25
3.1 Research Paper I.....	25
3.1.1 Objectives.....	25
3.1.2 Methodology.....	26
3.1.3 Main Findings.....	27
3.2 Research Paper II.....	27
3.2.1 Objectives.....	28
3.2.2 Methodology.....	28
3.2.3 Main Findings.....	29
3.3 Research Paper III.....	30
3.3.1 Objectives.....	30
3.3.2 Methodology.....	31
3.3.3 Main Findings.....	32
4 Conclusion.....	33
4.1 Main Contributions	33
4.2 Practical view in terms of sheep farming.....	34
4.3 Challenges for Future Research	35

References	37
Errata.....	43
Research Paper I-III (Individual page numbering)	

1 Introduction

The emergence of digital tracking technologies such as VHF, GPS, and Argos has simplified the acquisition of spatiotemporal information of moving objects including animal movements. Almost every process related to animal ecology, including foraging, migration, predation and predator avoidance is related to animal movements. Understanding of animal movement, including why, how (fast or slow), when and where animals move, can be useful in a wide range of biological topics, including population and community ecology, animal physiology, disease spread, gene flow, and wildlife management and conservation (1).

However, movement processes take place at different spatiotemporal scales, from coarse temporal scales in migration research studies to fine temporal scales in home range analysis and even finer in transitions between different behavioral states (e.g., resting, moving, foraging) (2).

Additionally, autocorrelation is another concern that we must take into account in tracking data analysis. On one hand, it is an obstacle in ecological studies as it violates the basic assumptions of standard statistical hypothesis testing. On the other hand, autocorrelation is often the consequence of ecological processes and also high frequency sampling data (3). Therefore, there is a need for adequate methods for analyzing movement data and extracting relevant information. The emergence of similar challenges in data analysis of tracking data has encouraged us to offer an interdisciplinary perspective between *geography* and *animal movement study* to trigger innovative approaches and research direction.

This thesis presents research on the development of methods applicable to movement of all kinds of animals, but our focus has been on livestock. We start this chapter with a brief history of Norwegian sheep farming and then, we mention our motivations and all the research questions and objectives for this thesis.

1.1 Introduction to Norwegian Sheep Farming

Domestic sheep are the most abundant large herbivore in alpine pastures during the Norwegian summer grazing season (4). Summer farming was formerly an important part

of Norwegian husbandry and mountain areas have been used for grazing for hundreds of years (5). Lambs are born during late winter or spring while the sheep are fed indoors. Sheep and lambs are being released in open forests or alpine ranges for about 100 days during summer. Most lamb growth occurs on the open ranges during summer. The animals are gathered in September, often using herding dogs, and selected lambs are then slaughtered (6). Supervision in the Norwegian context does not involve guarding or active herding, rather it represents patrolling of the grazing area to look for signs of dead or injured sheep (7).

Each day the animal decides where and when to graze, ruminate, rest and drink based on its own needs and the kind of available resource. Biotic factors, such as forage quality and quantity, and abiotic factors, such as topography, water availability, and thermal cover, affect animal performance and uniformity of grazing (8, 9).

One of today's most crucial agricultural dilemmas is: how to find a balance between a fast-growing global demand for food and the need to sustain the natural resource base of land, water, air and biological diversity (10).

As stated in (11), sustainable sheep grazing management must be considered from different aspects as follows:

1. Animal productivity (weight gained per unit time grazed in open pastures)
2. Sustainable production of forage plants
3. Erosion
4. Biodiversity

Although high animal productivity is both economically and socially beneficial to human beings and food demand in the short term, it might lead to ecological issues. For example, high animal density over a longer timescale (i.e. overgrazing) could reduce forage quality and quantity, lose biodiversity and lead to erosion. Also under-grazing might lead to ecological issues. For example, decline in sheep farming practices in the alpine areas of Norway has caused a large-scale encroachment of trees and shrubs leading to extensive landscape changes and decreased biodiversity (5).

In order to enhance the positive impact of livestock on the environment and contribute to the sustainable use of natural resources, the impact of management factors such as the number of sheep per unit area, breeds, grazing season, supplemental feeding, relative use

of different habitats and vulnerability to soil erosion, to name but a few, must be considered for any experiment. The control of the above-mentioned management factors is out of the scope of this thesis.

This thesis intends to primarily contribute to animal movement studies, figuring out livestock distribution and its spatiotemporal variation in open ranges. By distinguishing different behaviors from sheep movement data, we can contribute to the understanding of how sheep move and react.

1.2 Motivation

The ability to track sheep as moving objects with tracking units provide researchers (and potentially farmers) with valuable information about how sheep use their environment. Animal movement data is fundamental to many different branches of research including movement pattern mining, habitat use and selection analysis using vegetation maps, to name but a few.

From the geographical point of view, analyzing sequences of positional data and finding spatiotemporal movement patterns is an interesting challenge. Knowledge discovery techniques and statistical inference in addition to spatiotemporal analyses make it possible to explore the daily movement of sheep in order to detect pronounced/abnormal movement.

Given a dataset and general movement pattern mining questions, the practical starting point is to first carry out some exploratory analyses to gain a better understanding of data and then come up with different hypotheses and research questions. Finally, available methods can be adapted to the application or new methods can be developed.

In the following, we mention the process that motivated us to carry out this study and formulate our Research Questions.

First we got hold of half hourly position data of sheep tracking for a few months duration and did some exploratory analysis to gain insight into sheep movement. One of the main undertakings was to figure out the daily spatial distribution of sheep using Standard Deviational Ellipses (SDE)(12).

According to (13), this daily topographical trend of movement of sheep can be due to the existence of insects. Sheep tend to select upland or open areas with more wind for resting to avoid insects. The result of this exploratory analysis encouraged us to identify the resting pattern of sheep. Where and when is the favorite place and time for sheep to be inactive? According to the literature, the resting pattern has mostly been identified using activity sensor values and field observation (14). Due to the lack of this kind of extra information for our dataset, we tried to model resting through the spatiotemporal distribution of positional data by taking into account measuring error. To this end, we formulated our first Research Objective (see Section 1.3) and conducted stage I (see Section 1.4).

If we had had access to a dataset with significantly higher temporal sampling frequency, more accurate inference about the behavior between consecutive observations would be attainable. Straight-line connections between consecutive observations would then better represent the movement path of the animal. Analysis of various combinations of the derived movement feature values such as speed and turning angles could allow the identification of movement behavior type. Our premise was that the overall distribution of speed and turning angles for the observations must be the basis for our reasoning and clustering. A significant body of work has been done to distinguish between different movement behaviors (15, 16). Our interest was to find a solution that takes into account the spatiotemporal autocorrelation structure of the observations while considering the overall distribution of speed and turning angle values. To this end, we formulated our second Research Objective (see Section 1.3) and conducted stage II (see Section 1.4).

Although there is literature that investigates the accuracy of movement parameter estimation, there has been considerably less work on how movement parameter estimation varies with temporal sampling frequency (17, 18). Additionally, battery capacity and the length of the monitoring period put constraints on the temporal sampling frequency in practical field studies. We were therefore motivated to assess the impact of temporal sampling frequency on movement parameter estimation. Our goal was to design a field experiment with appropriate temporal sampling frequency such that the results of the performed analyses should not be affected much by the sampling regime. To this end, we formulated our third Research Objective (see Section 1.3) and conducted stage III (see Section 1.4).

1.3 Research Questions and Research Objectives

Research Question: Five Research Questions have been formulated:

1. How do animals move over time and what kind of movement patterns can be detected by exploring the recorded dataset?
2. What is the impact of positioning error in the animal movement recordings? How can we detect the impact of positioning error in the recordings? For which behavior states are the influence of positioning error pronounced?
3. How can we figure out different movement behaviors along an animal trajectory? How can we segment a trajectory into sub-trajectories according to movement behavior? How can we find segments that are similar with respect to movement behavior? How can we label the animal trajectory (annotating) according to the type of movement behavior?
4. To what extent are animal recordings auto-correlated in time and space? How can we take this into account in animal movement analysis?
5. To what degree do movement parameters such as speed, sinuosity, and turning angle vary when derived at variable temporal scales?

Research Objectives: This thesis intends to contribute to animal movement analysis by developing clustering methods on spatiotemporal recordings. The purpose of our first spatiotemporal clustering method is to distinguish between inactive and active periods along the animal trajectory. Our second spatiotemporal clustering method aims to partition and label animal trajectories according to foraging, resting and walking behavior. This thesis intends to investigate positing errors of the measurement units and assesses the requirement for temporal sampling frequency for trajectory data.

Two dataset from Norwegian sheep have been available for us. One dataset with half (or one) hour time interval and the other one with 10 second time interval. We consider the former one as low frequency tracking data and the latter as high frequency tracking data.

The aforementioned research questions are embodied in the following research objectives:

Objective 1: Given a user-specified minimum duration of inactivity, the research shall find Inactive Periods of an animal trajectory based on low frequency tracking data. By considering each Inactive Period as a dense cluster of recording, this research shall develop a density-based clustering method to both detect Inactive Periods and estimate positioning error.

Objective 2: The research shall develop a method to distinguish between three types of animal movement behavior (foraging, resting and walking) based on high frequency tracking data.

Objective 3: The research shall investigate the temporal interval between recordings and assess over-sampling or under-sampling in terms of different research questions, for example movement parameter estimation. Finding the temporal sampling frequency for which the difference between the reconstructed movement path and the true movement path is the minimum in terms of movement parameter estimation.

1.4 Research Process and Research Papers

The core of this thesis is presented in three scientific Research Papers. This thesis follows a three-stage process in order to achieve the main objectives and address the research questions. Each stage of this process is presented in a Research Paper, and is associated directly with one or two research questions and objectives. The thesis reports on the research published (or submitted) in peer-reviewed international scientific journals.

Stage I: Inactivity and Activity Period Detection from Low Frequency Tracking Data

Research Paper I presents the developed method and the outcome of this stage. This Research Paper is based on low frequency tracking data of non-differentially corrected GPS positions and discriminate between inactive and active periods of the animal.

In the context of this Research Paper, previous research from different areas is relevant: (i) current methods for identifying POIs (points of interest) in movement trajectories discussed in GIScience; (ii) those branches of animal study literature that are devoted to

identifying Inactive Periods; (iii) handling GNSS inaccuracy for biotelemetry. It treats Research Objective 1 and the associated Research Questions 1-3 of this thesis.

Stage II: Movement Behavior Detection from High Frequency Tracking Data

The main research question of Research Paper II is: given a sequence of locations of animals sampled at high frequency, how to make inferences about animal behavior. Where and when does the animal engage in a specific behavior and how does the behavior change over time?

One common reason for analyzing animal movement data is to discover latent information about behavior that cannot be observed directly. The results of behavior detection can significantly contribute to other important issues such as pasture management.

Stage III: Intermediate-Frequency Tracking for Movement Parameter Estimation

Movement parameters can enrich trajectories with additional information on the characteristics of the movement. The inaccuracy of GNSS measurements, sampling frequency, and species-specific movement characteristics influence movement parameter estimation. Research Paper III tries to model the contributing factors on movement parameter estimation and gives a guideline for designing experiments according to constraints.

1.5 Structure of the Thesis

The content of this thesis is presented in two parts. Part (I) - Synopsis; and Part (II) - Research Papers.

We start the Synopsis part with the theoretical background and a review of state of the art of animal movement analysis (Chapter 2). We also look into the basic definitions of telemetry and movement trajectories.

The Synopsis proceeds with Methods and Results (Chapter 3) that covers the research process, the objectives, the methods used in each research paper, and the main findings of the research papers. Finally, the Synopsis ends with conclusions and outlook for future research (Chapter 4).

In part II (Research Papers), the papers are presented as they were submitted or published. To obtain a comprehensive insight into the methods, problems, and achievements of this work, studying the full papers provided in Part II is recommended.

2 Theoretical Background and State of the Art

The overall aim of this chapter is fivefold: First, to introduce animal tracking and different issues and concerns for GNSS tracking units (Section 2.1). Second, to illustrate the trajectory concept and describe the different aspects of trajectories that are central to this research (Section 2.2). Third, to introduce two interchangeable concepts (*Inactive Period* from the animal movement study and *Stop* from GIScience) and report the relevant literature (Section 2.3). Fourth, to give an overview of earlier and current trends in animal movement behavior research (Section 2.4). And finally, to describe a recently developed stochastic animal movement model and take advantage of it to decide optimal temporal frequencies for research experiments based on the research questions (Section 2.5).

2.1 Animal Tracking

Usually, we cannot observe the complete, continuous movement path of an animal through space and time. Instead, we sample a set of discrete locations to approximate the animals' actual movement path (19). The resulting sequence of consecutive discrete locations of the animal (e.g., spatiotemporal coordinates, ordered by time) is termed a movement path (trajectory).

The emergence of digital tracking technologies have strongly facilitated the study of the movement and behavior of animals in their natural environment (20, 21). GNSS tracking have been applied to a wide range of taxa, including mammals (22, 23), reptiles (24), fish (25), and birds (26).

Measurement Accuracy

Recorded GNSS locations are subject to errors, which might be a concern based upon the application. GNSS accuracy for a given record may be affected by atmospheric conditions, satellite or receiver clock error, satellite orbit-error and multipath effects (bounced signals) (27), as well as topography, overhead canopies, or adjacent structures. Additionally, the accuracy of each GNSS location depends on the number of and position of visible satellites. Each location has a specific dilution of precision value, which is a measure of satellite geometry quality. The Position Dilution of Precision (PDOP) value and the *fix rates* (the

number of recorded locations vs. the number of attempted recorded locations (28)) are proxies for the quality of GNSS locations (29). Inaccurate locations are more likely when PDOP values are large, indicating poor satellite geometry for triangulation between the GNSS receiver and three or more satellites. The influence of satellite geometry on accuracy can be considerably reduced when applying differential correction (30). Differential correction involves calibration of measured locations using the measured location of a base station relative to its true location (31). However, GNSS based locations in biotelemetry are still inaccurate due to environmental parameters (canopy and topographic obstruction), weather conditions and animal behavior (32, 33). We can be certain that closed canopies and steep terrains interfere with satellite signals, inducing habitat-biased errors that include both missing and spatially imprecise animal locations (14).

Location measurement accuracy also depends on the type of tracking, *intermittent mode* versus *continuous mode*. In intermittent mode, the tracking unit is turned off between two consecutive recordings to save battery. In continuous mode, the tracking unit is on all the time, and the positions are expected to be more accurate.

Stationary Test

In order to quantify GNSS accuracy, it is common to perform a stationary test before the experiment. For a stationary test, the GNSS receiver is positioned at an accurately surveyed location and left in place for a period of time (usually 24 h or more). The device attempts to acquire locations (or 'fixes') at a pre-determined temporal interval. The number of successful fixes (stored locations) divided by the number of attempts yields the *fix rate*. The accuracy is quantified as the average distance between the estimated locations and the 'true' location of the device (the latter determined using fixed geodetic markers, differential-correction or a large-sample average)(14, 34).

Several different accuracy measures have been explored in various applications (Figure 1). 2drms (twice Distance Root Mean Squared), *CEP* (Circular Error Probability) and *R95* (an extension of CEP) are the most commonly used two-dimensional accuracy measures (35). *CEP* is the radius of the smallest circle centered at the true position that covers 50% of the observations. *R95* is a similarly defined quantity covering 95% of the observations.

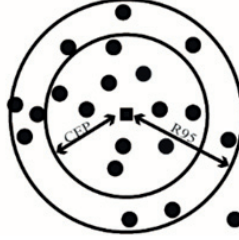


Figure 1. Points marked with a solid circle are the recorded points for the stationary test and the point with the solid square is the true location.

The $2drms$ (Twice Distance Root Mean Squared) is defined as two times the Root-Mean-Square (RMS) of the radial errors as follows:

$$2drms = 2 * \sqrt{\frac{\sum_{i=1}^k ((x_i - \mu_x)^2 + (y_i - \mu_y)^2)}{k}} \quad (1)$$

Where k is the number of recorded positions for the *stationary test*. Here $[(x_i, y_i) | i=1, \dots, k]$ represents coordinates of k recorded positions for each *stationary test* and (μ_x, μ_y) is the true stationary site.

For optimal performance, tests should deploy the GNSS devices on a medium that mimics an animal's body.

We know from stationary tests that canopy closure (33, 36), topography of the area (37, 38), habitat selection (39, 40), and collar orientation (41, 42) in addition to GNSS intrinsic error sources (as previously mentioned) influence both *fix rate* and location precision (see (14) for details). The effects of canopy closure on GNSS errors remain predominant (33, 43).

Temporal Sampling Frequency

Due to battery limitations, GNSS users are still confronted with a trade-off between the overall duration of monitoring and the number of locations per day, making the latter a key limiting factor for field studies (30). Generally, trade-offs between GNSS location collection intervals and unit longevity lead to fewer monitored individuals and shorter study

durations (14). Decreasing the interval between recorded locations increases the level of autocorrelation in the resulting data.

Longer time between data logging means that less fine-grained information about an animal path will be available, but the lifetime of the device in the field will increase.

Both memory and power could be used more efficiently if the data collecting device could change its temporal sampling frequency given the activity state of the animal. For example, during Inactive Periods there is little to benefit from a high temporal sampling frequency, while during active periods, a high temporal sampling frequency is required to identify prominent properties of an animal's path (44). It is important to carefully design an experiment in terms of sampling frequency, memory considerations, and battery limitations, making sure that the data can help to answer the specific experimental questions.

Throughout this thesis, we use two different GNSS recorded datasets with a priori sampling schedules. For Research Paper I, GPS collars were configured to record animal locations at intervals of 30 minutes (low sampling frequency). For the second dataset, the sampling interval is about 10 seconds (high sampling frequency). In the third paper, based on high frequency data, we develop a method to pre-design the sampling frequency for a specific research question.

2.2 Animal Movement Trajectories

Continuous animal movement in two-dimensional space can be represented as a discrete, ordered set of locations and associated times.

Definition 1: Trajectory (Traj) - A sequence of spatiotemporal points recording the trace of a moving object, i.e. $Traj = (P_1, P_2, \dots, P_n)$ where $P_i = (x; y; t)$ is a tuple including a position $(x; y)$ and a timestamp (t) .

Trajectories also differ with regard to the regularity of the time intervals between successive steps. Irregular data commonly result from missing relocation fixes or varying sampling frequencies throughout a study period (45). (A regularly sampled trajectory

(either recorded using a sampling frequency or resampled using interpolation techniques) is a necessity for Research Paper I, while it is not mandatory for Research Paper II&III) For the sake of clarity, let us recall some basic principles. Calenge et al. (46) categorize trajectories by using the following movement parameters (Figure 2):

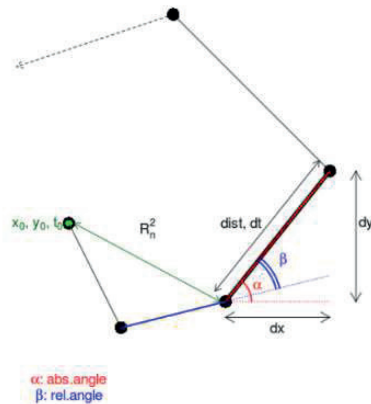


Figure 2. Definition of movement parameters on a trajectory (46)

dx , dy , dt : these parameters describe the increments in the x and y directions and the increment in time between the locations i and $i+1$.

dist: the spatial distance between successive locations.

abs.angle: the absolute angle α_i between the x axis and the line from location i to $i+1$.

rel.angle: the relative angle β_i measures the change of direction between the line from location $i-1$ to location i and the line from location i to location $i+1$ (often called turning angle)

R_n^2 (Net Squared Displacement - NSD): the squared distance between the first location of the trajectory and the current location (47).

2.3 Identifying Active and Inactive States from Animal Movement Trajectories

In this section, we want to show the possibility of identifying 'Inactive periods' (inspired from the animal study literature) / 'Stops' (inspired from GIScience literature) for half-hourly sampled data (low frequency). The problem can be formally described as a

trajectory segmentation problem, where the trajectory is divided into a series of static and moving segment groups. A wide range of methods exist for segmenting trajectories (48, 49); however, trajectory segmentation is mostly performed using ad-hoc methods developed for the dataset in question (17). Our segmentation method for animal movement trajectories is based on the detection of Inactive Periods. For subsequent animal movement analysis, the identified Inactive Periods could be excluded from the trajectory or given special treatment.

Definition 2. Inactive Period (IP): A state where the animal is stationary or has limited variation in geographic location.

When an animal sleeps or lies down and rests at one location over a period of time, the collected positions can be misplaced, due to positioning error, becoming scattered around the true resting position. A stationary test (see Section 2.1) is often used to estimate the positioning error in advance.

There are several strategies proposed in the animal study literature to identify Inactive Periods using activity sensors combined with GNSS (30, 44, 50), minimum distance or acceleration thresholds (34, 51), or combinations of activity sensors, movements and other variables (34, 50, 52) (see (14) for details). Predictive models that combine information from various sensors with field observations can also be used to detect IPs (53, 54).

The choice of a cut off distance or an acceleration threshold that can differentiate active from inactive behavior is subjective and will influence the results. Given location data with neither activity sensor values nor field observation, none of the above-mentioned strategies for detecting Inactive Periods are applicable. To address this shortcoming, we developed a clustering method, inspired by stop detection from GIScience. In GIScience research, the term 'stop' is, to some extent, the same concept as an Inactive Period.

Definition 3. Stop: the important places of the trajectory where the moving object stayed for at least a given minimal amount of time (55).

According to Spaccapietra et al. (56), a raw trajectory may semantically be segmented into 'move' and 'stop' parts (an event-based perspective) (57). In the conceptual framework of an application, the spatial range of the trajectory for each 'stop' part is a single point (56).

The original definition of “stop” was emphasized as “not moving” (56). However, real-life stops are not purely depending on whether it is moving or not. For example, ‘shopping at a mall’ with wandering around in a store can be considered as a stop even if the user is moving. In our *IP* detection algorithm, we consider ‘stop’ as not moving or a limited variation in the location (for example, when the animal reorients itself).

Anderienko et al. (58) classified stop detection methods based on movement parameters such as speed (59, 60), distance (61), speed and duration (62), speed and distance (63) and direction (64).

Gong et al. (65), in another study, categorized the identification of stop locations into the following five groups: centroid-based methods (66), speed-based methods (67), duration-based methods (68), density-based methods (69), and hybrid methods (58). Hybrid methods use two of the variables such as speed, duration, density, etc., together.

Recently, Luo et al. (70) divided stop discovery algorithms into two categories, i.e. *static* and *dynamic* methods.

Typically, the *static* methods, e.g. Algorithm SMoT (Stops and Moves of Trajectories) (55), are built on some a priori known POI (Points Of Interest) or Regions Of Interest (ROI). In these methods, first, the spatial intersections of the POI/ROI with the trajectory are computed to find stop candidates, and then additional stop constraints (e.g. time duration) are applied to filter out non-meaningful stop candidates. Difficulty in getting well-defined POI data significantly limits these *static* methods.

Dynamic methods, without using any predefined POI/ROI, consider different aspects of mobility characteristics (70), such as velocity characteristics. Typically, general clustering algorithms (such as k-means clustering, DBSCAN (71) and OPTICS (72)) are adopted in the dynamic solutions (58, 65-69).

In Research Paper I, inspired by clustering-based stop discovery methods, we consider each *IP* as a dense cluster of recordings in space-time, and assume that the location error is normally distributed. We test our algorithm using both simulated movement trajectories with well-defined characteristics and a real movement dataset. The only parameter required for our proposed algorithm is a user-specified minimum duration of inactivity, that will depend on the application.

2.4 Identifying Movement Behavior States from Animal Movement Trajectories

In the second stage of this research, first, we report on the relevant literature about movement behavior state detection, and we proceed by developing a movement behavior state detection algorithm. Research Paper II presents the substance of this stage. Movement behavior detection is often an early and critical component of movement data analysis that guides further analysis pathways.

A segmentation algorithm decomposes a trajectory into a set of segments (sub-trajectories) with homogeneous characteristics with respect to some application-dependent criteria (73, 74). Then, the segments are grouped together corresponding to movement behavior (75). Behavioral movement analysis methods, according Gurarie et al. (16), can be categorized into four groups: **Metric-based methods**, **Classification methods**, **Mechanistic movement models** and **Phenomenological methods**.

Metric-based methods are built on various *Movement Parameters*. *Movement Parameters* such as Net/Mean Squared Displacement (76), Fractal Dimension (77), (Multi-Scale) Straightness Index/Sinuosity (78), and First Passage Time (79) are defined to identify the different behaviors (see (19) for the details).

Classification methods apply clustering algorithms and selection criteria to determine the number of partitions a dataset can be subdivided into. For example, Van Moorter *et al.* (52) distinguished four different states of behavior of elk by k-means clustering, using turning angle, step length derived from GPS location data and activity sensor values for free-ranging elk.

Mechanistic movement models include the large family of random walk models with variations (see Section 2.5), in which movement of an animal is broken down into two components: step length and turning angle between successive steps (80). Different probability density functions for step length and turning angle distribution (15, 81) are assumed as the underlying movement model for the different behaviors. A limitation of this method is that it is ultimately only as good as the underlying movement model and the behavioral hypotheses associated with it (82). In case of no a priori anticipated behavioral states, this kind of model becomes inadequate.

Phenomenological methods, with origins in time series and signal processing, identifies the structure or periodicity of movement paths through time series of movement parameters. Our Research Paper II is built on a phenomenological method called 'Behavioral Change Point Analysis' (BCPA(83)). In BCPA, two time series are created: (i) 'Persistence velocity', the product of the estimated speed and the cosine of the turning angle, and (ii) 'Turning velocity', the product of the estimated speed and the sine of the turning angle. These two continuous auto-correlated time series are described by three parameters: a mean, a variance and a continuous autocorrelation. A change in these parameters corresponds to a change in behavior. For example, an increase in the mean of the 'Persistence velocity' component corresponds to a combination of faster and more directed movement. An increase in variance of the 'Persistence velocity' component indicates more variable movement, e.g. bouts of movement and stopping and sudden changes in direction. A higher autocorrelation indicates more directed and correlated movements, whether fast or slow (83). Naming the time at which change in the movement parameter occurs as a *change point*, there are eight possible models to consider when analyzing each side of a possible *change point*: 1) no significant change for mean, variance and autocorrelation, 2) significant change for mean, 3) significant change for variance, 4) significant change for autocorrelation, 5) significant change for mean and variance, 6) significant change for mean and autocorrelation, 7) significant change for variance and autocorrelation and 8) significant change for all the parameter values.

Given a sweeping analysis window of fixed size (this size is the only tunable variable), BCPA uses maximum likelihood to find the most likely change point within the window, corresponding to shifts in behavior, and determines which combination (if any) of the three parameters (mean, variance and autocorrelation) most parsimoniously describes the separation in the data according to BIC (Bayesian Information Criterion). BIC is a criterion for model selection among a finite set of models (84). The interested reader is encouraged to study the original paper for more details on the derivation of formulas and theory behind BCPA (83).

The developed methodology of Research Paper II consists of the following processes: Trajectory segmentation and sub-trajectory clustering. Trajectory segmentation partitions

the trajectory into reasonably homogenous sub-trajectories using BCPA and then these sub-trajectories are grouped according to similarities in movement behavior.

Grouping sub-trajectories is closely related to clustering analysis and similarity measures. We refer the reader to (85) for a complete review of similarity measures. The most important similarity measures defined for assessing the similarity of (sub)trajectories are **Spatial similarity measures** and **Spatiotemporal similarity measures** (86) as follows:

Spatial similarity measures: These similarity measures ignore the temporal aspect of (sub)trajectories and consider just geometric shapes. The simplest one is '**Euclidean distance**' which only works on trajectories with the same duration and sampling schedule. '**Hausdorff distance**' (85, 86) is defined as the maximum of the minimum Euclidean distances between each point of one trajectory to the set of points constituting the other trajectory. '**Fréchet distance**' (75), also called the dog-walking distance, takes into account the location and ordering of trajectory points and represents the minimum length of a leash between two objects, i.e. a person and its dog, that move along their respective trajectories without backtracking.

Spatiotemporal similarity measures: These similarity measures compare movement with respect to both spatial and temporal aspects. The simplest and most straightforward spatiotemporal distance measure is the sum of all Euclidean distances between the positions of the two trajectories that match in time divided by the common duration of the two trajectories. '**DTW (Dynamic Time Warping)**' (87), a classic speech recognition tool (88), calculates the similarity between two trajectories by finding an optimal match between all their points, even if the trajectories are not identical in size. The similarity measure is called 'time warping' since the two trajectories are compressed or expanded in time to find the optimal match between them. DTW finds a monotonically increasing matching between all pair of points in the two trajectories with the minimal cost. '**LCSS (Longest Common Subsequence)**' (87) matches two trajectories by allowing them to stretch, without rearranging the sequence of the elements and allows some elements to be unmatched in order to handle outliers. '**Edit distance**' (89), with the origin in pattern matching for alphanumeric datasets (86), encodes movement parameters for each trajectory into a symbolic sequence, and then calculates the cost of the string matching between the two generated symbolic sequences.

According to (15), animal movement behavioral states can be categorized by probability distribution functions of movement parameters. In Research Paper II, we establish the concept of similarity between different animal movement behaviors based on the cumulative distribution function of movement parameters. We do a spatiotemporal similarity analysis of the cumulative distribution function of movement parameters to detect sub-trajectories with similar dynamic behavior.

2.5 Stochastic Animal Movement Models

The final stage of this research intends to model animal movement data using continuous-time stochastic movement modeling, and to assess the sampling frequency requirements for estimating movement parameters correctly from a discrete-time movement trajectory. Movement of an individual organism is an interplay of four major components: (i) internal state (why move?), (ii) motion capacity (how to move?), (iii) navigation capacity (when and where to move?), and (iv) external factors (the abiotic and biotic environment influencing movement) (75, 90). Since modeling all of these interactions for movement of organisms is infeasible in practice, various analytic techniques, including Simple Random Walk models or Composite Random Walk models (80, 91, 92), have long served as the standard approach to wildlife movement modeling (82).

Uncorrelated and unbiased (isotropic) random walk is the most simple stochastic movement model. Uncorrelated means the direction of movement is completely independent of the previous directions (92). Unbiased means there is no preferred direction; the direction moved at each step is completely random. **Correlated random walk (CRW)**, frequently used to model animal paths in various contexts (e.g. (80, 93)), involves a correlation between successive step directions, which is termed 'persistence' (94). Persistence is the tendency to continue moving in the same direction (95). It means that the animal changes its direction gradually from step to step. **Biased correlated random walk (BCRW)** (92) has a consistent direction bias in addition to persistence. Codling et al. (92) provide a comprehensive review of different random walk models applied in ecology.

RW analyses focus on quantifying step length distributions and turn-angle distributions, and comparing the *observed mean squared displacement (MSD)* to the *model-predicted MSD* (92, 96).

Assuming that an observed trajectory is comprised of N successive points as follows:

$$P_i = (x_i, y_i), \quad i = 1, \dots, N$$

Mean Squared displacement (MSD) as a function of time duration (corresponding to n consecutive steps) is defined as the mean of the squared distances between pairs of points separated by that amount of time duration (80, 97):

$$MSD_n = \frac{1}{N-n} \sum_{i=1}^{N-n} (p_{i+n} - p_i)^2, \quad n=1, \dots, N-1$$

MSD can be used to distinguish between diffusive, super-diffusive, and sub-diffusive movements. In diffuse movement, *MSD* scales linearly with time or path length, but sub- or super-diffusive movements have a power-law relationship with time (98).

Parameter estimation for RW and its variations depend on the sampling schedule, making it difficult to draw sampling-independent inferences about the underlying movement process (99, 100). On the other hand, RW and its variations are based on locations that are collected at regular intervals, making subsampling, interpolation, or aggregation necessary for irregular sampling data (101). Furthermore, memory, avoidance, repetition and other biological complexities introduce long-term autocorrelations into the movement paths of individuals. Long-term autocorrelations in the movement paths of individuals cannot be modeled by using RW (102).

Recent developments in modeling movement as a *continuous-time stochastic process (CTSP)* (99, 100) solve the above mentioned problems (irregular sampling, sampling schedule dependency and long-term autocorrelation). *CTSP* models separate the continuous-time movement process from the discrete-time sampling process. *CTSP* models feature *position autocorrelation*, *velocity autocorrelation*, and *range residency*. *Position autocorrelation* means that position observations that are close in time tend to be more similar than those farther apart in time. *Velocity autocorrelation* means that an individual's direction and speed at one point in time tend to be correlated with those quantities at adjacent times.

Velocity autocorrelation is often observable for finely sampled datasets. *Range residency* exhibits the tendency of the animal to remain in a defined region or 'home range'. *CTSP* models currently in use in the ecologic literature (103), including ***IID***, ***BM***, ***OU***, ***IOU***, and ***OUF***, feature different combinations of *position autocorrelation*, *velocity autocorrelation* and *range residency*. In the following, we give a general explanation of the most frequently used *CTSP* models for animal movement in the ecologic literature and refer the reader to (99) for more details.

Independent Identically Distributed (IID): A purely random movement model which ignores autocorrelation completely in both positions and velocities. It is generally not recognized as a suitable movement model for organisms.

Brownian Motion (BM) (104): BM is a continuous counterpart to the Random Walk model (20). It features correlation in positions but not in velocities. It does not limit space use. BM is not able to model velocity autocorrelation and range residency. BM is appropriate for data with too coarse sampling rate or too short sampling duration, as such data will not reveal velocity autocorrelation and range residency, respectively.

Ornstein-Uhlenbeck (OU) (105): OU combines BM with a tendency to remain in a particular area. The OU process is therefore appropriate for data with too coarse sampling to reveal velocity autocorrelation, and with sampling duration that is long enough to show the restricted space use.

Integrated OU (IOU) (106): Similar to BM, IOU assumes that space use is not limited, but it assumes autocorrelation in velocities. IOU is therefore appropriate for finely sampled datasets that reveal velocity autocorrelation, but that have a sampling duration that is too brief to show range residency.

Ornstein-Uhlenbeck Foraging (OUF) (99): OUF is a hybrid of the OU and IOU processes that features both correlated velocities and restricted space use. The OUF process is appropriate for finely sampled data that reveal velocity autocorrelation and that have a long enough duration to reveal restricted space use.

Comparable to *MSD* for RW modeling, Semi-variance, a nonparametric autocorrelation estimator, contains information about the step length distribution for all possible time lags in the data.

The mathematical definition of semi-variance is as follows (99):

$$\gamma(\tau) = \frac{1}{2} * E((x(t + \tau) - x(t))^2)$$

Which measures the variability in the distance between two locations (noted as $x(.)$) as a function of time separation (τ) between them. Plotting the estimated semi-variance as a function of lag yields the empirical semi-variogram, which contains information about the mix of processes represented in a relocation dataset.

All *CSTP* models can be expressed in terms of their *SVFs* as follows (see Figure 3):

- IID: Since autocorrelation is zero for all time lags other than 0, the semivariance is constant for all time lags other than zero.
- BM: Semivariance is a linear function of time.
- OU: Semivariance is a mixture of a linear function of time over short scales, $\tau < \tau_H$, and approaches a constant for longer time lags ($\tau > \tau_H$)
- IOU: Semivariance is a mixture of a power law function (curve upward) for small time lags, $\tau < \tau_F$, and a linear function for longer time lags, $\tau > \tau_F$
- OUF: Semivariance is a mix of a power law function for small time lags, $\tau < \tau_F$, a linear function for intermediate time lags, $\tau_F < \tau < \tau_H$, and then approaches a constant for longer time lags ($\tau > \tau_H$)

The theoretical *SVFs* of *CSTP* models can be fit to the empirical semi-variogram via regression and then compared using standard model selection techniques to identify the mix of movement processes best supported by the relocation data (99).

Uncertainties in animal movement data, owing to issues such as measuring accuracy and sampling frequency, may strongly influence interpretations (17, 107).

We can't get GNSS data with very high temporal sampling frequency for a long monitoring duration due to battery constraints. We also lose a lot of information when recording data with a low temporal sampling frequency. In Research Paper III, we try to assess the impact of sampling frequency on the estimation of the movement parameter of speed, and develop a method to find the most effective sampling frequency in terms of the accuracy of speed estimation.

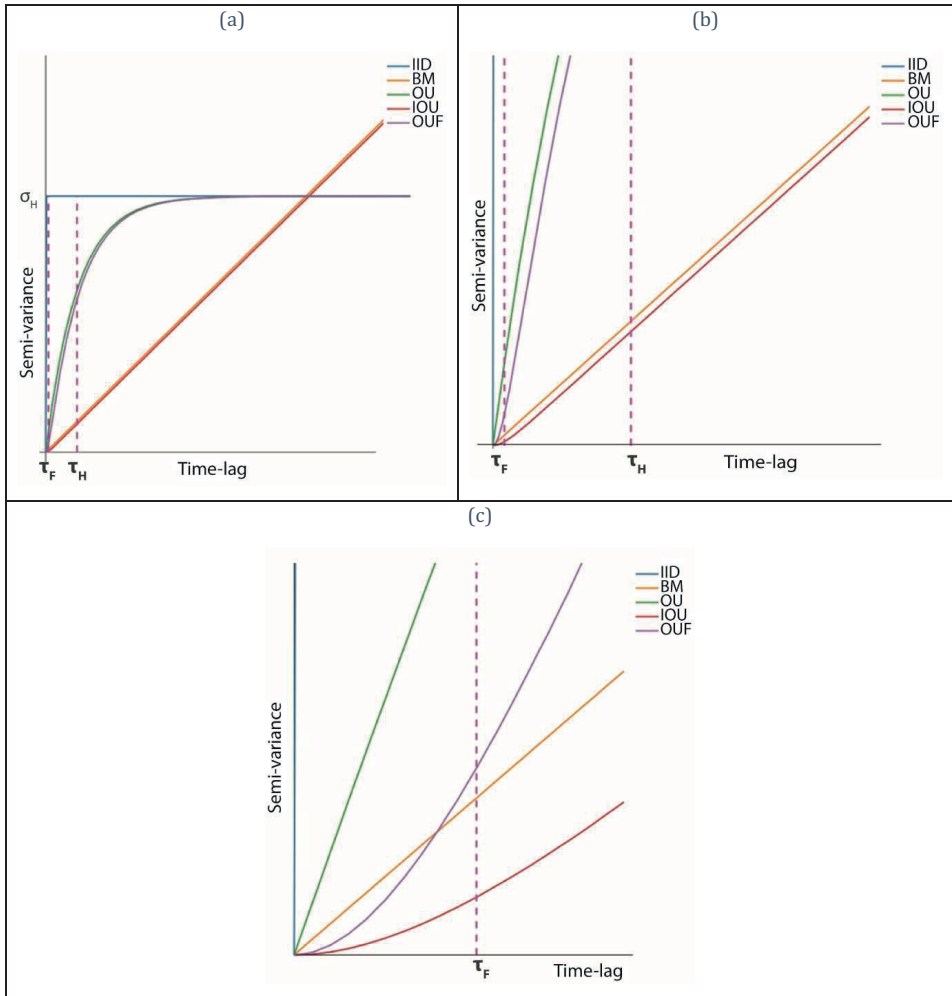


Figure 3. An exemplary illustration of CTSP models (in different zoomed-in views ranging from the general view (a) to more detailed views (b) and (c)). τ_F is the typical duration of a foraging bout. τ_H is the timescale associated with the animal covering its home range σ_H .

3 Methods and Results

This chapter summarizes the three research papers corresponding to the three stages of the research process introduced in chapter 1. For every paper the objectives, the methods, the main contributions, and the key findings are highlighted to provide a basis for the subsequent discussion chapter. However, this does not provide a substitute for reading the full papers. Studying the full papers provided in Part II is recommended.

3.1 Research Paper I

Teimouri, M., Indahl, U. G., & Tveite, H. (2016). A method to detect inactive periods in animal movement using density-based clustering. *Applied Geography*, 73, 102-112.

Research Paper I responds to Research Objective 1 and Research Questions 1-2 of the thesis. In Research Paper I, we focus on the implication of positioning error for animal movement studies over a range of sampling. When an animal such as sheep is generally inactive (standing immobile, grooming, ruminating or sleeping), the same location coordinate should be recorded for that time period. However, positioning errors create variations in the recorded locations, potentially leading to erroneous biological interpretation (30). The presence of positioning errors leads to bias in distance and velocity, and can lead to misclassification of animal activity (50).

3.1.1 Objectives

Segmenting animal trajectories by detecting Inactive Periods in the presence of positioning error is the main goal of Research Paper I.

Research Paper I aims to provide a framework to fulfill the following goals:

- Spatiotemporal clustering of points recorded during Inactive Periods
- Estimation of positioning error

Accordingly, the research framework includes two tasks as follows:

- An adapted version of the DBSCAN method for spatiotemporal clustering
- Estimation of positioning error with the aid of the Rayleigh distribution

3.1.2 Methodology

Due to positioning errors, an Inactive Period (IP) is characterized by the recorded positions being scattered around the true resting position. Therefore, detection of an IP resembles a stationary test, where the mean center of the recordings is an estimate of the stationary site (the true resting position), and its radius is determined by the positioning accuracy. We chose to use the *2drms* accuracy measure (See Section 2.1) to be able to statistically estimate measurement errors while detecting IPs. Each IP is considered to be a spatiotemporal dense cluster of recordings. Therefore, we take advantage of a density-based clustering method called DBSCAN (71). DBSCAN finds a group of positions located close together in Euclidean space and well separated from other positions with respect to the same distance measure. The user input of DBSCAN is the magnitude of the neighborhood (defined by a given radius, Eps) and a minimum number of points (MinPts) in the neighborhood, i.e. the density in the neighborhood has to exceed some threshold. Our adapted version of DBSCAN, called DBCIP (Density-Based Clustering method for Inactive Periods) takes advantage of the trajectory aspect in our recordings. The DBCIP method combines the spatial and temporal dimensions into a joint neighborhood definition as follows: (i) the Euclidean distance (geographical proximity) between points (Eps_e), and (ii) the distance between points along the trajectory (Eps_t). Eps_t introduces spatiotemporal constraints in the clustering. The only user input parameter that DBCIP requires is the minimum duration of an IP.

Since $Eps_e=2drms$ is usually unknown in practice, it was necessary to include the Eps_e -estimation as a part of the suggested method. For the minimum number of recordings (derived from the user input of minimum duration of inactivity) and some suggested a priori value for the *2drms*, the temporal neighborhood in terms of distance along the trajectory threshold is estimated by Gaussian simulation. A feasible Eps_e is estimated while detecting IPs in a repetitive approach. In each iteration, the radial errors for the detected IPs are calculated. By fitting a Rayleigh distribution to the radial error values, we derive a confidence interval for the associated scale parameter. We relate the unknown Eps_e to the scale parameter of the Rayleigh distribution. The candidate value Eps_e (i.e. the a priori Eps_e , increased by a step size of e.g. 1 meter) is assessed within the confidence interval of the

scale parameter for the Rayleigh distribution. A feasible estimate of Eps_e is smaller than the lower limit of the confidence interval of the scale parameter for the Rayleigh distribution.

3.1.3 Main Findings

As long as the step lengths of the animal are long enough to be distinguishable from positioning errors, segmentation of the animal tracking trajectory is possible through IP detection.

- The proposed clustering method, DBCIP, can be successfully applied to detect Inactive Periods of an animal using only a temporally ordered sequence of position recordings.
- Assuming that the coordinates for each IP is drawn from some unknown bivariate normal distribution, the radial errors are Rayleigh distributed. DBCIP can be used to estimate a positioning error.
- The experimental results of applying DBCIP on sheep trajectories suggest that their normal daily inactivity pattern contains around one-hour inactivity periods during daytimes and a few hours inactivity period around midnights.

3.2 Research Paper II

Teimouri, M., Indahl, U. G., Sickel, H., & Tveite, H. (2018). Deriving Animal Movement Behaviors Using Movement Parameters Extracted from Location Data. *ISPRS International Journal of Geo-Information*, 7(2), 78.

To respond to Research Objective 2 of the thesis and Research Questions 3-4 of the thesis, we develop a method to distinguish movement behavioral states for trajectories with high sampling frequency in Research Paper II. We characterize different movement behaviors using both the distribution of speed and turning angle values.

With respect to the distribution of speed and turning angle values, movement behavior can be grouped into four categories: (i) resting (low mean turning angle and low mean speed), (ii) walking (low mean turning angle and high mean speed), (iii) foraging (high mean turning angle and low mean speed), and possibly (iv) undefined movement behavior (high mean turning angle and high mean speed).

3.2.1 Objectives

Given a sequence of locations of an animal sampled at high frequency, we aim to detect different movement behaviors along the trajectory. First, we intend to segment the movement trajectory according to potential changes in the underlying movement behavior. Then, segments that are assumed to reflect similar underlying behavior are grouped together.

Accordingly, the research framework includes the following two tasks:

- Trajectory segmentation
- Sub-trajectory clustering

The way that we do each task is based upon a review of the corresponding literature. We concluded that a modified version of BCPA is the most appropriate trajectory segmentation method for our auto-correlated high frequency data. However, BCPA, like many trajectory segmentation methods, only provides information on significant change-points along the trajectory, leading to trajectory segmentation. Assuming that the number of distinct movement behaviors is an unknown number of one to four movement behaviors (as introduced earlier), we suggest using a hierarchical clustering method rather than k-means in order to consider the overall distribution of the movement parameter values in the clustering process, rather than a specific statistic like central tendency or median as used in (108) for the k-means clustering method.

3.2.2 Methodology

The proposed framework for detecting movement behavioral states according to changes in the movement parameters consists of two parts: (i) using a slightly modified version of the recently developed behavioral change point analysis (BCPA) method (83), we partition an individual trajectory into segments of homogeneous movement parameters. (ii) using the Kolmogorov–Smirnov distance metric, we compute a distance matrix for all pairs of segments, in order to generate an agglomerative hierarchical clustering model of the segments (109).

BCPA requires a moving window size as input, and determining this is a challenge. To identify a good value for the size of the moving window, exploring diagnostic plots of the residuals, comparing them to a standard normal distribution and inspecting their autocorrelation function in a trial-and-error fashion is suggested in Research Paper II.

'Persistence velocity' is typically used as a response variable in BCPA. We suggest using 'Turning velocity' as another response variable as well in order to improve the reliability of BCPA in the detection of significant change-points along the trajectory. We do not have any a priori assumptions regarding the number of change points. Since we expect that there are shared change points between the two resulting sets of change points, we combine the two sets of change points and remove duplicates. Using the resulting change points, we partition the trajectory into segments.

To group the segments according to behavior based on consistency in their movement parameters, we suggest hierarchical clustering (110, 111) with the Kolmogorov-Smirnov (ks) distance (112) in Research Paper II.

Unlike the segmentation part of our method, in which both 'Persistence velocity' and 'Turning velocity' are used, the clustering part of our method uses only the 'Persistence velocity' values. 'Persistence velocity' values can discriminate between the predefined behaviors mentioned above: walking is characterized by high persistence velocity values, foraging by low values, and resting by intermediate values. Additionally, investigating the directional histogram of turning angle values for each class of behavior increased our confidence in the classification.

In Research Paper II, we walk through the main analysis steps with data for a single sheep. However, we demonstrate the performance of our method using all the data (for the four available sheep) for which corresponding field observations are available.

3.2.3 Main Findings

As long as the movement behaviors of the animal in question can be distinguished by a combination of speed and turning angle values, our method for behavior detection is applicable.

A confusion matrix is used to quantify the accuracy of behavior detection for our methodology in terms of the number of minutes of correctly predicted movement divided by the whole period of observation.

- Analysis of seven GPS recording days where corresponding observed movement behavior data are available indicates an average accuracy for our dataset of 80.75% \pm 5.9% (SE).
- The method for discriminating between different animal behaviors requires neither a temporal regularization of the data nor a strict a priori value for the number of movement behavioral states, nor the specification of the movement model.

Adding context to the analysis would improve movement behavior prediction. Therefore, a future extension of our proposed approach should consider the incorporation of contextual information (e.g. environmental conditions) for the assessment of similarity between the movement characteristics of objects.

3.3 Research Paper III

Teimouri, M., Indahl, U. G., Sickel, H., & Tveite, H. (2018). Estimation of Temporal Sampling Frequency in Movement Parameter Estimation: A Stochastic Movement Modeling Approach. *Submitted to Ecological Applications*

Research Paper III responds to Research Objective 3 and Research Questions 2 and 4-5 of the thesis. The final stage of this thesis intends, firstly, to model animal movement paths stochastically in a continuous-time domain, secondly, to evaluate the sensitivity of the movement parameters derived from discrete-time movement paths to the sampling frequency of the observations, and thirdly, to find the best sampling frequency in terms of minimizing the impact of errors on the derived movement parameter(s).

3.3.1 Objectives

Sampling points from animal movement tracking can be connected by straight lines to reconstruct a discrete-time animal movement path. The accuracy of the reconstructed movement path depends on several factors, including temporal sampling frequency, accuracy of the measuring device and species-specific movement characteristics. Given a short observation period with high temporal sampling frequency, we can stochastically

model the animal movement path. The animal movement path can be viewed as a realization of a stochastic process that takes into account the accuracy of the locations and species-specific movement characteristics. The impact of sampling frequency can then be assessed through a cross-scale analysis. Therefore, our main research question is as follows: Given an animal species and a device for location measurement, what is the optimal temporal sampling frequency when the goal is that the reconstructed movement path shall have the least difference from the true movement path in terms of the estimated movement parameter (for example speed)?

3.3.2 Methodology

Our method for assessing the impact of temporal sampling frequency on movement parameter estimation is applicable as long as the following requirements are met: (i) an observation period with high temporal frequency recordings of the movement path is available, and (ii) the movement behavior during the observation period is representative for the long term movement behavior.

By calculating the empirical variogram for the available high frequency sampling dataset, the main features of the movement are revealed. The empirical variogram visualizes the autocorrelation structure and facilitates model identification. The empirical variogram is fitted to an acceptable semi-variance model selected from a palette of models of CTSP (continuous-time stochastic process). The selected parameterized stochastic movement model can be used as a simulator across different (sub)sampling schedules to figure out the impact of sampling frequency on the movement parameter in question.

Since we lack the true movement path for our experiment, we split the dataset into training and test datasets. We do model fitting based on the training trajectory dataset and figure out the impact of sampling frequency on the movement parameter in question by a cross-scale analysis based on the test dataset.

We compare statistically the simulated trajectories with the corresponding observed trajectory in terms of the movement parameter (here speed) over a range of sampling frequencies and find the Representative Temporal Sampling Frequency (RTSF). With RTSF, the difference between the true movement path and the reconstructed path in terms of movement parameter estimation is minimized.

3.3.3 Main Findings

- At high temporal sampling frequencies, location measurement errors cause overestimation of speed. At low temporal sampling frequencies, interpolation errors cause underestimation of speed.
- By viewing animal movement paths as realizations of a continuous-time stochastic process, we can incorporate the effects of position autocorrelation, velocity autocorrelation, range residency, measurement error and species-specific movement characteristics in the movement model selection.
- The proposed method is used to find a Representative Temporal Sampling Frequency (RTSF) that avoids both over- and under-sampling in terms of speed calculation.

4 Conclusion

This thesis followed a three-stage research process towards the main goals. The achievements of each stage contribute to the main goals of the thesis. The general aim of this chapter is to highlight the outcomes and the main contribution of our research and draw attention to some outlooks on future work. Finally, we mention how sheep farming may benefit from our outcomes.

4.1 Main Contributions

The first main outcome of this thesis is the development of quantitative methods to detect animal movement behavior for different temporal sampling frequencies of the observations. For observations with high temporal sampling frequency, detailed movement behaviors, such as foraging, resting and walking are distinguishable, while for observation with a lower temporal sampling frequency, only more generalized movement behaviors, such as Inactive and Active Periods are distinguishable.

To do animal movement behavior detection, we used existing standard clustering techniques, such as DBSCAN and hierarchical clustering, statistical reasoning based on the Rayleigh distribution and the Wilcoxon–Mann–Whitney test, and existing animal movement tools such as the BCPA and AdehabitatLT R packages.

We developed our generic methods based on sequences of location and time coordinates for animal movement paths. We investigated the applicability of the proposed methods on real movement datasets for sheep. This does not mean that the applicability of our methods is tied to the specific examples or equipment described herein. Our proposed methods for behavior detection provide useful tools for domain experts to gain insights into the behavior of animals in their habitat.

Finding a Representative Temporal Sampling Frequency (RTSF) that maximizes the monitoring period while providing guarantees for the accuracy of movement parameter estimation, is the second main contribution of this thesis. We studied the dependency of movement parameter estimation on different influencing factors, including sampling frequency, accuracy of the measuring device and species-specific movement characteristics. We proposed a novel method based on stochastic movement modeling

reflecting and quantifying measurement accuracy and species-specific movement characteristics. Then, we examined the impact of sampling frequency on movement parameter estimation. To the best of our knowledge, there is currently no method to quantify the contributions of all influencing factors, on movement parameter estimation. With our method, we tried to consider as many influencing factors as possible.

All in all, the main achievements of the thesis are:

1. Understanding the nature and implications of measurement errors on tracking paths and the derived movement parameters (Research Paper I & III).
2. Developing trajectory segmentation methods for different temporal sampling frequencies (Research Paper I & II).
3. Developing a similarity measure for animal movement behavioral state detection (Research paper II).
4. Studying the effects of the temporal sampling frequency on the tracking path and the derived movement parameters (Research paper III).
5. Assessing temporal sampling frequency requirements for specific research questions (Research paper III).

4.2 Practical view in terms of sheep farming

In Norwegian rangelands, sheep are not usually inspected regularly. In the case of disease and injury, they may be undiagnosed for a long period. Relying upon real-time monitoring and classification of animal movement behavior, we can detect deviations from normal behavior early and offer treatment on time to prevent more pain to the animal, or even animal loss.

The long-term objective of this work is to gain insight into spatiotemporal movement patterns of animals in their habitat. This can form the basis for developing sustainable sheep farming practices. For example, by proper placement of stock water and mineral licks (13, 113) or virtual fencing (114, 115), livestock movement patterns can be manipulated. Livestock can be attracted away from environmentally critical areas or into areas targeted for grazing in order to decrease the adverse effects from livestock in the grazed area, including biodiversity loss and soil erosion (13).

At a higher level, understanding when and where sheep move and discovering the movement patterns of animals within a habitat context is valuable for allocating resources or assessing the impact of utilization by livestock. With a particular emphasis on land management, the outcomes of this thesis can contribute to advance sheep grazing and increase the uniformity of grazing.

From a telemetry perspective, finding a sampling frequency for telemetry which respects animal welfare and provides sufficient information about animal movement is another achievement of this research that could benefit sheep farming in particular. In telemetry, high frequency sampling requires high battery capacity or frequent battery replacement. High battery capacity might impose a weight that exceeds the carrying capacity of the animal. Frequent battery replacement requires recapturing, which might be infeasible or pose ethical concerns about animal welfare.

All in all, our proposed methods can be used either solely in any decision making about sheep farming in particular or along with other information for further analyses in the animal study domain and in policy decision making.

4.3 Challenges for Future Research

The way that we detected different movement behaviors depended on the temporal sampling frequency of observations. We suggest that future research should do a comparative analysis of our methods for different temporal sampling frequencies. For example, the impact of different sampling rates on the detected Inactive Periods. What are the upper and lower limits for sampling rate for reliable detection of Inactive Periods using our proposed method? How high sampling frequency is needed to distinguish foraging, resting and walking behavior with acceptable accuracy?

We believe that by repeating our proposed methods on a reasonably large dataset, we can obtain a general view of the movement patterns of the animal in its habitat. Detection of abnormal behavior (that deviates from the normal movement patterns) could then be possible.

Future research could involve on-line monitoring of animals, in order to do real-time analyses on the data. Then we can monitor what is happening in the pasture and make on-time reaction to disturbances possible.

Assessing the impact of temporal sampling frequency on the movement parameter of speed and finding the RTSF for speed calculation is another major contribution of our research to the field of animal movement studies. The proposed method could be extended to assess the impact of sampling frequency on other movement parameters, for instance turning angle.

References

1. Kays R, Crofoot MC, Jetz W, Wikelski M. Terrestrial animal tracking as an eye on life and planet. *Science*. 2015;348(6240):aaa2478.
2. Bastille-Rousseau G, Potts JR, Yackulic CB, Frair JL, Ellington EH, Blake S. Flexible characterization of animal movement pattern using net squared displacement and a latent state model. *Movement ecology*. 2016;4(1):15.
3. Dray S, Royer-Carenzi M, Calenge C. The exploratory analysis of autocorrelation in animal-movement studies. *Ecological Research*. 2010;25(3):673-81.
4. Mobæk R. Density dependent foraging ecology and performance of domestic sheep on alpine ranges [Doctoral thesis]: Norwegian University of Life Sciences (NMBU); 2012.
5. Sickel H, Ihse M, Norderhaug A, Sickel MA. How to monitor semi-natural key habitats in relation to grazing preferences of cattle in mountain summer farming areas: An aerial photo and GPS method study. *Landscape and Urban Planning*. 2004;67(1):67-77.
6. Asheim L, Mysterud I. The Norwegian sheep farming production system. *Options Méditerranéennes Série A: Séminaires Méditerranéens (CIHEAM)*. 1999.
7. Mabile G, Stien A, Tveraa T, Mysterud A, Brøseth H, Linnell JDC. Sheep farming and large carnivores: What are the factors influencing claimed losses? *Ecosphere*. 2015;6(5):1-17.
8. Bailey DW. Identification and creation of optimum habitat conditions for livestock. *Rangeland Ecology & Management*. 2005;58(2):109-18.
9. Bailey D. Management strategies for optimal grazing distribution and use of arid rangelands. *Journal of Animal Science*. 2004;82(13_suppl):E147-E53.
10. Steinfeld H, De Haan C, Blackburn H. Livestock-environment interactions. *Issues and Options WRENmedia*, Suffolk. 1997.
11. Austrheim G, Asheim L, Bjarnason G, Feilberg J, Fosaa A, Holand Ø, et al. Sheep grazing in the North Atlantic region—a long term perspective on management, resource economy and ecology. *Rapport zoologisk serie*. 2008;3:86.
12. Yuill RS. The standard deviational ellipse; an updated tool for spatial description. *Geografiska Annaler: Series B, Human Geography*. 1971;53(1):28-39.
13. Ganskopp D, George M, Bailey D, Borman M, Surber G, Harris N. Factors and practices that influence livestock distribution. *University of California Division of Agriculture and Natural Resources*. 2007;8217:20.
14. Frair JL, Fieberg J, Hebblewhite M, Cagnacci F, DeCesare NJ, Pedrotti L. 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*. 2010;365(1550):2187-200.
15. Morales JM, Haydon DT, Frair J, Holsinger KE, Fryxell JM. Extracting more out of relocalational data: Building movement models as mixtures of random walks. *Ecology*. 2004;85(9):2436-45.
16. Gurarie E, Bracis C, Delgado M, Meckley TD, Kojola I, Wagner CM. What is the animal doing? Tools for exploring behavioural structure in animal movements. *Journal of Animal Ecology*. 2016;85(1):69-84.
17. Laube P, Purves RS. How fast is a cow? cross-scale analysis of movement data. *Transactions in GIS*. 2011;15(3):401-18.
18. Alexander G, Maritz B. Sampling interval affects the estimation of movement parameters in four species of African snakes. *Journal of Zoology*. 2015;297(4):309-18.
19. Edelhoff H, Signer J, Balkenhol N. Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns. *Movement Ecology*. 2016;4(1):21.
20. Horne JS, Garton EO, Krone SM, Lewis JS. Analyzing animal movements using Brownian bridges. *Ecology*. 2007;88(9):2354-63.
21. Pride IG, Swift SM. *Wildlife telemetry: remote monitoring and tracking animals*: Ellis Horwood; 1992.

22. Bishop J, Last J. Global radionavigation and location systems for tracking marine mammals. *International Journal of Remote Sensing*. 1995;16(11):1943-56.
23. Adrados C, Verheyden-Tixier H, Cargnelutti B, Pépin D, Janeau G. GPS approach to study fine-scale site use by wild red deer during active and inactive behaviors. *Wildlife Society Bulletin*. 2003:544-52.
24. Yasuda T, Arai N. Fine-scale tracking of marine turtles using GPS-Argos PTTs. *Zoological science*. 2005;22(5):547-53.
25. Gudbjornsson S, Godo O, Palsson A. Mini GPS fish tags contributing to fisheries management. *Sea Technol*. 2004;45(6):23-7.
26. Fritz H, Said S, Weimerskirch H. Scale-dependent hierarchical adjustments of movement patterns in a long-range foraging seabird. *Proceedings of the Royal Society of London B: Biological Sciences*. 2003;270(1520):1143-8.
27. Kaplan E, Hegarty C. *Understanding GPS: principles and applications*: Artech house; 2005.
28. Jerde CL, Visscher DR. GPS measurement error influences on movement model parameterization. *Ecological Applications*. 2005;15(3):806-10.
29. Hurford A. GPS measurement error gives rise to spurious 180 turning angles and strong directional biases in animal movement data. *Plos One*. 2009;4(5):e5632.
30. Pépin D, Adrados C, Mann C, Janeau G. Assessing real daily distance traveled by ungulates using differential GPS locations. *Journal of Mammalogy*. 2004;85(4):774-80.
31. Dussault C, Courtois R, Ouellet J-P, Huot J. Influence of satellite geometry and differential correction on GPS location accuracy. *Wildlife Society Bulletin*. 2001:171-9.
32. Hulbert IA, French J. The accuracy of GPS for wildlife telemetry and habitat mapping. *Journal of Applied Ecology*. 2001;38(4):869-78.
33. D'Eon RG, Serrouya R, Smith G, Kochanny CO. GPS radiotelemetry error and bias in mountainous terrain. *Wildlife Society Bulletin*. 2002:430-9.
34. Ganskopp DC, Johnson DD. GPS error in studies addressing animal movements and activities. *Rangeland Ecology & Management*. 2007;60(4):350-8.
35. Chin GY. Two-dimensional measures of accuracy in navigational systems. 1987.
36. Moen R, Pastor J, Cohen Y, Schwartz CC. Effects of moose movement and habitat use on GPS collar performance. *The Journal of wildlife management*. 1996:659-68.
37. Cain JW, Krausman PR, Jansen BD, Morgart JR. Influence of topography and GPS fix interval on GPS collar performance. *Wildlife Soc B*. 2005;33(3):926-34.
38. D'Eon RG, Serrouya R, Smith G, Kochanny CO. GPS radiotelemetry error and bias in mountainous terrain. *Wildlife Soc B*. 2002;30(2):430-9.
39. Di Orio AP, Callas R, Schaefer RJ. Performance of two GPS telemetry collars under different habitat conditions. *Wildlife Soc B*. 2003;31(2):372-9.
40. Moen R, Pastor J, Cohen Y, Schwartz CC. Effects of moose movement and habitat use on GPS collar performance. *J Wildlife Manage*. 1996;60(3):659-68.
41. Bowman JL, Kochanny CO, Demarais S, Leopold BD. Evaluation of a GPS collar for white-tailed deer. *Wildlife Soc B*. 2000;28(1):141-5.
42. D'EON RG, Delparte D. Effects of radio-collar position and orientation on GPS radio-collar performance, and the implications of PDOP in data screening. *Journal of Applied Ecology*. 2005;42(2):383-8.
43. Hebblewhite M, Percy M, Merrill E. Are all global positioning system collars created equal? Correcting habitat-induced bias using three brands in the central Canadian Rockies. *The Journal of Wildlife Management*. 2007;71(6):2026-33.
44. Schwager M, Anderson DM, Butler Z, Rus D. Robust classification of animal tracking data. *Computers and Electronics in Agriculture*. 2007;56(1):46-59.
45. Jonsen ID, Myers RA, James MC. Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. *Marine Ecology Progress Series*. 2007;337:255-64.

46. Calenge C, Dray S, Royer-Carenzi M. The concept of animals' trajectories from a data analysis perspective. *Ecological informatics*. 2009;4(1):34-41.
47. Singh NJ, Allen AM, Ericsson G. Quantifying migration behaviour using net squared displacement approach: Clarifications and caveats. *Plos One*. 2016;11(3):e0149594.
48. Buchin M, Driemel A, Van Kreveld M, Sacristán V, editors. An algorithmic framework for segmenting trajectories based on spatio-temporal criteria. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*; 2010: ACM.
49. Giannotti F, Pedreschi D, Theodoridis Y, editors. Geographic privacy-aware knowledge discovery and delivery. *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*; 2009: ACM.
50. Ungar ED, Henkin Z, Gutman M, Dolev A, Genizi A, Ganskopp D. Inference of animal activity from GPS collar data on free-ranging cattle. *Rangeland Ecology & Management*. 2005;58(3):256-66.
51. Frair JL, Merrill EH, Visscher DR, Fortin D, Beyer HL, Morales JM. Scales of movement by elk (*Cervus elaphus*) in response to heterogeneity in forage resources and predation risk. *Landscape ecology*. 2005;20(3):273-87.
52. Van Moorter B, Visscher DR, Jerde CL, Frair JL, Merrill EH. Identifying movement states from location data using cluster analysis. *Journal of Wildlife Management*. 2010;74(3):588-94.
53. Grünewälder S, Broekhuis F, Macdonald DW, Wilson AM, McNutt JW, Shawe-Taylor J, et al. Movement activity based classification of animal behaviour with an application to data from cheetah (*Acinonyx jubatus*). *Plos One*. 2012;7(11):e49120.
54. Body G, Weladji R, Holand Ø. The recursive model as a new approach to validate and monitor activity sensors. *Behav Ecol Sociobiol*. 2012;66(11):1531-41.
55. Alvares LO, Bogorny V, Kuijpers B, de Macedo JAF, Moelans B, Vaisman A, editors. A model for enriching trajectories with semantic geographical information. *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*; 2007: ACM.
56. Spaccapietra S, Parent C, Damiani ML, de Macedo JA, Porto F, Vangenot C. A conceptual view on trajectories. *Data & knowledge engineering*. 2008;65(1):126-46.
57. Hornsby KS, Cole S. Modeling Moving Geospatial Objects from an Event-based Perspective. *Transactions in GIS*. 2007;11(4):555-73.
58. Andrienko GL, Andrienko NV, Fuchs G, Raimond A-MO, Symanzik J, Ziemlicki C, editors. Extracting Semantics of Individual Places from Movement Data by Analyzing Temporal Patterns of Visits. *COMP@ SIGSPATIAL*; 2013.
59. Palma AT, Bogorny V, Kuijpers B, Alvares LO. A Clustering-based Approach for Discovering Interesting Places in Trajectories. *Applied Computing 2008, Vols 1-3*. 2008:863-8.
60. Yan Z, Parent C, Spaccapietra S, Chakraborty D. A hybrid model and computing platform for spatio-temporal trajectories. *The Semantic Web: Research and Applications*: Springer; 2010. p. 60-75.
61. Phithakkitnukoon S, Horanont T, Di Lorenzo G, Shibasaki R, Ratti C. Activity-aware map: Identifying human daily activity pattern using mobile phone data. *Human Behavior Understanding*: Springer; 2010. p. 14-25.
62. Zimmermann M, Kirste T, Spiliopoulou M. Finding stops in error-prone trajectories of moving objects with time-based clustering. *Intelligent interactive assistance and mobile multimedia computing*: Springer; 2009. p. 275-86.
63. Buard E. Pratiques spatiales des populations animales: analyses par les trajectoires. *Actes de Dixièmes Rencontres de Théo Quant*. 2011.
64. Rocha JAM, Oliveira G, Alvares LO, Bogorny V, Times VC, editors. DB-SMoT: A direction-based spatio-temporal clustering method. *Intelligent systems (IS)*, 2010 5th IEEE international conference; 2010: IEEE.
65. Gong L, Sato H, Yamamoto T, Miwa T, Morikawa T. Identification of activity stop locations in GPS trajectories by density-based clustering method combined with support vector machines. *Journal of Modern Transportation*. 2015;23(3):202-13.

66. Ashbrook D, Starner T. Using GPS to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous computing*. 2003;7(5):275-86.
67. Agamennoni G, Nieto J, Nebot E, editors. Mining GPS data for extracting significant places. *Robotics and Automation, 2009 ICRA'09 IEEE International Conference on*; 2009: IEEE.
68. Palma AT, Bogorny V, Kuijpers B, Alvares LO, editors. A clustering-based approach for discovering interesting places in trajectories. *Proceedings of the 2008 ACM symposium on Applied computing*; 2008: ACM.
69. Kami N, Enomoto N, Baba T, Yoshikawa T, editors. Algorithm for detecting significant locations from raw GPS data. *International Conference on Discovery Science*; 2010: Springer.
70. Luo T, Zheng X, Xu G, Fu K, Ren W. An improved DBSCAN algorithm to detect stops in individual trajectories. *ISPRS International Journal of Geo-Information*. 2017;6(3):63.
71. Ester M, Kriegel H-P, Sander J, Xu X, editors. A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd*; 1996.
72. Ankerst M, Breunig MM, Kriegel H-P, Sander J, editors. OPTICS: ordering points to identify the clustering structure. *ACM Sigmod record*; 1999: ACM.
73. Gudmundsson J, Laube P, Wolle T. Computational movement analysis. *Springer handbook of geographic information*: Springer; 2011. p. 423-38.
74. Yoon H, Shahabi C, editors. Robust time-referenced segmentation of moving object trajectories. *Data Mining, 2008 ICDM'08 Eighth IEEE International Conference on*; 2008: IEEE.
75. Demšar U, Buchin K, Cagnacci F, Safi K, Speckmann B, Van de Weghe N, et al. Analysis and visualisation of movement: an interdisciplinary review. *Movement ecology*. 2015;3(1):5.
76. Gutenkunst R, Newlands N, Lutcavage M, Edelstein-Keshet L. Inferring resource distributions from Atlantic bluefin tuna movements: An analysis based on net displacement and length of track. *Journal of theoretical biology*. 2007;245(2):243-57.
77. Tremblay Y, Roberts AJ, Costa DP. Fractal landscape method: an alternative approach to measuring area-restricted searching behavior. *Journal of Experimental Biology*. 2007;210(6):935-45.
78. Postlethwaite CM, Brown P, Dennis TE. A new multi-scale measure for analysing animal movement data. *Journal of Theoretical Biology*. 2013;317:175-85.
79. Fauchald P, Tveraa T. Using first-passage time in the analysis of area-restricted search and habitat selection. *Ecology*. 2003;84(2):282-8.
80. Kareiva P, Shigesada N. Analyzing insect movement as a correlated random walk. *Oecologia*. 1983;56(2):234-8.
81. Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns.
82. Gurarie E. Models and analysis of animal movements: From individual tracks to mass dispersal [Doctoral Thesis]: University of Washington; 2008.
83. Gurarie E, Andrews RD, Laidre KL. A novel method for identifying behavioural changes in animal movement data. *Ecology letters*. 2009;12(5):395-408.
84. Schwarz G. Estimating the dimension of a model. *The annals of statistics*. 1978;6(2):461-4.
85. Ranacher P, Tzavella K. How to compare movement? A review of physical movement similarity measures in geographic information science and beyond. *Cartography and geographic information science*. 2014;41(3):286-307.
86. Dodge S, Laube P, Weibel R. Movement similarity assessment using symbolic representation of trajectories. *Int J Geogr Inf Sci*. 2012;26(9):1563-88.
87. Vlachos M, Kollios G, Gunopulos D, editors. Discovering similar multidimensional trajectories. *Data Engineering, 2002 Proceedings 18th International Conference on*; 2002: IEEE.
88. Berndt DJ, Clifford J, editors. Using dynamic time warping to find patterns in time series. *KDD workshop*; 1994: Seattle, WA.
89. Chen L, Özsu MT, Oria V, editors. Symbolic representation and retrieval of moving object trajectories. *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*; 2004: ACM.

90. Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, et al. A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*. 2008;105(49):19052-9.
91. Marsh LM, Jones RE. The form and consequences of random walk movement models. *Journal of Theoretical Biology*. 1988;133(1):113-31.
92. Codling EA, Plank MJ, Benhamou S. Random walk models in biology. *Journal of the Royal Society Interface*. 2008;5(25):813-34.
93. Bovet P, Benhamou S. Spatial analysis of animals' movements using a correlated random walk model. *Journal of theoretical biology*. 1988;131(4):419-33.
94. Patlak CS. Random walk with persistence and external bias. *Bulletin of Mathematical Biology*. 1953;15(3):311-38.
95. Bartumeus F, da Luz ME, Viswanathan G, Catalan J. Animal search strategies: A quantitative random-walk analysis. *Ecology*. 2005;86(11):3078-87.
96. Turchin P. *Quantitative Analysis of Movement: measuring and modeling population redistribution in plants and animals*. Sinauer Associates, Sunderland, MA; 1998.
97. Michalet X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Physical Review E*. 2010;82(4):041914.
98. Almeida PJ, Vieira MV, Kajin M, Forero-Medina G, Cerqueira R. Indices of movement behaviour: conceptual background, effects of scale and location errors. *Zoologia (Curitiba)*. 2010;27(5):674-80.
99. Fleming CH, Calabrese JM, Mueller T, Olson KA, Leimgruber P, Fagan WF. From fine-scale foraging to home ranges: a semivariance approach to identifying movement modes across spatiotemporal scales. *The American Naturalist*. 2014;183(5):E154-E67.
100. Fleming CH, Calabrese JM, Mueller T, Olson KA, Leimgruber P, Fagan WF. Non-Markovian maximum likelihood estimation of autocorrelated movement processes. *Methods in Ecology and Evolution*. 2014;5(5):462-72.
101. Johnson DS, London JM, Lea M-A, Durban JW. Continuous-time correlated random walk model for animal telemetry data. *Ecology*. 2008;89(5):1208-15.
102. Fagan WF, Calabrese JM. The correlated random walk and the rise of movement ecology. *The Bulletin of the Ecological Society of America*. 2014;95(3):204-6.
103. Calabrese JM, Fleming CH, Gurarie E. ctm: an r package for analyzing animal relocation data as a continuous-time stochastic process. *Methods in Ecology and Evolution*. 2016;7(9):1124-32.
104. Pozdnyakov V, Meyer T, Wang Y-B, Yan J. On modeling animal movements using Brownian motion with measurement error. *Ecology*. 2014;95(2):247-53.
105. Dunn JE, Gipson PS. Analysis of radio telemetry data in studies of home range. *Biometrics*. 1977:85-101.
106. Gurarie E, Ovaskainen O. Characteristic spatial and temporal scales unify models of animal movement. *The American Naturalist*. 2011;178(1):113-23.
107. Shamoun-Baranes J, van Loon EE, Purves RS, Speckmann B, Weiskopf D, Camphuysen CJ. Analysis and visualization of animal movement. *The Royal Society*; 2011.
108. Zhang J, O'Reilly KM, Perry GL, Taylor GA, Dennis TE. Extending the functionality of behavioural change-point analysis with k-means clustering: a case study with the little penguin (*eudyptula minor*). *Plos One*. 2015;10(4):e0122811.
109. Han J, Lee J-G, Kamber M. An overview of clustering methods in geographic data analysis. *Geographic data mining and knowledge discovery*. 2009;2:149-70.
110. Bridges Jr CC. Hierarchical cluster analysis. *Psychological reports*. 1966;18(3):851-4.
111. Xu R, Wunsch D. Survey of clustering algorithms. *IEEE Transactions on neural networks*. 2005;16(3):645-78.
112. Rachev ST, Klebanov LB, Stoyanov SV, Fabozzi FJ. Probability distances and probability metrics: definitions. *The Methods of Distances in the Theory of Probability and Statistics*: Springer; 2013. p. 11-31.

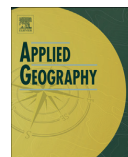
113. Ganskopp D. Manipulating cattle distribution with salt and water in large arid-land pastures: a GPS/GIS assessment. *Applied Animal Behaviour Science*. 2001;73(4):251-62.
114. Umstatter C. Review: The evolution of virtual fences: A review. *Comput Electron Agric*. 2011;75(1):10-22.
115. Anderson DM. Virtual fencing—past, present and future1. *The Rangeland Journal*. 2007;29(1):65-78.

Errata

Side	Line	Original text	Corrected text
Page # 2	3	on open forested	in open forests
Page # 2	19	Sustainable of forage plants	Sustainable production of forage plants
Page #4	18	Our interest is	Our interest was
Page #4	23	how movement parameter estimation vary	how movement parameter estimation varies
Page # 9	3	describe the aspects trajectories	describe the different aspects of trajectories
Page #10	Whole second and third paragraph	Wrong font ('Times New Roman')	Changed font to 'Cambria'
Page #17	3	Our Research paper III is	Our Research paper II is
Page # 23-24	Figure 3	Figure 3 contains 3 sub-Figures which were out of place in one page	I changed the legend font and layout of Figure 3 to make it fit in one page
Page # 23-24	Figure3 Caption	(in different zoomed-in views)	(in different zoomed-in views ranging from the general view (a) to more detailed views (b) and (c))

Research Paper I

Teimouri, M., Indahl, U. G., & Tveite, H. (2016). A method to detect inactive periods in animal movement using density-based clustering. *Applied Geography*, 73, 102-112.



A method to detect inactive periods in animal movement using density-based clustering



Maryam Teimouri*, Ulf Geir Indahl, Håvard Tveite

Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences (NBMU), N-1432, Norway

ARTICLE INFO

Article history:

Received 18 November 2015

Received in revised form

3 June 2016

Accepted 24 June 2016

Available online 7 July 2016

Keywords:

Animal tracking

Inactive periods

Positioning error

Density-based clustering method

ABSTRACT

In this paper, we propose a method to find inactive periods of a trajectory and employ it to livestock tracking.

In contrast to the existing methods to find inactive periods in the domain of animal movement studies, the proposed method estimates inactive periods based on the position recordings only, without involving information from activity sensors or field observations. The only parameter that the proposed method requires is the minimum duration of inactivity. Inactivity means being stationary or having limited variation in position. The results have been verified by applying the method to a dataset where activity sensor recordings are also available.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Recent developments in location-aware technologies (including GNSS “Global Navigation System Satellite” (e.g. (Awange, 2012)), RFID “Radio Frequency Identification” (e.g. (Kritzler, Raubal, & Krüger, 2007)), and cell phone tracking (e.g. (Abedi, Bhaskar, & Chung, 2014; Versichele, Neutens, Delafontaine, & Van de Weghe, 2012)) have significantly increased their use in data collection for moving object applications. Analyses and methods for extracting useful information from these increasingly large dataset have lagged behind the technology for generating them (Long & Nelson, 2013).

Wearable tracking collars have simplified monitoring of animal locations for many research projects in within ecology (Cagnacci, Boitani, Powell, & Boyce, 2010) and geography (Laube & Purves, 2011; Stewart, Nelson, Wulder, Nielsen, & Stenhouse, 2012; Technitis, Othman, Safi, & Weibel, 2015). Tracking an animal using tracking collars results in a trajectory which is a sequence of ordered records in time depicting the movement of the object (Gudmundsson, Laube, & Wolle, 2011; Long & Nelson, 2013). However, the so-called raw trajectory lacks semantic interpretation (Bogorny, Renso, Aquino, Lucca Siqueira, & Alvares, 2014).

According to Spaccapietra et al. (2008), a raw trajectory may semantically be segmented into ‘move’ and ‘stop’ parts, an event-based perspective (Hornsby & Cole, 2007). In the conceptual framework of an application, the spatial range of the trajectory for each ‘stop’ part is a single point (Spaccapietra et al., 2008). In animal movement studies, ‘stop’ and ‘move’ parts of the raw trajectory correspond to inactive periods (e.g., resting bouts) and active periods (e.g., foraging bouts) (Frair et al., 2010; Schwager, Anderson, Butler, & Rus, 2007).

Segmenting animal trajectories by detecting inactive periods in the presence of positioning error is the main goal of this paper. Positioning error is still considered an important concern in different applications, including animal movement studies ((Ganskopp & Johnson, 2007; Hurford, 2009; Jerde & Visscher, 2005), to name but a few). When an animal sleeps or lies down and rests at one location for a period, the collected positions can be misplaced, due to positioning error, complicating the detection of inactivity. Undetected periods of inactivity may lead to biological misinterpretation (Pepin, Adrados, Mann, & Janeau, 2004), e.g. degrade the estimation of spatial habitat use pattern or the evaluation of energetic requirement of animals.

In this paper, we investigate how well active and inactive periods of animals can be distinguished relying only on positions from non-differentially corrected GPS ‘Global Positioning System’ recordings. Positioning error is generally not available for the individual recordings, so we suggest a method for estimating the positioning error for the detected inactive periods. Due to the lack

* Corresponding author.

E-mail addresses: maryam.teimouri@nmbu.no (M. Teimouri), ulf.indahl@nmbu.no (U.G. Indahl), havard.tveite@nmbu.no (H. Tveite).

of empirical evidence (e.g., field observations) to validate the results, findings of the study are verified by utilizing available activity sensor data. A simulation is also conducted to evaluate the percentage of the inactive periods detected correctly by the proposed method.

The paper is structured as follows. In Section 2, we provide a brief literature review on methods for extracting inactive periods of GPS-tracked animals. In Section 3 (Methodology), a detailed description of the proposed method is given. In Section 4 (Case study), an explanation of our equipment to track domestic sheep on an alpine range is given. In Section 5 (Result), we apply the proposed method to our dataset and evaluate the outcomes based on the available activity sensor recordings. In Section 6 (Validation), we use a Random Walk model and a Correlated Random Walk model to simulate animal paths and evaluate what percentage of the simulated inactive periods that can be detected by the proposed algorithm.

2. Background and related works

An *Inactive Period (IP)* is defined as a state where the animal is stationary or has limited variation in geographic location. With error-free location data, defining a threshold value for the variation in geographic location would be sufficient to detect *IPs*. However, with errors in the location data, more sophisticated approaches are required to identify *IPs*.

In the literature of animal movement studies, inactive periods have been identified by combining activity sensor values and movement variables based on GPS recordings (Frair et al., 2005, 2010; Ganskopp & Johnson, 2007; Pepin et al., 2004; Schwager et al., 2007; Ungar et al., 2005; Van Moorter, Visscher, Jerde, Frair, & Merrill, 2010). Inactive periods (bedded, standing) and active states (feeding, moving) of animals have been identified using predictive models by including information from various sensors along with field observations (Body, Weladji, & Holand, 2012; Grunewalder et al., 2012).

Gervasi, Brunberg, & Swenson (2006)) tested an individual-based method to discriminate active and inactive periods for brown bears using dual-axis motion sensors mounted on GPS (Global Positioning Systems) collars. The motion sensors mounted on the collars “separately measure the true acceleration of the collar in 2 orthogonal directions 6–8 times per second”. The acceleration values acquired between 2 consecutive recordings were averaged, ranging from 0 to 255, and assigned to each direction. They found that the frequency distribution of motion sensor values was bimodal, and they identified a separation point and considered all activity values lower than the separation point as inactive.

Ganskopp & Johnson (2007) tried to select an activity sensor value threshold between active and inactive periods based on left-right sensor values of GPS collars worn by cattle. The activity sensor values, ranging from 0 to 255, were acquired at 5-min intervals. They visually identified the first point of inflection of the cumulative frequency curve of increasing activity values as a break point separating inactive and active periods. They found out that they needed to combine the activity value and a minimum distance threshold to filter out inactive periods from active periods.

Schwager et al. (2007) used the k-mean classification algorithm to categorize tracking data from cows into two groups corresponding to active and inactive periods. They used position and head angle data to demonstrate how the algorithm can be employed in a behavioral study.

Adrados, Baltzinger, Janeau, & Pepin (2008) proposed an individual-based relative method using the count provided by a GPS collar activity sensor to separate active from inactive periods. A dataset from free-ranging red deer (*Cervus elaphus*) was used to

pinpoint locations where animals were inactive versus active. For each individual and day of measurement, the mean activity during a 24-hr period was defined as the referential slope (a_0). Then the slope of each pair of successive activity values (a) was compared with (a_0). The animal was considered inactive during the time interval under consideration if $a < a_0$.

To our knowledge, most prior studies in this area have been conducted based on the existence of activity sensor values and field observations. With only location data available, it is a challenge to discriminate between active and inactive periods. Due to the lack of sufficiently fine resolution of activity sensor values and field observations in our data (see the case study section for details about the current dataset), we were interested in developing methods that only requires location data.

In the study of movement data, which is the forefront of Geographic information science research (Alvares et al., 2007; Andrienko et al., 2013; Long & Nelson, 2013; Palma, Bogorny, Kuijpers, & Alvares, 2008; Tran, Nguyen, Do, & Yan, 2011; Zimmermann, Kirste, & Spiliopoulou, 2009), the term ‘stop’ is used as a synonym for an inactive period (*IP*). Many analytical approaches in this area rely on geometrical properties and movement parameters of the trajectory, such as distance, duration and speed. The Euclidean distances between recordings in a fixed time window could be used to detect *IPs*, but identifying appropriate values for the Euclidean distances and time window length is challenging. In addition, such an approach is sensitive to outliers. Another approach could be to use the average Euclidean distance in a time window (Laube & Purves, 2011), but that reduces the sensitivity to the spatial distribution of the recordings.

Andrienko et al. (2013) classify stop detection methods based on movement parameters such as speed (Palma et al., 2008; Yan, Parent, Spaccapietra, & Chakraborty, 2010), distance (Phithakkitnukoon, Horanont, Di Lorenzo, Shibasaki, & Ratti, 2010), speed and duration (Zimmermann et al., 2009), speed and distance (Buard, 2011) and direction (Rocha et al., 2010). Among the different approaches to detect stops, density based clustering methods has attracted the attention of many researchers (Palma et al., 2008; Tran et al., 2011; Zimmermann et al., 2009). They develop adapted versions of ‘Density-Based Spatial Clustering of Applications with Noise’ (DBSCAN) (Ester et al., 1996). Based on a given set of points, DBSCAN cluster points that are located closely together in space. Our DBSCAN based method is original in the sense that it takes particular advantage of the trajectory aspect in our recordings (dataset) and detect *IPs* using a combined time-space distance measure.

Birant & Kut (2007) presented a spatio-temporal clustering method based on DBSCAN named ST-DBSCAN. This method has the ability to discover clusters according to non-spatial, spatial and temporal values of objects while DBSCAN finds clusters according to only the spatial values of objects. However, ST-DBSCAN handles spatiotemporal data stored as temporal slices, and it is not suitable for trajectory data. Spatial and temporal distances are defined separately, and the similarity of objects are defined by a conjunction of the two metrics.

Zimmermann et al. (2009) propose an interactive density-based clustering algorithm based on OPTICS (Ankerst, Breunig, Kriegel, & Sander, 1999) to discover stops in a trajectory. The density is defined based on the spatial and temporal properties of the trajectory and the potential stops are extracted interactively, which requires some domain knowledge. The advantage of the method is its applicability for trajectories of different transportation modes, so-called heterogeneous trajectories.

More relevant to our idea is the approach (CB-SMoT) of Palma et al. (Palma et al., 2008) suggesting a spatio-temporal clustering method for finding interesting places on the trajectory of a moving

object. Based on CB-SMoT, one can say that there are parts of a trajectory (i.e., interesting places) where the speed of the moving object is lower than in other parts of the same trajectory. However, finding this speed threshold is a non-trivial task.

3. Detection of inactive periods (IPs)

In order to detect IPs, one needs first to determine the magnitude of the positioning error for the collected location data. A *stationary test* is often used to estimate positioning error for animal studies (Lewis, Rachlow, Garton, & Vierling, 2007). A *stationary test* is usually based on a GPS collar of a known position that is left in place for a given period. Precision determines to what extent the observations are close to each other while accuracy defines the distribution of the difference between the “true” location and the measured locations. Therefore, if the true value of a collar position is known, the accuracy of the collar in terms of position estimation can be quantified; otherwise, only the precision can be estimated and hence quantified.

Several different accuracy measures have been explored in various applications. 2drms (twice Distance Root Mean Squared), CEP (Circular Error Probability) and R95 (an extension of CEP) are the most commonly used two-dimensional accuracy measures (Chin, 1987). CEP is the radius of the smallest circle centered at the true position that covers 50% of the observations. R95 is a similarly defined quantity covering 95% of the observations. The 2drms (Twice Distance Root Mean Squared) is defined as two times the Root-Mean-Square (RMS) of the radial errors as follows:

$$2drms = 2 * \sqrt{\frac{\sum_{i=1}^k \left((x_i - \mu_x)^2 + (y_i - \mu_y)^2 \right)}{k}} \quad (1)$$

Where k is the number of recorded positions included in a *stationary test*. Here $\{(x_i, y_i) | i = 1, \dots, k\}$ represent coordinates of k recorded positions for each *stationary test* and (μ_x, μ_y) is the true stationary site.

Experience with *stationary tests* has shown that the accuracy of collars depends on different criteria such as the topography of the area (Cain, Krausman, Jansen, & Morgart, 2005; D’eon, Serrouya, Smith, & Kochanny, 2002), habitat selection (Di Orio, Callas, & Schaefer, 2003; Moen, Pastor, Cohen, & Schwartz, 1996), and collar orientation (Bowman, Kochanny, Demarais, & Leopold, 2000; D’eon & Delparte, 2005) in addition to GPS intrinsic error sources (for example atmospheric conditions). Hansen & Riggs (2008) point out the importance of checking out the manufacturers’ GPS collar programming criteria for acquisition time before employing them in the field because longer satellite-acquisition times for GPS collars generally results in higher precision but reduced battery life.

Based on terrestrial deployments, Frair et al. (2010) conclude that GPS imprecision is up to 30 m on average, and the factors that contribute most to errors are canopy closures and unpredictable animal interactions with the local habitats. Cargnelutti et al. (2007) mention that GPS positioning error is highly variable, even for a given PDOP (Positional Dilution of Precision) value and a given number of tracked satellites.

Due to positioning errors, an IP is characterized by the recorded positions being scattered around the true resting position. Therefore, detection of an IP resembles a *stationary test*, where the mean center (\bar{X}, \bar{Y}) of the recordings is an estimate of the stationary site (the true resting position). A circle centered on the mean center with a radius of 2drms is then expected to contain between 95 and 98% (Specht & Szot, 2012) of the recordings for an IP. By choosing 2drms for measuring accuracy, we relate the unknown positioning error to the Rayleigh distribution of the radial errors for the

detected IPs (see formula 5 below).

In order to detect IPs from a trajectory, we will start by assuming that the positioning error (expressed as 2drms) is known. Later a technique for estimating the positioning error is presented as part of the IP detection process.

We propose two methods for detection of IPs for individual animals: i) *the naïve method* and ii) *the Density Based Clustering method for Inactive Periods* (DBCIP). The user-specified minimum duration of inactivity (T) is the only user input required for the DBCIP method, whereas the naïve method needs 2drms as well.

Both methods are designed to work on *regular trajectories*, characterized by a constant time lag between successive recordings (SI) (Calenge, 2006).

Given the user-specified minimum duration of inactivity (T) and the known fixed sampling interval (SI), the associated minimum number of recordings (N) required to be in sufficient spatio-temporal proximity is calculated as follows:

$$N = \left\lfloor \frac{T}{SI} \right\rfloor + 1 \quad (2)$$

Here $\lfloor \cdot \rfloor$ denotes rounding down to the nearest integer.

3.1. The naïve method

The naïve method detects IPs for a trajectory of n time ordered recordings $\mathbf{P} = \{p_1, p_2, \dots, p_n\}$ (where each p_i is a point) based on two input parameters: i) positioning error (2drms) and ii) the minimum duration of inactivity (T). Pseudo code for the naïve method is provided in appendix A. The naïve method shifts a time window of width T through the recordings in \mathbf{P} . An *initial IP* is defined as a sequence of points in the neighborhood of p_i ($p_{i-j}, \dots, p_{i-1}, p_i, p_{i+1}, \dots, p_{i+j}$) where $j = (N-1)/2$ (here $N > 0$ should be an odd integer requiring the user’s choice of T to be an even multiple of SI according to formula 2) and $EDist_{max}(p_i) < 2drms$. $EDist_{max}(p_i) = \max(EDist_1, \dots, EDist_2, \dots, EDist_{N-1})$ is the maximum Euclidean Distance between p_i and the set of the j previous and the j next points of p_i . If *initial IPs* overlap, they are combined into one IP. Fig. 1 illustrates a simple application of the naïve method, in which a polyline is used to visualize the time consecutive recordings. The circles with radius of $r = 2drms$ show the inactive periods detected by the naïve method. Recordings are represented with black dots and the recordings of the detected inactive periods are marked by black squares. Fig. 2 illustrates the potential inactive periods detected by DBCIP with the same notation as Fig. 1.

For the naïve IP-detection method, both T (the minimum duration of inactivity) and the positioning error are required inputs. With this method, one must expect that some fraction of “true” IPs remain undetected (compare Fig. 1 with Fig. 2) due to the outliers caused by positioning errors. Below we suggest a more robust version capable of detecting IPs in the presence of outliers. The more robust method is closely related to DBSCAN (Density-Based Spatial Clustering of Applications with Noise) (Ester et al., 1996).

3.2. The density based clustering method for inactive periods (DBCIP)

For a spatial database, a density-based cluster is defined as a group of positions located closely together and well separated from other positions with respect to some distance measure. In addition, each cluster must contain at least a critical number of recordings. Well separated from the clusters, there may be “noisy recordings” sparsely occupying the domain to be considered. Since each IP is considered as a dense cluster of recordings, first, we briefly provide a background of the basic concepts of DBSCAN through the

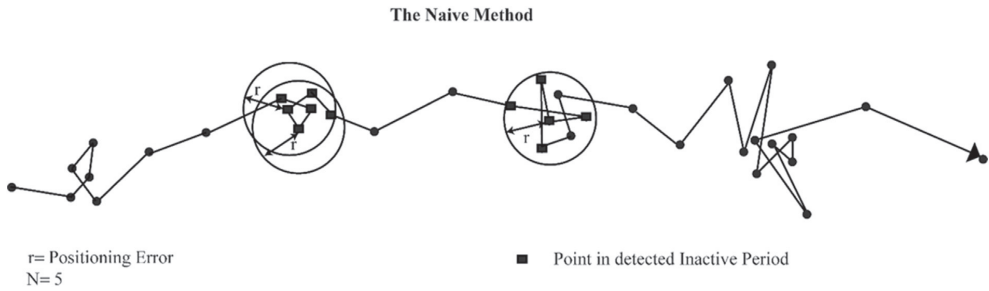


Fig. 1. Basic concepts of the naive method and the corresponding elements.

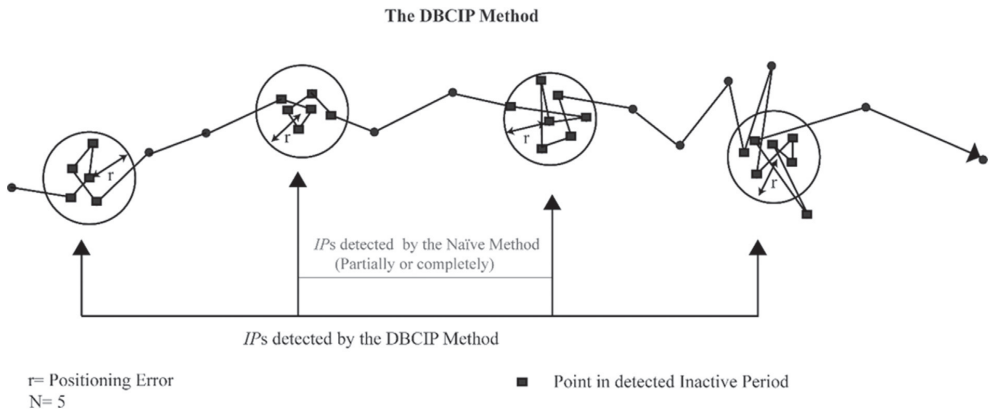


Fig. 2. IPs detected by the naive method versus IPs detected by the DBCIP method.

definitions. Then, we propose our modified version of DBSCAN for detecting IPs in trajectory data.

The DBSCAN algorithm (Ester et al., 1996) was designed to discover arbitrary-shaped clusters from noisy points in spatial datasets. The main idea of DBSCAN is that the neighborhood (defined by a given radius, Eps) of each cluster core point (see definition 2 below) is required to contain a minimum number of points ($MinPts$), i.e. the density in the neighborhood has to exceed some threshold.

The DBSCAN definitions based on a dataset D of points in 2-dimensional space using Euclidean distance ($EDist$) are as follows:

Definition 1: The Eps neighborhood of a point $p \in D$ is the subset

$$Q(p, Eps) = \{q \in D | EDist(p, q) \leq Eps\}.$$

Definition 2: (Core point) A point $p \in D$ whose Eps neighborhood contains at least $MinPts$ other points is called a core point.

Definition 3: (Directly density-reachable) A point $q \in D$ is said to be directly density-reachable from the point $p \in D$ if p is a core point and q is inside the Eps neighborhood of p (Fig. 3a).

Definition 4: (Density reachable) A point $p \in D$ is density-reachable from the core point $q \in D$ with respect to Eps and $MinPts$, if

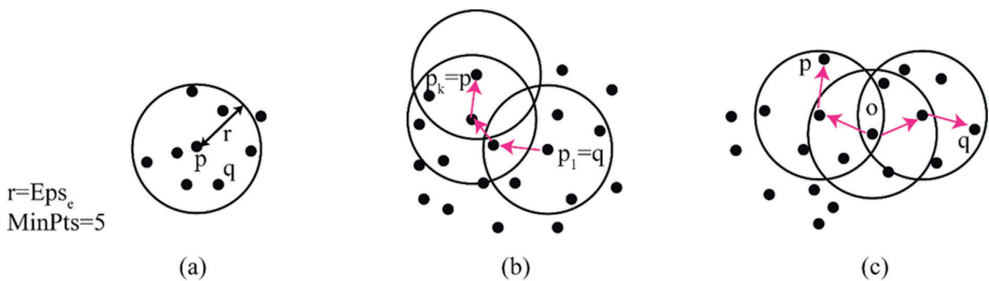


Fig. 3. Basic concepts of DBSCAN (a) q is directly density-reachable from the point p (b) p is density-reachable from q and (c) p and q are density-connected to each other through o (Birant & Kut, 2007; Ester et al., 1996).

there is a chain of points p_1, \dots, p_k with $p_1 = q$ and $p_k = p$, where each p_{i+1} is directly density-reachable from p_i (so all the points on the chain must be core points, with the possible exception of p) (see Fig. 3b).

Definition 5: (Density-connected) A point $p \in D$ is density-connected to a point $q \in D$, with respect to Eps and $MinPts$, if there is a point $o \in D$ such that both p and q are density-reachable from o with respect to Eps and $MinPts$ (see Fig. 3c).

Definition 6: (Cluster) A cluster C , with respect to Eps and $MinPts$, is a non-empty subset of D satisfying the following two conditions:

(Maximality) $\forall p, q \in D$: if $p \in C$ and q is density-reachable from p with respect to Eps and $MinPts$ then $q \in C$.

(Connectivity) $\forall p, q \in C$: is density-connected to q with respect to Eps and $MinPts$.

In our particular context, an IP is considered as a cluster of recordings (or positions) located closely together, both in the space and time dimension that contains at least a minimum number of recordings (N) (see formula 2). DBCIP is going to extend the notion of closeness in space to include the time dimension.

In contrast to the DBSCAN algorithm, which only considers the spatial dimension for the clustering ($EDist$), see the definitions 1–6, the DBCIP method combines the spatial and temporal dimensions into a joint neighborhood definition. In addition to the $EDist$ term that is the Euclidean distance (geographical proximity) between points, DBCIP introduces the $TDist$ term that is the distance between points along the trajectory (P). The $TDist$ term introduces temporal constraints in the clustering. Inclusion of both the Euclidean distance threshold, Eps_e , and the distance along the trajectory threshold, Eps_t , are required for the neighborhood definition in the DBCIP method.

Calculation of $EDist$ for two points is straightforward. For $TDist$, prior to running the algorithm, a list of cumulative step lengths ($CSLP = \{L_1, L_2, \dots, L_n\}$) from the first point to every point of the trajectory is created, where $L_1 = 0$, $L_2 = EDist(p_1, p_2)$, $L_3 = L_2 + EDist(p_2, p_3)$, ..., $L_n = L_{n-1} + EDist(p_{n-1}, p_n)$. Then $TDist$ for two points can be calculated by subtracting the cumulative step lengths associated with the two points.

Definition 7: (Trajectory distance) the trajectory distance ($TDist$) between two points p_i and p_j ($i \leq j$) on the path $P = \{p_1, p_2, \dots, p_n\}$ is defined as $TDist(p_i, p_j) = L_j - L_i$, where the cumulative step lengths $L_j, L_i \in CSLP$.

The $TDist$ term also allows ‘reasonable’ outliers, defined as follows, to be included in the cluster:

$$\forall p_j \in IP : \text{if } EDist(p_j, \text{core point}) > Eps_e \text{ then } TDist(p_j, \text{core point}) < Eps_t$$

To discover IPs using DBCIP for a trajectory P of n time ordered

recordings, let $MinPts = N-1$ (where N is the minimum number of recordings to detect an IP) and let Eps_e denote the $2drms$ in Euclidean space.

For the minimum number of recordings (N) and some suggested a priori value for the $2drms$, when it is unavailable for the application, the temporal neighborhood in terms of distance along the trajectory threshold is estimated from a bivariate normal distribution (see 3.2.1). Using N , the a priori $2drms$ and the estimated Eps_t , the DBCIP method (pseudo code is provided in appendix B) is able to cluster the recordings. Each detected IP is considered as the result of a stationary test and the associated positioning error estimate is compared to the a priori value of the $2drms$, as described in 3.2.2. Therefore, the $2drms$, and consequently the trajectory distance threshold, is estimated by a repetitive approach.

3.2.1. Estimation of Eps_t

An appropriate estimate for the trajectory distance threshold (Eps_t) can be obtained by simulation. The simulation is executed by drawing $\{\sigma_i = (x_i, y_i) | i = 1, \dots, N\}$ from a bivariate normal distribution with $Z \sim N(\mathbf{0}, \Sigma)$; where the mean ($\mathbf{0}$) is the vector (0,0), and the covariance matrix Σ (2×2) is diagonal with identical entries of $(Eps_e/2)^2$ (Four simulated trajectories are shown in Fig. 4).

The N simulated points $(\sigma_1, \dots, \sigma_N)$ are used to create a trajectory, $TSim$, and then the length of $TSim$ is calculated.

Finally, an average trajectory length is calculated from a large number (10,000) of simulated trajectories ($TSim$) and taken as the Eps_t estimate.

3.2.2. Estimation of Eps_e

Knowledge of the threshold value Eps_e (positioning error) is required in the proposed method for detecting IPs (see appendix B for pseudo code). If a stationary test before real animal tracking were available, or if the positioning error of the device were known, estimation of Eps_e by the following proposed procedure would be superfluous. Because this quantity is usually unknown in practice, it is necessary to include the Eps_e -estimation as a part of the suggested method based on some prior candidate value (Eps_{e0}). In the present case study we use $Eps_{e0} = 10$ (ordinary GPS collar technology with non-differentially corrected positions does not guarantee a $2drms$ less than 10 m). Starting our algorithm from Eps_{e0} with an error step size of 1 m, increasing up to $Eps_{emax} = 30$, our method explores all candidate values of Eps_{ei} , together with the spatial distribution of the points of the corresponding detected IPs . The smallest ‘feasible’ (as specified below) candidate value for the unknown Eps_e is taken as the final estimate.

The statistical justification of our strategy is based on the Rayleigh distribution:

Definition 8: $R \sim \text{Rayleigh}(\sigma)$ is Rayleigh distributed with the scale parameter σ if $R = \sqrt{X^2 + Y^2}$, where $X \sim N(0, \sigma^2)$, $Y \sim N(0, \sigma^2)$ are normal random variables, see p. 140 in (Papoulis, 1991).

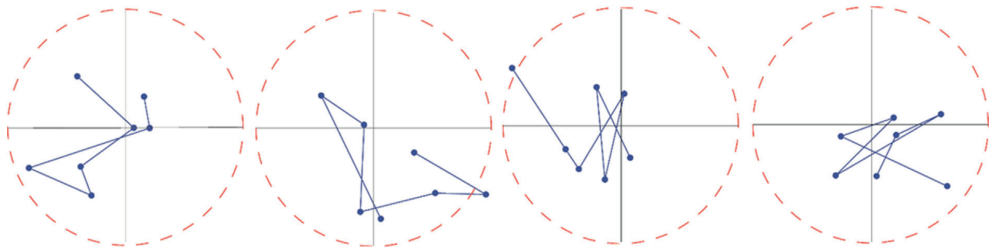


Fig. 4. Four simulated trajectories ($TSim$) are shown ($N = 7$). The circles are centered at zero and have radius of Eps_e .

According to our assumptions, the coordinates (x, y) for each IP are considered to be drawn from a bivariate normal distribution with a covariance matrix (2×2) having zero off-diagonal entries and diagonal entries of $(Eps_{\theta}/2)^2$. The distributions for the x - and y -coordinates in each IP are assumed to be

$$\left(2 * (X - \bar{X}) / \sqrt{(k + 1)/k} \right) \sim N(0, Eps_e^2) \tag{3}$$

$$\left(2 * (Y - \bar{Y}) / \sqrt{(k + 1)/k} \right) \sim N(0, Eps_e^2) \tag{4}$$

where k is the number of coordinates associated with the particular IP . Accordingly the statistic:

$$R = 2 * \sqrt{(X - \bar{X})^2 / ((k + 1)/k) + (Y - \bar{Y})^2 / ((k + 1)/k)} \sim \text{Rayleigh}(Eps_e) \tag{5}$$

Starting with the a priori value for Eps_{e0} , we proceed by calculating the radial errors for all detected IPs associated with the realizations of the statistic (R) in equation (Long & Nelson, 2013). By fitting the Rayleigh distribution to the radial errors, we derive a confidence interval for the associated scale parameter Eps_e . We expect the true unknown scale parameter (Eps_e) to be equal to (from the assumption that all IPs are equally accurate) or less than the candidate value Eps_{ei} . A candidate value Eps_{ei} is considered as a feasible estimate of Eps_e if it is smaller than the lower limit of the confidence interval of the scale parameter for the Rayleigh distribution.

For each i , we repeat the confidence interval calculations for the

scale parameter and compare with the corresponding Eps_{ei} . If j is the smallest integer resulting in Eps_{ei} (for all $i > j$) being strictly smaller than the estimated lower limit of the associated confidence interval of the scale parameter for the Rayleigh distribution, then Eps_{ej} is taken as the smallest “feasible” estimate for Eps_e .

Note that by increasing the Eps_{ei} 's, the neighborhoods in terms of $EDist$ and $TDist$ are extended correspondingly. This means that more points, possibly irrelevant to the real IPs , are taken into consideration with the effect of increasing the Rayleigh distribution's scale parameter estimates.

4. Case study

The dataset used in this paper is a part of a larger dataset from a long-term study of the ecological effects of sheep grazing in an alpine ecosystem (Mobæk, 2012). The research project is conducted on an established experimental site in Hol municipality, Buskerud county in the southern part of Norway (60°40'N, 7°55'E) (Fig. 5a). The site has been grazed by domestic sheep from 2002 onward, is roughly 2.7 km² in size (Fig. 5b) and covers an alpine habitat with sub-continental alpine climate. The mean annual temperature and precipitation of the enclosure are respectively -1.5 °C and about 1000 mm (Evju, Austrheim, Halvorsen, & Mysterud, 2009; Kausrud, Mysterud, Rekdal, HolandØ, & Austrheim, 2006). Fig. 5c illustrates an example trajectory that is used throughout this paper. The trajectory contains 1851 position recordings over 39 days. A sub-trajectory for one day with 48 recordings is also included.

Every summer 23 to 26 lactating ewes and 44 to 49 lambs belonging to the same sheep farmer are released into the experimental enclosures (Mobæk, 2012). Ewes are tracked using Tellus collars from Followit (URL, <http://wildlife.followit.se/>) during the

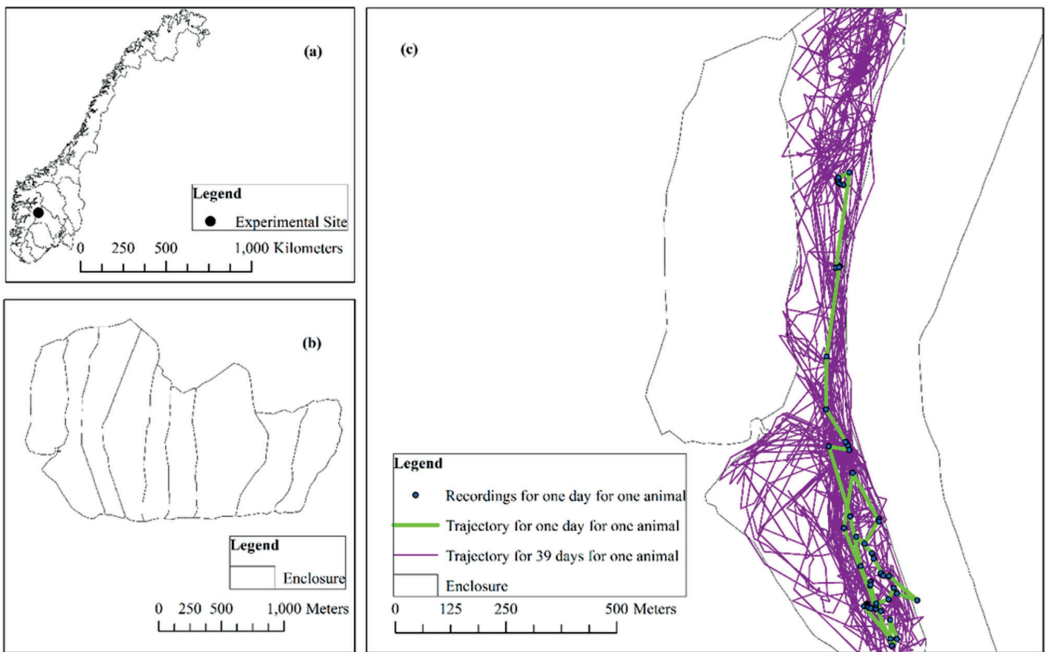


Fig. 5. (a) Map of Norway in a UTM projection (Source: Norwegian Mapping Authority) and the experimental site in Hol Municipality, (b) Map of Enclosures for the experimental site and (c) Representation of the trajectory of one specific ewe over 39 days and over one day.

grazing seasons (i.e., late June to late August/early September) within the enclosures. The collars record sheep locations once every half an hour. For each recording, the collars are programmed to search for satellites for 30–90 s. Time To Fix (TTF), the time the collar uses to calculate a position varies. The collars are equipped with dual axis activity sensors, which record the acceleration on two axes (x and y), i.e. the back-forth and left-right movement of the neck of ewes. The activity sensor value for each axis is the number of seconds within the TTF for which the accelerometer exceeds a user defined threshold. In order to control the variability of the TTF, the recorded activity sensor value for each axis is divided by the TTF for the particular attempt (Body et al., 2012), and the average of the x- and y-axis $((x + y)/(2 \times \text{TTF}))$ is considered as the Mean Activity Sensor Value (MASV). MASV does not capture the activities of the animal between recordings, but zero values (or values relatively close to zero) for consecutive activity sensor recordings may be used as a proxy for inactivity. Based on domain expert knowledge, the low quarter of the MASVs could represent inactivity.

5. Results

In this section, we first describe how the method works for a couple of cases. We use the MASV (Mean Activity Sensor Values) to verify our results. Then the effects of different minimum durations of inactivity (T) will be shown on the detected IP.

Since the proposed method requires a trajectory with constant time lag between successive recordings, linear interpolation was used for situations with one or two missing recordings and the trajectory was split when there were three or more missing recordings.

Three parameters $MinPts$, Eps_s and Eps_c are required for the DBCIP method (see appendix B). Based on the user input of T , 3-h for the following illustrative example, $MinPts = N-1$ is calculated by

formula 2, and Eps_c is calculated based on the explained simulation in (3.2.1). Table 1 and Fig. 6 summarize the method proposed in (3.2.2) for estimating Eps_c . The method results in an Eps_c estimate of 20 m for the current data.

Fig. 7 shows an IP that is ignored by the naïve method but detected by the DBCIP method. Fig. 8 shows an IP that is partly detected by the naïve method but detected by the DBCIP method. The corresponding details of recording time, coordinates, activity sensor values and more are provided with Table 2 and 3.

Fig. 9 shows histograms for MASV for all the recordings (a) and for the IPs detected by the DBCIP method (b) and the naïve method (c). The figure shows that the MASV for many of the recordings in the dataset is zero, and that the total number of recordings with a non-zero MASV is approximately two times the total number of recordings with zero MASV. The MASV histograms for the detected IPs confirm that the majority of IPs detected by both the presented methods are valid in terms of MASV. The results show that 84% of the IPs detected by the DBCIP method have MASV of zero or relatively close to zero.

MASVs close to zero for the recordings of detected IPs support our confidence in the clustering method. However, MASVs do not have to be close to zero for an IP. Because sensors are placed around the neck of animals, and therefore, are affected by head movements even at resting bouts. Therefore, higher MASVs than usual for that period may occur (for example see the fifth point for the detected IP in Fig. 8) (Gervasi et al., 2006).

Our results show that 79% of the IPs detected based on only MASVs are also detected by the DBCIP method, and 66% of the IPs detected by the DBCIP method are detected based on only MASVs.

It should be noted that the DBCIP method detected more IPs than the naïve method (29% and 17% of the recordings, respectively). Moreover, 55% of the recordings detected by the DBCIP method as IPs, were also detected by the naïve method, and 94% of the recordings which were detected by the naïve method as IPs,

Table 1
Recordings for the estimation of Eps_c for a dataset. The candidate values ($Eps_{c,i}$) and corresponding scale parameter confidence interval estimates based on the detected IPs. The candidate values $Eps_{c,i}$ 8 and 9 are included in the table just for demonstration.

Candidate values $Eps_{c,i}$ (2drms)	Confidence Interval for the Rayleigh distributed scale parameter	Percentage of recordings detected as IPs
8	[7.8 , 8.8]	13%
9	[8.4 , 9.4]	16%
10	[9.9 , 11.0]	18%
11	[10.6 , 11.7]	19%
12	[11.0 , 12.2]	20%
13	[11.9 , 13.1]	21%
14	[12.4 , 13.7]	22%
15	[13.2 , 14.5]	23%
16	[14.1 , 15.4]	24%
17	[14.9 , 16.3]	25%
18	[16.5 , 17.9]	26%
19	[18.0 , 19.5]	27%
20	[19.9 , 21.6]	29%
21	[21.5 , 23.2]	31%
22	[23.5 , 25.3]	33%
23	[24.1 , 25.9]	35%
24	[25.1 , 27.0]	37%
25	[25.7 , 27.6]	38%
26	[26.9 , 28.9]	39%

Table 2

The dataset corresponds to the recordings for the IP detected by DBCIP in Fig. 7.

label	Date and Time	x coordinate	y coordinate	Alt	DOP	SVs	Activity sensor X	Activity sensor Y	TTF
a	2012-07-19 00:00:00	441311.4	6730428.1	1298	1.1	7	0	0	30
b	2012-07-19 00:30:00	441317.4	6730427.4	1302	0.8	10	0	0	30
c	2012-07-19 01:00:00	441294.7	6730453.9	1290	0.8	9	0	0	30
d	2012-07-19 01:30:00	441320.5	6730430.9	1312	0.7	10	0	0	43
e	2012-07-19 02:00:00	441322.6	6730426.2	1307	0.8	10	2	0	30
f	2012-07-19 02:30:00	441327.7	6730436.4	1322	1	9	0	0	30
g	2012-07-19 03:00:00	441326.5	6730443.3	1311	1	9	0	0	30
h	2012-07-19 03:30:00	441330.0	6730429.1	1321	0.9	10	1	0	30
i	2012-07-19 04:00:00	441335.2	6730421.4	1319	1.1	9	1	1	30

Table 3

The dataset corresponds to the recordings for the IP detected by DBCIP in Fig. 8.

label	Date and Time	x coordinate	y coordinate	Alt	DOP	SVs	Activity sensor X	Activity sensor Y	TTF
a	2012-07-22 21:00:00	441275.6	6729579.2	1176	0.8	10	6	0	30
b	2012-07-22 21:30:00	441283.2	6729580.1	1176	0.8	10	2	0	43
c	2012-07-22 22:00:00	441282.2	6729574.9	1163	1	9	0	0	30
d	2012-07-22 22:30:00	441282.7	6729573.0	1161	0.8	8	0	1	30
e	2012-07-22 23:00:00	441282.2	6729573.4	1168	1.3	5	8	2	30
f	2012-07-22 23:30:00	441273.2	6729578.0	1162	1.1	6	3	0	44
g	2012-07-23 00:00:00	441274.4	6729577.2	1163	0.9	6	0	0	30
h	2012-07-23 00:30:00	441261.7	6729552.0	1169	0.9	8	0	0	30
i	2012-07-23 01:00:00	441284.1	6729581.5	1185	0.8	9	1	3	43
j	2012-07-23 01:30:00	441285.2	6729581.8	1182	0.8	9	2	2	30
k	2012-07-23 02:00:00	441292.3	6729591.2	1211	1	8	1	1	30
l	2012-07-23 02:30:00	441284.1	6729585.4	1188	0.9	9	1	0	43
m	2012-07-23 03:00:00	441281.7	6729578.2	1175	1	9	1	0	30

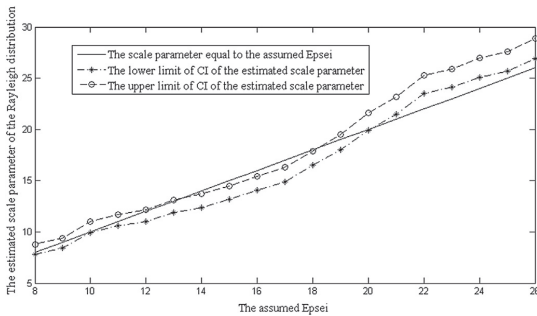


Fig. 6. The confidence intervals (CI) of the Rayleigh distribution scale parameters for increasing Eps_{ci} candidate values.

were also detected by the DBCIP method. The average step length during the detected IPs for the naïve method and DBCIP method were $7.11 \text{ m} \pm 5.30 \text{ (SE)}$ and $12.01 \text{ m} \pm 11.25 \text{ (SE)}$, respectively (in comparison, the average step length for the whole trajectory is $50.14 \text{ m} \pm 66.57 \text{ (SE)}$).

Fig. 10 shows the detected IPs for different values of T (1–6 h), using the DBCIP method. There are IPs around midnight for almost all of the 39 days (Fig. 10a) and also some IPs during the day. As T increases, the number of days with IPs decreases. Fig. 10c ($T = 3 \text{ h}$) shows very few detected IPs during daytime, and almost none during the daytime when $T = 4, 5$ or 6 h (Fig. 10d–f).

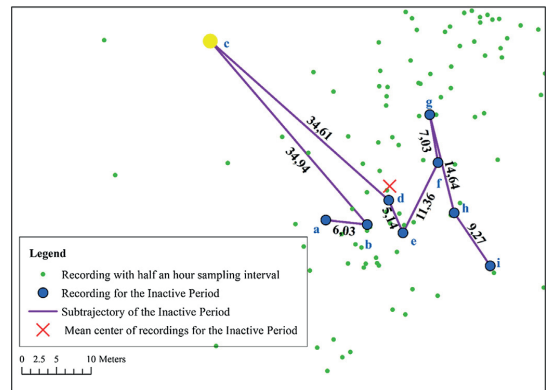


Fig. 7. An IP with MASV close to zero that is undetected by the naïve method but detected by the DBCIP method. The highlighted record in Table 2 corresponds to the outlier point, c. The labeled points from a to i show the IP detected by the DBCIP method centered at the red x with $T = 4 \text{ h}$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6. Validation

Besides using available activity sensor values to verify the IPs detected by the proposed method, we also used simulation techniques for a precise quantitative evaluation of the proposed method.

With the R-package ‘adehabitat’ (Calenge, 2006) it is possible to

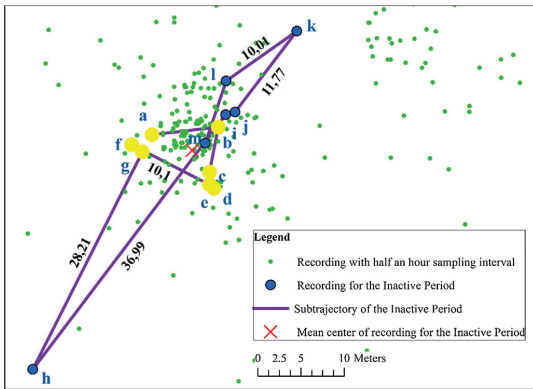


Fig. 8. An *IP* partly detected by the naïve method, and fully detected by the DBCIP method with *MASV* close to zero. The highlighted records in Table 3 correspond to the highlighted points on the map and show the *IP* detected by the naïve method. The labeled points from a to m show the *IP* detected by the DBCIP method centered at the red x with 6 h of inactivity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

centers representing true resting positions) as well as the duration of inactivity reflected by the sets of simulated local positions.

For the simple random walk model based on 10,000 simulations, the average percentage of the inactive periods correctly detected by the DBCIP algorithm and the naïve algorithm is $91\% \pm 4\%$ (SE) and $17\% \pm 3\%$ (SE) respectively.

For the correlated random walk model based on 10,000 simulations, the average percentage of the inactive periods correctly detected by the DBCIP algorithm and the naïve algorithm is $86\% \pm 4\%$ (SE) and $15\% \pm 5\%$ (SE) respectively.

Our experiments with the simple random walk and correlated random walk models indicate that when the average step length is less than $2.5 \times (Eps_s)$, we cannot trust the result of the proposed algorithm. As already mentioned in (Jerde & Visscher, 2005), it is difficult to conclude if an animal is moving or not at step lengths less than 5 error standard deviations.

7. Conclusion

In this paper, a method (DBCIP) is introduced to detect when and where a GPS tracked animal is inactive versus active. Contrary to existing work from the animal movement literature, which are mostly based on the existence of relatively high sampled activity

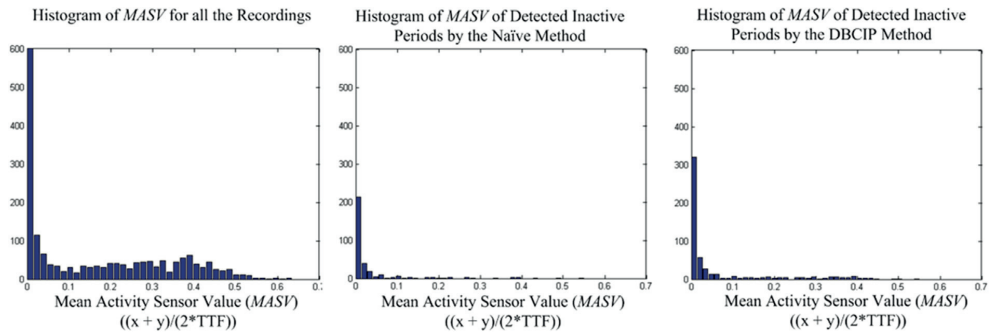


Fig. 9. Histogram of *MASVs* for (a) the whole dataset (mean = 0.17) (b) the *IPs* detected by the naïve method (mean = 0.03) and (c) the *IPs* detected by the DBCIP (mean = 0.05). *MASV* ranging from 0 to 0.7.

simulate individual animal movement trajectories based on both a simple random walk model (i.e. uncorrelated and unbiased) and a correlated random walk model¹ (Kareiva & Shigesada, 1983). The number of positions and the average step length for each simulated trajectory is chosen to be similar to the number of positions and average step length found from the real dataset.

Between 20% and 30% of the simulated trajectory points were chosen randomly as inactive locations (*IP* centers). For each *IP* center a set of local positions, satisfying the requirements for the minimum inactivity period (*T*), were drawn from a bivariate normal distribution with mean (μ) equal to the coordinates of the *IP* center, and covariance matrix (Σ) with zero off-diagonal entries and diagonal entries equal to $(Eps_s/2)^2$. The simulated trajectories to be tested with the DBCIP method were obtained by inserting simulated local positions for all *IP* centers (see Fig. 11). In these simulations, we control both the location of each inactive site (the *IP*

sensor values, the proposed method requires only the positions of the tracked animal. Another advantage of DBCIP is its estimation of the positioning error. Compared to existing studies in space-time analysis of movement data, DBCIP is original by being able to detect inactive periods by only requiring the minimum duration of inactivity (*T*) as input from the user.

An inactive period (*IP*) is defined as a set of points with high spatial and temporal similarity, where the exact similarity threshold is unknown in advance. The basic idea of the proposed method for detecting *IPs* is that both the distance along the trajectory and the distances between positions for an *IP* part of a trajectory should be very small when compared with the non-*IP* parts of the same trajectory. The required thresholds for the distance along the trajectory and the Euclidean distance between points are directly derived from the user specified *T* and the bivariate normal distribution of positions in the *IP* neighborhoods. Results have been verified using Mean Activity Sensor Values, domain expert knowledge, and simulations.

The usefulness of the proposed method has been demonstrated in an animal movement context, but it can also be employed for all homogeneous trajectories which are formed based on error-prone

¹ The correlated random walk models assume a dependency between the directions of successive steps. In the simple random walk models, such dependencies are exchanged with independence.

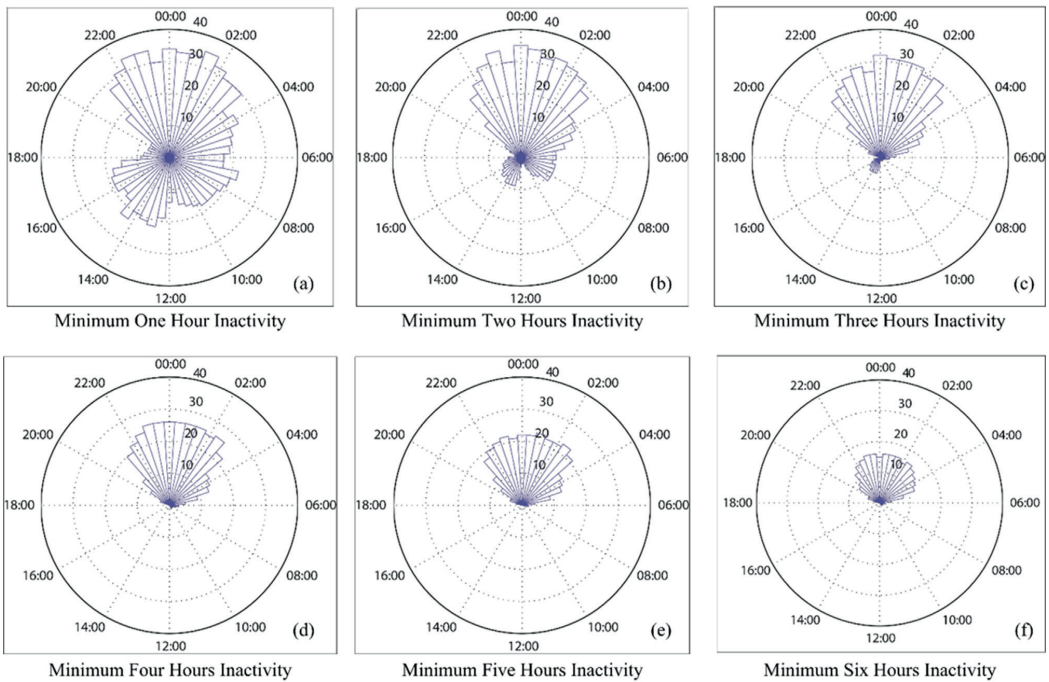


Fig. 10. 24-hour histograms of the IPs detected by the DBCIP method with $T = 1$ h (a), $T = 2$ h (b), $T = 3$ h (c), $T = 4$ h (d), $T = 5$ h (e), and $T = 6$ h (f) of inactivity.

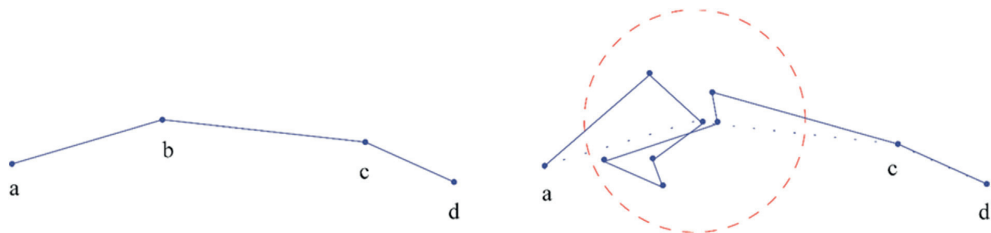


Fig. 11. The trajectory fragment (to the left) has 4 simulated points (i.e. a,b,c,d) of a random walk. Its second point (b) is drawn as an IP center and replaced by 7 local positions generated from the bivariate normal distribution. The resulting trajectory fragment (to the right) containing 10 points used for validation of the DBCIP method.

recordings in different application domains.

When increasing the sampling rate for a fixed user input value of T , more recordings (N) are required to be within a sufficient spatio-temporal proximity to be detected as an IP and then, the estimated trajectory distance threshold (Eps_t) will increase. By the same token, a decrease in sampling rate implies a reduction of N and Eps_t for the detection of IPs. In the end, the sampling rate does not influence the detected IPs significantly since an increase in the number of recordings requires a larger threshold Eps_t and vice versa. It should be noted that for homogeneous trajectories, there is a tight correspondence between sampling rate and the step lengths (Euclidean distance between consecutive recordings). An increased (finer) sampling rate may lead to unwanted problems when the step lengths become close to the positioning error. According to (Jerde & Visscher, 2005), it is difficult to classify movements of an object at step lengths less than 2.5 times the positioning error. The simulation results confirm that DBCIP can detect IPs with high

accuracy for step lengths larger than 2.5 times the positioning error. With exact knowledge about the average speed of a moving object and the positioning error, it would be possible to calculate the step lengths for various sampling rates and finally estimate the highest practical sampling rate for the DBCIP method.

Further investigations of the DBCIP method may consider the impact of sampling rate on the detected IPs. By considering trajectories of high sampling rate and their various associated down sampled versions, it is possible to study the effects (if any) of the sampling rate in the detection of IPs.

Acknowledgement

We are grateful to our colleagues Prof. Øystein Holand and his PhD student (Nicolai Hermann Jørgensen) from department of ‘Animal and Aquacultural Sciences’ who shared their expert knowledge with us and provided the dataset. The Research Council

of Norway (project number 239070) provided financial support for one of the co-authors (Indahl). This research was supported by the Norwegian University of Life Sciences under project number 1301051701.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.apgeog.2016.06.009>.

References


- Abedi, N., Bhaskar, A., & Chung, E. (2014). Tracking spatio-temporal movement of human in terms of space utilization using Media-Access-Control address data. *Applied Geography*, 51, 72–81.
- Adrados, C., Baltzinger, C., Janeau, G., & Pepin, D. (2008). Red deer *Cervus elaphus* resting place characteristics obtained from differential GPS data in a forest habitat. *European Journal of Wildlife Research*, 54(3), 487–494.
- A model for enriching trajectories with semantic geographical information. In Alvares, L. O., Bogorny, V., Kuijpers, B., de Macedo, J. A. F., Moelans, B., & Vaisman, A. (Eds.), *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, (2007). ACM.
- Andrienko, G. L., Andrienko, N. V., Fuchs, G., Raimond, A.-M. O., Symanzik, J., & Ziemlicki, C. (Eds.). (2013). *Extracting semantics of individual places from movement data by analyzing temporal patterns of visits*. COMP@ SIGSPATIAL.
- Ankerst, M., Breunig, M. M., Kriegel, H. P., & Sander, J. (June 1999). OPTICS: Ordering points to identify the clustering structure. *Sigmod Record*, 28(No 2), 49–60.
- Awange, J. L. (2012). *Environmental monitoring using GNSS: Global navigation satellite systems*. Springer Science & Business Media.
- Birant, D., & Kut, A. (2007). S. T.-D. B. S. C. A. N.: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering*, 60(1), 208–221.
- Body, G., Weladji, R. B., & Holand, O. (2012). The recursive model as a new approach to validate and monitor activity sensors. *Behavioral Ecology and Sociobiology*, 66(11), 1531–1541.
- Bogorny, V., Renso, C., Aquino, A. R., Lucca Siqueira, F., & Alvares, L. O. (2014). CONSTAnT—a conceptual data model for semantic trajectories of moving objects. *Transactions in GIS*, 18(1), 66–88.
- Bowman, J. L., Kochanny, C. O., Demaris, S., & Leopold, B. D. (2000). Evaluation of a GPS collar for white-tailed deer. *Wildlife Society Bulletin*, 28(1), 141–145.
- Buard, E. (2011). *Pratiques spatiales des populations animales: Analyses par les trajectoires*. In *Actes de Dixième Rencontres de Théo Quant*.
- Cagnacci, F., Boitani, L., Powell, R. A., & Boyce, M. S. (2010). Animal ecology meets GPS-based radiotelemetry: A perfect storm of opportunities and challenges. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1550), 2157–2162.
- Cain, J. W., Krausman, P. R., Jansen, B. D., & Morgart, J. R. (2005). Influence of topography and GPS fix interval on GPS collar performance. *Wildlife Society Bulletin*, 33(3), 926–934.
- Calenge, C. (2006). The package “adehabitat” for the R software: A tool for the analysis of space and habitat use by animals. *Ecology Model*, 197(3–4), 516–519.
- Cargnelutti, B., Coulon, A., Hewison, A. J. M., Goulard, M., Angibault, J. M., & Morellet, N. (2007). Testing Global Positioning System performance for wildlife monitoring using mobile collars and known reference points. *Journal of Wildlife Manage*, 71(4), 1380–1387.
- Chin, G. Y. (1987). *Two-dimensional measures of accuracy in navigational systems*. US Department of Transportation, Research and Special Programs Administration, Office of Program Management and Administration.
- D’eon, R. G., & Delporte, D. (2005). Effects of radio-collar position and orientation on GPS radio-collar performance, and the implications of PDOP in data screening. *Journal of Applied Ecology*, 42(2), 383–388.
- D’eon, R. G., Serrouya, R., Smith, G., & Kochanny, C. O. (2002). GPS radiotelemetry error and bias in mountainous terrain. *Wildlife Society Bulletin*, 30(2), 430–439.
- Di Orio, A. P., Callas, R., & Schaefer, R. J. (2003). Performance of two GPS telemetry collars under different habitat conditions. *Wildlife Society Bulletin*, 31(2), 372–379.
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (Eds.). (1996). *A density-based algorithm for discovering clusters in large spatial databases with noise*. *Kdd*.
- Evju, M., Austrheim, G., Halvorsen, R., & Mysterud, A. (2009). Grazing responses in herbs in relation to herbivore selectivity and plant traits in an alpine ecosystem. *Oecologia*, 161(1), 77–85.
- Frair, J. L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N. J., & Pedrotti, L. (2010). Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. *Philosophical Transaction of the Royal Society B*, 365(1550), 2187–2200.
- Frair, J. L., Merrill, E. H., Visscher, D. R., Fortin, D., Beyer, H. L., & Morales, J. M. (2005). Scales of movement by elk (*Cervus elaphus*) in response to heterogeneity in forage resources and predation risk. *Landscape Ecology*, 20(3), 273–287.
- Ganskopp, D. C., & Johnson, D. D. (2007). GPS error in studies addressing animal movements and activities. *Rangeland Ecology & Management*, 60(4), 350–358.
- Gervasi, V., Brunberg, S., & Swenson, J. E. (2006). An individual-based method to measure animal activity levels: A test on brown bears. *Wildlife Society Bulletin*, 34(5), 1314–1319.
- Grunevalder, S., Broekhuis, F., Macdonald, D. W., Wilson, A. M., McNutt, J. W., Shawe-Taylor, J., et al. (2012). Movement activity based classification of animal behaviour with an application to data from cheetah (*Acinonyx jubatus*). *Plos One*, 7(11).
- Gudmundsson, J., Laube, P., & Wolle, T. (2011). Computational movement analysis. In *Springer handbook of geographic information* (pp. 423–438). Springer.
- Hansen, M. C., & Riggs, R. A. (2008). Accuracy, precision, and observation rates of global positioning system telemetry collars. *Journal of Wildlife Management*, 72(2), 518–526.
- Hornsby, K. S., & Cole, S. (2007). Modeling moving geospatial objects from an event-based perspective. *Transactions in GIS*, 11(4), 555–573.
- Hurfurd, A. G. P. S. (2009). Measurement error gives rise to spurious 180 degrees turning angles and strong directional biases in animal movement data. *Plos One*, 4(5).
- Jerde, C. L., & Visscher, D. R. (2005). GPS measurement error influences on movement model parameterization. *Ecological Applications*, 15(3), 806–810.
- Kareiva, P. M., & Shigesada, N. (1983). Analyzing insect movement as a correlated random-walk. *Oecologia*, 56(2–3), 234–238.
- Kausrud, K., Mysterud, A., Rekdal, Y., Holand, Ø., & Austrheim, G. (2006). Density-dependent foraging behaviour of sheep on alpine pastures: Effects of scale. *Journal of Zoology*, 270(1), 63–71.
- Kritzler, M., Raubal, M., & Krüger, A. (2007). A GIS framework for spatio-temporal analysis and visualization of laboratory mice tracking data. *Transactions in GIS*, 11(5), 765–782.
- Laube, P., & Purves, R. S. (2011). How fast is a cow? cross-scale analysis of movement data. *Transactions in GIS*, 15(3), 401–418.
- Lewis, J. S., Rachlow, J. L., Garton, E. O., & Vierling, L. A. (2007). Effects of habitat on GPS collar performance: Using data screening to reduce location error. *Journal of Applied Ecology*, 44(3), 663–671.
- Long, J. A., & Nelson, T. A. (2013). A review of quantitative methods for movement data. *International Journal of Geographical Information Science*, 27(2), 292–318.
- Mobæk, R. (2012). *Density dependent foraging ecology and performance of domestic sheep on alpine ranges* [Doctoral thesis]. Norwegian University of Life Sciences (NMBU).
- Moen, R., Pastor, J., Cohen, Y., & Schwartz, C. C. (1996). Effects of moose movement and habitat use on GPS collar performance. *Journal of Wildlife Management*, 60(3), 659–668.
- Palma, A. T., Bogorny, V., Kuijpers, B., & Alvares, L. O. (2008). A clustering-based approach for discovering interesting places in trajectories. *Applied Computing*, 1–3, 863–868.
- Papoulis, A. (1991). *Probability, random variables, and stochastic processes* (3rd ed.). New York: McGraw-Hill.
- Pepin, D., Adrados, C., Mann, C., & Janeau, G. (2004). Assessing real daily distance traveled by ungulates using differential GPS locations. *Journal of Mammalogy*, 85(4), 774–780.
- Phithakkitkunoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., & Ratti, C. (2010). Activity-aware map: Identifying human daily activity pattern using mobile phone data. In *Human behavior understanding* (pp. 14–25). Springer.
- DB-SMoT: A direction-based spatio-temporal clustering method. In Rocha, J. A. M., Oliveira, G., Alvares, L. O., Bogorny, V., & Times, V. C. (Eds.), *Intelligent systems (IS)*, 2010 5th IEEE international conference, (2010). IEEE.
- Schwager, M., Anderson, D. M., Butler, Z., & Rus, D. (2007). Robust classification of animal tracking data. *Computers and Electronics in Agriculture*, 56(1), 46–59.
- Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Portoa, F., & Vangenot, C. (2008). A conceptual view on trajectories. *Data & Knowledge Engineering*, 65(1), 126–146.
- Specht, M., & Szoit, T. (2012). Accuracy analysis of GPS sport receivers in dynamic measurements. *Annual of Navigation*, 19(1), 165–176.
- Stewart, B. P., Nelson, T. A., Wulder, M. A., Nielsen, S. E., & Stenhouse, G. (2012). Impact of disturbance characteristics and age on grizzly bear habitat selection. *Applied Geography*, 34, 614–625.
- Technitis, G., Othman, W., Safi, K., & Weibel, R. (2015). From A to B, randomly: A point-to-point random trajectory generator for animal movement. *International Journal of Geographical Information Science*, 29(6), 912–934.
- Tran, L. H., Nguyen, Q. V. H., Do, N. H., & Yan, Z. (2011). *Robust and hierarchical stop discovery in sparse and diverse trajectories*.
- Ungar, E. D., Henkin, Z., Gutman, M., Dolev, A., Genizi, A., & Ganskopp, D. (2005). Inference of animal activity from GPS collar data on free-ranging cattle. *Rangeland Ecology & Management*, 58(3), 256–266.
- Van Moorter, B., Visscher, D. R., Jerde, C. L., Frair, J. L., & Merrill, E. H. (2010). Identifying movement states from location data using cluster analysis. *Journal of Wildlife Management*, 74(3), 588–594.
- Versichele, M., Neutens, T., Delafontaine, M., & Van de Weghe, N. (2012). The use of bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. *Applied Geography*, 32(2), 208–220.
- Yan, Z., Parent, C., Spaccapietra, S., & Chakraborty, D. (2010). A hybrid model and computing platform for spatio-semantic trajectories. In *The semantic Web: Research and applications* (pp. 60–75). Springer.
- Zimmermann, M., Kirste, T., & Spiliopoulou, M. (2009). Finding stops in error-prone trajectories of moving objects with time-based clustering. In *Intelligent interactive assistance and mobile multimedia computing* (pp. 275–286). Springer.

Research Paper II

Teimouri, M., Indahl, U. G., Sickel, H., & Tveite, H. (2018). Deriving Animal Movement Behaviors Using Movement Parameters Extracted from Location Data. *ISPRS International Journal of Geo-Information*, 7(2), 78.

Article

Deriving Animal Movement Behaviors Using Movement Parameters Extracted from Location Data

Maryam Teimouri ^{1,*}, Ulf Geir Indahl ¹, Hanne Sickel ² and Håvard Tveite ¹ 

¹ Faculty of Science and Technology, Norwegian University of Life Sciences, Drøbakveien 31, NO-1433 Ås, Norway; ulf.indahl@nmbu.no (U.G.I.); havard.tveite@nmbu.no (H.T.)

² Norwegian Institute of Bioeconomy Research, Postboks 115, NO-1431 Ås, Norway; hanne.sickel@nibio.no

* Correspondence: maryam.teimouri@nmbu.no; Tel.: +47-67-231547

Received: 14 December 2017; Accepted: 18 February 2018; Published: 24 February 2018

Abstract: We present a methodology for distinguishing between three types of animal movement behavior (foraging, resting, and walking) based on high-frequency tracking data. For each animal we quantify an individual movement path. A movement path is a temporal sequence consisting of the steps through space taken by an animal. By selecting a set of appropriate movement parameters, we develop a method to assess movement behavioral states, reflected by changes in the movement parameters. The two fundamental tasks of our study are segmentation and clustering. By segmentation, we mean the partitioning of the trajectory into segments, which are homogeneous in terms of their movement parameters. By clustering, we mean grouping similar segments together according to their estimated movement parameters. The proposed method is evaluated using field observations (done by humans) of movement behavior. We found that on average, our method agreed with the observational data (ground truth) at a level of $80.75\% \pm 5.9\%$ (SE).

Keywords: behavioral change point analysis (BCPA); hierarchical clustering; Kolmogorov-Smirnov (ks) distance

1. Introduction

Animal movement analysis is being revolutionized by the increasing positional accuracy and temporal frequency of tracking devices, such as ARGOS tags, RFID (Radio Frequency Identification) tags, Geotags, and GNSS (Global Navigation Satellite System) tags [1]. Inexpensive and ubiquitous positioning technologies and the development of methods to characterize and classify movement behavioral states from captured location data have received considerable attention in movement ecology.

Given an animal movement path [2–4], as a sequence of time-stamped locations, we focus on how to make inferences about animal movement behavior. Where and when does the animal engage in a specific movement behavior, and how does the movement behavior change over time?

Quantities like speed, step length (straight-line distance between successive locations), direction, and turning angle (change of direction between successive steps) that can be calculated from the raw location data are called movement parameters [5,6]. The movement parameters can be good proxies for movement behavioral states along an animal movement path [4].

The main research goal is this: Given sampled locations of an animal movement path, we aim at partitioning the movement path according to changes in movement behavior. Movement behavior is defined based on various combinations of the movement parameters, including turning angle and speed. According to [7], movement behaviors can be (i) resting (low mean turning angle and low mean speed), (ii) walking (low mean turning angle and high mean speed), (iii) foraging (high mean turning angle and low mean speed), and possibly (iv) undefined movement behavior (high mean turning

angle and high mean speed). The movement behavior of high turning angle and high speed is unlikely, as demonstrated by the Classification and Regression Tree method proposed in [8].

The proposed framework for detecting movement behavioral states according to changes in the movement parameters consists of two parts: (i) Using a slightly modified version of the recently developed behavioral change point analysis (BCPA) method [5,9], we partition an individual trajectory into segments of homogeneous movement parameters. (ii) Using the Kolmogorov–Smirnov distance metric, we compute a distance matrix for all pairs of segments, in order to generate an agglomerative hierarchical clustering model of the segments [10].

The remaining parts of the paper are structured as follows. We start by reviewing some relevant animal movement behavioral studies, which attempt to distinguish between different movement behavioral states based on location data. Then we describe our dataset and the field observations. Finally, we describe our methodological approach, including the modified version of BCPA and the proposed hierarchical clustering method, and validate the resulting model, using the field observations as ground truth.

2. Background

Recent developments in tracking devices and the increasing availability of movement data provide new opportunities for the inference of movement behavior from animal movement paths [4,11–13]. There are different approaches for distinguishing movement behaviors from animal movement paths, including statistical modelling, data mining techniques, mixtures of random walk models, and movement-derived parameters [3–5,9,14–16].

Gurarie et al. [12] group behavioral movement analysis methods into four categories: (1) metric-based, (2) classification and segmentation, (3) phenomenological time series analysis and (4) mechanistic movement modelling. They compare the categories in terms of complexity of the results and the intrinsic differences in the output, using one method from each category.

Edelhoff et al. [4] outline three broad types of research questions that are commonly addressed using path segmentation methods: (1) the quantitative description of movement patterns, (2) the detection of significant change points, and (3) the identification of underlying processes or hidden states.

The first type is based on geometric analysis of movement parameters. The movement paths are split into segments that are assumed to reflect different underlying movement behavior. Movement parameters include mean squared displacement [17], first passage time [18], (multi-scale) straightness index [14] and fractal dimension [19].

State–space modeling is closely related to the latter type, and seems promising. State–space models [16] are models that allow unobservable, true states to be inferred from observed data, by accounting for errors arising from imprecise observations and from stochasticity in the process being studied [20]. Specifying appropriate prior distributions and underlying model parameters for state–space models can be a challenging task for non-experts.

The second, and most interesting type for our framework, assesses movement parameters along the time axis, and identifies the structure or correlation of movement data using time series analysis. Relying on significant change points along the animal movement path, we selected the BCPA method [4,9] to segment the path, taking the temporal autocorrelation of the movement data into account. BCPA is acknowledged in the literature as a good method for segmenting movement paths efficiently, and it is relatively straightforward to implement. The method developed by Zhang et al. [21] is directly comparable to our method, since it also applies BCPA. Zhang et al. [21] used a three-step framework, including BCPA, hierarchical multivariate cluster analysis, and k-means clustering. Hierarchical multivariate cluster analysis is required to determine the number of clusters (k) before doing k-mean clustering. We wanted to consider the overall distribution of the movement parameter values in the clustering process, rather than a specific statistic like central tendency or median as used in [21]. With our approach, we also do not have to worry about seed selection, which k-means is sensitive to.

BCPA, like many segmentation methods (e.g., [22]), only provides information on significant change points along the movement path, without any further ecological context. That is why subsequent analysis (see Section 3.2 below) is required to gain particular insight into the movement behavioral states.

Having found the significant change points along the movement path, the main challenge is to cluster the segments, in order to identify different types of movement behavior. The clustering task includes two important steps: first, an appropriate similarity measure must be chosen, and second, an appropriate algorithm for grouping the observations based on the chosen similarity measure is required.

Several of the proposed trajectory clustering approaches [23–25] rely on the similarity of geometric shape, ignoring the temporal dimension. Recent works have developed spatiotemporal similarity measures for trajectory data, considering both the spatial and temporal dimensions [26,27]. As stated in [28], related work on similarity measures for trajectories include time series similarity measures, such as a variation of Minkowski distance (Lp-Norm family) [29], dynamic time warping (DTW) [30], edit distance [31], longest common subsequence (LCSS) [32], geometric shape matching techniques (such as Hausdorff distance) [33], and Fréchet distance [34]. It is worth noting that trajectories may be considered as similar in different respects—they may fully or partly coincide in space, have similar shapes, be fully or partly synchronous, or they may be disjointed in time but with similar movement behavior (speed, acceleration, etc.) [35,36].

We will assume that some unknown probability distribution functions of movement parameters can be used to characterize animal movement behavioral states [3]. The corresponding research goal is therefore to establish the concept of similarity between different animal movement behaviors, based on probability distribution functions estimated from the movement parameters.

3. Methodology

Classifying a heterogeneous trajectory according to movement behavior starts with a segmentation of the trajectory into reasonably homogenous parts, using relevant movement parameters. Then, these segments are grouped according to similarities in movement behavior. In this context, segmentation means the partitioning of a trajectory, T , into an unknown number (l) of sub-trajectories (T_1, T_2, \dots, T_l) referred to as segments [4].

In Section 3.1, we demonstrate how to apply the BCPA for trajectory segmentation. In Section 3.2, we explain how hierarchical clustering can be applied to obtain a grouping of the data, yielding satisfactory results for real-world, auto-correlated, high temporal frequency data (see Section 4).

3.1. Behavioral Change Point Analysis (BCPA)

Spatio-temporal autocorrelation is an intrinsic property of animal movement data that separates it from a random set of location data [37]. As the time interval between location observations decreases, the dependency between successive observations increases. In other words, the higher the temporal frequency of position sampling, the stronger the spatio-temporal autocorrelation (function of temporal and spatial distance between observations) [38]. Autocorrelation means that values from observations taken close to each other tend to be either more similar (positive autocorrelation) or less similar (negative autocorrelation) than would be expected from a random arrangement [2]. Nearness can be defined in space (spatial autocorrelation), or in time (temporal autocorrelation), or in both space and time (spatio-temporal autocorrelation).

Our first main goal is to identify points on the animal movement path where significant changes occur in movement behavior. We do not have any a priori assumptions about the number of change points. The temporal correlation (autocorrelation) of the movement parameters has to be considered to avoid spurious change points [12]. According to the guidelines for method selection in [4], “time-series based analysis” is the most appropriate segmentation method for our specific research goal and data. Recently, Gurarie et al. [9] introduced BCPA to statistically determine changes in movement behavioral

states along an animal's movement path. In BCPA, the autocorrelation is modelled explicitly as part of the likelihood function. Changes in the movement behavioral state are identified based on pronounced changes in the movement parameters (in our case, "persistence velocity" and "turning velocity", see Section 3.1.2 below).

The BCPA identifies changes in movement parameter values across a dataset, by using likelihood comparisons in a moving window over the time series. Within the window, the most likely change point is located according to the Bayesian Information Criterion (BIC) [12], which is a criterion for model selection among a finite set of models [39]. For the mathematical details of BCPA, readers are referred to [9]. For the present analysis, only the principles (without technical details) are required.

3.1.1. Moving Window Size

Choosing the moving window size is an important part of the data analysis, as it drives the resulting segmentation. Exploring the results obtained by using different window sizes is therefore an essential aspect in adapting the methodology for different studies.

The most important statistical assumptions in the modeling of BCPA are the Gaussian error structure and the exponential decay in the autocorrelation. To assess the assumptions of the BCPA, a diagnostic plot should be used both for comparing the residuals to a standard normal distribution and inspecting the autocorrelation function of the residuals [12,40]. Determining the moving window size is a challenge in BCPA. Based on the diagnostic plot of the residuals, BCPA is run with increasing window sizes (including at least 20 samples [9]), in a trial and error fashion, to identify a good value for the size of the moving window (see Section 5, results).

3.1.2. Movement Parameters and Movement Behavior

Movement parameters can be calculated based on either consecutive positions (stepwise) or multiple steps [4]. Most movement parameters are stepwise, including speed, step length, turning angle [41], persistence/turning velocity [9], net/mean squared displacement [17] and first passage time [18]. Sinuosity/tortuosity [42], fractal dimension [19] and multi-scale straightness index [14] are examples of movement parameters that are calculated over multiple steps.

Speed and turning angle are basic movement parameters that can be used to describe and analyze movement paths [6,41]. Laube and Purves [43] investigated how the temporal scale of locational data affects the calculation of movement parameters, such as the speed and turning angle of animal trajectories. By combining speed and turning angle into "persistence velocity" (or "turning velocity"), possible biases caused by varying sampling intervals are avoided, by relating speed to the observed turning angles [4]. Persistence (or turning) velocity is defined as the product of the estimated speed and the cosine (or sine) of the turning angle.

Influenced by [37], we segmented the animal movement path twice using BCPA (first for "persistence velocity" and then for "turning velocity") for improved reliability in the detection of real changes in the movement parameters. Because we expected that the two resulting sets of time points have some change points in common, we combined and ordered their change points and remove duplicates. Further analysis is required to group the segments according to behaviors (based on consistency in their movement parameters).

3.2. Hierarchical Clustering

Grouping segments into distinct groups based on some measure of similarity (or distance) is the essence of a cluster analysis [10]. The procedure consists of two steps: first we need to use a similarity measure that is defined for pairs of objects in the data domain, and second, we need to apply a clustering algorithm to partition the data into distinct groups using the similarity measure [13]. A self-contained review of cluster analysis in general terms is provided in [44,45], and we refer the reader to those for a thorough review.

From the animal movement behavior perspective, clustering analysis can be applied as either a main analysis [46–48] or a subsequent analysis [7,21] to infer movement behavior components from bio-logged movement data, including location data and activity sensor data.

As already mentioned, the walking behavior of animals is expected to be characterized by high speed and small turning angles (high directional persistence) [3,7]. Low speed and low directional persistence, usually with mean of zero, is characteristic for resting behavior. Low-directional persistence while resting is commonly attributed to GNSS errors [49]. Foraging behavior is expected to be characterized by low to moderate speed values and low directional persistence, with many sharp changes in direction [7] (usually with a mean value different from zero). Turning angles and associated parameters and graphs are analyzed using circular statistics (the key concept and formulation of circular statistics is found in [50,51]).

For segmenting a trajectory, we considered both “persistence velocity” and “turning velocity” when identifying behavioral change points. However, for clustering we considered only the persistence velocity values, because their associated distribution, unlike those of turning velocity, can discriminate between the three behaviors mentioned above (see Section 5, results): walking is characterized by high persistence velocity values, foraging by low values, and resting by intermediate values. Additionally, the directional histogram of turning angle values for each behavior increases our confidence to draw comparison across behaviors.

Unlike the approach in [21], for each segment we consider the overall distribution of the movement parameters (in our case, persistence velocity), rather than a specific parameter like central tendency. We use the Kolmogorov-Smirnov (ks) distance [52] to determine the distance between the segments’ persistence velocity distributions. No assumptions about the distributions are required [52]. Unlike the “Pearson correlation measure” or the “cosine measure” [53], the ks-distance metric can compare vectors of different lengths. Using the R-package “adehabitat” [54], we simulated a number of trajectories based on correlated random walk (CRW) models with different characteristics [55], and figured out how well the ks-metric differentiated between the different CRW models.

Given the observed persistence velocities $V_p(t_1), V_p(t_2), \dots, V_p(t_n)$ of a particular segment containing n samples, the empirical distribution function $F_n(V_p)$ is defined as the fraction of the observations that have values that are less than or equal to the value V_p . Thus $F_n(V_p)$ is the empirical cumulative distribution function, with respect to the n observations of the segment [52]. The ks metric is defined on the space of distribution functions, and the ks-distance between two empirical distribution functions [52,56,57] is obtained by calculating the supremum over the set of differences. Figure 1 indicates the ks-distance between two empirical persistence velocity distribution functions for two different segments.

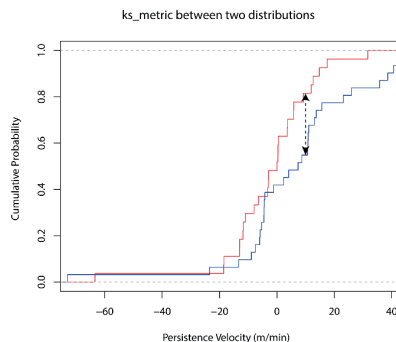


Figure 1. Illustration of Kolmogorov-Smirnov (ks)-metric on the two empirical distribution functions of two different segments’ persistence velocity (red and blue lines). The dashed line shows the ks-distance between two segments’ persistence velocities.

From the ks-metric, we generated a distance matrix to be used for the agglomerative hierarchical clustering approach using Ward's minimum variance method [58]. The grouping of segments into distinct movement behaviors was obtained as follows: each segment was assigned to its own cluster, and then the algorithm proceeded iteratively, so that at each stage the cluster pair to be merged was the one whose merger minimized the increase in the total within-group-error sum of squares. This continued until all segments were merged into a single cluster. The associated tree-like structure obtained by the merging process is known as a dendrogram [59–61].

The “hclust” function in the R programming language with the “ward.D2” criterion for hierarchical clustering was used in our method. Hierarchical methods produce not a single partitioning, but a hierarchy of nested partitions, which allows the user to consider different partitions according to the desired similarity level [53]. An essential part of the cluster analysis are the interpretations of the various clusters by a human analyst to acquire their meaning and value [23]. Careful inspection of the dendrogram and the distribution (and mean) of the persistence velocity values and the turning angle values is needed to determine an appropriate number of distinct movement behaviors. The number of distinct movement behaviors in our method is a priori considered to be unknown, except for the three movement behavioral states mentioned above.

4. Case Study

The dataset investigated in this paper belongs to an ongoing project, in which the main objective is to determine the influence of grazing on biodiversity. Four individual sheep were followed during the foraging season, in an established experimental site in Valdres in Oppland county, Norway (61.06° N, 9.40° E; Figure 2 (left)). The animals were tracked with a sampling frequency of 10 s, using Canmore GT-740FL GPS units (http://www.canmore.com.tw/pdf/GT-740FL%20DataSheet_V6.1.pdf). These units have a battery capacity of approximately 24 h when sampling data using this frequency. The GPS units were equipped with extra battery capacity, which extended the capacity to 3–4 days. The GPS position accuracy is specified as 2.5 m CEP (CEP, circular error probable, which is the radius of the smallest circle centered at the true position that covers 50% of the observations) under optimal conditions. The waterproof box with the GPS unit and batteries weighed approximately 400 g, and the size of the box was 13 cm × 10 cm × 5 cm.

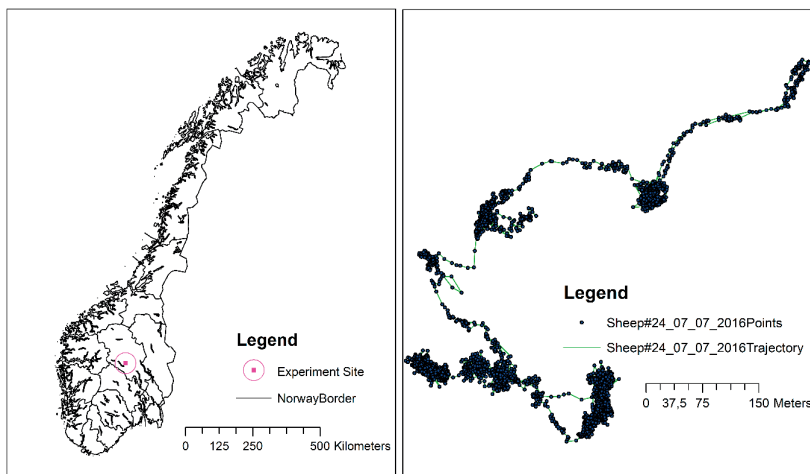


Figure 2. Map of Norway (source: Norwegian Mapping Authority; projection: UTM 33N) with the location of the experimental site (left) and an example trajectory for one sheep over one day (right).

The animals were also equipped with a radio bell (Telespor) which was set to transmit position signals every 15 min, so that it would be easy to find the animals. Before each observation period the sheep were captured for a few minutes, wherever they were found using radio bell signals, to replace the batteries.

The usable position data for the four sheep were collected on 15, 14, 9, and 4 different days respectively, totaling up to approximately 26 days of observations. The total number of position observations was 227,050.

The Android application “observationlogger” (<https://play.google.com/store/apps/details?id=com.mortensickel.obslogger>), developed by Morten Sichel, with available source code on github (https://github.com/sickel/observationlogger/releases/tag/version2_0), was used to log animal behavior (grazing, walking, resting, and other behavior) when observing the sheep. The registration times from the “observationlogger” and the tracking units were set up to automatically synchronize.

“Grazing” was registered when the animal clearly was eating, and “resting” if the animal was lying down. “Walking” was registered if the walking was associated with displacements that were not part of a grazing event. Each animal was observed for a few hours per day during GPS logging. The observations were made at a distance of approximately 25–50 m, attempting to avoid any influence on the behavior of the animals.

To reduce the amount of noise due to GNSS position errors, the dataset was resampled to 1 min intervals by averaging. For our analysis, the part of the valid position recordings that had corresponding “observationlogger” recordings was used.

5. Results

We demonstrated our method using a part of the trajectory of one of the sheep in the Valdres dataset. The trajectory part contained 4960 GPS positions logged between 00:01 a.m. and 8:20 p.m. on 7 July 2016, and is shown in Figure 2 (right). The corresponding ground truth movement behavior was recorded by a human observer from 7:20 a.m. until 3:33 p.m., and classified into “foraging behavior”, “resting behavior”, “walking behavior” and “other behavior”. The observer recorded the behaviors quite regularly, but when a specific behavior lasted for a while, observations were generally not recorded until a change in behavior took place. The model performance was assessed by comparing the predicted behaviors using our model with the ground truth (field) observations, using a confusion matrix.

By doing the BCPA for various moving window sizes and inspecting their diagnostic plots, we concluded that 30 consecutive sampling points (equivalent to considering a time interval of 30 min) was a feasible moving window size for our dataset. Figure 3 illustrates the corresponding diagnostic plots, including the qq-norm plot, the histogram, and the auto-correlation function of the standardized example dataset for persistence velocity (panel a) and turning velocity (panel b). The result is consistent with the required assumption of normality. The residual plots show that the model has captured the dominant patterns of the data quite well, although there is a small amount of autocorrelation left in the residuals (indicated by some significant spikes in the autocorrelation function plotted in the rightmost parts of Figure 3a,b).

The dendrogram resulting from the hierarchical clustering procedure (Figure 4) guides the choice of the number of clusters in the observed data. One might expect that there would be a certain number of clusters (e.g., 1, 2, or 3 for Figure 4). With reference to Figure 4, one cluster means that the whole dataset is a group, two clusters means that the leftmost (red) box is one group and the two rightmost boxes (colored blue and green) are the other group, while three clusters means that each colored box is a separate group. A subsequent analysis, based on visual inspection of the distribution of the turning angle and persistence velocity values, is recommended in order to confirm the appropriate number of clusters. The preferable number is the alternative that produces the most homogeneous clusters with respect to distinguishable persistence velocity boxplots and turning angle histograms (see Section 3.2). In other words, the behavior with the highest values for persistence velocity should have a mean

turning angle value of around zero radian, with low variance to correspond to walking. The behavior with intermediate values for persistence velocity should have a mean turning angle value of around zero radian with high variance, to correspond to resting. The behavior with the lowest values for persistence velocity should have a mean turning angle value different from zero with low variance, to correspond to grazing.

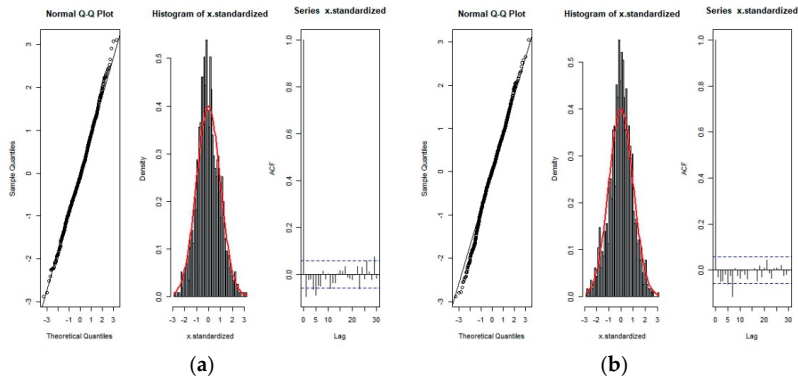


Figure 3. Diagnostic Plots for residuals for “persistence velocity” (a) and “turning velocity” (b).

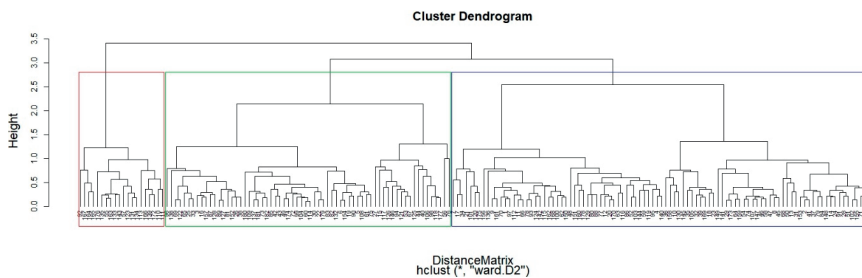


Figure 4. Dendrogram representation of the clustered sub-trajectories; the numbers in the bottom of the plot represent the segment numbers from the trajectory segmentation.

When using three clusters (corresponding to the three colored boxes in Figure 4), the textual output of our algorithm, including the start and the end times for each cluster, is shown in Figure 5.

Behaviour 1	start time	End time	Behaviour 2	start time	End time	Behaviour 3	start time	End time
[1,]	"2016-07-07"	"2016-07-07 00:27:00"	[1,]	"2016-07-07 00:27:00"	"2016-07-07 00:33:00"	[1,]	"2016-07-07 06:45:00"	"2016-07-07 06:53:00"
[2,]	"2016-07-07 00:33:00"	"2016-07-07 00:36:00"	[2,]	"2016-07-07 00:36:00"	"2016-07-07 00:55:00"	[2,]	"2016-07-07 06:57:00"	"2016-07-07 06:59:00"
[3,]	"2016-07-07 00:55:00"	"2016-07-07 02:08:00"	[3,]	"2016-07-07 02:08:00"	"2016-07-07 02:21:00"	[3,]	"2016-07-07 10:41:00"	"2016-07-07 10:42:00"
[4,]	"2016-07-07 02:21:00"	"2016-07-07 02:32:00"	[4,]	"2016-07-07 02:32:00"	"2016-07-07 02:46:00"	[4,]	"2016-07-07 12:25:00"	"2016-07-07 12:34:00"
[5,]	"2016-07-07 02:46:00"	"2016-07-07 02:54:00"	[5,]	"2016-07-07 02:54:00"	"2016-07-07 03:12:00"	[5,]	"2016-07-07 12:46:00"	"2016-07-07 12:47:00"
[6,]	"2016-07-07 03:12:00"	"2016-07-07 03:27:00"	[6,]	"2016-07-07 03:27:00"	"2016-07-07 03:39:00"	[6,]	"2016-07-07 13:47:00"	"2016-07-07 13:53:30"
[7,]	"2016-07-07 03:39:00"	"2016-07-07 04:06:00"	[7,]	"2016-07-07 04:06:00"	"2016-07-07 04:14:00"	[7,]	"2016-07-07 14:37:00"	"2016-07-07 14:46:00"
[8,]	"2016-07-07 04:14:00"	"2016-07-07 04:25:00"	[8,]	"2016-07-07 04:25:00"	"2016-07-07 04:37:00"	[8,]	"2016-07-07 14:49:00"	"2016-07-07 15:09:00"
[9,]	"2016-07-07 04:37:00"	"2016-07-07 04:51:00"	[9,]	"2016-07-07 04:51:00"	"2016-07-07 04:56:00"	[9,]	"2016-07-07 15:28:00"	"2016-07-07 15:32:00"
[10,]	"2016-07-07 04:56:00"	"2016-07-07 05:02:00"	[10,]	"2016-07-07 05:02:00"	"2016-07-07 05:03:00"	[10,]	"2016-07-07 15:53:00"	"2016-07-07 15:57:00"
[11,]	"2016-07-07 05:03:00"	"2016-07-07 05:10:00"	[11,]	"2016-07-07 05:10:00"	"2016-07-07 05:23:00"	[11,]	"2016-07-07 17:53:00"	"2016-07-07 18:29:00"
[12,]	"2016-07-07 05:23:00"	"2016-07-07 05:58:00"	[12,]	"2016-07-07 05:58:00"	"2016-07-07 06:06:00"	[12,]	"2016-07-07 18:52:00"	"2016-07-07 19:06:00"
[13,]	"2016-07-07 06:06:00"	"2016-07-07 06:24:00"	[13,]	"2016-07-07 06:24:00"	"2016-07-07 06:29:00"	[13,]	"2016-07-07 19:31:00"	"2016-07-07 19:32:00"

Figure 5. A part of the typical text output of the algorithm (behavior 1 is resting, behavior 2 is foraging, and behavior 3 is walking).

As already mentioned, for clustering we considered only the persistence velocity values, because their distribution (Figure 6), unlike the turning velocity values (Figure 7), can discriminate between the three behaviors.

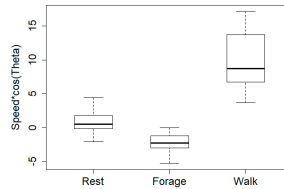


Figure 6. Boxplot representation of persistence velocity values for each cluster. There is a significant difference between the persistence velocity values, according to the Kruskal-Wallis rank sum test (chi-square distribution with 2 degrees of freedom ($X^2(2) = 58.22, p < 0.01$)).

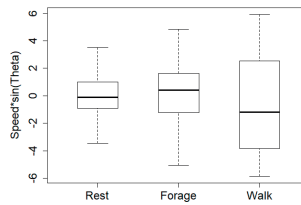


Figure 7. Boxplot representation of turning velocity values for each cluster. There is no significant difference between the turning velocity values, according to the Kruskal-Wallis rank sum test (chi-square distribution with 2 degrees of freedom ($X^2(2) = 4.29, p = 0.12$)).

The boxplots of persistence velocity values for each cluster in Figure 6, and the directional histograms summarizing the turning angle values for each cluster in Figure 8, are considered in combination, in order to assign meaningful movement behavior names to the clusters.

Figure 8c shows a movement behavior where most turning angles are small, suggesting that the overall direction is persistent. A comparison with the corresponding boxplot of persistence velocity values in Figure 6 (Walk) suggests that the particular movement behavior is consistent with walking.

Figure 8a shows a movement behavior in which the mean of the turning angles is close to zero with a directional persistence of around zero. A comparison with the corresponding boxplot of persistence velocity values in Figure 6 (Rest), suggests that the particular movement behavior is consistent with resting. By similar reasoning and visual inspection of Figure 6 (Forage) and Figure 8b, we can conclude that the remaining movement behavior is consistent with foraging behavior. The latter observations are characterized by low 'Persistence velocity' values and an overall mean direction vector with a non-zero angle, i.e., a sinuous type of motion, indicating a tendency to reverse direction.

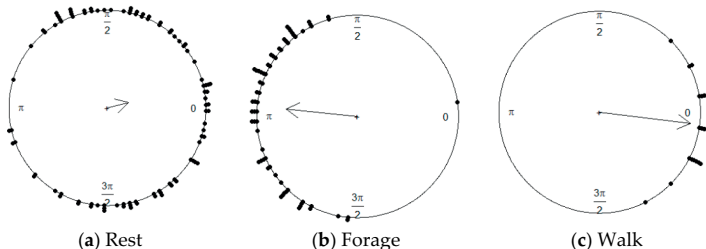


Figure 8. Distribution of turning angles for each detected movement behavior; each dot on the perimeter of the circles represent the mean of the turning angles for one segment. The mean value of the turning angles for foraging behavior (b) tends to be near π radians, meaning sinuous segments. While the mean value of turning angles for resting (a) and walking (c) movement behaviors tends to zero, with high variance and low variance, respectively (inverse relationship with the arrow length).

By comparing the derived movement behaviors to the ground truth movement behavior recorded by a human observer, we obtain a validation of the resulting model in terms of its discriminative power between behavioral states. The model performance is presented by a confusion matrix (Table 1) that shows how the number of minutes for each type of observed behavior is distributed according to the predictions of our model. Each row in the confusion matrix represents an observed movement behavior, while each column represents a particular movement behavior predicted by our model. Hence, each cell counts the number of minutes in the intersection of these two observed and predicted movement behaviors. The diagonal entries of the confusion matrix represent the correct predictions, and off-diagonal values represent the misclassifications. The sum of values on the diagonal divided by the sum of all the matrix values shows the accuracy. Thus, using the demonstration data, the accuracy obtained from the confusion matrix in Table 1 is 77%. Analysis of seven GPS recording days where corresponding “observationlogger” data are available indicates an averaged accuracy for our dataset of $80.75\% \pm 5.9\%$ (SE). For these seven days, the number of hours of corresponding “observationlogger” data varied from 2 h to 8 h, with an average of 4.6 h. The total number of valid GPS points was 39,123.

Table 1. Confusion matrix showing the performance of the clustering. Precision = True Positives/(True Positives + False Positives). Precision (Forage) = 87.6%, Precision (Rest) = 68.7%, Precision (Walk) = 70%.

		Predicted		
		Forage	Rest	Walk
Real	Forage	184 (min)	73 (min)	15 (min)
	Rest	26 (min)	160 (min)	0
	Walk	0	0	35 (min)

6. Concluding Remarks

With this paper, our goal has been to identify activity patterns of individual animals over a 24 h time span. We demonstrate how changes in movement behavior can be inferred from the tracks of individual animals, and how the movement behavior of individual animals varies over time. The method for discriminating between different animal behaviors described in this paper requires no specialist coding experience or statistical background. The source code is provided as a Supplementary File.

In contrast to other methods proposed for the same purpose, our method requires neither a temporal regularization of the data nor a strict a priori value for the number of movement behavioral states, nor the specification of the movement model. The proposed method is exploratory and depends only on location in space and time, with no need for ancillary data like accelerator sensor values. The heuristics described for choosing the moving window size (see Section 3.1.1) and the number of clusters (see Section 5) seem difficult to replace by completely objective criteria. Validation of the selection procedure by using the proposed diagnostic plots (Figure 3), box plots (Figure 6), and histograms (Figure 8) is therefore urgent for our confidence in the goodness of these values.

Consideration of the diurnal activity pattern for individual sheep over several days should improve the possibilities of detecting deviant movement behavior for individual animals. Some cautionary steps are, however, required before applying our method to large datasets. It is well-known from the literature that applying hierarchical clustering to large datasets is challenging, but some advanced and efficient alternatives do exist (e.g., [62]). An alternative is running the hierarchical clustering on a manageable subset of the included animals. The mean/overall distribution for each of the (k) resulting clusters can then be used to initialize a k-means clustering process (using the ks-metric) to cluster the complete dataset into k clusters.

In spite of its accuracy, being sensitive to the uncertainty in the location data and the derived movement parameters, we claim that our method for identifying movement behavioral states from tracking data has proven successful in practice. There are, however, uncertainties associated with the identification and recording of activities during field observations. In addition, the matching of field

observations and tracking data must be executed carefully, to prevent loss of accuracy in the resulting models. Further work is required to address these issues in the best possible way.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2220-9964/7/2/78/s1>. Source code S1 in the R programming language: ‘manuscript-supplementary.R’.

Acknowledgments: We want to thank the farmer for his inspiring cooperation and valuable help regarding use of his animals and help to change GPS collars during the fieldwork period. We would also like to thank Morten Sickel for facilitating data collection and making the data collected available for the analyses in this paper, and colleagues in NIBIO for conscientious collection of observation data. We are very thankful to The Norwegian Agriculture Agency, which was the main financier of the project that provided data to this work. This research was supported by the Norwegian University of Life Sciences under project number 1301051701.

Author Contributions: H.S. designed, performed the field experiments. M.T. developed the theoretical framework, did the programming task and drafted the manuscript. H.T. and U.G.I. assisted in writing the manuscript. H.T. supervised the findings of this work and was in charge of overall direction and coordination between all authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cooke, S.J.; Hinch, S.G.; Wikelski, M.; Andrews, R.D.; Kuchel, L.J.; Wolcott, T.G.; Butler, P.J. Biotelemetry: A mechanistic approach to ecology. *Trends Ecol. Evol.* **2004**, *19*, 334–343. [[CrossRef](#)] [[PubMed](#)]
2. Dray, S.; Royer-Carenzi, M.; Calenge, C. The exploratory analysis of autocorrelation in animal-movement studies. *Ecol. Res.* **2010**, *25*, 673–681. [[CrossRef](#)]
3. Morales, J.M.; Haydon, D.T.; Frair, J.; Holsinger, K.E.; Fryxell, J.M. Extracting more out of relocation data: Building movement models as mixtures of random walks. *Ecology* **2004**, *85*, 2436–2445. [[CrossRef](#)]
4. Edelhoff, H.; Signer, J.; Balkenhol, N. Path segmentation for beginners: An overview of current methods for detecting changes in animal movement patterns. *Mov. Ecol.* **2016**, *4*, 21. [[CrossRef](#)] [[PubMed](#)]
5. Patterson, T.A.; Thomas, L.; Wilcox, C.; Ovaskainen, O.; Matthiopoulos, J. State-space models of individual animal movement. *Trends Ecol. Evol.* **2008**, *23*, 87–94. [[CrossRef](#)] [[PubMed](#)]
6. Dodge, S.; Weibel, R.; Lautensch, A.-K. Towards a taxonomy of movement patterns. *Inf. Vis.* **2008**, *7*, 240–252. [[CrossRef](#)]
7. Guo, Y.; Poulton, G.; Corke, P.; Bishop-Hurley, G.J.; Wark, T.; Swain, D.L. Using accelerometer, high sample rate GPS and magnetometer data to develop a cattle movement and behaviour model. *Ecol. Model.* **2009**, *220*, 2068–2075. [[CrossRef](#)]
8. de Weerd, N.; van Langevelde, F.; van Oeveren, H.; Nolet, B.A.; Kölzsch, A.; Prins, H.H.; de Boer, W.F. Deriving animal behaviour from high-frequency GPS: Tracking cows in open and forested habitat. *PLoS ONE* **2015**, *10*, e0129030. [[CrossRef](#)] [[PubMed](#)]
9. Gurarie, E.; Andrews, R.D.; Laidre, K.L. A novel method for identifying behavioural changes in animal movement data. *Ecol. Lett.* **2009**, *12*, 395–408. [[CrossRef](#)] [[PubMed](#)]
10. Han, J.; Lee, J.-G.; Kamber, M. An overview of clustering methods in geographic data analysis. In *Geographic Data Mining and Knowledge Discovery*, 2nd ed.; Miller, H.J., Han, J., Eds.; CRC Press: Boca Raton, FL, USA, 2009; pp. 149–187.
11. Schick, R.S.; Loarie, S.R.; Colchero, F.; Best, B.D.; Boustany, A.; Conde, D.A.; Halpin, P.N.; Joppa, L.N.; McClellan, C.M.; Clark, J.S. Understanding movement data and movement processes: Current and emerging directions. *Ecol. Lett.* **2008**, *11*, 1338–1350. [[CrossRef](#)] [[PubMed](#)]
12. Gurarie, E.; Bracis, C.; Delgado, M.; Meckley, T.D.; Kojola, I.; Wagner, C.M. What is the animal doing? Tools for exploring behavioural structure in animal movements. *J. Anim. Ecol.* **2016**, *85*, 69–84. [[CrossRef](#)] [[PubMed](#)]
13. Demšar, U.; Buchin, K.; Cagnacci, F.; Safi, K.; Speckmann, B.; Van de Weghe, N.; Weiskopf, D.; Weibel, R. Analysis and visualisation of movement: An interdisciplinary review. *Mov. Ecol.* **2015**, *3*, 5. [[CrossRef](#)] [[PubMed](#)]
14. Postlethwaite, C.M.; Brown, P.; Dennis, T.E. A new multi-scale measure for analysing animal movement data. *J. Theor. Biol.* **2013**, *317*, 175–185. [[CrossRef](#)] [[PubMed](#)]
15. Schwager, M.; Anderson, D.M.; Butler, Z.; Rus, D. Robust classification of animal tracking data. *Comput. Electron. Agric.* **2007**, *56*, 46–59. [[CrossRef](#)]

16. Jonsen, I.D.; Flemming, J.M.; Myers, R.A. Robust state-space modeling of animal movement data. *Ecology* **2005**, *86*, 2874–2880. [[CrossRef](#)]
17. Gutenkunst, R.; Newlands, N.; Lutcavage, M.; Edelstein-Keshet, L. Inferring resource distributions from Atlantic bluefin tuna movements: An analysis based on net displacement and length of track. *J. Theor. Biol.* **2007**, *245*, 243–257. [[CrossRef](#)] [[PubMed](#)]
18. Fauchald, P.; Tveraa, T. Using first-passage time in the analysis of area-restricted search and habitat selection. *Ecology* **2003**, *84*, 282–288. [[CrossRef](#)]
19. Tremblay, Y.; Roberts, A.J.; Costa, D.P. Fractal landscape method: An alternative approach to measuring area-restricted searching behavior. *J. Exp. Biol.* **2007**, *210*, 935–945. [[CrossRef](#)] [[PubMed](#)]
20. Jonsen, I.D.; Myers, R.A.; Flemming, J.M. Meta-analysis of animal movement using state-space models. *Ecology* **2003**, *84*, 3055–3063. [[CrossRef](#)]
21. Zhang, J.; O'Reilly, K.M.; Perry, G.L.W.; Taylor, G.A.; Dennis, T.E. Extending the Functionality of Behavioural Change-Point Analysis with k-Means Clustering: A Case Study with the Little Penguin (*Eudyptula minor*). *PLoS ONE* **2015**, *10*, e0122811. [[CrossRef](#)] [[PubMed](#)]
22. Lavielle, M. Detection of multiple changes in a sequence of dependent variables. *Stoch. Process. Their Appl.* **1999**, *83*, 79–102. [[CrossRef](#)]
23. Rinzivillo, S.; Pedreschi, D.; Nanni, M.; Giannotti, F.; Andrienko, N.; Andrienko, G. Visually driven analysis of movement data by progressive clustering. *Inf. Vis.* **2008**, *7*, 225–239. [[CrossRef](#)]
24. Lee, J.-G.; Han, J.; Whang, K.-Y. Trajectory clustering: A partition-and-group framework. In Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, Beijing, China, 11–14 June 2007; ACM: New York, NY, USA, 2007; pp. 593–604.
25. Miller, H.J.; Han, J. *Geographic Data Mining and Knowledge Discovery*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2009.
26. Etienne, L.; Devoegele, T.; Bouju, A. Spatio-temporal trajectory analysis of mobile objects following the same itinerary. *Adv. Geo-Spat. Inf. Sci.* **2012**, *10*, 47–57.
27. Nanni, M.; Pedreschi, D. Time-focused clustering of trajectories of moving objects. *J. Intell. Inf. Syst.* **2006**, *27*, 267–289. [[CrossRef](#)]
28. Dodge, S.; Laube, P.; Weibel, R. Movement similarity assessment using symbolic representation of trajectories. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1563–1588. [[CrossRef](#)]
29. Yanagisawa, Y.; Akahani, J.-I.; Satoh, T. Shape-Based Similarity Query for Trajectory of Mobile Objects. In Proceedings of the 4th International Conference on Mobile Data Management, Melbourne, Australia, 21–24 January 2003; Chen, M.-S., Chrysanthis, P.K., Sloman, M., Zaslavsky, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 63–77.
30. Vlachos, M.; Gunopulos, D.; Das, G. Rotation invariant distance measures for trajectories. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 22–25 August 2004; ACM: New York, NY, USA, 2004; pp. 707–712.
31. Chen, L.; Özsu, M.T.; Oria, V. Robust and fast similarity search for moving object trajectories. In Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, Baltimore, Maryland, 14–16 June 2005; ACM: New York, NY, USA, 2005; pp. 491–502.
32. Vlachos, M.; Gunopoulos, D.; Kollios, G. Discovering Similar Multidimensional Trajectories. In Proceedings of the 18th International Conference on Data Engineering, San Jose, CA, USA, 26 February–1 March 2002; p. 673.
33. Alt, H.; Guibas, L.J. Discrete geometric shapes: Matching, interpolation, and approximation. *Handb. Comput. Geom.* **1999**, *1*, 121–153.
34. Alt, H.; Godau, M. Computing the Fréchet distance between two polygonal curves. *Int. J. Comput. Geom. Appl.* **1995**, *5*, 75–91. [[CrossRef](#)]
35. Andrienko, G.; Andrienko, N.; Wrobel, S. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explor. Newsl.* **2007**, *9*, 38–46. [[CrossRef](#)]
36. Pelekis, N.; Andrienko, G.; Andrienko, N.; Kopanakis, I.; Marketos, G.; Theodoridis, Y. Visually exploring movement data via similarity-based analysis. *J. Intell. Inf. Syst.* **2012**, *38*, 343–391. [[CrossRef](#)]
37. Thiebault, A.; Tremblay, Y. Splitting animal trajectories into fine-scale behaviorally consistent movement units: Breaking points relate to external stimuli in a foraging seabird. *Behav. Ecol. Sociobiol.* **2013**, *67*, 1013–1026. [[CrossRef](#)]

38. Cagnacci, F.; Boitani, L.; Powell, R.A.; Boyce, M.S. Animal ecology meets GPS-based radiotelemetry: A perfect storm of opportunities and challenges. *R. Soc.* **2010**. [[CrossRef](#)] [[PubMed](#)]
39. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
40. Gurarie, E. Behavioral Change Point Analysis in R: The bcpa package (R package version 1.1). Available online: <https://pdfs.semanticscholar.org/dc3e/3c9baac39d228f1dd2de4b35431395c76fd3.pdf> (accessed on 22 February 2018).
41. Calenge, C.; Dray, S.; Royer-Carenzi, M. The concept of animals' trajectories from a data analysis perspective. *Ecol. Inf.* **2009**, *4*, 34–41. [[CrossRef](#)]
42. Benhamou, S. How to reliably estimate the tortuosity of an animal's path: Straightness, sinuosity, or fractal dimension? *J. Theor. Biol.* **2004**, *229*, 209–220. [[CrossRef](#)] [[PubMed](#)]
43. Laube, P.; Purves, R.S. How fast is a cow? Cross-scale analysis of movement data. *Trans. GIS* **2011**, *15*, 401–418. [[CrossRef](#)]
44. Murtagh, F. A survey of recent advances in hierarchical clustering algorithms. *Comput. J.* **1983**, *26*, 354–359. [[CrossRef](#)]
45. Berkhin, P. A survey of clustering data mining techniques. *Group. Multidimens. Data* **2006**, *25*, 71.
46. Van Moorter, B.; Visscher, D.R.; Jerde, C.L.; Frair, J.L.; Merrill, E.H. Identifying movement states from location data using cluster analysis. *J. Wildl. Manag.* **2010**, *74*, 588–594. [[CrossRef](#)]
47. Garriga, J.; Palmer, J.R.; Oltra, A.; Bartumeus, F. Expectation-maximization binary clustering for behavioural annotation. *PLoS ONE* **2016**, *11*, e0151984. [[CrossRef](#)] [[PubMed](#)]
48. Braun, E.; Geurten, B.; Egelhaaf, M. Identifying prototypical components in behaviour using clustering algorithms. *PLoS ONE* **2010**, *5*, e9361. [[CrossRef](#)] [[PubMed](#)]
49. Hurford, A. GPS measurement error gives rise to spurious 180 turning angles and strong directional biases in animal movement data. *PLoS ONE* **2009**, *4*, e5632. [[CrossRef](#)] [[PubMed](#)]
50. Pewsey, A.; Neuhäuser, M.; Ruxton, G.D. *Circular Statistics in R*; Oxford University Press: Oxford, UK, 2013.
51. Jammalamadaka, S.R.; Sengupta, A. *Topics in Circular Statistics*; World Scientific: Singapore, 2001.
52. Rachev, S.T.; Klebanov, L.B.; Stoyanov, S.V.; Fabozzi, F.J. Probability distances and probability metrics: Definitions. In *The Methods of Distances in the Theory of Probability and Statistics*; Springer: Cham, Switzerland, 2013; pp. 11–31.
53. Rokach, L. A survey of clustering algorithms. In *Data Mining and Knowledge Discovery Handbook*; Maimon, O., Rokach, L., Eds.; Springer: New York, NY, USA, 2010; pp. 269–298.
54. Calenge, C. The package “adehabitat” for the R software: A tool for the analysis of space and habitat use by animals. *Ecol. Model.* **2006**, *197*, 516–519. [[CrossRef](#)]
55. Kareiva, P.; Shigesada, N. Analyzing insect movement as a correlated random walk. *Oecologia* **1983**, *56*, 234–238. [[CrossRef](#)] [[PubMed](#)]
56. Deza, M.M.; Deza, E. Encyclopedia of distances. In *Encyclopedia of Distances*; Springer: Cham, Switzerland, 2009; pp. 1–583.
57. Gibbs, A.L.; Su, F.E. On choosing and bounding probability metrics. *Int. Stat. Rev.* **2002**, *70*, 419–435. [[CrossRef](#)]
58. Murtagh, F.; Legendre, P. Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *J. Classif.* **2014**, *31*, 274–295. [[CrossRef](#)]
59. Bridges, C.C., Jr. Hierarchical cluster analysis. *Psychol. Rep.* **1966**, *18*, 851–854. [[CrossRef](#)]
60. Xu, R.; Wunsch, D. Survey of clustering algorithms. *IEEE Trans. Neural Netw.* **2005**, *16*, 645–678. [[CrossRef](#)] [[PubMed](#)]
61. Köhn, H.-F.; Hubert, L.J. Hierarchical Cluster Analysis. In *Wiley StatsRef: Statistics Reference Online*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2014.
62. Vijaya, P.; Murty, M.N.; Subramanian, D. Leaders—Subleaders: An efficient hierarchical clustering algorithm for large data sets. *Pattern Recognit. Lett.* **2004**, *25*, 505–513. [[CrossRef](#)]



Research Paper III

Teimouri, M., Indahl, U. G., Sickel, H., & Tveite, H. (2018). Estimation of Temporal Sampling Frequency in Movement Parameter Estimation: A Stochastic Movement Modeling Approach. Submitted to *Ecological Applications*.

Estimation of Temporal Sampling Frequency in Movement Parameter Estimation: A Stochastic Movement Modeling Approach

Maryam Teimouri ^{1,*}, Ulf Geir Indahl ¹, Hanne Sickel ² and Håvard Tveite ¹

¹ Faculty of Science and Technology, Norwegian University of Life Sciences, Drøbakveien 31, 1433 Ås, Norway; Maryam.teimouri@nmbu.no; ulf.indahl@nmbu.no; havard.tveite@nmbu.no

² Norwegian Institute of Bioeconomy Research, Postboks 115, NO-1431 Ås, Norway; hanne.sickel@nibio.no

* Correspondence: maryam.teimouri@nmbu.no; Tel.: +47-67231547

Abstract

A discrete-time animal movement path can be reconstructed by connecting the sampled points with straight lines. The accuracy of movement parameters derived from the reconstructed movement path depends on several factors, including temporal sampling frequency, accuracy of the measuring device and species-specific movement characteristics. We quantify the accuracy of the measuring device and species-specific movement characteristics in a stochastic movement model. By using the stochastic movement model, we explore the sensitivity of the derived movement parameters to the temporal sampling frequency of the observations. Finally, we propose a method to estimate the best temporal sampling frequency (Representative Temporal Sampling Frequency, RTSF) in terms of minimizing the difference between the derived movement parameter and the true movement parameter.

1 Introduction

The technological revolution in tracking devices allows us to capture the trajectories of moving objects with increasing spatial accuracy and temporal sampling frequency. Spatial accuracy is a measure of how close recorded positions are to the “true” positions.

High temporal sampling frequency requires high battery capacity or frequent battery replacement. High battery capacity might impose a too heavy weight, exceeding the carrying capacity of the animal (at most 5% of the animal's body mass (1)). Frequent battery replacement requires recapturing, which might be either infeasible, or pose ethical concerns about the animal welfare (2). Therefore, battery capacity and battery weight limitations still put constraints on the temporal sampling frequency in practical field studies. Hence, prescheduling temporal sampling frequency when using animal-borne tracking devices for any research question is in high demand. Although the literature indicates that sampling frequency is selected based on the experience of domain experts (3, 4), the impact of sampling frequency has been explored in various research fields such as (i) animal movement distribution and behavior (4-8) (ii) fishing vessels (9) (iii) transportation research (10) and (vi) pedestrian movement (11). Depending on the research question, the complexity of movement and the accuracy of the measuring device, the optimal temporal sampling frequency can vary considerably (6, 7, 9, 10).

Our main research question is the following: Given an animal species and a device for location measurement, what is the best temporal sampling frequency when the goal is that the reconstructed movement path shall have smallest possible difference from the true movement path in terms of movement parameter estimation (for example speed)?

To decide the required temporal sampling frequency, given the measuring device, we will utilize a species-specific stochastic movement model reflecting and quantifying species-specific movement paths and accuracy of locations. The model includes a set of parameters that are computed from high frequency sampling of the movement path of the species of interest (in its habitat) for a short observation period. We use a fitted model to simulate the species-specific movements at varying sampling frequencies, to determine the sampling frequency that minimizes the difference between observed and estimated data in terms of speed calculations.

We start our investigation by exploring different sources of errors affecting animal movement path reconstruction. Section 3 reviews the most frequently used ecology movement models in the literature. Section 4 presents the proposed method to determine the best sampling frequency for animal movement path reconstruction using stochastic

movement modeling and cross-scale analysis. In Section 5 we explore the applicability of our method on real tracking datasets and evaluates the implications of sampling frequency on speed calculations, and finally section 6 concludes the article.

2 Background

Animal movement results in a continuous trajectory or path through space and time. Correspondingly, animal movement data consist of sequences of sampled point locations (x_i, y_i) of an individual collected at time t_i . Methods for estimating the expected movement path between consecutive samples differ vastly with respect to the underlying statistical assumptions and the output produced (12-14). The Brownian Bridge Movement Model (12) and the Continuous-Time Correlated Random Walk Model (14) represent some of the alternatives. Frequently, animal trajectories are reconstructed by connecting consecutive observation points with straight lines (15, 16). There are two types of errors affecting the accuracy of path reconstruction on a discrete-time scale (17, 18): *measurement error* and *interpolation error*. *Measurement error* (19) refers to the spatial uncertainty associated with the observed samples while *interpolation error* is influenced by the temporal sampling frequency and the movement characteristics of the observed object. It is widely agreed that path reconstruction from tracking data is affected by *measurement error* and *interpolation error* at high and low temporal sampling frequencies, respectively (6, 7, 9, 10). With high temporal sampling frequencies, *measurement error* becomes significant, leading to positive bias in the straight-line distance calculation (18). The overestimation depends on the spatiotemporal autocorrelation of the *measurement error* (larger positive autocorrelations means reduced distance calculation bias and larger negative autocorrelation means increased distance calculation bias), the variance of the *measurement error* (larger variance means larger distance calculation bias) and the Euclidean distance between consecutive points (longer distance means smaller distance calculation bias) (18, 20-22). Conversely, with low temporal sampling frequencies, the effect of the *measurement error* is ignorable, but the impact of *interpolation error* is significant. At low temporal sampling frequencies, tortuous (meandering) paths are more realistic than straight lines between points (23). At some intermediate sampling frequency, we expect the impact of the combined

measurement and *interpolation error* to be minimized, so that the straight line would deviate the least from the true path.

Many movement parameters are based on the reconstructed path (24), including step length (straight-line distance between successive samples), speed (step length divided by time interval), direction, and turning angle (change of direction between successive steps). Movement parameters are used in many types of studies ranging from wildlife movement behavior studies (25) to wildlife movement simulations (26) and wildlife movement modeling (27).

Laube and Purves (7) explored the sensitivity of the movement parameters to *measurement error* and the chosen sampling frequency, which they termed '*granularity grief*'.

Pépin *et al.* (4) quantified the dependence of the estimated daily distance travelled by red deer *Cervus elaphus* on sampling frequency. They fit an asymptotic curve to the relationship between apparent distance and sampling frequency. The estimated asymptote is claimed to be the true representation of daily distance travelled.

Jerde and Visscher (28) quantified the impact of *measurement error* on the estimation of movement parameters. High sampling frequencies with inaccurate samples introduces large errors in movement parameter estimation.

Therefore, we consider temporal sampling frequencies and the accuracy of the measuring devices together with the movement characteristics of the animal of interest as fundamental factors in sensible movement path reconstruction and movement parameter estimation.

A short period of location recordings provides us with valuable information about the accuracy of the measuring device and the movement characteristics of the animal of interest. Our ambition is to present general guidelines for choosing the sampling frequency that best explains the true movement path traveled by the animal in terms of movement parameter (speed) estimation.

By identifying Representative Temporal Sampling Frequency (RTSF), movement path reconstruction and movement parameter estimation will be manageable task.

Correspondingly, the longest possible duration of monitoring can be found conditional on the available battery capacity and the carrying capacity of the animal.

3 Related work

The movement of an individual organism is influenced by four major components: 1. internal state (why move?), 2. motion capacity (how to move?), 3. navigation capacity (when and where to move?), and 4. the (abiotic and biotic) environmental factors affecting the movement (29, 30). The first three components relate to the organism and the last one relates to the environment within which the movement takes place. Since modeling all the constituting components of movement is infeasible in practice, various analytic techniques, including Simple Random Walk models or Composite Random Walk models (31-35), have long served as the standard approach to wildlife movement modelling.

In the following we will first consider traditional discrete-time modeling and discuss its limitations. Then, we will explain the advantages of continuous-time stochastic modeling.

3.1 Discrete-Time Stochastic Process

Discrete-time stochastic movement models such as Random walk models are breaking the movement of an animal down into two components: 1. the step lengths, and 2. the turning angles between successive steps along with their associated distributions (32). Correlated Random Walk and Random Walk models are distinguished by the presence or absence of correlation between successive step directions, termed local 'persistence' in direction (36). Persistence is defined as the tendency of an animal to continue moving in the same direction (37). Note that at a sufficiently high sampling frequency, all movements are necessarily correlated, and at a low enough sampling frequency, all movements tend to be uncorrelated (38).

Biased Correlated Random Walk models (31) are characterized by a global directional bias to maintain a global direction of the movement from an origin toward a destination, in addition to the local persistence in direction.

Typically, a well-defined probability density function of random velocities and turning angles is used to characterize the movement stochastically (38). The directional persistence

is controlled by the probability distribution of the turning angles (37). Random Walk models and their variations require regularly sampled movement data. Thus, pre-processing is required for irregularly sampled data (14). Another limitation of variations of Random Walk models is that the required parameter estimation depends on the sampling schedule. For example, variations of Random Walk models can be fitted to exactly the same movement path by considering different temporal frequencies in the sampled data (39, 40). Therefore, drawing sampling-independent inferences from movement data is challenging. The shortcomings of Random Walk models and their variations get worse when long-term autocorrelation must be taken into account in order to reveal latent movement patterns, including repetition and memory (40, 41).

3.2 Continuous-Time Stochastic Processes (CTSPs)

Continuous-time stochastic processes (CTSP) (39, 40) aim at resolving several limitations of their discrete-time counterparts. Recent CTSP models consider the continuous-time movement process and the discrete-time sampling process separately (40, 42). They model movement path as a stochastic process on a continuous time scale, in which the irregularly spaced data points can simply be seen as discrete samples in time. In the following, we first briefly present different variation of CTSPs, and then formulate them in terms of semivariance functions.

3.2.1 CTSP models

To classify CTSP models (39) the following three features are used: 1. position autocorrelation, 2. velocity autocorrelation, and 3. range residency. Samples that are close together in time have a stronger tendency to resemble each other than those farther apart. Therefore, relocation datasets mostly feature ‘position autocorrelation’. Finely sampled datasets also usually feature ‘velocity autocorrelation’ resulting in *directional persistence*. The latter means that an individual’s direction of motion at one point in time tends to be correlated with those at adjacent times. Finally, at larger time scales, most animals will exhibit a tendency to remain in a defined region or ‘home range’. In the following table, the most frequently used CTSP models for different features of animal movement in the ecological literature are categorized by whether or not they feature position

autocorrelation, velocity autocorrelation and home range. The interested reader is referred to (40, 43-45) for more details.

Table 1. A summary of the most frequently used CTSP models, categorized according to whether or not position autocorrelation (Pos. AC), velocity autocorrelation (Vel. AC) and home range (H. Range)(39) are included.

Movement Model	Pos. AC	Vel. AC	H. Range
Independent Identically Distributed process (IID)	No	No	Yes
Brownian Motion (BM)	Yes	No	No
Ornstein-Uhlenbeck (OU)	Yes	No	Yes
Integrated OU (IOU)	Yes	Yes	No
Ornstein-Uhlenbeck Foraging (OUF)	Yes	Yes	Yes

3.2.2 Mean Squared Displacement (MSD) versus Semivariance Function (SVF)

Classical RW analyses focus on quantifying step length distributions, turn-angle distributions, and comparing the observed mean squared displacement (MSD) to the model-predicted MSD (27, 31). MSD for a trajectory is a function of time that for each time duration provides the mean of the squared distances between pairs of points separated by that amount of time (32, 46). MSD can be used to distinguish between diffusive, super-diffusive, and sub-diffusive movements. In diffuse movement, MSD scales linearly with time or path length, but sub- or super-diffusive movement has a power-law relationship with time (47). As already mentioned, an issue that seems to be ignored in RW analyses is that both turning angle distributions and step length distributions are dependent on the choice of sampling frequency.

The semivariance function (SVF), a nonparametric autocorrelation estimator that is very similar in nature to MSD, quantifies the variability in the distance between two samples as a function of time lag τ (Equation 1). $x(t)$ represents the location at time t (40).

$$\gamma(\tau) = \frac{1}{2} * E((x(t + \tau) - x(t))^2) \quad (1)$$

Plotting the estimated semivariance as a function of time lag yields an ‘empirical variogram’ (48), which contains information on the autocorrelation structure in a relocation dataset.

CTSP models can be expressed in terms of their SVFs. The SVFs of the most frequently used CTSP models in the literature are the following:

- IID: Since autocorrelation is zero for all time lags other than 0, the semivariance is constant for all time lags other than zero.
- BM: Semivariance is a linear function of time.
- OU: Semivariance is a mixture of a linear function of time over short scales, $\tau < \tau_H$, and approaches a constant for longer time lags ($\tau > \tau_H$)
- IOU: Semivariance is a mixture of a power law function (Curve upward) for small time lags, $\tau < \tau_F$, and a linear function for longer time lags, $\tau > \tau_F$
- OUF: Semivariance is a mix of a power law function for small time lags, $\tau < \tau_F$, a linear function for intermediate time lags, $\tau_F < \tau < \tau_H$, and then approaches a constant for longer time lags ($\tau > \tau_H$)

For the sake of movement modeling, CTSP semivariance models are fitted to the ‘empirical variogram’ via regression (49), and the best-fit model is selected based on standard model selection techniques (40).

4 The proposed method

The choice of Representative Temporal Sampling Frequency (RTSF) should obviously be close to the timescale that governs the movement process. In the following, we explain how to determine the RTSF by using a continuous-time stochastic process and cross-scale analysis to study the sensitiveness of reconstructed movement paths to sampling frequency.

4.1 Model parameterization

A location dataset for an individual animal can be considered as a realization of a stochastic process. According to the Semivariance approach for animal movement modeling initiated by (40), we use the continuous-time movement modelling (ctmm) package¹ (39) for the R programming language to find the underlying stochastic process. First, the ‘empirical

¹ Github repository URL: <https://github.com/ctmm-initiative/ctmm>

variogram' for the available dataset is used to reveal the main features of the movement, including velocity autocorrelation (short-lag feature), position autocorrelation (intermediate-lag feature), and range residency (long-lag feature). Using maximum likelihood and AIC (Akaike Information Criterion)-based model selection, a semivariance model is fitted to the 'empirical variogram' and the appropriate movement model is determined (40). Once the best model for the data in terms of AIC has been selected, the parameterized stochastic movement model can be used as a simulator across different (sub)sampling schedules to do cross-scale analysis to derive the appropriate RTSF.

4.2 Measurement error

For densely sampled tracking data, it is necessary to consider *measurement error* in the model fitting. HDOP is a standard measure of GNSS spatial accuracy, based on the number of satellites used for each fix and their relative position in the sky (6). GNSS animal tracking devices usually return HDOP values. HDOP values can be used to translate the device's UERE (User Equivalent Range Errors) to telemetry errors. UERE is device specific and represents the RMS error given ideal satellite conditions. Since information about the UERE for the devices used in our study were unavailable, we used data from a stationary test (where the device has been left fixed over a period of time (50)) to estimate the UERE from our calibration data by the *uere()* function in the *ctmm* package (39). The residuals of the UERE calibration model are calculated by the associated *residuals()* function. The output of *residuals()* function can be processed further by the *correlogram()* function to assess the error autocorrelation. A correlogram (autocorrelation plot) is used to show the error autocorrelation versus the time lags.

4.3 Cross-scale analysis and simulation

The search for a RTSF requires a reference trajectory, ideally calculated with a more accurate measuring system (10). Due to the lack of a reference movement path for our experiment, we decided to split our dataset into a training- and test set. The subsequent model fittings were based on the training trajectory dataset and cross-scale analysis using the test trajectory dataset.

The observed trajectory: $\{(t_1, x_1, y_1), (t_2, x_2, y_2), \dots, (t_n, x_n, y_n)\}$ is split into training (the initial 80%) and test (the final 20%) dataset as follows:

$$\text{Training dataset: } \{(t_1, x_1, y_1), (t_2, x_2, y_2), \dots, (t_i, x_i, y_i); i = \lceil 0.8n \rceil\}$$

Where $\lceil \cdot \rceil$ denotes rounding up to the nearest integer.

$$\text{Test dataset: } \{(t_{i+1}, x_{i+1}, y_{i+1}), (t_{i+2}, x_{i+2}, y_{i+2}), \dots, (t_n, x_n, y_n)\}$$

The speed values for each segment of the test dataset is calculated as follows:

$$V_{j\text{-observed}} = \sqrt{(x_{j+1} - x_j)^2 + (y_{j+1} - y_j)^2} / (t_{j+1} - t_j) \text{ for } j = i + 1, \dots, n - 1$$

A stochastic movement model is fitted to the coordinates and times of the training trajectory dataset (see 3.2 and 4.1). From the fitted model we simulate trajectories $\{(t_{i+1}, x'_{i+1}, y'_{i+1}), (t_{i+2}, x'_{i+2}, y'_{i+2}), \dots, (t_n, x'_n, y'_n)\}$ to be compared with the true observation in the test dataset.

The simulated speed values for each segment of the simulated trajectory is calculated in the same way as for the observed trajectory:

$$V'_{j\text{-simulated}} = \sqrt{(x'_{j+1} - x'_j)^2 + (y'_{j+1} - y'_j)^2} / (t_{j+1} - t_j) \text{ for } j = i + 1, \dots, n - 1$$

The simulated trajectories are generated through conditional simulation based on the test dataset (Figure 1). Therefore, the simulated trajectories have the same sampling schedule as the test dataset.

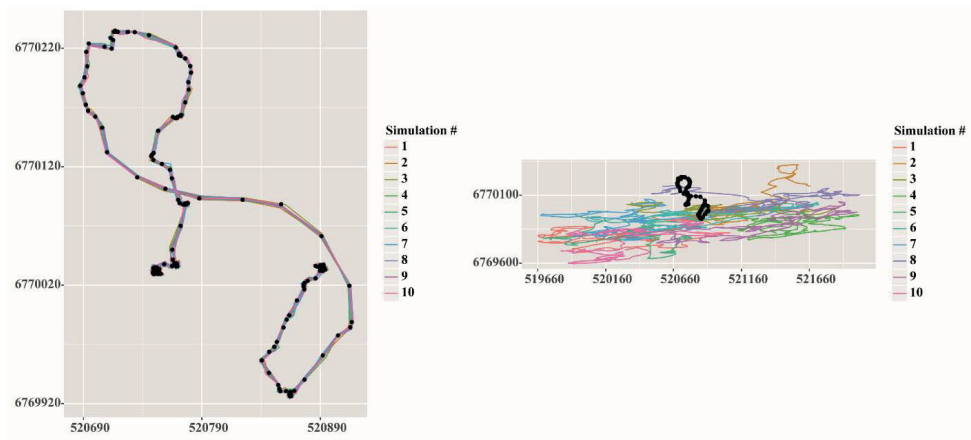


Figure 1. 10 simulated movement paths by Conditional Gaussian Simulation (left) and Purely Gaussian Simulation (right) according to a common test dataset (as shown by black dots). Color is used to differentiate between the simulations.

For each sampling frequency we simulate 1000 trajectories and calculate the average speed for each segment.

The impact of sampling frequency on the movement parameter of interest, i.e. speed, is quantified by comparing the speed values of the observed trajectory and the mean simulated speed values for corresponding segments of the simulated trajectories across the sampling frequencies. Our selection rule for the RTSF is simply to choose the finest among the available sampling frequencies where the average of the simulated speed values and corresponding observed speed values are statistically indistinguishable.

5 Experimental Result

In the following, we first describe our dataset, and then demonstrate how the proposed method works in practice. We will also discuss the effect of sampling frequency on the temporal variation of speed to estimate how fast an animal moves.

5.1 Experimental dataset

The dataset analyzed in this paper is part of a project where the main objective is to determine the influence of grazing on biodiversity.

Four individual sheep were followed during the foraging season in an established experimental site in Valdres in Oppland county, Norway (61.06° N, 9.40° E) - Figure 2 (left). All animals were tracked with a sampling interval of 10 seconds using Canmore GT-740FL GPS devices². These devices have a battery capacity of approximately 24 hours when sampling data using this sampling interval. The GPS devices were equipped with extra battery which extended the capacity to 3-4 days. The GPS spatial accuracy is specified as 2.5 m CEP³ under optimal conditions.

The usable position data for four sheep (#24, #10001, #30011 and #8029) were collected on 15, 14, 9 and 4 different days respectively, totaling up to approximately 26 days of observations. The total number of position observations was 227050.

We recorded calibration data with the Canmore GT-740FL GPS device to estimate both UERE (see 4.2 for details), and the shortest temporal interval for which we can ignore the error autocorrelation in our observations (the `ctmm` package does not yet include modelling of error autocorrelation). Therefore, we must assess the error autocorrelation in our observations before proceeding to the model fitting. To limit the effects of error autocorrelation in the observations, we therefore perform subsamplings of the observations to identify a sampling interval in which the error autocorrelation is negligible. According to our calibration data, recorded between 17:10 and 19:40 on 2018/08/22 with approximately 10 second time intervals, some error autocorrelation are present in the observations, see Figure 3 (top). The top left panel of Figure 3 shows that there is significant error autocorrelation for the lags up to 2 minutes. The top right panel of Figure 3 shows a zoomed-in view of the same error autocorrelation plot over the shorter time lags. The bottom panel of Figure 3 shows that the error autocorrelations for the 1 minute time interval subsample of the observations for all time-lags up to 30 minutes are statistically ignorable, i.e. within the upper and lower confidence bands drawn by red dotted dash-lines.

² http://www.canmore.com.tw/pdf/GT-740FL%20DataSheet_V6.1.pdf

³ CEP(Circular Error Probable) is the radius of the smallest circle centered at the true position that covers 50% of the observations

We therefore conclude that by subsampling our data to a 1 minute time interval, the impact of the error autocorrelation in our calculations can be ignored.

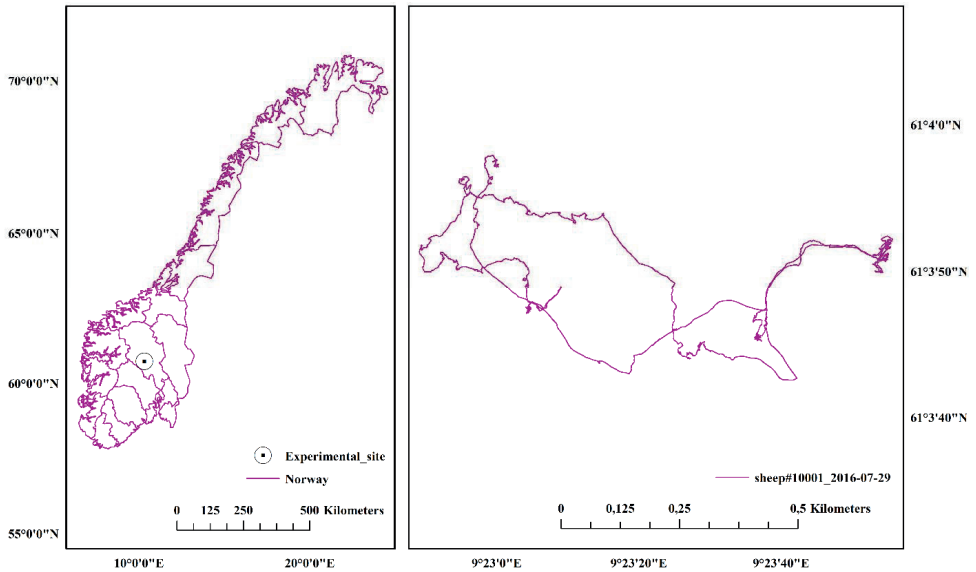


Figure 2. Map of Norway (source: Statens kartverk) with the location of the experimental site (left), an example trajectory for sheep#10001 over one day (2016 July 29th) (right)

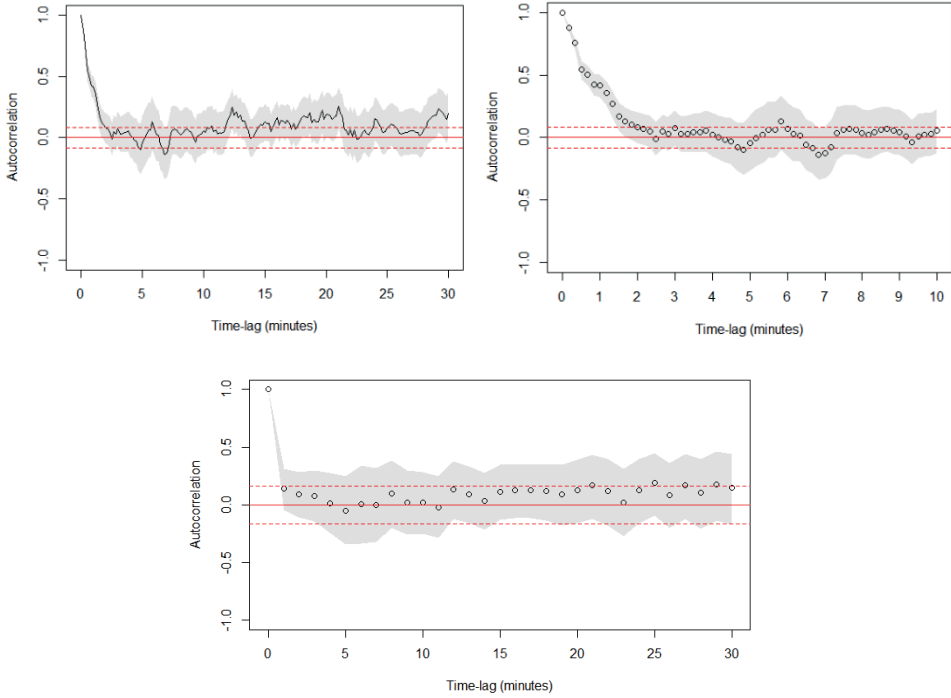


Figure 3. Error autocorrelation plot for the observations with 10 second sampling interval (top-left) and its zoomed-in view (top right), Error autocorrelation plot for the observations with 1 minute sampling interval (bottom).

5.2 Demonstration of the proposed method

To demonstrate the proposed method, we use all the data for sheep#10001 (14 observation days including 4 observation periods with some gaps, recorded in 2016 from June 28th until August 26th) and a small version providing continuous coverage of a period of 4 days (2016 July 28th-July 31st). The total number of observed data points for the whole dataset is 14717. The training dataset is chosen to contain the initial 80% of observations (11774 data points) leaving the test dataset with the final 20% of observations, i.e. 2943 data points. The number of observation in the small dataset is 4247, with a corresponding split into 3398 training data points and 849 test data points.

By using the `variogram()` function from the `ctmm` package (39), an *empirical variogram* for the training dataset is created. Fitting a semivariance model to an empirical variogram via interactive sliders on a plot using the `ctmm.fit()` function provides us with some rough

initial parameters for the movement model. Thereafter, an appropriate movement model using maximum likelihood and AIC (Akaike Information Criterion)-based model selection is calculated.

The top panel of Figure 4 shows the empirical variogram (black signal) and the fitted semivariance function (red curve) obtained from the training part of the whole dataset for sheep#10001.

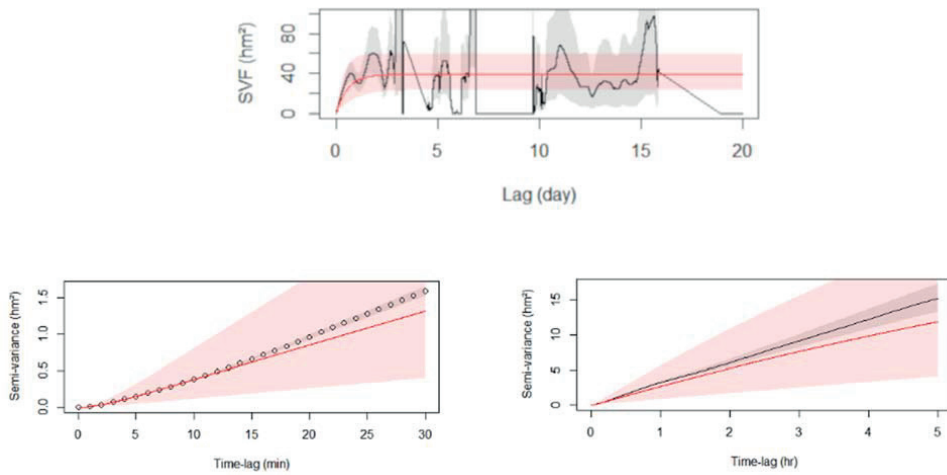


Figure 4. The fitted semivariance function (red curve) against the empirical variogram (black curve) of sheep#10001 for the whole monitoring period (top), and two zoomed in views produced by the ctmm package with the range of 30 minutes (bottom left) and 5 hours (bottom right). The y-axis shows the samivariance estimate (Equation 1) at different time lags, represented on the x-axis. The gray shade shows the confidence interval of estimation.

The selected model is an OUF type of model which is the only alternative that fits the entire variogram (while the other model alternatives listed in Table 1 are missing at least one of the variogram’s key features). Note that the shortest lags represent the most reliable semivariance estimations, so one should pay most attention to shorter time-lags when selecting a model, see the bottom panel of Figure 4.

The selected movement model is shown by the red curve in Figure 4. This model is used to simulate coordinate values at various (sub)sampling frequencies (as described in 4.3) to be compared with the test dataset at corresponding (sub)sampling frequencies.

We subsampled the test dataset into 2, 3, 4, 5, 6, 7, 8, 9 and 10 minute intervals by including every 2nd, 3rd, 4th, 5th, 6th, 7th, 8th, 9th, and 10th of the observed points, respectively. For each subsample, we simulated trajectories for the corresponding sampling intervals. For each choice of sampling interval, we then calculated the observed speed for each segment of the associated observed trajectory dataset and the average of the corresponding simulated trajectory speeds. Then, we compared their distribution statistically to test for statistical significance in the differences between the two.

Because there is temporal autocorrelation in the speed values obtained from the samples, the assumption of independence which underlies most of the traditional statistical test (such as t-test) is violated. Alternative non-parametric statistical tests of significance such as the Wilcoxon–Mann–Whitney test (51) do not assume any specific distribution, and are therefore more useful for our purposes. We have therefore used the Wilcoxon–Mann–Whitney test with a ranked comparison of two independent samples in our analysis. We tested sampling intervals with data point sequences ranging from 1 minute to 10 minutes (in steps of 1 minute) to explore the difference between the mean simulated speed values and observed speed values at the corresponding sampling frequency when using a significance level of 0.01. The null hypothesis (H_0) is that the distributions of the observed speed values and the average of simulated speed values for the various segments are indistinguishable. The alternative hypothesis (H_1) is that the distributions of the observed speed values and the mean simulated speed values are distinguishable.

Table 2 shows the result of the Wilcoxon–Mann–Whitney test for the whole dataset at different sampling intervals. At sampling intervals shorter than 8 minutes we reject the null hypothesis and accept the alternative hypothesis, meaning there is a significant difference in the distribution of observed and simulated speed values (p -value < 0.01). The shortest sampling interval in which the Wilcoxon–Mann–Whitney test results in not rejecting the null hypothesis (p -value > 0.01) is 8 minutes. Therefore, we conclude that a temporal sampling frequency of 7.5 fixes per hour (8 minute time interval) is the RTSF for the particular animal, habitat and accuracy of measuring device.

Table 2. The result of the Wilcoxon–Mann–Whitney test at different sampling intervals for sheep#10001 for the whole monitoring period

Sampling interval (in minutes)	H_0 = The distributions of the observed and simulated speed values are indistinguishable H_1 = The distributions of the observed and simulated speed values are distinguishable
1,2,3,4,5,6,7	p-value<0.01
8+	p-value>0.01

Large datasets are currently not available, but we were able to explore the potential of our method for smaller datasets by using the tracking data of four consecutive days (between 2016 July 28th and July 31st). Figure 5 shows the ctm model obtained from the training dataset.

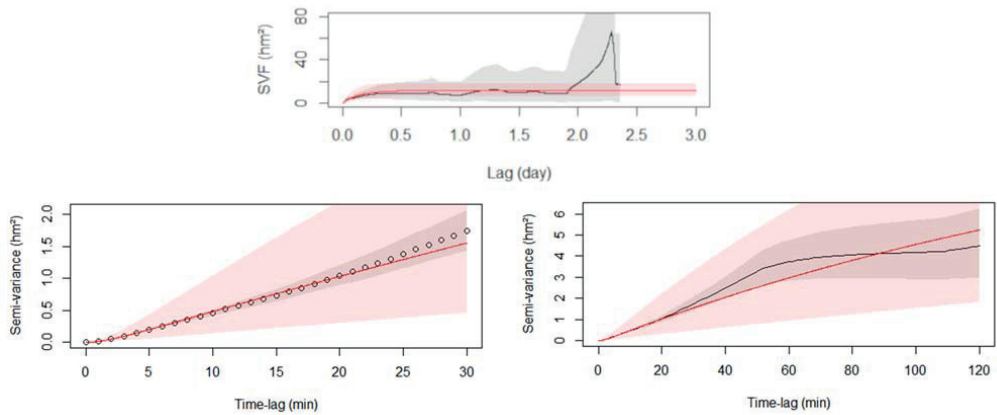


Figure 5. The fitted semivariance function (red curve) against the empirical variogram (black curve) of sheep#10001 for the short monitoring period (2016 July 28th-July 31st) (top), and two zoomed in views produced by the ctm package with the range of 30 minutes (bottom left) and 2 hours (bottom right). The definition of the x-axis and y-axis are as in Figure 4.

For sampling intervals up to and including 6-minutes there is a significant difference (p-value<0.01) in the distribution of observed speed values and mean simulated speed values (Table 3) due to measurement error. However, for sampling interval longer than 6 minutes our tests conclude no significant difference (p-value>0.01) in the distribution of observed speed values and mean simulated speed values. We therefore conclude the RTSF is 8.6 fixes per hour (time interval of 7 minutes).

Table 3. The result of the Wilcoxon–Mann–Whitney test at different sampling intervals for sheep#10001 for the specific observation period (2016 July 28th-July 31st)

Sampling interval (in minutes)	H ₀ = The distributions of the observed and simulated speed values are indistinguishable
	H ₁ = The distributions of the observed and simulated speed values are distinguishable
1,2,3,4,5,6	p-value<0.01
7+	p-value>0.01

We have compared these results with similar results obtained for the other three available sheep. They all suggests that 7 or 8 minutes is the appropriate sampling interval for this specific type of sheep in this habitat with our measuring device.

5.3 Speed calculation

In order to demonstrate the influence of sampling interval on calculated speed, we decided to investigate the trajectory of sheep #10001 for one specific day (2016-07-29), see the right panel of Figure 2. Figure 6 shows the daily temporal variation of speed for the indicated sampling time intervals. The y-axis represents the speed of the sheep. The x-axis represents the time of day. Figure 6 shows that by increasing the sampling interval, the level of detail for speed variation and the magnitude of speed values decreases (Figure 6 (a, b, c, d)). The magnitude is clearly reduced by at least 50% when the sampling interval is increasing from 1 minute (Figure 6 (a)) to 10 minutes (Figure 6 (d)). This does not necessarily mean that the speed variation at higher sampling frequencies is representative of the truth. For example, Figure 6 (a) indicates fluctuations in speed at all times, but there should be at least some consecutive zero speed values corresponding to resting periods during a typical day. Non-zero speed values during a whole day are not reasonable. According to our analysis and simulations, the representative temporal sampling frequency for the combination of this type of sheep in its specific habitat and with our measuring device is 8.6 or 7.5 fixes per hour (time intervals of 7 or 8 minutes).

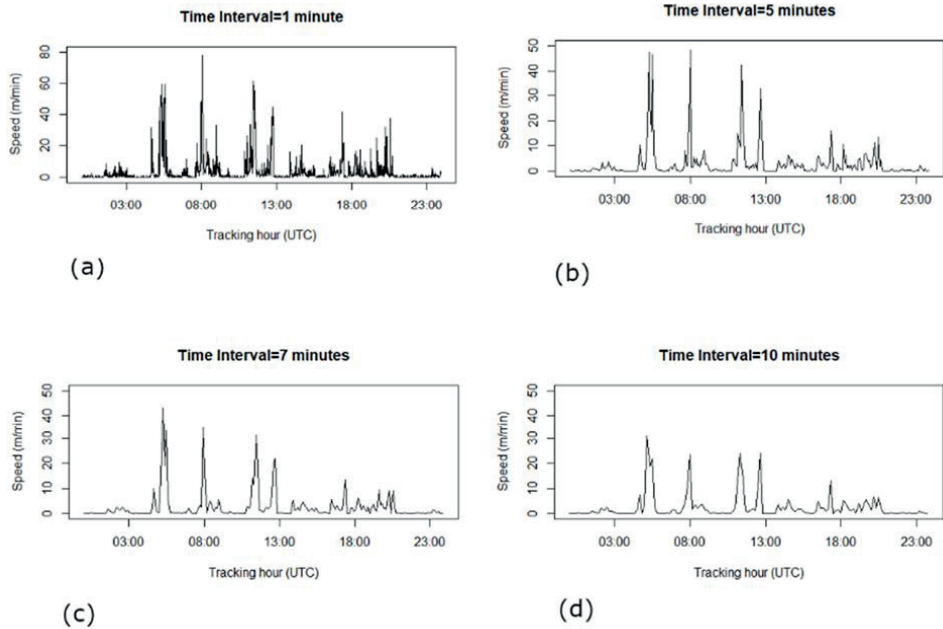


Figure 6. The daily speed variation for sheep#10001 on a specific date (2016_07_29) for four selected sampling intervals: (a) 1-minute time interval, (b) 5 minutes time interval, (c) 7 minutes time interval and (d) 10 minutes time interval

6 Conclusion

Densely sampled data allows detailed insight into fine-scale behaviors, but are more sensitive to potential *measurement errors*. In contrast, too coarsely sampled data will lose fine-scale behaviors but be less affected by *measurement error*. The *measurement error* at high sampling frequencies and *interpolation error* at lower sampling frequencies cause the speed to be overestimated and underestimated, respectively. Finding a Representative Temporal Sampling Frequency (RTSF) that avoids both over- and under-sampling in terms of speed calculation was the motivation behind this article.

By viewing animal movement paths as realizations of a continuous stochastic process, we were able to incorporate the effects of position autocorrelation, velocity autocorrelation, range residency and *measurement error* in the movement model selection. By using the selected movement model as a simulation device, we have been able to demonstrate how to calculate Representative Temporal Sampling Frequency for a combination of a specific

species in its habitat given a specific type of tracking device. The calculated sampling frequency reflects the scale of the animal movement in general, the degree of autocorrelation in its movements and the accuracy of the tracking device.

Our approach is quite simple and general, thus it could also be useful for other animal movement data and tracking devices. The strength of our approach is that the model is parameterized using tracking data of the animal of interest in its habitat. By assuming that the movement behaviors persist over longer spans of time, the resulting movement model(s) should be appropriate for the design of sensible and useful logging schedules.

We believe that our method should also be useful for investigating the impact of sampling frequency on other movement parameters such as turning angle.

Acknowledgements

We would like to thank Dr. Christen H. Fleming from University of Maryland, one of the developers of the `ctmm` R package, who provided us with technical assistance in working with the `ctmm` package.

This research was supported by the Norwegian University of Life Sciences under project number 1301051701.

References

1. Technitis G, Othman W, Safi K, Weibel R. From A to B, randomly: a point-to-point random trajectory generator for animal movement. *Int J Geogr Inf Sci.* 2015;29(6):912-34.
2. Urbano F, Cagnacci F. *Spatial Database for GPS Wildlife Tracking Data: A Practical Guide to Creating a Data Management System with PostgreSQL/PostGIS and R*: Springer Science & Business Media; 2014.
3. Ungar ED, Henkin Z, Gutman M, Dolev A, Genizi A, Ganskopp D. Inference of animal activity from GPS collar data on free-ranging cattle. *Rangeland Ecology & Management.* 2005;58(3):256-66.
4. Pépin D, Adrados C, Mann C, Janeau G. Assessing real daily distance traveled by ungulates using differential GPS locations. *Journal of Mammalogy.* 2004;85(4):774-80.
5. Johnson DD, Ganskopp DC. GPS collar sampling frequency: effects on measures of resource use. *Rangeland ecology & management.* 2008;61(2):226-31.
6. Ryan P, Petersen S, Peters G, Grémillet D. GPS tracking a marine predator: the effects of precision, resolution and sampling rate on foraging tracks of African Penguins. *Marine biology.* 2004;145(2):215-23.

7. Laube P, Purves RS. How fast is a cow? cross-scale analysis of movement data. *Transactions in GIS*. 2011;15(3):401-18.
8. Codling E, Hill N. Sampling rate effects on measurements of correlated and biased random walks. *Journal of Theoretical Biology*. 2005;233(4):573-88.
9. Palmer MC. Calculation of distance traveled by fishing vessels using GPS positional data: A theoretical evaluation of the sources of error. *Fisheries Research*. 2008;89(1):57-64.
10. Ranacher P, Brunauer R, van der Spek S, Reich S. What is an appropriate temporal sampling rate to record floating car data with a GPS? *ISPRS International Journal of Geo-Information*. 2016;5(1):1.
11. Schneider C, Zutz S, Rehr K, Brunauer R, Gröchenig S. Evaluating GPS sampling rates for pedestrian assistant systems. *Journal of Location Based Services*. 2016;10(3):212-39.
12. Horne JS, Garton EO, Krone SM, Lewis JS. Analyzing animal movements using Brownian bridges. *Ecology*. 2007;88(9):2354-63.
13. Winter S, Yin Z-C. Directed movements in probabilistic time geography. *Int J Geogr Inf Sci*. 2010;24(9):1349-65.
14. Johnson DS, London JM, Lea M-A, Durban JW. Continuous-time correlated random walk model for animal telemetry data. *Ecology*. 2008;89(5):1208-15.
15. Calenge C, Dray S, Royer-Carenzi M. The concept of animals' trajectories from a data analysis perspective. *Ecological informatics*. 2009;4(1):34-41.
16. Long JA, Nelson TA. A review of quantitative methods for movement data. *Int J Geogr Inf Sci*. 2013;27(2):292-318.
17. Schneider M, editor *Uncertainty management for spatial datain databases: Fuzzy spatial data types*. International Symposium on Spatial Databases; 1999: Springer.
18. Ranacher P, Brunauer R, Trutschnig W, Van der Spek S, Reich S. Why GPS makes distances bigger than they are. *Int J Geogr Inf Sci*. 2016;30(2):316-33.
19. Hofmann-Wellenhof B, Legat K, Wieser M. *Navigation: principles of positioning and guidance*: Springer Science & Business Media; 2011.
20. El-Rabbany A, Kleusberg A. Effect of temporal physical correlation on accuracy estimation in GPS relative positioning. *Journal of surveying engineering*. 2003;129(1):28-32.
21. Howind J, Kutterer H, Heck B. Impact of temporal correlations on GPS-derived relative point positions. *Journal of Geodesy*. 1999;73(5):246-58.
22. Wang J, Satirapod C, Rizos C. Stochastic assessment of GPS carrier phase measurements for precise static relative positioning. *Journal of Geodesy*. 2002;76(2):95-104.
23. Marcus Rowcliffe J, Carbone C, Kays R, Kranstauber B, Jansen PA. Bias in estimating animal travel distance: the effect of sampling frequency. *Methods in Ecology and Evolution*. 2012;3(4):653-62.
24. Fleming CH, Fagan WF, Mueller T, Olson KA, Leimgruber P, Calabrese JM. Estimating where and how animals travel: an optimal framework for path reconstruction from autocorrelated tracking data. *Ecology*. 2016;97(3):576-82.
25. Edelhoff H, Signer J, Balkenhol N. Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns. *Movement Ecology*. 2016;4(1):21.

26. Ahearn SC, Dodge S, Simcharoen A, Xavier G, Smith JL. A context-sensitive correlated random walk: a new simulation model for movement. *Int J Geogr Inf Sci.* 2017;31(5):867-83.
27. Turchin P. *Quantitative Analysis of Movement: measuring and modeling population redistribution in plants and animals.* Sinauer Associates, Sunderland, MA; 1998.
28. Jerde CL, Visscher DR. GPS measurement error influences on movement model parameterization. *Ecological Applications.* 2005;15(3):806-10.
29. Demšar U, Buchin K, Cagnacci F, Safi K, Speckmann B, Van de Weghe N, et al. Analysis and visualisation of movement: an interdisciplinary review. *Movement ecology.* 2015;3(1):5.
30. Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, et al. A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences.* 2008;105(49):19052-9.
31. Codling EA, Plank MJ, Benhamou S. Random walk models in biology. *Journal of the Royal Society Interface.* 2008;5(25):813-34.
32. Kareiva P, Shigesada N. Analyzing insect movement as a correlated random walk. *Oecologia.* 1983;56(2):234-8.
33. Viswanathan GM, Afanasyev V, Buldyrev S, Murphy E, Prince P, Stanley HE. Lévy flight search patterns of wandering albatrosses. *Nature.* 1996;381(6581):413-5.
34. Marsh LM, Jones RE. The form and consequences of random walk movement models. *Journal of Theoretical Biology.* 1988;133(1):113-31.
35. Morales JM, Haydon DT, Frair J, Holsinger KE, Fryxell JM. Extracting more out of relocation data: Building movement models as mixtures of random walks. *Ecology.* 2004;85(9):2436-45.
36. Patlak CS. Random walk with persistence and external bias. *Bulletin of Mathematical Biology.* 1953;15(3):311-38.
37. Bartumeus F, da Luz ME, Viswanathan G, Catalan J. Animal search strategies: A quantitative random-walk analysis. *Ecology.* 2005;86(11):3078-87.
38. Gurarie E. *Models and analysis of animal movements: From individual tracks to mass dispersal [Doctoral Thesis]: University of Washington; 2008.*
39. Calabrese JM, Fleming CH, Gurarie E. ctmm: an r package for analyzing animal relocation data as a continuous-time stochastic process. *Methods in Ecology and Evolution.* 2016;7(9):1124-32.
40. Fleming CH, Calabrese JM, Mueller T, Olson KA, Leimgruber P, Fagan WF. From fine-scale foraging to home ranges: a semivariance approach to identifying movement modes across spatiotemporal scales. *The American Naturalist.* 2014;183(5):E154-E67.
41. Fagan WF, Calabrese JM. The correlated random walk and the rise of movement ecology. *The Bulletin of the Ecological Society of America.* 2014;95(3):204-6.
42. Fleming CH, Calabrese JM, Mueller T, Olson KA, Leimgruber P, Fagan WF. Non-Markovian maximum likelihood estimation of autocorrelated movement processes. *Methods in Ecology and Evolution.* 2014;5(5):462-72.
43. Dunn JE, Gipson PS. Analysis of radio telemetry data in studies of home range. *Biometrics.* 1977:85-101.
44. Gurarie E, Ovaskainen O. Characteristic spatial and temporal scales unify models of animal movement. *The American Naturalist.* 2011;178(1):113-23.

45. Pozdnyakov V, Meyer T, Wang Y-B, Yan J. On modeling animal movements using Brownian motion with measurement error. *Ecology*. 2014;95(2):247-53.
46. Morales JM, Ellner SP. Scaling up animal movements in heterogeneous landscapes: the importance of behavior. *Ecology*. 2002;83(8):2240-7.
47. Almeida PJ, Vieira MV, Kajin M, Forero-Medina G, Cerqueira R. Indices of movement behaviour: conceptual background, effects of scale and location errors. *Zoologia (Curitiba)*. 2010;27(5):674-80.
48. Barnes R. Variogram tutorial. Golden Software, Inc. 2004.
49. Cressie N. *Statistics for spatial data*: John Wiley & Sons; 2015.
50. Frair JL, Fieberg J, Hebblewhite M, Cagnacci F, DeCesare NJ, Pedrotti L. 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*. 2010;365(1550):2187-200.
51. Neuhäuser M. Wilcoxon–mann–whitney test. *International encyclopedia of statistical science*: Springer; 2011. p. 1656-8.

ISBN: 978-82-575-1564-5

ISSN: 1894-6402



Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no