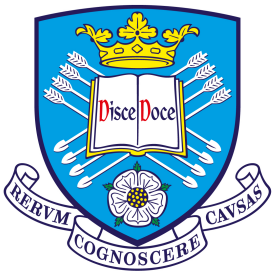

The University Of Sheffield.

Dude, where's my cow? Using high-frequency movement data to quantify animal space use

Rhys Munden

Submitted for the degree of Doctor of Philosophy

School of Mathematics and Statistics

September 2020

Supervisor: Dr. Jonathan Potts

University of Sheffield

Acknowledgements

Firstly, I would like to thank the Leverhulme Trust Studentship for funding this project. I also offer my eternal gratitude to my supervisor Jonathan Potts for all his help throughout my PhD and particularly for his patience with me.

I thank too my co-supervisors in Swansea, Luca Börger and Rory Wilson, for all of their help in this project and for always making me feel welcome whenever I visited. Also from Swansea I thank James Redcliffe for collecting then dead reckoning the cattle and goat data used in Chapters 3 and 4. My thanks also goes to Natasha Ellison who provided helpful feedback on this thesis and spurred me on throughout the writing process.

I am also most grateful to all those who have prayed for me from those in the university's Christian Union and my Growth Group at Christ Church Endcliffe. Most of all I thank God for sustaining me throughout my PhD.

Finally I thank my parents and sister for their support and putting up with me throughout the writing process during lockdown.

Abstract

In this thesis I investigate the use of high frequency ($\geq$1Hz) location data for examining animal movements. This high frequency means that an animal's movement path can often be considered pseudo-continuous, which removes the necessity for interpolating between successive recordings. On the other hand, many existing techniques used to analyse animal movement data are either too computationally expensive or not directly applicable to high frequency data. I introduce novel adaptations of two techniques, namely the residence time metric and Step Selection Analysis (SSA), which I tailor for use on high frequency data.

The residence time metric is used to estimate how long an animal spends in different areas. I explain how using high frequency data enables us to accurately calculate how much time an animal spends in a clearly delimited area. I also show how, through adaptations to the existing method used to calculate residence times, we can greatly reduce its computational time. Furthermore, this adapted method enables us to identify specific areas that are extensively used by an animal, which provides further opportunities to examine why an animal chooses to use those areas so much.

SSA enables a user to infer the reasons underpinning an animal's decisions to move from one location to another. I explain how it is not useful to apply SSA directly to high frequency data, since there is often little interesting information contained in movements between consecutive recordings that are a second or less apart. Therefore, I first show how high frequency data can be rarefied such that the rarefied points represent times when the animal has made key movement decisions. I then adapt the SSA method for use with such rarefied paths. In comparing the existing and adapted SSA methods I find that the previous SSA method can lead to inaccurate results.

Contents

List of Tables

List of Figures

Nomenclature

List of Abbreviations

- ARS - Area Restricted Search
- CLR - Conditional Logistic Regression
- EHU site - Extremely high-use site
- GPS - Global Positioning System
- iSSA - integrated Step Selection Analysis
- KDE - Kernel Density Estimation
- OU - Ornstein-Uhlenbeck
- PDF(s) - Probability Distribution Function(s)
- SCSD - Squared Circular Standard Deviation
- SSA - Step Selection Analysis
- tiSSA - time-varying integrated Step Selection Analysis
- UD - Utilisation Distribution
- VHF - Very High Frequency

Chapter 1

Introduction

Movement ecology is a key aspect of a diverse range of research areas, such as epidemiology (Altizer et al., 2011), wildlife management (Allen and Singh, 2016) and biodiversity conservation (Jeltsch et al., 2013). Being able to identify movement corridors is critical to conservation studies (Chetkiewicz and Boyce, 2009; Sawyer et al., 2009). It is important to incorporate the movements of hosts in disease models (Merkle et al., 2018). Fundamental to these applications of movement ecology is understanding the underlying motivations for animals' movements.

Animals move for a whole host of different reasons. Their decisions to move may be affected by their internal state, interactions with other animals or the environment (Nathan et al., 2008). The challenge we are faced with is in identifying these decisions and their causes from data on animal locations over time.

Until recently, animal tracking has typically been collected at frequencies of one every few minutes or hours (see Table 1 in Thurfjell et al. (2014)), but due to recent technological advances (Williams et al., 2020) it is possible to obtain data on animal locations at much higher frequencies ($\geq 1\,\text{Hz}$), often gathered over very long time intervals (sometimes up to a year). The fine resolution of such data allows for movements on very small spatio-temporal scales (e.g. micro-movements; Wilson et al. (2014)) to be studied, whilst the great length of such time-series enable decisions over larger spatio-temporal scales to be studied using the same datasets (e.g. migration; Hedenström et al. (2016)). Hence such data enable us to analyse movements across a wide range of spatio-temporal scales. The aim of this thesis is to show how such data can enable us to gain greater insight into the motivations underlying an animal's decisions to move.

Many pre-existing techniques for analysing animals' movements were designed with lower resolution (on the order of minutes or hours) data in mind. In this thesis I adapt two of these techniques, namely the residence time metric (Barraquand and Benhamou, 2008) and Step Selection Analysis (SSA; Fortin et al. (2005); Rhodes et al. (2005)), specifically for use on high resolution data. The residence time metric provides an estimate of how much time an animal spends in a particular area, so it is used to identify areas that are extensively used. In calculating the residence time Barraquand

and Benhamou (2008) assumes that areas where an animal spends long periods of time are used for a specific purpose (e.g. foraging), therefore identifying these areas is a first step in determining their use.

Other methods such as state-space modelling (Jonsen et al., 2005, 2007; Patterson et al., 2008) and hidden Markov models (Morales et al., 2004; Langrock et al., 2012) not only identify areas where an animal is exhibiting interesting behaviours, but they also determine which behaviours are being exhibited. Both state-space modelling and hidden Markov models require prior knowledge on how an animal's underlying behaviour links to the movement mode identified from different movement metrics (e.g. see Table 1 in Edelhoff et al. (2016)). In contrast, the residence time metric requires no such prior knowledge, so it is a useful way of determining the areas that are most used by an animal without imposing any assumptions on why these areas are being used. Although the residence time is therefore a useful metric, it turns out that applying existing algorithms to calculate this time take a prohibitively long time to run. Furthermore, to calculate the residence time, in Barraquand and Benhamou (2008) they interpolate between consecutive recordings, which is not useful when recordings are a second or less apart. Therefore I adapt their method used to calculate residence times for higher frequency data, since it is unnecessary to interpolate and the computational costs need reducing.

The other technique I adapt for high frequency data is SSA. This is used to understand animals' underlying movement drivers between consecutive relocations by testing hypotheses concerning why an animal chooses to move to one location rather than any other available to them. However, it makes little sense to apply this to locations that are less than a second apart as there is likely to be little interesting information contained within such very small "steps". Often an animal will simply be carrying on with a decision it made sometime earlier. In this thesis I therefore present an adaptation of SSA that makes use of the high resolution of modern data sets to find places where decisions to move are likely to have been made and infer information about the drivers behind these decisions.

§ 1.1 Ultra-High-Resolution Animal Movement Data

Animal movement data consists of sequences of recorded locations that I write as $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N$, where $\mathbf{x}_i$ is recorded at time t_i. This sequence of locations is termed either a *path* or a *trajectory*. Locations can be collected using a wide variety of methods (Kenward, 2000), including very high frequency (VHF) radiotracking, Argos and global positioning systems (GPS). The movement between two consecutive recorded locations is often referred to as a *step*, each characterised by its *step length* which is the distance between the two locations. The *turning angles* are defined to be the difference in headings of successive steps (Fig. 1.1).

VHF radiotracking was one of the first methods used to remotely record animal locations (Craighead, 1979; Craighead et al., 1995), which enabled the tracking of wild animals in their natural habitat without being too obtrusive. VHF radiotracking has two components; a tracking device, which is light enough for it to even be attached to

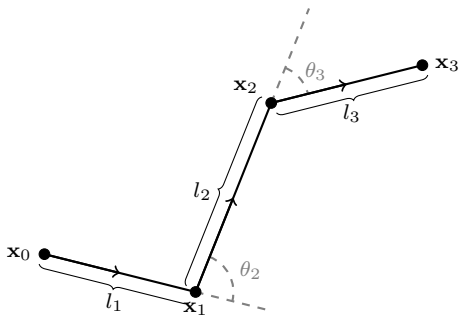

Figure 1.1: **Demonstration of movement metrics**
A segment of a movement trajectory, with recorded locations $\mathbf{x}_i$. The step length l_i is the distance from $\mathbf{x}_{i-1}$ to $\mathbf{x}_i$. The turning angle at each $\mathbf{x}_i$ is denoted as θ_i.

some small insects (e.g. dragonflies; Wikelski et al. (2006)) and a receiver to collect the recordings. Therefore the VHF system is limited to areas near to a receiver, which can either be mobile (Wikelski et al., 2003) or fixed to a tower (Kays et al., 2011). If the receiver is fixed then the spatial range is limited, which makes it impossible to track large scale movements, whereas mobile receivers can often be impractical since they require a researcher to follow the animal.

Argos and GPS are forms of satellite tracking systems, where locations are determined by triangulation from satellites orbiting the Earth. Therefore these systems provide global coverage, which enables them to be used over larger spatial scales than VHF. On the other hand, Argos and GPS devices tend to weigh far more than VHF radiotracking tags (Kays et al., 2011). The larger weight of Argos and GPS tags limits the range of animals that can be tracked using these tags, since there is a recommended tag weight limit of 3-5% of the animal's body weight (Kenward, 2000; Bodey et al., 2018). However, there have been recent technological advances which have decreased the size of GPS devices enabling smaller animals to be tracked (Kays et al., 2015), but they still typically weigh more than VHF (e.g. in Smith et al. (2018) the GPS tags they use weigh twice as much as the VHF tags).

The Argos system allows for real-time tracking making it an expensive system (Bridge et al., 2011). On the other hand, there is no need to recapture the animal to retrieve the Argos tag and the data stored within it, which can be a limitation of some GPS tags (Moen et al., 1996; Haines et al., 2006; Steinfurth et al., 2008). This has made Argos tags the ideal choice for tracking animal's long range movements (e.g migration; Luschi et al. (1996); Kjellén et al. (1997)), whereas GPS tags have typically been used to answer questions concerning more localised movements, such as foraging and habitat selection (see Table 1 in Wilmers et al. (2015)).

More recent technological advances have enabled researchers to gather animal relocation data at sub-second resolutions using magnetometers and accelerometers (Wilson et al., 2008). Accelero-magnetometer data can be processed using Dead Reckoning

(Bidder et al., 2015) to obtain estimates of an animal's speed and heading, from which their path is reconstructed. There is a known locational error in the procedure which can accumulate over time, so a ground truthing method is used to correct for these errors. Often this ground truthing method will involve using GPS recordings (Dewhirst et al., 2016), but other methods (e.g. VHF or video recordings; (Bidder et al., 2012, 2015)) can be used when GPS is unsuitable. For example, GPS transmissions can fail due to thick canopy cover and terrain obstructions (Rempel et al., 1995; Moen et al., 1996; Dussault et al., 1999; D'Eon et al., 2002; Hebblewhite et al., 2007).

Since the frequency of these accelero-magnetometer data are so high, the distances between successive relocations are often less than the animal's body length, which means the paths from these data can be considered to be pseudo-continuous. For example the average distances between successive locations gathered at a resolution of 1Hz for cattle and goats are 0.082 and 0.096m respectively. Figure 1.2 shows how intricately detailed these paths are and the inset shows how close these locations typically are. Usually when Argos or GPS recordings are used there is a need to interpolate between successive recordings if it is necessary to infer the animal's location between recordings. However, the distances between accelero-magnetometer recordings are so short it removes the necessity to interpolate. This can be beneficial for various applications, including calculating residence time and SSA, as I show in this thesis.

There are great advantages in using accelero-magnetometer data over GPS data. The transmission of GPS data can be limited by the environment (e.g. thick canopy cover), which limits the usefulness of GPS tags on taxa living in these environments. Accelero-magnetometer data can easily be used across a wide range of taxa: terrestrial (Bidder et al., 2012; Street et al., 2018), marine (Tanaka et al., 2001; Noda et al., 2014) and aerial (Shepard et al., 2008; Williams et al., 2015). The power necessary to transmit GPS data can be great, which means that either the study duration is shortened or a larger battery is used. Compared to GPS devices, accelero-magnetometer data can be collected using a far smaller battery. For example, a GPS device collecting locations at 1Hz may require 30-50 mA of current, whereas accelero-magnetometer data collected at 40Hz require 5-10 mA of current (Bidder et al., 2015). Therefore, accelero-magnetometer data can be collected for long periods of time (e.g. weeks or months; Wilson et al. (2013*a*); Street et al. (2018)), without compromising on the battery size or sampling frequency.

§ 1.2 Site Fidelity

Due to the heterogeneous nature of the environment in which animals live, certain areas may be more suitable for an animal than others (Barraquand and Benhamou, 2008). An animal's needs may change through time (Getz and Saltz, 2008) thus also changing the suitability of these areas. For example, the most suitable area for foraging may be different from the most suitable resting area, where avoidance of predators may be the key concern of the animal. Temporal variations in the environment can also drive animal movement. For example, many ungulates migrate in response to seasonal changes in vegetation (Leimgruber et al., 2001; Boone et al., 2006; Hebblewhite

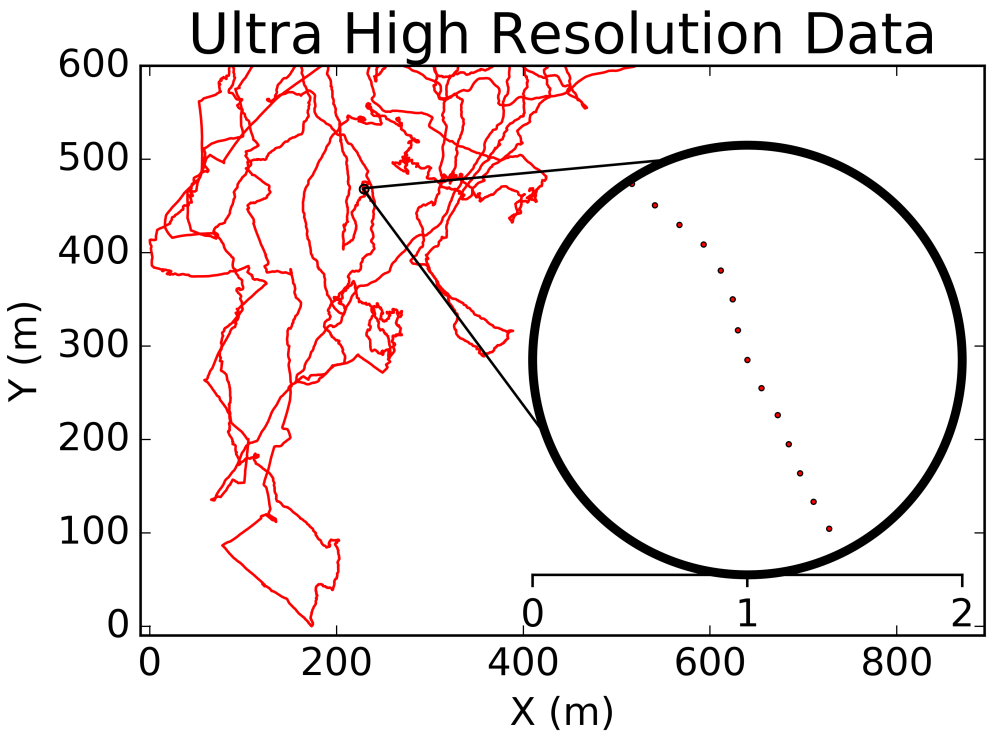

Figure 1.2: **Example of ultra high resolution data**
Example path segment of a goat at a resolution of 1Hz.

et al., 2008). Changes in an animal's internal state or environment can motivate their movements from one location to another, which creates space-use patterns. Therefore, describing an animal's space-use patterns is a useful initial step prior to inferring the motivations underlying their movements.

One of the most common methods for describing an animal's space-use pattern is to calculate its *utilisation distribution* (UD). The UD of an animal is the relative frequency distribution of the animal's location (Van Winkle, 1975), which gives a general description of how an animal uses the space available to them. One of the most common methods used to calculate the UD is Kernel Density Estimation (KDE; Silverman (1986); Worton (1989); Benhamou and Cornélis (2010)). In KDE circular bivariate probability density functions (PDFs) are placed at each recording to estimate an animal's space use between consecutive recordings. The UD is calculated by summing and then normalising these PDFs. Therefore KDE is unsuitable for data gathered at $\sim$1Hz or more frequently since it does not make sense to use PDFs to estimate an animal's space-use when an animal's location is known is pseudo-continuous time. Alternatively a researcher could first rarefy these high resolution data by subsampling at a constant rate then use KDE. However, this would mean removing potentially valuable informa-

tion about the animal's location between subsampled locations and replacing it with estimations of the animal's location.

The UD gives a general description of an animal's space use patterns, but depending on the research question (e.g. identifying resource patches for use in Optimal Foraging studies Bond (1980)) it may be more useful to identify specific areas which are profitable for the animal (e.g. food patches). I term these areas as *sites of interest*. These sites of interest may not be known *a priori* due to limited environmental information (Dodge et al., 2013), which means that techniques to infer their locations from the animal's movements are needed. For example, foraging patches correspond to time periods when the animal reduces its speed and increases the frequency of turns resulting in a very tortuous path (Benhamou and Bovet, 1989).

The *residence time* metric (Barraquand and Benhamou, 2008) is used to determine the amount of time spent in the vicinity of an area, so it can be used to identify areas of high use (i.e. sites of interest). Residence times are calculated by centring circles at every location on the animal's path and estimating the amount of time spent in these circles. However, when relocation data is sparse, the animal's location is unknown between recordings, which means that there is a large margin of error when calculating the amount of time an animal spends inside the circles from low resolution data. Therefore, when Barraquand and Benhamou (2008) estimated the residence time of a circle, they included any time periods between visits to the circle that are less than a given threshold (see Fig. 1 in Barraquand and Benhamou (2008)). With data gathered at high frequencies (e.g. $\geq$1Hz) it is unnecessary to include these time periods between visits to the circle when calculating the amount of time spent inside the circles.

However, given the magnitude of such high frequency datasets, it is necessary to adapt the residence-time algorithm to ensure it can be executed in a reasonable time-frame. Therefore, in Chapter 2 I introduce a new algorithm designed to calculate residence time (which I call *usage time*) from high resolution data. From the usage times of the circles I identify sites of interest as a distinct subset of the circles, each of which has a relatively higher usage time than all of the circles not in the subset. The algorithm used to identify these sites is termed the Sites of Interest Algorithm.

§ 1.3 Step Selection Analysis

Step Selection Analysis (SSA; Fortin et al. (2005); Rhodes et al. (2005)) is a principal tool for understanding the motivations underlying an animal's decisions to move. In simple terms, SSA works by comparing the steps that an animal made with steps that they could have taken, referred to as *case* and *control steps*, respectively. By making this comparison the motivations for choosing one step over another are inferred. This was later generalised and termed integrated-SSA (iSSA, Avgar et al. (2016)), which enables a user simultaneously to test hypotheses about the factors affecting where an animal chooses to move and how they move there. The iSSA method has been used to assess the effects that roads have on a wide variety of animals' movements, including elephants (Wadey et al., 2018), elk (Prokopenko et al., 2017*a*,*b*) and wolverines (Scrafford et al., 2018). It has also been used to understand the effects of other linear features, such as

pipelines and seismic lines (DeMars and Boutin, 2018). Also, in Ladle et al. (2019) they use iSSA to examine the affects that human recreational activities have on grizzly bear movements. In Raynor et al. (2017) they investigate how bison respond to variations in forage quality.

Inference from iSSA is often limited by the data, since case steps are usually defined as the movements between consecutive relocations in data gathered at low frequencies (Fig. 1.1). Much existing data (e.g. from VHF, Argos and GPS) are obtained at evenly spaced points in time. However, these points in time are very unlikely to correspond to times when the animal makes actual movement decisions. Instead there is a need for methods that analyse animal movements on the same scale as the decisions they make, rather than the scale at which the data are collected (Bastille-Rousseau et al., 2018).

When data are gathered at sufficiently high frequencies, however, we have sufficient information to infer the precise points (in space and time) at which an animal turns (Potts et al., 2018). Since turning is energetically costly, an animal would not change direction unless it were beneficial for them (Wilson et al., 2013*b*). Therefore turning points can be considered as places where the animal has made a distinct movement decision. As such, steps between successive turning points are the natural objects on which to perform iSSA.

However, it is not possible to use iSSA directly on steps from turning point to turning point, since iSSA is designed for use on data where the time interval between successive points (a.k.a. *step time*) is the same across the whole trajectory. In contrast, the time intervals between successive turning points are highly unlikely to be the same for all steps (Potts et al., 2018). I therefore adapt iSSA in Chapter 4 to cope with non-constant step-times, calling the resulting method time-varying iSSA (tiSSA).

§ 1.4 Thesis Outline

Chapter 2 explains in detail the differences between the residence and usage time metrics. I also compare the runtime and accuracy of calculating these metrics on high frequency data. I provide details of how the Sites of Interest Algorithm uses the usage times to identify specific areas that are extensively used by an animal.

In Chapter 3 I apply the Sites of Interest Algorithm to data from free-ranging cattle and goats in the French Alps. I use these data as examples to demonstrate some post-processing methods and extensions to the Sites of Interest Algorithm that can be used to extract interesting results from the whole output of the algorithm. In particular, I demonstrate an extension to the algorithm enabling it to be applied to group data as well as individual trajectories. I also describe several post-processing methods used to identify sites from the usage times, which differ from the method described in Chapter 2.

Chapter 4 adapts the iSSA method for use with high frequency data. I use data on goat movements (also used in Chapter 3) to show the advantages of this technique, over using iSSA with regularly sampled location data. Chapter 5 contains some discussion and concluding remarks.

Chapter 2

An Algorithm to Infer Sites of Interest

In this chapter I describe a method termed the Sites of Interest Algorithm to identify locations where an animal spends a disproportionately long time. This method is specifically designed for use on ultra-high-resolution data ($\geq$1 Hz) and therefore enables these long and complex paths to be broken down into segments of intra and inter site movements.

The Sites of Interest Algorithm was published in Munden et al. (2019) and adapted from a method in Barraquand and Benhamou (2008) and Benhamou and Riotte-Lambert (2012). One aim of Benhamou and Riotte-Lambert (2012)'s algorithm is to understand an animal's space use patterns by calculating the *residence time*, which is the amount of time an animal spends in the vicinity of a particular area. As well as the residence time, they also calculate the number of times an animal visits an area and the average visit duration. These different metrics (residence time, number of visits and average visit duration) are used to identify areas that may be of interest to the animal for a variety of reasons.

However, the method in Barraquand and Benhamou (2008) was designed with lower resolution data in mind (e.g. GPS data). Consequently it requires interpolation of the data, whereby equidistant points on the straight lines joining successive relocations are included in the trajectory. For the high resolution data we are interested in, however, such interpolation is not necessary because the distances between successive relocations are so short.

The algorithm I present here capitalises on the benefits of ultra-high-resolution data, which obviate the need to interpolate between successive recordings. However, a drawback of having an increased amount of data is that it greatly increases the computational time of the algorithm in Barraquand and Benhamou (2008). Here, I introduce two adaptations to speed up this algorithm. The first is to limit the number of areas for which the residence time is calculated. The second is to decrease the amount of computational time needed to calculate the residence times.

§ 2.1 Sites of Interest Algorithm

2.1.1 Usage Time

Suppose an animal's trajectory is given by a sequence of recorded locations $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, where $\mathbf{x}_i$ is the recorded location of the animal at time t_i. The key idea behind the Sites of Interest Algorithm is to slide a circle along the trajectory and then calculate the *usage time* within each circle. The radius of the circle is denoted by R (details of how to choose R are given in Section 2.6). The set of circle centres is a subset of $\mathbf{X}$ and denoted as $\left\{\mathbf{x}_1^{(c)}, \mathbf{x}_2^{(c)}, \ldots, \mathbf{x}_{N^{(c)}}^{(c)}\right\}$ and $t_i^{(c)}$ is the time when the location $\mathbf{x}_i^{(c)}$ was recorded.

One might naively centre the circle at every location along a trajectory (Barraquand and Benhamou, 2008), but this is very computationally expensive when recordings are gathered at a very high frequency (see Section 2.5.1). Instead, the Sites of Interest Algorithm centres the first circle at the start of the trajectory (i.e. $\mathbf{x}_1^{(c)} = \mathbf{x}_1$) then each other circle is centred at the first time point when the animal leaves the previous circle. In other words, $t_k^{(c)}$ is the first time point after $t_{k-1}^{(c)}$ at which the animal leaves the circle centred at $\mathbf{x}_{k-1}^{(c)}$. This reduces the number of circles and thus decreases the algorithm's runtime.

The usage time of a circle is calculated as the sum of *passage* durations, where a passage is a movement through the circle. A passage can be defined by its *entrance* and *exit* times, which are the times at which the animal enters and leaves the circle, i.e. when they cross the circle's boundary. Together I term these entrances and exits as *crossings*. The duration of a passage is the difference between its exit and entrance times. I also define two types of passage; *forward* and *backward* passages. Forward passages occur after the time point at which the circle is centred and backward passages occur before this point. For example if the usage time of the circle centred at $\mathbf{x}_k^{(c)}$ is being calculated then forward crossings occur after $t_k^{(c)}$ and backward crossings before $t_k^{(c)}$.

A crossing occurs if one point is inside the circle and the next is outside it. To check whether a point is inside or outside a circle I use the Euclidean metric, assuming that paths range over a sufficiently small area so that we can ignore the Earth's curvature. I denote the Euclidean metric distance as $d\left(\mathbf{x}_k^{(c)}, \mathbf{x}_j\right)$.

It would be natural to identify crossings by checking between all consecutive points in the data (Barraquand and Benhamou, 2008), but for high resolution data this is very computationally expensive (see Section 2.5.1). To decrease the runtime when identifying crossings I introduce a *crossing interval*, whereby the algorithm checks for crossings between t_i and t_{i+s} then between t_{i+s} and t_{i+2s} and so on, for some fixed integer $s \geq 1$. The most appropriate value of s will be dependent on the time requirements of the user, but I find that using $s = 10$ gives accurate results in a reasonable amount of time for the purposes of the data I examine (see Section 2.5). Crossings occur if one of the interval's end points is inside the circle and the other is outside the circle.

Entrances into and exits from a circle are taken to be the recorded location closest to the point where the animal crosses the circle's boundary.

In the following subsections I give rigorous definitions of the various concepts used in the algorithm.

Forward Entrance

Forward crossings occur between $t_k^{(c)}$ and t_N. A forward entrance occurs between t_i and t_{i+s} if

$$d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_i\Big) \geq R \text{ and } d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_{i+s}\Big) \leq R.$$

Next, the algorithm finds the crossing time between t_i and t_{i+s}. The exact crossing time occurs between t_j and t_{j+1} (where $j \in \{i, i+1, \ldots, i+s-1\}$) if

$$d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_j\Big) \geq R \text{ and } d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_{j+1}\Big) \leq R.$$

Then the crossing time, t^*, is defined to be

$$t^* = \begin{cases} t_j & \text{if } R - d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_j\Big) \leq d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_{j+1}\Big) - R \\ t_{j+1} & \text{if } R - d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_j\Big) > d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_{j+1}\Big) - R. \end{cases}$$

Forward Exit

A forward exit occurs between t_i and t_{i+s} if

$$d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_i\Big) \leq R \text{ and } d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_{i+s}\Big) \geq R.$$

Next, the algorithm finds the crossing time between t_i and t_{i+s}. The crossing time is chosen to be either t_j or t_{j+1} (where $j \in \{i, i+1, \ldots, i+s-1\}$) if

$$d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_j\Big) \leq R \text{ and } d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_{j+1}\Big) \geq R.$$

Then the crossing time, t^*, is defined to be

$$t^* = \begin{cases} t_j & \text{if } R - d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_j\Big) \leq d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_{j+1}\Big) - R \\ t_{j+1} & \text{if } R - d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_j\Big) > d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_{j+1}\Big) - R. \end{cases}$$

Backward Entrance

Backward crossings occur between $t_k^{(c)}$ and t_1. A backward entrance occurs between t_i and t_{i-s} if

$$d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_i\Big) \leq R \text{ and } d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_{i-s}\Big) \geq R.$$

Next, the algorithm finds the crossing time between t_i and t_{i-s}. The crossing time is chosen to be either t_j or t_{j+1} (where $j \in \{i-s+1, i-s+2, \ldots, i\}$) if

$$d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_j\Big) \leq R \text{ and } d\Big(\mathbf{x}_k^{(c)}, \mathbf{x}_{j-1}\Big) \geq R.$$

Then the crossing time, t^*, is defined to be

$$t^* = \begin{cases} t_j & \text{if } R - d\left(\mathbf{x}_k^{(c)}, \mathbf{x}_j\right) \leq d\left(\mathbf{x}_k^{(c)}, \mathbf{x}_{j-1}\right) - R \\ t_{j-1} & \text{if } R - d\left(\mathbf{x}_k^{(c)}, \mathbf{x}_j\right) > d\left(\mathbf{x}_k^{(c)}, \mathbf{x}_{j-1}\right) - R. \end{cases}$$

Backward Exit

A backward exit occurs between t_i and t_{i-s} if

$$d\left(\mathbf{x}_k^{(c)}, \mathbf{x}_i\right) \geq R \text{ and } d\left(\mathbf{x}_k^{(c)}, \mathbf{x}_{i-s}\right) \leq R.$$

Next, the algorithm finds the crossing time between t_i and t_{i-s}. The crossing time is chosen to be either t_j or t_{j+1} (where $j \in \{i-s, i-s+1, \ldots, i-1\}$) if

$$d\left(\mathbf{x}_k^{(c)}, \mathbf{x}_j\right) \geq R \text{ and } d\left(\mathbf{x}_k^{(c)}, \mathbf{x}_{j-1}\right) \leq R.$$

Then the crossing time, t^*, is defined to be

$$t^* = \begin{cases} t_j & \text{if } R - d\left(\mathbf{x}_k^{(c)}, \mathbf{x}_j\right) \leq d\left(\mathbf{x}_k^{(c)}, \mathbf{x}_{j-1}\right) - R \\ t_{j-1} & \text{if } R - d\left(\mathbf{x}_k^{(c)}, \mathbf{x}_j\right) > d\left(\mathbf{x}_k^{(c)}, \mathbf{x}_{j-1}\right) - R. \end{cases}$$

I define the sets of crossing times for the circle centred at $t_k^{(c)}$ for forward and backward passages as

$$F\left(t_k^{(c)}\right) = \left\{F_1\left(t_k^{(c)}\right), F_2\left(t_k^{(c)}\right), \ldots, F_{2\varphi_k+1}\left(t_k^{(c)}\right)\right\} \text{ where } F_i\left(t_k^{(c)}\right) < F_{i+1}\left(t_k^{(c)}\right)$$
$$\text{and } B\left(t_k^{(c)}\right) = \left\{B_1\left(t_k^{(c)}\right), B_2\left(t_k^{(c)}\right), \ldots, B_{2\beta_k+1}\left(t_k^{(c)}\right)\right\} \text{ where } B_i\left(t_k^{(c)}\right) > B_{i+1}\left(t_k^{(c)}\right)$$

where φ_k and β_k are the numbers of forward and backward passages, respectively. So $B_1\left(t_k^{(c)}\right)$ and $F_1\left(t_k^{(c)}\right)$ are the entrance and exit times of the passage that the circle is centred on. The next circle is centred at the first forward exit, i.e. $t_{k+1}^{(c)} = F_1\left(t_k^{(c)}\right)$. Then $F_{2i}\left(t_k^{(c)}\right)$ and $F_{2i+1}\left(t_k^{(c)}\right)$ are the entrance and exit times for the i^{th} forward passage, respectively. Also $B_{2i}\left(t_k^{(c)}\right)$ and $B_{2i+1}\left(t_k^{(c)}\right)$ are the exit and entrance times for the i^{th} backward passage, respectively.

Illustrative Example

I use the trajectory segment from Figure 2.1 as an illustrative toy example of how the algorithm identifies crossings. Suppose that the circle centred on the time point t_{11} is the k^{th} circle along this trajectory, then $t_k^{(c)} = t_{11}$. I use the notation $[t_i, t_j]$ to denote the crossing interval between t_i and t_j. I.e. a crossing occurs in the interval $[t_i, t_j]$ if precisely one of t_i or t_j is inside the circle and the other is outside the circle. For this example I set $s = 5$.

In identifying crossings for the circle centred at t_{11}, the algorithm checks for forward crossings in the crossing intervals $[t_{11}, t_{16}]$, $[t_{16}, t_{21}]$, $[t_{21}, t_{26}]$, $[t_{26}, t_{31}]$, $[t_{31}, t_{36}]$ and $[t_{36}, t_{41}]$. Backward crossings are checked for in the crossing intervals $[t_{11}, t_6]$ and $[t_6, t_1]$.

Forward crossings occur within the intervals $[t_{11}, t_{16}], [t_{26}, t_{31}]$ and $[t_{31}, t_{36}]$. Exact crossings occur within the intervals $[t_{13}, t_{14}], [t_{30}, t_{31}]$ and $[t_{35}, t_{36}]$. The first forward crossing is found to be t_{14}, since it is closer to the circle's boundary than t_{13}. Since t_{14} is the first forward exit, it is the centre of the next circle, i.e. $t_{k+1}^{(c)} = t_{14}$. The other forward crossing points are found to be t_{31} and t_{36}.

Similarly the algorithm checks for backward crossings occurring in the intervals $[t_{11}, t_6]$ and $[t_6, t_1]$. The algorithm finds that there is a crossing in the interval $[t_{11}, t_6]$, but not $[t_6, t_1]$, with the exact crossing occurring in the interval $[t_7, t_6]$. Then t_7 is found to be the crossing point since it is closer to the circle's boundary than t_6.

Usage Time Calculation

Recall that the usage time is calculated as the sum of all passage durations. So the usage time of the circle centred at $t_k^{(c)}$ is defined as

$$\psi_k = F_1\left(t_k^{(c)}\right) - B_1\left(t_k^{(c)}\right) + \sum_{i=1}^{\varphi_k}\left(F_{2i+1}\left(t_k^{(c)}\right) - F_{2i}\left(t_k^{(c)}\right)\right) + \sum_{i=1}^{\beta_k}\left(B_{2i}\left(t_k^{(c)}\right) - B_{2i+1}\left(t_k^{(c)}\right)\right). \tag{2.1}$$

Returning to the example in Figure 2.1, I write the sets of forward and backward crossing times as $F\left(t_k^{(c)}\right) = \{t_{14}, t_{31}, t_{36}\}$ and $B\left(t_k^{(c)}\right) = \{t_7\}$. Therefore the number of forward passages, φ_k, is one and the number of backward passages, β_k, is zero. Substituting these values into Equation 2.1 results in a usage time of $\psi_k = t_{14} - t_7 + t_{36} - t_{31}$.

2.1.2 Rarefication Procedure

It is possible for the set of circles centred at $\{\mathbf{x}_1^{(c)}, \ldots \mathbf{x}_{N^{(c)}}^{(c)}\}$ to overlap with one another, meaning that the same movements will contribute to the usage times of multiple circles. It does not make sense to compare the usage times of two overlapping circles, so it is necessary to remove these overlaps. Since sites are areas of high usage times, the rarefied set should contain the circles of highest usage time that do not overlap. I refer to the method of creating this rarefied set as the *rarefication procedure*.

I define the set of usage times as $\Psi = \left\{\psi_1, \psi_2, \ldots, \psi_{N^{(c)}}\right\}$, where each ψ_i represents a particular circle. This set is then ordered in decreasing order giving a new set, denoted $\Psi^{(o)} = \left\{\psi_1^{(o)}, \psi_2^{(o)}, \ldots, \psi_{N^{(c)}}^{(o)}\right\}$, such that $\psi_i^{(o)} \geq \psi_{i+1}^{(o)}$ for all $i \in \left\{1, 2, \ldots, N^{(c)} - 1\right\}$. The circle with usage time $\psi_i^{(o)}$ is centred on the point recorded at $t_i^{(o)}$. Next, a rarefied set of usage times, given as Ψ', is constructed using the following steps:

(Step 1) set $\Psi' = \left\{\psi_1^{(o)}\right\}$ and $i = 2$

(Step 2) if the circle centred at $t_i^{(o)}$ does not intersect with any of the circles so far represented by elements in Ψ' (i.e. the centre of this circle is a distance greater than

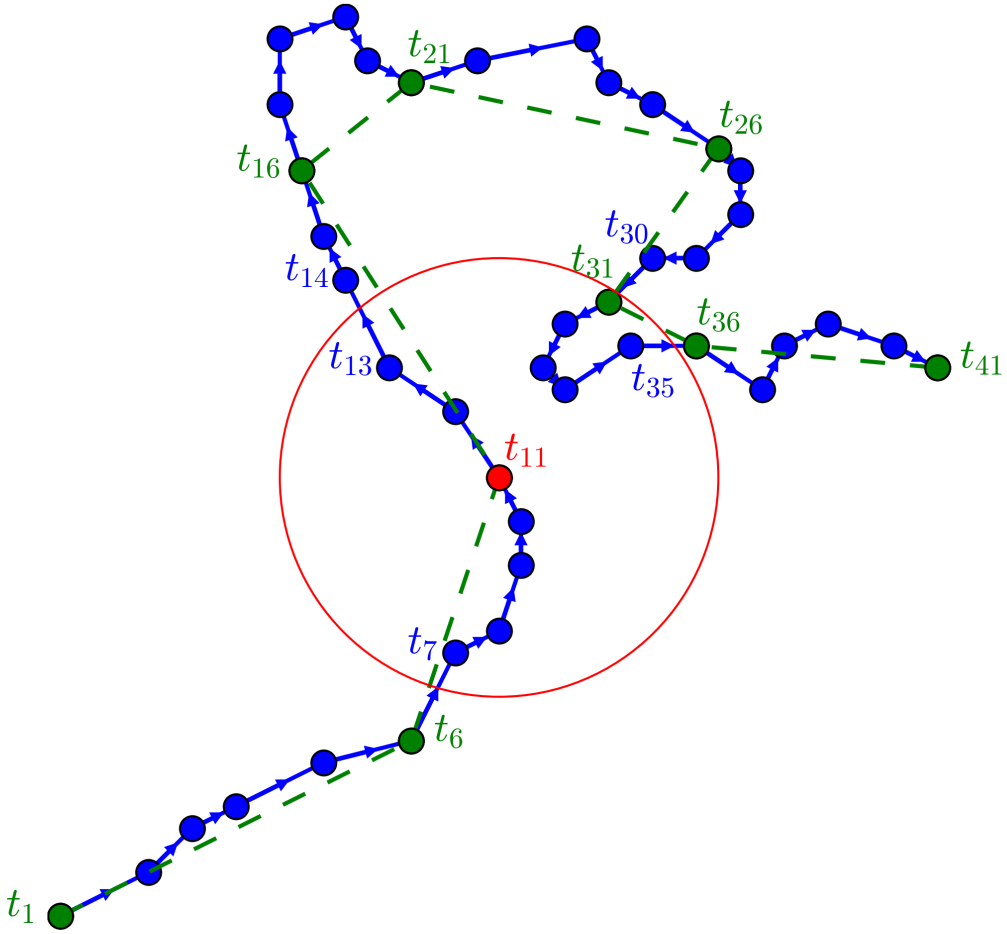

Figure 2.1: **Annotated example trajectory segment**
An annotated example trajectory segment (blue lines), where arrows show the direction of movement. Points at the ends of crossing intervals are given in green and labelled with the times of when they are recorded. Dashed green lines join the end points of the crossing intervals. Blue dots either side of a crossing are labelled with the time they are recorded. The red dot is the circle's centre and labelled with the time it is recorded, t_{11}. Forward and backward crossings occur at time points t_i where $i > 11$ and $i < 11$, respectively.

$2R$ from the centre of all the circles so far in Ψ') then include $\psi_i^{(o)}$ in the set Ψ'

(Step 3) if $i = N^{(c)}$ then end, else set $i = i + 1$ and return to Step 2

I denote the elements of this rarefied set, Ψ', as $\psi'_1, \psi'_2, \ldots, \psi'_{N'}$, where $\psi'_i > \psi'_{i+1}$ and N' is the number of circles left after the rarefication procedure.

Figure 2.2 gives an illustrated example of how the rarefication procedure works, where the path I use comes from an Ornstein-Uhlenbeck simulation (OU, Section 2.2). I index the usage time of a circle with the circle's lettered label in Figure 2.2 (e.g. the usage time of circle A is ψ_A). Therefore, following the steps, the rarefication procedure starts with $\Psi' = \{\psi_A\}$. Circle B overlaps with Circle A, so is not included in Ψ'. Circle C does not overlap with Circle A, so is included, meaning that $\Psi' = \{\psi_A, \psi_C\}$. Circle D overlaps with Circle C, so is not included in Ψ'. Circle E neither overlaps with Circles A or C, so is included. Thus $\Psi' = \{\psi_A, \psi_C, \psi_E\}$.

2.1.3 Identifying sites

Having rarefied the set of circles, the problem now is to identify which of the circles are sites of interest. Notice that after using the rarefication procedure there are two distinct groups of circles that arise; those with higher usage times and those with smaller usage times. This difference can be seen from comparing panels (a) and (c) in Figure 2.3. In panel (c) there is a clear drop after the fifth bar, but there is no similarly distinct drop in panel (a).

Identifying the sites can be seen as a partitioning problem, where the rarefied set of circles is partitioned into two subsets; one of sites and the other of non-sites. Circles in the non-sites subset typically correspond to movements when the animal is transitioning between sites, so have lower usage times. Since I define a site of interest to be an area with a disproportionately larger usage time than other areas, the partition between the two subsets should be somewhere in the ordered list of usage times, Ψ'. The algorithm still needs a method to determine the best partition.

For this, I use a method which I term the *percent drop*, which is the relative difference in usage times between the minimum of the subset of "sites" and the maximum of the subset of "non-sites". The algorithm finds the partition which maximises this measure, which I term the *maximum percent drop* and is defined as

$$MPD_R = \max\left(1 - \frac{\psi'_{i+1}}{\psi'_i}\middle| i \in \{1, 2, \ldots, N' - 1\}\right) \tag{2.2}$$

and is dependent on the radius value R. If $MPD_R = 1 - \frac{\psi'_{n_R+1}}{\psi'_{n_R}}$ the sites of interest are the circles corresponding to usage times, $\psi'_1, \ldots, \psi'_{n_R-1}$ and ψ_{n_R}. So n_R is the number of sites identified from using a radius of R.

Returning to the example in Figure 2.2 the algorithm calculates the percent drops as $100(1 - \psi_A/\psi_C)\%$ and $100(1 - \psi_C/\psi_E)\%$, where the former is the larger of the two. Therefore, the maximum percent drop occurs between ψ_A and ψ_C, so the only site is

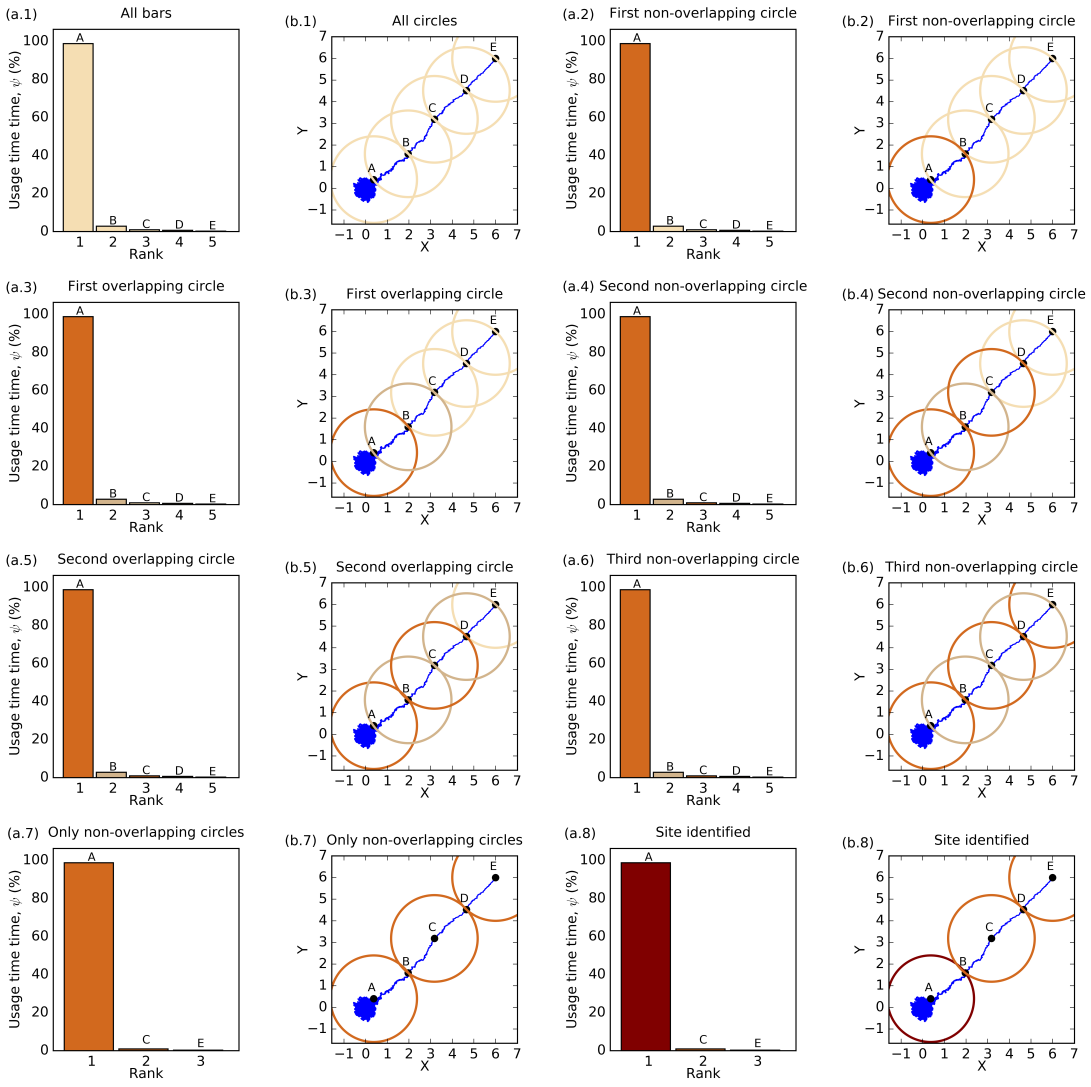

Figure 2.2: **How to identify sites of interest from the circle usage times** All of the (a.1-8) panels show the usage times (as percentages of the simulation duration) of the circles. All of the (b.1-8) panels show the positions of the circles. Panels (a.1) and (b.1) show the usage times and the locations of all circles, respectively. Panels (a.2) and (b.2) show the first circle (A) added to the set of non-overlapping circles, Ψ'. Panels (a.3) and (b.3) show the first circle (B) to be removed, because it overlaps with circle A. Panels (a.4) and (b.4) show the next non-overlapping circle (C). Panels (a.5) and (b.5) show the next circle which overlaps (D), so is removed. Panels (a.6) and (b.6) show the last non-overlapping circle (E). Panels (a.7) and (b.7) show the circles that remain after the rarefication procedure. Panels (a.8) and (b.8) show which of the circles (A, C and E) are identified as sites of interest (A).

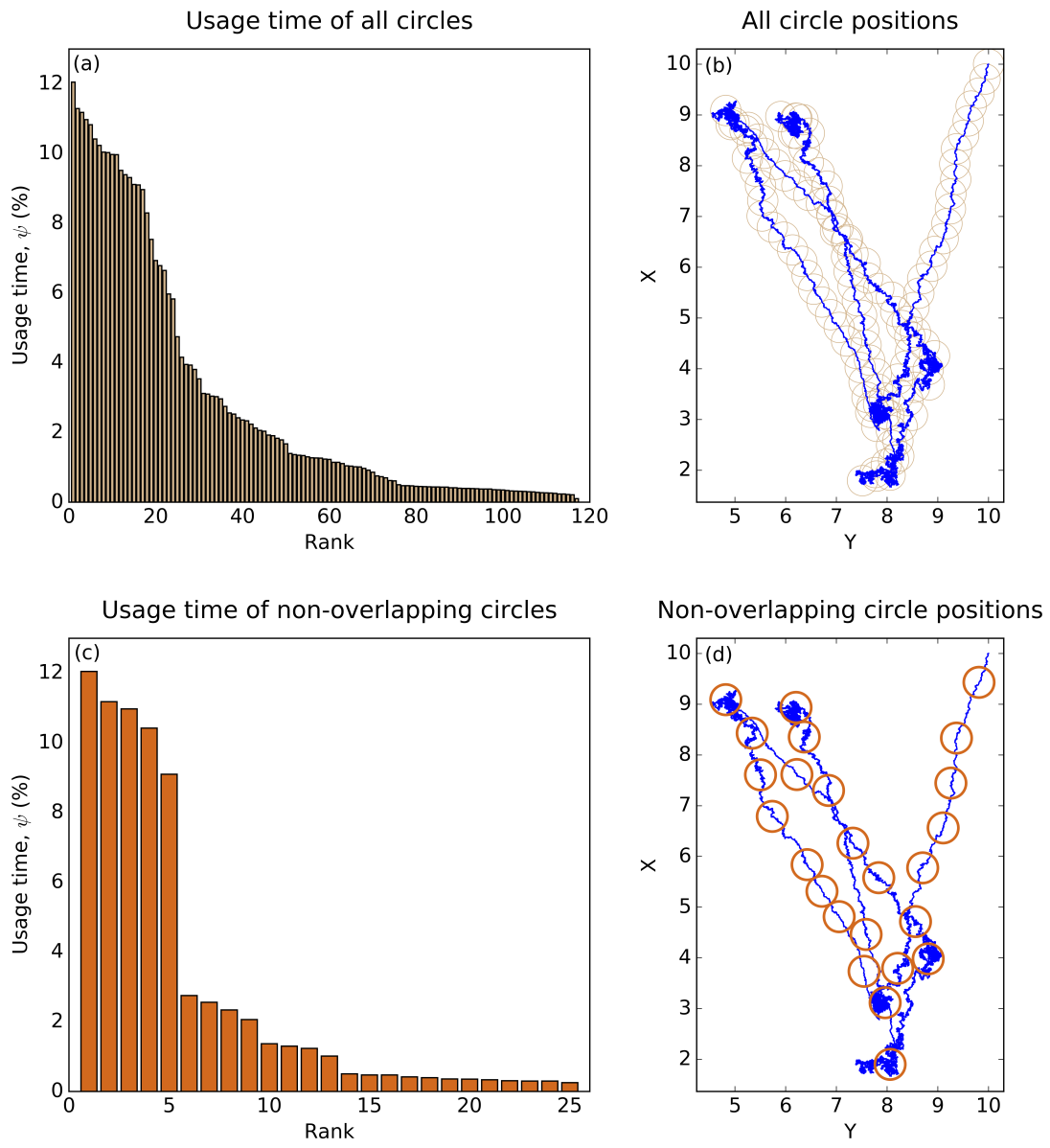

Figure 2.3: **Comparison from before and after rarefying the set of circles** Panels (a) and (b) show the results before rarefying the set of circles (Section 2.1.2), whereas panels (c) and (d) show the results after the rarefication procedure. Panels (a) and (c) show the sets of usage times (as percentages of the simulation duration) in decreasing order of usage time. Panels (b) and (d) show the locations of the sets of circles.

circle A and the non-sites are circles C and D. Note that this method of identifying sites from the maximum percent drop is similar to a scree plot (Cattell, 1966) from Principal Component Analysis (Jolliffe, 1986). The problem of subdividing the set of all circles into subsets of Sites and Non-sites is also similar to problems found in Cluster Analysis (Edwards and Cavalli-Sforza, 1965).

2.1.4 Colour Assignment

This process of calculating usage times, rarefying the set of circles and identifying sites can be repeated for different radius values, with the aim of using the two measures – the maximum percent drop and number of identified sites – to ascertain the most appropriate radius value. The radius values that should be tested will depend on the study species, but some suggestions of how to choose a test set of radius values are given in Section 2.6.

In an ideal situation the algorithm would be applied using a vast array of radius values, but it can be time consuming to choose the most appropriate radius value from inspecting all of the results by eye. Instead, I look for objective criteria to use as baseline rules for choosing the most appropriate radius value. I therefore introduce two criteria for choosing a radius value, which I term the *threshold* and *stability* criteria.

Both criteria find a *local maximum* radius value, which is one resulting in a higher maximum percent drop than other nearby radius values. In practice, these nearby radius values are the ones either side of the candidate value in the ordered set of radius values used. In other words, if $\{R_1, R_2, \dots\}$ is the ordered set of radius values used, where $R_i < R_{i+1}$, then R_i is a local maximum if $MPD_{R_{i-1}} < MPD_{R_i} > MPD_{R_{i+1}}$.

The algorithm searches for local maximums, since MPD_R is a measure of how distinct the sites are from non-sites in terms of usage times, and should therefore be maximised across different values of R. The algorithm does not use a global maximum, because that would occur when $MPD_R = 100\%$ which is only possible when there is one circle left after the rarefication procedure.

Threshold Criterion

The threshold criterion first creates a set of all radius values, R_i, for which the associated MPD_{R_i} is a local maximum and greater than a given threshold value, T_{MPD} (some advice on how to choose T_{MPD} is given in Section 2.3). Next, the threshold criterion chooses the minimum radius value from this set. In other words, if R_t is the radius value found from using the threshold criterion then $R_t = \min(R_i | MPD_{R_{i-1}} < MPD_{R_i} > MPD_{R_{i+1}}$ and $MPD_{R_i} > T_{MPD})$.

The maximum percent drop is an indication of how distinct the sites are from non-sites. Therefore, I include the maximum percent drop threshold, T_{MPD}, to ensure that the maximum percent drop is large enough to indicate two clearly distinct sets of circles.

I use the results presented in Figure 2.4 as examples of how the threshold criterion works, where the threshold, T_{MPD}, is 65% and the set of radii is $\{0.2, 0.3, \dots, 3.8\}$. For example from the results in panel (a) the threshold criterion chooses a radius value

of 0.8, since the maximum percent drop from using $R = 0.8$ is greater than 65% and higher than using either $R = 0.7$ or $R = 0.9$. For the results in panel (b) $MPD_1 > 65\%$ and $MPD_{0.9} < MPD_1 > MPD_{1.1}$, so $R_t = 1$. For the results in (c) $MPD_{0.3} > 65\%$ and $MPD_{0.2} < MPD_{0.3} > MPD_{0.4}$, so $R_t = 0.3$.

Stability Criterion

Firstly, the stability criterion finds the set of all radius values, R_i, which identify the same number of sites as the radius values either side of it (i.e. R_{i-1} and R_{i+1}) and for which the associated maximum percent drop, MPD_{R_i}, is a local maximum. Then the stability criterion finds the minimum of this set. More precisely the radius value found from using the stability criterion is defined to be $R_s = \min(R_i | n_{R_{i-1}} = n_{R_i} = n_{R_{i+1}}$ and $MPD_{R_{i-1}} < MPD_{R_i} > MPD_{R_{i+1}})$.

Using the results presented in Figure 2.4 I show how the stability criterion works, where the set of radii is $\{0.2, 0.3, \ldots, 3.8\}$. For the results shown in panel (a) using the stability criterion returns a radius value of 0.8, since 0.8 is a local maximum and using radius values of 0.7, 0.8 and 0.9 result in three sites being identified. For the results in panel (b) $MPD_{0.5} < MPD_{0.6} > MPD_{0.7}$ and $n_{R_{0.5}} = n_{R_{0.6}} = n_{R_{0.7}}$, so $R_s = 0.6$. For the results in panel (c) $MPD_{0.7} < MPD_{0.8} > MPD_{0.9}$ and $n_{R_{0.7}} = n_{R_{0.8}} = n_{R_{0.9}}$, so $R_s = 0.8$.

Note that both criteria (Threshold and Stability) find the minimum radius value subject to other conditions since as the radius value increases, the number of non-overlapping circles left after the rarefication procedure decreases to the point of there only being one left.

Combining the criteria together

The *colour assignment* assigns to a trajectory a traffic-light colour, which gives an indication of how much confidence the algorithm has in the output. The colour assignment works as a simple goodness-of-fit test, which combines the two criteria. Neither criteria works perfectly every time, so I use the colour assignment as a two check procedure to filter out situations when one or both criteria return incorrect results.

The options of colours are Red, Amber and Green, which correspond to having weak, intermediate and strong confidence, respectively. A trajectory is assigned to the Green category if both criteria identify the same number of sites and radius value. Amber is assigned if only the numbers of identified sites are the same for the two criteria. Red is assigned if neither the numbers of sites nor the radius values are the same for the two criteria. How the colour assignment works is summarised in Table 2.1.

Red	Amber	Green
$n_{R_t} \neq n_{R_s}$	$n_{R_t} = n_{R_s}$	$n_{R_t} = n_{R_s}$
$R_t \neq R_s$	$R_t \neq R_s$	$R_t = R_s$

Table 2.1: **Conditions for assigning the different colours.**
The number of sites found by using a radius value of R is n_R. Also R_t and R_s are the radii found from using the threshold and stability criteria, respectively.

For example, in Figure 2.4 (a) both the stability and threshold criteria return the same radius value, thus this trajectory is assigned to the Green category. In panel (b) the radii from the two criteria are different, but the number of sites identified are the same, therefore this trajectory is given the Amber assignment. In panel (c) the radii identified are different and so are the numbers of identified sites, so this trajectory is assigned to the Red category.

The colour assignment works by comparing the results from each criterion since it is possible that one criterion identifies the correct number of sites and the other does not, as can be seen in Figure 2.4 (c), where the stability criterion identifies the correct number of sites whereas the threshold criterion does not. I compare the number of sites identified since even when the number identified is not the same as the number of "true" sites, the correct locations are still found. In the cases where the number of sites is overestimated often the identified sites will be clustered around the true sites (Fig. 2.5b). Similarly if the number of sites is underestimated, the true sites are often contained within the boundaries of the identified sites (Fig. 2.5d).

A flow chart demonstrating the process for implementing the algorithm is given in Figure 2.6. To make the technique as accessible as possible I created an `R` package of the method, called `SitesInterest`, which is freely available on `CRAN` (`https://cran.r-project.org/web/packages/SitesInterest/index.html`). A starting guide to using the package is included in Appendix C.

2.1.5 Possible extensions

A user of this algorithm may be interested in applying it to data for more than one individual. Applying the algorithm to a group of individuals identifies shared areas of interest across the group. For applying the algorithm to a group of individuals there is a simple extension to the method, whereby circles are placed all along each individual path in the same way as before (Section 2.1.1), then the algorithm finds the usage time of each circle. These usage times may be contributed to by any of the individuals, not just the individual on whose path the circle is centred on. The process of rarefying circles (Section 2.1.2), identifying sites (Section 2.1.3) and the colour assignment (Section 2.1.4) are all exactly the same when applied to group data as when to an individual. See Section 3.4 for a comparison between applying the algorithm to a single cow and a group of cows.

This same extension for group data can also be applied to discontinuous data, where each segment is treated as an individual trajectory. This may be useful if the tag recording the animal's movements fails temporarily, resulting in missing data. Also for some data it may be appropriate to exclude certain segments of the trajectory. For example, some behaviours can cause a particular site to have an extremely high usage time, which leads to no other sites being identified. Therefore, it may be appropriate to remove segments of the trajectory corresponding to that particular behaviour. For example, the sleeping site of several cattle trajectories obscures other potential sites from being identified (Section 3.5).

It may be of interest to segment a long trajectory into separate movements for each

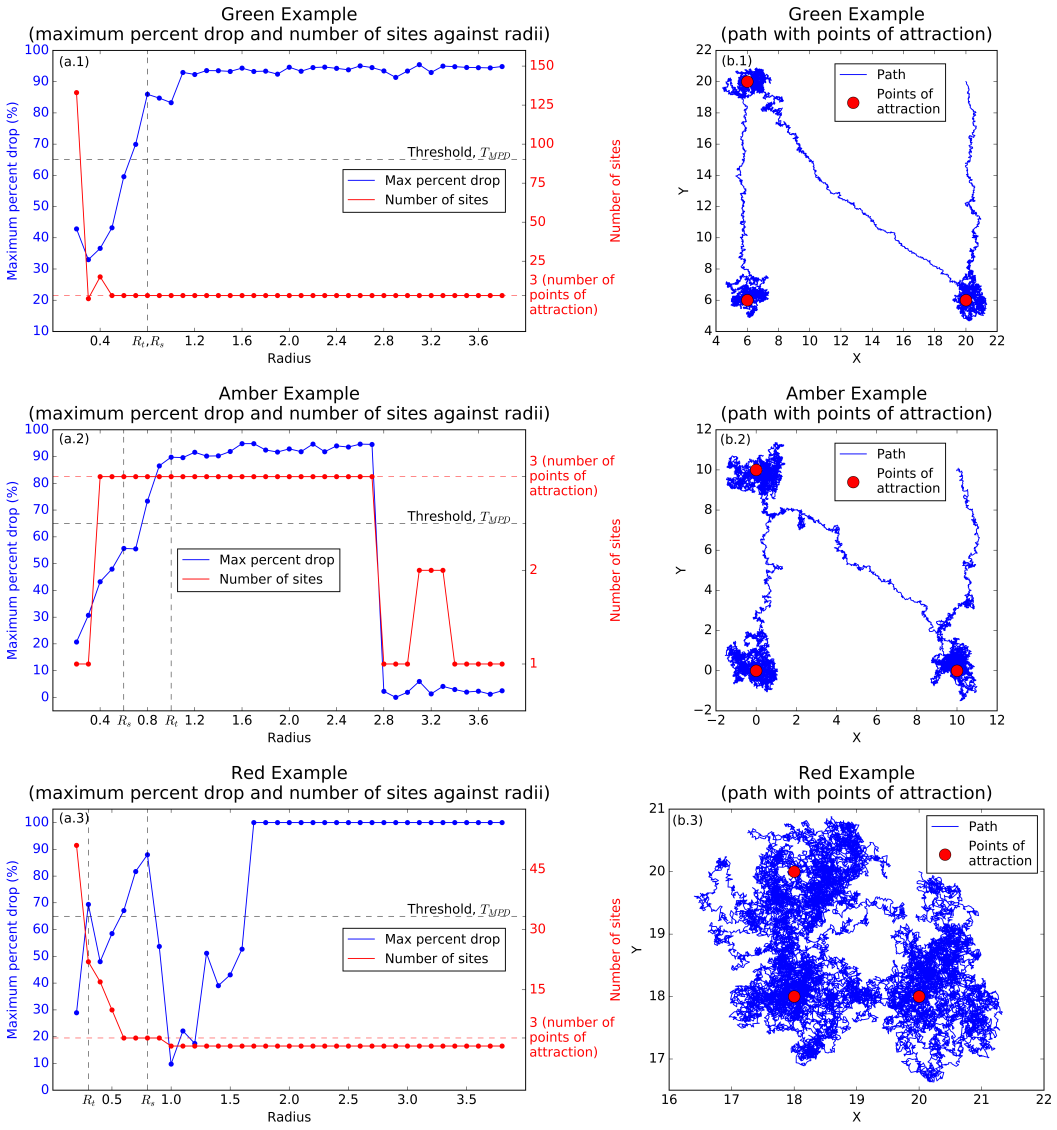

Figure 2.4: **Examples of how the criteria and colour assignment work** Results (maximum percent drops in blue and number of identified sites in red) of applying the algorithm to three example trajectories for various radius values in panels (a.1-3). The trajectories in panels (b.1-3) are assigned to the Green, Amber and Red categories, respectively. In panels (a.1-3) the maximum percent drop threshold value, T_{MPD}, and the actual number of sites are also indicated with horizontal dashed lines. Radii identified from the threshold and stability criteria are indicated on the x-axis by R_t and R_s, respectively. The threshold criterion finds the minimum radius value which results in a maximum percent drop greater than the threshold, T_{MPD} and a local maximum. The stability criterion finds the minimum radius value such that the number of identified sites is same for nearby radius values and the maximum percent drop is a local maximum.

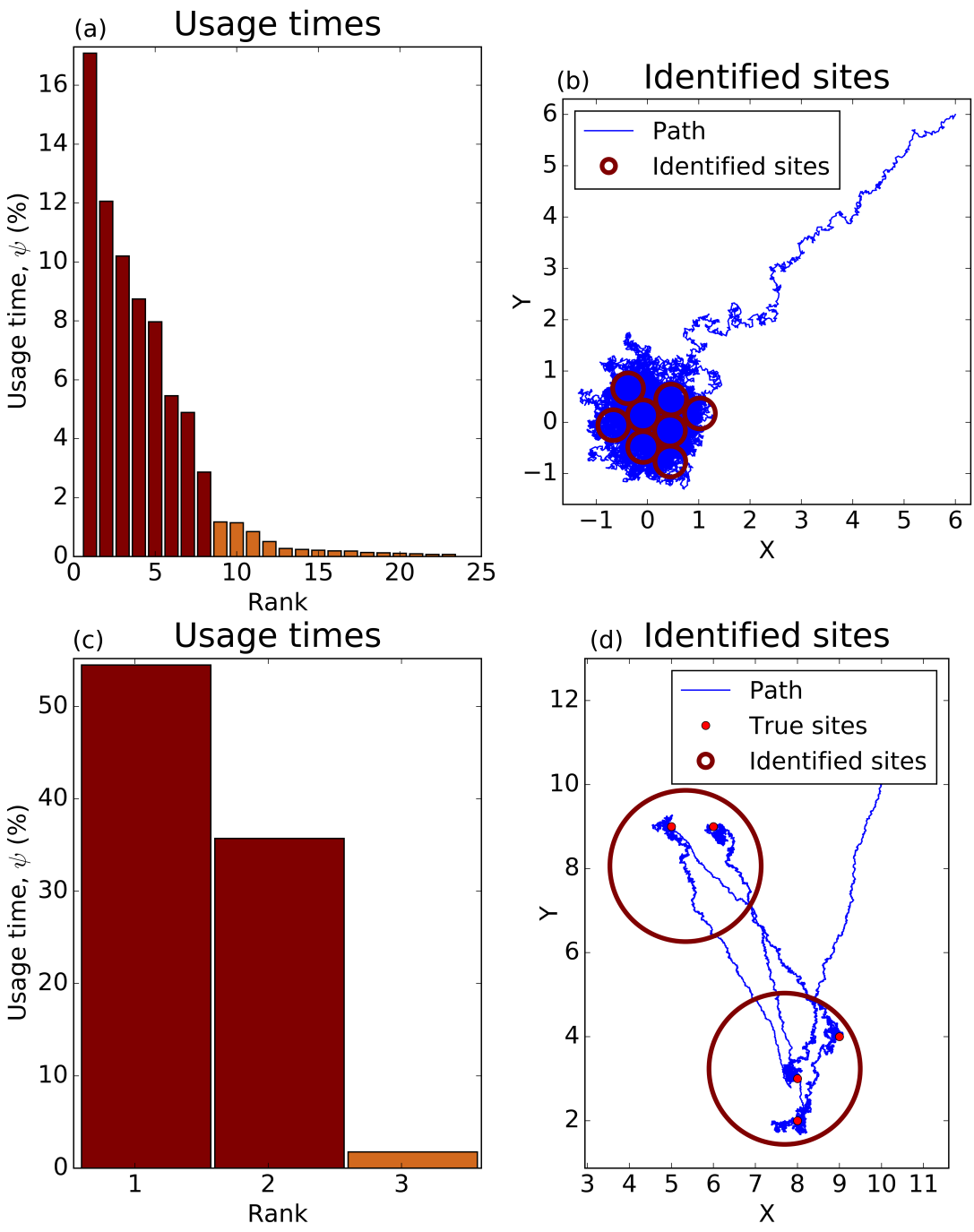

Figure 2.5: **Examples of over and under estimating the number of sites** Panels (a) and (b) correspond to a scenario where the number of sites is overestimated because the radius value is too small. Panels (c) and (d) correspond to the situation of the number of sites being underestimated because too large a radius value is used. Panels (a) and (c) show the usage times as percentages of the simulation's duration. Panels (b) and (d) show the locations of identified sites on the paths and the red points represent the "true" sites. Both of these paths are examples of when a different radius value would be more appropriate.

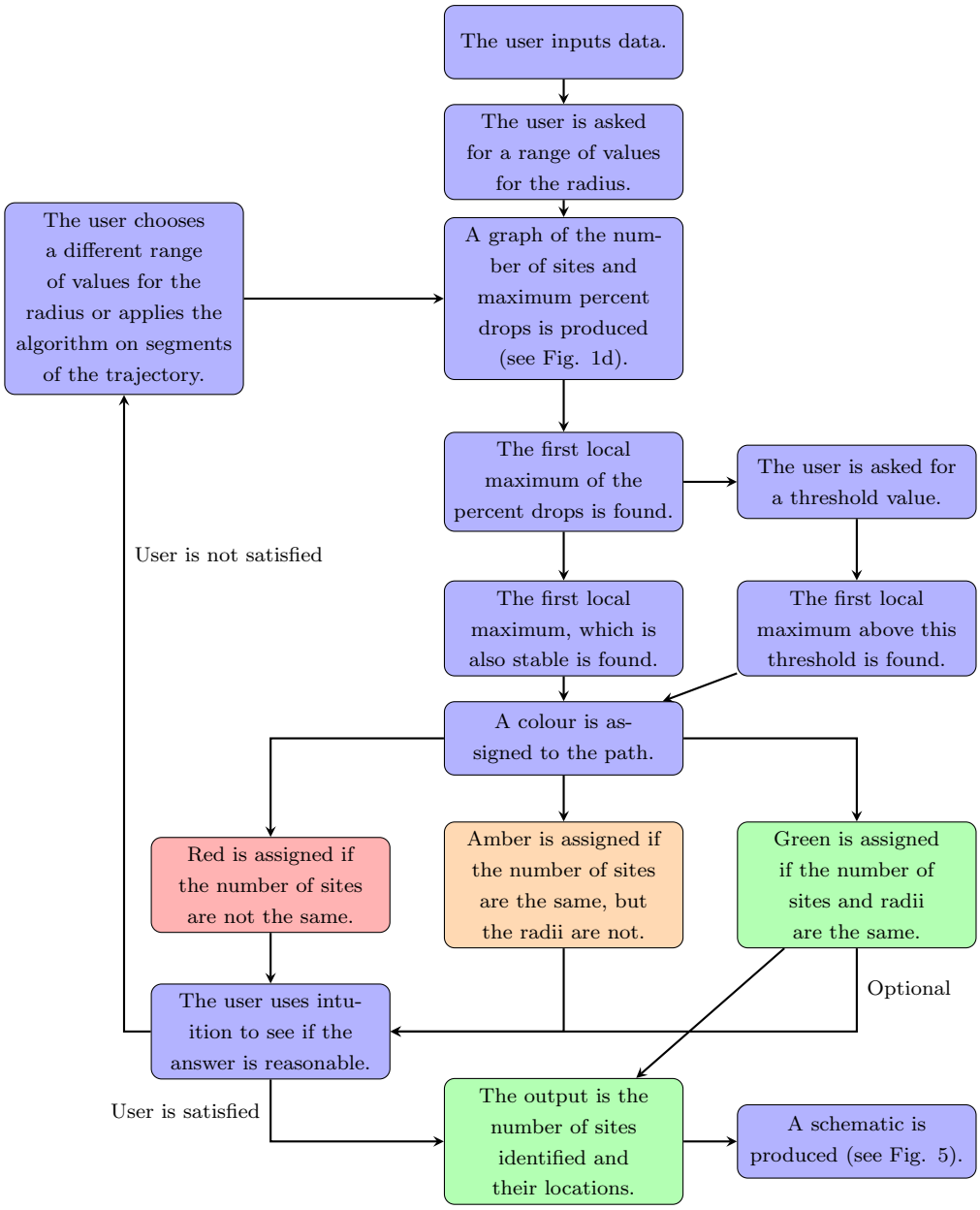

Figure 2.6: **Sites of Interest Algorithm implementation flow chart** A flow chart showing how to implement the Sites of Interest Algorithm.

day. This highlights any day to day variations, whereas applying the algorithm to the full trajectory reveals the commonly used locations from day to day. Daily variations could arise from changing environmental factors. For example, resources may be depleted and therefore not used until the resource is replenished.

§ 2.2 Ornstein-Uhlenbeck Simulations

To test the efficacy of the Sites of Interest Algorithm I apply the method to a selection of switching Ornstein-Ulhenbeck (OU) simulations. An animal following an OU process (Uhlenbeck and Ornstein, 1930) will move with some amount of (Gaussian) randomness toward a particular point, termed the *point of attraction*. For a switching OU process (Blackwell, 1997; Taylor and Karlin, 2014) the location of the point of attraction changes over time, but at each point in time there is only ever one point of attraction.

Using switching OU processes allows me to define the "true" sites of interest as the points of attraction. The test is to see if the sites of interest identified from the algorithm encompass the points of attraction. Across all simulations I vary the number of points of attraction, their locations, the strength of attraction and amount of randomness. I limit the number of points of attraction to a maximum of 10 and their locations are chosen from a 10 by 10 grid of possible locations. Details of how the simulations vary are given in Tables 5.1 - 5.4 and can be seen visually in Figure 5.1.

As an example of how to use the method, I present the results from applying the algorithm to one of the OU paths, shown in Figure 2.7. In panel (b) the algorithm not only identifies the correct number of sites, but also identifies their locations. When identifying the sites of interest, the algorithm is searching for a clear drop in usage times, like the one in panel (c). If this drop is not clear then it may be the case that a different radius value would be more appropriate. The results from using various radius values are presented in panel (d). Here, there is no clear relationship between the radius values and the numbers of sites identified or the maximum percent drops.

I also note that I found that sometimes applying the algorithm to the OU simulations resulted in every circle being identified as a site except one, which is the first circle on the trajectory. In this situation the algorithm can be seen as identifying an area of particularly low usage time, rather than identifying the high usage times. Naturally I would expect the first circle to have a particularly low usage time, since the trajectory is at least R in length, whereas the length of the trajectory in every other circle is at least $2R$. We see this since the trajectory starts at the centre of the first circle and then moves to the boundary, whereas for every other circle the trajectory moves from the boundary to the centre and then to the boundary again. as such whenever this case occurs the algorithm can be seen as either removing that circle and then calculating the maximum percent drop from the new set or as calculating the second highest percent drop.

I measure the efficacy of the algorithm for each OU path by the number of identified sites of interest compared to the number of points of attraction. I use the number of identified sites to measure efficacy since the locations of the points of attraction are almost always correctly identified, but sometimes the number of sites is under or

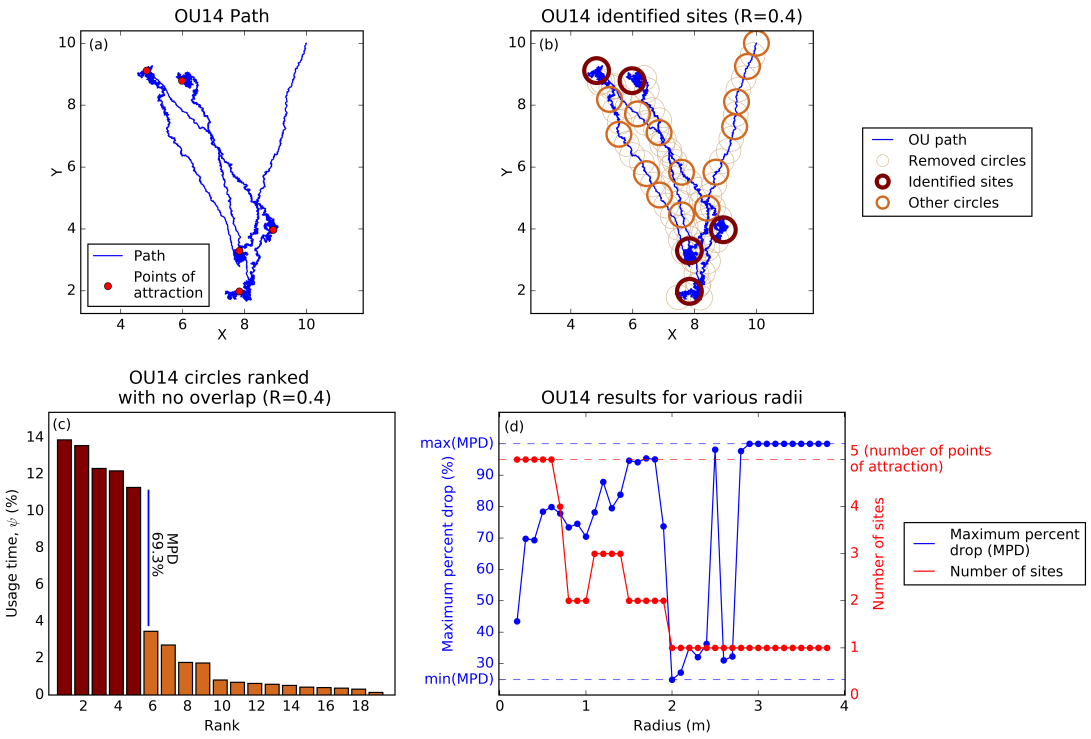

Figure 2.7: **Ornstein-Uhlenbeck simulation example**
Results from applying the Sites of Interest Algorithm to an Ornstein-Uhlenbeck example path (a). Panel (b) shows which circles were removed (yellow), identified as sites (brown) and the non-sites (orange). Panel (c) shows the usage times of all non-overlapping circles, with those associated to the sites in brown and the non-sites in orange. The maximum percent drop (MPD) is also highlighted between the fifth and sixth bars. Panel (d) shows the results from applying the algorithm with different radius values, where the results are the maximum percent drops (blue) and number of identified sites (red).

overestimated (Fig. 2.5).

Across all OU simulations, the colour assignment (Section 2.1.4) works well. For 72% of the 110 OU paths the number of sites identified is correct (Fig. 2.8). There is only one false positive result (assigned to the Green category but returned the incorrect number of sites), which means that for 99% of paths assigned to the Green category the number of sites identified is correct. Only two paths assigned to the Amber category identified the incorrect number of sites. The colours are assigned using a threshold value of 65%, which was chosen to be greater than the lower limit found in Section 2.3 and also to minimise the number of paths assigned to the Green category with the incorrect number of sites identified. The full results are given in Tables 5.5-5.8.

The colour assignment works as a warning system of when further verification is

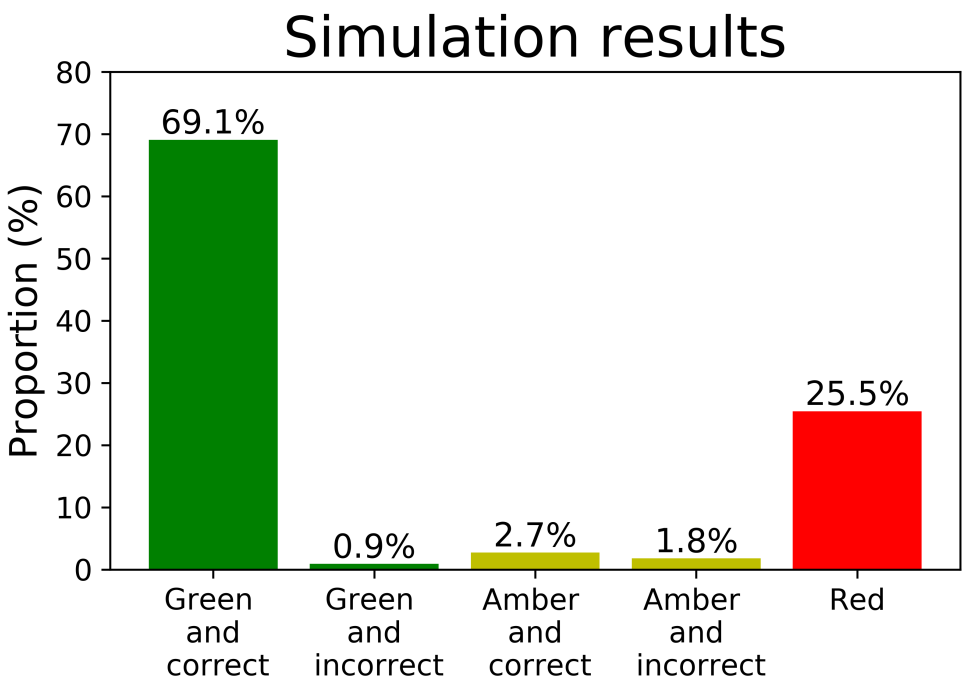

Figure 2.8: **Ornstein-Uhlenbeck simulation colour assignment results** Results from the colour assignment part of the algorithm (see Section 2.1.4) as applied to the Ornstein-Uhlenbeck simulations, where the correct and incorrect categories refer to the number sites identified compared to the number of points of attraction.

recommended. For example inspecting the results of some of the paths assigned to the Red category reveals that there are instances where one of the two criteria identifies the correct number of sites and the other does not (see Tables 5.5-5.8). Since Red is assigned when different numbers of sites are identified from using the two criteria this does not necessarily mean that they are both incorrect. Also the paths assigned to the Green category vary in terms of the number of points of attraction, distances between the points and strength of attraction (Tables 5.5 - 5.8 and Figure 5.1). This shows the versatility of the algorithm not just to assign paths of one particular type to the Green category.

§ 2.3 Choosing the threshold value, T_{MPD}

Recall that the threshold value, T_{MPD}, is used as part of the criterion to identify a radius value, R, which results in a significant maximum percent drop, MPD_R. Here, I explain how Brownian motion (Brown, 1828) simulations can be used to determine a lower limit for T_{MPD}. Brownian motion has no points of attraction so one would expect the usage times histogram to have no steep drops between consecutive bars, which means that the maximum percent drop is small. Hence there is less of a clear

distinction between the circles identified as sites and all other circles. In Figure 2.9 the maximum percent drop is not as visually striking in the Brownian motion simulation (panel a) as it is in the OU simulation (panel c).

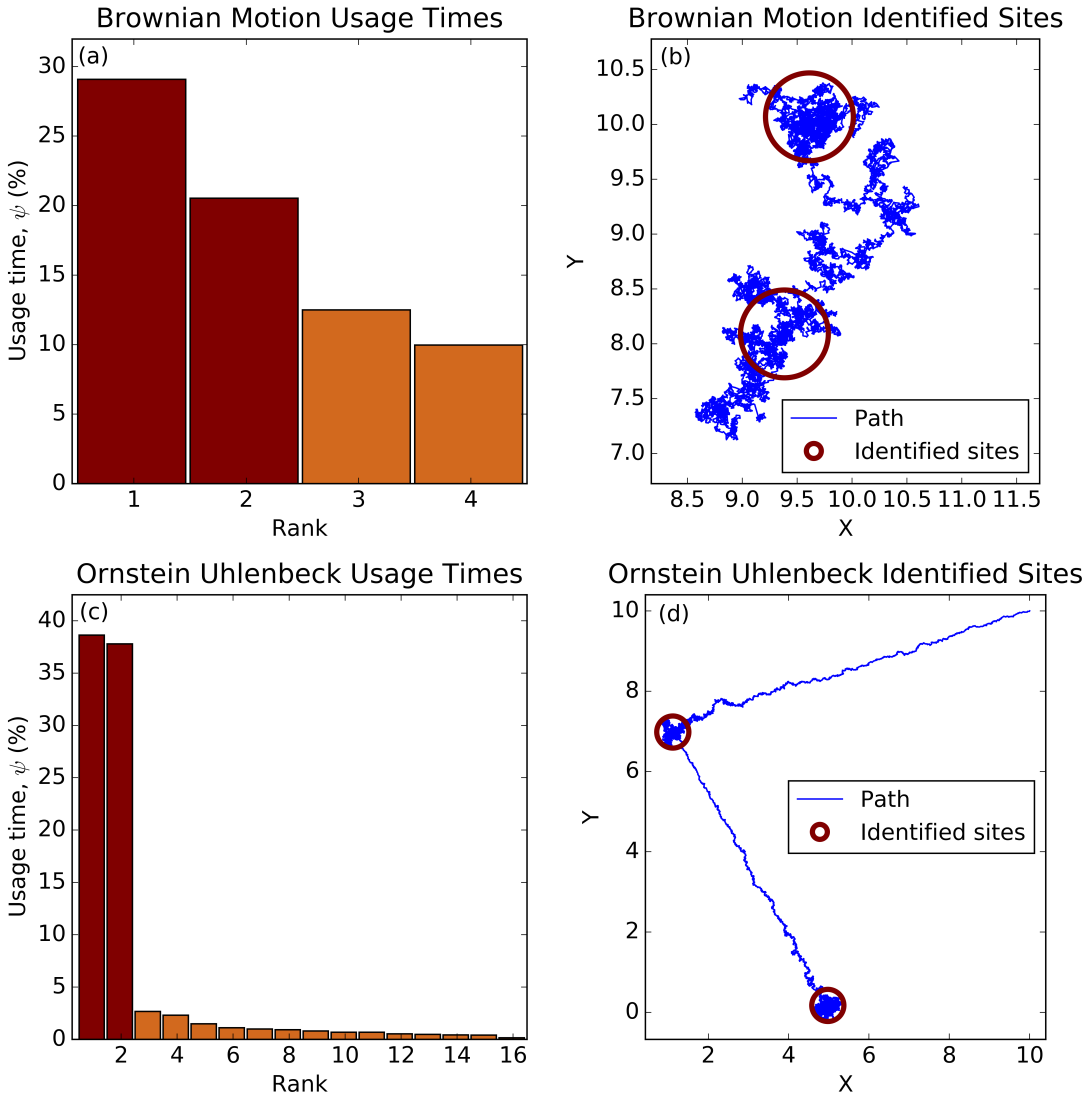

Figure 2.9: **Comparison of the results from a Brownian Motion simulation with its corresponding OU simulation**
Panels (a) and (b) correspond to the Brownian motion simulation, whereas panels (c) and (d) correspond to the OU simulation. Panels (a) and (c) show the usage times as percentages of the simulations' durations. Panels (b) and (d) show the locations of the identified sites of interest.

I use the OU simulations data set as an example of how to obtain a lower limit for T_{MPD}. For each OU simulation I create a corresponding Brownian motion simulation,

where both are created using a discrete-time approximation (as explained in Iacus (2009)), which means that I can set both trajectories to have the same mean step length. Therefore, both trajectories are approximately equal in length. However, for some of the larger radius values only one circle is left after the rarefication procedure resulting in $MPD_R = 100\%$, so I exclude these radius values when calculating the lower limit for T_{MPD}. To obtain the lower limit of T_{MPD} I calculate the mean MPD_R across all Brownian motion trajectories and radius values which do not result in $MPD_R = 100\%$. For the OU simulations this gives a lower limit for T_{MPD} of 52.2%. For the results presented in Section 2.2 I use $T_{MPD} = 65\%$, which was chosen since it is greater than the lower limit and reduced the number of paths assigned to the Green category with the incorrect number of sites identified.

§ 2.4 Schematic Representation

The Sites of Interest Algorithm enables a complex movement trajectory to be simplified into a set of transitions between sites (Fig. 2.10). This simplification opens up opportunities to answer questions concerning why an animal transitions between sites and why at the particular time points it was observed to do so.

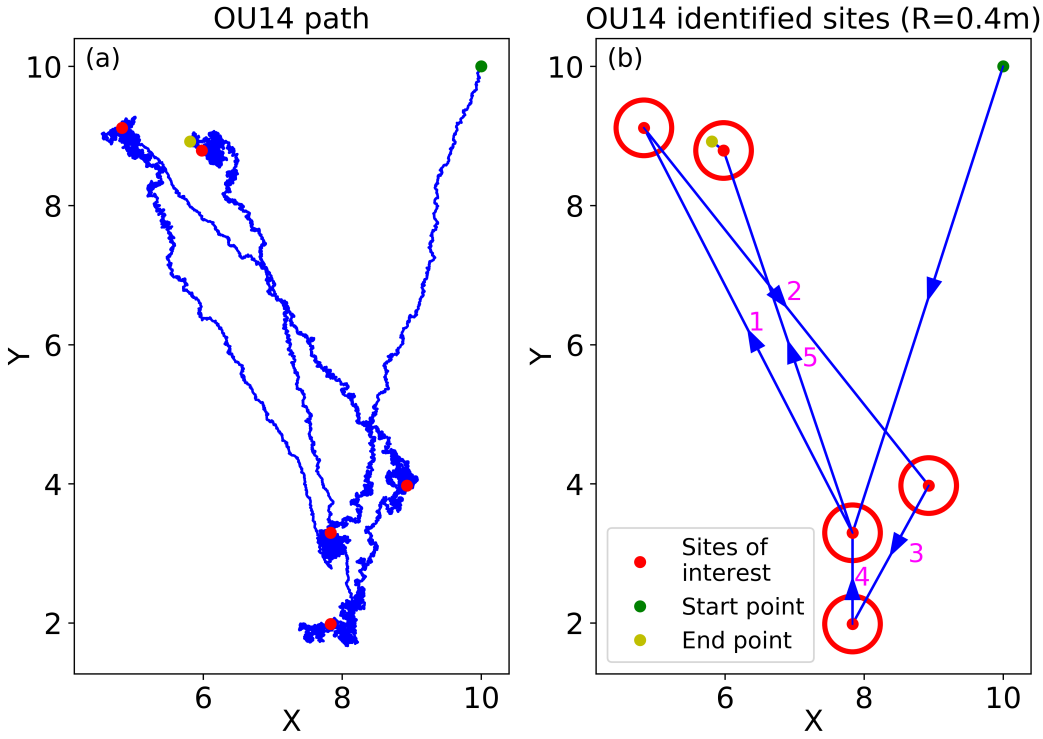

Figure 2.10: **OU Schematic**
A path from an Ornstein-Uhlenbeck simulation (a) and the schematic representation of movements between sites of interest (b). Arrows represent the movement direction from one site to another. Numbers represent the order of transitions between sites.

This schematic (Fig. 2.10b) can be seen as representing a Markov-chain, which is made up of a sequence of decisions to either stay at the same site or transition to another site, where the probability of each potential decision made at time, t, is based on the current state of the system at t. The current state of the system encompasses both the internal state of the animal and external state of the environment. For example, an animal may transition between sites because it is thirsty or hungry or tired and which site it transitions to may be effected by the distance it would need to travel.

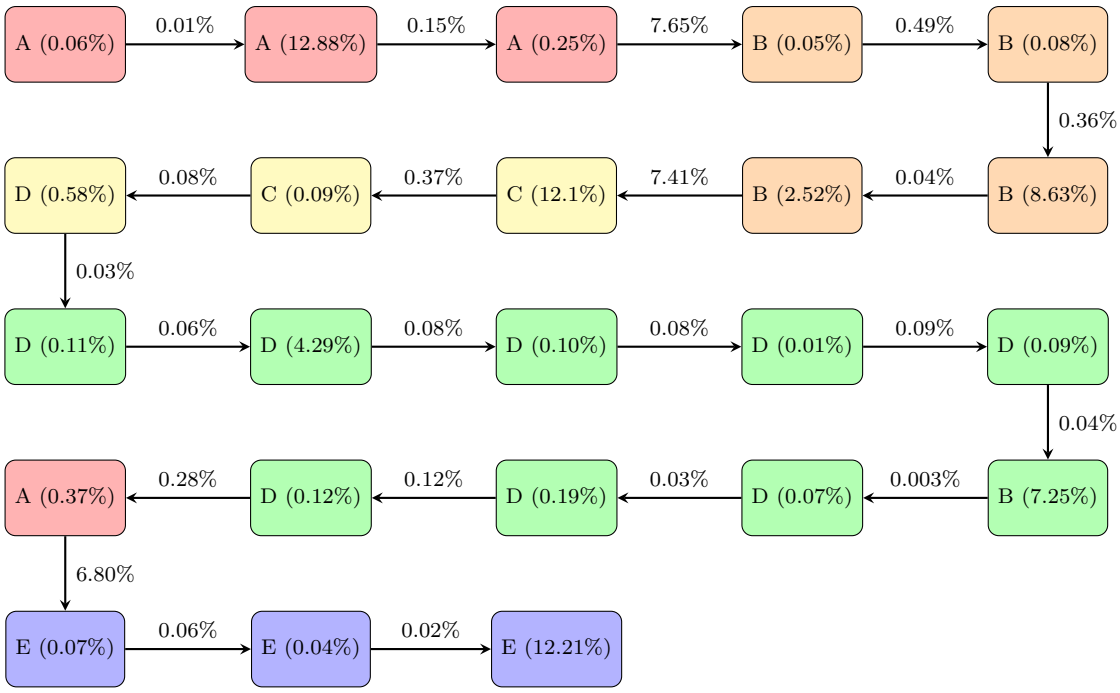

Figure 2.11: **Markov-chain-like flow chart of an OU path**
Flow chart of movements from site to site, where each box represents a visit to a site and a visit is defined to be a continuous time period spent within the site's boundary. Consecutive visits to the same site represent times when the simulated animal has left the site then re-entered it sometime later before visiting any other site. The percentage in brackets gives the amount of time spent at that site relative to the duration of the simulation. The percentages attached to arrows are the amounts of time spent transitioning between sites relative to the simulation duration.

The simplified trajectory can also be visualised as a flow chart showing the movements from site to site as well as the duration of each visit and the amount of time spent transitioning between sites (Fig. 2.11). This flow chart visualises immediate return visits to the same site before visiting any other site which cannot be seen in the schematic of Figure 2.10 (b). The flow chart can help determine when a significant visit has been made or whether the animal is just moving through a site. For example,

in Figure 2.11 there is a considerable initial visit to site A, but later the visit to site A between visits to sites D and E is only brief suggesting a transitory phase is being exhibited then. By visually inspecting the flow chart a user is able to determine if there is a particular movement routine that the animal is following, where they tend to visit sites in a particular order. Alternatively conditional entropy (Riotte-Lambert et al., 2016) can be used to detect repeating subsequences within the sequence of sites visited. More details are given in Section 3.7.

§ 2.5 Algorithm Comparison

Here, I compare the runtime and accuracy of calculating the usage and residence times from using the Sites of Interest Algorithm (referred to as the "Adapted Algorithm") and the method of Barraquand and Benhamou (2008) (referred to as the "Original Algorithm"), respectively. The Original Algorithm places circles centred at every location on the trajectory and checks between all consecutive points for either an entrance into or exit from the circle. In contrast, the adapted algorithm centres each circle on the boundary of the previous circle and initially checks for crossings between data points which are s points apart before checking between consecutive points within the interval if a crossing has occurred (see Section 2.1.1 for more details). Since the Original Algorithm checks for crossings between all consecutive points this can be viewed as using a crossing interval with $s = 1$. I also test how varying s affects the algorithm's runtime and accuracy.

The usage time is the exact amount of time spent inside a circle, whereas the residence time is the time spent in the vicinity of the circle. The residence time includes the usage time and any movements where the time between consecutive exit and entrance times is less than a given threshold, T_C. For example if the animal leaves the circle and then re-enters it 1 minute later then this 1 minute will be included in the residence time if $T_C \geq 1$ minute and not if $T_C < 1$ minute. Note that the usage time and residence time are equal when $T_C = 0$, thus making the usage time a special case of residence time. Hence, to compare the residence and usage times I set $T_C = 0$ when calculating the residence times.

2.5.1 Runtime

I use a straight line trajectory to compare runtimes of the two algorithms, since in the adapted algorithm this trajectory maximises the number of circles and thus can be considered a worst case scenario. All trajectories have the same length and the same radius value is used, since changing either of these would affect the number of circles used in the adapted algorithm and therefore also the runtime. Instead, I vary the number of points, i.e. the resolution of the trajectory.

The runtime is greatly decreased (approximately by two orders of magnitude) when using the adapted algorithm (with $s = 10$) in comparison to the Original Algorithm (note the different scales for the y-axes in Fig. 2.12). When the Original Algorithm takes hours to complete the Adapted Algorithm takes seconds. Figure 2.12 also shows that the runtime of the Adapted Algorithm increases linearly with the number of points,

whereas the Original Algorithm increases exponentially. This is because when the number of points increases the number of circles used in the Original Algorithm is greater and the amount of time necessary to calculate the residence times increases too. For the Adapted Algorithm the computational time of calculating the usage time similarly increases because of the higher resolution, but the number of circles is unaffected.

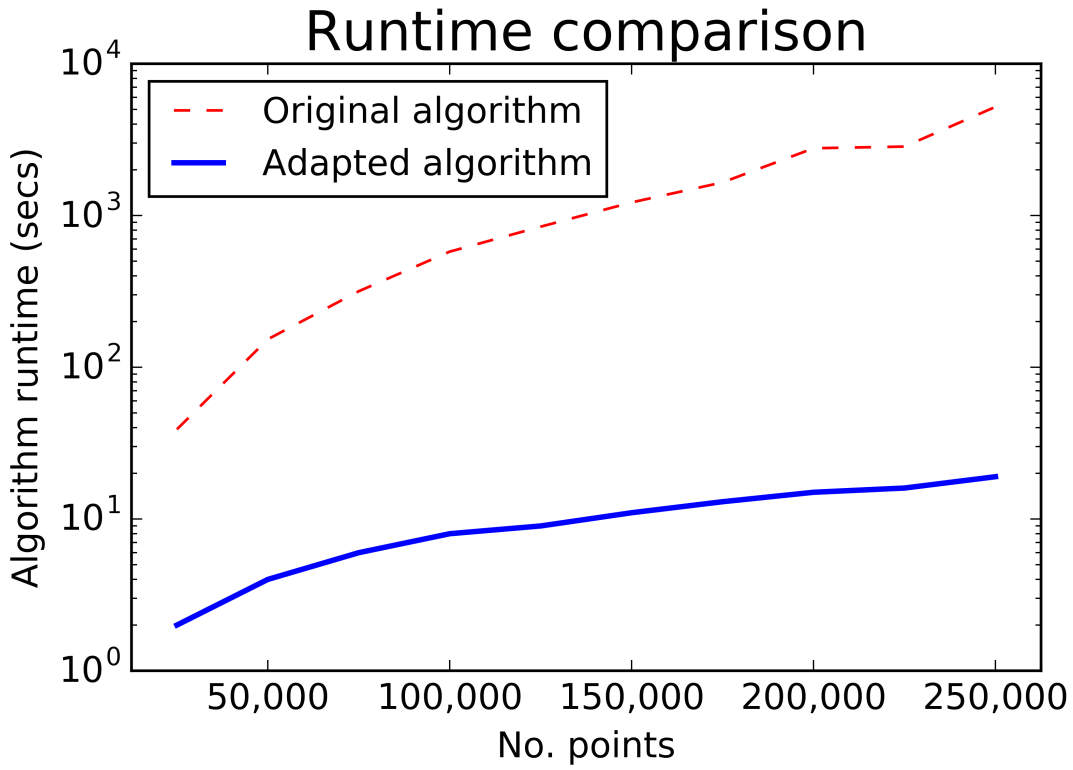

Figure 2.12: **Runtime comparison**
Runtime comparison of the Original Algorithm (Barraquand and Benhamou, 2008) in red and Adapted Algorithm (Munden et al., 2019) with $s = 10$ in blue as applied to straight line trajectories of differing resolutions as indicated by the number of points.

I find that changing s from 1 to 10 when using the Adapted Algorithm reduces the runtime by 90.85% on average for the OU paths. The runtime reduction is slightly greater than 90% since for some paths when $s = 10$ there are circles that are not included when compared to using $s = 1$. This happens when the first exit from a circle is missed. The change to s will roughly be inversely proportional to the runtime. In other words, the runtime from using a crossing interval of length s^* will be $1/s^*$ of the runtime from using $s = 1$.

Possible Further Reductions to the Runtime

It is possible to reduce the algorithm's runtime further by decreasing the number of circles. The algorithm could be adapted so that the algorithm only calculates the usage

time of every other circle in the set (i.e. circles with centres at $\mathbf{x}_{2i+1}^{(c)}$) and for the other circles (i.e. circles with centres at $\mathbf{x}_{2i}^{(c)}$) it would only calculate the first forward exit, which gives the location of the next circle. In this new set of circles, the boundaries of consecutive circles would touch rather than each circle being centred on the boundary of the previous circle. Reducing the number of circles would have a knock-on effect on the rarefication process, such that the same sites may not be identified when there are less circles compared to having more circles. Reducing the number of circles could lead to a situation where a true site is located between two circles, which means that the usage time at that site is split between the two circles and therefore may not be identified as a site. A site between two circles could be identified by changing the location of the first circle, but this would be equivalent to running the algorithm with more circles.

Another way to reduce the runtime could be to speed up the process of identifying crossing times. One suggestion could be to use the bisection method once a crossing has been identified as occurring within a crossing interval. However, it is possible that there are multiple crossings that occur within one crossing interval and the bisection method would only identify one of these. Also the computational time would not be greatly decreased when the size of the crossing interval is quite small. It would be more beneficial when the crossing interval is large, but when the crossing interval is large there is also likely to be more instances where there are multiple crossings within the same crossing interval.

2.5.2 Accuracy

Recall that s is the size of the crossing interval over which entrances and exits from the circles are checked for. This means that when $s = 1$ the algorithm checks between all consecutive points and the usage time in this case is correct up to the resolution of the data. Therefore, I measure the accuracy of using $s = 10$ against $s = 1$. I check the accuracy for all OU paths and radius values. I quantify the accuracy in several ways. The first is the number of circles that are included and their positions. The second is in the usage times of the circles with the same positions. Finally I quantify the accuracy with the number of sites identified.

Comparing using $s = 10$ to $s = 1$ I find that 74.7% of cases (across OU simulations and radius values) have the same number of circles, $N^{(c)}$, before applying the rarefication procedure. Of the paths with the same number of circles, 32% of these have the exact same positions, i.e. $\{\mathbf{x}_1^{(c)}, \ldots, \mathbf{x}_{N^{(c)}}^{(c)}\}$ is the same when $s = 1$ and $s = 10$.

The average usage time error per circle is 1.9 time units. For example, this means that the usage time for data collected at a resolution of 1Hz would be on average inaccurate by 1.9 seconds. Note, this is less than the size of the crossing interval ($s = 10$), which is the maximum of any error per passage. In the cases when the usage time error is significantly large, then it may in fact be the case that a larger radius value would be more appropriate. The error may come from there being a lot of short time periods when the animal leaves the circle before immediately returning to the same circle which

could be missed if they are shorter than the crossing interval length, s.

For 92.51% of cases the same number of sites (n_R) are identified from using $s = 10$ compared to $s = 1$. In the vast majority of cases, even if there is a change in the number of circles or their positions or usage times, the same number of sites is identified.

When choosing the value of s, there is a trade-off between accuracy and runtime. Larger values of s decrease the runtime, but increase the chance of errors. The best value of s will be dependent on the particular data of a user. Even using a value of $s = 1$, which may make the algorithm's runtime larger than any other value of s will still be quicker than the original algorithm and just as accurate.

§ 2.6 Choosing the radius value, R

It is easy to put initial limits on the radius values since it neither makes sense to use a radius smaller than half the animal's body length, nor would it be logical to use a radius value such that only one circle is left after the rarefication procedure. If the path is short enough and the algorithm runtime low enough, it would be possible to repeat the algorithm many times with various radius values within these limits. The most appropriate radius value can then be chosen from the set based upon the results. Choosing the most appropriate radius value can be done by visually inspecting the usage time histograms, looking for one with the most striking maximum percent drop. For example, significant drops can be seen from the OU simulations and particularly in contrast to the Brownian Motion simulations, as seen in Figure 2.9.

There exist other methods (Fauchald and Tveraa, 2003) which aim to identify periods when an animal is performing an area restricted search (ARS) and specifically to determine the size of these ARS areas. These same techniques for determining the size of these ARS areas can be used to choose a radius value to use in the Sites of Interest Algorithm. ARS areas are also sites of interest, but correspond to a particular behaviour. ARS typically corresponds to foraging behaviour, which is typified by slower speeds and high turning frequency resulting in a very constricted path.

Fauchald and Tveraa (2003) suggests using variance-scale plots calculated from the first passage time (Redner, 2001), which is the time from when the animal first enters the circle to when it first leaves (i.e. $F_1 - B_1$). The variance-scale plot is constructed as the variance in the logarithm of first passage times as a function of R. The radius value with maximum variance is chosen as the radius of the sites. This is an easily applicable method (Pinaud and Weimerskirch, 2005; Bailey and Thompson, 2006; Pinaud and Weimerskirch, 2007), but it was criticised in Barraquand and Benhamou (2008) because the radius estimate can be unreliable for certain scenarios. This led to the method being revisited in Kapota et al. (2017), where the criticisms raised in Barraquand and Benhamou (2008) were specifically addressed. In particular, in Kapota et al. (2017) they simulate the scenario where the scale of sites varies. They show how these different scales are identified by multiple peaks in the variance scale plots. Variance-scale plots could therefore be used alongside the Sites of Interest Algorithm if the size of the sites is of particular concern to a user of the algorithm.

Another possible approach to estimate the size of a site could be to simulate a correlated random walk and approximate the size of the site from the simulated path. The radius can be chosen such that the circle centred at the point of attraction encompasses the entire path. To create the correlated random walk it is necessary to estimate the strength of attraction to the site. This strength of attraction can be estimated using the mean vector length of the wrapped Cauchy distribution (McClintock et al., 2012). Also the step length distribution of the correlated random walk should be approximately the same as the step length distribution from the empirical data.

Alternatively, a trial and error method can be used, where an initial radius value is tested then the results from it are used to infer whether a larger or smaller radius value would be more appropriate. There are several ways to infer whether a radius value is too small or too large. For example, clusters of sites indicate that a larger R would be more appropriate, which is clearly seen in Figure 2.5 (b). Conversely, if there are large areas of the identified sites not used then the radius value may be too large (Fig. 2.5d). If there are a lot of instances where the animal leaves the circle and then re-enters shortly afterwards, this may indicate that R is too small. On the other hand, if there are very few circles left after the rarefication procedure this could suggest that R is too large. The size of sites can also be estimated from a heat map of usage times (Section 2.7), which is constructed ideally using as small a value for R as possible.

It is possible to use various radius values in the same iteration of the algorithm. This would mean that sites of different sizes can be identified without rerunning the algorithm with a different radius value. However this would require extra computational time to compute the best radius value to use for each circle. If the algorithm is adapted to have varying radius values then it would also be necessary to normalise the usage times by the area of their respective circles. It would be unfair to compare the usage times of different size circles, since it would be natural for a larger circle to have a larger usage time. Once the location of a site has been identified the size of the site could be estimated by maximising the normalised usage time across different radius values.

§ 2.7 Heat maps

The results from the Sites of Interest Algorithm can be used to create heat maps based on the usage time (ψ_i), number of visits ($1 + \varphi_i + \beta_i$) or average usage time per visit ($\psi_i/(1 + \varphi_i + \beta_i)$). These are similar to the Utilisation, Recursion and Intensity Distributions calculated in Benhamou and Riotte-Lambert (2012), though they are not the same due to the differences between the usage and residence time. The residence time includes time periods spent outside the circle between visits that are below a given threshold, whereas the usage time does not include these (Section 2.1.1)

Heat maps are constructed by assigning every point within a circle the attribute value (i.e. usage time, number of visits or average usage time per visit) of that circle. For any point in the intersection of multiple circles the value assigned at that point is the mean attribute value of all circles the point is contained within.

Examples of heat maps of usage time are shown in Figure 2.13 for radius values of 0.05 and 1.5. The smaller radius value more accurately captures the shape of the space

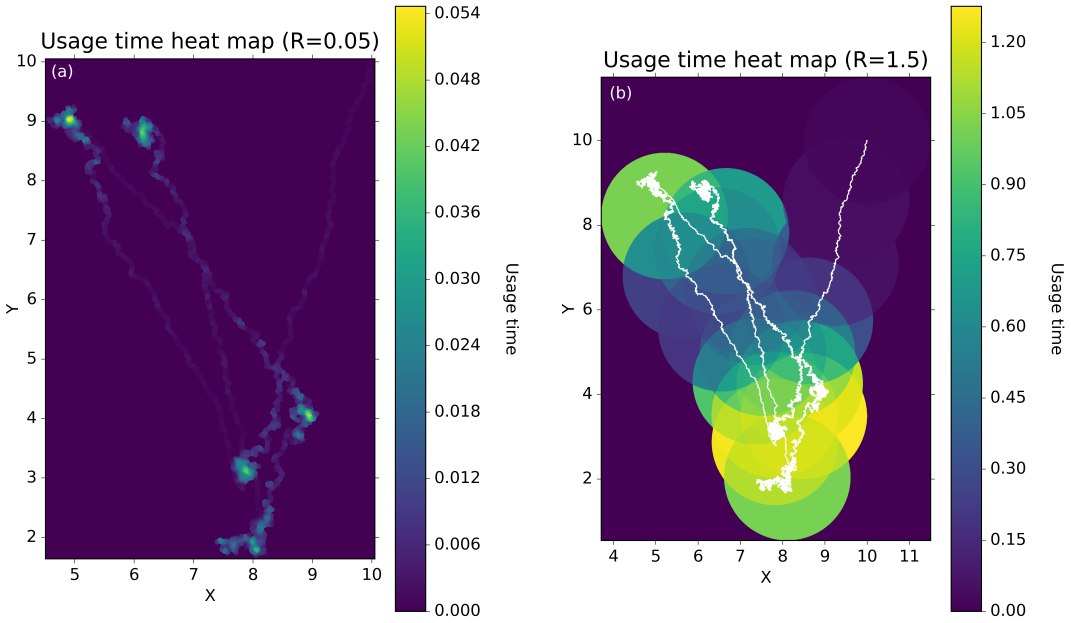

Figure 2.13: **Orstein-Uhlenbeck example heat map**
Heat maps of the usage time calculated using the Sites of Interest Algorithm as applied to the same Ornstein-Uhlenbeck path (in white) with radius values of 0.05 and 1.5 in panels (a) and (b), respectively.

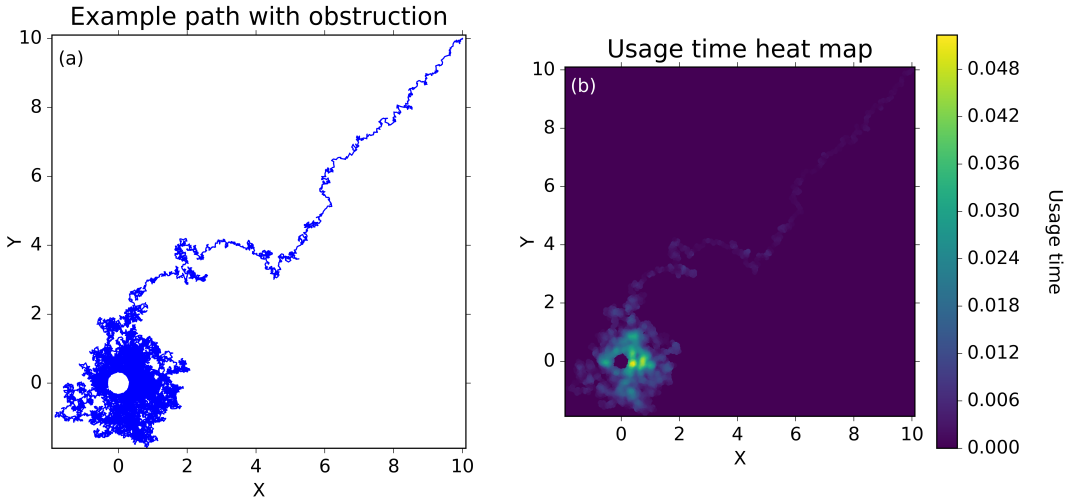

Figure 2.14: **Example of a non-circular site**
An example of a trajectory with an obstruction within the the site of interest (a), which creates an annulus shaped site of interest. This annulus shape can be seen in the heat map of usage times (b).

used, since areas within a circle are assigned the attribute value of that circle even if they are not used (Fig. 2.13b). However, the algorithm will take longer to complete when the radius value is smaller, so using a larger radius value may be more useful when a researcher is only interested in the broad-scale space-use patterns.

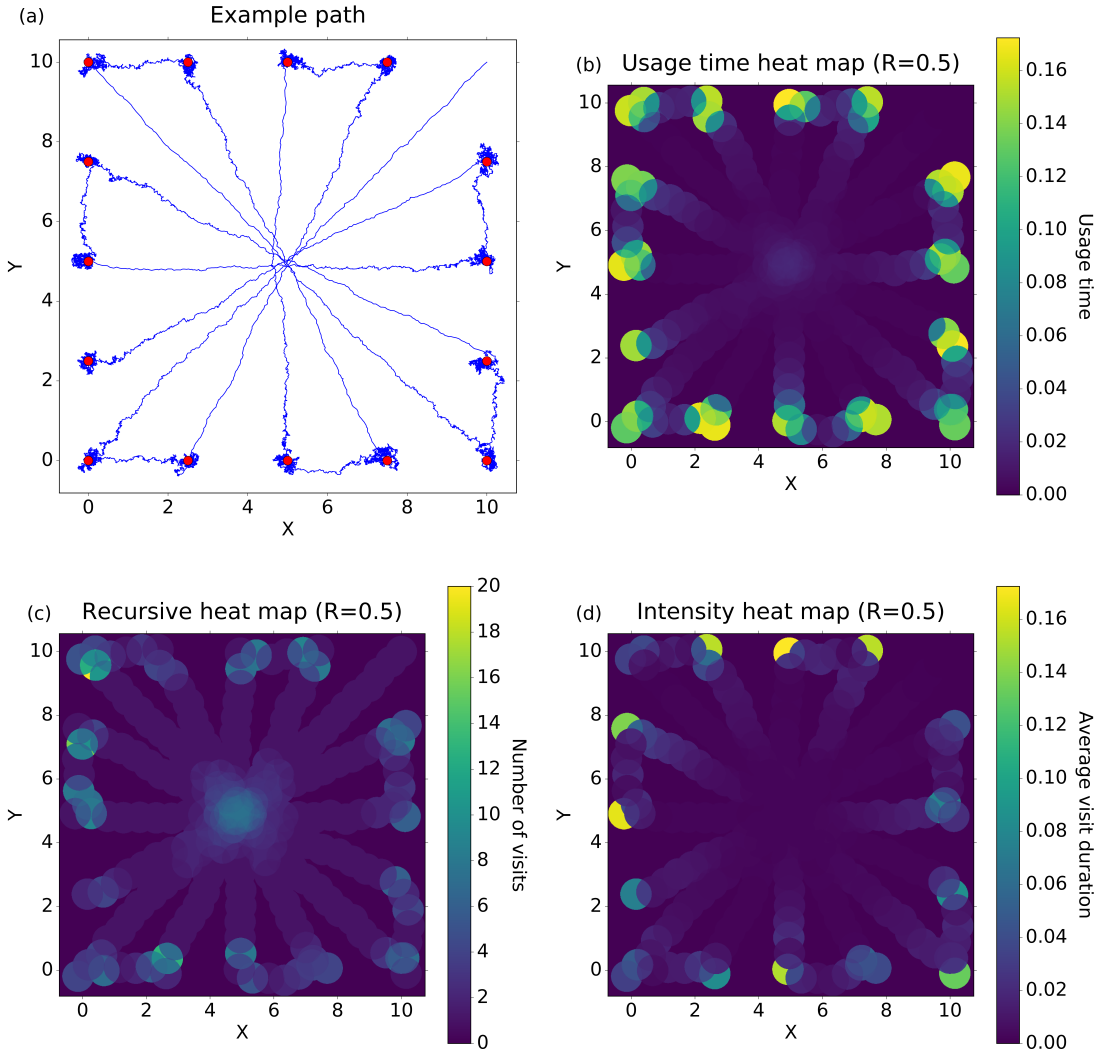

Figure 2.15: **Heat maps of different distributions**
Different distribution heat maps of an example path (a). The different distributions are similar to the utilisation distribution (b), recursion distribution (c) and intensity distribution (d), which are constructed from the usage time, number of visits and average visit duration, respectively.

The usage time heat map can be used to identify the shape of a particular site of interest if it is non-circular. The shape of a site may be determined by the landscape. For example, an impassable landscape feature (e.g. fences) will prevent the animal from

using that section of the identified site. In Figure 2.14 there is an unavailable area at the centre of the site, which can be seen in the heat map. As well as the shape of a site, the usage time heat map could also be used as a visual check that the algorithm has correctly identified the location and size of the sites.

The heat map of the number of visits can be used to identify areas with a high number of visits, but not usage time. These areas can be identified by visually comparing the heat maps of usage time and number of visits. Figure 2.15 demonstrates the differences between using the three heat maps. In panel (b) the points of attraction are easily identified from using the usage time, whereas using the number of visits in panel (c) highlights the intersection point at the centre of the plot.

Using the average usage time per visit could be used to identify if any sites are overwhelmingly dominant, such as a sleeping site. It may be necessary to remove such sites in order to identify less obvious and unexpected sites as discussed in Section 2.1.5.

§ 2.8 Discussion

The Sites of Interest Algorithm is effective in identifying the known areas of interest. Unlike other model-based methods, such as State Space Modelling (Jonsen et al., 2005, 2007; Patterson et al., 2008) or Hidden Markov Models (Morales et al., 2004; Patterson et al., 2009; Langrock et al., 2012), the Sites of Interest Algorithm makes no assumptions about why an area might be of interest. For example, foraging patches can be identified from a change in the distribution of step-lengths and turning angles (Benhamou and Bovet, 1989). In contrast, the Sites of Interest Algorithm can first be used to sift through long data streams to identify the areas of interest, then the motivations underlying the animal's choice of those particular areas can be identified. Motivations for using a site can be inferred by inspection of environmental data or analysing the finer scale movements within a site to infer the type of behaviour that is being exhibited. It is often the case that only limited environmental information is available due to the difficulty of gathering such data across a wide area (Dodge et al., 2013), so as such the Sites of Interest Algorithm could be used to identify the particular areas covered by the animal that could be analysed in more detail. For example, food patches could be identified then the availability of food in these patches can be quantified.

How well an animal exploits an area can be analysed using techniques such as optimal foraging theory (Charnov et al., 1976; Pyke, 1984). The model described in Charnov et al. (1976) and Bond (1980) assumes that an animal moves through a patchy habitat, where these patches represent sources of food. However, in order to apply the model to a real world environment, the patches need to be known *a priori*. Therefore, there is a need to first identify these patches, which can be done using the Sites of Interest Algorithm, since often only limited environmental data is available (Dodge et al., 2013) and especially when the area available to the animal is quite large. It is particularly useful to study when the optimal time is for an animal to leave a patch, this can then be compared with the time the animal is observed to have left the identified site of interest, which is part of the output from the Sites of Interest Algorithm.

Behaviours can also be inferred using behavioural change point analysis (Yoda et al.,

1999; Nathan et al., 2012; Williams et al., 2017). This can involve using the accelero-magnetometer data, whereby postures can be identified. From these postures the internal state of the animal can then be inferred (Wilson et al., 2014). For example, the disease state of an animal may be inferred for certain diseases such as Parkinson's which has a clear effect on animal posture.

The Sites of Interest Algorithm is inspired by work in Barraquand and Benhamou (2008) and Benhamou and Riotte-Lambert (2012), but adapted specifically for use on high resolution data. I show how decreasing the number of circles and the time taken to calculate the usage times greatly reduce the algorithm's runtime. The runtime is decreased by approximately two orders of magnitude (see Section 2.5.1). As these high resolution data sets rise in popularity and availability the need for such adaptations become ever more needed.

As well as the adaptations to speed up the algorithm in Benhamou and Riotte-Lambert (2012), their aims also differ from those of the Sites of Interest Algorithm. The method presented here aims to identify specific areas which are delimited from the rest of the landscape, whereas in Benhamou and Riotte-Lambert (2012), they give a more general description of the animal's space use patterns. They use heat maps of the utilisation, intensity and recursion distributions to identify space use patterns. Similar heat maps (see Section 2.7) can be constructed as part of a post processing analysis of the results from the Sites of Interest Algorithm to identify non-circular sites (Fig. 2.15). Though creating heat maps from the results of the Sites of Interest Algorithm is faster than the method of Benhamou and Riotte-Lambert (2012) I would recommend using their method if the resolution of the data is not high enough. A rule of thumb for how high the resolution should be to use the Sites of Interest Algorithm is that the maximum distance between consecutive points is less than the animal's body length.

These heat maps can also be used to identify areas which are important to the animal, but do not necessarily have high usage times. For example, an area with a large number of visits but not usage time could indicate a crossing point (Fig. 2.15c), such as a ford in a river. Places with a high average visit duration could be sleeping sites, which may not be visited often but still have high usage times. It may be important to identify areas of high visit duration, since these could be too dominant and therefore obscure other sites from being identified (see Section 3.2).

The colour assignment gives a simple goodness-of-fit test, which is a novel aspect of the Sites of Interest Algorithm. Indeed, goodness-of-fit tests are often missing from statistical studies of animal movement (Potts et al., 2014). I show how the colour assignment works well at filtering out incorrect results, since when applied to the OU simulations, only one path gives a false positive result (Fig. 2.8). The colour assignment acts as a warning system of when further analysis may be necessary. For example, a Red colour assignment may mean that a different set of radius values would be more appropriate or that the algorithm should only be applied to a subsection of the path due to an overwhelmingly dominant site (e.g. a sleeping site; see Section 3.5).

The Sites of Interest Algorithm simplifies a complex animal movement trajectory down to the key broad-scale movement decisions that an animal makes. This leads to

a Markov-chain like description of the animal's movement decisions (see Section 2.4). This simplified description of the movement trajectory enables a user to more easily answer questions concerning the animal's movement routine. These questions may be answered using pre-existing methods, such as conditional entropy (Riotte-Lambert et al., 2013) and sequence analysis techniques (De Groeve et al., 2016). Recursive movement patterns may arise from a particular periodic behaviour such as traplining (Thomson et al., 1997; Lihoreau et al., 2011). More details are given in Section 3.7.

In summary, the Sites of Interest Algorithm enables a user to fully utilise the great wealth of information contained within ultra-high-resolution data by simplifying an animal's path, thus opening up the opportunity for other methods to be used such as step selection analysis (SSA, Fortin et al. (2005); Rhodes et al. (2005)). SSA is used to understand the motivations underlying decisions that the animal makes. Traditionally in SSA (Fortin et al., 2005; Rhodes et al., 2005) these decisions are implicitly assumed to occur between successive recordings, but when data are gathered at a very high frequency (e.g. 1 Hz) it is no longer appropriate to assume the animal is making a different decision at each recorded location. Instead, simplifying a trajectory enables SSA to be applied to the movements between sites.

§ 2.9 Glossary

Symbol	Explanation
$\mathbf{x}_i$	recorded location of the animal at time, t_i
t_i	sequence of times at which the animal's location is recorded
N	number of recorded locations
$\mathbf{x}_i^{(c)}$	centre of the circle centred at time point, $t_i^{(c)}$
$t_i^{(c)}$	sequence of times indicating where the circles are centred
$N^{(c)}$	number of circles
R	circle radius
$d(\mathbf{x}, \mathbf{y})$	Euclidean metric distance between $\mathbf{x}$ and $\mathbf{y}$
s	length of the crossing interval
F, B	sets of forward and backward passage crossings, respectively
ψ_k, β_k	number of forward and backward passages for the k^{th} circle, respectively
φ_k	usage time of the k^{th} circle
Φ	set of usage times
$\varphi_k^{(o)}$	usage time of the k^{th} ordered circle
$\Phi^{(o)}$	ordered set of usage times
φ_k'	usage time of the k^{th} ordered and reduced circle
Φ'	ordered and reduced set of usage times
n_R	number of sites identified when a radius of R is used
R_s, R_t	radii identified from using the stability and threshold criteria, respectively
MPD_R	maximum percent drop resulting from using a radius of R
T_{MPD}	maximum percent drop threshold value

Table 2.2: **Chapter 2 Glossary**
Glossary of symbols used in Chapter 2

Chapter 3

Applications of the Sites of Interest Algorithm

The Sites of Interest Algorithm Munden et al. (2019) is primarily used to identify clearly delimited areas along an animal's path where they spend a disproportionately large amount of their time. Recall that the algorithm works by sliding a circle of radius R along an animal's path and calculating the time they spend in each circle, termed the usage time. The algorithm is applied using multiple radius values, then two criteria are used to choose the most suitable radius value.

In this chapter I test the efficacy of the algorithm when it is applied to cattle (*Bos taurus*) and goat (*Capra aegagrus hircus*) data. In these empirical data sets there are sites which are known *a priori*, such as a watering hole, two milking station locations, a goat pen and salt blocks, so I test whether the algorithm is able to identify these sites.

I also use these data as demonstrations of how to apply the algorithm practically. In some cases post-processing methods are needed to identify all of the sites. I find for the cattle data that there is one overwhelmingly dominant site that obscures other sites from being identified, hence this necessitates further examination of the algorithm's results to identify all of the sites (described in Section 3.2).

In section 2.1.5 I introduce an extension to the algorithm enabling it to be applied across a collection of paths, so I use the cattle data to highlight the differences between applying the algorithm to individual paths and a group of paths. Applying the algorithm to a group identifies shared sites of interest, whereas applying it to each trajectory separately highlights individual variation.

Identifying sites of interest enables a long complex trajectory to be simplified down to a sequence of site residencies and transitions between sites. I therefore discuss some of the applications of these simplified trajectories to identify periodic patterns in an animal's movement routine and understand the motivations underlying their decisions to transition from one site to another.

§ 3.1 Data

Here I introduce the two data sets that I use to test the Sites of Interest Algorithm. In both cases the data were collected from magnetometers and accelerometers housed in Daily Diary tags (Wildbytes Technologies `http://www.wildbyte-technologies.com` and Gipsy-5 tags; TechnoSmArt Tracking Systems `http://www.technosmart.eu`), placed in custom made 3D printed ABS plastic housing and attached to commercial nylon collars (Fearing Lifestyles, Durham, UK). From these readings the trajectories are reconstructed using the Dead Reckoning procedure (Bidder et al., 2015) and the Framework4 software (Walker et al., 2015).

I separate the cattle data into two groups based on the duration of the paths, where paths in Group 1 are roughly 10 hours long and those in Group 2 are roughly a week long. Groups 1 and 2 contain seven and five individual paths, respectively. I denote the individual paths from Group 1 and 2 as 1.1-7 and 2.1-5, respectively. Accelerometers recorded at a frequency of 20Hz, whereas magnetometer readings were recorded at 6Hz. The accelero-magnetometer data were then subsampled at 1Hz, so the Dead Reckoned paths are made up of relocations with a frequency of 1Hz. GPS recordings were taken every 15 minutes and used to correct for errors in the Dead Reckoned paths.

In the goat data set there are seven individual paths, each with a rough duration of 7 days. Accelerometer, magnetometer and temperature readings were gathered at resolutions of 20, 8 and 2Hz, respectively. GPS recordings were made every 15 minutes. The accelero-magnetometer data were subsampled at 1Hz, so the Dead Reckoned paths also have a frequency of 1Hz.

Both the cattle and goats live in the Bauges Mountains, which is situated in the French Alps. The cattle have several sites of interest known *a priori*, including a watering hole and two milking station locations. We observed that the cattle tend to move between these known sites as well as areas for grazing and resting. The goats spend most of their time in a central area which encompasses a goat pen and nearby salt blocks. They tend to rest in this central area and make foraging excursions out into the surrounding area.

§ 3.2 Cattle Results

3.2.1 Colour assignment Results

Following Section 2.3 I use Brownian motion simulations to find a minimum value for the maximum percent drop threshold value (T_{MPD}), which is used to choose a radius value that results in sites that are clearly distinct from all other areas. For the cattle paths from Groups 1 and 2 I find the lower bound of T_{MPD} to be 47.5% and 45.6%, respectively. I use a threshold value of 50% in the colour assignment (see Section 2.1.4 for details), which gives an indication of the confidence the algorithm has in the results, where the colours are Green, Amber and Red, which represent strong, intermediate and weak levels of confidence, respectively.

Figure 3.1 shows the colour assignment results from applying the algorithm to the

cows individually from Group 1 (panel a) and Group 2 (panel b). Full details are given in Table 5.9. The Amber and Red colour assignments suggest that it may be necessary to analyse the results further, which I explain in the following sections (Sections 3.2.2 and 3.2.3).

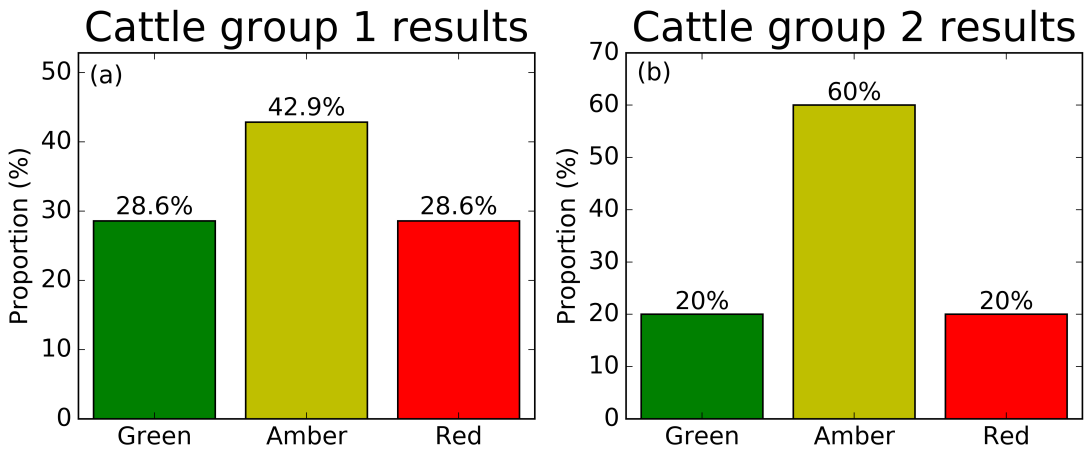

Figure 3.1: **Cattle colour assignment results**
Colour assignment results from Cattle Groups 1 and 2 in panels (a) and (b), respectively. Green, Amber and Red assignments represent strong, intermediate and weak confidence in the algorithm's results, respectively. I use a threshold value, T_{MPD}, of 50%.

3.2.2 Cattle Group 1

Only two paths from Cattle Group 1 are assigned to the Green category (Fig. 3.1a). On the other hand, when they are analysed collectively the algorithm identifies two sites that are known *a priori*, which are a watering hole (site A2 in Fig. 3.2d) and a milking station (site B2 in Fig. 3.2d). Other less expected sites (sites C2 and D2 in Fig. 3.2d) are also identified. From a researcher's visual observations of the cattle's movements I hypothesise that sites C2 and D2 are a grazing and resting site, respectively. Note that when the radius value is 100m the algorithm identifies the maximum percent drop after the sixth circle, but from the histogram of usage times there is a clearly significant drop after the fourth circle (Fig. 3.2c), so I take these four circles to be the sites of interest. The algorithm segments all of the circles into either sites or non-sites, where the sites have disproportionately large usage times. However, in this case when six sites are identified this is because the circles assigned to be non-sites have disproportionately low usage times compared to all over circles (i.e. there is a significant drop after the sixth circle).

We see in Figure 3.2 how changing the radius value affects the sites that are found. The results from using $R = 100\,\text{m}$ (panels c and d) show that some of the sites found from using $R = 20\,\text{m}$ that are quite close together may be clustered into one site (e.g. the site to the South-East). However, the site to the South-West (sites B1 and F1 in

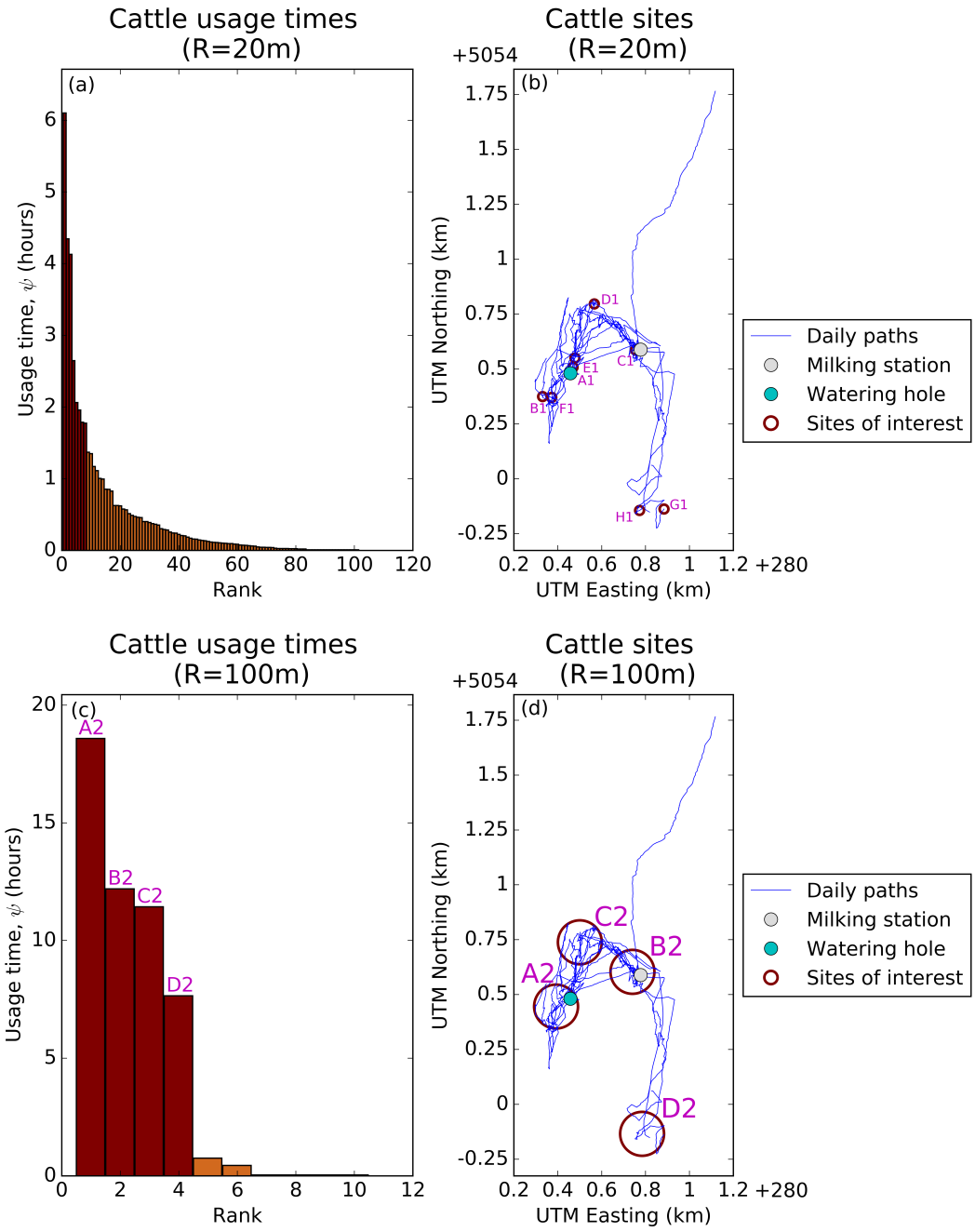

Figure 3.2: **Cattle group 1 results**
Results from applying the algorithm as applied to the cattle paths as a group using a radius value of 20m in panels (a) and (b) and 100m in panels (c) and (d). Panels (a) and (c) give the usage times from the rarefied list of circles, with the sites in brown and non-sites in orange. Panels (b) and (d) give all of the cattle paths in blue and the sites in brown. The sites are labelled from A1 to H1 for panels (a) and (b) and from A2 to D2 for panels (c) and (d), where sites A1 and A2 have the highest usage times when the radius is 20m and 100m, respectively.

panel b) is missed when using the $R = 100\,\mathrm{m}$, since this site seems to have conflated with the site around the watering hole. Therefore we see how sites can be of different sizes and it can be helpful to use different radius values to identify them.

3.2.3 Cattle Group 2

Since only one path from Cattle Group 2 is assigned to the Green category (Fig. 3.1b) I examine the results further and find that for several of the paths only one site is identified, which has an overwhelmingly large usage time. I term these sites as *extreme high-use sites* (EHU sites) and hypothesise that they have such large usage times because they are used for sleeping, since the EHU sites tend to be visited during the night (Fig. 3.11). This site can clearly be seen in the usage time heat maps (Fig. 3.4), which give a general description of the cattle's space-use patterns (details of constructing the heat maps are given in Section 2.7).

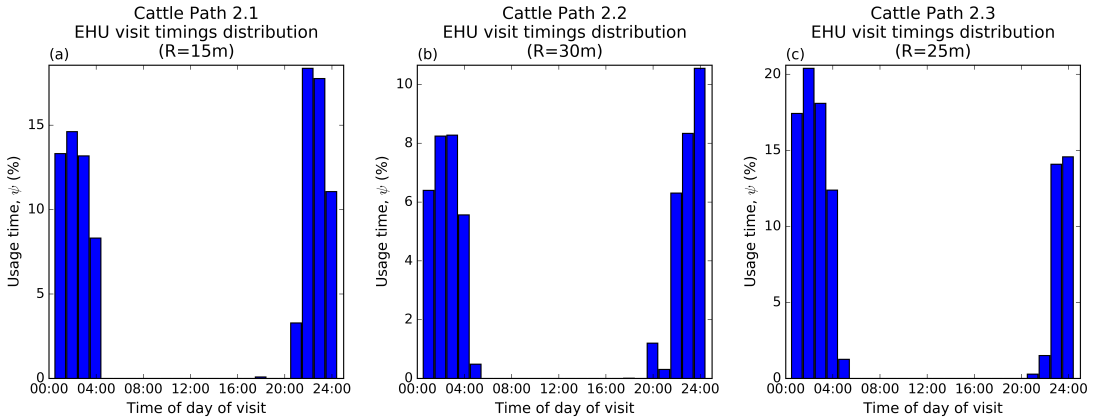

Figure 3.3: **EHU visit distribution**
Histograms of the time spent in the extreme high-use sites for different cattle paths, where the usage time is given as a proportion of the study duration (roughly 1 week). Panels (a)-(c) show the histograms for Cattle Paths 2.1-3, respectively.

The high usage times of the EHU sites means that no other sites are identified, so in these instances it is valuable to examine the second and third highest percent drops rather than the maximum, as described in Section 2.1.3, to see if these capture any known sites of interest. This leads to several more sites being identified including the watering hole and the two milking station locations (Fig. 3.5). There are two milking station locations since the milking station was moved during the study, which explains why the milking station at the second location is only used by Cattle Paths 2.4 and 2.5.

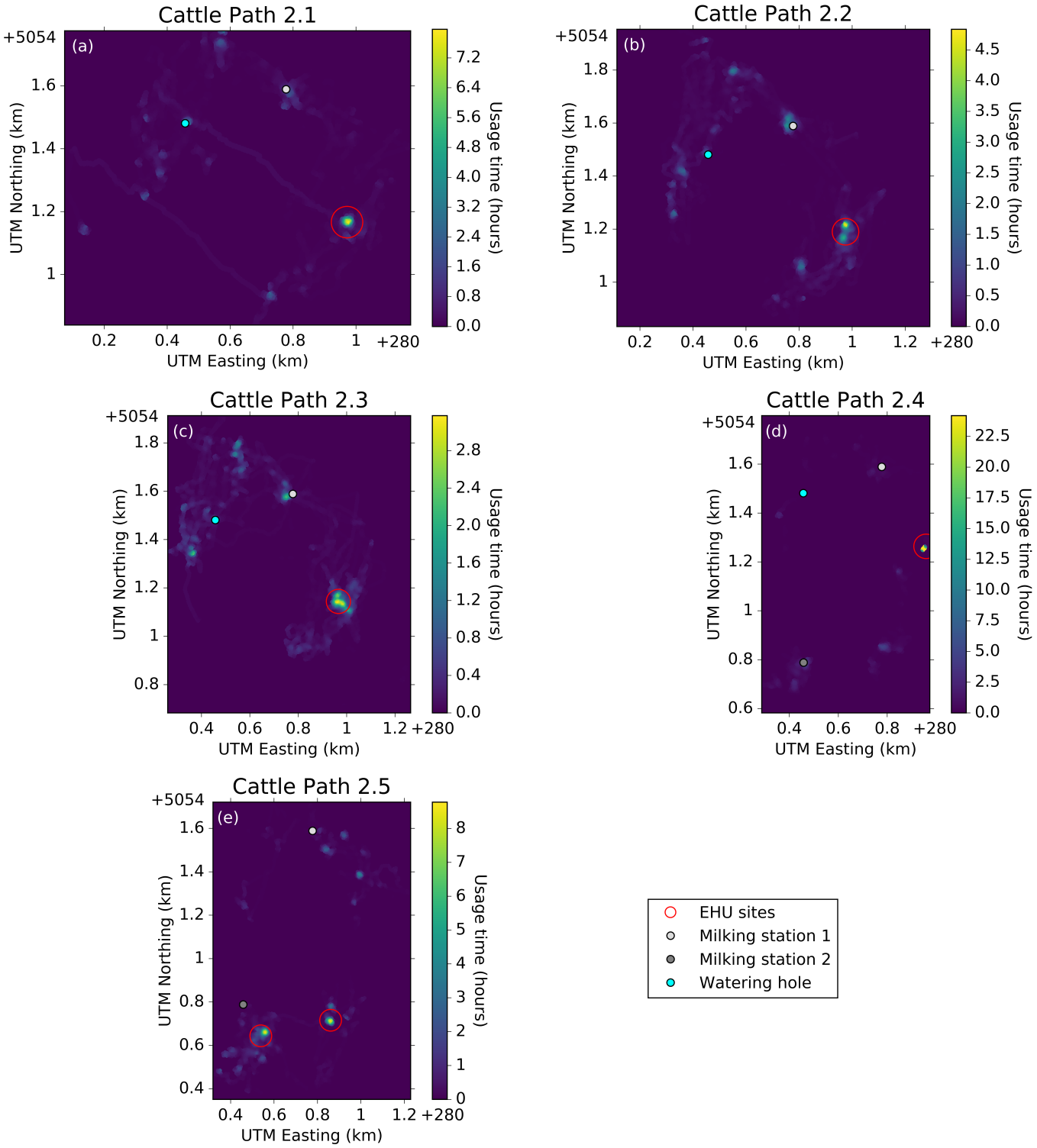

Figure 3.4: **Cattle Group 2 heat maps**
Heat maps of usage time for cattle paths 2.1-5 in panels (a)-(e), respectively. Extreme high-use (EHU) sites of interest are represented by the brown circles.

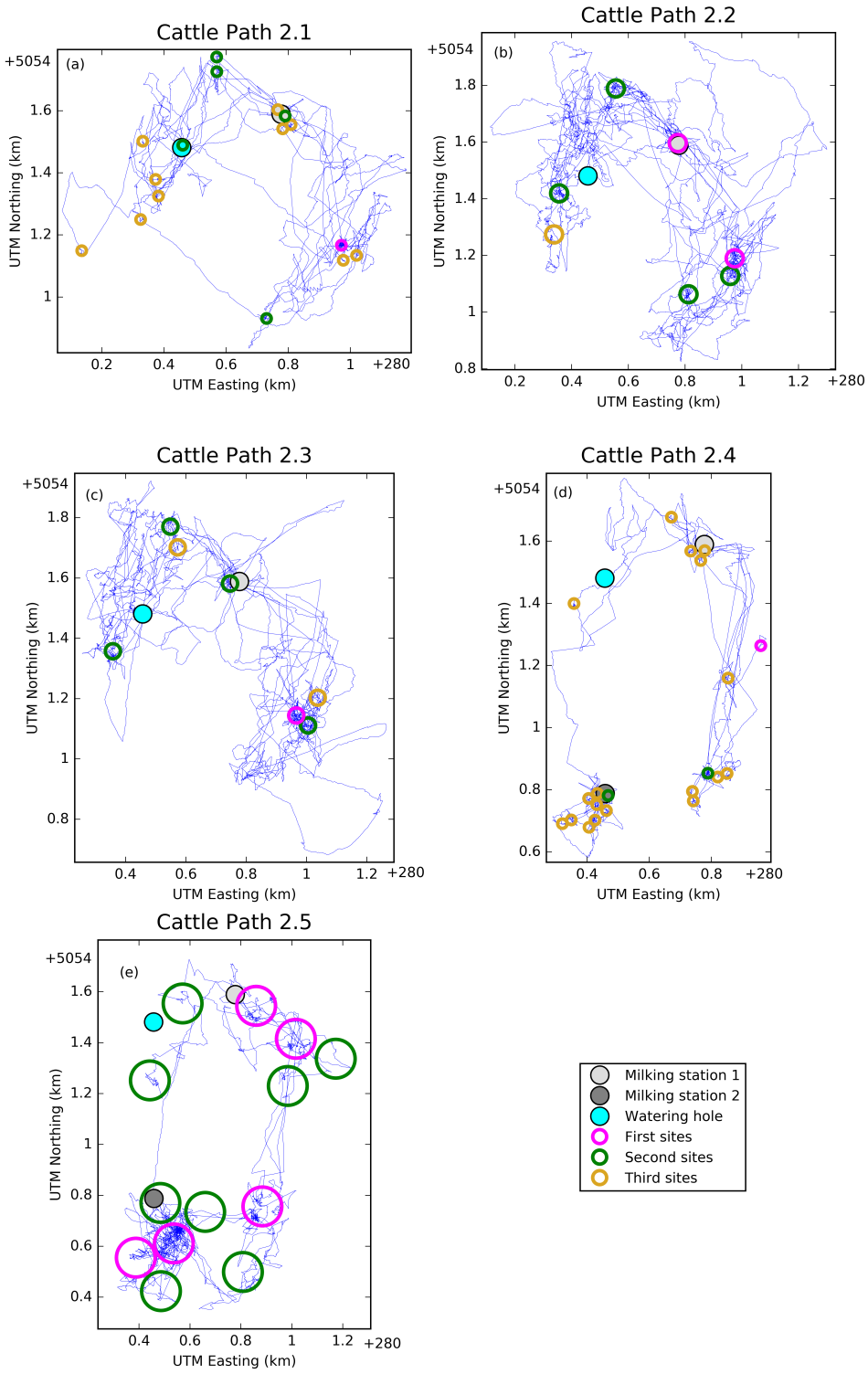

Figure 3.5: **Cattle Group 2 sites**
Sites identified for Cattle Group 2 paths 2.1-5 in panels (a)-(e). First, second and third sites are identified from the maximum, second and third highest percent drops, respectively.

§ 3.3 Goats Results

When I apply the algorithm to the goat data it is clear that there is only one site of interest. This is seen from the paths, where a lot of the goats' movements are in a constrained central area and from the usage time heat maps (e.g. Fig. 3.6 and 5.2). We see that this central area encompasses a goat pen and nearby salt blocks, which are sites known *a priori*.

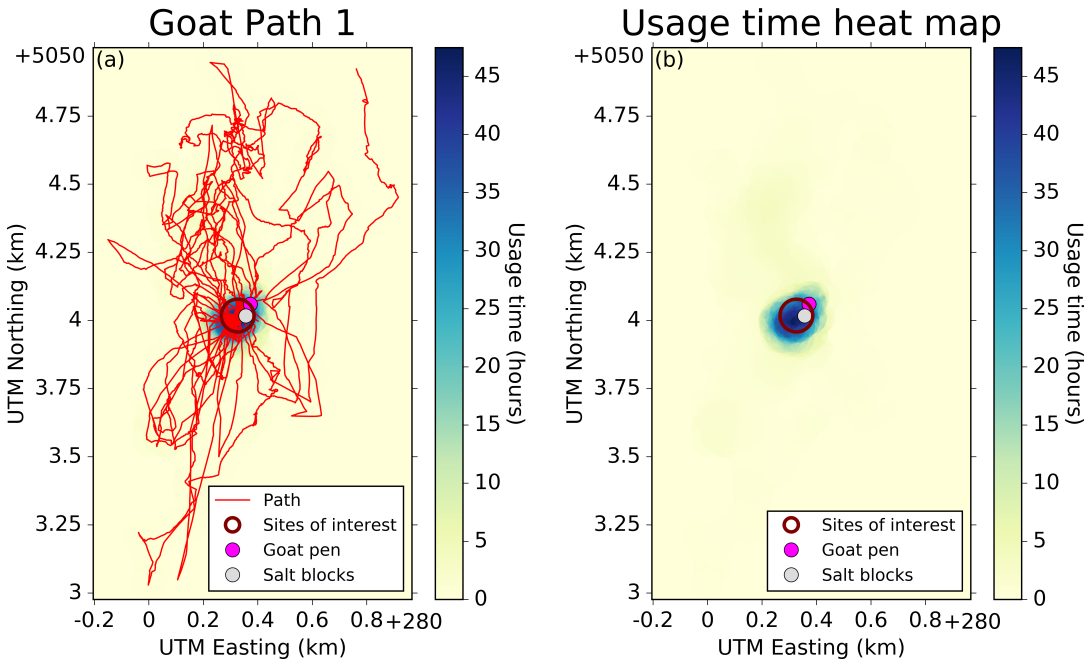

Figure 3.6: **Goat path heat map of usage times**
Heat map of usage times showing only one site for Goat Path 1 calculated using the Sites of Interest Algorithm. Panel (a) shows the heat map of usage times with the goat path overlaid, whereas panel (b) just shows the heat map.

I find from the Sites of Interest Algorithm that almost always only one site is identified (for 72% of cases across all radius values and goat paths only one site is identified). There is almost always a very distinct drop after the first bar in the histogram of usage times (Fig. 3.7a.2-3). In Figure 3.7 the same area encompassing the goat pen and salt blocks is always identified as a site of interest no matter which radius value is used. In the cases when more than one site is identified then these sites are clustered together around the central area (Fig. 3.7b.1). As such, I choose the radius value that results in only one site being identified and of these I choose the one with the largest maximum percent drop (i.e. R_i such that $n_{R_i} = 1$ and $MPD_{R_i} > MPD_{R_j}$ where $n_{R_j} = 1$).

Similar to the results from Cattle Group 2, where there are extreme high use (EHU) sites, I test to see whether the single central area site identified for each goat path is an EHU site that is obscuring other sites from being identified. When I look at the sites

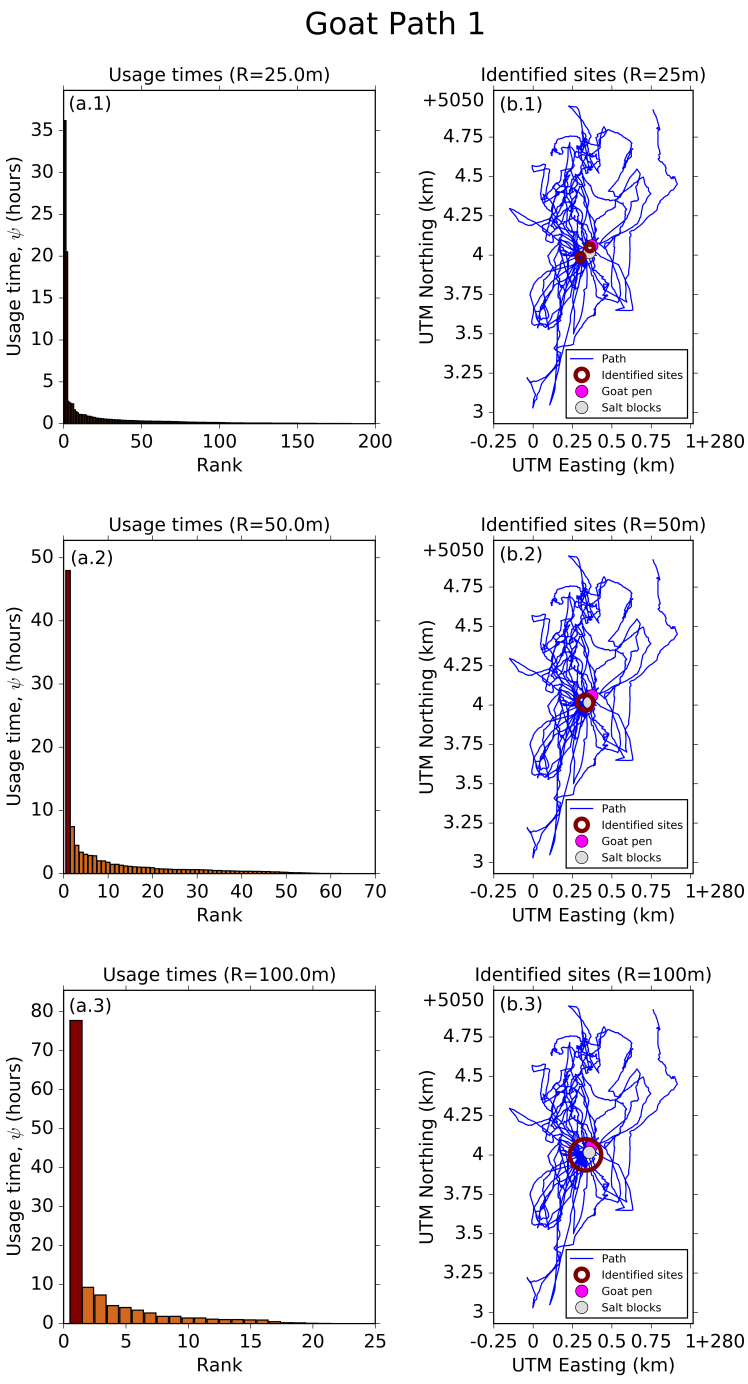

Figure 3.7: **Goat's identified site for various radius values**
Results from various radius values identify the same area. Panels (a.1-3) show the usage times. Panels (b.1-3) show the locations of the identified sites of interest. Panels (a,b.1), (a,b.2) and (a,b.3) correspond to using radius values of 25, 50 and 100m, respectively.

identified by the second highest percent drop I see that almost all of the path is covered by sites (Fig. 3.8 and 5.3), which suggests that there are no areas of particularly high usage time other than the central area.

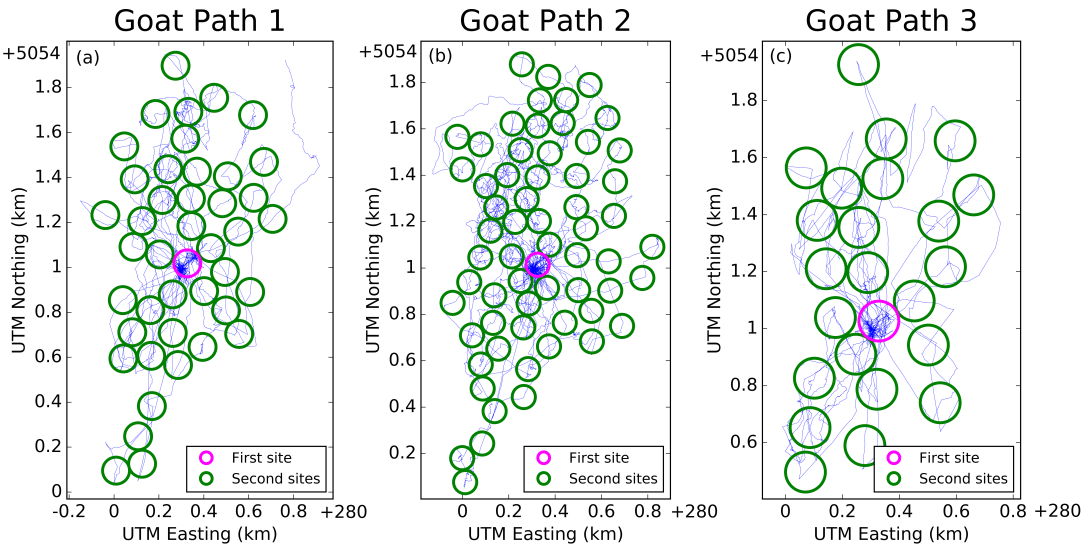

Figure 3.8: **Identified goat sites**
Sites identified from the maximum (pink circle) and second highest percent drop (green circles) for Goat Paths 1-3 in panels (a-c), respectively.

§ 3.4 Individual or Group

It is possible to apply the Sites of Interest Algorithm to individual animal paths separately or together as one group (see Section 2.1.5 for details). The former can be used to identify individual variation between the members of a group, whereas the latter reveals the shared sites of interest. Here I use the cattle paths from Group 1 to demonstrate the differences between applying the algorithm to the individual paths and all of them together.

From the plot of all the cattle paths together we can see the individual variation between the cattle, for example there is one cow which heads North-East (Fig. 3.9a) where none of the other cows venture. However, differences may not always be so clear and especially if an area is a site for one individual but another individual only passes through that area rather than spending a significant amount of time there. For example, for Cattle Paths 1.4 and 1.6 the watering hole is not identified as a site of interest but they both pass by it (Fig. 3.9c and d).

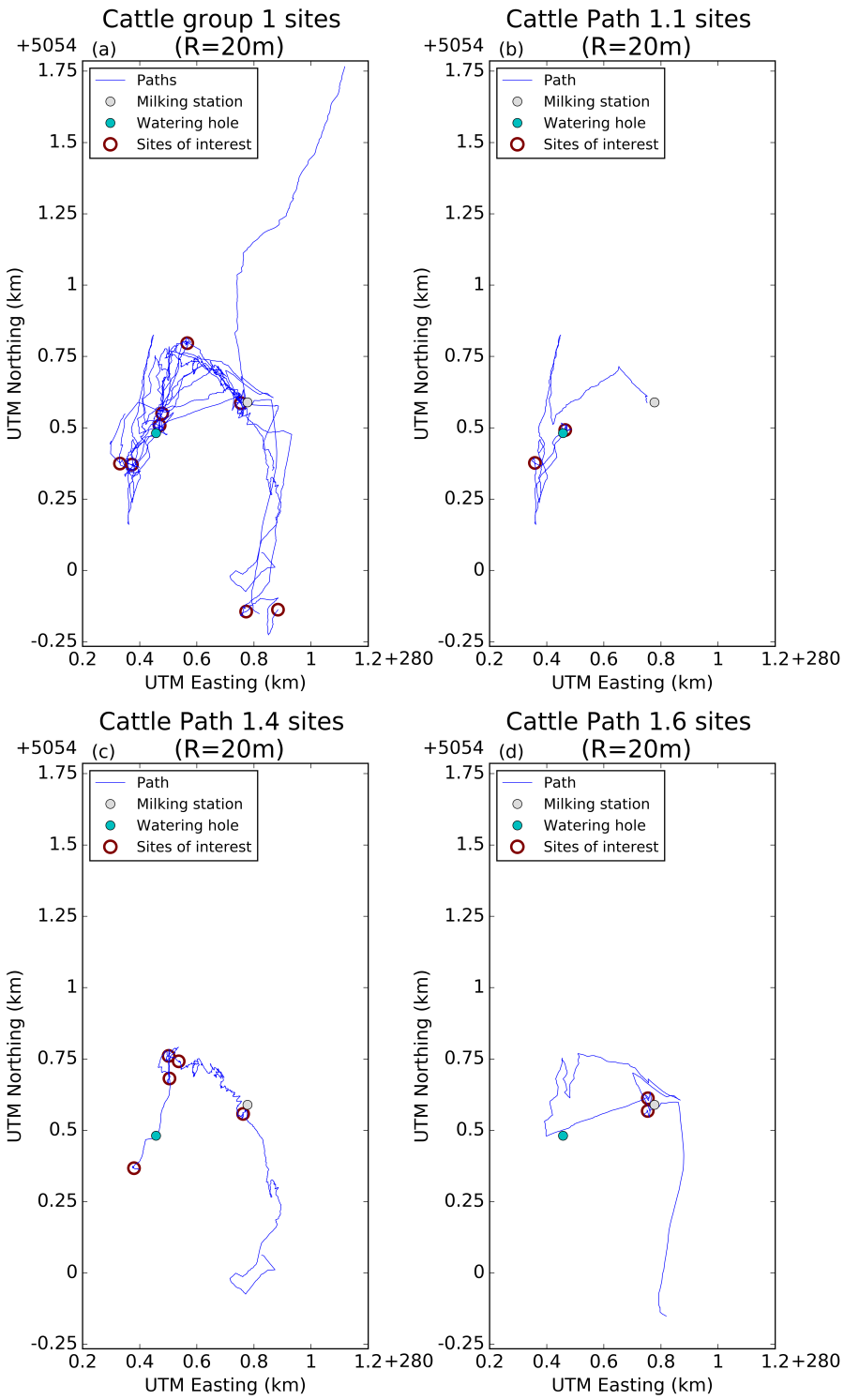

Figure 3.9: **Comparison of results from an individual and a group** Sites identified from all cattle paths (a) and individual Cattle Paths 1.1 (b), 1.4 (c) and 1.6 (d). Paths are given in blue and the sites identified using the Sites of Interest algorithm are represented by brown circles.

§ 3.5 Discontinuous Data

I have shown how an extension to the Sites of Interest Algorithm enables it be applied to groups of paths as well as to individual paths (see Section 3.4). The same extension to apply the algorithm to group data can be used on discontinuous data. It may be the case that a path is discontinuous because of a failure in the tracking method, or in certain situations it may be necessary to remove certain segments of the trajectory, thus making the data discontinuous. For example, for diurnal or nocturnal animals it may be helpful to remove time periods corresponding to night-time or daytime, respectively, since these may be less interesting time periods to study.

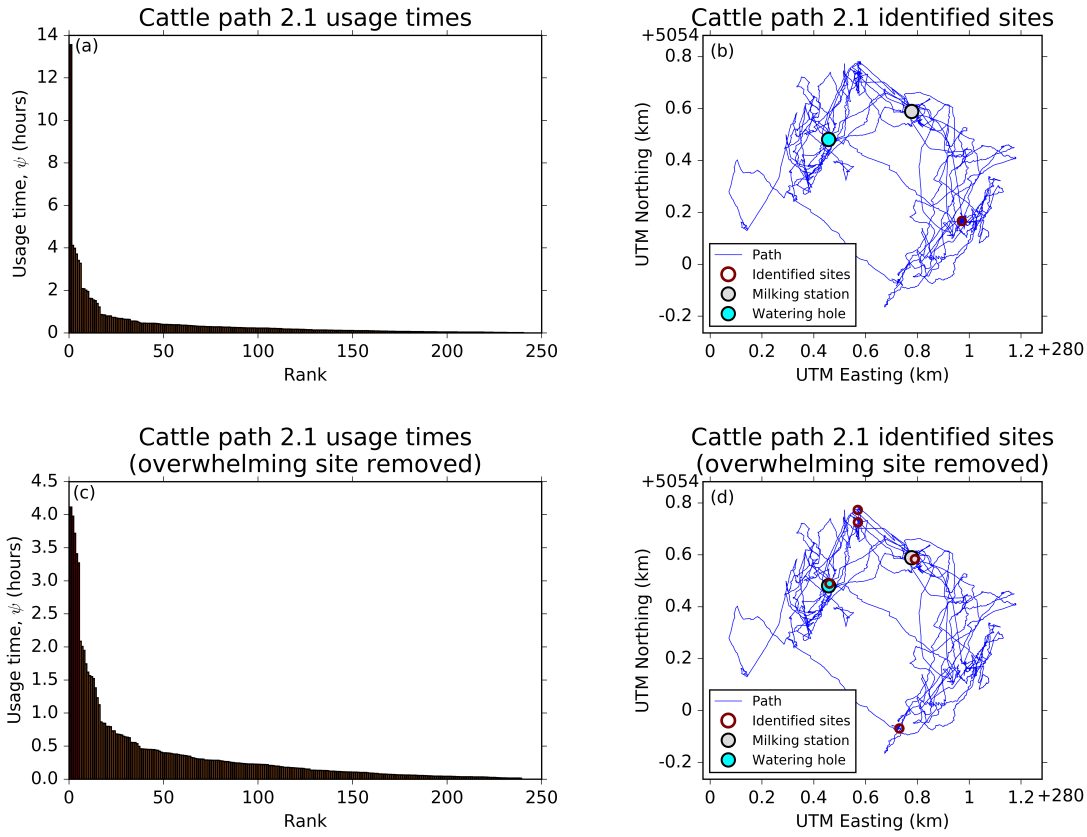

Figure 3.10: **Extreme high-use site example**
Results of applying the Sites of Interest Algorithm to a full cattle trajectory (panels a and b) and the same trajectory with the extreme high-use (EHU) site removed (panels c and d). The EHU site is the only site identified from applying the algorithm to the full trajectory.

For several of the paths from Cattle Group 2 only one site of interest is identified (see Section 3.2), but from examining usage time heat maps (Fig. 3.4) we see that in these cases there is one EHU site of interest, which is used for sleeping and is obscuring

other sites from being identified. Depending on what the EHU site is used for these may correspond to time periods of uninteresting movement, such as sleeping sites. Sites other than EHU sites may be more interesting and they can be identified using the second or third highest percent drop rather than the maximum, or alternatively by removing the path segments within the EHU site. As an example of applying the algorithm to discontinuous data I apply the algorithm to Cattle Path 2.1 after removing all of the path segments within the EHU site.

Figure 3.10 shows the difference in applying the algorithm to the full trajectory (panels a and b) compared to the trajectory with movements in the EHU site removed (panels c and d). We see that either identifying the second highest percent drop or removing movements within the EHU site result in known sites of interest being identified, such as the watering hole and milking station.

§ 3.6 Algorithm Comparison

The Sites of Interest Algorithm is adapted from the method of Barraquand and Benhamou (2008) for use on high frequency data. I compare the runtime and accuracy of applying the Sites of Interest Algorithm (referred to as the "Adapted Algorithm") to the method of Barraquand and Benhamou (2008) (referred to as the "Original Algorithm"), which is similar to the comparison made in Section 2.5. I also compare the results from varying the crossing-interval length, denoted s, which is the size of the moving interval over which entrances into and exits from the circle are checked for (see Section 2.1.1 for more details).

3.6.1 Runtime

Table 3.1 shows how the Sites of Interest Algorithm vastly reduces the runtime by two orders of magnitude. The Adapted Algorithm takes seconds to complete, whereas the Original Algorithm takes hours. This is similar to the results from the simulated data (Section 2.5.1). Even when the crossing interval length, s, is one, the runtime is greatly reduced and just as accurate as using the Original Algorithm. As expected the runtime of using $s = 10$ is roughly a tenth of the runtime when $s = 1$.

3.6.2 Accuracy

I compare the accuracy of using $s = 10$ against $s = 1$, where s is the length of the crossing interval, which is used to search for movements across the circle's boundary. Therefore s is the largest time length of a passage that might be missed, so when $s = 1$ the results are completely accurate up to the resolution of the data.

I find that for 88.6% of cases the number of circles, $N^{(c)}$, is unchanged by using $s = 10$ as opposed to $s = 1$. Of these cases where the number of circles is the same 88.7% of them have the exact same circle positions (i.e. $\{\mathbf{x}_1^{(c)}, \ldots, \mathbf{x}_{N^{(c)}}^{(c)}\}$). I also find that the average usage time error per circle is 0.24 seconds, which is less than the time difference between relocations. Finally, I find that in all cases the same number of sites is identified no matter whether s is 10 or 1. Hence using $s = 10$ is a suitable choice, since it enables the algorithm to complete within seconds while having a minimal effect

on the algorithm's accuracy.

	Path 1.1	Path 1.2	Path 1.3	Path 1.4	Path 1.5	Path 1.6	Path 1.7	Average
No. points	36913	39601	35964	36898	30605	31504	35736	35317.3
Original algorithm	11:22:31	12:56:32	13:51:53	11:08:42	08:29:20	07:11:02	13:20:59	11:11:34
Average ($s = 1$)	00:03:51	00:04:58	00:04:11	00:06:07	00:05:23	00:05:52	00:04:44	00:05:00
Average ($s = 10$)	00:00:39	00:00:31	00:00:26	00:00:39	00:00:34	00:00:37	00:00:18	00:00:32

Table 3.1: **Algorithm runtimes comparison**
Runtimes (given as hours:minutes:seconds) for the cattle paths from group 1 using the Original Algorithm of Barraquand and Benhamou (2008) and the Sites of Interest Algorithm (referred to as the Adapted Algorithm) with different values for s (1 and 10), which are averaged across all radius values used. Full details are given in Tables 5.10 and 5.11.

§ 3.7 Identifying Periodic Behaviour

There exists several tools and techniques which may be used to identify periodic patterns in an animal's movement routine, such as conditional entropy (Riotte-Lambert et al., 2013). The first step in the method described in (Riotte-Lambert et al., 2013) is to define areas in which to check for periodicity, so we see how the Sites of Interest Algorithm can be used to define such areas. Since the Sites of Interest Algorithm results in clearly delimited sites it is easy to define when an animal is inside or outside a site. The second step is to identify entrance and exit times into and from the area, which are part of the output from the Sites of Interest Algorithm, referred to as crossing times and described in Section 2.1.1. Thirdly either wavelet analysis or Fast Fourier transforms are used to detect periodicities in the three time series of presence/absence, entrance times or exit times. Finally, any detected periodicities are tested for significance against the null hypothesis that there is no periodic pattern.

A simple method for identifying periodicity involves creating histograms showing the distribution of when an animal spends time in a site of interest (Fig. 3.11). For the goats we identify a bimodal pattern, where the time periods of high use are over night (8pm-4am roughly) and around midday. This visualisation is most useful when there is only one site of interest as is the case with the goat data (Section 3.3). However, if there are multiple sites then histograms can be created for each site. Identifying when a site is most used can help in determining what the site is used for. For example, in this case I infer that the goats' site of interest is used for resting. In particular, I assume that over night the goats are sleeping and around midday they are staying in

the shade of the goat pen and other nearby buildings, since it is the warmest part of the day.

We see in Figure 3.12 how the Sites of Interest Algorithm leads to a simplification of an animal's movement path, where only the key broad-scale movement decisions are highlighted. We can consider these movements from site to site as indicating distinct movement decisions. This simplification can be used to identify periodic patterns in an

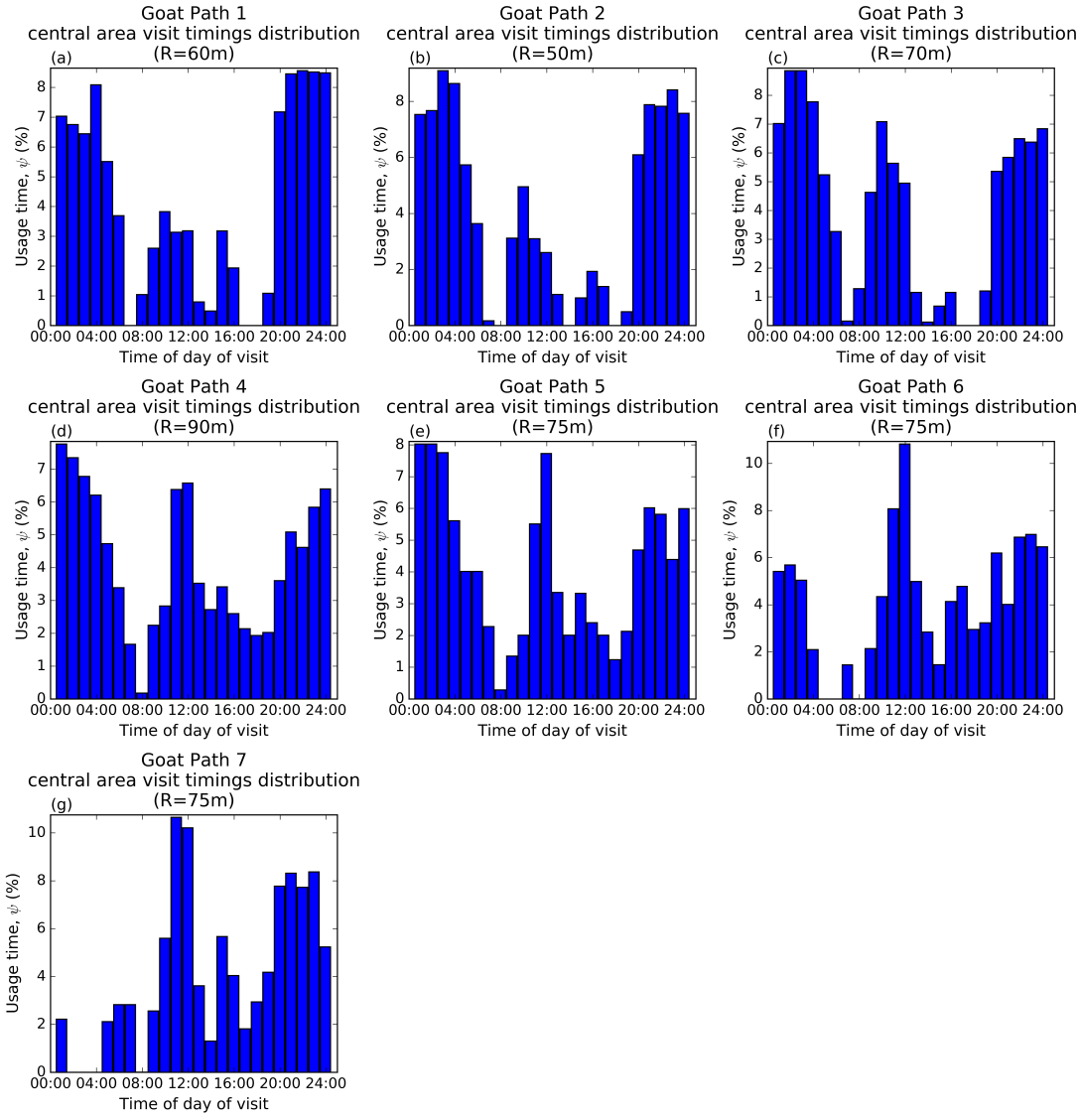

Figure 3.11: **Central area visit distribution**
Histograms of the time spent in the central area for each goat path, where the usage time is given as a proportion of the study duration (roughly 1 week). Panels (a)-(g) show the histograms for Goat Paths 1-7, respectively.

animal's movement routine and to infer the reasons for an animal to transition from one site to another using methods such as Step Selection Analysis (SSA; Fortin et al. (2005); Rhodes et al. (2005)).

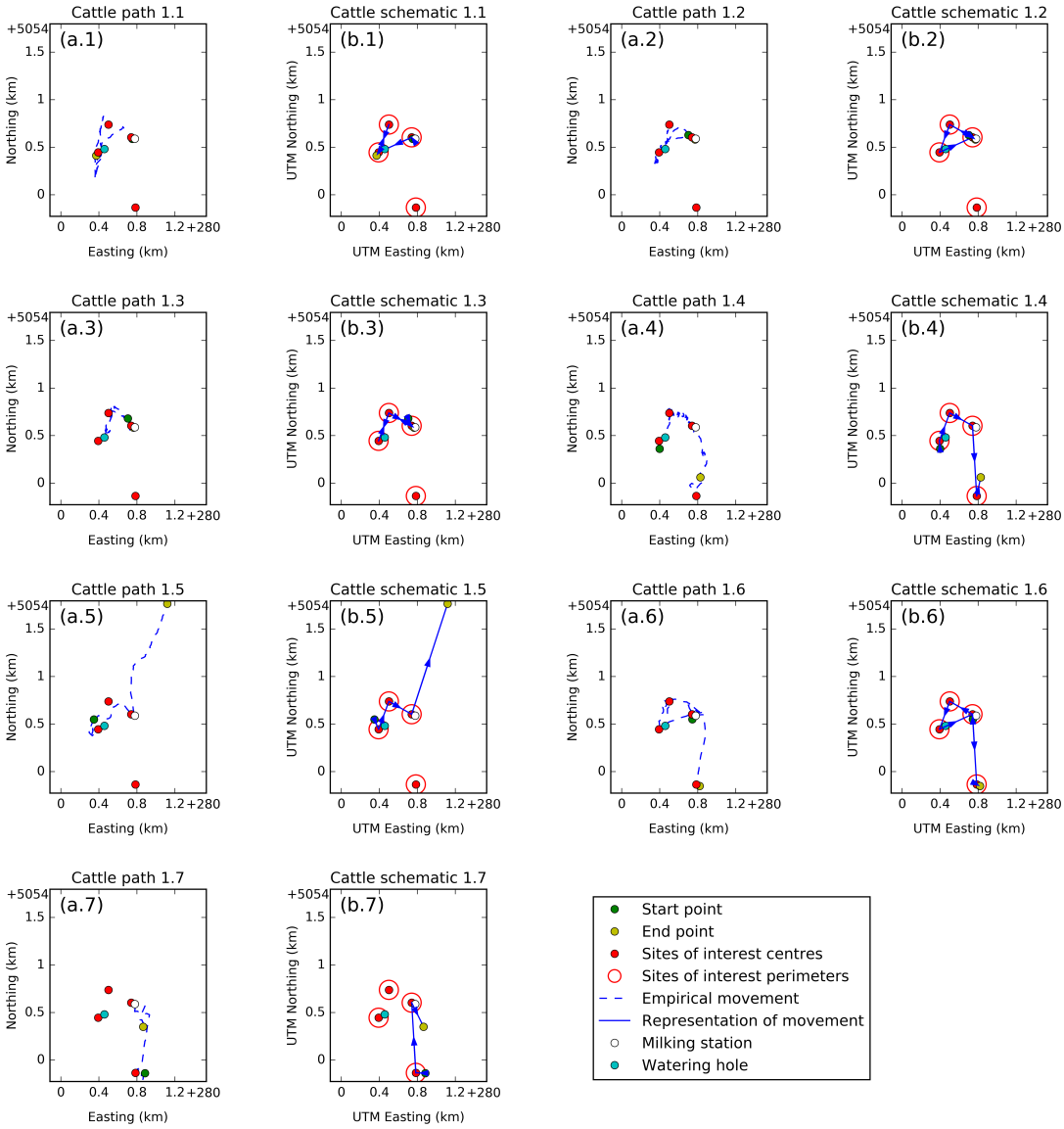

Figure 3.12: **Cattle path schematic representations**
The i^{th} original cattle path from group 1 (blue dashed line) is presented in panel (a.i) and its corresponding schematic representation (solid blue line) is presented in panel (b.i). Sites of interest are in red. Arrows indicate the order that sites are visited in from the starting point (green point) to the end point (yellow point). The milking station (grey point) and watering hole (light blue point) are also included.

SSA enables the inference of the motivations underlying an animal's habitat selection decisions, so we can consider an animal's movements where they transition from one site to another as indications that they have made a habitat selection decision. These transitions between sites can be used as the steps in SSA, however usually in SSA studies the steps are defined as the movements between low frequency relocation data, which are evenly spaced in time, whereas transitions between sites are not. Hence this necessitates an adaptation to accommodate for data which is unevenly spaced temporally, which I describe in Chapter 4,

§ 3.8 Discussion

When applying the Sites of Interest Algorithm to empirical data I find that the run-time of the algorithm is far less compared to the method of Barraquand and Benhamou (2008) (Table 3.1) without compromising on accuracy (see Section 3.6.2). The Sites of Interest Algorithm is able to identify known sites, such as the watering hole, milking station locations, goat pen and salt blocks, as well as to reveal other less expected sites (e.g. sites C2 and D2 in Fig. 3.2d), which allows us to further investigate why these areas are of interest by examining environmental data and intra-site movements. Therefore the Sites of Interest Algorithm would be most useful when these areas of interest are unknown and environmental information is either unavailable or too coarse to identify these areas of interest

In identifying known sites of interest, I have demonstrated several post-processing methods to identify the number of sites and choose a suitable radius value. Instead of identifying the sites from the maximum percent drop it may be helpful to look at the second or third highest percent drop, which may identify other sites when there is one extreme high-use (EHU) site (e.g. Cattle Group 2; see Section 3.2). Also, a researcher can look at the reasons behind the single EHU site and perhaps use these to remove data that is not of interest for their purposes (e.g. when the animal is sleeping, locations are not of interest for understanding movement decisions). When the number of sites is known, as in the case of the goat paths where there is clearly only one site (see Section 3.3), then a suitable radius can be chosen such that it has the highest maximum percent drop of those radius values that identify the correct number of sites.

Applying the algorithm to the data on goat movements reveals only one site of interest for each goat, which leads me to apply Step Selection Analysis (SSA; Fortin et al. (2005); Rhodes et al. (2005)) to the goats' movements, using the identified site of interest as an area the goats are attracted to (see Section 4.2). SSA enables me to test hypotheses concerning the goats attraction the site of interest in the central area, which encompasses the goat pen and nearby salt blocks. More specifically I test whether the goats' strength of attraction is proportional to their distance from the central area and if they tend to turn more or less often when moving toward the central area than when they are moving away from it. I am able to use the centre of the identified site of interest as the point of attraction representing the central area. More details of how I apply SSA to the goat data are presented in Section 4.2.

Extending the Sites of Interest Algorithm for use with groups of individuals allows a

user to answer different questions, such as identifying shared sites of interest or individual variation. It may be interesting to examine if there is any particular reason for the individual variation. For example, the algorithm might be applied across different subgroups to identify if there are any sites which are particular to one subgroup, where these subgroups are unified under one particular trait, such as gender or age.

The Sites of Interest Algorithm identifies sites which are clearly delimited from the rest of the environment, which enables a quick segmentation of a movement path into bouts of intra and inter-site movement. These movements can be analysed in further detail to identify the type of movement that is being exhibited, which is in contrast to other methods, such as state-space modelling (Patterson et al., 2008) and Hidden Markov Models (Morales et al., 2004), where sites are identified from the movement behaviour. Therefore the Sites of Interest Algorithm does not rest on being able to define and identify different behaviours, thus allowing it to identify sites that may be used for a wide variety of reasons.

§ 3.9 Glossary

Symbol	Explanation
$\mathbf{x}_i$	recorded location of the animal at time, t_i
t_i	sequence of times at which the animal's location is recorded
N	number of recorded locations
$\mathbf{x}_i^{(c)}$	centre of the circle centred at time point, $t_i^{(c)}$
$t_i^{(c)}$	sequence of times indicating where the circles are centred
$N^{(c)}$	number of circles
R	circle radius
$d(\mathbf{x}, \mathbf{y})$	Euclidean metric distance between $\mathbf{x}$ and $\mathbf{y}$
s	length of the crossing interval
F, B	sets of forward and backward passage crossings, respectively
ψ_k, β_k	number of forward and backward passages for the k^{th} circle, respectively
φ_k	usage time of the k^{th} circle
Φ	set of usage times
$\varphi_k^{(o)}$	usage time of the k^{th} ordered circle
$\Phi^{(o)}$	ordered set of usage times
φ_k'	usage time of the k^{th} ordered and reduced circle
Φ'	ordered and reduced set of usage times
n_R	number of sites identified when a radius of R is used
R_s, R_t	radii identified from using the stability and threshold criteria, respectively
MPD_R	maximum percent drop resulting from using a radius of R
T_{MPD}	maximum percent drop threshold value

Table 3.2: **Chapter 3 Glossary**
Glossary of symbols used in Chapter 3

Chapter 4

Step Selection Analysis for High Frequency Data

Step Selection Analysis (Fortin et al. (2005); Rhodes et al. (2005); SSA) is a key technique used to infer the behavioural motivations underlying an animal's decisions to move. Avgar et al. (2016) generalised the SSA method, termed integrated-SSA (iSSA), which enables the inference of how an animal's movement and habitat selection decisions are affected by environmental factors. However, up to now SSA studies have been constrained by the data that they typically use. Animal relocations are usually recorded at low fixed rates (e.g. from every 15 minutes (Dickson et al., 2005; Gillies et al., 2011) to daily recordings (Richard and Armstrong, 2010)). Typically in SSA studies inference is made about the movements between consecutive relocations in low frequency data, but it is unlikely that these recordings correspond to the times when the animal makes a distinct decision to move. The accelero-magnetometer data in this thesis, however, are collected at such high frequencies (e.g. $\geq$ 1 Hz) that it is highly unlikely for an animal to make distinct movement decisions at each recorded location. The problem is to identify the points within such high resolution data when actual movement decisions are made by the animal.

One method for identifying distinct behaviourally meaningful movement decisions from high frequency data is to use the Turning Points Algorithm (Potts et al., 2018), which detects turning points as changes in an animal's heading. Turning points can be considered to correspond with behaviourally meaningful movement decisions since turns are energetically costly, so would not be made unless it were beneficial to the animal (Wilson et al., 2013*b*). Identifying turning points enables us to use the movements between successive turning points as the steps in iSSA. However, the time intervals between consecutive turning points, termed *step-times*, are unlikely to be the same for all steps. In iSSA all steps are assumed to have the same step-time since iSSA was designed for data that is evenly spaced in time. In this chapter I describe a generalisation of iSSA, termed time-varying iSSA (tiSSA), designed to accommodate for these data with variations in step-time, which I term *non-constant step-time data*.

I also give an illustrative example of applying tiSSA to empirical data.

§ 4.1 Time-varying Integrated Step Selection Analysis

I start by defining the probability density of the animal making a turn at $\mathbf{y}$ after a time period of τ since it made its last turn at $\mathbf{x}$. This takes the following form

$$f(\mathbf{y}, \tau|\mathbf{x}, \alpha_{\mathbf{x}}) = C^{-1}\varphi(\mathbf{y}, \tau|\mathbf{x})W(\mathbf{y}, \tau, \mathbf{x}, \alpha_{\mathbf{x}}) \tag{4.1}$$

where $\alpha_{\mathbf{x}}$ is the heading upon which the animal arrived at $\mathbf{x}$, $\varphi(\mathbf{y}, \tau|\mathbf{x})$ is the *sampling kernel*, $W(\mathbf{y}, \tau, \mathbf{x}, \alpha_{\mathbf{x}})$ is the *weighting function* and C is the *normalising constant*, which is defined as $\int_\Omega \int_0^\infty \varphi(\mathbf{z}, \tau|\mathbf{x})W(\mathbf{z}, \tau, \mathbf{x}, \alpha_{\mathbf{x}})\mathrm{d}\tau\mathrm{d}\mathbf{z}$. The normalising constant ensures that $f(\mathbf{y}, \tau|\mathbf{x}, \alpha_{\mathbf{x}})$ integrates to one and is therefore a probability density function (PDF).

Usually in iSSA studies the sampling kernel is a distribution of step-lengths and headings. However, for data with non-constant step-times it is also necessary to incorporate step-times into the sampling kernel. Hence I define the sampling kernel as the product of probability density functions (PDFs) for each of the heading, step-time and step-length dependent on the step-time. Therefore the sampling kernel has the following form

$$\varphi(\mathbf{y}, \tau|\mathbf{x}) = h(\alpha)g_1(\tau)g_2(l|\tau) \tag{4.2}$$

where $h(\alpha), g_1(\tau)$ and $g_2(l|\tau)$ are PDFs of the heading, α, step-time, τ, and step-length, l, from $\mathbf{x}$ to $\mathbf{y}$, respectively. The step-length PDF, $g_2(l|\tau)$, is dependent on the step-time, since an animal is able to move further (i.e. have a larger step-length) when the step-time is greater. However the heading is assumed to be independent of step-length and time.

Similar to Avgar et al. (2016), I suggest that $g_1(\tau)$ and $g_2(l|\tau)$ be PDFs of gamma distributions. Both the exponential and χ^2 distributions are special cases of the gamma distribution, which can be written as

$$P(\sigma|k, \theta) \propto \exp(-\sigma/\theta + (k-1)\ln\sigma) \tag{4.3}$$

where k and θ are the shape and scale parameters, respectively and σ represents either the step-length or step-time. I also suggest that $h(\alpha)$ be the PDF of a uniform distribution, since any potential directional bias can be tested for in the weighting function.

The weighting function is defined as

$$W(\mathbf{y}, \tau, \mathbf{x}, \alpha_{\mathbf{x}}) = \exp\left(\mathbf{B}_n\mathbf{Z}^{\mathrm{T}} + \beta_{n+1}\frac{l}{\tau} + \beta_{n+2}\tau + \beta_{n+3}\ln\left(\frac{l}{\tau}\right) + \beta_{n+4}\ln(\tau)\right) \tag{4.4}$$

where $\mathbf{Z} = [Z_1, Z_2, \dots Z_n]$ and $\mathbf{B}_n = [\beta_1, \beta_2, \dots \beta_n]$ are vectors of covariates, $Z_i(\mathbf{y}, \mathbf{x}, \alpha_{\mathbf{x}}, \tau)$, and their respective coefficients, β_i, respectively. Notice that the covariates may be dependent on the step-time, τ, which is novel to the tiSSA method. For example, later I test how the step-times of goats are affected by the terrain they are moving in and the temperature. The weighting function also includes correcting factors

for the step-speed, β_{n+1} and β_{n+3}, and the step-time, β_{n+2} and β_{n+4}. These factors are included to limit potential bias in the sampling kernel (Forester et al., 2009; Avgar et al., 2016). Both the step-speed and step-time have two correcting factors each since I use gamma distributions for $g_1(\tau)$ and $g_2(l|\tau)$, which have two parameters each (θ and k in Eq. 4.3). Writing the gamma PDF in the way of Equation 4.3 shows the necessity of having two correcting factors, one for σ and the other for $\ln \sigma$.

§ 4.2 Application to Fine-Scale Movements of Goats

4.2.1 Goat Data

Note this is the same data set as used in Chapter 3. I summarise the details of how the data was collected in Section 3.1. From inspecting the goat paths I make several observations concerning their movement tendencies that I use to formulate hypotheses that I test using tiSSA. The results from applying the Sites of Interest Algorithm (see Section 3.1) show that the goats spend a large proportion of their time ($\sim$40% of the study duration) in a relatively small region ($\sim$70m radius circle), which I refer to as the *central area*. The central area encompasses a goat pen and salt blocks, which explains why they are attracted to this area. I therefore test the hypothesis that they have an attraction to this central area and that the strength of this attraction is proportional to their distance from the central area. I also test whether they tend to make smaller angle turns rather than larger (i.e. whether the bearings of successive steps are correlated), since turning is energetically costly (Wilson et al., 2013*b*). The goats live in a narrow steep-sided valley (Fig. 4.1), so I test whether they avoid the steeper upward slopes, since moving up these slopes requires more energy (Minetti et al., 2002; Ardigò et al., 2003).

4.2.2 Turning Points Algorithm

It does not make sense to directly apply iSSA to high frequency data, since successive relocations are so close that there is little interesting biological information contained within the movements between consecutive relocations. We need a different method for defining the steps, which involves rarefying the data such that the rarefied data points represent times when the animal has made a behaviourally meaningful movement decision.

I use the Turning Points Algorithm (Potts et al., 2018) to identify the turning points on the goat paths. These turning points represent times when the goats have made significant movement decisions, since turning is energetically costly so would not be made unless it was beneficial to them (Wilson et al., 2013*b*). The algorithm works by calculating the squared circular standard deviation (SCSD) across a moving window. Peaks in the SCSD signal are identified as candidate turning points. Any candidate turning point is discarded if the turning-angle at that point is below a given threshold. Therefore, a user of this algorithm has to make a choice of window size and threshold value. Across all paths I use a window size of 40 seconds and a threshold of 30°, since these values result in simplified paths which closely resemble the unrarefied paths (Fig. 4.2a-b). On the other hand, there is a lot of detail missing from paths constructed by

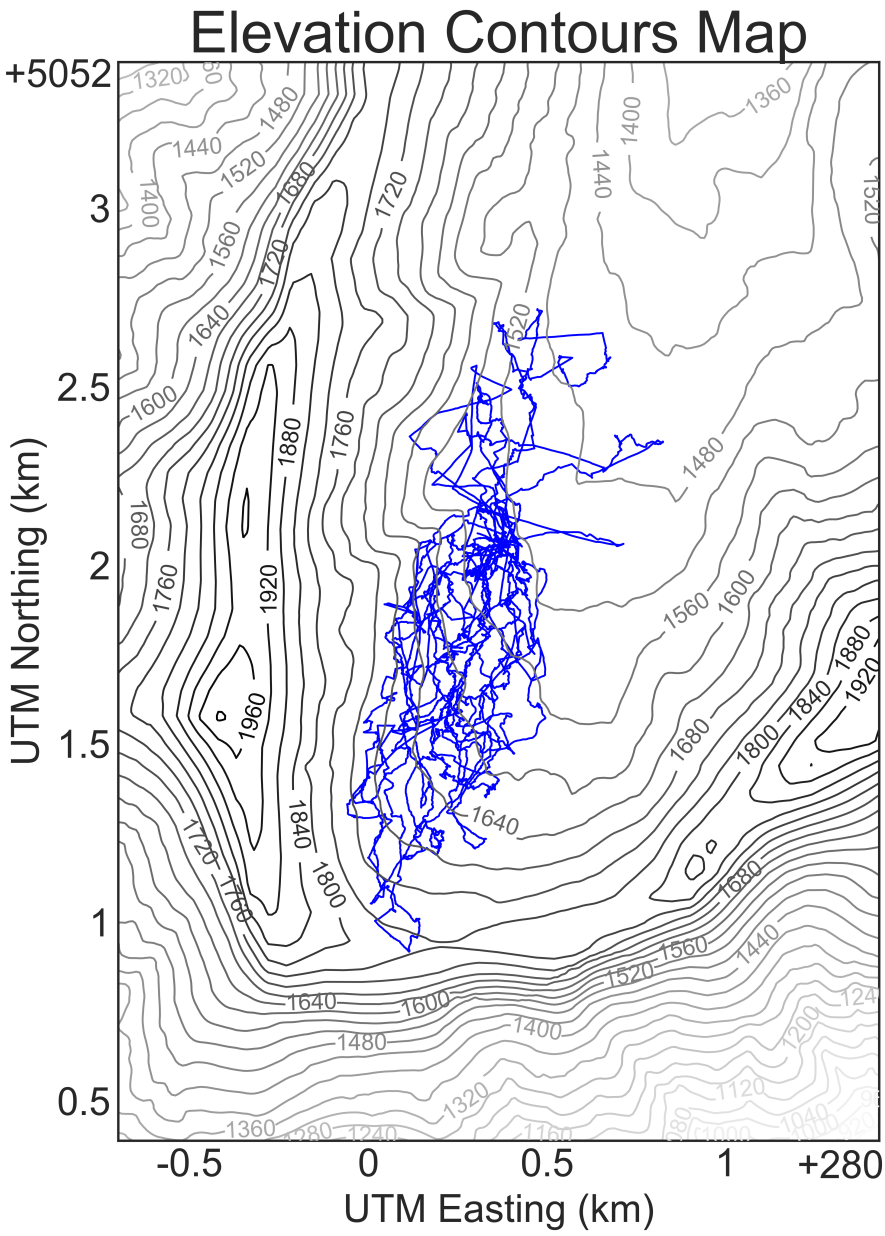

Figure 4.1: **Elevation contour map**
A contour map of elevations for the valley in which the goats live. Darker contour lines represent higher elevations and the elevation values are given in metres. Goat Path 1 is given as the blue line.

sampling at regular time intervals, mimicking those that would be obtained from lower frequency data (panels c-f).

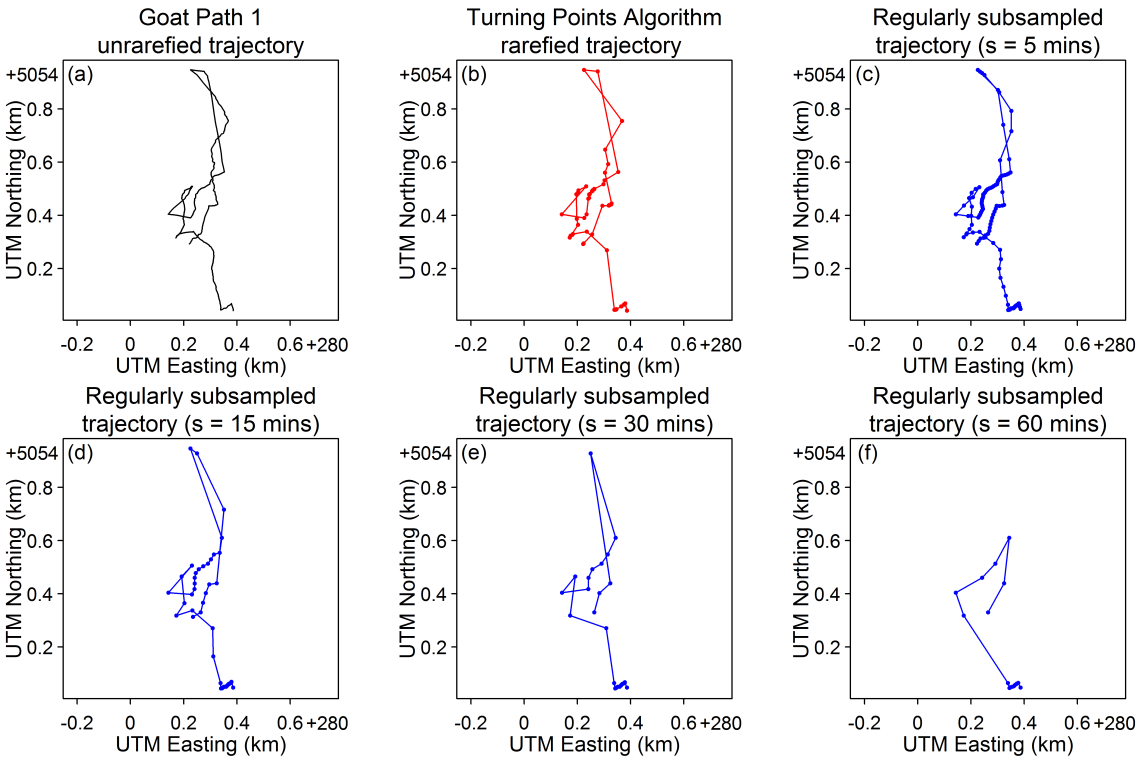

Figure 4.2: **Visual comparison of rarefied paths**
A path segment taken from Goat Path 1 (panel a), which is rarefied using the Turning Points Algorithm (panel b) and regularly subsampled (panels c-f) at rates of 5, 15, 30 and 60 minutes, respectively.

4.2.3 Model formulation

Non-Constant Step-Time Data

After applying the method of Potts et al. (2018) the goat's path is described by a sequence of turning points, $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N$ and corresponding step-times $\tau_1, \tau_2, \ldots \tau_N$, where τ_i is the length of time taken by them to move from $\mathbf{x}_{i-1}$ to $\mathbf{x}_i$, l_i is the distance between $\mathbf{x}_{i-1}$ and $\mathbf{x}_i$ and the step-speeds are defined as $s_i = l_i/\tau_i$.

I assume that the headings are uniformly distributed, so the PDF of headings is defined as

$$h^{(\mathrm{r})}(\alpha) = \frac{1}{2\pi} \tag{4.5}$$

where the superscript (r) represents any function relating to the tiSSA method.

I fit gamma distributions for the step-time and step-length dependent on step-time,

so their respective PDFs are defined as

$$g_1^{(\mathrm{r})}(\tau) = \frac{\tau^{k_1-1}\exp(-\tau/\theta_1)}{\Gamma(k_1)\theta_1^{k_1}} \quad \text{and} \quad g_2^{(\mathrm{r})}(l|\tau) = \frac{l^{k_2-1}\exp(-l/(\theta_2\tau))}{\Gamma(k_2)(\theta_2\tau)^{k_2}} \tag{4.6}$$

where k_1 and k_2 are shape values fitted for step-times and step-lengths, respectively, and θ_1 and θ_2 are scale parameters fitted to the step-times and step-lengths, respectively.

The best fit k_1 and θ_1 in $g_1^{(\mathrm{r})}(\tau)$ can be found by maximising the following likelihood function

$$L_\tau(k_1, \theta_1|\tau_1, \ldots, \tau_N) = \prod_{i=1}^{N} \frac{\tau_i^{k_1-1}\exp(-\tau_i/\theta_1)}{\Gamma(k_1)\theta_1^{k_1}}. \tag{4.7}$$

Similarly, the likelihood function used to fit parameters k_2 and θ_2 to $g_2^{(\mathrm{r})}(l|\tau)$ is

$$\begin{aligned} L_l(k_2, \theta_2|l_1, \ldots, l_N, \tau_1, \ldots, \tau_N) &= \prod_{i=1}^{N} \frac{l_i^{k_2-1}\exp(-l_i/(\theta_2\tau_i))}{\Gamma(k_2)(\theta_2\tau_i)^{k_2}} \\ &= \prod_{j=1}^{N}\frac{1}{\tau_j}\prod_{i=1}^{N} \frac{s_i^{k_2-1}\exp(-s_i/\theta_2)}{\Gamma(k_2)\theta_2^{k_2}} \end{aligned} \tag{4.8}$$

where the best fit k_2 and θ_2 given the empirical step-lengths and step-times are those which maximise L_l.

Notice that Equation 4.8 is maximised for the same parameter values as the following likelihood function for the step speeds, s_i

$$L_s(k_2, \theta_2|s_1, \ldots, s_N) = \prod_{i=1}^{N} \frac{s_i^{k_2-1}\exp(s_i/\theta_2)}{\Gamma(k_2)(\theta_2\tau_i)^{k_2}}. \tag{4.9}$$

Hence, the best fit k_2 and θ_2 for $g_2(l|\tau)$ can be found by fitting a gamma distribution to the step speeds instead (e.g. using `fitdist` in `R`).

The sampling kernel is defined as the product of $h^{(r)}(\alpha), g_1^{(r)}(\tau)$ and $g_2^{(r)}(l|\tau)$, so it has the following functional form

$$\varphi^{(\mathrm{r})}(\mathbf{y}, \tau|\mathbf{x}) = \frac{\tau^{k_1-1}l^{k_2-1}\exp(-\tau/\theta_1 - l/(\theta_2\tau))}{2\pi\Gamma(k_1)\Gamma(k_2)\theta_1^{k_1}(\theta_2\tau)^{k_2}}. \tag{4.10}$$

Based on the observations of the goats' movements (see Section 4.2.1) I formulate several hypotheses concerning their movement tendencies. I hypothesise that they tend to

(A1) move toward the central area with strength of attraction proportional to their distance from the central area,

(A2) be attracted to the central area with strength proportional to their distance from the central area and also have more directed movement when moving toward the central area,

(B) move in the same general direction,

(C) avoid moving up steeper gradients,

(D) have shorter step-times when the temperature is higher,

(E) and have shorter step-times when moving in rocky terrain.

Note that hypotheses A1 and A2 compete with each other, which explains why I label the hypotheses in this way.

I construct covariates to test each of these hypotheses, which are defined as

(A1) $Z_1(\mathbf{y}, \mathbf{x}) = |\mathbf{x}_{\text{cp}} - \mathbf{x}| \cos(\psi - H(\mathbf{x}_{\text{cp}}, \mathbf{x}))$,

(A2) $\tilde{Z}_1(\mathbf{y}, \mathbf{x}, \tau) = \tau |\mathbf{x}_{\text{cp}} - \mathbf{x}| \cos(\psi - H(\mathbf{x}_{\text{cp}}, \mathbf{x}))$,

(B) $Z_2(\mathbf{y}, \mathbf{x}, \alpha_{\mathbf{y}}) = \cos(\psi - \alpha_{\mathbf{x}})$,

(C) $Z_3(\mathbf{y}, \mathbf{x}) = (\text{Elev}(\mathbf{y}) - \text{Elev}(\mathbf{x}))/l$,

(D) $Z_4(\mathbf{y}, \mathbf{x}, \tau) = -\tau \left(\frac{\text{Temp}(\mathbf{x}) + \text{Temp}(\mathbf{y})}{2} - \overline{\text{Temp}} \right)$,

(E) $Z_5(\mathbf{y}, \tau) = -\tau I(\mathbf{y})$.

where $\mathbf{x}_{cp}$ represents the central area and is defined as the centre of the single site of interest, $H(\mathbf{y}, \mathbf{x})$ is the heading from $\mathbf{x}$ to $\mathbf{y}$, $\text{Elev}(\mathbf{x})$ is the elevation at $\mathbf{x}$, $\text{Temp}(\mathbf{x})$ is the temperature recorded when the goat was at $\mathbf{x}$, $\overline{\text{Temp}}$ is the average temperature across all steps and $I(\mathbf{x})$ is an indicator function, equalling 1 when $\mathbf{x}$ is in rocky terrain and 0 otherwise.

From this set of covariates I build four separate models, which include different covariates in the weighting function, but have the same sampling kernel (Eqn. 4.10). The first model is the *Base Model*, which includes covariates Z_1, Z_2 and Z_3. By adapting the Base Model I build the other models, where the *Returns Model* replaces Z_1 with $\tilde{Z}_1$, the *Temperature Model* adds Z_4 and the *Rocky Model* adds Z_5. The movement kernels for the four models are given as

$$f_1^{(r)}(\mathbf{y}, \tau | \mathbf{x}, \alpha_{\mathbf{x}}) \propto \frac{1}{\tau} \exp\left(\beta_1^{(r)} Z_1(\mathbf{y}, \mathbf{x}) + \beta_2^{(r)} Z_2(\mathbf{y}, \mathbf{x}, \alpha_{\mathbf{x}}) + \beta_3^{(r)} Z_3(\mathbf{y}, \mathbf{x}) + \left(\beta_6^{(r)} - \frac{1}{\theta_1} \right) \tau + \left(\beta_7^{(r)} - \frac{1}{\theta_2} \right) \frac{|\mathbf{y} - \mathbf{x}|}{\tau} + (\beta_8^{(r)} + k_1 - 1) \ln \tau + (\beta_9^{(r)} + k_2 - 1) \ln \left(\frac{|\mathbf{y} - \mathbf{x}|}{\tau} \right) \right) \quad (4.11)$$

$$f_2^{(r)}(\mathbf{y},\tau|\mathbf{x},\alpha_{\mathbf{x}}) \propto \frac{1}{\tau}\exp\left(\tilde{\beta}_1^{(r)}\tilde{Z}_1(\mathbf{y},\mathbf{x},\tau)+\beta_2^{(r)}Z_2(\mathbf{y},\mathbf{x},\alpha_{\mathbf{x}})+\beta_3^{(r)}Z_3(\mathbf{y},\mathbf{x})+\left(\beta_6^{(r)}-\frac{1}{\theta_1}\right)\tau+\left(\beta_7^{(r)}-\frac{1}{\theta_2}\right)\frac{|\mathbf{y}-\mathbf{x}|}{\tau}+(\beta_8^{(r)}+k_1-1)\ln\tau+(\beta_9^{(r)}+k_2-1)\ln\left(\frac{|\mathbf{y}-\mathbf{x}|}{\tau}\right)\right) \quad (4.12)$$

$$f_3^{(r)}(\mathbf{y},\tau|\mathbf{x},\alpha_{\mathbf{x}}) \propto \frac{1}{\tau}\exp\left(\beta_1^{(r)}Z_1(\mathbf{y},\mathbf{x})+\beta_2^{(r)}Z_2(\mathbf{y},\mathbf{x},\alpha_{\mathbf{x}})+\beta_3^{(r)}Z_3(\mathbf{y},\mathbf{x})+\beta_4^{(r)}Z_4(\mathbf{y},\mathbf{x},\tau)+\left(\beta_6^{(r)}-\frac{1}{\theta_1}\right)\tau+\left(\beta_7^{(r)}-\frac{1}{\theta_2}\right)\frac{|\mathbf{y}-\mathbf{x}|}{\tau}+(\beta_8^{(r)}+k_1-1)\ln\tau+(\beta_9^{(r)}+k_2-1)\ln\left(\frac{|\mathbf{y}-\mathbf{x}|}{\tau}\right)\right) \quad (4.13)$$

$$f_4^{(r)}(\mathbf{y},\tau|\mathbf{x},\alpha_{\mathbf{x}}) \propto \frac{1}{\tau}\exp\left(\beta_1^{(r)}Z_1(\mathbf{y},\mathbf{x})+\beta_2^{(r)}Z_2(\mathbf{y},\mathbf{x},\alpha_{\mathbf{x}})+\beta_3^{(r)}Z_3(\mathbf{y},\mathbf{x})+\beta_5^{(r)}Z_5(\mathbf{y},\tau)+\left(\beta_6^{(r)}-\frac{1}{\theta_1}\right)\tau+\left(\beta_7^{(r)}-\frac{1}{\theta_2}\right)\frac{|\mathbf{y}-\mathbf{x}|}{\tau}+(\beta_8^{(r)}+k_1-1)\ln\tau+(\beta_9^{(r)}+k_2-1)\ln\left(\frac{|\mathbf{y}-\mathbf{x}|}{\tau}\right)\right) \quad (4.14)$$

where $f_1^{(r)}, f_2^{(r)}, f_3^{(r)}$ and $f_4^{(r)}$ are the movement kernels from the Base, Returns, Temperature and Rocky Models, respectively.

Deriving β_i-values using conditional logistic regression

The β_i-values in the movement kernels for each model given the empirical data, $\mathbf{X}$, can be found by maximising the following likelihood function

$$\begin{aligned} L_I(\mathbf{B}|\mathbf{X}) &= \prod_{i=2}^{N} f(\mathbf{x}_i,\tau_i|\mathbf{x}_{i-1}) \\ &= \prod_{i=2}^{N} \frac{\varphi(\mathbf{x}_i,\tau_i|\mathbf{x}_{i-1})W(\mathbf{x}_i,\tau_i,\mathbf{x}_{i-1},\alpha_{\mathbf{x}_{i-1}})}{\int_\Omega\int_0^\infty \varphi(\mathbf{z},\tau|\mathbf{x}_{i-1})W(\mathbf{z},\tau,\mathbf{x}_{i-1},\alpha_{\mathbf{x}_{i-1}})\mathrm{d}\mathbf{z}\mathrm{d}\tau}. \end{aligned} \quad (4.15)$$

where $\alpha_{\mathbf{x}_i}$ is the heading from $\mathbf{x}_{i-1}$ to $\mathbf{x}_i$, therefore the indexing in the product starts at two since there is no $\alpha_{\mathbf{x}_0}$ or $\alpha_{\mathbf{x}_1}$.

Ideally the normalising constant may be calculated analytically (Matthiopoulos, 2003), but this can be computationally intensive and technically demanding. In contrast, I will show that it is possible to use Conditional Logistic Regression (CLR) to approximate the $\mathbf{B}$ that maximises Equation 4.15, which is a much faster process and well-used by ecological practitioners. CLR involves maximising the following

$$L_{CLR}(\mathbf{B}|\mathbf{X}) = \prod_{i=2}^{N} \frac{W(\mathbf{x}_i,\tau_i,\mathbf{x}_{i-1},\alpha_{\mathbf{x}_{i-1}})}{\sum_{j=0}^{M} W(\mathbf{y}'_{i,j},\tau'_{i,j},\mathbf{x}_{i-1},\alpha_{\mathbf{x}_{i-1}})} \quad (4.16)$$

where $\mathbf{y}'_{i,j}$ and $\tau'_{i,j}$ are the end point and step-time, respectively, of the j^{th} control step drawn from $\varphi(\mathbf{y},\tau|\mathbf{x}_{i-1})$. Sampling control steps from φ is a two step procedure,

where firstly a set of headings, $\alpha'_{i,j}$, and step-times, $\tau'_{i,j}$, is drawn from $h(\alpha)$ and $g_1(\tau)$, respectively. Secondly, a step length, $l'_{i,j}$, is drawn for each step-time from $g_2(l|\tau'_{i,j})$. The control steps are constructed such that $\mathbf{y}'_{i,j} = \mathbf{x}_{i-1} + l'_{i,j}[\cos(\alpha'_{i,j})], \sin(\alpha'_{i,j})]$. I set $\mathbf{y}'_{i,0} = \mathbf{x}_i$ and $\tau'_{i,0} = \tau_i$.

Notice that the denominator in Equation 4.16 approximates the normalising constant in Equation 4.15 up to a constant with respect to $\beta_1, \ldots, \beta_{n+4}$ (by the theory of Monte Carlo Integration), whereby the integral is approximated as the mean of the integrand evaluated at the end points and step times of the sampled control steps. Also, since $\varphi(\mathbf{x}_i, \tau_i|\mathbf{x}_{i-1})$ does not depend on $\beta_1, \ldots, \beta_{n+4}$ the following holds

$$\frac{\varphi(\mathbf{x}_i, \tau_i|\mathbf{x}_{i-1})W(\mathbf{x}_i, \tau_i, \mathbf{x}_{i-1}, \alpha_{\mathbf{x}_{i-1}})}{\int_\Omega \int_0^\infty \varphi(\mathbf{z}, \tau|\mathbf{x}_{i-1})W(\mathbf{z}, \tau, \mathbf{x}_{i-1}, \alpha_{\mathbf{x}_{i-1}})\mathrm{d}\mathbf{z}\mathrm{d}\tau} \approx K_i \frac{W(\mathbf{x}_i, \tau_i, \mathbf{x}_{i-1}, \alpha_{\mathbf{x}_{i-1}})}{\sum_{j=0}^{M} W(\mathbf{y}'_{i,j}, \tau'_{i,j}, \mathbf{x}_{i-1}, \alpha_{\mathbf{x}_{i-1}})} \tag{4.17}$$

where K_i is constant with respect to $\beta_1, \ldots, \beta_{n+4}$. Equation 4.17 shows that the β_i-values that maximise L_{CLR} are approximately the same as those that maximise L_I. This approximation becomes exact as $M \to \infty$ (i.e. as more control steps are sampled). Hence, CLR can be used to approximate the best fit β_i-values to parametrise the movement kernel (Eqn. 4.1).

Constant Step-Time Data

As well as using the Turning Points Algorithm I also subsample the goat paths regularly in time to have a comparison of iSSA and tiSSA applied to the same data set. I use constant subsampling rates varying from 5 seconds to 7 hours.

I use a uniform distribution for the heading and a gamma distribution for the step-lengths, which means the sampling kernel is defined as

$$\varphi^{(c)}(\mathbf{y}|\mathbf{x}) = \frac{l^{k_3-1}\exp(-l/\theta_3)}{2\pi\Gamma(k_3)\theta_3^{k_3}} \tag{4.18}$$

where k_3 and θ_3 are the shape and scale values from fitting a gamma distribution to the step-lengths, respectively. Note the difference between $\varphi^{(c)}(\mathbf{y}|\mathbf{x})$ and $\varphi^{(r)}(\mathbf{y}, \tau|\mathbf{x})$; there is no τ in $\varphi^{(c)}$ since the step-time is the same for all of the steps in the regularly subsampled paths.

The weighting function for the regularly subsampled paths includes the same covariates as I use in the Base model (i.e. Z_1, Z_2 and Z_3), since this is the only model which does not include step-time as a variable, so the weighting function is defined as

$$W^{(c)}(\mathbf{y}, \mathbf{x}, \alpha_x) = \exp\Big(\beta_1^{(c)} Z_1(\mathbf{y}, \mathbf{x}) + \beta_2^{(c)} Z_2(\mathbf{y}, \mathbf{x}, \alpha_x) + \beta_3^{(c)} Z_3(\mathbf{y}, \mathbf{x}) + \beta_{10}^{(c)}|\mathbf{y} - \mathbf{x}| + \beta_{11}^{(c)} \ln|\mathbf{y} - \mathbf{x}|\Big). \tag{4.19}$$

Note that here I correct for the step-length and logarithm of the step-length, but I find that for some paths and regular subsampling frequencies that there are steps with lengths of zero, so in these cases I remove $\beta_{11}^{(c)} \ln|\mathbf{y} - \mathbf{x}|$ from $W^{(c)}$.

The movement kernel for the constant step-time data has the following functional form

$$f^{(c)}(\mathbf{y}|\mathbf{x},\alpha_{\mathbf{x}}) \propto \exp\Bigg(\beta_1^{(c)} Z_1(\mathbf{y},\mathbf{x}) + \beta_2^{(c)} Z_2(\mathbf{y},\mathbf{x},\alpha_{\mathbf{x}}) + \beta_3^{(c)} Z_3(\mathbf{y},\mathbf{x}) + \left(\beta_{10}^{(c)} - \frac{1}{\theta_3}\right)|\mathbf{y}-\mathbf{x}| + (\beta_{11}^{(c)} + k_3 - 1)\ln|\mathbf{y}-\mathbf{x}|\Bigg). \tag{4.20}$$

4.2.4 Results

I find little evidence that the goats are avoiding steep upward slopes, since the inferred β_3-values are insignificant for the regularly subsampled paths and the paths rarefied using the Turning Points Algorithm. The goats sometimes use areas of high altitude as places of safety, which may explain these insignificant β_3-values. Therefore, in all of the subsequent analyses the covariate corresponding to the hypothesis of avoiding steep upward slopes (Z_3) is removed.

Constant Subsampling vs. the Turning Points Algorithm

I compare using the Base model on regularly subsampled data with iSSA against data rarefied using the Turning Points Algorithm with tiSSA. Figures 4.3-4.6 show the results from using data rarefied from constant subsampling (in blue) and the Turning Points Algorithm (in red). The precision of the two methods is given by the size of the confidence intervals. The accuracy of the iSSA method is given by the difference between the inferred $\beta_i^{(c)}$-values and the corresponding $\beta_i^{(r)}$-value inferred from tiSSA.

Figure 4.3a shows the inferred β_1-values, which represent the tendency for the goats to move toward the central area. For six of the seven goat paths the inferred $\beta_1^{(r)}$-value is significant. The $\beta_1^{(c)}$-values inferred from the regularly subsampled paths are greatly affected by the frequency of subsampled locations. The accuracy of the inferred $\beta_1^{(c)}$-values decreases as the frequency of constant subsampling decreases.

For the highest constant subsampling frequencies the $\beta_1^{(c)}$-values are negative for six of the seven paths, which might suggest that the goats are not attracted to the central area and instead tend to be repelled away from it. This is in apparent contradiction with the results from both the data rarefied using the Turning Points Algorithm and lower frequency data. This suggests there is an alternative explanation for these negative $\beta_1^{(c)}$-values is that goats tend to have more directed movement when moving toward the central area than when they are moving away from it. This means that at high subsampling frequencies there are more steps directed away from the central area than toward it, which could be the cause of the negative $\beta_1^{(c)}$-values that we see with the high frequency subsampled data. To further investigate the cause of these negative $\beta_1^{(c)}$-values I test the hypothesis that the goats have longer step-times when moving toward the central area using the $\tilde{Z}_1$ covariate.

In Figure 4.3 (b) we see the inferred β_2-values, which represent the tendency for the goats to make smaller turns. The inferred $\beta_2^{(c)}$-values are inaccurate for almost

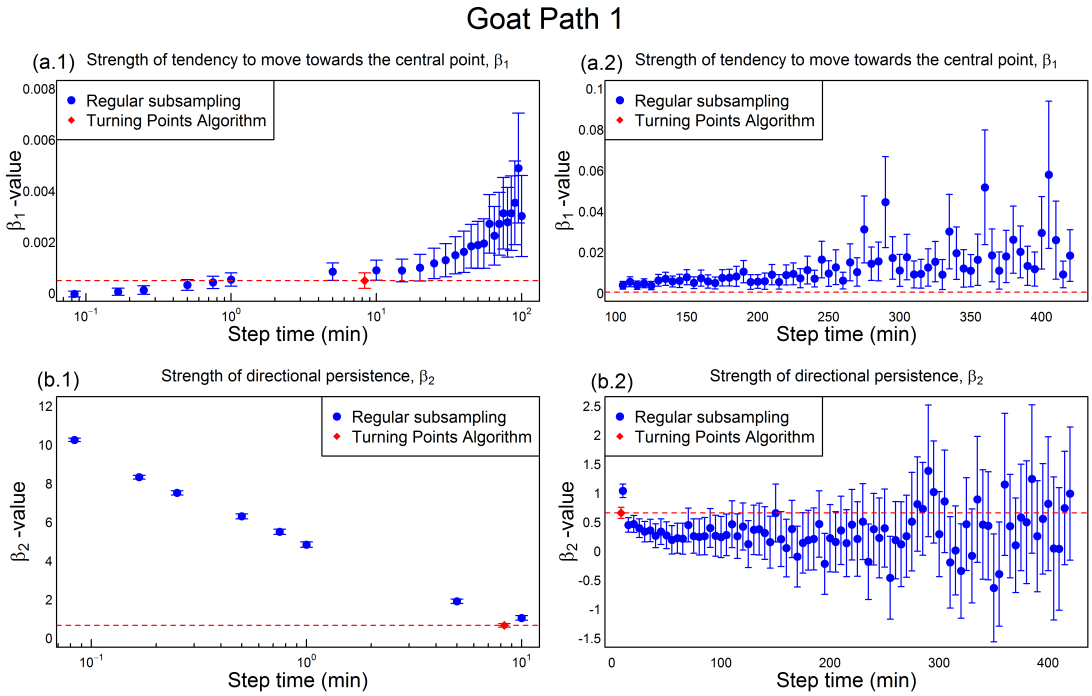

Figure 4.3: **The β_i-values for Goat Path 1**
The β_i-values found from regular subsampling (in blue) and non-constant step time paths (in red), with 95% confidence intervals. The x-axis location of the non-constant step time path is its average step time. Panels (a.1-2) and (b.1-2) show the inferred β_1 and β_2-values, respectively.

all of the regularly subsampled paths when compared to the values from using tiSSA. For high constant subsampling frequencies we see that as the subsampling frequency increases the $\beta_2^{(c)}$-values rise swiftly, but this means that the goats tend to move in straight lines over short time intervals, so this is not very informative about the goats' tendency to prefer smaller turning-angles. When the subsampling frequency is low the confidence intervals around the $\beta_2^{(c)}$-values are so large that the inferred $\beta_2^{(c)}$-values are insignificant.

Time-dependent Covariates

Figure 4.7 shows the inferred $\beta_i^{(r)}$-values for the different models. The results in panels (a) and (c) are from the Base model, whereas panels (b), (d) and (e) are from the Returns, Temperature and Rocky models, respectively. For four of the seven goat paths we see significant inferred $\tilde{\beta}_1^{(r)}$-values (Fig. 4.7b), which represent the tendency for the goats to have longer step-times when moving in the direction of the central area. All of the goats tend to have higher turning frequencies in higher temperatures (Fig. 4.7d), which is probably due to them restricting their movements to shaded areas caused by the goat pen and other nearby buildings in the central area. Also all of the

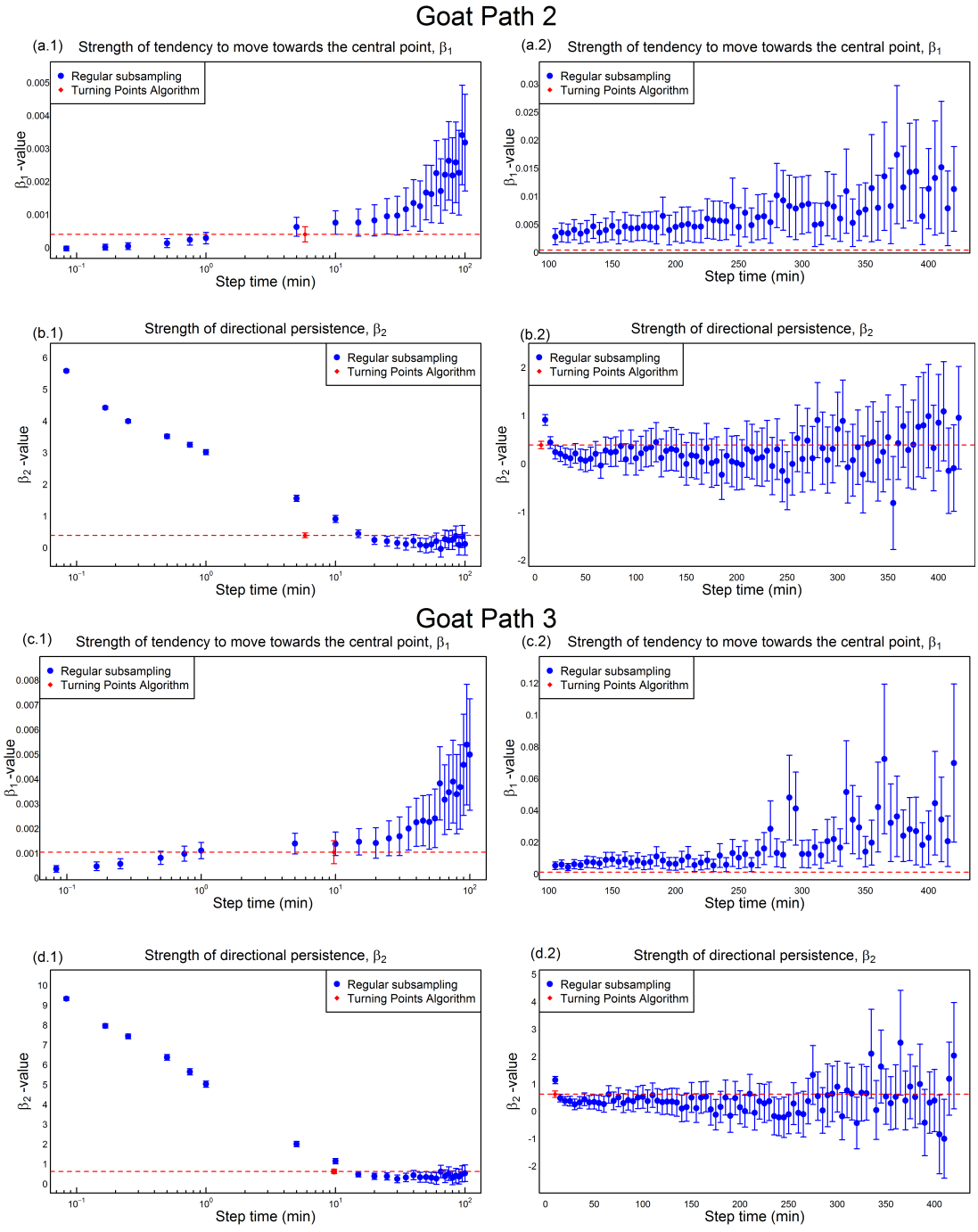

Figure 4.4: **β-values for Goat Paths 2 and 3**
The β_i-values inferred from Goat Path 2 (panels a-b.1-2) and 3 (panels c-d.1-2) presented in the same way as in Figure 4.3

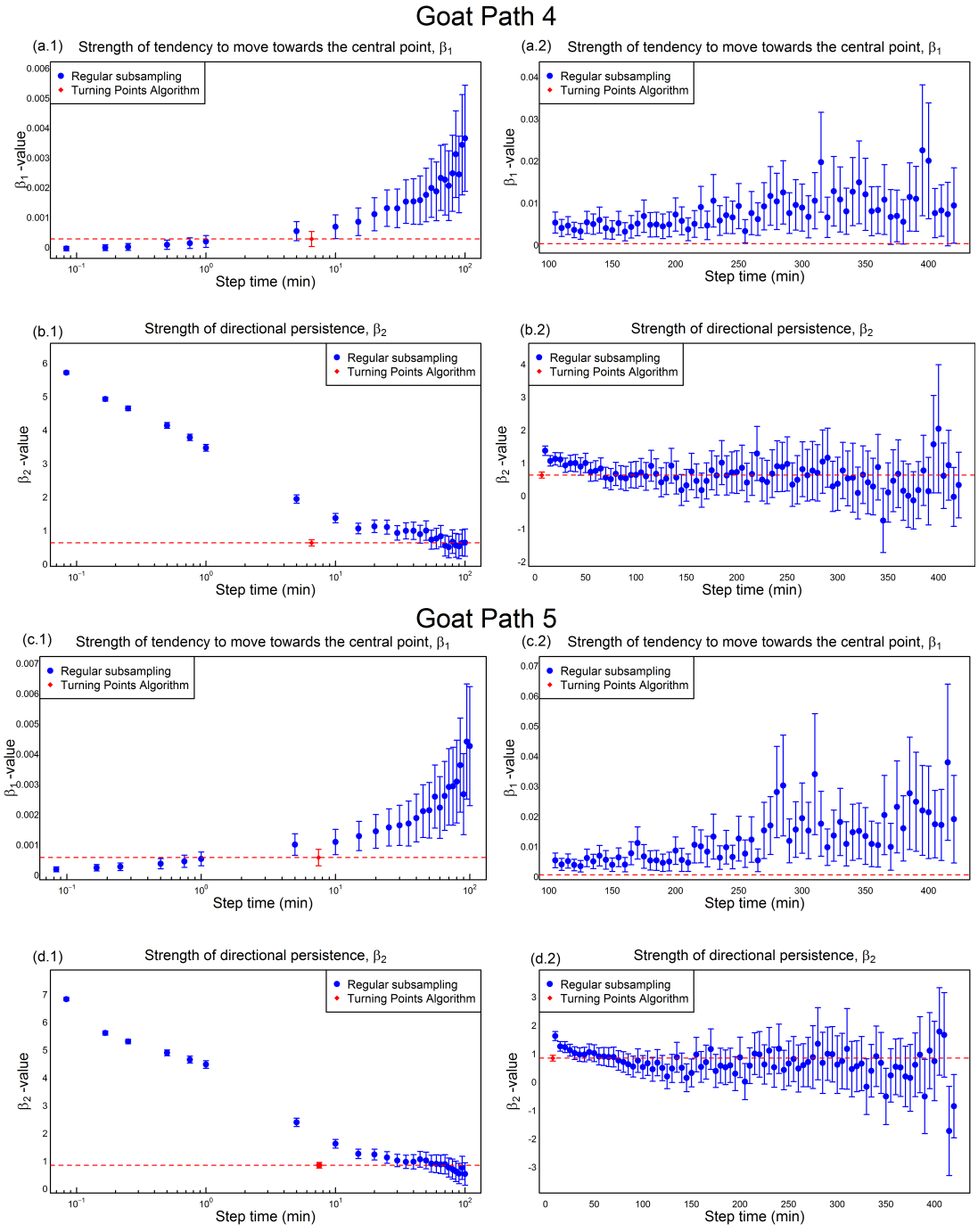

Figure 4.5: **β-values for Goat Paths 3 and 4**
The β_i-values inferred from Goat Path 4 (panels a-b.1-2) and 5 (panels c-d.1-2) presented in the same way as in Figure 4.3

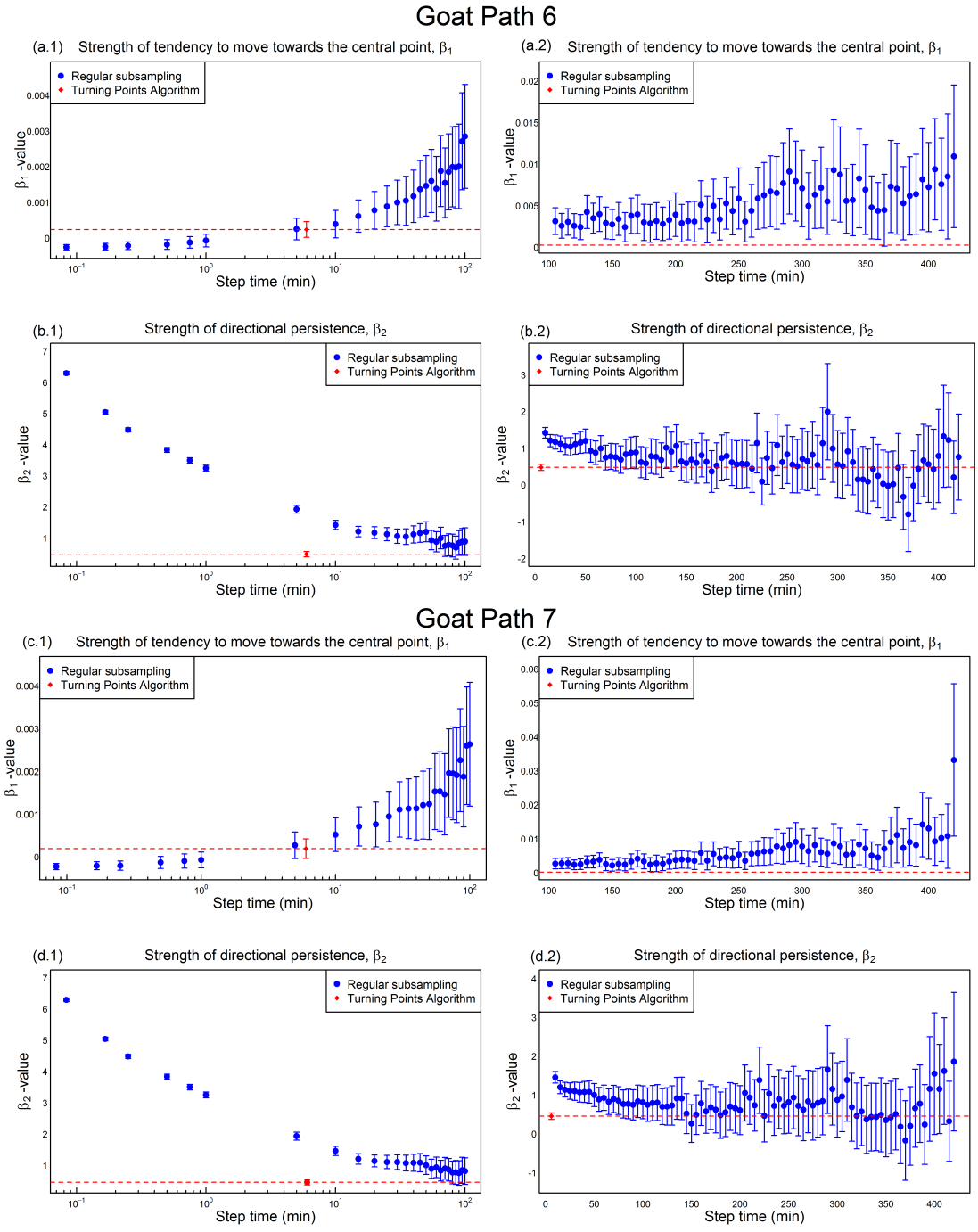

Figure 4.6: **β-values for Goat Path 4 and 5**
The β_i-values inferred from Goat Path 6 (panels a-b.1-2) and 7 (panels c-d.1-2) presented in the same way as in Figure 4.3

goats tend to have higher turning frequencies when moving in rocky terrain (Fig. 4.7e), which is probably due to the fact that it is more difficult to move in a straight line across uneven ground, which results in them taking more turns.

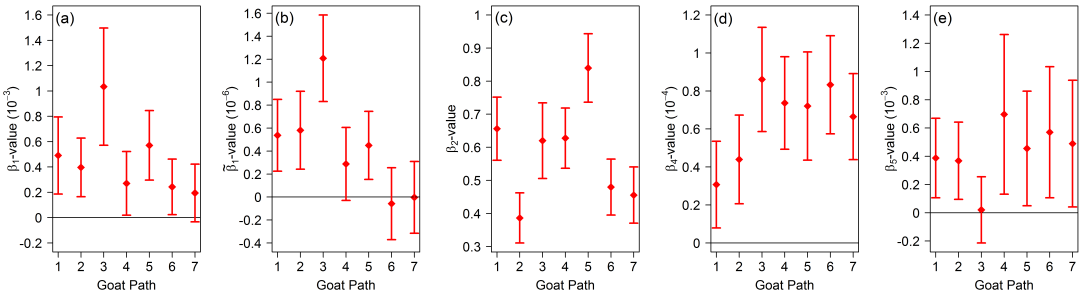

Figure 4.7: **Inferred $\beta_i^{(\mathbf{r})}$-values for different covariates**

Inferred $\beta_i^{(\mathrm{r})}$-values from non-constant step-time paths, where $\beta_1^{(\mathrm{r})}$ represents a tendency to move toward the central area, with strength proportional to the distance from the central area, $\tilde{\beta}_1^{(\mathrm{r})}$ represents the same as $\beta_1^{(\mathrm{r})}$, but with the added assumption that the step times are greater when moving toward the central area, $\beta_2^{(\mathrm{r})}$ represents a tendency for the goats to move in the same general direction, $\beta_4^{(\mathrm{r})}$ represents a tendency for the goats to have shorter step times when the temperature is higher and $\beta_5^{(\mathrm{r})}$ represents the tendency for the goats to have shorter step times when traversing rocky terrain.

§ 4.3 Discussion

In this chapter I have adapted the method of integrated Step Selection Analysis (iSSA) for use on high frequency data (of the order of ≥ 1Hz). In particular I have rarefied such data using the Turning Points Algorithm, which results in a sequence of locations whereby the times between successive locations is not constant (non-constant step-times), but the locations represent places where the animal has actually made a turn. To deal with such non-constant step-times required me to extend the iSSA technique. I termed the resulting method time-varying iSSA (tiSSA).

It is possible to identify an animal's key movement decisions in other ways, such as using the Sites of Interest Algorithm (Munden et al., 2019), which identifies areas in which an animal spends a disproportionately large amount of its time. The movements of an animal when they transition from one site to another could potentially be considered as the steps in tiSSA, enabling the inference of the reasons behind an animal transitioning from one site to another. Performing tiSSA on transitions between sites would be most useful when there are more transitions than I found in the data I used, which would involve using data gathered over longer periods of time than I use and with enough identified sites. It would not be very interesting to perform tiSSA on the transitions between sites if there are only a couple of sites.

Instead of collecting high frequency data which is then rarefied using methods such

as the Turning Points Algorithm, it may be possible to collect data in the first place that reflects the key movement decisions of an animal. For example, in the case of passerines, relocations may be collected when a bird is observed to move from one tree to another (Napper and Hatchwell, 2016; Ellison et al., 2020). Key movement decisions could also be identified from animal-borne video cameras (Rutz et al., 2007; Watanabe and Takahashi, 2013; Tremblay et al., 2014).

It is beneficial to collect data at significant movement decisions, either directly from the animal or rarefied high frequency data. To show these benefits I compare the results from the tiSSA and iSSA methods. For high frequency regular subsampling the results can be misleading (e.g. negative $\beta_1^{(c)}$-values) or uninformative (e.g. extremely high $\beta_2^{(c)}$-values when the constant step-time is small are to be expected and reveal nothing about the goat's tendency to make larger or smaller turns). On the other hand, low frequency regular subsampling leads to results which are both inaccurate and imprecise.

Since tiSSA is a generalisation of SSA it can be used to answer as wide a range of questions as SSA (Thurfjell et al., 2014). For example, SSA has been used to assess the impact that roads (Prokopenko et al., 2017*a*,*b*; Scrafford et al., 2018; Wadey et al., 2018) and other linear features (DeMars and Boutin, 2018; Dickie et al., 2019) have on animal movements, as well as the effects of other anthropogenic disturbances (Ladle et al., 2019). It has also been used to test the effects of other individuals, both in terms of competitors (Vanak et al., 2013) and prey (Latombe et al., 2014). However, unique to the tiSSA method, we can include time dependent covariates, which enables investigation into the factors causing an animal to make movement decisions more or less often. For example, I have shown that both higher temperatures and if the goats are moving through rocky terrain they tend to make turns more often (i.e. have shorter step-times).

§ 4.4 Glossary

Symbol	Explanation
$\mathbf{x}_i$	i^{th} turning point
τ_i	i^{th} step-time, i.e. time taken to move from $\mathbf{x}_{i-1}$ to $\mathbf{x}_i$
l_i	i^{th} step-length, i.e. distance between $\mathbf{x}_{i-1}$ and $\mathbf{x}_i$
s_i	i^{th} step-speed, i.e. l_i/τ_i
$\alpha_{\mathbf{x}}$	heading of arriving at $\mathbf{x}$
N	number of case steps
$\mathbf{X}$	empirical data of turning-points and step-times, ie. $\{\{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N\}, \{\tau_1, \tau_2, \ldots, \tau_N\}\}$
Ω	domain of availability
$f(\mathbf{y}, \tau \vert \mathbf{x}, \alpha_{\mathbf{x}})$	general movement kernel for tiSSA
$f_1^{(\text{r})}$, $f_2^{(\text{r})}$, $f_3^{(\text{r})}$, $f_4^{(\text{r})}$	movement kernels for the Base, Returns, Temperature and Rocky models, respectively for tiSSA
C	normalising constant of the movement kernel, $f(\mathbf{y}, \tau \vert \mathbf{x}, \alpha_{\mathbf{x}})$
$\varphi(\mathbf{y}, \tau \vert \mathbf{x})$	general sampling kernel
$h(\alpha), g_1(\tau), g_2(l \vert \tau)$	general PDFs of headings, step-time and step-length dependent on step-time, respectively
$W(\mathbf{y}, \tau, \mathbf{x}, \alpha_{\mathbf{x}})$	general weighting function for tiSSA
$Z_i(\mathbf{y}, \tau, \mathbf{x}, \alpha_{\mathbf{x}})$	i^{th} covariate function
β_i	coefficient of $Z_i(\mathbf{y}, \tau, \mathbf{x}, \alpha_{\mathbf{x}})$ in $W(\mathbf{y}, \tau, \mathbf{x}, \alpha_{\mathbf{x}})$
$\mathbf{Z}$	vector of covariates, $Z_i(\mathbf{y}, \tau, \mathbf{x}, \alpha_{\mathbf{x}})$
$\mathbf{B}$	vector of covariate coefficients, β_i
$L_{CLR}(\mathbf{B} \vert \mathbf{X})$	likelihood function from conditional logistic regression
$L_I(\mathbf{B} \vert \mathbf{X})$	likelihood function from integration
K_i	constants used to show direct proportionality
$\mathbf{y}'_{m,i}$	m^{th} control step end point as an alternative to the case step ending at $\mathbf{x}_i$
$\tau'_{m,i}$, $\alpha'_{m,i}$	m^{th} control step-time and heading from $\mathbf{x}_{i-1}$ to $\mathbf{y}'_{m,i}$, respectively
M	number of control steps for each case step
$\mathbf{x}_{\text{cp}}$	location of the central point
$H(\mathbf{a}, \mathbf{b})$	measure of heading from $\mathbf{b}$ to $\mathbf{a}$.
$\text{elev}(\mathbf{x})$	measure of the elevation at $\mathbf{x}$
$\text{Temp}(\mathbf{x})$	temperature measured when at $\mathbf{x}$
$\overline{\text{Temp}}$	average temperature across all steps
$I(\mathbf{x})$	indicator function, equalling 1 when $\mathbf{x}$ is in rocky terrain and 0 otherwise

Table 4.1: **Chapter 4 Glossary**
Glossary of symbols used in Chapter 4

Chapter 5

Discussion and Conclusion

As high frequency ($\geq$1 Hz) data gain in popularity and become more widespread (Williams et al., 2020), the need for methods to sift efficiently through these long data streams to answer interesting ecological questions becomes ever more vital. Such high frequency data has the advantage that paths can be considered to be pseudo-continuous since often the distances between successive relocations are so short that they are less than the animal's body length. This removes the necessity to interpolate (i.e. add extra points) between recordings, which is often necessary for calculating metrics from lower resolution data (e.g. utilisation distributions (Benhamou and Cornélis, 2010) and residence times (Barraquand and Benhamou, 2008)).

Despite these advantages, we are often faced with problems when trying to apply existing techniques to high frequency data that may have been designed for much lower resolution data. Some techniques, such as the residence time metric (Barraquand and Benhamou, 2008), are prohibitively slow when applied directly to high frequency data, due to the magnitude of the resulting datasets (see Sections 2.5.1 and 3.6.1). For other techniques, such as Step Selection Analysis (SSA; Fortin et al. (2005); Rhodes et al. (2005)), we find that it is no longer sensible to directly apply SSA to high frequency data, since there is usually little interesting information contained in movements between locations that are a second or less apart. In this thesis I introduce adaptations of these two techniques specifically for use on high frequency data.

In Chapters 2 and 3 I describe a method adapted from the residence time metric, termed the Sites of Interest Algorithm and designed to identify areas that are used extensively. I then give illustrative examples from empirical data of how it can be applied. Chapter 4 explains how SSA can be applied to high frequency data and includes an example of its application to empirical data compared to an application of SSA to the same data. In the next two sections, I discuss the implications of the results described in Chapters 2-4.

§ 5.1 Sites of Interest Algorithm

In Chapter 2, I constructed an algorithm that makes full use of high frequency data to calculate the amount of time an animal spends in a clearly delimited area, termed the usage time. The previous method of Barraquand and Benhamou (2008) was limited by the available data to estimate the amount of time spent in and near to an area (which they termed residence time). Also the existing method for calculating residence time, given in Barraquand and Benhamou (2008), proves to be prohibitively slow when applied to sub-second resolution data. Consequently I take advantage of the extra information high resolution data affords us to decrease the runtime of the algorithm by two orders of magnitude (see Sections 2.5.1 and 3.6.1). I call the resulting algorithm the Sites of Interest Algorithm.

The Sites of Interest Algorithm calculates the usage times of circles placed at various points along the animal's path, so a user of the algorithm is required to make a choice about the radius of the circles. I recommend that the usage time of a path be calculated for multiple radius values, since changing the radius value can result in different outputs. Initial limits can be placed on the chosen radius values. A radius value should be at least half the animal's body length otherwise the site would be smaller than the animal. Also a radius value should not be so large that the whole path is contained within one circle, since this would be uninformative about which areas along the path are most used. When analysing group data compared to individual trajectories a larger minimum radius value than half the animal's body length may be more appropriate and especially if the animals use the sites simultaneously, since then the site should be large enough to contain multiple individuals from the group. A wide variety of radius values can be used within these limits. Then, from examining the results, an appropriate value can be chosen by a trial and error method as explained in the next paragraph.

There are several signs to look out for when choosing a suitable radius value, which may indicate that the radius is too large or too small. Sometimes when too small a radius value is used the identified sites can be clustered together (Fig. 2.5b), which means that the number of sites identified is overestimated. On the other hand, if there are extensive parts of the identified sites that are not used then the radius value may be too large (Fig. 2.5d). The maximum percent drop is a measure of difference between the identified sites and other circles, so it is generally best to maximise the maximum percent drop when choosing a radius value (see Section 3.3). However, if the maximum percent drop is 100% then this indicates that there is only one circle left after the rarefication procedure, since there are no other circles to compare it with. If the maximum percent drop is 100% then the radius value is clearly too large, because having only one circle provides no comparison of which areas have relatively high usage times. I also include in the Sites of Interest Algorithm a colour assignment (see Section 2.1.4) which gives an indication of the confidence that the algorithm has in the results by comparing the outputs from two criteria (threshold and stability criteria) used to choose a radius value. Since when the maximum percent drop is 100% the results are uninformative the two criteria find maximum percent drops that are local maximums rather than global maximums across the different radius values used. This simple

feature provides a baseline rule for choosing an appropriate radius value and indicates when further analyses may be appropriate

In Section 2.6 I describe an alternative method to the colour assignment that can be used to choose the radius value using variance-scale plots calculated from the first passage time (Fauchald and Tveraa, 2003; Kapota et al., 2017). The radius value can be chosen as the one resulting in the highest variance in first passage times. Though this method was specifically designed to identify the size of an area in which an animal performs an area restricted search it could also be used to estimate the size of a site of interest if a user of the Sites of Interest Algorithm is particularly interested in determining the size of the sites of interest.

I have shown that it is possible for sites to be of different sizes, so a natural question to ask might be if a variable radius value could be used, but this could add computational time in calculating what radius to use at each point. The maximum percent drop would also need adapting since it does not make sense to compare the usage times of different sized circles. In most cases the larger circles will have larger usage times, so instead the algorithm would need to normalise the usage time by dividing by the circle's area.

It is possible for there to be several significant drops in usage time between consecutive bars in the histograms of usage times (Fig. 3.2c), which have quite large percent drop values (i.e. relative differences in usage times of consecutive bars). Therefore, it may be beneficial to identify the second and third highest percent drops as well as the maximum (Fig. 3.5), since the maximum percent drop can distinguish a small number of extremely high use sites or extremely low use circles from all other circles. The extremely high use sites (e.g. sleeping sites) may obscure other interesting sites from being identified, whereas extremely low use areas often correspond to uninteresting behaviours (e.g. transitioning between sites).

I have shown the efficacy of the Sites of Interest Algorithm by testing its ability to identify sites of interest known *a priori* from simulated (see Section 2.2) and empirical data (see Sections 3.2 and 3.3). I use switching Ornstein-Uhlenbeck processes (Blackwell, 1997; Taylor and Karlin, 2014) to create simulated trajectories each of which has points of attraction that the simulated particle moves toward, so I aimed to identify these points of attraction as sites of interest using the Sites of Interest Algorithm. In the empirical data, these known sites constitute features of known interest to the animals, such as the watering hole and milking station locations for cattle or salt blocks and a goat pen for goats.

As well as examining sites of interest inferred from a single path, the algorithm can also find sites across multiple trajectories. This may be useful if we want to find sites of shared interest to a group of animals (Fig. 3.2), or when the data is discontinuous because certain segments have been removed. Depending on the question at hand, it is not always valuable to use the entire path for identifying specific types of sites of interest. For example, if we are interested in foraging behaviour then locations recorded whilst the animal is asleep are not useful. Indeed, when I applied the Sites of Interest Algorithm to cattle paths, I often found a single site which is used overwhelmingly more than other sites (see Fig. 3.4). On further examination, I found that this site was

the place where the cattle slept. Only after removing the recordings in this extremely high use site was I able to identify the grazing sites of interest correctly (see Section 3.5).

The Sites of Interest Algorithm fits in the wider field of site fidelity as a method for describing how an animal uses its environment. However, unlike other ways of describing an animal's space-use pattern, such as the Utilisation Distribution (UD; Van Winkle (1975)), the Sites of Interest Algorithm specifically identifies high use areas. These high-use areas are delimited from the rest of the environment available to the animal, which enables a quick segmentation of the path into inter and intra-site movements, whereas the UD gives a more general description of space-use. Though it is not the primary purpose of the Sites of Interest Algorithm it can also be used to create a usage time heat map (see Section 2.7), which is analogous to the UD. This usage time heat map can be used instead of Kernel Density Estimation (Silverman, 1986; Worton, 1989; Benhamou and Cornélis, 2010) for calculating the UD from high frequency data. From the heat map we can identify the shapes of non-circular sites (Fig. 2.14), since there may be areas within an identified site that are less used because they are too energetically costly or dangerous to traverse.

As part of the calculation of the usage time, the number of visits to each circle is also recorded, which can be used to identify interesting areas other than high-use areas that may be visited often but not for significant amounts of time. For example, a repeatedly revisited area may be used to traverse topographically challenging landscape features, such as a ford in a river. Identifying these repeatedly revisited areas is useful in understanding how an animal interacts with the environment and uses particular environmental features to their advantage. Areas with a high number of visits can be identified from a visual inspection of the heat map of the number of visits (Fig. 2.15c). Alternatively the same method as I use to identify sites of interest can be used to identify intensively revisited areas, where we find the maximum percent drop in the number of visits to each circle instead of the usage time.

One can attain accurate representations of an animal's space use through creating heat maps from the results of the Sites of Interest Algorithm since the animal's location is known in pseudo-continuous time. When the resolution is high enough there is no uncertainty in the animal's location between consecutive recordings. However, this does require that ultra-high-resolution data be used, which may not be available to a researcher so other methods (e.g. Kernel Density Estimation; Benhamou and Cornélis (2010)) could be used instead when the resolution of the data is not sufficiently high enough.

§ 5.2 Time-varying integrated Step Selection Analysis

Chapter 4 explains how integrated Step Selection Analysis (iSSA; Avgar et al. (2016)) can be applied to high frequency data, which is first rarefied using the Turning Points Algorithm (Potts et al., 2018) enabling us to consider an animal's movements between consecutive turning points to be steps. The advantage of this is that turning points likely correspond to times when an animal has made a decision to move, since turning

is energetically costly so would not be made unless they benefited from turning (Wilson et al., 2013*b*). Consequently, the methods developed in Chapter 4 take advantage of high frequency data to put Step Selection Analysis on a more behaviourally-meaningful footing.

However, it is not possible to use iSSA directly on steps between turning points, as iSSA is designed for data whereby the time intervals between successive steps (called step-time) are constant throughout the path. This is not the case with the time intervals between actual turning points of animals. This therefore necessitates the adaptation of iSSA to incorporate steps with varying step-times. I term this adapted method time-varying iSSA (tiSSA).

By applying tiSSA to data on goat movements, I showed that using iSSA with lower frequency data gathered at constant step-times can lead to inaccurate results compared with tiSSA applied to the actual turning points of the animal (as inferred by the Turning Points Algorithm). The tiSSA method also allows for the inclusion of time-dependent covariates examining how the frequency of turns may be affected by landscape features or environmental conditions. For example, in Section 4.2.4 I found that the goats from our study tend to turn more frequently when the temperature is higher or when they are moving through rocky terrain.

I suggest that the tiSSA method only be applied to data where recordings occur when the animal has made key movement decisions. These recordings may be attained from identifying when movement decisions are made from high resolution data or from directly collecting relocation data when key movement decisions are made. If the locations in the data do not occur when key movement decisions are made then it makes more sense to use iSSA. It is possible for data to be unevenly spaced in time but for the recordings not to occur when key movement decisions are made, such as when there are failures in the tracking method. In these cases I would suggest using continuous time methods (Wang et al., 2019) which are well suited to such data where movement decisions are made between times when recordings are collected (Blackwell et al., 2016).

§ 5.3 Future Work

5.3.1 Sites of Interest Algorithm

The Sites of Interest Algorithm opens up opportunities to analyse periodic patterns in an animal's movement routine (see Section 3.7). The output from the Sites of Interest Algorithm includes the times when the animal visits and then leaves a site (i.e. crossing times). From these crossing times we can calculate how long an animal spends inside a site for each visit and the length of time between visits. It may be possible to use this information, alongside the method described in Riotte-Lambert et al. (2013) to determine if there is a particular period with which an animal tends to revisit a site. In short, the method of Riotte-Lambert et al. (2013) works by applying Fast Fourier transforms to time series representing either entrance or exit times or presence/absence. Animals may have a particular movement routine, where they tend to visit sites in a certain order. It may be possible to determine such a routine from the flow chart (Fig.

2.11) of movements between sites using conditional entropy (Riotte-Lambert et al., 2016), whereby repeating subsequences are detected in the sequence of sites visited. This is done by creating a site-lag upper triangular matrix (see Figure 2 in Riotte-Lambert et al. (2016)), which gives a visualisation of which subsequences are repeated and the time between occurrences of the subsequence (i.e. lag).

It is possible to infer shared sites of interest (see Section 2.1.5) across subgroups unified under a particular trait, such as gender or age. This may enable a researcher to assess whether there is any variation in the sites identified from different subgroups, hence allowing for the inference of the particular traits that cause such variation. For example, there are sex-specific variations in wandering albatross's foraging strategies that are dependent on wind conditions (Clay et al., 2020). In this situation and other similar scenarios the Sites of Interest Algorithm can be applied more generally to identify sites used for different reasons (e.g. foraging and resting). Also the identified sites highlight areas where it would be valuable to gather environmental data, which in general can be quite coarse (Dodge et al., 2013). It is also possible to compare the results from individuals to the group as a whole to identify the reasons for individuals to move differently to the rest of the group. For example, there is one cattle path from Group 1 that moves to the North-East where no other individuals go (see Section 3.4).

The typically-large intervals between GPS relocations mean that these data are often unable to capture the full range of social interactions (Wilmers et al., 2015), but some of the inherent issues can be alleviated by using data gathered at sub-second resolutions. This is particularly true in conspecific interactions where animals are more likely to interact directly with each other than predator-prey interactions where the prey is actively trying to avoid any interactions with the predator. Such direct interactions are important in understanding the spread of disease from particularly sociable individuals (i.e. super spreaders; Lloyd-Smith et al. (2005)). Disease transmissions will more likely take place when animals spend extended periods of time together as they can do at sites of interest, so identifying sites can help trace other individuals that may be infected by one infected individual visiting the same site.

By comparing usage time heat maps from different conspecific groups (e.g. herds of ungulates or flocks of birds) we may see territorial patterns emerge as they compete for the same resources and therefore the same sites. It is valuable to use high frequency data in these situations to observe direct interactions (e.g. fighting; Smith (1974)), since most studies are based on stigmergent interactions (Giuggioli et al., 2013), such as scent marking (Giuggioli et al., 2011). High frequency data also enables the examination of edge effects between populations (Ries et al., 2004) at a fine-scale using usage time heat maps from each group. Knowing the animals' locations in pseudo-continuous time enables us to clearly define boundaries, since we know exactly where an animal has been and where they have not been.

Social (or dominance) hierarchies (Drews, 1993) may exist within a group that determine the quality of the sites used by an individual dependent on their position in the hierarchy (Morse, 1974; Syme, 1974). From the output of the Sites of Interest Algorithm we know the exact times when an animal visits a site, so from this we can identify

the order in which individuals visit a site. As the site's resources are depleted by the more dominant individuals there are less resources for the less dominant individuals when they arrive later. From the order in which animals visit a site we may be able to determine their rank in the hierarchy.

One could also assess the effects of inter-species interactions, particularly in a predator-prey system. The usage time heat map of a predator can be used to create a map of fear (Gallagher et al., 2017) for their prey, who will tend to avoid areas most used by the predator. This heat map can then be used in SSA (Latombe et al., 2014) to test hypotheses concerning an animal's movement decisions as they weigh up different factors affecting these decisions, such as food quality and predator presence (Brown and Kotler, 2004).

As well as an animal changing their sites of interest in response to the movements of other individuals, there are numerous temporal changes in the environment that could be contributing factors in their decisions of which sites to use, such as the depletion of resources in a foraging patch (Ohashi and Thomson, 2005). Future work could identify sites of interest for different time periods then infer the reasons for an animal to change their sites for particular time periods. Temporal variations may also inform an animal's movement routine as they wait until resources are replenished at a site before revisiting it for example.

Often it can be time consuming to examine an animal's entire trajectory in search of time periods when they exhibit interesting behaviour. The Sites of Interest Algorithm highlights segments of the trajectory corresponding to interesting behaviours, which can be analysed in more detail using accelero-magnetometer data (Moreau et al., 2009; Nathan et al., 2012; Williams et al., 2017). This enables us to infer what an animal uses a site for (e.g. foraging or sleeping). For example, these techniques could be used to infer what use the cattle from Group 2 have for the identified sites that are not the known sites of interest (i.e. milking station locations and watering hole). Alternatively we could examine the environmental data to identify why they visit a site and why they choose that particular site over other areas. It can sometimes be impractical to gather environmental data over the whole landscape available to an animal, so any data that is available will often be quite coarse (Dodge et al., 2013). Instead a researcher can gather environmental data of the areas within and around the identified sites. For example, examining the landscape that the cattle live in I see that there is a wooded area that they tend to avoid, which causes them to take the longer route around the woods when they transition between the sites to East and West of the woods.

5.3.2 Time-varying integrated Step Selection Analysis

The tiSSA method results in a movement kernel, $f(\mathbf{y}, \tau|\mathbf{x}, \alpha_{\mathbf{x}})$ (Eqn. 4.1), which gives the probability density of the animal making successive movement decisions at $\mathbf{x}$ then $\mathbf{y}$ after a time period of τ and arriving at $\mathbf{x}$ on a heading of $\alpha_{\mathbf{x}}$. From this movement kernel simulations can be created (Signer et al., 2019) enabling us to not only describe current space-use patterns, but also to predict future patterns, such as the utilisation distribution (Signer et al., 2017) and home ranges (Moorcroft and Barnett,

2008). From using tiSSA for these same applications it may be possible to achieve more accurate descriptions of these space-use patterns and also to assess the effect of time-dependent covariates in their formation. Since the data used in tiSSA is based on movements between key movement decisions we can gain greater insight compared to iSSA into the mechanistic underpinnings of these space-use patterns.

Though I introduce tiSSA as an extension enabling iSSA to be applied to high frequency data rarefied using the Turning Points Algorithm, it can also be applied to other data where the steps reflect periods when the animal is making a key movement decision. For example, steps might be defined as the movements when an animal transitions from one site of interest to another. Using tiSSA one could infer the reasons behind an animal's decisions to transition between sites at the times they have been observed to. These transitions represent broader scale decisions, rather than the smaller scale decisions made at turning points. For example, an animal may turn to avoid an obstacle as it transitions between sites, but its broader aim of visiting the site is still the same. Alternatively, instead of collecting high frequency data then identifying movement decisions, relocations can be collected at time points that the animal is observed to make movement decisions. For example, animal-borne video cameras (Rutz et al., 2007; Watanabe and Takahashi, 2013; Tremblay et al., 2014) can be used to identify when a movement decision has been made. It is important to apply tiSSA to these data to avoid potentially inaccurate inference, as seen from the example application to goat data (see Section 4.2.4). Also applying tiSSA to data with steps between movement decisions enables the inclusion of time-dependent covariates, which can be used to test hypotheses concerning why an animal makes movement decisions at the times they have been observed to. For example, an animal may choose to leave a patch once the resources there have been depleted below a certain point.

§ 5.4 Summary

In this thesis I have demonstrated the benefits of gathering ultra-high-resolution data and described two tools for utilising these benefits. These tools efficiently extract key information used to identify the motivations underlying an animal's decisions to move.

The Sites of Interest Algorithm is able to identify high use areas along an animal's path, which opens up opportunities to explore questions concerning an animal's movement routine. This algorithm also pin-points areas that may be examined in greater detail either through their intra-site movements or environmental data. Through the use of empirical data I have demonstrated some of the challenges faced when applying the algorithm and given advice on how to overcome these.

The tiSSA method enables iSSA to be applied to high frequency data rarefied using the Turning Points Algorithm, which helps us to avoid potentially inaccurate results. The tiSSA method can also be applied more generally to other non-constant step-time data where recordings are collected when key movement decisions are made. The tiSSA method also enables the inclusion of time-dependent covariates, which can be used to test hypotheses concerning the factors causing an animal to make movement decisions more of less frequently.

Methods previously designed for use on lower resolution data need adapting if they are to fully utilise the benefits afforded to us by using high frequency data. As such data become more widespread the need for methods specifically designed to analyse them will be more vital than ever before.

References

Allen, A. M. and Singh, N. J. 2016. Linking movement ecology with wildlife management and conservation. Frontiers in Ecology and Evolution, **3**:155.

Altizer, S., Bartel, R., and Han, B. A. 2011. Animal migration and infectious disease risk. science, **331**:296–302.

Ardigò, L., Saibene, F., and Minetti, A. 2003. The optimal locomotion on gradients: walking, running or cycling? European journal of applied physiology, **90**:365–371.

Avgar, T., Potts, J. R., Lewis, M. A., and Boyce, M. S. 2016. Integrated step selection analysis: bridging the gap between resource selection and animal movement. Methods in Ecology and Evolution, **7**:619–630.

Bailey, H. and Thompson, P. 2006. Quantitative analysis of bottlenose dolphin movement patterns and their relationship with foraging. Journal of Animal Ecology, **75**:456–465.

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

Bastille-Rousseau, G., Murray, D. L., Schaefer, J. A., Lewis, M. A., Mahoney, S. P., and Potts, J. R. 2018. Spatial scales of habitat selection decisions: implications for telemetry-based movement modelling. Ecography, **41**:437–443.

Benhamou, S. and Bovet, P. 1989. How animals use their environment: a new look at kinesis. Animal Behaviour, **38**:375–383.

Benhamou, S. and Cornélis, D. 2010. Incorporating movement behavior and barriers to improve kernel home range space use estimates. The Journal of Wildlife Management, **74**:1353–1360.

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

Bidder, O., Walker, J., Jones, M., Holton, M., Urge, P., Scantlebury, D., Marks, N., Magowan, E., Maguire, I., and Wilson, R. 2015. Step by step: reconstruction of terrestrial animal movement paths by dead-reckoning. Movement Ecology, **3**:23.

Bidder, O. R., Soresina, M., Shepard, E. L., Halsey, L. G., Quintana, F., Gómez-Laich, A., and Wilson, R. P. 2012. The need for speed: testing acceleration for estimating animal travel rates in terrestrial dead-reckoning systems. Zoology, **115**:58–64.

Blackwell, P. 1997. Random diffusion models for animal movement. Ecological Modelling, **100**:87–102.

Blackwell, P. G., Niu, M., Lambert, M. S., and LaPoint, S. D. 2016. Exact Bayesian inference for animal movement in continuous time. Methods in Ecology and Evolution, **7**:184–195.

Bodey, T. W., Cleasby, I. R., Bell, F., Parr, N., Schultz, A., Votier, S. C., and Bearhop, S. 2018. A phylogenetically controlled meta-analysis of biologging device effects on birds: Deleterious effects and a call for more standardized reporting of study data. Methods in Ecology and Evolution, **9**:946–955.

Bond, A. B. 1980. Optimal foraging in a uniform habitat: the search mechanism of the green lacewing. Animal Behaviour, **28**:10–19.

Boone, R. B., Thirgood, S. J., and Hopcraft, J. G. C. 2006. Serengeti wildebeest migratory patterns modeled from rainfall and new vegetation growth. Ecology, **87**:1987–1994.

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

Brown, J. S. and Kotler, B. P. 2004. Hazardous duty pay and the foraging cost of predation. Ecology letters, **7**:999–1014.

Brown, R. 1828. XXVII. A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. The Philosophical Magazine, **4**:161–173.

Cattell, R. B. 1966. The scree test for the number of factors. Multivariate behavioral research, **1**:245–276.

Charnov, E. L. et al. 1976. Optimal foraging, the marginal value theorem.

Chetkiewicz, C.-L. B. and Boyce, M. S. 2009. Use of resource selection functions to identify conservation corridors. Journal of Applied Ecology, pages 1036–1047.

Clay, T. A., Joo, R., Weimerskirch, H., Phillips, R. A., Den Ouden, O., Basille, M., Clusella-Trullas, S., Assink, J. D., and Patrick, S. C. 2020. Sex-specific effects of wind on the flight decisions of a sexually dimorphic soaring bird. Journal of Animal Ecology, **89**:1811–1823.

Craighead, F. C. 1979. Track of the grizzly. Random House (NY).

Craighead, J. J., Sumner, J. S., and Mitchell, J. A. 1995. The grizzly bears of Yellowstone: their ecology in the Yellowstone ecosystem, 1959-1992. Island Press.

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

DeMars, C. A. and Boutin, S. 2018. Nowhere to hide: effects of linear features on predator–prey dynamics in a large mammal system. Journal of Animal Ecology, **87**:274–284.

D'Eon, R. G., Serrouya, R., Smith, G., and Kochanny, C. O. 2002. GPS radiotelemetry error and bias in mountainous terrain. Wildlife Society Bulletin, pages 430–439.

Dewhirst, O. P., Evans, H. K., Roskilly, K., Harvey, R. J., Hubel, T. Y., and Wilson, A. M. 2016. Improving the accuracy of estimates of animal path and travel distance using GPS drift-corrected dead reckoning. Ecology and evolution, **6**:6210–6222.

Dickie, M., McNay, R., Sutherland, G., Cody, M., and Avgar, T. 2019. Corridors or risk? Movement along, and use of, linear features vary predictably among large mammal predator and prey species. Journal of Animal Ecology.

Dickson, B. G., Jenness, J. S., and Beier, P. 2005. Influence of vegetation, topography, and roads on cougar movement in southern California. The Journal of Wildlife Management, **69**:264–276.

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

Drews, C. 1993. The concept and definition of dominance in animal behaviour. Behaviour, **125**:283–313.

Dussault, C., Courtois, R., Ouellet, J.-P., and Huot, J. 1999. Evaluation of GPS telemetry collar performance for habitat studies in the boreal forest. Wildlife Society Bulletin, pages 965–972.

Edelhoff, H., Signer, J., and Balkenhol, N. 2016. Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns. Movement ecology, **4**:21.

Edwards, A. W. and Cavalli-Sforza, L. L. 1965. A method for cluster analysis. Biometrics, pages 362–375.

Ellison, N., Hatchwell, B. J., Biddiscombe, S. J., Napper, C. J., and Potts, J. R. 2020. Mechanistic home range analysis reveals drivers of space use patterns for a non-territorial passerine. Journal of Animal Ecology.

Fauchald, P. and Tveraa, T. 2003. Using first-passage time in the analysis of area-restricted search and habitat selection. Ecology, **84**:282–288.

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

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

Gallagher, A. J., Creel, S., Wilson, R. P., and Cooke, S. J. 2017. Energy landscapes and the landscape of fear. Trends in Ecology & Evolution, **32**:88–96.

Getz, W. M. and Saltz, D. 2008. A framework for generating and analyzing movement paths on ecological landscapes. Proceedings of the National Academy of Sciences, **105**:19066–19071.

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

Giuggioli, L., Potts, J. R., and Harris, S. 2011. Animal interactions and the emergence of territoriality. PLoS computational biology, **7**:e1002008.

Giuggioli, L., Potts, J. R., Rubenstein, D. I., and Levin, S. A. 2013. Stigmergy, collective actions, and animal social spacing. Proceedings of the National Academy of Sciences, **110**:16904–16909.

Haines, A. M., Grassman, L. I., Tewes, M. E., and Janečka, J. E. 2006. First ocelot (Leopardus pardalis) monitored with GPS telemetry. European Journal of Wildlife Research, **52**:216–218.

Hebblewhite, M., Merrill, E., and McDermid, G. 2008. A multi-scale test of the forage maturation hypothesis in a partially migratory ungulate population. Ecological monographs, **78**:141–166.

Hebblewhite, M., Percy, M., and Merrill, E. 2007. 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, **71**:2026–2033.

Hedenström, A., Norevik, G., Warfvinge, K., Andersson, A., Bäckman, J., and Åkesson, S. 2016. Annual 10-month aerial life phase in the common swift Apus apus. Current Biology, **26**:3066–3070.

Iacus, S. M. 2009. Simulation and inference for stochastic differential equations: with R examples. Springer Science & Business Media.

Jeltsch, F., Bonte, D., Pe'er, G., Reineking, B., Leimgruber, P., Balkenhol, N., Schröder, B., Buchmann, C. M., Mueller, T., Blaum, N., et al. 2013. Integrating movement ecology with biodiversity research-exploring new avenues to address spatiotemporal biodiversity dynamics. Movement Ecology, **1**:1–13.

Jolliffe, I. T. 1986. Principal Component Analysis and Factor Analysis. In Principal component analysis, pages 115–128. Springer.

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

Jonsen, I. D., Myers, R. A., and James, M. C. 2007. Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. Marine Ecology Progress Series, **337**:255–264.

Kapota, D., Dolev, A., and Saltz, D. 2017. Inferring detailed space use from movement paths: A unifying, residence time-based framework. Ecology and Evolution, **7**:8507–8514.

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

Kays, R., Tilak, S., Crofoot, M., Fountain, T., Obando, D., Ortega, A., Kuemmeth, F., Mandel, J., Swenson, G., Lambert, T., et al. 2011. Tracking animal location and activity with an automated radio telemetry system in a tropical rainforest. The Computer Journal, **54**:1931–1948.

Kenward, R. E. 2000. A manual for wildlife radio tagging. Academic press.

Kjellén, N., Hake, M., and Alerstam, T. 1997. Strategies of two Ospreys Pandion haliaetus migrating between Sweden and tropical Africa as revealed by satellite tracking. Journal of Avian Biology, pages 15–23.

Ladle, A., Avgar, T., Wheatley, M., Stenhouse, G. B., Nielsen, S. E., and Boyce, M. S. 2019. Grizzly bear response to spatio-temporal variability in human recreational activity. Journal of Applied Ecology, **56**:375–386.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., and Morales, J. M. 2012. Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. Ecology, **93**:2336–2342.

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

Leimgruber, P., McShea, W. J., Brookes, C. J., Bolor-Erdene, L., Wemmer, C., and Larson, C. 2001. Spatial patterns in relative primary productivity and gazelle migration in the Eastern Steppes of Mongolia. Biological Conservation, **102**:205–212.

Lihoreau, M., Chittka, L., and Raine, N. E. 2011. Trade-off between travel distance and prioritization of high-reward sites in traplining bumblebees. Functional ecology, **25**:1284–1292.

Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., and Getz, W. M. 2005. Superspreading and the effect of individual variation on disease emergence. Nature, **438**:355–359.

Luschi, P., Papi, F., Liew, H., Chan, E., and Bonadonna, F. 1996. Long-distance migration and homing after displacement in the green turtle (Chelonia mydas): a satellite tracking study. Journal of Comparative Physiology A, **178**:447–452.

Matthiopoulos, J. 2003. The use of space by animals as a function of accessibility and preference. Ecological Modelling, **159**:239–268.

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

Merkle, J. A., Cross, P. C., Scurlock, B. M., Cole, E. K., Courtemanch, A. B., Dewey, S. R., and Kauffman, M. J. 2018. Linking spring phenology with mechanistic models of host movement to predict disease transmission risk. Journal of applied ecology, **55**:810–819.

Minetti, A. E., Moia, C., Roi, G. S., Susta, D., and Ferretti, G. 2002. Energy cost of walking and running at extreme uphill and downhill slopes. Journal of applied physiology, **93**:1039–1046.

Moen, R., Pastor, J., Cohen, Y., and Schwartz, C. C. 1996. Effects of moose movement and habitat use on GPS collar performance. The Journal of wildlife management, pages 659–668.

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

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

Moreau, M., Siebert, S., Buerkert, A., and Schlecht, E. 2009. Use of a tri-axial accelerometer for automated recording and classification of goats' grazing behaviour. Applied Animal Behaviour Science, **119**:158–170.

Morse, D. H. 1974. Niche breadth as a function of social dominance. The American Naturalist, **108**:818–830.

Munden, R., Börger, L., Wilson, R. P., Redcliffe, J., Loison, A., Garel, M., and Potts, J. R. 2019. Making sense of ultrahigh-resolution movement data: A new algorithm for inferring sites of interest. Ecology and evolution, **9**:265–274.

Napper, C. J. and Hatchwell, B. J. 2016. Social dynamics in nonbreeding flocks of a cooperatively breeding bird: causes and consequences of kin associations. Animal Behaviour, **122**:23–35.

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

Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M., and Getz, W. M. 2012. Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures. Journal of Experimental Biology, **215**:986–996.

Noda, T., Kawabata, Y., Arai, N., Mitamura, H., and Watanabe, S. 2014. Animal-mounted gyroscope/accelerometer/magnetometer: In situ measurement of the movement performance of fast-start behaviour in fish. Journal of experimental marine biology and ecology, **451**:55–68.

Ohashi, K. and Thomson, J. D. 2005. Efficient harvesting of renewing resources. Behavioral Ecology, **16**:592–605.

Patterson, T. A., Basson, M., Bravington, M. V., and Gunn, J. S. 2009. Classifying movement behaviour in relation to environmental conditions using hidden Markov models. Journal of Animal Ecology, **78**:1113–1123.

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. 2008. State–space models of individual animal movement. Trends in ecology & evolution, **23**:87–94.

Pinaud, D. and Weimerskirch, H. 2005. Scale-dependent habitat use in a long-ranging central place predator. Journal of Animal Ecology, **74**:852–863.

Pinaud, D. and Weimerskirch, H. 2007. At-sea distribution and scale-dependent foraging behaviour of petrels and albatrosses: a comparative study. Journal of Animal Ecology, **76**:9–19.

Potts, J. R., Auger-Méthé, M., Mokross, K., and Lewis, M. A. 2014. A generalized residual technique for analysing complex movement models using earth mover's distance. Methods in Ecology and Evolution, **5**:1012–1022.

Potts, J. R., Börger, L., Scantlebury, D. M., Bennett, N. C., Alagaili, A., and Wilson, R. P. 2018. Finding turning-points in ultra-high-resolution animal movement data. Methods in Ecology and Evolution.

Prokopenko, C. M., Boyce, M. S., and Avgar, T. 2017*a*. Characterizing wildlife behavioural responses to roads using integrated step selection analysis. Journal of Applied Ecology, **54**:470–479.

Prokopenko, C. M., Boyce, M. S., and Avgar, T. 2017*b*. Extent-dependent habitat selection in a migratory large herbivore: road avoidance across scales. Landscape ecology, **32**:313–325.

Pyke, G. H. 1984. Optimal foraging theory: a critical review. Annual review of ecology and systematics, **15**:523–575.

Raynor, E. J., Beyer, H. L., Briggs, J. M., and Joern, A. 2017. Complex variation in habitat selection strategies among individuals driven by extrinsic factors. Ecology and Evolution, **7**:1802–1822.

Redner, S. 2001. A guide to first-passage processes. Cambridge University Press.

Rempel, R. S., Rodgers, A. R., and Abraham, K. F. 1995. Performance of a GPS animal location system under boreal forest canopy. The Journal of wildlife management, pages 543–551.

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

Richard, Y. and Armstrong, D. P. 2010. Cost distance modelling of landscape connectivity and gap-crossing ability using radio-tracking data. Journal of Applied Ecology, **47**:603–610.

Ries, L., Fletcher Jr, R. J., Battin, J., and Sisk, T. D. 2004. Ecological responses to habitat edges: mechanisms, models, and variability explained. Annu. Rev. Ecol. Evol. Syst., **35**:491–522.

Riotte-Lambert, L., Benhamou, S., and Chamaillé-Jammes, S. 2013. Periodicity analysis of movement recursions. Journal of Theoretical Biology, **317**:238–243.

Riotte-Lambert, L., Benhamou, S., and Chamaillé-Jammes, S. 2016. From randomness to traplining: a framework for the study of routine movement behavior. Behavioral Ecology, page arw154.

Rutz, C., Bluff, L. A., Weir, A. A., and Kacelnik, A. 2007. Video cameras on wild birds. Science, **318**:765–765.

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

Scrafford, M. A., Avgar, T., Heeres, R., and Boyce, M. S. 2018. Roads elicit negative movement and habitat-selection responses by wolverines (Gulo gulo luscus). Behavioral Ecology, **29**:534–542.

Shepard, E. L., Wilson, R. P., Halsey, L. G., Quintana, F., Laich, A. G., Gleiss, A. C., Liebsch, N., Myers, A. E., and Norman, B. 2008. Derivation of body motion via appropriate smoothing of acceleration data. Aquatic Biology, **4**:235–241.

Signer, J., Fieberg, J., and Avgar, T. 2017. Estimating utilization distributions from fitted step-selection functions. Ecosphere, **8**:e01771.

Signer, J., Fieberg, J., and Avgar, T. 2019. Animal movement tools (amt): R package for managing tracking data and conducting habitat selection analyses. Ecology and evolution, **9**:880–890.

Silverman, B. W. 1986. Density Estimation for Statistics and Data Analysis, volume 26. CRC Press.

Smith, B. J., Hart, K. M., Mazzotti, F. J., Basille, M., and Romagosa, C. M. 2018. Evaluating GPS biologging technology for studying spatial ecology of large constricting snakes. Animal Biotelemetry, **6**:1.

Smith, J. M. 1974. The theory of games and the evolution of animal conflicts. Journal of theoretical biology, **47**:209–221.

Steinfurth, A., Vargas, F. H., Wilson, R. P., Spindler, M., and Macdonald, D. W. 2008. Space use by foraging Galápagos penguins during chick rearing. Endangered Species Research, **4**:105–112.

Street, G. M., Avgar, T., and Börger, L. 2018. Net displacement and temporal scaling: Model fitting, interpretation and implementation. Methods in Ecology and Evolution, **9**:1503–1517.

Syme, G. 1974. Competitive orders as measures of social dominance. Animal Behaviour, **22**:931–940.

Tanaka, H., Takagi, Y., and Naito, Y. 2001. Swimming speeds and buoyancy compensation of migrating adult chum salmon Oncorhynchus keta revealed by speed/depth/acceleration data logger. Journal of Experimental Biology, **204**:3895–3904.

Taylor, H. M. and Karlin, S. 2014. An introduction to stochastic modeling. Academic press.

Thomson, J. D., Slatkin, M., and Thomson, B. A. 1997. Trapline foraging by bumble bees: II. Definition and detection from sequence data. Behavioral Ecology, **8**:199–210.

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

Tremblay, Y., Thiebault, A., Mullers, R., and Pistorius, P. 2014. Bird-borne video-cameras show that seabird movement patterns relate to previously unrevealed proximate environment, not prey. PLoS One, **9**:e88424.

Uhlenbeck, G. E. and Ornstein, L. S. 1930. On the theory of the Brownian motion. Physical review, **36**:823.

Van Winkle, W. 1975. Comparison of several probabilistic home-range models. The Journal of wildlife management, pages 118–123.

Vanak, A. T., Fortin, D., Thaker, M., Ogden, M., Owen, C., Greatwood, S., and Slotow, R. 2013. Moving to stay in place: behavioral mechanisms for coexistence of African large carnivores. Ecology, **94**:2619–2631.

Wadey, J., Beyer, H. L., Saaban, S., Othman, N., Leimgruber, P., and Campos-Arceiz, A. 2018. Why did the elephant cross the road? The complex response of wild elephants to a major road in Peninsular Malaysia. Biological Conservation, **218**:91–98.

Walker, J. S., Jones, M. W., Laramee, R. S., Holton, M. D., Shepard, E. L., Williams, H. J., Scantlebury, D. M., Nikki, J. M., Magowan, E. A., Maguire, I. E., et al. 2015. Prying into the intimate secrets of animal lives; software beyond hardware for comprehensive annotation in 'Daily Diary' tags. Movement Ecology, **3**:29.

Wang, Y., Blackwell, P., Merkle, J., and Potts, J. 2019. Continuous time resource selection analysis for moving animals. Methods in Ecology and Evolution.

Watanabe, Y. Y. and Takahashi, A. 2013. Linking animal-borne video to accelerometers reveals prey capture variability. Proceedings of the National Academy of Sciences, **110**:2199–2204.

Wikelski, M., Moskowitz, D., Adelman, J. S., Cochran, J., Wilcove, D. S., and May, M. L. 2006. Simple rules guide dragonfly migration. Biology letters, **2**:325–329.

Wikelski, M., Tarlow, E. M., Raim, A., Diehl, R. H., Larkin, R. P., and Visser, G. H. 2003. Costs of migration in free-flying songbirds. Nature, **423**:704–704.

Williams, H., Shepard, E., Duriez, O., and Lambertucci, S. A. 2015. Can accelerometry be used to distinguish between flight types in soaring birds? Animal Biotelemetry, **3**:45.

Williams, H. J., Holton, M. D., Shepard, E. L., Largey, N., Norman, B., Ryan, P. G., Duriez, O., Scantlebury, M., Quintana, F., Magowan, E. A., et al. 2017. Identification of animal movement patterns using tri-axial magnetometry. Movement ecology, **5**:6.

Williams, H. J., Taylor, L. A., Benhamou, S., Bijleveld, A. I., Clay, T. A., de Grissac, S., Demšar, U., English, H. M., Franconi, N., Gómez-Laich, A., et al. 2020. Optimizing the use of biologgers for movement ecology research. Journal of Animal Ecology, **89**:186–206.

Wilmers, C. C., Nickel, B., Bryce, C. M., Smith, J. A., Wheat, R. E., and Yovovich, V. 2015. The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology. Ecology, **96**:1741–1753.

Wilson, A. M., Lowe, J., Roskilly, K., Hudson, P. E., Golabek, K., and McNutt, J. 2013*a*. Locomotion dynamics of hunting in wild cheetahs. Nature, **498**:185.

Wilson, R., Griffiths, I., Legg, P., Friswell, M., Bidder, O., Halsey, L., Lambertucci, S. A., and Shepard, E. 2013*b*. Turn costs change the value of animal search paths. Ecology Letters, **16**:1145–1150.

Wilson, R. P., Grundy, E., Massy, R., Soltis, J., Tysse, B., Holton, M., Cai, Y., Parrott, A., Downey, L. A., Qasem, L., et al. 2014. Wild state secrets: ultra-sensitive measurement of micro-movement can reveal internal processes in animals. Frontiers in Ecology and the Environment, **12**:582–587.

Wilson, R. P., Shepard, E., and Liebsch, N. 2008. Prying into the intimate details of animal lives: use of a daily diary on animals. Endangered Species Research, **4**:123–137.

Worton, B. J. 1989. Kernel methods for estimating the utilization distribution in home-range studies. Ecology, **70**:164–168.

Yoda, K., Sato, K., Niizuma, Y., Kurita, M., Bost, C., Le Maho, Y., and Naito, Y. 1999. Precise monitoring of porpoising behaviour of Adélie penguins determined using acceleration data loggers. Journal of Experimental Biology, **202**:3121–3126.

Appendix A: Additional Figures

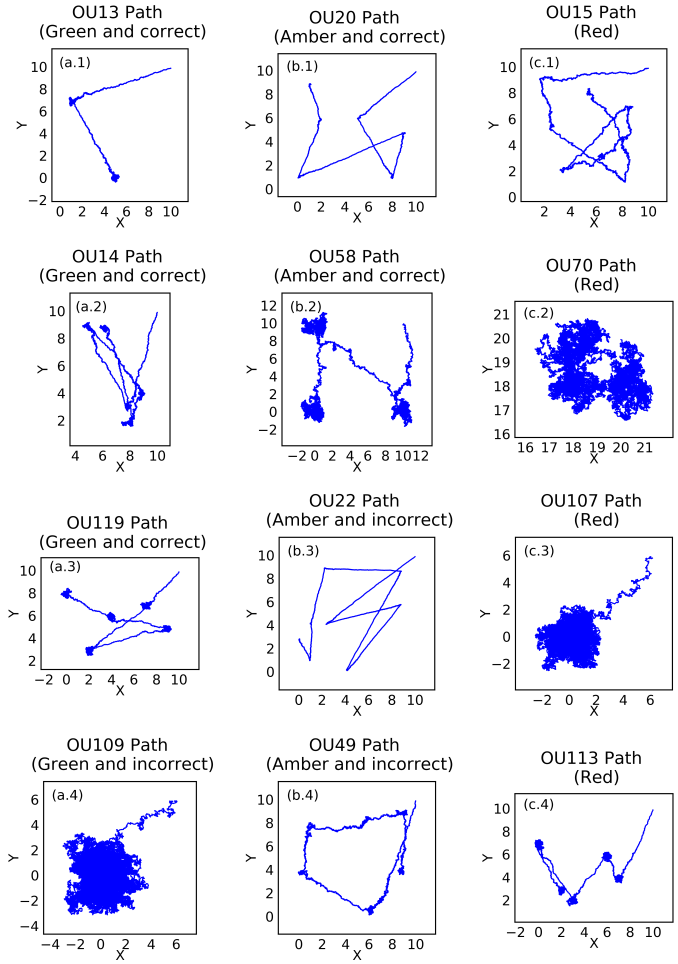

Figure 5.1: **A Selection of Ornstein-Uhlenbeck simulation paths**
A Selection of Ornstein-Uhlenbeck simulation paths with an indication of the colour assigned to the path using the Sites of Interest algorithm and whether the correct number of sites of interest were identified.

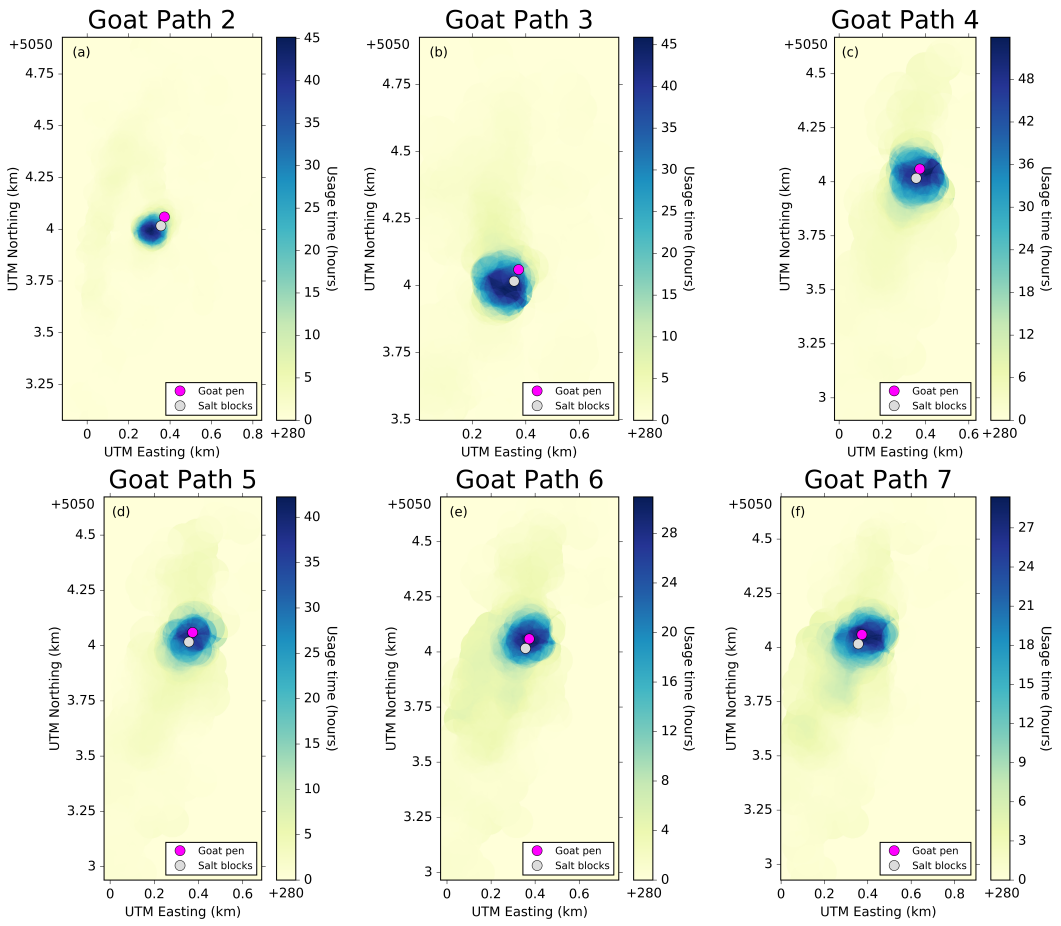

Figure 5.2: **Usage time heat maps for goat path**
Heat maps of usage time for goat paths 2-7 in panels a-f, respectively.

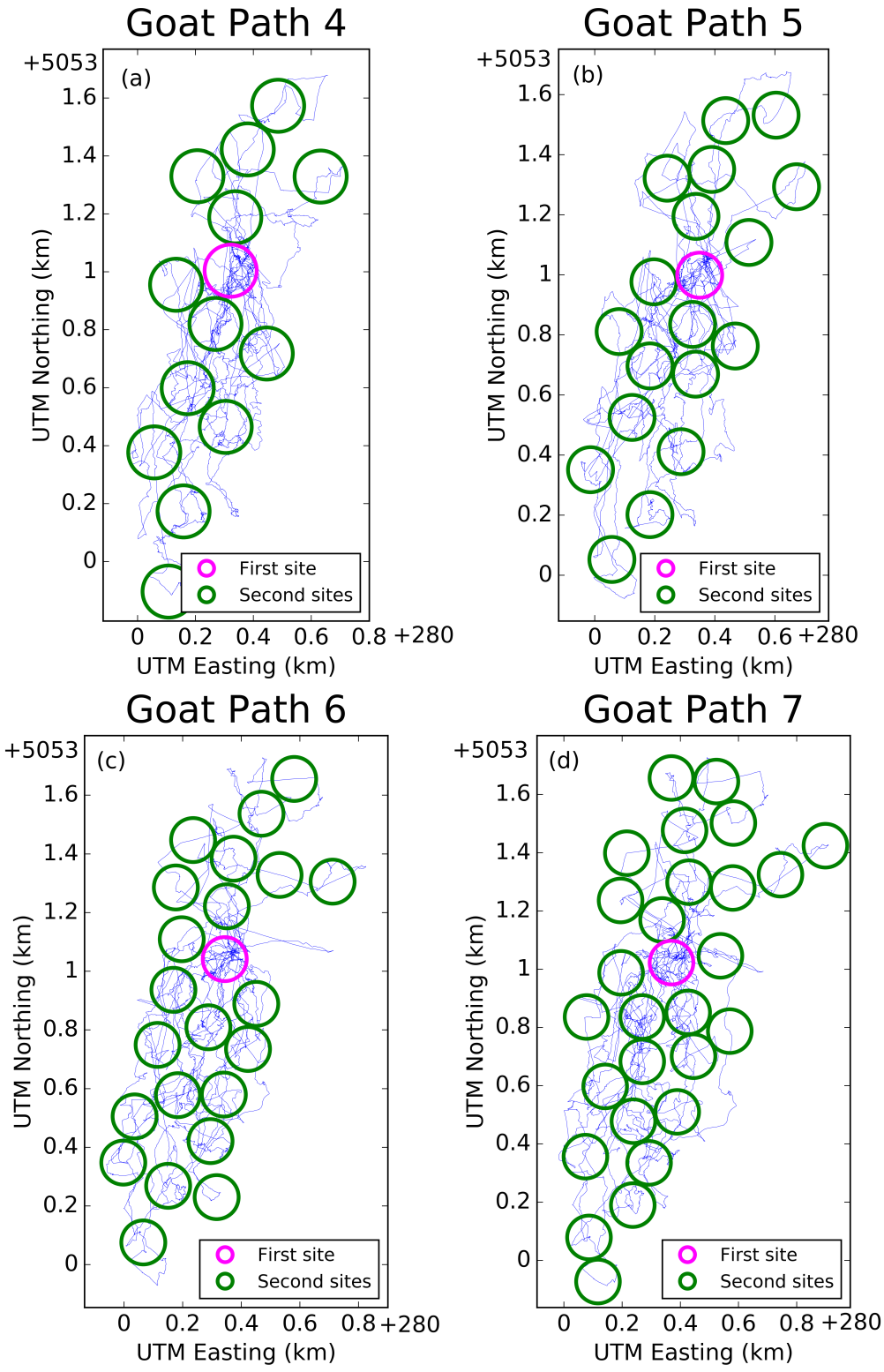

Figure 5.3: **Identified goat sites**
Sites identified from the maximum (pink circle) and second highest percent drop (green circles) for Goat Paths 4-7 in panels (a-d), respectively.

Appendix B: Additional Tables

OU number	Number of points of attraction	Long term standard deviation	Average distance between points of attraction
11	2	17.32050808	4.123105626
12	1	25.29822128	0
13	2	17.5	8.062257748
14	5	15.68929081	4.083655616
15	9	21.38089935	5.425134867
16	3	14.43375673	7.038119745
17	7	18.89822365	4.49114932
18	9	8.528028654	3.399783517
19	8	15.11857892	3.85324933
20	6	7.071067812	5.382584744
21	2	2.357022604	3.605551276
22	8	4.714045208	5.556525929
23	8	6.123724357	5.076562436
24	1	18.89822365	0
25	7	8.660254038	4.912798158
26	5	10	5.725744631
27	2	22.36067978	8.062257748
28	4	15.68929081	7.095093289
29	5	28.86751346	4.976695174
30	9	6.123724357	4.729544651
31	6	17.32050808	4.064343555
32	3	2.236067978	3.954624491
33	3	14.14213562	4.717943447
34	4	18.85618083	3.903077327
35	3	2.236067978	8.398751316

Table 5.1: **Characteristics of OU simulations 11 - 35**
Table of how the OU simulations (11-35) vary, including the number of points of attraction, the long term standard deviation about these points and the average distance between points of attraction.

OU number	Number of points of attraction	Long term standard deviation	Average distance between points of attraction
36	5	2.886751346	3.127913027
37	7	7.844645406	5.889149455
38	7	9.805806757	4.612522825
39	6	8.944271910	4.671099195
40	4	25.29822128	4.059632853
41	5	8.528028654	2.601533972
42	1	26.72612419	0
43	9	25	6.194382572
44	2	5.773502692	4.472135955
45	7	26.72612419	6.407601193
46	7	6.123724357	4.562102360
47	4	14.28869017	7.570496844
48	5	17.00840129	3.370118293
49	7	22.36067978	5.864138580
50	1	16.03567451	0
51	3	18.25741858	11.38071187
52	3	25.56038602	11.38071187
53	3	32.86335345	11.38071187
54	3	40.16632088	11.38071187
55	3	47.46928832	11.38071187
56	3	54.77225575	11.38071187
57	3	62.07522318	11.38071187
58	3	69.37819062	11.38071187
59	3	76.68115805	11.38071187
60	3	83.98412548	11.38071187
61	3	54.77225575	22.76142375
62	3	54.77225575	20.48528137
63	3	54.77225575	18.20913900
64	3	54.77225575	15.93299662

Table 5.2: **Characteristics of OU simulations 36- 64**
Table of how the OU simulations (39-66) vary, including the number of points of attraction, the long term standard deviation about these points and the average distance between points of attraction.

OU number	Number of points of attraction	Long term standard deviation	Average distance between points of attraction
65	3	54.77225575	13.65685425
66	3	54.77225575	11.38071187
67	3	54.77225575	9.104569500
68	3	54.77225575	6.828427125
69	3	54.77225575	4.552284750
70	3	54.77225575	2.276142375
71	3	17.67766953	1.885618083
72	3	17.67766953	3.771236166
73	3	17.67766953	3.771236166
74	3	17.67766953	3.771236166
75	3	17.67766953	3.771236166
76	3	17.67766953	3.771236166
77	3	17.67766953	3.771236166
78	3	17.67766953	3.771236166
79	3	17.67766953	3.771236166
80	3	17.67766953	3.771236166
81	3	27.38612788	5.841619253
82	3	27.38612788	6.222238329
83	3	27.38612788	4.066930952
84	3	27.38612788	4.520207766
85	3	27.38612788	5.997850602
86	3	27.38612788	6.632054515
87	3	27.38612788	4.976702199
88	3	27.38612788	4.867042053
89	3	27.38612788	5.575760139
90	3	27.38612788	5.079001964
91	3	27.38612788	22.76142375
92	3	27.38612788	20.48528137

Table 5.3: **Characteristics of OU simulations 65 - 92**
Table of how the OU simulations (67-93) vary, including the number of points of attraction, the long term standard deviation about these points and the average distance between points of attraction.

OU number	Number of points of attraction	Long term standard deviation	Average distance between points of attraction
93	3	27.38612788	18.20913900
94	3	27.38612788	15.93299662
95	3	27.38612788	13.65685425
96	3	27.38612788	11.38071187
97	3	27.38612788	9.1045695
98	3	27.38612788	6.828427125
99	3	27.38612788	4.55228475
100	3	27.38612788	2.276142375
101	1	0.1	0
102	1	0.3	0
103	1	0.5	0
104	1	0.7	0
105	1	0.9	0
106	1	1.1	0
107	1	1.3	0
108	1	1.5	0
109	1	1.7	0
110	1	1.9	0
111	5	13.69306394	6.178002262
112	5	13.69306394	4.563538614
113	5	13.69306394	3.473467008
114	5	13.69306394	5.326197512
115	5	13.69306394	3.219039168
116	5	13.69306394	5.891979101
117	5	13.69306394	6.526701595
118	5	13.69306394	4.674765813
119	5	13.69306394	6.039779245
120	5	13.69306394	3.711774791

Table 5.4: **Characteristics of OU simulations 93 - 120**
Table of how the OU simulations (94-120) vary, including the number of points of attraction, the long term standard deviation about these points and the average distance between points of attraction.

OU number	Number of points of attraction	Identified sites from threshold criterion (65%)	Identified sites from stablility criterion	Radius from threshold criterion	Radius from stability criterion	Colour category
11	2	2	2	0.4	0.4	Green
12	1	1	1	0.5	0.5	Green
13	2	2	2	0.4	0.4	Green
14	5	5	5	0.3	0.3	Green
15	9	7	2	1.1	0.3	Red
16	3	3	3	0.5	0.5	Green
17	7	7	7	0.4	0.4	Green
18	9	5	1	0.9	2	Red
19	8	7	1	0.7	1.8	Red
20	6	6	6	0.6	0.3	Amber
21	2	2	1	0.2	1.3	Red
22	8	1	1	3.5	1.6	Amber
23	8	8	1	1	2	Red
24	1	1	1	0.9	0.9	Green
25	7	5	3	1.2	1.5	Red
26	5	5	5	0.4	0.4	Green
27	2	2	2	0.3	0.3	Green
28	4	4	4	0.5	0.5	Green
29	5	3	2	1.1	1.7	Red
30	9	6	2	0.9	1.9	Red
31	6	4	1	1.2	1.9	Red
32	3	3	3	0.4	0.4	Green
33	3	3	3	0.2	0.4	Amber
34	4	6	2	0.4	1	Red
35	3	3	3	0.4	0.4	Green
36	5	5	3	0.5	1.4	Red
37	7	7	6	0.4	1.1	Red
38	7	5	1	0.6	2.3	Red

Table 5.5: **OU simulations 11 - 38 results**
The results from applying the Sites of Interest algorithm to OU simulations 11 - 38.

OU number	Number of points of attraction	Identified sites from threshold criterion (65%)	Identified sites from stability criterion	Radius from threshold criterion	Radius from stability criterion	Colour category
39	6	5	6	0.8	0.4	Red
40	4	5	4	0.9	1.2	Red
41	5	5	5	0.3	0.3	Green
42	1	1	1	0.8	0.8	Green
43	9	8	1	1	2.7	Red
44	2	2	2	0.3	0.3	Green
45	7	8	3	0.9	1.7	Red
46	7	15	1	0.5	1.2	Red
47	4	4	4	0.5	0.5	Green
48	5	5	1	0.4	2	Red
49	7	6	6	1.1	0.8	Amber
50	1	1	1	0.6	0.6	Green
51	3	3	3	0.5	0.5	Green
52	3	3	3	0.5	0.5	Green
53	3	3	3	0.7	0.7	Green
54	3	3	3	0.8	0.8	Green
55	3	3	3	0.8	0.8	Green
56	3	3	3	0.6	0.6	Green
57	3	3	3	0.8	0.8	Green
58	3	3	3	1	0.6	Amber
59	3	3	3	1.2	1.2	Green
60	3	3	3	0.7	0.7	Green
61	3	3	3	1	1	Green
62	3	3	3	1.1	1.1	Green
63	3	3	3	0.9	0.9	Green
64	3	3	3	0.8	0.8	Green
65	3	3	3	0.9	0.9	Green
66	3	3	3	1	1	Green

Table 5.6: **OU simulations 39 - 66 results**
The results from applying the Sites of Interest algorithm to OU simulations 39 - 66.

OU number	Number of points of attraction	Identified sites from threshold criterion (65%)	Identified sites from stablility criterion	Radius from threshold criterion	Radius from stability criterion	Colour category
67	3	3	3	0.7	0.7	Green
68	3	3	3	0.8	0.8	Green
69	3	3	3	0.7	0.7	Green
70	3	22	3	0.3	0.8	Red
71	1	1	1	0.8	0.8	Green
72	2	2	2	0.7	0.7	Green
73	3	3	3	0.3	0.3	Green
74	4	4	4	0.3	0.3	Green
75	5	5	5	0.3	0.3	Green
76	6	6	6	0.3	0.3	Green
77	7	7	7	0.3	0.3	Green
78	8	8	8	0.7	0.7	Green
79	9	9	9	0.7	0.7	Green
80	10	10	10	0.4	0.4	Green
81	3	3	3	0.6	0.6	Green
82	3	3	3	0.5	0.5	Green
83	3	3	3	0.5	0.5	Green
84	3	3	3	0.7	0.7	Green
85	3	3	3	0.6	0.6	Green
86	3	3	3	0.7	0.7	Green
87	3	3	3	0.7	0.7	Green
88	3	3	3	0.7	0.7	Green
89	3	3	3	0.6	0.6	Green
90	3	3	3	0.6	0.6	Green
91	3	3	3	0.3	0.3	Green
92	3	3	3	0.4	0.4	Green
93	3	3	3	0.4	0.4	Green

Table 5.7: **OU simulations 67 -93 results**
The results from applying the Sites of Interest algorithm to OU simulations 67 - 93.

OU number	Number of points of attraction	Identified sites from threshold criterion (65%)	Identified sites from stablility criterion	Radius from threshold criterion	Radius from stability criterion	Colour category
94	3	3	3	0.3	0.3	Green
95	3	3	3	0.5	0.5	Green
96	3	3	3	0.5	0.5	Green
97	3	3	3	0.4	0.4	Green
98	3	3	3	0.5	0.5	Green
99	3	3	3	0.3	0.3	Green
100	3	3	3	0.3	0.3	Green
101	1	1	1	0.4	0.4	Green
102	1	1	1	0.4	0.4	Green
103	1	5	1	0.3	0.6	Red
104	1	7	1	0.4	0.9	Red
105	1	6	1	0.4	1	Red
106	1	6	1	0.6	1.2	Red
107	1	7	1	0.7	1.5	Red
108	1	12	1	0.7	1.3	Red
109	1	6	6	1	1	Green
110	1	8	1	0.9	2	Red
111	5	5	5	0.4	0.4	Green
112	5	5	5	0.4	0.4	Green
113	5	5	3	0.6	1.3	Red
114	5	5	5	0.5	0.5	Green
115	5	5	5	0.3	0.3	Green
116	5	5	5	0.5	0.5	Green
117	5	5	5	0.5	0.5	Green
118	5	5	5	0.7	0.7	Green
119	5	5	5	0.4	0.4	Green
120	5	5	5	0.5	0.5	Green

Table 5.8: **OU simulations 94 -120 results**
The results from applying the Sites of Interest algorithm to OU simulations 94 -120.

Cattle path	Identified sites from threshold criterion (50%)	Identified sites from stability criterion	Radius from Threshold	Radius from stability criterion	Colour category
1.1	3	3	11.4	11.4	Green
1.2	2	2	14	16	Amber
1.3	2	2	12	12	Green
1.4	1	1	10	11	Amber
1.5	24	7	12	36	Red
1.6	2	2	10	13	Amber
1.7	2	7	10.5	20	Red
2.1	1	1	15	15	Green
2.2	2	3	28	18	Red
2.3	1	1	23	18	Amber
2.4	1	1	10	13	Amber
2.5	2	2	10	12	Amber

Table 5.9: **Results from applying Sites of Interest Algorithm to cattle data**
The results from applying the algorithm to cattle data for seven paths.

Radius (m)	Path 1.1	Path 1.2	Path 1.3	Path 1.4	Path 1.5	Path 1.6	Path 1.7
10	02:37.0	01:56.6	01:37.8	03:07.0	02:16.4	02:44.2	01:00.5
20	01:21.3	01:00.5	00:51.1	01:17.8	01:05.6	01:15.6	00:31.2
30	00:49.6	00:37.8	00:29.9	00:41.3	00:40.6	00:39.4	00:20.8
40	00:27.5	00:25.8	00:23.6	00:22.8	00:25.9	00:23.5	00:17.4
50	00:20.5	00:17.7	00:14.1	00:18.1	00:19.0	00:17.6	00:12.5
60	00:15.4	00:15.0	00:10.5	00:12.4	00:13.2	00:13.7	00:09.2
70	00:12.4	00:10.8	00:09.4	00:10.4	00:11.3	00:10.9	00:09.2
80	00:10.4	00:08.4	00:07.3	00:06.8	00:09.0	00:08.6	00:07.5
90	00:08.1	00:06.7	00:07.7	00:06.3	00:08.5	00:06.9	00:07.0
100	00:06.7	00:05.7	00:06.7	00:05.6	00:07.8	00:05.8	00:05.0
Average	00:38.9	00:30.5	00:25.8	00:38.9	00:33.7	00:36.6	00:18.0

Table 5.10: **Runtime of Sites of Interest Algorithm applied to cattle paths (with $s = 10$)**
The runtimes (min:sec) from applying the algorithm to cattle data for seven paths and using $s = 10$, which is the length of interval that the algorithm looks over.

Radius (m)	Path 1.1	Path 1.2	Path 1.3	Path 1.4	Path 1.5	Path 1.6	Path 1.7
10	14:22.5	20:54.4	16:24.3	29:45.8	21:28.4	26:03.4	18:42.4
20	06:47.2	09:43.0	07:48.1	13:04.8	10:26.6	12:14.4	08:38.2
30	04:25.4	05:49.5	04:55.1	06:15.9	06:44.9	06:45.1	05:36.9
40	03:05.4	03:58.8	03:35.0	03:19.1	04:26.7	04:07.7	03:54.9
50	02:28.1	02:34.1	02:18.2	02:30.5	03:06.3	02:34.1	02:49.4
60	02:00.8	01:58.3	01:48.1	01:43.4	02:17.4	01:58.0	02:01.4
70	01:39.8	01:41.3	01:34.6	01:33.3	01:42.8	01:34.7	01:39.4
80	01:24.9	01:22.5	01:10.7	01:05.5	01:18.9	01:17.7	01:31.4
90	01:13.9	00:53.2	01:10.2	00:58.1	01:09.6	01:08.8	01:18.1
100	00:58.7	00:43.5	01:06.0	00:51.6	01:05.8	00:57.4	01:04.6
Average	03:50.7	04:57.9	04:11.0	06:06.8	05:22.7	05:52.1	04:43.7
Ratio between averages for s=1 and 10	5.93	9.77	9.73	9.43	9.57	9.62	15.76

Table 5.11: **Runtime of Sites of Interest Algorithm applied to cattle paths (with $s = 1$)**
The runtimes (min:sec) from applying the algorithm to cattle data for seven paths and using $s = 1$, which is the length of interval that the algorithm looks over.

Appendix C: SitesInterest R Package User Guide

Here, I include code with examples of the different functions contained in the `R` package, `SitesInterest`, which is meant as a user friendly tool for identifying sites of interest. The package is available on `CRAN`, where further information can be found in the package reference manual (`https://cran.r-project.org/web/packages/SitesInterest/index.html`).

```
##Find the current working directory
wd = getwd()
##Set the working directory as the temporary one
setwd(tempdir())
##Load the data
data(OU_14)
t=unlist(OU_14['t'])  #time series of recording times
X=unlist(OU_14['X']) #time series of X coordinate locations
Y=unlist(OU_14['Y']) #time series of Y coordinate locations


###Single Trajectory###

#Example 1 - plot.displacement
##Change the class of X and Y
class(X) = "displacement"
class(Y) = "displacement"

##Plot the displacement from  t=0 to t=2.9999  with
## a line indicating a displacement of 0.3
plot(X, Y, "OU14", t, 0.3, 0, 0, 2.9999)


#Example 2 - Alt_Alg
##Run the Sites of Interest Algorithm for multiple radii values. Note
##also that we do not allow the first circle to be the only site. The
```

```
##results are also saved in csv files
Radii=seq(0.2,1.0,0.1) #set of radii values
for (R in Radii){
        Alt_Alg("OU_14",t,X,Y,R,first='y',save='y')}


#Example 3 - Sites
##Calculate all the necessary information to be used elsewhere. This can
##only be used when the csv files from using Alt_Alg have been saved.
Sites("OU_14",0.3,first='y')
##This includes:
## - sites_index which is an array of the indices of the identified sites
##   among all circles
## - N_no_overlap which is the number of non-overlapping circles
## - X_no_overlap which is an array of the x-coordinates of all
##   non-overlapping circles
## - Y_no_overlap which is an array of the y-coordinates of all
##   non-overlapping circles
## - X_sites which is an array of the x-coordinates of the site centres
## - Y_sites which is an array of the y-coordinates of the site centres
## - max_percent_drop is the maximum percent drop
## - number_identified_sites is the number of identified sites
## - psi_sort_no_overlap2 is an array of non-overlapping usage times
##   sorted into descending order

#Example 4 - plot_bar_chart
##Plot the bar chart of ranked non-overlapping usage times. This can
##only be used when the csv files from using Alt_Alg have been saved.
plot_bar_chart("OU_14",0.3,first='y')

##It is possible to choose manually where the cut off between sites and
##non-sites should be
plot_bar_chart("OU_14",0.3,first='y' number_sites=4)

##The colours used in the plot can also be changed
plot_bar_chart("OU_14",0.3,first='y',colours=c("darkgreen","red"))


#Example 5 - plot.hoops
##Change the class of X and Y
class(X) = "hoops"
class(Y) = "hoops"

##Plot the positions of the identified sites as well as the non-
```

```
##overlapping circles. This can only be used when the csv
##files from using Alt_Alg have been saved.
plot(X,Y,"OU_14",0.3)

##The colours for the hoops can be changed
plot(X,Y,"OU_14",0.3,first='y',colours=c('tan','chocolate','maroon'))

##The thickness of hoops can also be changed
plot(X,Y,"OU_14",0.3,first='y',lwds=c(0.5,2,3.5))

##It is also possible to choose manually the number of sites
plot(X,Y,"OU_14",0.3,first='y',number_sites=4)


#Example 6 - plot_bars_and_hoops
##Change the class of X and Y.
class(X) <- "bars_and_hoops"
class(Y) <- "bars_and_hoops"

##Plot the bar chart of ranked non-overlapping usage times and the
##plot showing the positions of these circles. This can only be used
##when the csv files from using Alt_Alg have been saved.
plot_bars_and_hoops(X,Y,"OU_14",0.3,first='y')


#Example 7 - plot_radii_results
##Plot the results (i.e. maximum percent drop and number of sites
##identified) from applying the algorithm with various radius values.
##This can only be used when the csv files from using Alt_Alg have
##been saved.
plot_radii_results("OU_14",Radii,first='y')

##The location of the legend can be changed.
plot_radii_results("OU_14",Radii,first='y',legend_loc='bottomleft')


#Example 8 - print_colour_assignment
##Print a summary table of the results from the two criteria (threshold
## and stability)  and the colour assigned. Here, a threshold
##value of 65% is used. This can only be used when
##the csv files from using Alt_Alg have been saved.
print_colour_assignment("OU_14",65,Radii,first='y')


```

```
#Example 9 - print_sites_pos
##Print the coordinates of the centres of all identified sites. This can
##only be used when the csv files from using Alt_Alg have been saved.
print_sites_pos("OU_14",0.3,first='y')

##There is also the option of saving the results as a csv file
print_sites_pos("OU_14",0.3,first='y',save='y')


#Example 10 - print_site_visits
##Prints a summary of the site visitation results. This can only be
##used when the csv files from using Alt_Alg have been saved.
print_site_visits("OU_14",X,Y,0.3,first='y')

##There is also the option of saving the results as a csv file
print_site_visits("OU_14",X,Y,0.3,first='y',save='y')


#Example 11 - plot.schematic
##Change the class of X and Y.
class(X) = "schematic"
class(Y) = "schematic"

##Plot the schematic representation of movements between sites
plot(X,Y,"OU_14",0.3,first='y')

##There is also the option to make changes to:
## - the length of the arrow head
plot(X,Y,"OU_14",0.3,first='y',len_arrow=0.25)

## - the thickness of the arrow
plot(X,Y,"OU_14",0.3,first='y',lwd_arrow=2)

## - the thickness of the hoops
plot(X,Y,"OU_14",0.3,first='y',lwd_r=2)

## - the size of the arrow labels
plot(X,Y,"OU_14",0.3,first='y',text_size=2)

## - the location of the legend
plot(X,Y,"OU_14",0.3,first='y',legend_loc='bottomleft')


```

```
159
160
161 ###Group/Discontinuous Data###
162
163 #Example 12 - Alt_Alg_discont
164 n=5 #Number of path sections
165
166 N = length(t) #Number of recorded locations
167
168 ##Here, the trajectory is split into 5 segments, but only 3 will be used.
169 ##A list of arrays of the time recordings for the 3 trajectory segments
170 t_all = list(t[seq(1,floor(N/n))],t[seq(floor(N/n)*2,floor(N/n)*3)],
171         t[seq(floor(N/n)*4,floor(N/n)*5)])
172
173 ##A list of arrays of the x-coordinates for the 3 trajectory segments
174 X_all = list(X[seq(1,floor(N/n))],X[seq(floor(N/n)*2,floor(N/n)*3)],
175         X[seq(floor(N/n)*4,floor(N/n)*5)])
176
177 ##A list of arrays of the y-coordinates for the 3 trajectory segments
178 Y_all = list(Y[seq(1,floor(N/n))],Y[seq(floor(N/n)*2,floor(N/n)*3)],
179         Y[seq(floor(N/n)*4,floor(N/n)*5)])
180
181 ##The calculation of the usage time for discontinuous data
182 Alt_Alg_discont("OU_14_discont",c("OU_14.1","OU_14.3","OU14.5"),t_all,
183         X_all,Y_all,0.3,save='y')
184 ##where "OU_14_discont" is the name of the discontinuous path
185
186 ##It is possible to use each of the functions contained in Alt_Alg_discont
187 ##separately:
188 ## - Alt_Alg_mini runs the algorithm across a particular path and a
189 ##   particular set of circles, which may not necessarily lie on the path
190 ## - Combining collects all of the usage times for the same circle from
191 ##   different paths together
192
193 ##Reset the original working directory
194 setwd(wd)
```

Index