

# **Mathematical Analysis of Fine-Scale Badger Movement Data**

**Jessica Rachel Furber**

PhD Confirmation Report submitted to the University of Surrey

*Department of Mathematics  
University of Surrey  
Guildford GU2 7XH, United Kingdom*



Copyright © 2022 by Jessica Rachel Furber. All rights reserved.

E-mail address: [j.furber@surrey.ac.uk](mailto:j.furber@surrey.ac.uk)



## **Scientific abstract**

The purpose of the current study is to establish a modelling framework to understand the movement of European badgers (*Meles meles*) within their environment. It is widely understood that badgers play a crucial role in the transmission of bovine tuberculosis (bTB), but the exact role is still not clear. An approach to modelling animal movement is using energy potentials, primarily seen in physics to describe particle motion, and geometric Brownian motion. Such an approach has been successful in describing the movements of free ranging Mountain elk (*Cervus elaphus*) and their avoidance of vehicles and humans but has yet to be applied to the study of badgers.

An exploratory modelling framework has been built that describes how badgers move around their landscape foraging, defaecating, and interacting with other badgers (including male and female differences). The model incorporates general badger behaviour, badger social interaction, and random walk/individual behaviour. The framework is extended through the use of data driven methods to parametrize the model using GPS data and interpret the level of noise within the system.

Key questions this modelling framework will allow us to answer are: How does badger movement affect the spread of bTB? How does the climate affect badger movement? With the role badgers play in the spread of bTB, the answer to these questions could be crucial in the role of strategy planning.

In this report, we present preliminary results on the fitted/parameterized model and explore some of the key modelling questions.

Keywords and AMS Classification Codes: coherent sets; data science; energy potential; KDE; movement ecology; stochastic models; transfer operators



## **Acknowledgements**

I give thanks to the University of Surrey's Department of Mathematics and the School of Veterinary Medicine for their support in the journey so far. I am extremely grateful to my supervisors, Prof. Mark Chambers, Prof. David Lloyd, Prof. Gianni Lo Iacono, and Prof. Stefan Klus for their invaluable advice, continuous support, and patience. I also wish to thank my family and friends for their continued support in my new adventure.



## Contents

1. Introduction . . . . .	1
1.1. Literature Review . . . . .	2
1.2. Outline of Report . . . . .	5
2. Methods . . . . .	7
2.1. Badger Moving in an Energy Potential . . . . .	7
2.2. Stochastic Differential Equations . . . . .	9
2.2.1. Trajectory in the Energy Potential . . . . .	9
2.2.2. Equations of Motion . . . . .	10
2.3. Large Deviation Theory . . . . .	16
2.4. Data Methods . . . . .	17
2.4.1. Kernel Density Estimation . . . . .	17
2.4.2. Kramers-Moyal Formulae . . . . .	21
2.4.3. Transfer Operators . . . . .	25
3. Modelling Strategies . . . . .	29
3.1. Badger Model . . . . .	29
3.2. Ecology Model . . . . .	32
3.3. Infection Model . . . . .	34
4. Badger Data . . . . .	37
4.1. Overview of Woodchester Park . . . . .	37
4.2. Preliminary Data Analysis . . . . .	42
4.2.1. Logistic Regression . . . . .	42
4.2.2. Weather Data . . . . .	43
4.3. Data Wrangling Issues . . . . .	45
4.4. Map and QGIS Data . . . . .	47
5. Generation of Badger Model with Data . . . . .	51
5.1. Generation of Energy Potential . . . . .	51
5.2. Calculating the Noise . . . . .	54
5.3. Location of Mountain Passes . . . . .	56
5.4. Simulation in the Energy Potential . . . . .	58
6. Conclusions and Outlook . . . . .	63
Appendices . . . . .	69
A. Methods . . . . .	69
A.1. Euler-Maruyama . . . . .	69

B. Logistic Regression . . . . .	71
C. Conclusions and Outlook - Gantt Chart . . . . .	73
D. Data Management Plan . . . . .	75
E. Seminar Hours and Extended Abstracts . . . . .	79
References . . . . .	87

---



# 1

## Introduction

Modelling animal movement has gained greater interest in the last couple of decades and has been researched in a variety of settings, where there has been a steady increase in the amount of published articles in movement ecology from the late 1900s [21]. Primarily, the movements of animals have been studied in order to understand ecological problems, such as animal migration and swarming [19, 25]. Within this work, we are wanting to understand the movements of European badgers (*Meles meles*), along with how the weather influences their movements. This is to aid in the understanding of the spread of bovine tuberculosis (bTB). The disease bTB, caused by *Mycobacterium bovis*, is a serious disease of cattle and has a significant economic impact on farmers. In the British Isles, its management in livestock is complicated by the role of badgers [49]. Although cattle are the principal hosts for *M. bovis*, wild animals, such as the European badger, may act as reservoirs of infection [6], which poses as an obstacle to eradicate the disease. Additionally, due to the natural activities of the badgers (such as foraging, defecation, and urination) then there is a possibility for the environment to play a role in the transmission of the disease [50, 20, 28], but this is not yet clear and needs to be studied further.

*M. bovis* was first identified in badgers in the south-west of England in 1971 [33], and since then, there have been lots of studies to understand and monitor the disease. One such study began in 1976 in Gloucestershire [10], with principal areas of work including TB epidemiology in badger populations. Understanding the ecology and social behaviour of badgers is important in an attempt to understand the transmission and interaction with cattle. Badgers are known territorial nocturnal mammals that live in social groups in underground setts that are located throughout their territory. During the night, they socialize with one another and also forage for food, with a predominant diet of earthworms [5], that are often found in pastures that cattle use. However, when food is scarce, badgers may supplement their diet with cattle feed. Thus, suggesting that there are locations that are shared by cattle that badgers are attracted to. This risk of close interaction is important in the transmission cycle of *M. bovis* between badgers and cattle. Hence, reiterating the importance

of understanding the movements of badgers.

One of the main research questions that we hope to answer is, can we identify coherent badger behaviour at fine scales? For example, do badgers remain as a group over time, or do they disperse. This question will help in the understanding of badger social behaviour, as well as being relevant for the spread of bTB. Another question that we hope to answer is, can we calculate transition times of badgers between social groups? Here, transition times relate to how long a badger will be in one social group before visiting another, before returning to the original sett. Again, being relevant to help understand the spread of bTB. Finally, we would like to answer a two part question; can we generate a dynamical model to explore badger movement and, how does behaviour change based on weather/season/climate? Models of badger movement do exist, such as the agent based model developed by Smith and Budgey (2001) [45], however, there are still questions unanswered that we hope to uncover. Through the use of data-driven methods, learning governing equations and transfer operators, we intend to create such a model that describes the evolution of the position of badgers over time from real GPS badger tracking data.

## 1.1. Literature Review

Mathematically modelling the movement of animals is a notion that has been developed substantially with the advancement of technology and machine learning. Siniff and Jessen state that ‘ecology will be greatly advanced if the perfection of simulation will allow evaluation of interactions within species or among species’ [44]. Their article is an example where telemetry data (a way of recording and transmitting data from a remote or inaccessible source to an IT system in a different location via radio, satellite, etc.) has been used to monitor animal movement (such as the red fox) to gain some understanding of the patterns that are produced. Yet, methods for the collection of data have developed vastly over the decades to allow for further advancements in models, such as collecting the data through the Global Positioning System (GPS), allowing for more accurate results.

Brillinger et al. [2] is another example where the advancement of technology has enhanced their mathematical model. They firstly considered in 2001 (and developed in later papers) free-ranging elk (*Cervus elaphus*) that have had radio collars fitted (Loran-C receivers) in a large, fenced experimental forest, with the aim to use the trajectories that have been simulated to estimate an energy potential function (described in Chapter 2). This function is then incorporated into a stochastic differential equation (SDE) to help understand the movements of the elk. When creating their potential surface they consider two factors: firstly the attraction towards unknown foraging and resting areas, and secondly, the reaction to human disturbance. A key difference between what they modelled and what we would like to model is they collected their data where boundaries have been enforced for the animals. Badgers do not take into account fences or farm boundaries, so

when considering the energy potential for our data, we will not necessarily observe such boundaries. Additionally, we are not specifically looking for the reaction of badgers to regions that see human disturbances, but this has been investigated in Ireland [17]. If there were such disturbances in the data, then the potential surface should represent this. Thus, the advancement of technology has allowed for such techniques in modelling animals, whilst taking into account features of the ecology.

Whilst the use of an energy potential has been seen to work in modelling animal movements by Brillinger et al., they have also identified a weakness [2]. The weakness, however, is associated with the quality of the data. Specifically, it was noted irregular time lags within the data, which could have an impact on the energy potential because the true behaviour of the animal is not seen. The collected data sometimes had one to four hours apart in observations for one animal. A reason for this could be due to the transmitter not detecting the animal's movement at the set capture time. For example, there has been research to show that forest canopy is a limiting factor for the accuracy of GPS measurements [35]. In the data set for the elks [2], it appeared that one elk had 'jumped' the boundary, where in reality, it would have had to go round. Unfortunately, the data does not show the movement in between the time captures, and hence, this could limit the energy potential and the model because not all locations are recorded. To overcome this, regular time captures for the data is necessary to determine true points of attraction/repulsion.

It has been stated by Schick et al. how the studies of animal movement generally have three lines of investigation [41]; modelling for realistic movement, animal-environment interactions, and inferring movement when the data are incomplete. Preisler et al. [38] suggest that through the use of a stochastic model with an incorporated energy potential function, they are able to study animal movement whilst addressing all these lines of investigation. For example, SDEs can be used to statistically model the ecology of animal movement, and the potential function surfaces can have underlying biological meaning. Whilst they do admit that it is difficult to derive ecological inferences due to the multi state, stochastic process, it is stated that SDEs derived from potential functions enhance the discussed approaches.

As it was mentioned in Presisler et al. [38], the development of GPS data has stimulated many new approaches for analysing modelling movement patterns of free-ranging animals. In essence, we need to ask what we would like from our energy potential? It is important to understand the concepts of territory and home range when answering this. The home range is considered to be the area that is used by any or all members of the group in the course of their normal activities, such as sleep in the sett. On the other hand, the territory relates to the area of the land which is used and defended. These concepts have been discussed for over a century, whereby when Burt [3] published an article on animal territories and home ranges in 1943 his latest Webster's dictionary (published in 1938) did not list a definition for the 'home range'. Ideally, the data collected will cover the territories of the badgers, hence giving a clear perception of what we would hope to see

in our energy potential.

There are multiple methods that have been used throughout the literature to delineate the home range of an animal. One of the first home range estimators developed in ecology was the minimum convex polygon [31]. This method draws the smallest polygon around points with all interior angles less than 180 degrees. Another method that has been developed is kernel density estimation (KDE) [43]. This method makes use of kernel smoothing for a utilization distribution, i.e. a probability distribution that gives the probability density of an animal being found at a given point in space. Downs and Horner (2009) [14] recognize that these two methods in particular have been traditionally favoured for home range analysis, but both have come under criticism. Primarily for the fault of overestimating the home range. The minimum convex polygon method overestimates by including area that is not used by the animal. Similarly, KDE method has been seen to overestimate the home range areas [24], when the chosen bandwidth is too small. The bandwidth is an important parameter that is a real positive number, and defines the smoothness of the density plot. So, when the bandwidth is too large, the smoother the plot becomes and features are lost. On the other hand, if the bandwidth is too small, then the plot can be rigid and not be realistic of the area the animal uses. Yet, within a review of methods by Walter et al. [48], KDE is both lauded and criticized. It is lauded for its use with GPS technology yet criticized for the errors in bandwidth selection. It is worth noting that different methods are appropriate for different scenarios. For example, KDE does not effectively represent animals that migrate several kilometres or a species that cover large areas. Fortunately, badgers are very local animals, and typically do not travel miles from their homes. When their circumstances change, then they are likely to move [20], such as when their population is perturbed in some way.

Silva et al. [42] investigate the choice of traditional estimators against another method called dynamic Brownian bridge movement models. The technique is intended to be used for GPS telemetry, which allows for efficient and repeatable analysis of high-resolution data. It is suggested in [42] that it is particularly useful for animals with behaviourally distinct movement patterns, such as the movements of reptiles. Additionally, when estimating the bandwidth for KDE they focus on two selection algorithms: reference bandwidth ( $h_{ref}$ ) and Least-Squares Cross-Validation (LSCV), which are commonly used within reptile studies [42]. It was shown that the algorithm  $h_{ref}$  overestimates areas, whilst LSCV tends to underestimate. Nevertheless, LSCV was preferred for calculating the bandwidth in comparison to  $h_{ref}$ . But this highlights the importance in choosing the right bandwidth, as the result will either swing one way or the other in terms of estimation.

Downs, Horner and Tucker [15] presented a new technique of animal home range analysis in 2011 called time-geographic density estimation. This method combines methodologies of time geography, which is a powerful concept for understanding human spatial behaviour [29], and statistical density estimation to create a continuous probability distribution of an object's spatial position over

time. Downs et al. in a later paper in 2018 [13] compare the accuracy of time-geographic density estimation against KDE and another method called characteristic-hull polygons. Characteristic-hull polygons vary from minimum convex polygons since they can have non-convex edges, as well as empty holes within their interiors [14], whereas minimum convex polygon include all areas (leading to the overestimation of the home range). Whilst it is concluded in [13] that time-geographic density estimation does come out best in comparison to other methods, KDE still remains a popular method due to it being easy to implement and is well studied (despite the drawbacks). With this conclusion, it appears KDE is a good starting point to generate an energy potential to study. On top of these state-of-the-art methods, we will be combining them with other mathematical tools such as transfer operators and data-driven methods to improve the results and a model.

## 1.2. Outline of Report

Within this report, we shall take a journey through the methods that will be used to creating a model with data. Firstly, in Chapter 2, we outline the methods to be used. We begin by discussing the concept of a badger in an energy potential and what this potential means for us. We progress by discussing large deviation theory, which will be used to identify mountain passes in our potentials. Then, we define our stochastic differential equation and explore trajectories in the potential and other equations of motion. We finish by going into detail of the data methods, including kernel density estimation, Kramers-Moyal Formulae and transfer operators. These methods are important when it comes to generating different aspects of our model.

Chapter 3 is an exploratory stage where different strategies for modelling are discussed. We begin by considering five badgers in a double-well potential with an added behaviour potential. This behaviour potential represents attraction and repulsion between the badgers, which is to be discussed in Chapter 2. This is progressed by taking ecological assumptions into the model, such as the attraction between males and females. Finally, the model is adapted to include an infection model, however this is to focus on technique rather than results.

In Chapter 4 we explore the obtained data. Primarily, we have two sets of data: one from the Animal and Plant Health Agency (APHA) and another from the Zoological Society of London (ZSL). Within the report, we focus on and apply the methods in Chapter 2 to the APHA data. We discuss how the data are collected, including a visit to the location. Preliminary data analysis is conducted on a subset of capture-mark-recapture data and weather data. This is followed by issues that we have had with data wrangling, including the cleaning of the data and transforming the coordinate units. The chapter ends with the plotting of the data in a data plotting tool called QGIS, where k-clustering of the data is also considered.

In Chapter 5 the methods outlined in Chapter 2 are employed on the data from APHA. We begin

with generating an energy potential from the data observed in Chapter 4, where three estimation algorithms are used to compare different bandwidth options. The noise is then estimated for the badgers in two different ways. Firstly, the noise for each badger is calculated separately, and then an average is taken. Secondly, the noise is estimated using the full data set. In comparison, they are approximately similar. To further investigate the noise, the data are split into metrological seasons. The chapter ends by initiating the investigation into large deviation theory and locating mountain passes, as well as generating some simulations using the found energy potential and estimated noise.

Finally, in Chapter 6 the work that has been completed in this report will be reflected on, and the future work of the project is discussed. This is accompanied by an outline of the future in the format of a Gantt chart.

# 2

## Methods

Within this chapter, the methodology that will be used for the report is outlined. This includes setting up of the model, and methods to be used to parametrize the model.

### 2.1. Badger Moving in an Energy Potential

Energy minimization is a classical concept in physics where objects move around an energy landscape trying to minimize their energy. Preisler, Ager and Wisdom provide a detailed explanation for the motivation of using energy potential functions for the movements of animals [38]. When used in mechanics, it is assumed that the particles are moving on a surface that contains points or regions of attraction and repulsion. When applying this to the movement of animals, the potential field can be seen to reflect the topographic surface under the influence of gravity. That is, objects are attracted to low points and repelled by high points. It is highlighted by Preisler, Ager and Wisdom that a 'foundation for this modelling framework is provided by a formal relationship between the potential surface and the velocity of the particle at a given location and time' [38]. The velocity of the particle at a given location on the surface is found by taking the negative of the gradient of the potential function at the given location. This is important, because whilst the potential may not be directly observable, the approximate velocity of the animal is. This allows us to estimate a potential function if we have observations of animal movements on consecutive locations. A benefit of having a potential energy allows for easy visualization and deeper understanding of the badger dynamics. Subsequently, complicated potential surfaces can also be estimated, since it is likely that some animals might be attracted to a specific region to forage (so we would see a deeper well in the potential), and may be repulsed by regions that see human disturbances.

To explain the basic ideas, we first look at a badger moving within a simple double-well potential

(Figure 2.1). The double-well potential is defined as

$$V(x_1, x_2) = (x_1^2 - 1)^2 + x_2^2, \quad (2.1)$$

where  $\mathbf{V} : \mathbb{R}^2 \rightarrow \mathbb{R}$ .

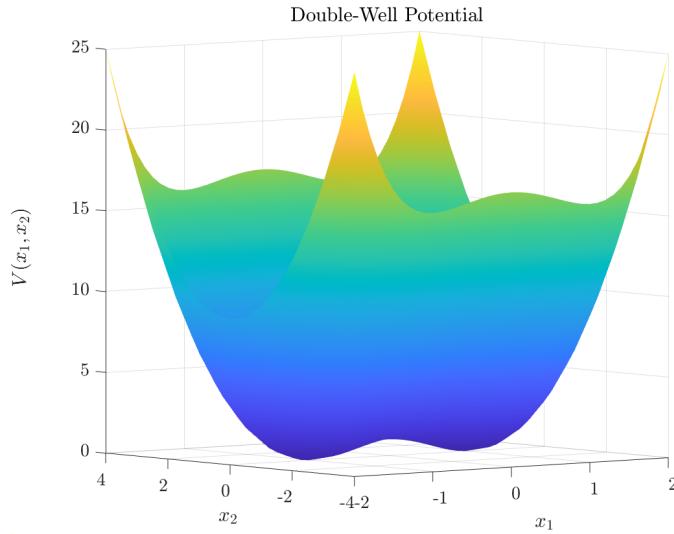


Figure 2.1: Double-well energy potential. The bases of the well are located at  $(x_1, x_2) = (-1, 0)$  and the other at  $(x_1, x_2) = (1, 0)$ .

To understand the meaning of the double-well for our context, one well is set as a badger territory and the second well as another badger territory that shares its border. The base of each well is the point of attraction, so in general a well could represent a lot of things for badgers, such as a sett, a mate (during mating season), latrine (used frequently to mark their territory), known food source (when they are foraging), etc. A particle (or badger, as it shall be referred to from here) is stuck in the potential and moves around the well, where it has low probability of moving where the energy increases for they want to minimize their energy, e.g., ball rolling downhill in a well. When noise is added to the system, the badger is more likely to move between the two wells, passing through the mountain pass. This noise can represent the variability of badgers, the need to go foraging and/or large excursions due to mating season. We are interested if a badger moves between territories, and therefore, has to overcome the energy barrier (i.e. the mountain pass) by expending energy. This is something we want to be able to calculate from the model.

## 2.2. Stochastic Differential Equations

A classic approach to modelling animal movement is through via geometric Brownian motion. Considering a stochastic differential equation (SDE) with a state space  $d$ , it takes the form

$$d\mathbf{X}_t = b(\mathbf{X}_t, t)dt + \sigma(\mathbf{X}_t, t)d\mathbf{W}_t, \quad (2.2)$$

where  $\mathbf{X} \in \mathbb{R}^d$  represents the animal of the particle at time  $t$ ,  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  represents the drift term,  $\sigma : M_d(\mathbb{R}) \rightarrow M_d(\mathbb{R})$  is a real valued matrix that represents the diffusion term, and  $\mathbf{W}_t$  a  $d$ -dimensional Wiener process. A Wiener Process (alternatively called Brownian motion) is a widely used random process and is used to describe a random, but continuous, motion of an animal. The assumption for the use of Brownian motion for the movement of badgers suggests that during simulation the badger will perform small random movements and is unlikely to perform large jumps.

The drift term,  $b$ , describes the directional movement of the badger based on a given landscape. The diffusion term,  $\sigma$ , also known as noise, allows for variability in the badgers, such as the need to go foraging. In particular, we are interested when the drift term takes the form,

$$b(x) = -\nabla V(x),$$

where  $V$  is an energy potential and  $\nabla$  is the gradient operator. An energy potential, in this scenario, could be the double-well potential, as seen in Figure 2.1, or an energy potential calculated from the data. It can be said that animals put in effort for a reward, where animals consider a multitude of ecological landscapes in addition to the physical energy when making movement decisions. Hence, we assume that badgers act based on reward. Thus, considering an energy potential as the drift term works well, as a badger needs to spend energy in order to cross a mountain pass to move to a different well and get its reward. In addition to the form of the drift term, we will be considering a diagonal matrix for  $\sigma$  (letting  $\mathbb{I}_n$  represent an  $n \times n$  identity matrix). For example, when considering  $d = 2$ , a diagonal  $\sigma$  would mean that there is the same noise in the  $x_1$  and  $x_2$  direction. This allows for simplicity when generating trajectories, but a diagonal matrix for the noise may not hold true when it comes to the data.

### 2.2.1. Trajectory in the Energy Potential

With a double-well energy potential established, we would like to add a badger into the system and analyse the behaviour for the different values of  $\sigma$ . It is expected to see more movement between wells for a larger sigma because this indicates a larger chance to move to the other well. Thus, the main purpose of this analysis is to test the code. The same initial condition,  $\mathbf{X}_0 = [0.6, 0.4]^T$

(chosen randomly), is used when considering the trajectories and the value of sigma will change.

Firstly, consider a random value for the noise,  $\sigma = 0.9\mathbb{I}_2$ . To obtain a trajectory, the Euler–Maruyama method is used for the given initial condition and sigma. See Appendix A for a description of the method. Figure 2.2 shows the plotted trajectory, where the left figure gives the trajectory of the  $x_1$  coordinate over time and the right image displays the trajectory from above. Clearly, the badger visits both wells, making the journey over the mountain pass three times.

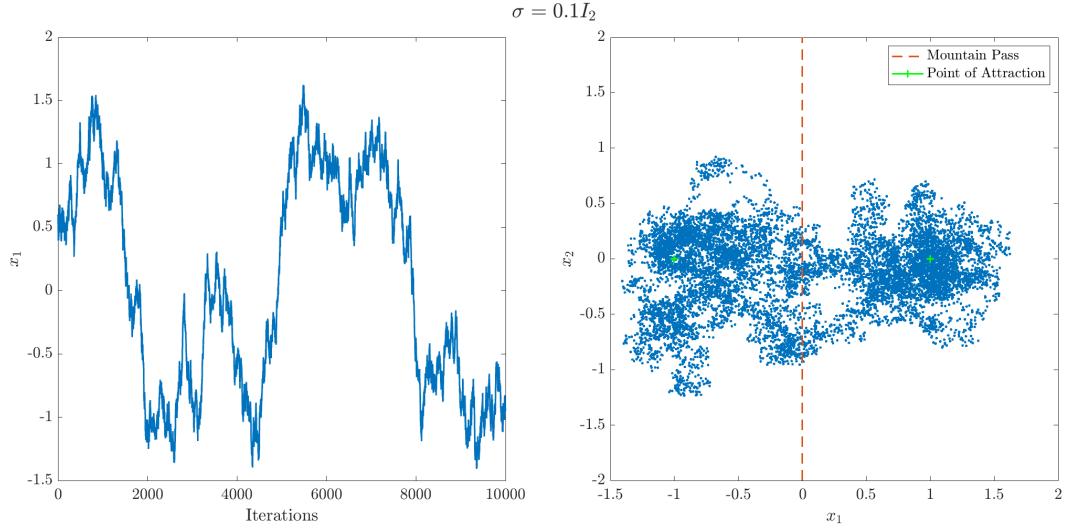


Figure 2.2: Generated trajectory in a double-well potential, with  $\sigma = 0.9\mathbb{I}_2$ . On the left, the  $x_1$  path is plotted against the iterations, so movement between the wells can be observed. On the right, the positions of the trajectory are plotted, alongside the coordinates of the mountain pass and the base of the wells (i.e. the point of attraction).

Next, the value of sigma will be decreased to  $\sigma = 0.1\mathbb{I}_2$  to represent a smaller noise, where we obtain the trajectory as seen in Figure 2.3. No movement between the two wells is observed, i.e. the badger remains in one territory. There is little movement within the territory centred around the point of attraction, but there is no pull to move to another territory. This could be considered very unrealistic when considering the movement of badgers, since it is known that, whilst they sometimes move far, it does not happen often. They leave their sett in order to use the latrine (found on the border of the territory), to forage for food and to find mates. Therefore, when considering this system for badgers, it is imperative that an appropriate value for the noise is considered.

### 2.2.2. Equations of Motion

Next, the attraction and repulsion of different points and badgers will be explored. So far, we have simulated a single badger moving between two territories. Badgers are social animals, so more

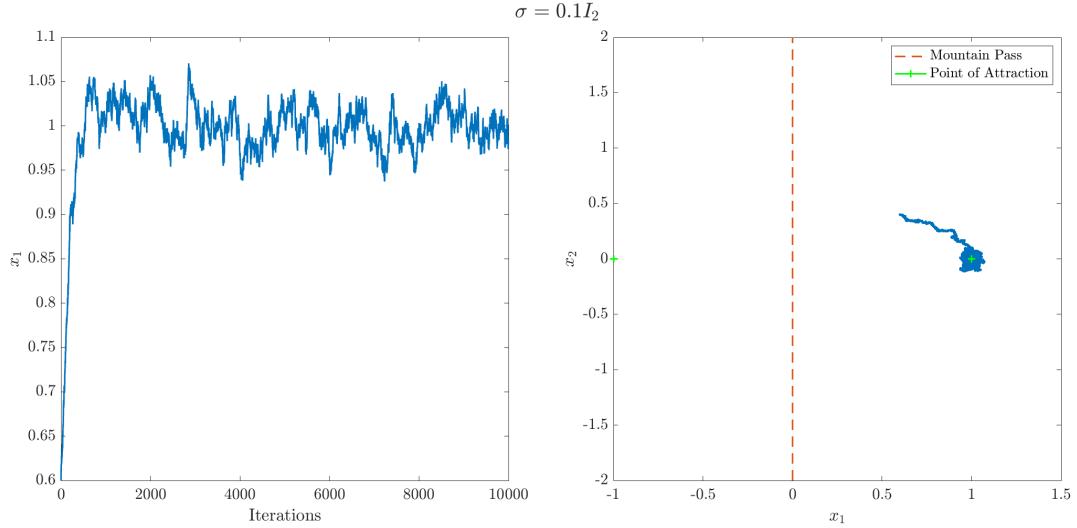


Figure 2.3: Generated trajectory in a double-well potential, with  $\sigma = 0.1\mathbb{I}_2$ . On the left, the  $x_1$  path is plotted against the iterations, so any movement between the wells would be observed. On the right, the positions of the trajectory are plotted. Clearly there are no movements between the two wells, the badger remains in the right well (attracted to the base  $(x_1, x_2) = (1, 0)$ ).

badgers should be introduced into the simulation and interaction terms need to be considered. It has been noted in [37] that modelling animal movements is one of the more challenging aspects of wildlife studies, and this should be taken into consideration when investigating the movements of the badgers. There is a question if the information about attraction and repulsion is obtainable from the data, but at this stage we are exploring the possibilities and techniques should they be employed.

So far, we have been considering the drift term in Equation (2.2) as the double-well potential function (Equation (2.1)). We would like to consider another type of potential function,  $\mathbf{H} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , where  $d$  is the state space. The potential function  $\mathbf{H}$ , similar to  $\mathbf{V}$ , is used to describe this force field at location  $\mathbf{X}$  and time  $t$  [37]. In our case, we would like to use this potential function  $\mathbf{H}$  to describe both attraction and repulsion. In order to implement this, we firstly consider the SDE without the double-well potential, and only simulate with one behaviour.

The attraction of a badger to a fixed point is the simplest case. This could, in our instance, represent a latrine, something badgers are naturally attracted to use to mark their territory. Our attraction potential function, when considered in a state space of  $d = 2$ , takes the form

$$\mathbf{H} = |\mathbf{X} - \mathbf{a}|^2 = (x_1 - a_1)^2 + (x_2 - a_2)^2, \quad (2.3)$$

where  $\mathbf{X} = (x_1(t), x_2(t))^T \in \mathbb{R}^2$  represents the location of a badger at time  $t$ , and  $\mathbf{a} = (a_1, a_2)^T \in \mathbb{R}^2$  is the fixed location. In order to generate a trajectory, the Euler-Maruyama method will be

used, whereby the initial position of the badger is set as  $\mathbf{X}_0 = [0.0, 0.0]^T$ , the fixed point as  $\mathbf{a} = [1.0, 1.0]^T$ , and the diffusion term takes the form of a diagonal matrix,  $\sigma = \Sigma \mathbb{I}_2$  (for some value  $\Sigma$ ).

Firstly, a trajectory with no noise is generated to show the attraction of the badger to the fixed point  $\mathbf{a}$ , observed in Figure 2.4a. The path of the trajectory is very straight, where no deviation is made from the path to reach the fixed point. This is to be expected, since there is no noise within the system. Next, changing the noise term, we expect the trajectory to become more random. Figure 2.4b represents the trajectory when  $\sigma = 0.5\mathbb{I}_2$ . The badger heads to the point of attraction and once there, it wanders around whilst staying close to this point of attraction. On the other hand, when we increase the noise term to  $\sigma = 0.9\mathbb{I}_2$  (Figure 2.4c), then we see a lot more randomness in the trajectory. It is difficult to see the path of the badger when there is noise, but it appears there is a pull to stay close to the point of attraction.

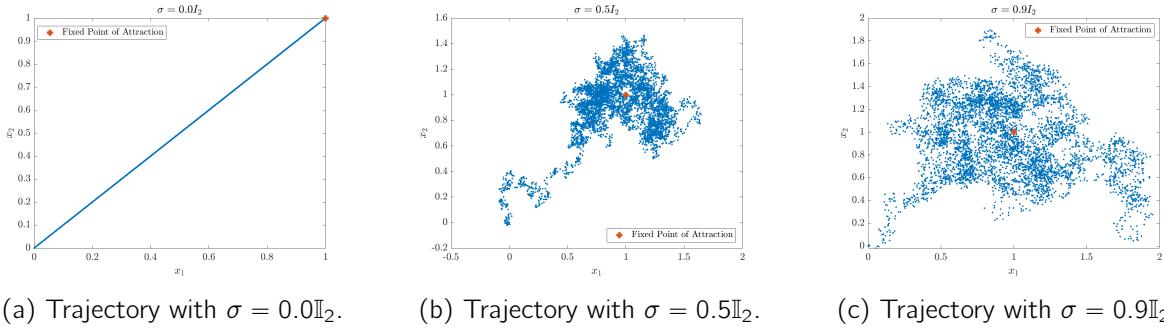


Figure 2.4: Trajectory comparing the increase in noise with a point of attraction. The trajectory is more random the larger the value of noise, where a less direct route is taken by the badger to get to the point of attraction. When there is noise in the system, the badger arrives at the point of attraction, then it wanders about in the same region of the point  $\mathbf{a}$ . The larger the noise, the further the badger deviates from that point.

The next behaviour to consider is the repulsion of the badger to a fixed point. For example, this could represent a household where badger repellents have been installed. Hence, considering a two-dimensional field, the potential takes the form

$$\mathbf{H} = \frac{1}{|\mathbf{X} - \mathbf{a}|^2} = \frac{1}{(x_1 - a_1)^2 + (x_2 - a_2)^2}, \quad (2.4)$$

where  $\mathbf{X} \in \mathbb{R}^2$  represents the location of the badger at time  $t$ , and  $\mathbf{a} \in \mathbb{R}^2$  is the fixed location. The same initial condition, fixed point and values of  $\sigma$  are considered as used for Figure 2.4. Firstly, repulsion to the point  $\mathbf{a}$  when there is no noise within the system is considered (Figure 2.5a), where the badger moves away from the point of repulsion in a very straightforward manner. This is to be expected since there is no noise in the system. Increasing the noise, we obtain Figures 2.5b and

**2.5c.** In both figures, the badger has an increase in movement and what appears to be a radius around the point of repulsion that they do not enter. This radius appears larger for the smaller noise value, whereas, when  $\sigma = 0.9\mathbb{I}_2$ , then the badger does get closer to  $\mathbf{a}$ , whilst keeping a certain distance. In this instance, it appears the noise of the badger is stronger than the repulsion of  $\mathbf{a}$  so it does not mind moving in that direction.

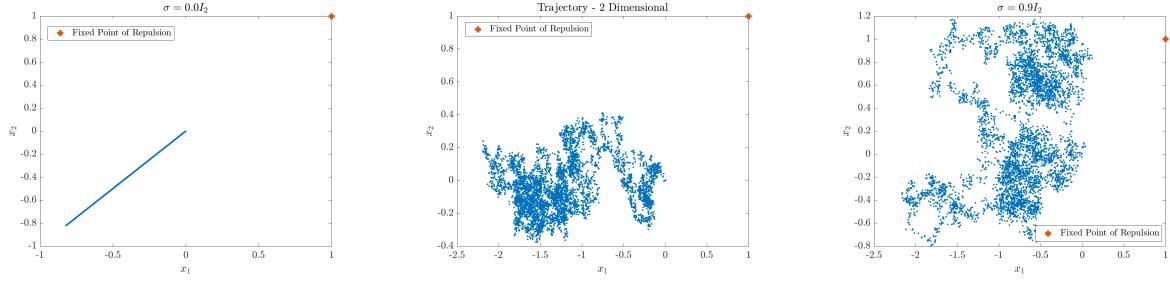


Figure 2.5: Trajectory comparing the increase in noise with a point of repulsion. The trajectory is more random when the noise increases, where the badger gets closer to the point of repulsion but still remains at a distance.

The next logical point to expand on is replacing the fixed point to the location of another badger. The state space of Equation (2.2) becomes  $d = 4$ . The attraction potential function with another animal included is defined as,

$$\mathbf{H}_{ij} = |\mathbf{X}^{(i)} - \mathbf{X}^{(j)}|^2 = (x_1^{(i)} - x_1^{(j)})^2 + (x_2^{(i)} - x_2^{(j)})^2 \quad (2.5)$$

where  $\mathbf{X}^{(i)} = (x_1^{(i)}, x_2^{(i)})^T$  represents the position of badger  $i$  at time  $t$  and  $\mathbf{X}^{(j)} = (x_1^{(j)}, x_2^{(j)})^T$  represents the position of badger  $j$  at time  $t$ . Additionally, let  $\sigma^{(i)}$  and  $\sigma^{(j)}$  represent the noise matrix for badger  $i$  and badger  $j$ , respectively.

Two initial conditions are needed since we have two badgers. Let  $\mathbf{X}_0^{(i)} = [0.0, 0.0]^T$  and  $\mathbf{X}_0^{(j)} = [1.0, 1.0]^T$  represent the initial position of badger  $i$  and badger  $j$ , respectively. Firstly, the badgers will be simulated when there is no noise present. We expect for the badgers to meet halfway between  $\mathbf{X}_0^{(i)}$  and  $\mathbf{X}_0^{(j)}$ , that is, they should meet at the position  $\mathbf{X}_{\text{middle}} = [0.5, 0.5]^T$ . This can be observed in Figure 2.6a. As the iterations increase, both badgers move towards each other and then, since there is no noise within the system, the badgers remain there. Introducing noise into the system, we set two different noises, since it has been observed that male badgers have a higher noise than females [39]. With regard to their actual behaviour, a higher noise value will indicate that the badger is more likely to move further. Let  $\sigma^{(i)} = 0.5\mathbb{I}_2$  be the noise term for badger  $i$  and  $\sigma^{(j)} = 0.9\mathbb{I}_2$  for badger  $j$ , i.e.  $\sigma = \text{diag}(0.5, 0.5, 0.9, 0.9)$ . It is hard to comment on behaviour in Figure 2.6b due to the noise but since badger  $j$  has a higher noise value, then we

do see the more random movements in their trajectory. Nevertheless, it appears the trajectories remain close to each other.

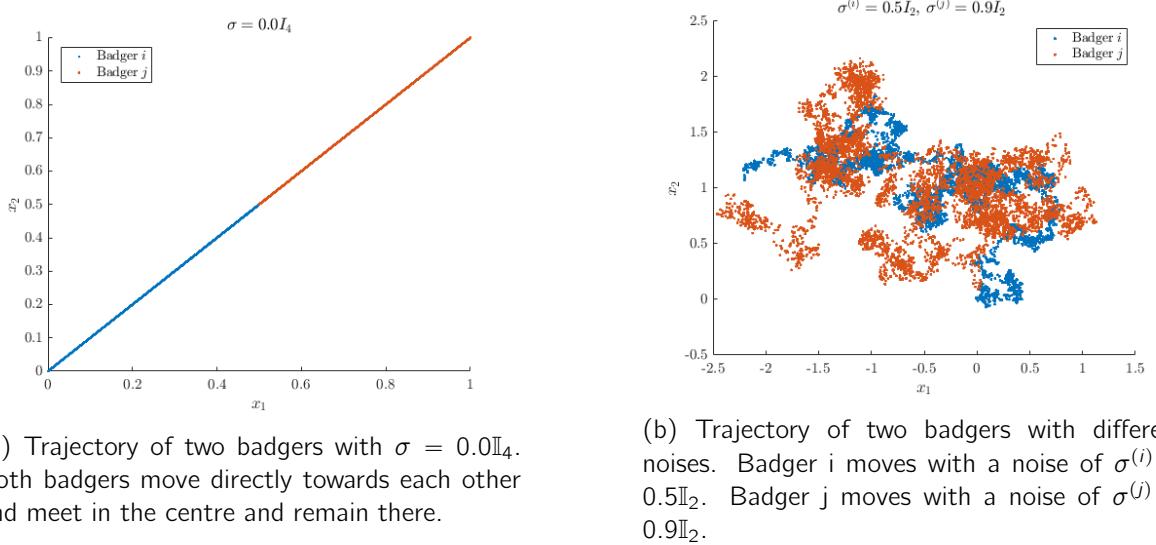


Figure 2.6: Trajectory of two badgers with Equation (2.5) employed as the potential.

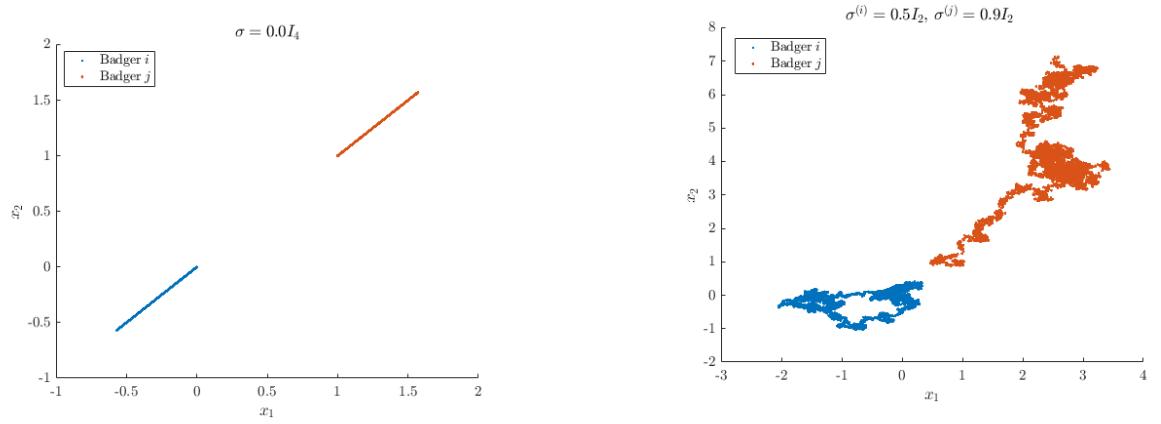
A disadvantage of this potential is that whilst it does show the attraction of one badger to another, it shows that once they have found each other, they don't move away. To simulate more realistic movements of badgers, it would be nice to see two badgers meet up, but then move to a different location for a different purpose, i.e. to forage for food, or to visit their sett.

Changing the potential to consider the situation where one badger is repulsed from another, the potential takes the form,

$$\mathbf{H}_{ij} = \frac{1}{|\mathbf{X}^{(i)} - \mathbf{X}^{(j)}|^2} = \frac{1}{(x_1^{(i)} - x_1^{(j)})^2 + (x_2^{(i)} - x_2^{(j)})^2}. \quad (2.6)$$

Upon generating trajectories, the initial conditions and the forms of  $\sigma$  are kept the same. Again, we firstly consider the system with no noise. Under this condition, we expect to see the badgers to just tend away from each other in a straightforward manner. This can be seen in Figure 2.7a, whereby the badgers move in opposite directions from each other. Then, noise is introduced, where each badger is different. The badgers persist on moving away from each other, and depending on the magnitude of noise, indicates on how random their walk is, as seen in Figure 2.7b.

This potential does exactly as expected, with the badgers being repulsed by each other. Even with noise in the system, this potential is not overly interesting because if there is no bound on the system, then the two badgers will continue to move away from each other. With the basics understood of how to simulate animal motion, these can be introduced on top of the energy



(a) Trajectory of two badgers with  $\sigma = 0.0I_4$ . Both badgers move in the opposite direction to each other.

(b) Trajectory of two badgers with different noises. Badger i moves with a noise of  $\sigma^{(i)} = 0.5I_2$ . Badger j moves with a noise of  $\sigma^{(j)} = 0.9I_2$ .

Figure 2.7: Trajectory of two badgers with Equation (2.6) employed as the potential.

potential, along with more badgers. This will be discussed in Chapter 3.

### 2.3. Large Deviation Theory

Rare events, that are often predictable, are important if they are extreme. They require computational approaches based on large deviation theory (LDT), which need the calculation of the least unlikely scenario (maximum likelihood pathway) that reduces to a deterministic optimization problem. In our case, the rare event is the transition between the wells that occur through minimum energy paths, i.e. mountain pass transitions. Hence, LDT provides tools for evaluating the probability of noise-induced transitions between the wells (also known as metastable states).

To find these minimum energy paths, we need to calculate the minimum points of the potential and then investigate the Hessian matrix at those points. If the matrix is indefinite, then it is a saddle point and the point is indeed a mountain pass. When the equation of the potential is known, then the process is quite simple. For example, consider the equation for the double-well potential (Equation (2.1)), i.e.,

$$V(x_1, x_2) = (x_1^2 - 1)^2 + x_2^2.$$

Finding the partial derivatives with respect to  $x_1$  and  $x_2$ , we obtain,

$$\frac{\partial V}{\partial x_1} = 4x_1(x_1 + 1)(x_1 - 1), \quad (2.7)$$

$$\frac{\partial V}{\partial x_2} = 2x_2, \quad (2.8)$$

respectively. Setting Equation (2.7) equal to zero, we obtain  $x_1 = -1, 0, 1$ . Similarly, for Equation (2.8). we obtain  $x_2 = 0$ . Next, computing the second partial derivatives, we have,

$$\frac{\partial^2 V}{\partial x_1^2} = 4(3x_1^2 - 1), \quad (2.9)$$

$$\frac{\partial^2 V}{\partial x_1 \partial x_2} = 0, \quad (2.10)$$

$$\frac{\partial^2 V}{\partial x_2^2} = 2. \quad (2.11)$$

Hence, our Hessian matrix is defined as

$$H = \begin{pmatrix} 4(3x_1^2 - 1) & 0 \\ 0 & 2 \end{pmatrix}, \quad (2.12)$$

where the eigenvalues of Equation (2.12) are  $\lambda_1 = 4(3x_1^2 - 1)$  and  $\lambda_2 = 2 > 0$ . In order to find the mountain pass, we need  $H$  to be indefinite, i.e. all eigenvalues are real and are not all the same sign. Therefore, we require  $\lambda_1 < 0$ . Clearly,  $\lambda_1 < 0$  iff  $x_1 = 0$ . Thus, the position  $(x_1, x_2) = (0, 0)$  is a mountain pass between the stable states  $(x_1, x_2) = (-1, 0)$  and  $(x_1, x_2) = (1, 0)$ , i.e. the base

of the wells. Once the mountain passes have been obtained, then we can find the probability of transition between wells, which can help us work out transition times. However, we will firstly focus on calculating the mountain passes in our generated energy potentials, since numerical methods would need to be applied due to the unknown equation.

## 2.4. Data Methods

Within this section, we will discuss more specific data methods that shall be employed.

### 2.4.1. Kernel Density Estimation

One step to creating the model is to generate the energy potential from the observations of random badger positions. These data can be either generated or collected from GPS. It is assumed that these positions are drawn from a certain, smooth function which describe the probability to observe a badger at a location  $(x_1, x_2)$  at time  $t$ . As highlighted in the literature review, there are lots of methods to delineate the home range of an animal, but since the key goal is to find this probability, we employ the method Kernel Density Estimation (KDE), despite the disadvantages identified. In essence, KDE is a smoothing technique, where we estimate the value of an unknown quantity based on measurements from neighbour observed badger points with a certain weight. The weight is given by a Gaussian function, which has an important parameter - the bandwidth ( $h$ ). This is an important mathematical problem because there are problems when the wrong bandwidth is chosen. If  $h$  is too small then the curve is under-smoothed and rigid, and when  $h$  is too large then the curve is over-smoothed and features are lost. Unfortunately, there are no 'right' choices, but there are estimators (discussed below) that can help.

KDE estimates the density ( $\rho$ ), whereas we require the energy potential,  $V$ , in application. The density has the form,

$$\rho = e^{-V}. \quad (2.13)$$

Therefore, to obtain an estimate for  $V$ , we need to take the logarithm of Equation 2.17, i.e.

$$V = -\log(\rho). \quad (2.14)$$

The density is calculated with the concept of weighting the distances of our observations from a particular point,  $x$  and can be expressed informally as:

$$\rho(x) = \sum_{i=1}^n K(x - X_i; h), \quad (2.15)$$

where  $K : \mathbb{R}^2 \rightarrow \mathbb{R}$  represents the kernel function,  $X_i; i = 1 \dots n$  represents the observations, and  $h > 0$  is the smoothing parameter called the bandwidth. There are several choices of kernels, but the Gaussian kernel is most appropriate for the task to gain the wells, which has the form,

$$K(x; h) \propto \exp\left(-\frac{x^2}{2h^2}\right). \quad (2.16)$$

So, with a normalization constant of

$$c = \frac{1}{2h\sqrt{2\pi}},$$

the density  $\rho$  takes the form

$$\rho(x) = \frac{c}{n} \sum_{i=1}^n \exp\left(-\frac{\|x - X_i\|^2}{2h^2}\right), \quad (2.17)$$

with  $\|\cdot\|$  is the Euclidean norm.

As an example, using generated data, we will generate the energy potential. Specifically, one badger will be simulated, with a trajectory of length 100,000 iterations. There is no other badger and no other attraction or repulsion within the system, only the double-well potential. We expect to see this double-well when we run KDE. The trajectory we obtain can be seen in Figure 2.8, where it is quite clear that the badger visits both wells. Now, different choices of bandwidths will be investigated in order to highlight the faults that have been identified in literature.

It has been identified how important it is to choose the correct bandwidth. Silva et al. [42] compare different bandwidth selection algorithms and recorded that the Cross Validation Least-Squares (CVLS) method was preferred against normal reference. We compare bandwidth estimates with three different algorithms: normal reference, cross validation least-squares (CVLS), and cross validation maximum likelihood (CVML). The normal reference refers to *Silverman's rule of thumb*, where the data are assumed to be unimodal and close to normal (thus 'normal' reference). It is computationally very fast, but as stated, is derived under the assumption that the true density is normal. The cross validation maximum likelihood method dates back to the 1900s, where it was proposed by Hobbema, Hermans and Van den Broeck in 1971 [18] and by Duin in 1976 [16]. A decade later, Rudemo [40] and Bowman [1] proposed the least-squares cross validation (also known as the unbiased cross-validation method), which could be considered the most popular and the best studied one. As highlighted by Silva et al. [42], the CVLS was the preferred method out of the two that they presented, whereby they were writing in 2020, a mere thirty years after the proposal of the method. Thus, showing that the method is still relevant in today's working.

Table 2.1 provides the bandwidth estimates for the different methods. Due to the size of the data set (100,000 iterations) both CVML and CVLS were taking too long to calculate an estimate,

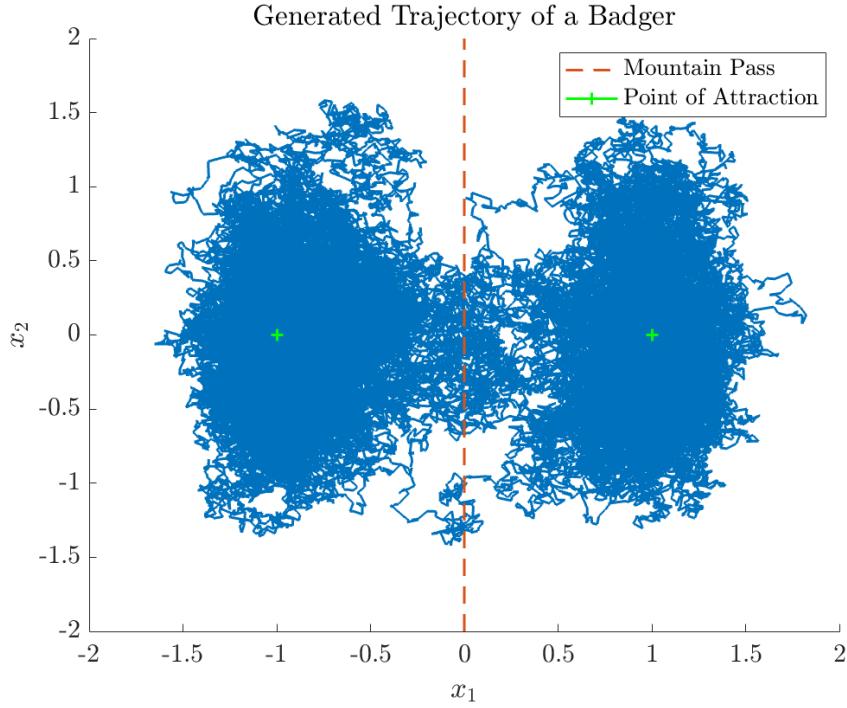


Figure 2.8: The generated trajectory of a single badger in a double-well energy potential, with noise  $\sigma = 0.9\mathbb{I}_2$ .

hence in this situation are not feasible. Yet, both methods were quick at giving an estimate for the downsampled data (i.e. a subset of the original data), where every 100 entries were taken. It can be observed from the table, the values for CVML and CVLS are very similar to the estimate for the normal reference rule of thumb rule for the full data set. On the contrary, if the downsampled data set was to be used with the normal reference method, then a larger estimate is given. Suggesting that if that method is to be used, then the full data set must be used in order to get a better estimate. Since the choice of bandwidth is up to the researcher, then it seems reasonable to choose  $h = 0.1$ , which is approximately the midpoint of the normal reference method. To check if this is a good fit, several potentials are plotted to compare, seen in Figure 2.9.

Firstly, the potential is calculated with a bandwidth  $h = 0.0648$ , the smallest value as estimated by CVLS. The generated potential can be observed in Figure 2.9a, with the two wells being clearly

Table 2.1: Comparison of Bandwidth Estimators using Python. Methods CVML and CVLS took too long to load an estimate for the full data set but took seconds to give an estimate for the downsampled data set.

Data Points	Normal Reference	CVLS	CVML
100,000	[0.07205407 0.14524988]	N/A	N/A
1,000	[0.15589465 0.31253089]	[0.06483719 0.17845769]	[0.08209628 0.18870059]

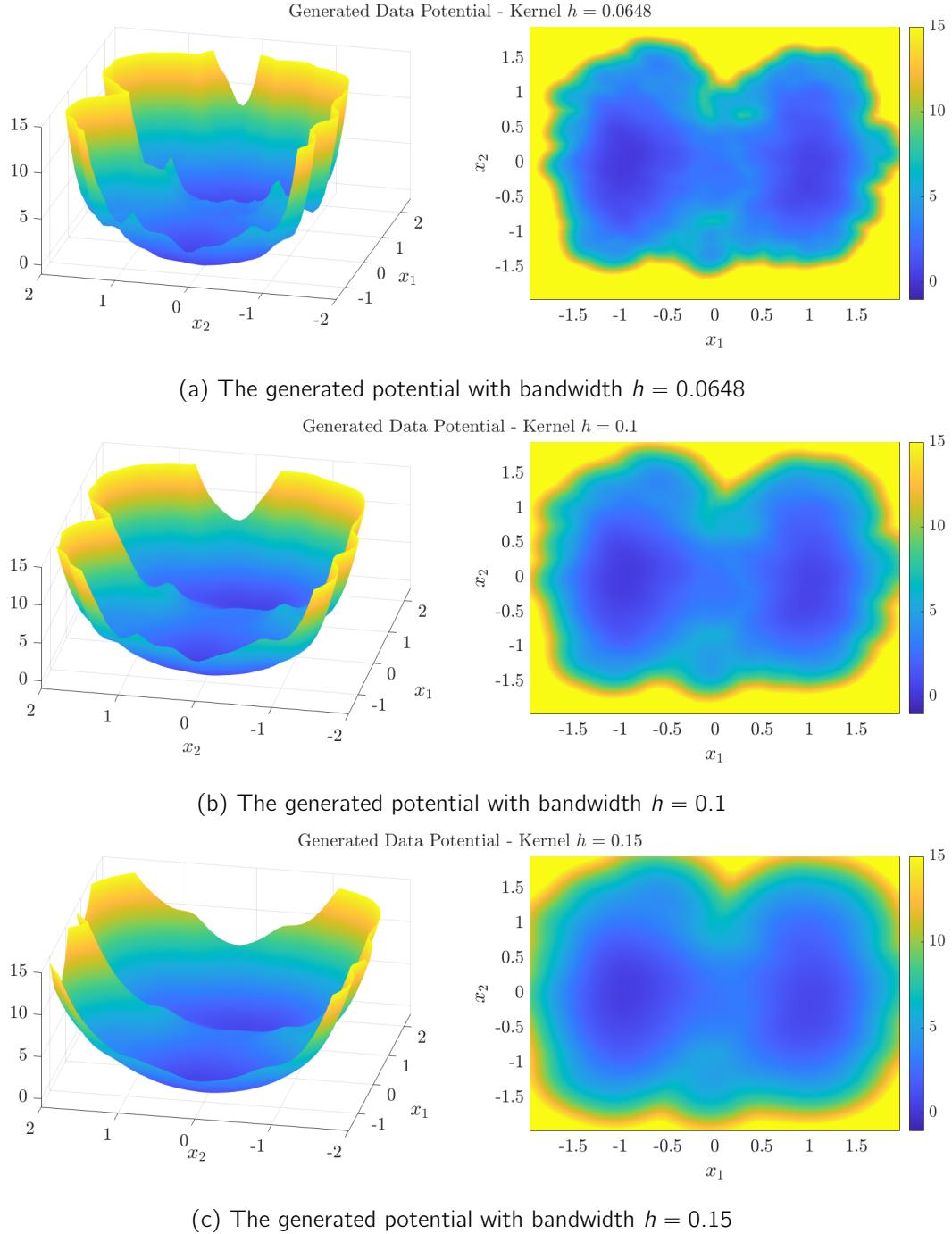


Figure 2.9: Using the generated trajectory from Figure 2.8, energy potentials have been generated using KDE with different values for the bandwidth. On the left, the 3-Dimensional image of the potential. On the right, the 2-Dimensional colour plot.

identified. Secondly, the bandwidth was set at  $h = 0.1$  (Figure 2.9b), which is approximately the midpoint of the normal reference estimation. Figure 2.9c represents the upper bound estimate for the normal reference method, where we are starting to see the potential becoming smoother. Hence, as discussed in literature, it is evident that the choice of the bandwidth is important, since a too large bandwidth can lead to over-smoothing and a too small bandwidth can lead to under-smoothing.

Comparing the exact double-well potential with the generated potential (with  $h = 0.1$ ) in Figure 2.10, we can observe that the exact potential is a lot smoother and looks more like two wells. As stated, KDE generates a potential based on where the badger has been, so it would be difficult to get the exact back. Regardless, it is clear that the generated potential is still a double-well. Next, comparing the exact invariant density of the double-well potential with the generated invariant density in Figure 2.11, we see that the generated invariant density has one well that is more likely to be visited than the other. Conversely, the exact invariant density plot (Figure 2.11a) gives equal chance of visiting either well. This is to be expected since the badger clearly spent more time in the left well and did not spend equal time across them both.

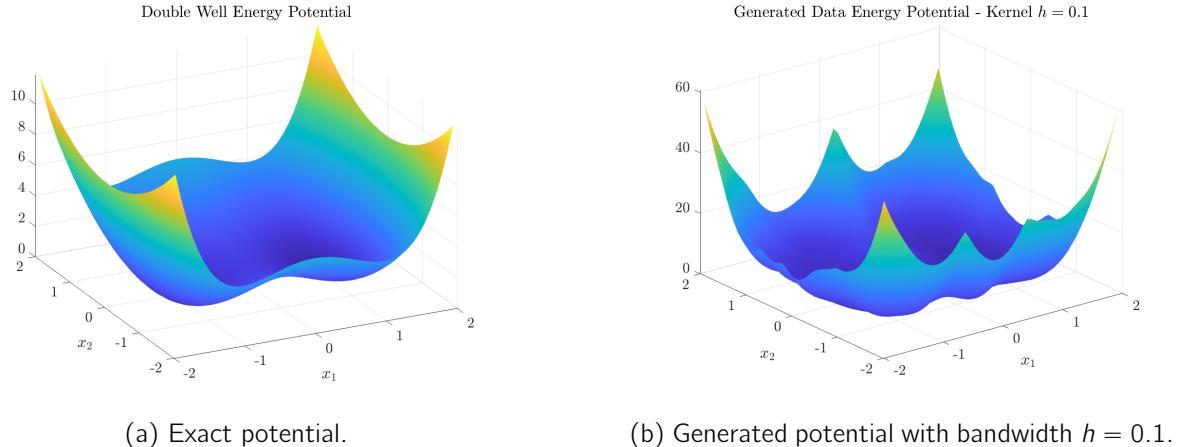


Figure 2.10: Comparison of Potential Plots.

Likewise, since we are able to plot the exact energy potential and invariant density, then we are able to compare them. However, this will not be the case when generating the potential with GPS data, as the potential is unique to the data. Hence, reinforcing the need to choose the correct bandwidth, so the potential is not over- or under-smoothed.

#### 2.4.2. Kramers-Moyal Formulae

Another goal for making the model is to recover the noise term of the badger, which will represent our diffusion term in the SDE (2.2). Niemann, Klus and Schütte [34] describe a method that can be

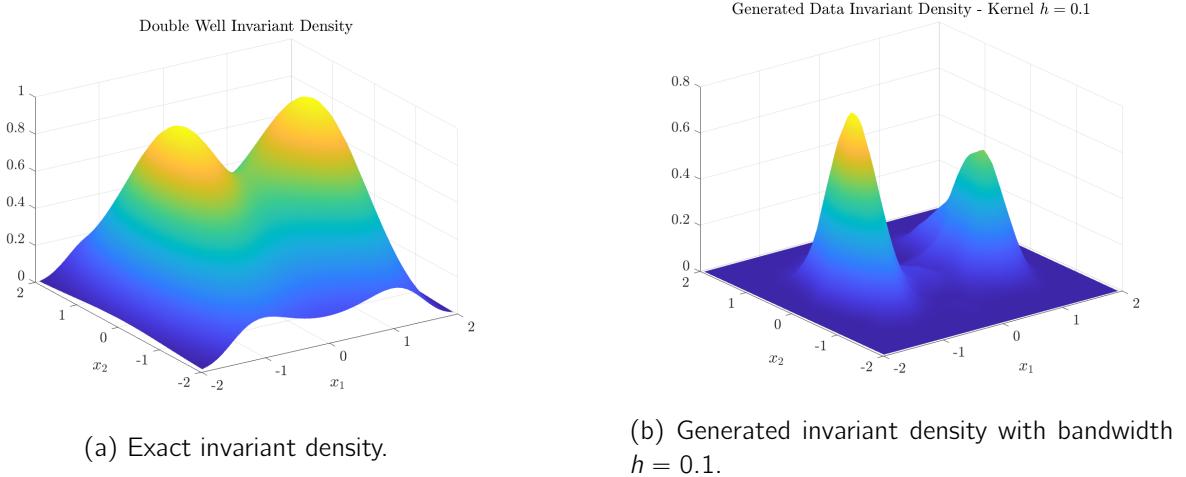


Figure 2.11: Comparison of Density Plots

used to estimate the diffusion pointwise via finite difference approximations for each measurement  $\{x_i\}_{i=1}^m$  using the Kramers-Moyal formulae,

$$a(x) := \lim_{\tau \rightarrow 0} \mathbb{E} \left[ \frac{1}{\tau} (X_\tau - x)(X_\tau - x)^T \middle| X_0 = x \right], \quad (2.18)$$

where  $a = \sigma\sigma^T$ , i.e. the diffusion term  $\sigma$  from Equation (2.2) multiplied by its transpose. This means that the method does not estimate the diffusion term directly. Let  $\mathbb{E}[\cdot]$  denote the expected value, and  $\tau$  the step size. Here,  $X_\tau$  represents the measurement  $x$  at a specific time step  $\tau$ . The algorithm firstly entails finding the difference in the coordinates and the time steps. Then, the differences of coordinates are multiplied by its transpose and one over the time step to obtain a two-by-two matrix. If there are  $m$  observations, we would obtain  $m - 1$   $2 \times 2$  matrices. These matrices are then summed and divided by how many data points are used. This matrix becomes the estimate  $a = \sigma\sigma^T$ . To obtain an estimate for  $\sigma$ , Cholesky decomposition is performed. Cholesky decomposition of a Hermitian positive-definite matrix  $A$ , is a decomposition of the form

$$A = LL^*,$$

where  $L$  is a lower triangular matrix with real and positive diagonal entries, and  $L^*$  denoting the conjugate transpose of  $L$ . Hence, since our estimation of the noise is  $a = \sigma\sigma^T$  then the Cholesky decomposition,  $L$ , is a good estimate to use within the model for  $\sigma$ .

The generated data from Figure 2.8 shall be used to practice this method of identifying the noise. For this set of data, the time difference ( $\tau$ ) is the same for each entry, so this could be factored out. However, for the GPS data, this does need to be considered as there are different

times recorded in between data points. Figure 2.12 visualizes the distribution, where the differences have been plotted in a 3-dimensional histogram.

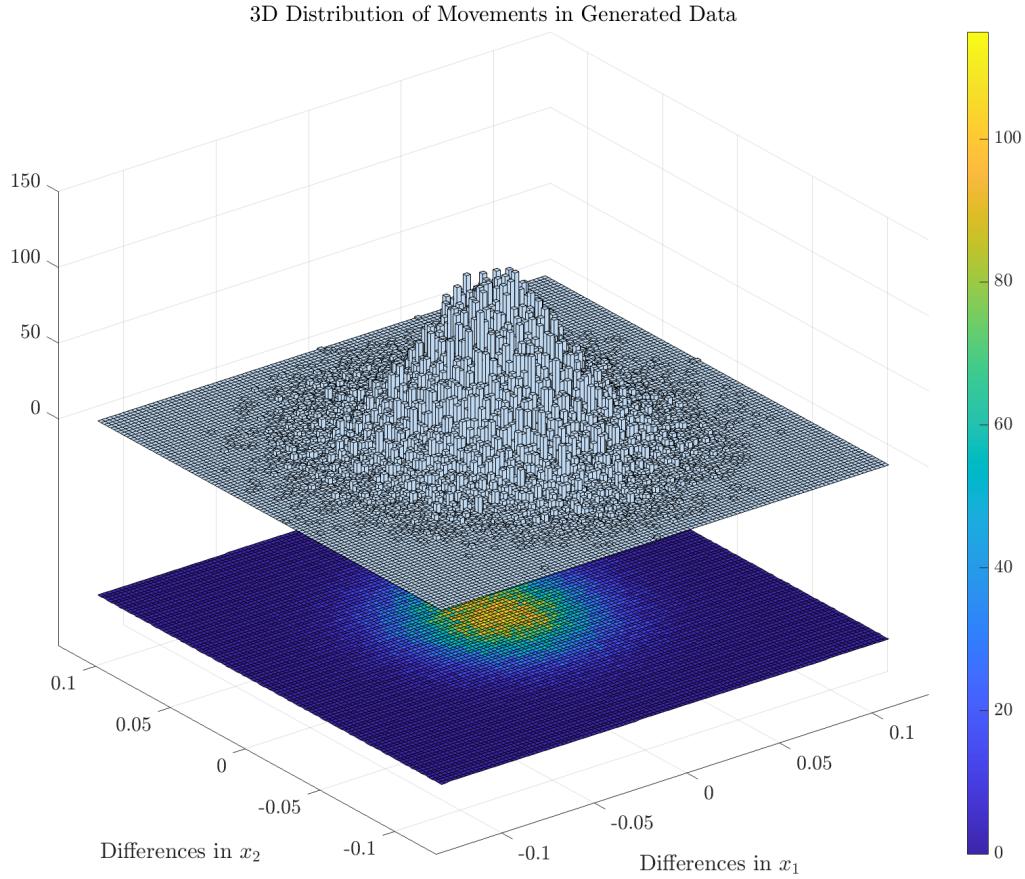


Figure 2.12: 3-dimensional histogram of the generated data with 106 bins in each direction.

We can calculate  $a$  directly for the generated data, given we know the specific sigma ( $\sigma = 0.9\mathbb{I}_2$ ). Evidently, when we estimate  $a$  for the GPS data, we do not have anything to compare to. The estimate is,

$$a = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix} \cdot \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}^T = \begin{pmatrix} 0.81 & 0 \\ 0 & 0.81 \end{pmatrix}.$$

Therefore, firstly, we investigate how the number of data points affects the estimate. Recall that the generated trajectory was made with 100,000 iterations. In the real data, we have no control of how many data points can be collected for a badger, so this needs to be taken into account. So, using the first one hundred entries of the trajectory, our parameters are  $m = 100$  and  $\tau = 0.001$  (the set value when we generated the trajectory). Applying Equation (2.18) and

Cholesky decomposition we obtain,

$$a \approx \begin{pmatrix} 0.9703708 & -0.00487441 \\ -0.00487441 & 0.8124079 \end{pmatrix} \quad \text{and} \quad L = \begin{pmatrix} 0.9851 & 0.0 \\ -0.0049 & 0.9013 \end{pmatrix},$$

respectively. This is a close approximation to what we expect. Next, if we increase the number of iterations to 1,000 entries, we obtain

$$a \approx \begin{pmatrix} 0.8550 & 0.0111 \\ 0.0112 & 0.7459 \end{pmatrix} \quad \text{and} \quad L = \begin{pmatrix} 0.9247 & 0.0 \\ 0.0121 & 0.8636 \end{pmatrix}.$$

Clearly, increasing the number of iterations in the trajectory obtains a better approximation. Nevertheless, the approximation is not too far out when there are only 100 entries.

Secondly, we use the same method but with the data downsampled to 5000 random entries. This will allow us to investigate how the approximation changes when  $\tau$  is not the same between data points, as it will be seen in the data. We obtain,

$$a \approx \begin{pmatrix} 0.7364 & 0.0066 \\ 0.0066 & 0.8116 \end{pmatrix} \quad \text{and} \quad L = \begin{pmatrix} 0.8581 & 0.0 \\ 0.0077 & 0.9009 \end{pmatrix}.$$

Clearly, the approximation is not too far out again, and it is a reasonable estimate for the noise.

Since it is currently unclear the form the matrix will take with the badger tracking data, we apply the same method on a second set of generated data where the noise is no longer a diagonal matrix, i.e.

$$\sigma = \begin{pmatrix} 1.5 & 0.9 \\ 0.1 & 1.0 \end{pmatrix}.$$

These data were generated with 1000 iterations and then downsampled to 750 random samples to have the effect of different step sizes. Applying the method to obtain an estimate for  $a$  and the Cholesky decomposition, we obtain

$$a \approx \begin{pmatrix} 2.9665 & 1.0266 \\ 1.0266 & 1.0124 \end{pmatrix} \quad \text{and} \quad L = \begin{pmatrix} 1.7223 & 0.0 \\ 0.5960 & 0.8106 \end{pmatrix},$$

respectively. Since  $L$  is a lower triangular matrix, then we don't have the full matrix returned. Since we know  $\sigma$  then we can calculate what we expect  $a$  to be,

$$a = \begin{pmatrix} 3.06 & 1.05 \\ 1.05 & 1.01 \end{pmatrix}.$$

As it can be observed, the estimation of  $a$  is very close to what it should be. Hence, using Cholesky decomposition will give us a good approximation to what  $\sigma$  could be and will be a good method to use on our data to retrieve an estimate for the diffusion.

### 2.4.3. Transfer Operators

We would like to obtain more information about our system and model on top of obtaining the energy potential, recovering the noise, and finding the mountain passes. One way we can obtain information about the behaviour of the dynamical system is through the use of transfer operators. A dynamical system describes the evolution of a quantity over time, which, in the current situation, corresponds to the positions of badgers in time. A transfer operator is a mathematical tool that is used to obtain these behaviours. A behaviour of the system that we hope to obtain is information about the metastable states (i.e. how many wells are in the energy potential and the coordinates of them), which we obtain by analysing the eigenvalues and corresponding eigenfunctions of these linear operators.

The two main operators for analysing such a dynamical system are the Perron-Frobenius and the Koopman operator. Let  $\mathbb{X} \subset \mathbb{R}^d$  be the state space and  $S : \mathbb{X} \rightarrow \mathbb{X}$  be a dynamical system. Here,  $S$  is a flow map for the movement of badgers in a fixed time interval. If not stated explicitly, time is considered to be discrete, where  $x \in \mathbb{X}$  evolves as  $\{x, S(x), S^2(x), \dots\}$ , such that it holds that  $x^{(k+1)} = S(x^k)$ . For context,  $x$  would represent the initial position of the badger, then, when applied to the dynamical system,  $S(x)$  becomes the next position of the badger, and so forth. This is known as the orbit of the dynamical system. In essence, it maps the movement of the badger.

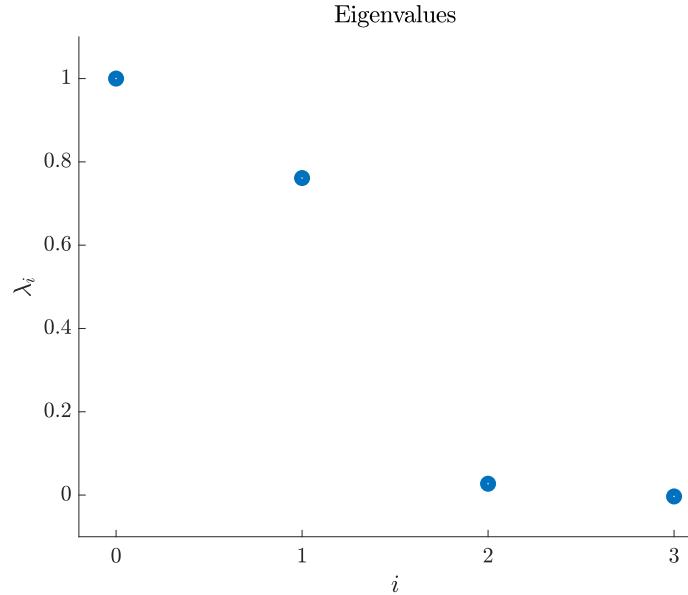
One of the main differences between the Perron-Frobenius operator and the Koopman operator is the latter describes the evolution of observables, whilst the Perron-Frobenius operator describes the evolution of densities [22]. A density is a positive function that belongs to the Lebesgue space with unit  $L^1$  norm. A probability density would give us the probability of finding a badger in a well. On the other hand, an observable is completely different, and is a function mapping the state space to  $\mathbb{R}^2$ . In other words, an observable is anything that can be measured, such as coordinates (lat/long) or weather. So, we consider the measurements  $\{f(x), f(S(x)), f(S^2(x)), \dots\}$ , where  $f \in L^\infty(\mathbb{X})$  is an observable function, instead of analysing the orbit. It has been stated in [22] that since the ‘two operators are adjoint to each other in appropriately defined function spaces’, it should not theoretically matter which operator is used to study the behaviour of a system. So, the Koopman operator  $\mathcal{K} : L^\infty(\mathbb{X}) \rightarrow L^\infty(\mathbb{X})$  is defined by

$$[\mathcal{K}f](x) = (f \circ S) = f(S(x)), \quad (2.19)$$

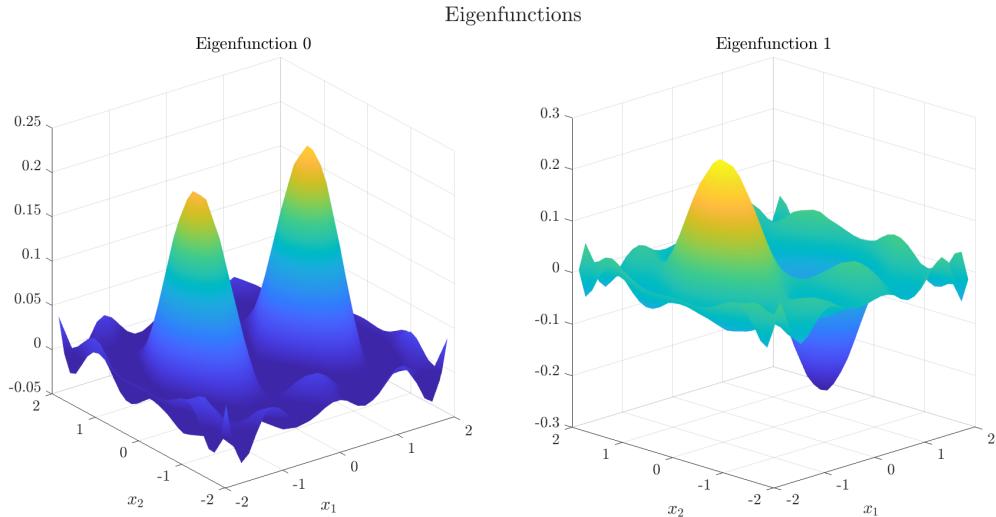
where  $f$  is an observable function. It can be observed that the operator depends on the dynamical

system.

The Koopman eigenvalues, eigenfunctions and eigenmodes can be computed approximately using a developed method called *Extended Dynamic Mode Decomposition* (EDMD). As an example, let's consider the double-well problem, where EDMD will be used to approximate the eigenvalues and corresponding eigenfunctions. Figure 2.13a plots the four eigenvalues found for the system. As it can be observed, there is a spectral gap between eigenvalues  $\lambda_1$  and  $\lambda_2$ , indicating that the double-well system has two metastable sets. To understand more about these sets, the eigenfunctions related to eigenvalues  $\lambda_0$  and  $\lambda_1$  should be computed. Figure 2.13b highlights the plots of the corresponding eigenfunctions. The eigenfunction, corresponding to eigenvalue  $\lambda_0$ , indicates the locations that the badger is most likely to be found, i.e. at the base of the well. As the badger moves away from the base of the well, the probability of locating the particle decreases. On the other hand, the eigenfunction, corresponding to eigenvalue  $\lambda_1$ , indicates to the reader that there are indeed two distinct wells within this system. One located at  $(x_1, x_2) = (-1, 0)$  and the other at  $(x_1, x_2) = (1, 0)$ . The significance of using transfer operators to approximate the eigenvalues and eigenfunctions of a dynamical system is that it will allow us to detect badger setts/social groups when using the data to generate the energy potential.



(a) Plot of the eigenvalues for the double-well system. There is a spectral gap between  $\lambda_1$  and  $\lambda_2$ , which implies the double-well system has two metastable sets. The eigenfunctions related to these eigenvalues contain information about these sets.



(b) The eigenfunctions of the double-well system, computed with the Koopman Operator. Eigenfunction 0 corresponds to  $\lambda_0$  and Eigenfunction 1 corresponds to  $\lambda_1$ .

Figure 2.13: Using EDMD to approximate the eigenvalues of the double-well problem, along with the corresponding eigenfunctions.



# 3

## Modelling Strategies

This chapter is a preliminary exploratory stage, where different modelling strategies are investigated. It is known that within badger ecology there are different territories, whereby each territory has a main sett that is occupied by the social group of that territory. Hence, straight away, the main sett can be set as a point of attraction when considering a potential function. To keep things simple when creating a model, we use a double-well energy potential (as described in Chapter 2), in conjunction with a behaviour potential. The behaviour potential only represents the attraction of badgers to one another and their repulsion, and no other forms of behaviour. This potential will then be developed to consider more ecological reasoning in the choices. It should be highlighted that this behaviour matrix is not something that we expect to obtain from the data (as it is expected to be included within the noise) but something that we are experimenting with within these models. Finally, we shall consider a simple infection model on top to visualize how the spread of bTB amongst the population might look.

### 3.1. Badger Model

Let there be  $n$  badgers in the dynamical system, where each badger has two positions,  $(x_1, x_2)$ . Therefore, we have a state space  $d = 2n$ . Defining our SDE, we have

$$d\mathbf{X}_t = b(\mathbf{X}_t)dt + \sigma(\mathbf{X}_t)d\mathbf{W}_t,$$

where  $\mathbf{X} \in \mathbb{R}^d$  represents the location of animals at time  $t$ ,  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  represents the drift term,  $\sigma \in M_d(\mathbb{R}) \rightarrow M_d(\mathbb{R})$  represents the diffusion and  $\mathbf{W}_t$  is a Wiener Process.

In order to determine the behaviour of the badger, i.e. whether it is attracted or repulsed by another, we have defined a matrix  $A \in M_n(\mathbb{R})$  with the entries 0, 1, and  $-1$ . If the entry  $a_{ij} = 1$  then badger  $i$  is attracted to badger  $j$ . If the entry  $a_{ij} = -1$  then badger  $i$  is repulsed by badger  $j$ .

Finally, if the entry  $a_{ij} = 0$  then badger  $i$  is impartial to badger  $j$ . Only the diagonal entries are to be set to 0 (i.e.  $i = j$ ), since one cannot be attracted or repulsed by oneself. This matrix will be generated randomly. Furthermore, we have weighted our potentials, so the double well potential is a bigger force on the badgers than the behavioural potentials. Ecologists have observed that when badgers come out of the sett for the night, they socialize a little and then go their own way to forage alone. Hence, this is why we expect the energy potential to have a larger influence on the movements than the behaviour potential.

We explore one random example together, where our system shall contain five badgers. The initial positions and matrix  $A$  will be kept the same for each simulation. Table 3.1 displays the randomly generated initial positions (between -2 and 2) and Matrix 3.1 displays the random behaviour matrix. Choosing an entry of  $A$ , to help understand the meaning, let  $a_{34} = -1$ , which means that badger 3 is repulsed by badger 4. We use the Euler–Maruyama method to solve the SDE with these conditions.

Table 3.1: Initial Positions of Badgers

	1	2	3	4	5
$x_1$	1.05	-0.29	-0.74	-0.14	0.87
$x_2$	-1.11	-0.43	-1.06	-1.19	-1.44

$$A = \begin{pmatrix} 0 & 1 & -1 & -1 & -1 \\ 1 & 0 & 1 & 1 & -1 \\ -1 & 1 & 0 & -1 & 1 \\ -1 & -1 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 0 \end{pmatrix} \quad (3.1)$$

To begin, the trajectories will be generated with the noise set as  $\sigma = 0.0\mathbb{I}_{10}$ , i.e. no noise in the system. No movement is expected between the wells, but we expect for the badgers to move toward the base of the wells. Matrix (3.1) clearly has a mix of behaviours within the system, which can be seen in Figure 3.1. For example, badger 1 (blue line) is repulsed by badger 5 (orange line) ( $a_{15} = -1$ ), whilst badger 5 is attracted to badger 1 ( $a_{51} = 1$ ). With this behaviour, we see that they both move towards the base of the well, but badger 1 keeps moving to keep a distance from badger 5. In the other well, we see similar behaviour, where badger 2 (yellow line) is attracted to badger 3 (green line), whereby the attraction is reciprocated. But since badger 4 (burgundy line) is repulsed by badger 2 then the badger 4 remains at a distance from the base of the well.

Next, we simulate the badgers with the noise term now as  $\sigma = 0.9\mathbb{I}_{10}$ . We run the simulation two times (Figure 3.2) to see how the behaviour changes. It can be observed in the first simulation, Figure 3.2a, that both badger 1 (blue) and badger 5 (orange) make the move across the mountain pass. It appears that badger 1 does spend more time in the right territory in comparison to badger 5. This is surprising, as by the Matrix (3.1), badger 5 is repulsed by all the badgers in the left territory. Hence, showing the force of the energy potential having a larger effect than the behaviours. It is hard to determine the rest of the activity, but it can be seen that all five badgers remain on the

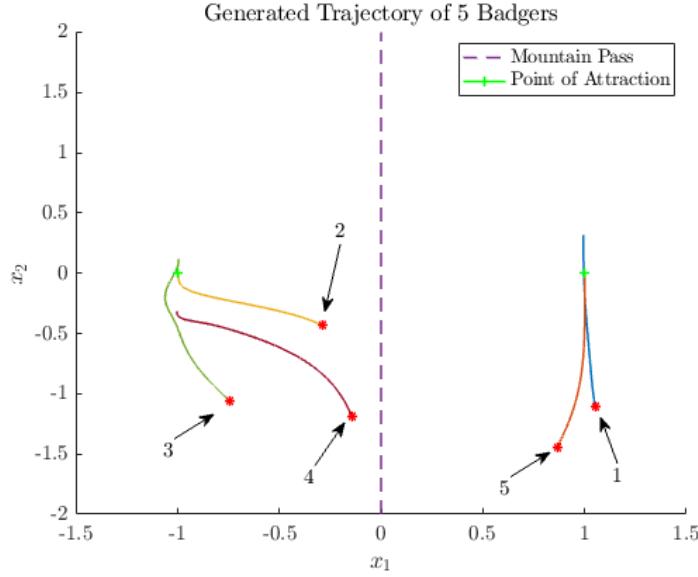


Figure 3.1: Simulation of the badgers when there is no noise. The red star indicates their initial positions. The badgers remain within the wells that they start in. Then, depending on the behaviour, they move towards each other or remain at a distance. Badger 1 does not stop at the well and continues since it is repulsed by badger 5. Similar behaviour can be seen by badger 4, where they do not go to the base due to being repulsed by badger 2, who is already at the base of the well.

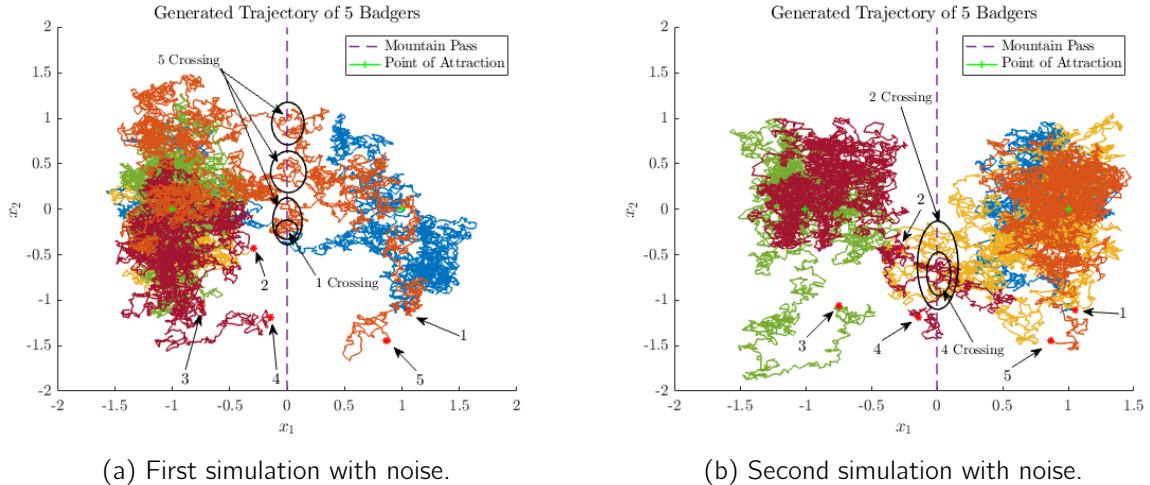


Figure 3.2: Simulation of badgers with  $\sigma = 0.9\mathbb{I}_{10}$ . The red star indicates their initial positions. There is a lot of movement within the system, and it is hard to determine all movements. In both simulations, there are well crossings by different badgers, such as badger 5 in Figure 3.2a, who starts in the right well and makes multiple crossings to the left well, as indicated by the circles.

left territory for the remainder of the iterations.

The second simulation (Figure 3.2b) offers a similar observation, but this time badgers 2 and 4 make the journey across to the right territory (as indicated by the circles). However, it can be

observed that badger 4 makes a short journey across to the right territory before moving back to the left territory. Badger 4 is repulsed by both badger 1 and 5, so this is not too surprising to see the badger return. Similar to Figure 3.2a, it is hard to determine the rest of the activity, but some simple behaviours can be observed from the figures.

## 3.2. Ecology Model

It is fair to say that the model in Section 3.1. does not take into account the ecology of badgers. Mainly, it is assumed that there is no difference between the male and female badgers. They all have the same noise term and the interaction matrix  $A$  is randomized. Whilst the noise is seen to change seasonally, in general, it is seen that the males are more likely to travel [39]. During mating season, however, the females could be seen to travel more (in order to gain more mates) since European badgers are one of the few mammal species that show [delayed implantation](#) and one of only five which are suggested to show [superfoetation](#) as a reproductive strategy [7]. This means the females practice cryptic [polyandry](#), which is where one female will mate with several males in a breeding season. This suggests that females will travel more during the mating season to obtain more mates. Hence, in order to make our simple model more ecology orientated, these are some things that we need to take into consideration.

To make the matrix  $A$  more ecology realistic, we can make some general assumptions. For  $n$  badgers, let the first  $f$  entries be females and the remainder to be males ( $n - f = m$ ), where age is not taken into consideration. Additionally, they can be split into two groups, to represent two different territories. However, it remains that the initial conditions are random, regardless of assigned groups to get a rough idea on how they would move in the energy potential with this behaviour potential. The first  $f/2$  females will be assigned to the first territory and the second half are part of the second territory. Then, with a similar approach to the males, the first  $m/2$  are part of the first territory and the remainder are part of the second. Then, it seems reasonable to assume that all females are attracted to all males, for mating purposes, regardless of social group. In return, all males are attracted to all females. Then, if the badger shares the same territory, then they will be attracted to them (since they are a social group) and all the others shall be repulsed. The ecology would be more complex, as from an evolutionary perspective in-breeding is not good, and it could be expected that a badger from a territory to prefer mating with others from different territories to ensure better gene mixing, but this will allow for a starting point for the model.

As an example, let there be ten badgers within the system, with four females, and six males.

Then, we obtain the following interaction matrix:

$$A = \begin{pmatrix} 0 & 1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 0 \end{pmatrix}. \quad (3.2)$$

When running this model, the male noise is set as 0.9 and the female noise set as 0.6. Hence, we expect to see larger movements from males. One such trajectory obtained can be observed in Figure 3.3.

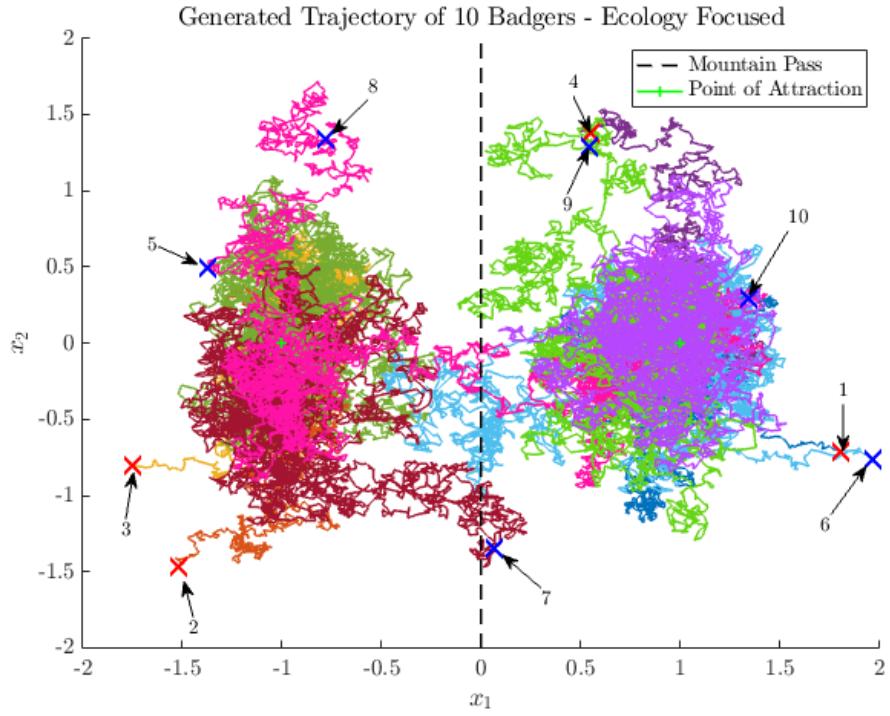


Figure 3.3: The trajectory of ten badgers over 5,000 iterations. Red crosses indicate the initial position of a female, and the blue cross indicate the initial position of a male. The following colours link the corresponding badgers: B1 dark blue, B2 dark orange, B3 orange, B4 dark purple, B5 dark green, B6 light blue, B7 burgundy, B8 pink, B9 light green, B10 light purple.

Evidently, it is hard to see the interactions of the trajectories when more badgers are added to the system. However, what we are able to observe from Figure 3.3 is that no females cross between the territories. But, badger 6 (light blue) and badger 8 (pink) do swap territories. It appears a large amount of time is spent in their initial territories before moving, which is what we are expecting. In order to overcome the difficulty to observe the behaviours, we can transform the trajectory into an animation so the movements can be observed at each time point.

### 3.3. Infection Model

The aim of this section is focusing on technique and how an infection could be modelled on top of the SDE. As noted, the spread of bTB amongst the badger population is an important topic and an issue that we would like to explore, if the data allow. So, to take the SDE model further, a simple susceptible-infectious (SI) model will be considered. Due to the volume of badgers that shall be generated, we return to the random generated matrix  $A$ , rather than the ecology focused matrix. On the one hand, this could be realistic, since we do not understand if and when badgers are attracted or repulsed to one another. On the other, the system that was used in Section 3.2. seemed to offer some realistic thinking into the relationships of badgers that we could incorporate into the model. But, since we won't be obtaining this matrix from the data, then at this moment the values in  $A$  do not matter.

To create the SI model, firstly, the pairwise distances between badgers needs to be computed, which will be stored in a matrix. The Euclidean distance is used, to keep it simple. Since we are considering an energy potential and not topography, another type of mathematical tool to calculate the distances between the badgers does not need to be considered. Then, for every time point  $t$ , the distance matrix will be stored in a list  $D$ , for later use. For instance,  $D[0]$  is the matrix of distances at the initial time point. All badgers, except one, will begin in the ‘susceptible’ compartment. The one chosen badger shall be in the ‘infectious’ compartment. Then, a certain distance is set, and if a susceptible badger is within that distance of the infectious badger, then with a random probability that badger will move to the infectious compartment. This process is repeated for all time points.

Evidently, this is a very simple SI model, and, hence, does not take into account all properties surrounding bTB transmission. Initially, the time the badger needs to be around the infectious badger before the susceptible badger becomes infected, i.e. contact time, was not taken into consideration. Caley et al. discuss the impact of contact time in developing TB in a school setting [4]. Following with the identification that one student had developed TB, it was then found that the students in the same school year were at a significantly higher risk of infection in comparison to others in the school. Thus, suggesting that contact time is important when considering the

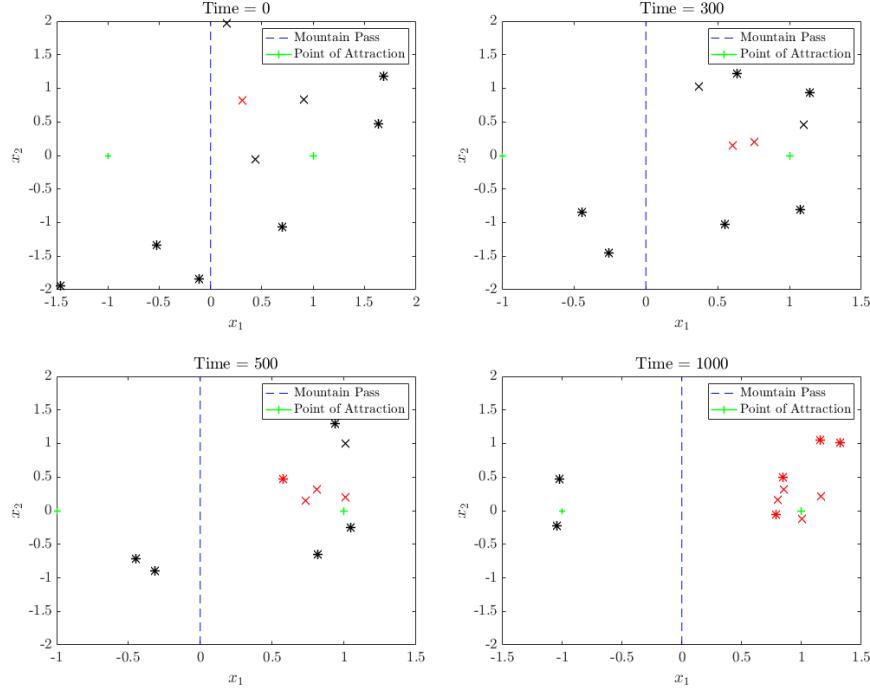
transmission of the infection.

The mechanics of the current model suggests when an animal is infected then it is instantly infectious, yet one can be infected but not infectious. In terms of bTB, when an animal is infected, this means that the bacterium has established within the host. When an animal is infectious, then the animal is now liable to transmit the infection to other animals. Additionally, once an animal becomes infected, there is every possibility that they could heal (in proportion) but it is impossible to infer time for how long this takes and how many animals actually heal. Hence, when developing the model further, infected and infectious needs to be taken into consideration.

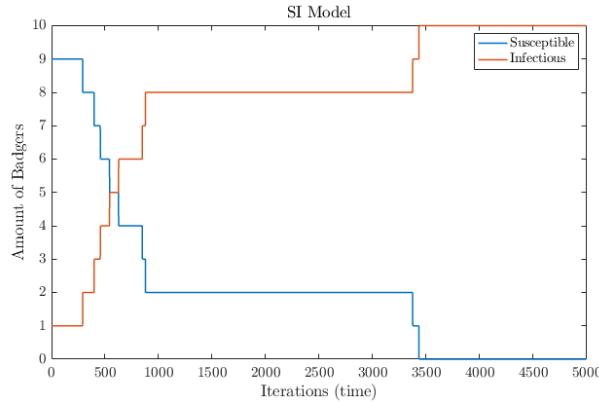
Taking contact time into consideration, the model was developed to include a new matrix  $C$ , which represents the time counter for each individual badger. When a badger is within the certain distance for infection, then the counter increases by one for that badger. When they have reached a certain count, then that badger will move from the susceptible compartment to the infectious compartment. Whilst this concept might still be lacking in certain bTB transmission dynamics, it gives the idea that contact time is important before a badger becomes infectious.

Finally, in order to see the transition from the susceptible compartment to the infectious compartment, we are able to make an animation. This time, the colour of the badgers depend on which compartment they are in; if they are a member of the susceptible compartment they are blue, and if they are a member of the infectious compartment they are red. This aids the viewer when watching the animation to see what compartment each badger is in at each time point, whilst also having features, such as different symbols for male and females, and also lines to indicate the territory border.

For example, Figure 3.4 highlights a simulation made with the infection model. There are ten badgers in the system (four females (cross) and six males (stars)) with random starting positions and a random behaviour matrix  $A$ . Similar to the ecology model, the females are given a lower noise value than the males, with the values  $\sigma = 0.5$  and  $\sigma = 0.9$ , respectively. The model was implemented with the delay, meaning a badger had to have at least twenty contacts with the infected badger (with a distance of 0.25 units) before becoming infected and infectious themselves. These values were chosen randomly with no link to bTB itself. The model was simulated with five thousand iterations, which could be considered as five thousand hours (approximately 208 days), where time is continuous, i.e. there is no break for daytime. By the 37th day, all the badgers in the right well are infected (with time = 1000 being the 42nd day (Figure 3.4a)). By the 140th day, all badgers in the system are infected. As stated, this is a very simple model, and is not a true representation of how bTB spreads. But, it has allowed for exploration of a technique for developing an infection model on top of a simulation.



(a) The infection model at four different time points. The crosses represent female badgers and the stars represent males. If the colour is black, the badger in the susceptible compartment, and if red they are infected and infectious. By the one thousandth iteration, all the badgers in the right well are infected.



(b) Details for how the infection spreads over time. One badger starts in the infectious compartment and begins to infect others. The two badgers in the left well (Figure 3.4a) remain in the susceptible compartment for a large amount of time, but by the end of the simulation, all badgers are infected.

Figure 3.4: A simulation run with the infection model, with a total of ten badgers (four females ( $\sigma = 0.5$ ) and six males ( $\sigma = 0.9$ )). In order to become infectious, the badger has to be within the distance of 0.25 units of the infected badger for twenty counts. The behaviour matrix  $A$  is random, and the simulation was 5000 iterations. The iterations can be considered as time, more specially, one hour is one iteration, so this simulation is approximately 208 days. By approximately the 140th day, all badgers are infected.

# 4

## Badger Data

The methods that we have established we would like to apply to the obtained data. Within this chapter, we discuss what data we have for the project, along with any issues we have had. We have been supplied with two sets of GPS data; a set from the Animal and Plant Health Agency (APHA) for badgers in Woodchester Park (Gloucestershire), and a second set of data for four different sites in Cornwall supplied by the Zoological Society of London (ZSL) [50]. In addition to these data sets, we have acquired some small sets of badger data from other researchers, including some data that have been recorded using the method of dead-reckoning [27]. Dead-reckoning is a procedure that allows the animal tracks to be elucidated through the combined use of tri-axial accelerometers, tri-axial magnetometers and GPS loggers. However, the main data sets that we have for the project were captured with GPS collars on the badgers. Within this report, the methods are applied only to the data from APHA, with further work to be done for the ZSL data and other sets.

### 4.1. Overview of Woodchester Park

Woodchester Park is a 521-acre biological Site of Special Scientific Interest, situated on the Cotswold sandstone escarpment in Gloucestershire between 47m and 210m above sea level. The Woodchester Park Study was established in 1976 to investigate bTB in a wild, naturally infected badger population. The study includes analysis of capture-mark-recapture data [39], and the analysis of GPS data. The study site itself covers c.  $11\text{ km}^2$  of predominantly permanent pasture and woodland [9].

I was invited to Woodchester Park to observe what happens at a trapping event by APHA in December 2021. I travelled the day before to see Woodchester Park in daylight, as the next morning was going to be an early one. Unaware of the locations of the badger territories, I walked past quite a few, which were pointed out the next morning. As it can be seen in Figure 4.1 and Figure 4.2, Woodchester Park is very scenic, with lots of wooded areas. The park features a manor

house, which remains unbuilt when bought in 1845 to be redesigned. Additionally, there is a chain of lakes, with natural bridges between them (Figure 4.2d) in the middle of the park with hills either side. This can be seen in Figure 4.2 where Figure 4.2a is taken near the lake and Figure 4.2b is taken looking down at the lake.

We were out early on the morning of the 7th December, and we were lucky to see two badgers being brought in (Figure 4.3). The first badger was a male, who was caught from the Honeywell social group, and the second badger was another male caught from the Yew social group. Both badgers were first caught earlier in the year as cubs. A few more badgers were caught from different territories by different field workers. Some features of badger ecology were highlighted on the trip, including entrances to badger setts (Figure 4.4a). Additionally, when walking through a field between territories, we saw a mole hill with a worm on top (Figure 4.4b), which would be the perfect meal for a badger.

As mentioned, APHA run a long-term study at Woodchester Park, so traps are left out during trapping seasons to catch the badgers. Sometimes the traps can be naturally damaged, i.e. if a tree falls on them. However, to stop them being damaged otherwise, a description is left on the cages to explain why they are there (Figures 4.5a and 4.5b). I was also lucky to see the process that the scientists do to test the badgers for their analysis, as seen in Figure 4.5c. In the figure, the badger has been given a general anaesthetic, and then a number of tests are run, including a sputum test (to test for bTB). This badger had a collar on, tracking its movements, so this is removed to check that it is all okay. This will be returned to the badger if there are no injuries. If this was a newly caught badger, then a tattoo for a unique identity would be given (this can be seen on the stomach of the badger in the image), and if no injuries have been sustained, a collar for GPS tracking would be added. Normally, once the anaesthesia has worn off, the badgers would be released the same day in the same area that they were caught. There was a storm coming in on the day I visited, so the badgers were released the next day when it was safer to do so. Overall, a very interesting visit, with lots of information gained of how the information for this project has been collected.



(a) National Trust information board in the car park.



(b) Path around the park, off road route to Woodchester Manor.



(c) Variety of levels around the park.



(d) Lots of wooded areas.

Figure 4.1: Scenery of Woodchester Park, taken on 6th December 2021.



(a) Looking up the hill to a badger territory.



(b) View down from Figure 4.2a. In the right of the image, there are forestry works happening. At the bottom of the hill is the Boat House and the bridge.



(c) Woodchester's Boat House.



(d) Natural bridge between the chain of lakes.

Figure 4.2: Natural features of Woodchester Park, captured on the morning 7th December 2021.



(a) Male badger caught in the Honeywell Sett. First caught in May 2021 as a cub.



(b) Male badger caught in the Yew Sett. First caught in August 2021 as a cub.

Figure 4.3: Two badgers that I saw being brought in during my visit to Woodchester from two different setts. Images caught on 7th December 2021.



(a) An entrance to a badger sett.



(b) A badger's best meal - must have been missed from the night before.

Figure 4.4: Images of a badger sett and perfect meal in a field, taken on 7th December 2021.



(a) Their main study being analysis of capture-mark-recapture data. The cages are left out all year round, so the badgers get used to seeing them.



(b) A note is left on the cages so the public who come across these cages can understand what they are used for.



(c) Sputum test on a badger - identifying if the badger is a carrier of bTB. Blood is also taken, and if the badger is well, then a collar will be fitted to track their movements.

Figure 4.5: Woodchester Study - the traps for capturing the badgers and at the facility testing the badgers. Images captured on 7th December 2021.

## 4.2. Preliminary Data Analysis

A starting point is to analyse a set of data from Woodchester Park that was obtained for an undergraduate project in the School of Veterinary Medicine at the University of Surrey, titled *Effects of extreme weather events on the demographic of a population of Eurasian Badgers (*Meles meles*)*. The data are from the long-term study of capture-mark-recapture program, which were collected to help understand the demographics of the population in the park and to test for bTB.

The data are a subset of the total collected data, where there are 2478 observations taken between 1992 and 2011. The data are made up of the variables; tattoo, code, population name, birth population, moved, sex, year, and age. Upon capture, each badger is allocated a unique tattoo to identify the badger on later captures. It was noted by McDonald, Robertson and Silk that ‘more than 85% of individuals are caught as cubs (in their first year), such that their natal social group and age are known’ [28]. This has allowed the researchers at Woodchester to identify the birth population name of the majority of the badgers. In the event that the badger was not caught as a cub, then the sett they were caught in would become their natal social group and an age would be estimated by their tooth wear [11]. Thus, allowing there to be a distinction if the badger has moved, where, upon recapture, the natal social group is different to the social group they are caught in.

Out of the total 2478 observations, we observe 906 individual badgers: 504 observations of females and 402 observations of males. Within the data, it is seen that the minimum age of a badger being caught is 1 (cub) and the maximum age is 13. The life expectancy of a Badger is 14 years, with the average lifespan 5-8 years, so it is rare to see badgers of that age in the wild [47].

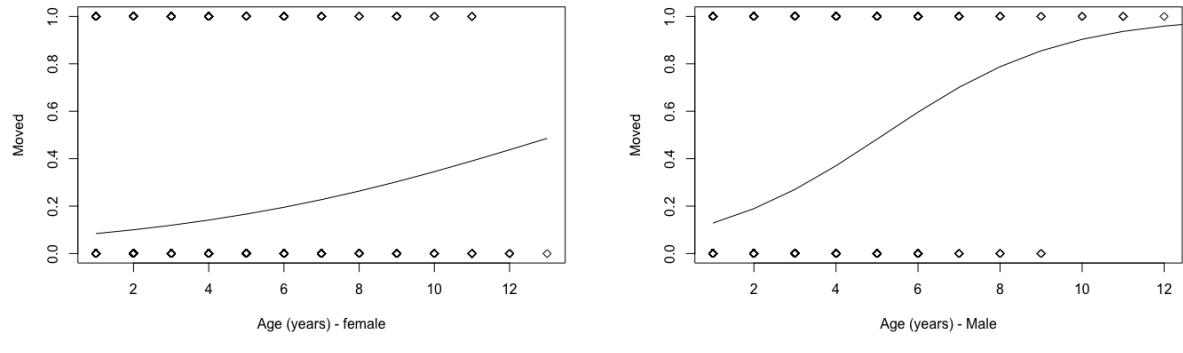
### 4.2.1. Logistic Regression

A logistic regression is performed on the set of data to see how the variables sex and age impact on whether the badger has moved or not. Table 4.1 gives a summary of the p-values from the logistic regression. See Appendix B for full output of the logistic regression. From Table 4.1, we are able to see that all the variables are significant when compared against the variable ‘moved’. Firstly, the sex of the badger and the age of the badger were considered separately, and the results showed significance. Secondly, the data were split into males and females, and then the test was run again. Again, both showed significance. This can be observed from Figure 4.6 that as the badger gets older (regardless of sex), then they are more likely to move. But, when comparing sex alone and if the badger has moved, then it is clear from the regression that males are more likely to move. This can be supported by [39], where, in general, it is observed that male badgers are more likely to move than females. Due to the format of the data, we are unable to take into consideration the seasons, but we would expect to see a difference between the genders in the different seasons.

Specifically, we would expect to see more movements from females during mating season.

Table 4.1: Logistic regression on data: Moved vs. Variable(s). Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Variable(s)	p-value
Sex	1.7e-15 ***
Age	<2e-16 ***
Female and Age	8.38e-12 ***
Male and Age	<2e-16 ***



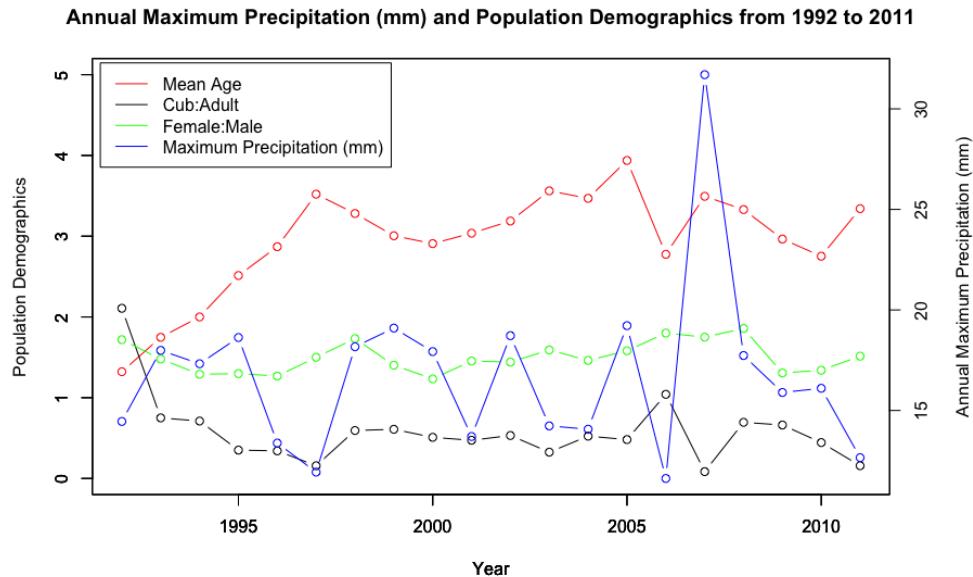
(a) Logistic regression: Female and Age vs. Moved. (b) Logistic regression: Male and Age vs. Moved.

Figure 4.6: Logistic regression of different variables from the data. The diamond symbols indicate observations, which are overlapped due to the amount of them.

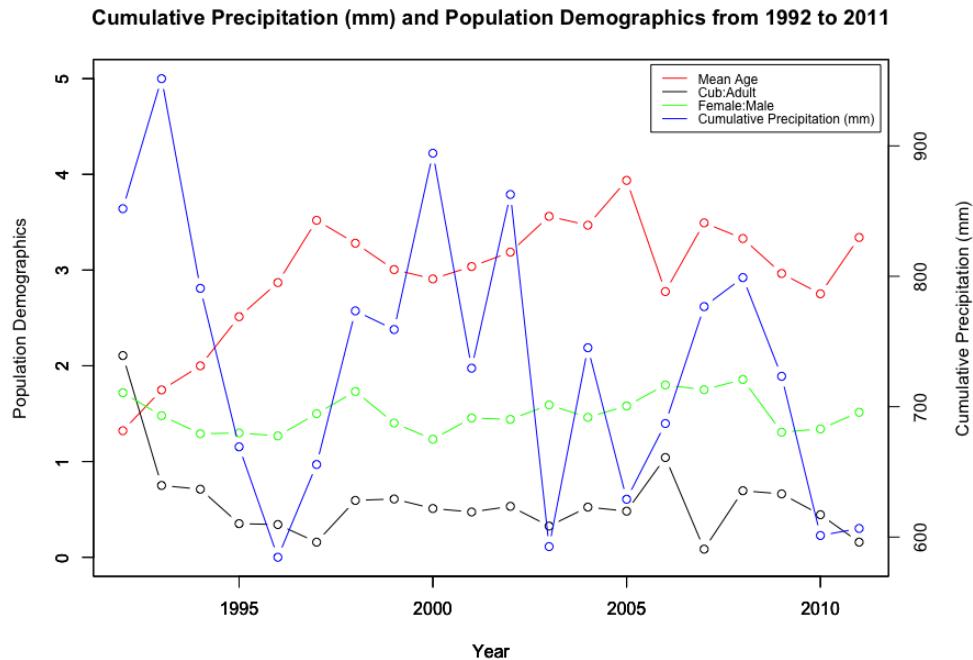
## 4.2.2. Weather Data

Weather records to accompany the capture-mark-recapture data are supplied by the Met Office for the time period 1988 to 2019. Three postcodes are supplied (GL1 3NN (Gloucester Hospital), GL53 7AN (Cheltenham Hospital), and GL10 3RF (Stonehouse)) and are made up of daily information on variables, such as precipitation and soil temperature. It has been shown that the simplest approach for spatial interpolation is to choose the nearest weather station and adjust for elevation [46]. The data have already been adjusted, hence, we use the altitude regression adjusted estimate for the postcode GL10 3RF.

Focusing on precipitation, we can calculate the annual maximum series and the annual cumulative sum. One notable day for precipitation is the 20 July 2007, where there was a measured value of 31.70 mm. This event coincides with the 2007 United Kingdom floods. Hence, placing this event as an extreme weather event. Despite this notable measure of precipitation on that day in 2007, the cumulative sum of precipitation for that year was not the highest in comparison to the



(a) Plot to compare the annual maximum precipitation (mm) and the population demographics.



(b) Plot to compare the annual cumulative precipitation (mm) and the population demographics.

Figure 4.7: Comparing precipitation from the postcode GL10 3RF with population demographics. The left axis is used for the population demographics, including the mean age of the badgers caught (years) and the ratio of cub:adult and female:male. The right axis details the precipitation (mm).

time frame 1992 to 2011. The year 1993 recorded the highest cumulative precipitation, regardless that there was no notable extreme weather event.

It has been identified that a key feature of a badger's diet are earthworms [5], depending on availability. Additionally, there are theories to suggest that earthworms are more likely to come to the surface when it is raining. Hence, it is more favourable for the weather to include rain to encourage the earthworms up to the surface for the badgers to eat (as seen in Figure 4.4b). Figure 4.7 details the comparison of precipitation recorded at the postcode GL10 3RF with the population dynamics. The impact of low precipitation on the badger population could be highlighted in Figure 4.7a, where in 2006 the maximum precipitation was quite low and increased very dramatically in 2007. As noted, the year 2007 is known for high precipitation, with floods seen in the UK. Additionally, from Figure 4.7b it can be seen in 2005 there was a drop in the cumulative precipitation, which could suggest implications for later years. Nevertheless, in 2007, the cub to adult ratio dropped in comparison to other years, suggesting there could be a link to the low maximum rainfall the year prior that influenced this ratio or the overall low cumulative rainfall in 2005.

Currently, there appears to be little known about the impact of weather/climate on fine-scale badger movements. A study, by Macdonald et al. [26], was conducted at Wytham Woods, 5 km north-west of Oxford, with the main objective to evaluate how the life history indices of badgers correlate with climatic indices. They observed a strong predictive relationship between survival and both temperature and late summer rainfall. However, the study was not interested in how the weather impacts the movements of the badgers, or how this could influence the spread of bTB. Therefore, leaving room for further research.

### 4.3. Data Wrangling Issues

The GPS data from APHA were collected between 2018 and 2021, where 73 badgers (41 females and 32 males) were monitored over the period across the Park. The data are made up of the variables; GMT Time, latitude, longitude, altitude, duration, dilution of precision (DOP), satellites, and cause of fix. The variable duration refers to how many seconds it takes for the collar to connect with the satellite to record the fix. It is programmed to take the location every 35 minutes, where the satellites attempt to connect for up to 70 seconds. If the duration is longer than 70 seconds, then the connection has effectively timed out and that record is skipped. The DOP is an indicator of the accuracy of the fix at the time it is obtained, where the higher the value, the less accurate the fix. The variable satellites detail how many satellites were connected during the fix. A Spearman's rank-order correlation was run to determine the relationship between the amount of satellites connected during a fix and the DOP. There was a strong, negative correlation between amount of

satellites and DOP, which was statistically significant ( $r_s = -.772$ ,  $p < .001$ ). This suggests the fewer satellites that connect, the higher the DOP and the location is less accurate. This is taken into consideration when cleaning the data, which is described in the next paragraph.

In total, 36,742 data points were collected over the period 2018 to 2021. However, the data had not yet been cleaned. To keep in line with how the data from ZSL were cleaned, which Woodroffe et al. discuss the reasoning in the supplementary of [50], we decided to filter in the same way. This means that any data points where the DOP was greater than four is excluded, along with any data points where less than four satellites connected. This also corresponds to the strong, negative correlation between the amount of satellites and DOP when fixing the location of the badger. With these filtered, we were left with 31,017 data points (15.6% excluded). Then, we looked to filter out any data points that were highly improbable places, for example in the lake (6 data points). Finally, in a study from 2007, badgers have been recorded as moving up to 26.2m/min [12], which is equivalent to an expected maximum movement distance of 917m between locations recorded 35 minutes apart (the location frequency used by APHA to collect the data). Hence, looking through the data, any distance over 900m in less than 35 minutes and any distance over 1000m in less than 38 minutes were excluded. Additionally, we calculated the distance and time interval between successive locations and excluded locations that were greater than 1500m from both the preceding and subsequent locations. In total, this leaves 30,996 data points with 15.64% of the data excluded.

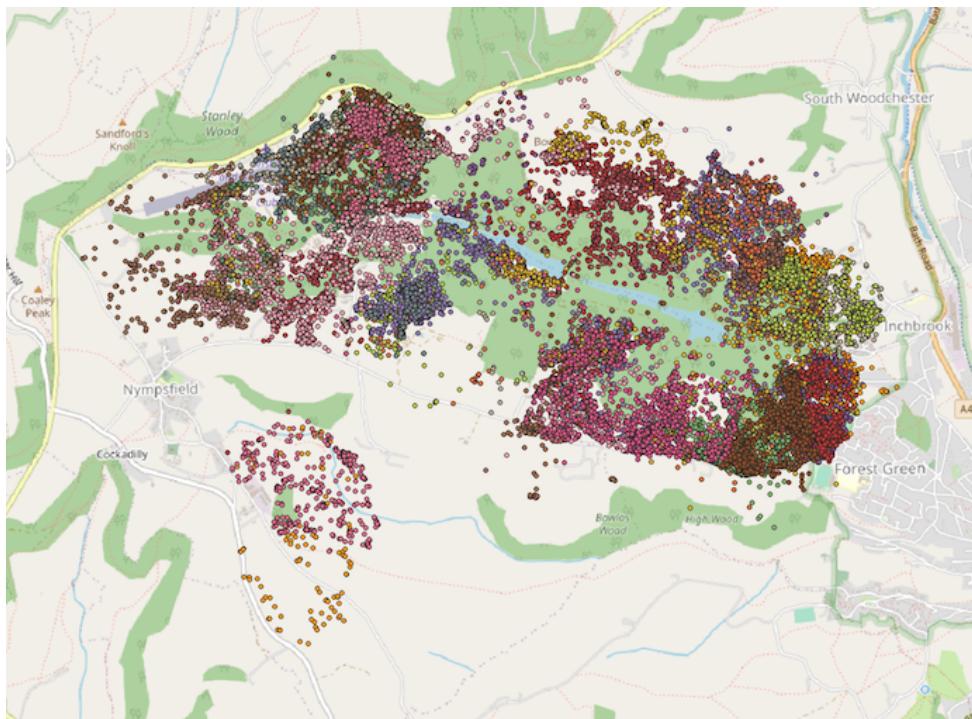
Furthermore, it was found when investigating the data that the months June and July are missing for any badger over the period 2018 to 2021. When conversing with the ecologists at APHA, it was found that they only deployed collars from August each year, as this is when they would receive new ones. Even though there would be a trapping event in May, very few or no badgers were collared at these events. Additionally, it is observed if the badgers were wearing collars from the previous season, then the battery stopped working by June. This could pose a problem if we would like to investigate if the noise changes seasonally. Luckily, we do not have this problem with the ZSL data, where all months are present within the data.

Additionally, the data from APHA use the longitude/latitude coordinate system. Milner, Blackwell and Niu aim to develop a model that takes into account a flexible social framework, i.e. looking at the impact of hierarchy in social animals [30]. They test their model with baboon data (available on Movebank [8]), whereby Milner, Blackwell and Niu convert the GPS coordinates to the UTM (Universal Transverse Mercator) coordinate system in R [36]. One main advantage of using the UTM coordinate system over angular coordinate systems like latitude and longitude, is that the coordinates are measured in meters, so simple Cartesian coordinate mathematics can be used over spherical trigonometry. For this reason, we adopt the same R method and convert the badger latitude and longitude coordinates into UTM coordinates (Northing and Easting, respectively).

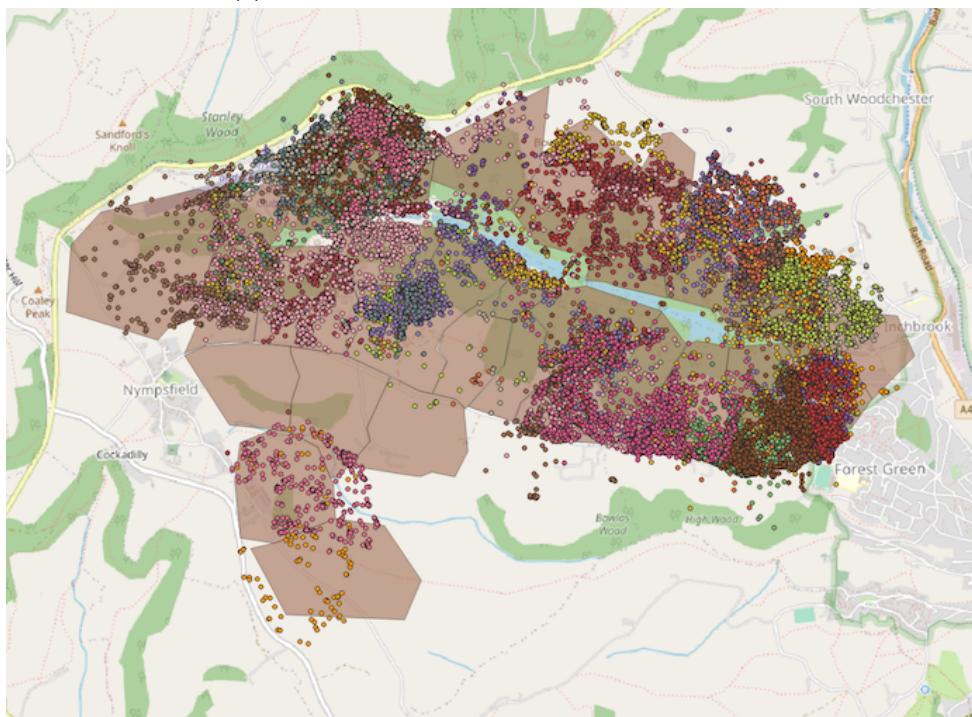
#### 4.4. Map and QGIS Data

Using a geographic information system, QGIS, we are able to view the geospatial data. QGIS also allows for editing and analysis of the data, for instance adding shapes to represent territories. Figure 4.8 is a graphic summary of the data, where a rough estimates of the territories from 2018 have been added (Figure 4.8b). It can be observed from the figures, the data are skewed to on the east of the park, with some parts of the park empty. When consulting with the ecologists at APHA it was stated that their method for collaring badgers was any healthy badger that got caught would have a collar put on. Clearly, more badgers were caught in the east than the west, so they had to change strategy in order to put more collars on badgers from the west. Furthermore, looking at Figure 4.8b, it can be seen there are three territories where very little data are obtained, which causes a split in the park. Consulting with the ecologists, it was stated that they did originally deploy a few collars in those territories, however it was hard to retrieve the collars (and therefore the data) from these locations. So, it was decided to not deploy collars in those areas and save them for other locations. Recently, they have obtained more collars so if badgers are now caught in those areas, they will be collared. This gap between the north of the park and the cluster in the bottom left should not cause too much problem for our analysis, as a clear energy well should be created from it.

We employ a mathematical tool, called  $k$ -means clustering, to partition the data into  $k$  clusters. This will allow for an estimate of territories (and wells) that we expect to see within the data. The Matlab code *kmeans\_opt* [23] will be used to determine the optimal number of clusters, which uses the Elbow method. The results of the test suggest that the optimal number of clusters in the data is  $k = 9$ . Running the built-in Matlab function *kmeans*, with 100 replicates, we can visualize how the clusters are formed, see Figure 4.9. Figure 4.9a represents the clustering for the optimal amount ( $k = 9$ ). Figure 4.9b highlights the clustering when  $k = 28$ . This was chosen since there were 31 territories identified in 2018, however, by Figure 4.8b, three territories have no data captured in them, so in theory we have 28 territories among the data. This clustering looks similar to how the territories are detailed but are not perfect. Nevertheless, clustering the data gives us an idea of how the energy potential shall look.

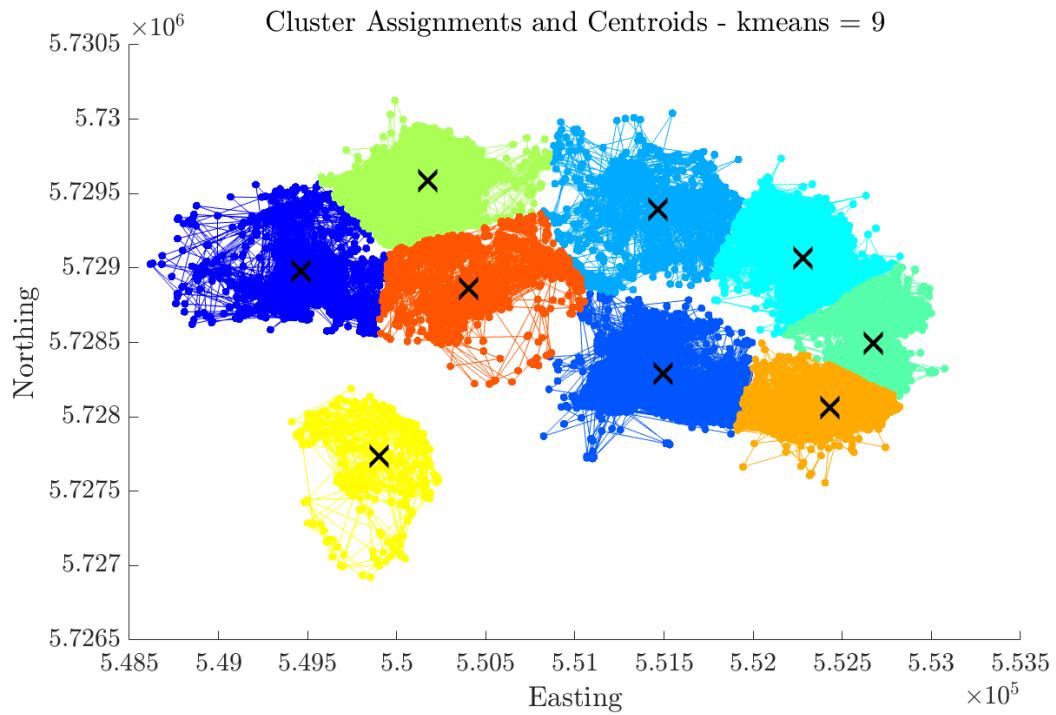


(a) All the data plotted on the map on QGIS.

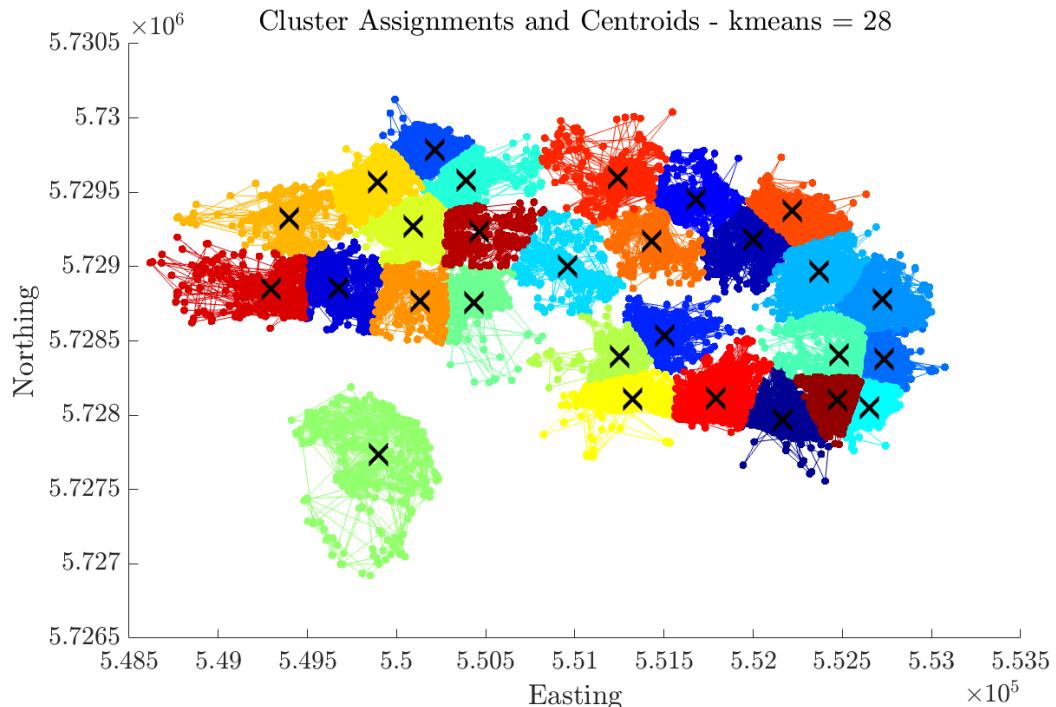


(b) Rough estimates of the territories from a 2018 bait map. Territories do split or merge from time to time, so today's map might look different.

Figure 4.8: 30,966 data points plotted on QGIS. Each colour represents a single badger's data set.



(a) K-means clustering with 9 clusters - the optimal number calculated with *kmeans\_opt* in Matlab.



(b) K-means clustering with 28 clusters - the number of territories in the park, minus three, in 2018.

Figure 4.9: K-means clustering, 100 replicates calculated with the best one chosen on Matlab.



## Generation of Badger Model with Data

The next step is to parametrize the model using the data described in Chapter 4 with the methods in Chapter 2. This will be explored and discussed in this chapter. We first generate the energy potential, followed by investigation into the calculation of the noise.

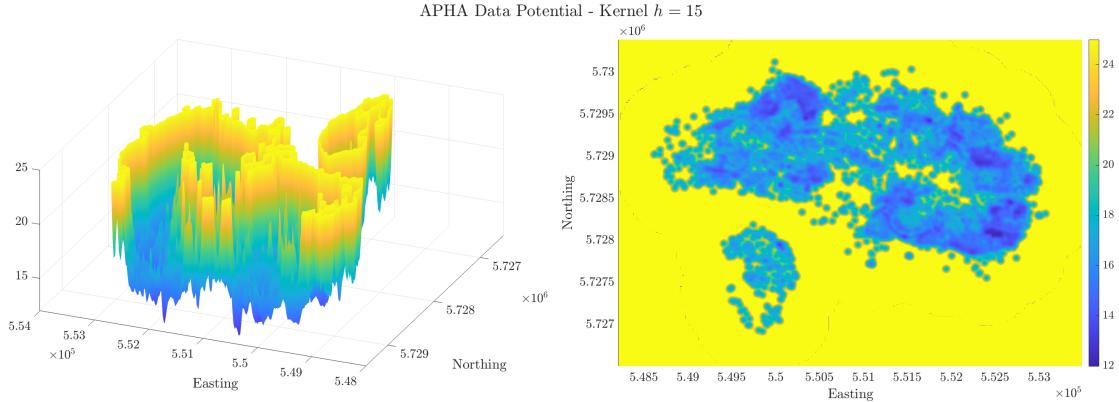
### 5.1. Generation of Energy Potential

First, the energy potential needs to be established from the data. The mathematical tool KDE is employed to establish the potential. Recall that the major criticism of the method is the error in the bandwidth selection. To overcome this, an estimate for the bandwidth has been calculated by the three different methods (Normal Reference, Cross-Validation Least Squares (CVLS), Cross-Validation Maximum Likelihood (CVML)), and can be seen in Table 5.1. All estimates have been calculated, including a downsampled sample where every other data point is kept. It took approximately 14.5 hours to calculate the estimate for the full data set using the CVLS method. As it can be observed from Table 5.1, the estimate for CVML is considerably smaller than both the normal reference and the top end of the CVLS estimate. Additionally, a large interval is given for the CVLS estimate. Nevertheless, the top end of this estimate does fall in line with the lower end of the normal reference estimates.

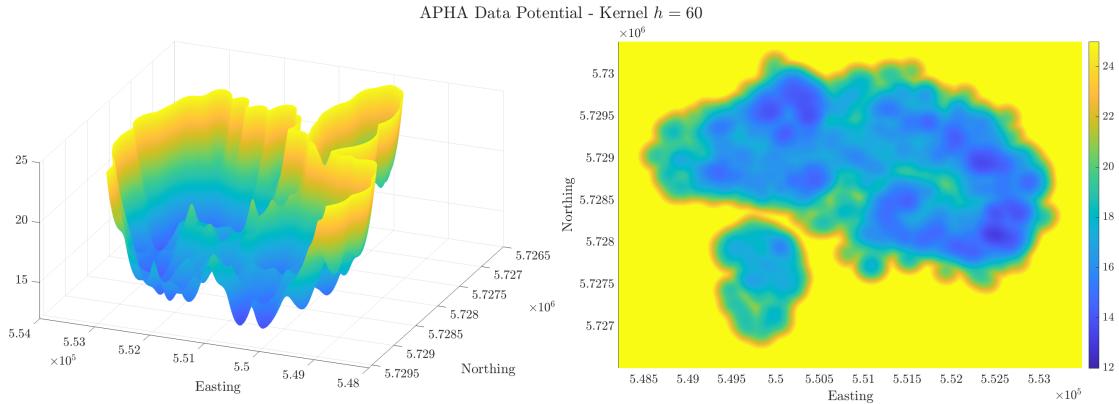
Table 5.1: Calculation of an appropriate Bandwidth for APHA Data using three different methods. These estimates have been calculated with the APHA data in UTM coordinates.

Data Points	Normal Reference	CVLS	CVML
30996	[113.5165, 208.4450]	[0.0000, 119.5537]	[13.8622, 15.3575]
15498	[127.3903, 233. 9909]	[0.0000, 134.1409]	[18.4416, 20.8666]

Since the drawback of KDE is the choice of the bandwidth, it is useful to compare different choices of kernels in order to help choose an appropriate energy potential. Again, it is useful to



(a) Generated energy potential with kernel  $h = 15$ . From the 3-Dimensional plot, it can be seen that the wells are very rigid - not smooth for a badger to move amongst.



(b) Generated energy potential with kernel  $h = 60$ . From the 3-Dimensional plot, it can be seen that the wells are less rigid. However, the base well is not connected to the main set of wells, which can also be seen in the colour plot.

Figure 5.1: Different choices of kernels to generate an energy potential are explored to find an appropriate choice for the APHA data.

plot the potential for different bandwidth choices to compare. Beginning with  $h = 15$ , which is approximately the top end of the CVML estimate, the potential generated is seen in Figure 5.1a. As it can be seen from this figure, it appears the energy potential is constructed from the individual data points. Hence, the wells are very rigid, i.e. it is under-smooth, and there is no connection between the set of data in the bottom left of the territory. Thus, this is an inappropriate choice for the kernel as it would not allow movements of the badgers between the setts.

Next, choosing  $h = 60$ , which is approximately halfway in the CVLS interval, we obtain Figure 5.1b. From Figure 5.1b we can see that the energy potential is smoother than when  $h = 15$ , however, the bottom set of data is still not connected to the main data set. This would be unrealistic for the data since there are two data spots from the main data set that visit the bottom left setts. Again,  $h = 60$  is an inappropriate choice for the kernel.

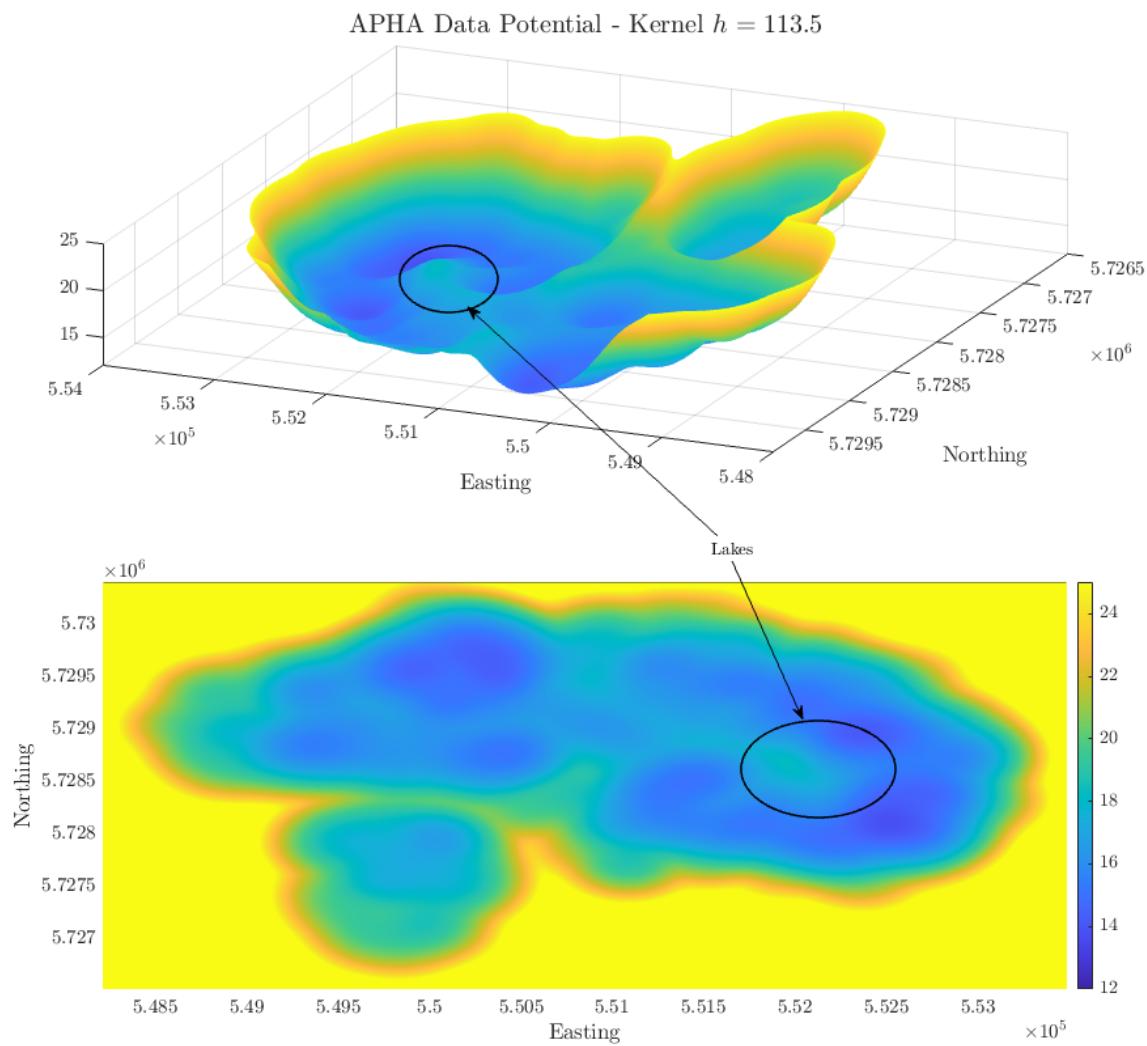


Figure 5.2: Generated energy potential using the data from APHA with a kernel of  $h = 113.5$ .

Increasing the kernel again, to the lower estimate of the normal reference method,  $h = 113.5$ , we obtain Figure 5.2. This appears to be the most ideal potential of three estimates. We are able to see wells, that are connected, and also a mound where the man-made lakes are. This is important, since we don't expect to see the badgers go into the lakes, but to take the lower mountain pass and go round. Additionally, in comparison to Figure 4.9a, it appears that there are similarities between the clusters and the number of wells. By increasing the value of the kernel, we will start to see increased smoothing of the energy potential and a loss of wells. Thus, this is the energy potential that will be used within the model.

## 5.2. Calculating the Noise

An advantage of GPS data is the amount of coordinate reference systems that are available, such as longitude/latitude and UTM. The data collected by APHA were given in longitude and latitude coordinates, so the coordinates were converted to the UTM system, then the noise estimated using the methodology stated in Chapter 2. The noise was firstly calculated whilst the data were in the original coordinate system. However, the numbers were very small, but in proportion with that system. On the one hand, this made it hard to discuss what this would mean ecologically as they appeared to be just numbers than a distance. On the other hand, when computing the noise in the UTM coordinates, we can relate the values to the distance the badgers travel in meters.

Figure 5.3 displays the distribution of the total data set. This includes a total of 24,551 data points (where approximately 20% have been discounted due to the large time gaps). It can be observed that the distribution of movements tends to the smaller movements, with the data looking zero-inflated. For now, we continue with the assumption that the data are normal and apply the methods in Chapter 2 to calculate the noise.

Using the methodology, we first try and approximate the matrix  $L$  for all individual badgers using all the data points and then take an average across all of them and repeat the process by sex. This is followed by approximating the same matrix, however, using the full data set. The results can be seen in Table 5.2. It can be observed that the averaged matrices are similar to the noise calculated using all the data. Additionally, it can be highlighted that, whilst there is not a lot of difference, the male noise is higher than female noise. This analysis is supported by the work by Rogers et al. [39]. However, upon running an independent t-test, the noise calculated is not statistically significant between gender. This was firstly tested elementwise, where only entry  $a_{11}$  was significant ( $t = -2.125, p = 0.037$ ). Secondly, calculating the Frobenius norm of each matrix and testing for difference with an independent t-test, we observe that it is not significant ( $t = -1.57, p = 0.122$ ). Overall, upon observation, whilst the male noise is higher, it is not statistically significantly higher.

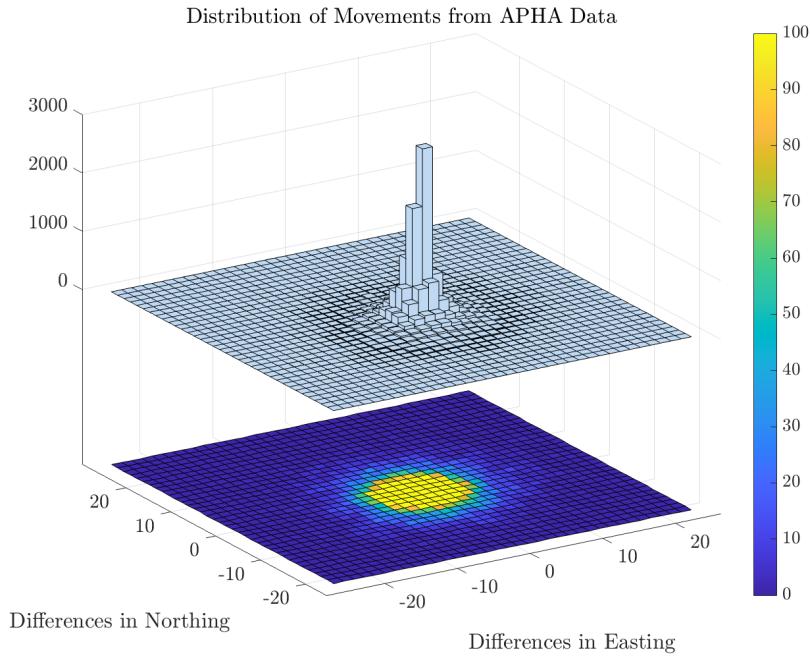


Figure 5.3: 3-dimensional histogram of the total APHA data, with 35 bins in each direction.

Table 5.2: Comparing the computed noise terms from individual noise averaged and the total data set. The matrices are the Cholesky decomposed from the estimated matrix  $a = \sigma\sigma^T$ .

	Average from individuals	Total Data Set
Total	$\begin{pmatrix} 21.1290 & 0.0000 \\ 1.2337 & 22.1221 \end{pmatrix}$	$\begin{pmatrix} 22.6361 & 0.0000 \\ 1.7451 & 22.3695 \end{pmatrix}$
Male	$\begin{pmatrix} 23.1140 & 0.0000 \\ 1.7506 & 23.0416 \end{pmatrix}$	$\begin{pmatrix} 23.4885 & 0.0000 \\ 2.0670 & 22.4880 \end{pmatrix}$
Female	$\begin{pmatrix} 19.5797 & 0.0000 \\ 0.8302 & 21.4044 \end{pmatrix}$	$\begin{pmatrix} 21.7238 & 0.0000 \\ 1.3899 & 22.2426 \end{pmatrix}$

To split this further, we can explore the differences in seasons and the estimated noise, seen in Table 5.3. The meteorological seasons are made up of three months each, which coincide with our Gregorian calendar. The seasons are defined as spring (March, April, May), summer (June, July, August), autumn (September, October, November) and winter (December, January, February). The estimations for the noise have been taken from all the data together which have been filtered for these months, rather than the individual badgers and an average being taken. It is clear from this table that the noise is higher in winter and summer in comparison to spring and autumn. It should be highlighted that summer only holds data for August, as due to the delivery of collars, no data were collected in June and July. If there were any collars already on the badgers, then the battery died by the end of May, so again, no data were collected for the months June and July.

Additionally, the data are not evenly spread across the seasons. Out of the total 30996 data points, 12.7% is from winter, 4.9% is from spring, 13.5% is from summer, and the remaining is caught in autumn with 68.9%. Again, due to obtaining the new collars in August, they begin collaring from then, so it is not surprising that most of the data are in autumn.

On comparison of sex within the seasons, it can be observed that it is only spring where the female noise is higher than the males. It has been stated by Neal and Cheeseman [32] that badgers can mate during any month of the year in Britain. However, it has been suggested in [51] that the most important period in the reproduction of British badgers is between February and May. It is during this period that observations of long-duration matings, both in the field and in captivity, are most frequent [32]. This period encompasses the season of spring; thus it is not surprising to see the females having a higher noise than males in this period.

Table 5.3: Comparing computed noise terms for the different seasons, where the whole data set has been used to estimate the noise rather than averages of individual badgers. The matrices are the Cholesky decomposed from the estimated matrix  $a = \sigma\sigma^T$ .

	Winter	Spring	Summer	Autumn
Total	$\begin{pmatrix} 25.9552 & 0.0000 \\ 0.3565 & 25.2002 \end{pmatrix}$	$\begin{pmatrix} 19.7398 & 0.0000 \\ 2.7093 & 21.9346 \end{pmatrix}$	$\begin{pmatrix} 24.8771 & 0.0000 \\ 2.3936 & 25.0332 \end{pmatrix}$	$\begin{pmatrix} 21.7964 & 0.0000 \\ 1.7930 & 21.3392 \end{pmatrix}$
Male	$\begin{pmatrix} 27.1587 & 0.0000 \\ 1.0455 & 24.0617 \end{pmatrix}$	$\begin{pmatrix} 18.4746 & 0.0000 \\ 2.7806 & 20.1317 \end{pmatrix}$	$\begin{pmatrix} 24.6466 & 0.0000 \\ 1.7188 & 25.2865 \end{pmatrix}$	$\begin{pmatrix} 22.9453 & 0.0000 \\ 2.3169 & 21.8471 \end{pmatrix}$
Female	$\begin{pmatrix} 23.9112 & 0.0000 \\ -0.8815 & 26.8527 \end{pmatrix}$	$\begin{pmatrix} 21.7512 & 0.0000 \\ 2.6227 & 24.7404 \end{pmatrix}$	$\begin{pmatrix} 25.0447 & 0.0000 \\ 2.8979 & 24.8214 \end{pmatrix}$	$\begin{pmatrix} 20.5958 & 0.0000 \\ 1.2211 & 20.8138 \end{pmatrix}$

A surprising element of Table 5.3 is the largest noise is seen in winter. Research suggests that badgers showed a marked reduction of activity in the winter period [12]. This would suggest that we should see a reduced noise matrix for winter. During winter, badgers do not hibernate, but they reduce their activity during periods of cold weather. The activity starts to increase at the beginning of the main mating season in late winter/early spring [32]. Yet, the noise has actually reduced for this period in comparison to winter. It will be interesting to see what observations will be made with other sets of data and if other locations are similar.

### 5.3. Location of Mountain Passes

To progress, we would like to locate the mountain passes in our potential. As expected, there are a lot of wells, so we focus on the bottom left of the potential, where there is a cluster of data away from the main body.

In order to locate the roots in the potential, we use the function `fsolve` on python, which is a package that returns the roots of the (non-linear) equations defined by  $function(x) = 0$ , given a

starting estimate. Recall from Chapter 2, Equation (2.14) and Equation (2.17), that an estimate for the potential  $V$  is

$$V = -\log(\rho),$$

with invariant density,

$$\rho = \frac{1}{2nh\sqrt{2\pi}} \cdot \sum_{i=1}^n \exp\left(-\frac{\|x - X_i\|^2}{2h^2}\right),$$

where  $n$  is the number of observations, and  $h$  is the bandwidth. Here, the function to use with *fso*lve is the negative gradient of the potential  $V$ , which is defined as,

$$\frac{\partial V}{\partial x} = \frac{1}{h^2 \sum_{i=0}^n k(x, X^{(i)})} \cdot \sum_{i=0}^n (x - X^{(i)}) K(x, X^{(i)}), \quad (5.1)$$

where,

$$K(x, X^{(i)}) = \exp\left(-\frac{\|x - X^{(i)}\|^2}{2h^2}\right).$$

Then, seven points are chosen around the pass between the bottom well and the main set of wells, which are detailed in Table 5.4. These are run through *fso*lve, where the results are contained in the same table and shown graphically in Figure 5.4. Arrows have been plotted onto the figure to help guide the estimation to the calculated root. The calculation of the root from estimation 2 is what we were expecting for a mountain pass. We would assume that there would be a possible mountain pass in the middle between the two sets of wells to allow for the badgers to move between both. However, we were not expecting the result from estimation 6, which lead to another root, inside the well. Similar for points 1, 3, 4 and 7, which also locate minimums not along the border. Further investigation is needed in order to determine if there are any more minimum points along the border and if the points located are indeed mountain passes. This will require calculating and coding the Hessian matrix for the energy potential. This is currently in the process of being done and is yet to be completed. Whilst the investigation into these mountain passes are estimates, it would be interesting to see ground-truth with another visit to Woodchester. Nevertheless, this is an interesting start to the investigation of large deviation theory in the energy potential.

Table 5.4: Estimation of roots followed by the results using the python package *fsolve*. This table corresponds to Figure 5.4, where the numbers by the red dots relate to the initial estimation and the red star are the roots found.

	Estimations		fsolve Roots	
	Easting	Northing	Easting	Northing
1	550000.00	5728500.00	549862.39	5728796.80
2	549928.00	5728290.00	549858.32	5728313.16
3	549758.00	5728160.00	549668.98	5727886.91
4	549567.00	5728150.00	549669.28	5727886.82
5	550161.00	5728020.00	550039.59	5727931.33
6	549864.00	5728380.00	550039.59	5727931.33
7	549673.00	5728430.00	550109.19	5728976.48

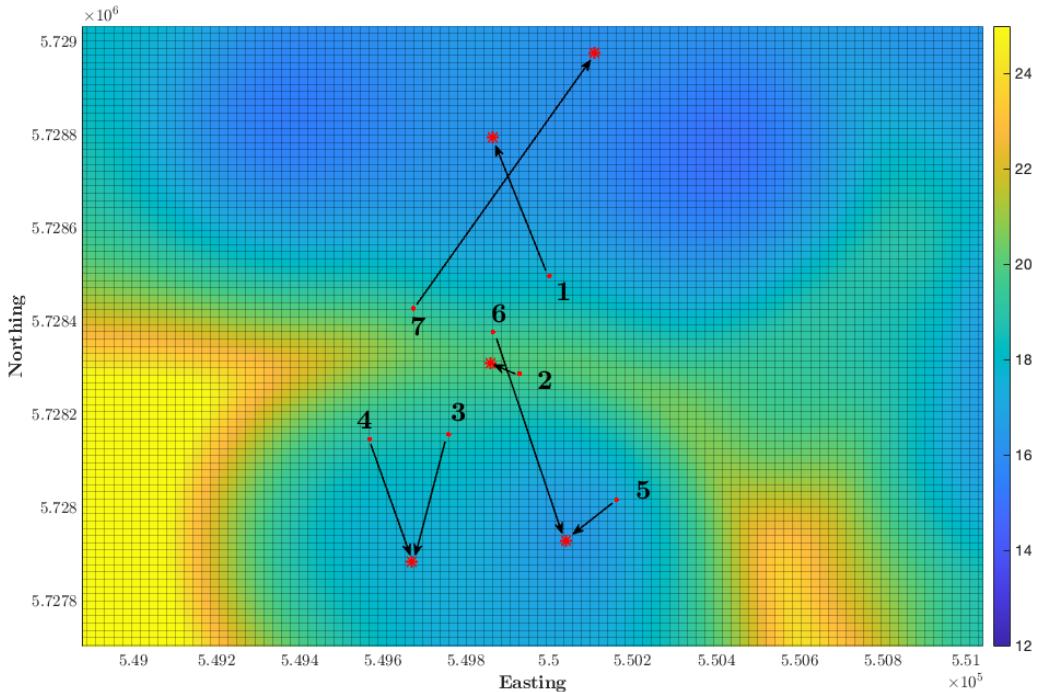


Figure 5.4: Estimation of roots followed by the results, calculated in python, for the bottom left of the potential. This figure corresponds to Table 5.4, where the numbers by the red dots relate to the initial estimation and the red star are the roots found. The arrows guide the estimation to the root found. It appears that estimation 3-6 found the base of wells and 1,2, and 7 found mountain passes. These are yet to be confirmed.

## 5.4. Simulation in the Energy Potential

Finally, we would like to run some simulations with the generated energy potential and estimated noise. We generate the trajectories with the same initial position ( $\mathbf{X}_0 = [550048, 5727980]$ ) and

noise,

$$\sigma = \begin{pmatrix} 22.63613 & 0.0 \\ 1.74510 & 22.36947 \end{pmatrix}.$$

The initial position was chosen randomly, so it lay in a well in the bottom left of the potential, but not at a base of a well. This choice was made because we would expect to see the badger to stay in the well when there are only a few iterations, and with a sufficient number of iterations we would hope to see the badger cross to the main set of wells. The noise has been chosen from Table 5.2. Using Euler-Maruyama, we obtain Figure 5.5 and Figure 5.6. Here, we consider one iteration as one minute. In one night (between 20:00 and 06:00 there are 10 hours, or 600 minutes), hence, we consider trajectories with 600 iterations.

However, firstly, we consider a long trajectory, as seen in Figure 5.5, where 100,000 iterations are completed. This is equivalent to approximately 167 nights, with the aim to see the behaviour of the badger in the energy potential over a long number of iterations. Calculating the Euclidean distance between each time step, it is found the total distance travelled by the badger in 100,000 iterations is 631,451.87m. In the space of one iteration, the smallest distance travelled is 0.018640m and the largest distance is 23.970204m. Recall, in a study from 2007, badgers have been recorded as moving up to 26.2m/min [12]. Hence, we do not want the longest distance travelled to be over the value 26.2m. Experimenting with the step size found that the value of 0.05 was ideal, as the farthest distance travelled in one iteration in Figure 5.5 is 23.97m. When the step size had

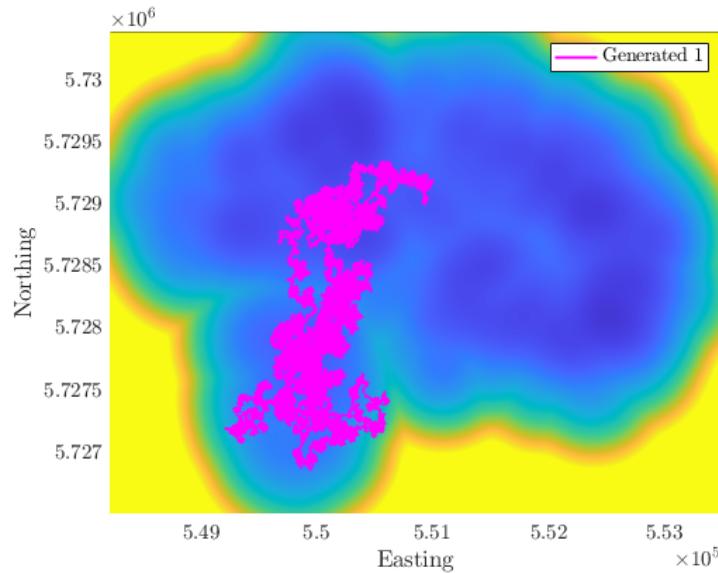


Figure 5.5: A long trajectory with 100,000 iterations. The badger moves between different wells, which is what we wanted to see. Total Euclidean distance travelled is 631,451.87m. The smallest distance travelled in one iteration is 0.018640m and the largest distance travelled is 23.970204m.

been set larger, i.e. 1.0, then the largest distance travelled became 94.5814m. Clearly, not being ecologically possible for a badger to travel this distance in one minute. Additionally, when the step size was set to large, then it was found that the badger traversed outside the energy potential, which shouldn't be possible. From Figure 5.5 we can see that the badger stays within the potential and does travel between different wells, which is what is expected. Dimensional analysis needs to be completed in order to determine the correct scaling for the step size, but current analysis suggests we are on the right track,

Figure 5.6 highlights several trajectories. The simulation was run ten times, for 600 iterations each, with Table 5.5 detailing the distance statistics for each run of the simulation. It should be noted that these simulations were generated one after the other, so there is no other interaction for the trajectory other than the energy potential. The white star in Figure 5.6 represents a calculated root, which is suspected to be the base of the well. It can be observed in this figure, that all generated trajectories remain in the well and close to the base, where Figure 5.6b allows for a zoom-in of Figure 5.6a to help see the trajectories a bit clearer. Additionally, it appears that the badger did not leave the well in any of the trajectories, but further analysis of the mountain passes will allow for further conclusion on this.

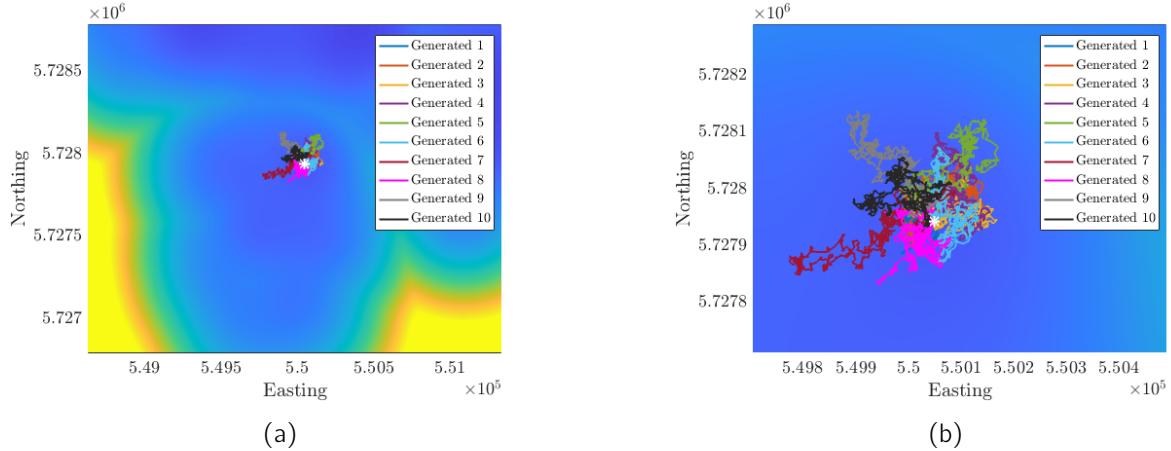


Figure 5.6: (a) Simulation of a badger for 600 iterations, repeated ten times. The white star represents the located root. It appears the trajectories stay around the base of the well. (b) Zoom in on (a).

Overall, we have been able to obtain a simulation of a badger in the energy potential, and now requires further analysis to gain useful ecological insight into the fine scale behaviour of badgers.

Table 5.5: Distance statistics relating to Figure 5.6. On average, the badger travelled 3731.18m across the 600 minutes (approximately 6.22 meters/minute).

Generated	Total Distance (m)	Distance in One Iteration (m)	
		Shortest	Longest
1	3764.5635	0.5185	18.3662
2	3883.0892	0.2718	19.4248
3	3664.8360	0.4823	16.7761
4	3774.9458	0.0784	17.8217
5	3701.3974	0.0698	17.6916
6	3736.2312	0.2955	17.9149
7	3634.6805	0.0893	18.1896
8	3728.2537	0.0757	18.6083
9	3708.0399	0.2641	17.3345
10	3715.7728	0.2329	18.4560



# 6

## Conclusions and Outlook

In this report, we have travelled through the beginning of the project on mathematical data analysis of fine scale badger movement. Firstly, in Chapter 2, we outlined the methods that have so far been used and shall be continuing to use within the project. Here, we began by discussing the concept of a badger in an energy potential and what this potential means for us. The concept of using an energy potential for modelling animal movement is not new, but it is novel within the badger community. We progressed by discussing large deviation theory, which we started to discuss further in Chapter 5. Within Chapter 2 we also defined our stochastic differential equation and explored trajectories in the potential and other equations of motion. We finished the chapter by going into detail of the data methods, including KDE, Kramers-Moyal Formulae and transfer operators. These methods have played a significant role in the make-up of the model, in calculating the energy potential from the data and estimating the noise.

The journey continued in an exploratory stage within Chapter 3, where different strategies for modelling were discussed. We started by considering five badgers in a double-well potential with an added behaviour potential. The behaviour potential represents attraction and repulsion between the badgers, as discussed in Chapter 2. However, as identified in the chapter, it is not clear if this behaviour matrix is obtainable from the data. We progressed this model by considering more ecological assumptions. For example, we considered the attraction between males and females within the behaviour matrix, rather than the matrix being random. We ended the chapter by adapting the model again to include an infection model. This was a very simple infection model that does not reflect the true nature of bTB, however, the aim of this analysis was to focus on the technique of creating an infection model on top of the movement model rather than results. If we are to do analysis on bTB, this will allow for a starting point.

Next, in Chapter 4, we explored the obtained data. Primarily, there are two sets of data, where the focus has been on the data from APHA for this report. We discussed how the data were collected, which included an account of the visit to the location in December 2021. Then,

we completed a preliminary data analysis on a subset of capture-mark-recapture data and weather data from the same location of Woodchester Park, Gloucester. This was followed by issues that we had with data wrangling, including the cleaning of the data and transforming the coordinate units from longitude and latitude to UTM. We ended the chapter with the plotting of the data in a data plotting tool called QGIS, where we also considered k-clustering of the data. On comparison of this clustering to the energy potential in Chapter 5, we are able to see a pattern in the location of the wells.

Finally, we have employed the methods outlined in Chapter 2 to the APHA data, in Chapter 5. Building the model, we firstly generated an energy potential from the data. On comparison of the three estimation algorithms for the bandwidth, it was found an ideal kernel size comes from the low end of the normal reference estimate. The method CVML gave an estimate that under-smoothed the data, and CVLS gave a broad estimate. Yet, the top end of CVLS did fall in line with the low end of normal reference, which helped in the choice. The final aspect required to build the model is the noise term. To do this, we first calculated the noise for each badger separately and then took an average. Then, we calculated the noise using the full data set together. By comparison, they are similar. We were also interested to see how the noise value changed seasonally, so, to further investigate the noise, we split the data into metrological seasons and estimated the noise. As expected, during spring, females were seen to have a higher noise than males, where spring is seen to be an important season for mating. Surprisingly, the noise term in winter was high, where it would be expected that badgers reduce activity in these months. Hence, further investigation is needed on this, and comparison with other data sets of badgers. We ended the chapter by starting the investigation into large deviation theory by locating a couple of mountain passes. Yet, these are to be confirmed through calculating the Hessian and determining the sign of the eigenvalues. Additionally, we also computed some simulations in the found energy potential and estimated noise. Again, further analysis is required to this chapter, but it is a start.

This leads to the outlook for the remaining of the project. The first stage is completing the analysis that has been started in Chapter 5. This includes using transfer operators, such as EDMD, on the data set to investigate the behaviour of the dynamical system and large deviation theory to confirm mountain passes. As mentioned in Chapter 4, we also have a set of data from Cornwall, where we will apply the methods used with the APHA data in Chapter 5, to the Cornwall data. This analysis partly answers the two part question; can we generate a dynamical model to explore badger movement and, how does behaviour change based on weather/season/climate? We have reflected on seasons within Chapter 5 but are yet to consider the weather/climate. Another question that we hope to answer is, can we calculate transition times of badgers between social groups? With further analysis in the current model, we should be able to calculate some transition times, which relate to how long a badger will be in one social group before visiting another, and before returning

to the original sett. We expect the main analysis on the data sets to be completed within six months, which is taken into account in the Gantt chart in Appendix C.

Once that analysis is completed, we would like to focus on coherent sets behaviour detection, using EDMD. Ecologically, this would be looking, for example, if there is a sett of badgers, do they remain as that sett or do they disperse. This is one of the main research questions that we hope to answer. This is to help in the understanding of badger social behaviour. Additionally, they can be useful and relevant in the discussion and understanding of the spread of bTB within a population. By following the Gantt chart in Appendix C, we aim for this analysis to take six months, whereby we also aim to write a paper if there are any interesting ecological findings.

Finally, we would like to apply different approaches to derive the model, which also looks at answering the question, can we generate a dynamical model to explore badger movement? The different approaches include data-driven methods, such as gEDMD (an extension of EDMD), and Ensemble Kalman filtering. Whilst gEDMD does require derivatives, which is not available for the badger data, there is a possibility that this could be estimated through mathematical tools like Koopman lifting. The other method, Ensemble Kalman filtering, would allow us to pose the inference of the SDE model as an inverse problem, where we would like to parametrize the model (both the energy potential and the noise) at the same time. These methods are to be followed with uncertainty quantification analysis, which aims at quantifying the variability of the output that is due to the variability of the input. We expect this to take longer to complete, with an estimation of taking a year. Again, we would like to publish a mathematical paper if there are any interesting findings.

Overall, the foundations have been set for a model that will allow for some analysis into the behaviour of fine scale badger movement. Further analysis on the model is yet to be completed, with an interesting path to be taken for the future.



## **Appendices**





## Methods

### A.1. Euler-Maruyama

The Euler–Maruyama method is a method for the approximate numerical solution of a stochastic differential equation (SDE). It is an extension of the Euler method for ordinary differential equations to stochastic differential equations.

Suppose we wish to solve SDE (2.2) on some interval of time  $[0, T]$ , with initial condition  $\mathbf{X}_0 = x_0$ , where  $\mathbf{W}_t$  stands for the Wiener process. Then the Euler–Maruyama approximation to the true solution  $\mathbf{X}$  is the Markov chain  $\mathbf{Y}$  defined as follows:

- partition the interval  $[0, T]$  into  $N$  equal subintervals of width  $\Delta t > 0$  ::

$$0 = \tau_0 < \tau_1 < \dots < \tau_N = T \text{ and } \Delta t = T/N;$$

- set  $\mathbf{Y}_0 = x_0$
- recursively define  $\mathbf{Y}_n$  for  $0 \leq n \leq N - 1$  by

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + b(\mathbf{Y}_n, \tau_n)\Delta t + \sigma(\mathbf{Y}_n, \tau_n)\Delta \mathbf{W}_n,$$

where

$$\Delta \mathbf{W}_n = \mathbf{W}_{\tau_{n+1}} - \mathbf{W}_{\tau_n}.$$

The random variables  $\Delta \mathbf{W}_n$  are independent and identically distributed normal random variables with expected value zero and variance  $\Delta t$ .





## Logistic Regression

Below are the outputs of the logistic regression for different combinations, where the p-values are summarized in Table 4.1.

Firstly, we investigated the statistical association between if they moved to the sex of the badger. From Table B.1, we can see that  $1.7e - 15 < 0.05$ , and hence, is significant. Suggesting that the sex is significant when it comes to if they have moved or not.

Table B.1: Logistic Regression results for 2478 badgers observations between 1992 and 2011, investigating statistical association between if they have moved and the sex of the badger.

Predictors	Estimates	Std. Error	z Value	$P >  z $
(Intercept)	-1.05021	0.07192	-14.603	$< 2e - 16$
Sex	0.10553	-7.962	-0.84016	$1.7e - 15$

Secondly, we investigated the statistical association between if they moved to the age of the badger. From Table B.2, we can see that  $(< 2e - 16) < 0.05$ , and hence, is significant. Suggesting that the age of the badger is significant when it comes to if they have moved or not.

Table B.2: Logistic Regression results for 2478 badgers observations between 1992 and 2011, investigating statistical association between if they have moved and the age of the badger.

Predictors	Estimates	Std. Error	z Value	$P >  z $
(Intercept)	-2.29104	0.09373	-24.44	$< 2e - 16$
Age	0.24232	0.02140	11.32	$< 2e - 16$

Thirdly, we investigated the statistical association between if they moved to the age of the female badgers. From Table B.3, we can see that  $8.38e - 12 < 0.05$ , and hence, is significant. Suggesting that there is a statistical association between the age of the female badgers in the sample and whether they have moved.

Finally, we investigated the statistical association between if they moved to the age of the

Table B.3: Logistic Regression results for 2478 badgers observations between 1992 and 2011, investigating statistical association between if they have moved and the age of the badger when female.

Predictors	Estimates	Std. Error	z Value	$P >  z $
(Intercept)	-2.58329	0.13784	-18.742	$< 2e - 16$
Age	0.19433	0.02844	6.832	$8.38e - 12$

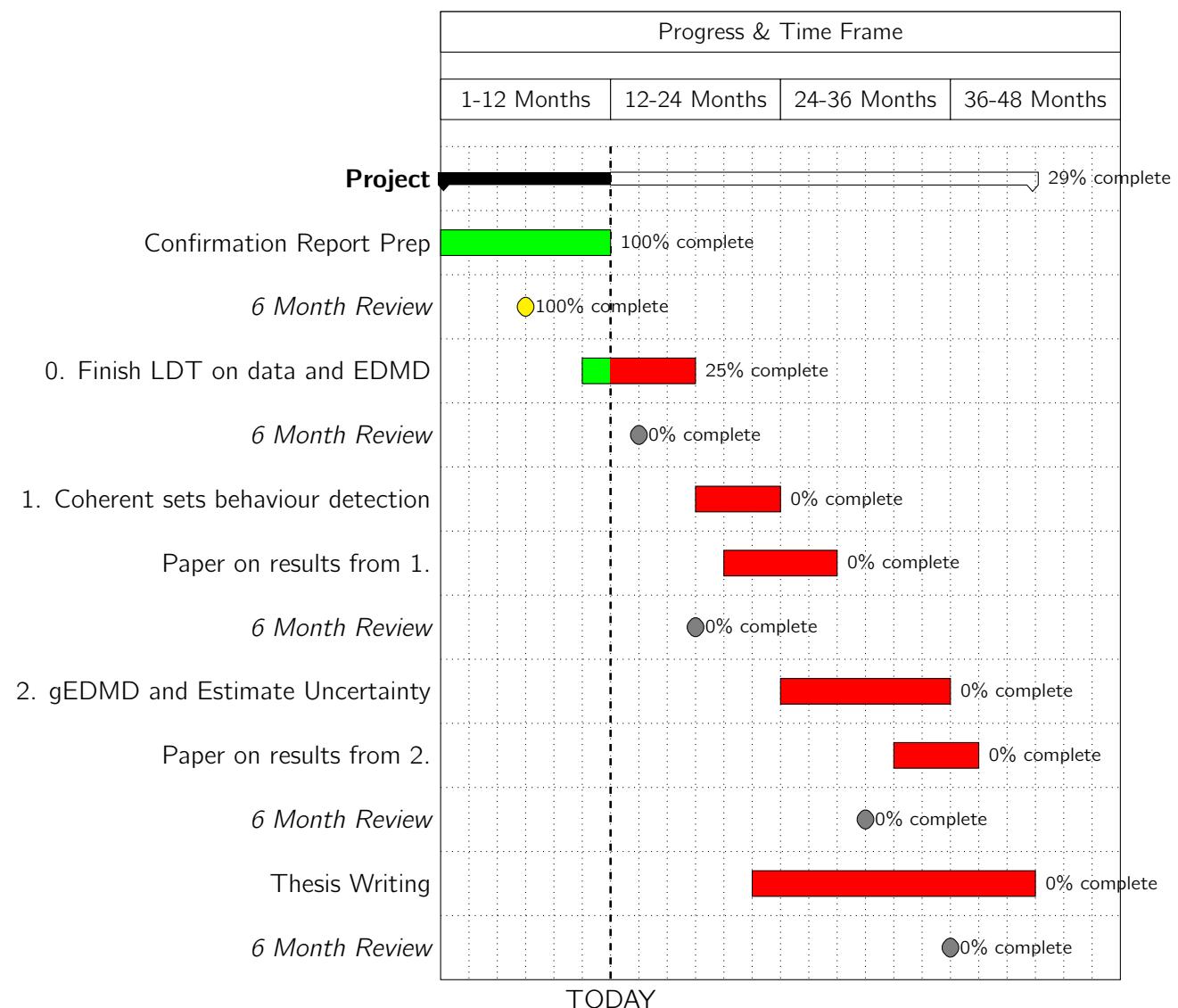
male badgers. From Table B.4, we can see that ( $< 2e - 16$ )  $< 0.05$ , and hence, is significant. Suggesting that there is a statistical association between the age of the male badgers in the sample and whether they have moved.

Table B.4: Logistic Regression results for 2478 badgers observations between 1992 and 2011, investigating statistical association between if they have moved and the age of the badger when male.

Predictors	Estimates	Std. Error	z Value	$P >  z $
(Intercept)	-2.37613	0.14780	-16.08	$< 2e - 16$
Age	0.46113	0.04182	11.03	$< 2e - 16$



## Conclusions and Outlook - Gantt Chart







## Data Management Plan

Below, find the Data Management Plan, which is a written document outlining how I am planning to manage my research data both during and after the project. The plan addresses what types of data will be collected and how the data will be documented, stored, shared and preserved.

# **Mathematical Data Analysis of Fine Scale Badger Movement Data**

---

## **Data Collection**

### **What data will you collect or create?**

GPS data of badger locations from different sites in the United Kingdom. When visiting these sites, I will collect images and videos. Programming code, numerical data and images shall also be created.

### **How will the data be collected or created?**

The GPS data has already been collected in previous years by chosen different organisations. When visiting sites, I shall collect the images/videos on my personal mobile phone. This shall be uploaded to OneDrive in a folder for that visit. Numerical results and images will be created as a result from programming code created for the report. This will be based on the GPS data already collected.

## **Documentation and Metadata**

### **What documentation and metadata will accompany the data?**

All code created has been clearly commented, so others will understand how it works and what it will produce. This will be alongside the code. Documentation of the GPS data shall be recorded within the report to include details on methodology of cleaning the data and units of measurement (and if they have changed).

## **Ethics and Legal Compliance**

### **How will you manage any ethical issues?**

From all organizations that have granted use of their data, no consent has been granted to share the data further. The data from ZSL hold sensitive location data, so the locations of the data shall be changed to have anonymity. All the data shall be stored securely on OneDrive.

### **How will you manage copyright and Intellectual Property Rights (IPR) issues?**

The data owned by ZSL has restrictions on, so no sharing of the locations or posting of the data is allowed. Any models created with this data will have coordinates changed. The data owned by APHA has similar restrictions on posting the data, however is not restricted on the sharing of the locations. Any data created by code is owned by myself or my supervisors.

## **Storage and Backup**

### **How will the data be stored and backed up during the research?**

The data will be stored on the University OneDrive and is backed up daily to keep the data safe. Code is stored on the University

Gitlab, and updated weekly or at change of code. A separate personal storage device is used as an external backup, which is done weekly.

#### **How will you manage access and security?**

Data is accessed on a machine which is logged in to the Onedrive (2-Factor Authentication) or Gitlab account, namely the work laptop. Nobody but the creator of the code has access to this machine. The personal storage device is password protected, so no access can be gained without it.

## **Selection and Preservation**

#### **Which data are of long-term value and should be retained, shared, and/or preserved?**

The code data written for the report is of long term value and will be retained and preserved, as it can be applied to other data sets. At the end of the project, the GPS data will need to be destroyed, unless granted use for further use.

#### **What is the long-term preservation plan for the dataset?**

The code data will be preserved in the cloud through the University and in the Github. The code will be commented so the use of it is clear.

## **Data Sharing**

#### **How will you share the data?**

The raw GPS data will not be shared outside the supervisor team. Within the report, it shall be stated where the data is from should it be requested, but it is not within my right to share it. Sharing between the supervisor team will be done via a secure OneDrive Link or Github for the code.

#### **Are any restrictions on data sharing required?**

Yes - the location of the ZSL data is not to be released. This shall be overcome by changing the locations within the model, so anonymity is kept.

## **Responsibilities and Resources**

#### **Who will be responsible for data management?**

I, Jessica Furber, is responsible for the DMP, ensuring it is reviewed and revised accordingly.

#### **What resources will you require to deliver your plan?**

No additional resources are required.



# *E*

## **Seminar Hours and Extended Abstracts**

Below, find the submission of the student seminar participation form for 20 hours towards the broadening training, followed by two extended abstracts, as required.

---

---

### PGR STUDENT SEMINAR PARTICIPATION FORM

---

Student Name: Jessica Furber

Supervisor Name: David Lloyd

Number of Hours Requested Towards Broadening Training (10 hours, 20 hours): 20

In the table below give the date of the talk, its title, the speaker's name and tick if you intend to submit an extended abstract for this talk. Ask an attending academic to initial the form to validate your attendance. Talks can include: colloquia, research group seminars, pre-viva talks, reading groups etc. by both internal and external speakers.

Date	Abbreviated Title	Speaker	Extended Abstract?	Academic initials
10/10/21	MoLSS: Mathematically Modelling Crime	David Lloyd and Laura Jones		DJBL
20/10/21	Dynamical Systems and Crime	Laura Jones	✓	AP
26/10/21	MoLSS: Mathematical Biology and Cells	Carina, Kieran and Josepine		DJBL
27/10/21	Deep Learning and Conjugate Maps	Jason Bramburger		DJBL
09/11/21	MoLSS: Sleep Cadence	Anne, Imran and Thalia		DJBL
17/11/21	Mechanobiology: A tense situation	Kieran Boniface		AL
23/11/21	MoLSS: Machine Learning and the Koopman Operator	Stefan Klus		DJBL
26/11/21	Hybrid Models for Multiscale reaction-diffusion dynamics	Stephanie Winkelman		DJBL
07/12/21	MoLSS: Modelling Influenza	Stephen Falconer		DJBL
28/01/22	Pattern search design informed by fitness landscape analysis	Ferrant Neri		DJBL

Date	Abbreviated Title	Speaker	Extended Abstract?	Academic initials
02/02/22	Studying dynamics using computational polynomial optimization	David Goluskin		DJBL
09/02/22	How much analytics can there be in advertising	Emily Weldon		AP
23/02/22	Potential impacts of climate change on peak river flows in GB and modelling crops	Ali Rudd and Matt Brown		AP
02/03/22	The effect of compliance on the stability of jets and wakes	Ryan Poole		AP
25/03/22	Machine Learning and Computation Modelling in Systems Biology	Francesca Buffa		DJBL
06/04/22	Modelling dynamical systems directly from data	Stephen Falconer		AP
04/05/22	Data visualization and analytics	Anna van der Vliet		AP
11/05/22	Mathematics, machine learning and medicine	Philip Aston		AP
17/05/22	PDI's and Measure Preserving Maps, Black Holes and Information Paradox	Elliot Sullinge-Farrall, Vaibhav Gautam		DJBL
05/07/22	Pre-viva talk	Imran Usmani	✓	DJBL

To claim 10/20 hours towards your broadening training requires 1/2 extended abstracts and this form to be submitted with your Confirmation Report.

---

---

**ASSESSMENT**

---

To be completed by the Confirmation Examiners following a discussion of the extended abstract/s submitted by the student with the Confirmation Report.

1. Did the student submit the requisite number of extended abstracts for the requested number of hours? (1 for 10 hours or 2 for 20 hours)

**YES**

**NO**

2. Was the student able to suitably discuss the content of the seminars for which the extended abstracts were submitted?

**YES**

**NO**

3. In your opinion should the student be awarded the requested number of hours toward their broadening training commitment?

**YES**

**NO**

4. If you have further comments then add them here.

**Signatures and Date:**

Examiner 1:

Examiner 2:

Date:

PGR Director:

Date:

**Extended Abstract 1: Laura Jones, 20th October 2022, Dynamical Systems and Crime**

A key question that is asked, that postulates the theory of collective efficacy is, what if residents of a neighbourhood could influence crime rates by their behaviour? Firstly, it should be asked, what is collective efficacy? Collective efficacy is the conviction shared by a group of people that they can work together to successfully complete a specific task. The idea is that the difference in neighbourhoods' inner structure leads to spatial variation in crime rates. For example, if something happens in your neighbourhood, would you trust your neighbours to intervene and help. There already exists many models that show the negative link between collective efficacy and crime, but the literature in studying the formation of patterns is still limited. Within the presentation, a novel convolution model of collective efficacy shall be presented, that shall allow for a mathematical investigation of neighbourhood and resource effects on the formation of collective efficacy and transitions between different regions of collective efficacy. The data that has been used was shared by Professor Brunton-Smith, who co-authored a paper on the relationship between collective efficacy and violence in London, where he constructed a collective efficacy score. Spatial clustering of the data is explored with k-means clustering, where it was found the optimal number of clusters is 3. The different types of kernels used in the model were also investigated, primarily looking at a door kernel and exponential kernel. To continue this research, we would like to expand the continuum model to 2-dimensions and incorporate a crime variable and study as a composition model.

**Extended Abstract 2: Imran Usmani, 5th July 2022, Parametric Models of the Circadian Pacemaker in Controlled Lighting Conditions**

Within the talk, the main results of the thesis shall be discussed. Within the mammalian circadian system there are many biological clocks, including individual cells and organs also function as clocks. But the focus is on the circadian pacemaker that is in the SCN of the brain. It uses light information from the eyes to entrain to the day-night cycle, and the clocks in the system are entrained by signals from the pacemaker. Firstly, we consider a dynamical equation in one-dimension, which is made up of the intrinsic velocity and the velocity response. Here, we consider three controlled lighting conditions: darkness, light pulse and light-dark cycles. We can then derive the response function from the phase response curve and apply it to Khalsa's (type 1) phase response curve. Yet, the solutions are over-fits to the data. So, the design requires a priori assumptions about the form of response curve. If we are to consider 'simple' clock models, then we shall assume a sinusoidal response because if the response function is sinusoidal, then the phase response curve is approximately sinusoidal. Secondly, we shall consider a two-dimensional model, commonly known as Kronauer's Model. It is very difficult to apply the data and we require simplifying assumptions to make it work, such as, we need a strongly attracted limit cycle. There are two future directions following this research. Namely circadian phase prediction in the field to improve more reliable models. Additionally, prediction of sleep timing. Whilst there are already models out there for combined parametric and sleep models, there is scope to improve them.

## Glossary

**delayed implantation** A reproductive strategy whereby the embryo does not immediately implant in the uterus, it remains in a state of dormancy. Also known as embryonic diapause.

**polyandry** A mating system where one female mates with several males in a breeding season. The benefit for badgers is the reduction in the risk of infanticide by males as they would be uncertain of their paternity of any offspring produced.

**superfoetation** A second, new pregnancy occurs during an initial pregnancy. I.e. another egg is fertilized and implanted in the womb days or weeks later than the first one.



## References

- [1] A. W. Bowman, *An alternative method of cross-validation for the smoothing of density estimates*, Biometrika **71** (1984), no. 2, 353–360, [doi](#).
- [2] D. R. Brillinger, H. K. Preisler, A. A. Ager, and J. G. Kie, *The use of potential functions in modelling animal movement*, pp. 385–409, Springer New York, New York, NY, 2012, [doi](#).
- [3] W. H. Burt, *Territoriality and home range concepts as applied to mammals*, Journal of Mammalogy **24** (1943), no. 3, 346–352, [doi](#).
- [4] M. Caley, T. Fowler, S. Welch, and A. Wood, *Risk of developing tuberculosis from a school contact: retrospective cohort study, united kingdom, 2009*, Eurosurveillance **15** (2010), no. 11, 1–4, [doi](#).
- [5] G. P. Cleary, L. A. L. Corner, J. O'Keeffe, and N. M. Marples, *The diet of the badger meles meles in the republic of ireland*, Mammalian Biology **74** (2009), no. 6, 438–447, [doi](#).
- [6] L. Corner, *The role of wild animal populations in the epidemiology of tuberculosis in domestic animals: How to assess the risk*, Veterinary Microbiology **112** (2006), no. 2, 303–312, 4th International Conference on Mycobacterium bovis, [doi](#).
- [7] L. A. L. Corner, L. J. Stuart, D. J. Kelly, and N. M. Marples, *Reproductive biology including evidence for superfetation in the european badger meles meles (carnivora: Mustelidae)*, PLoS ONE **10** (2015), no. 10, 1–18, [doi](#).
- [8] M. Crofoot, R. Kays, and M. Wikelski, *Data from: Shared decision-making drives collective movement in wild baboons*, 2015, [doi](#).
- [9] R. J. Delahay, S. P. Carter, G. J. Forrester, A. Mitchell, and C. L. Cheeseman, *Habitat correlates of group size, bodyweight and reproductive performance in a high-density eurasian badger (meles meles) population*, Journal of Zoology **270** (2006), no. 3, 437–447, [doi](#).
- [10] R. J. Delahay, *Research on badgers and tb at woodchester park*, Animal & Plant Health Agency, Retrieved November 15 2022, [https://www.gla.ac.uk/media/Media\\_538509\\_smxx.pdf](https://www.gla.ac.uk/media/Media_538509_smxx.pdf).
- [11] R. Delahay, N. Walker, M. Gunn, C. Christie, G. Wilson, C. Cheeseman, and R. McDonald, *Using lifetime tooth-wear scores to predict age in wild eurasian badgers: performance of a predictive model*, Journal of Zoology **284** (2011), no. 3, 183–191, [doi](#).
- [12] E. Do Linh San, N. Ferrari, and J.-M. Weber, *Spatio-temporal ecology and density of badgers meles meles in the swiss jura mountains*, European Journal of Wildlife Research **53** (2007), no. 4, 265–275, [doi](#).

- [13] J. Downs, M. Horner, D. Lamb, R. W. Loraamm, J. Anderson, and B. Wood, *Testing time-geographic density estimation for home range analysis using an agent-based model of animal movement*, International Journal of Geographical Information Science **32** (2018), no. 7, 1505–1522, [doi](#).
- [14] J. A. Downs and M. W. Horner, *A characteristic-hull based method for home range estimation*, Trans. GIS **13** (2009), 527–537, [doi](#).
- [15] J. A. Downs, M. W. Horner, and A. D. Tucker, *Time-geographic density estimation for home range analysis*, Annals of GIS **17** (2011), no. 3, 163–171, [doi](#).
- [16] R. P. W. Duin, *On the choice of smoothing parameters of parzen estimators of probability density functions*, IEEE Transactions on Computers **C-25** (1976), 1175–1179, [doi](#).
- [17] A. Gaughran, E. Mullen, T. MacWhite, P. Maher, D. J. Kelly, R. Kelly, M. Good, and N. M. Marples, *Badger territoriality maintained despite disturbance of major road construction.*, PLoS One **16** (2021), no. 9, e0242586 (eng), [doi](#).
- [18] J. Habbema, J. Hermans, and K. Van den Broek, *A stepwise discrimination analysis program using density estimation*, Compstat 1974: Proceedings in Computational Statistics, Physica Verlag, Vienna, 1974, pp. 101–110.
- [19] R. J. Hall, S. Altizer, S. J. Peacock, and A. K. Shaw, *Animal migration and infection dynamics: Recent advances and future frontiers*, Animal Behavior and Parasitism (2022), 111.
- [20] C. Ham, C. A. Donnelly, K. L. Astley, S. Y. B. Jackson, and R. Woodroffe, *Effect of culling on individual badger *meles meles* behaviour: Potential implications for bovine tuberculosis transmission*, Journal of Applied Ecology **56** (2019), no. 11, 2390–2399, [doi](#).
- [21] R. Joo, S. Picardi, M. E. Boone, T. A. Clay, S. C. Patrick, V. S. Romero-Romero, and M. Basille, *Recent trends in movement ecology of animals and human mobility*, Movement Ecology **10** (2022), no. 1, 26, [doi](#).
- [22] S. Klus, P. Koltai, and C. Schütte, *On the numerical approximation of the Perron-Frobenius and Koopman operator*, Journal of Computational Dynamics **3** (2016), no. 1, 51–79, [doi](#).
- [23] S. D. Landsheer, *kmeans\_opt*, MATLAB Central File Exchange., 2022, Retrieved August 12, 2022., [https://www.mathworks.com/matlabcentral/fileexchange/65823-kmeans\\_opt](https://www.mathworks.com/matlabcentral/fileexchange/65823-kmeans_opt).
- [24] R. Loraamm, K. Goodenough, C. Burch, L. Davenport, and T. Haugaasen, *A time-geographic approach to identifying daily habitat use patterns for amazonian black skimmers*, Applied Geography **118** (2020), 102189, [doi](#).

- [25] F. Lu, M. Zhong, S. Tang, and M. Maggioni, *Nonparametric inference of interaction laws in systems of agents from trajectory data*, Proceedings of the National Academy of Sciences **116** (2019), no. 29, 14424–14433, [doi](#).
- [26] D. Macdonald, C. Newman, C. Buesching, and P. Nouvellet, *Are badgers ‘under the weather’? direct and indirect impacts of climate variation on european badger (*meles meles*) population dynamics*, Global Change Biology **16** (2010), 2913 – 2922, [doi](#).
- [27] E. A. Magowan, I. E. Maguire, S. Smith, S. Redpath, N. J. Marks, R. P. Wilson, F. Menzies, M. O'Hagan, and D. M. Scantlebury, *Dead-reckoning elucidates fine-scale habitat use by European badgers *Meles meles**, Animal Biotelemetry **10** (2022), no. 1, 10, [doi](#).
- [28] J. L. McDonald, A. Robertson, and M. J. Silk, *Wildlife disease ecology from the individual to the population: Insights from a long-term study of a naturally infected european badger population*, Journal of Animal Ecology **87** (2018), no. 1, 101–112, [doi](#).
- [29] H. J. Miller, *A measurement theory for time geography*, Geographical analysis **37** (2005), no. 1, 17–45, [doi](#).
- [30] J. E. Milner, P. G. Blackwell, and M. Niu, *Modelling and inference for the movement of interacting animals*, Methods in Ecology and Evolution **12** (2021), no. 1, 54–69, [doi](#).
- [31] C. O. Mohr, *Table of equivalent populations of north american small mammals*, The American Midland Naturalist **37** (1947), no. 1, 223–249.
- [32] E. Neal and C. Cheeseman, *Badgers*, Natural History Series, T. & A.D. Poyser, 1996.
- [33] D. NíBhuachalla, L. A. Corner, S. J. More, and E. Gormley, *The role of badgers in the epidemiology of mycobacterium bovis infection (tuberculosis) in cattle in the united kingdom and the republic of ireland: current perspectives on control strategies.*, Vet Med (Auckl) **6** (2015), 27–38 (eng), [doi](#).
- [34] J.-H. Niemann, S. Klus, and C. Schütte, *Data-driven model reduction of agent-based systems using the Koopman generator*, PLOS ONE **16** (2021), no. 5, 1–23, [doi](#).
- [35] C. Ordóñez Galán, J. Rodríguez-Pérez, J. Martínez Torres, and P. García Nieto, *Analysis of the influence of forest environments on the accuracy of GPS measurements by using genetic algorithms*, Mathematical and Computer Modelling **54** (2011), no. 7, 1829–1834, Mathematical models of addictive behaviour, medicine & engineering, [doi](#).
- [36] E. Pebesma and R. S. Bivand, *S classes and methods for spatial data: the sp package*, R news **5** (2005), no. 2, 9–13.
- [37] H. Preisler, A. Ager, B. Johnson, and J. Kie, *Modeling animal movements using stochastic differential equations*, Environmetrics (2004), 643–657, [doi](#).

- [38] H. K. Preisler, A. A. Ager, and M. J. Wisdom, *Analyzing animal movement patterns using potential functions*, *Ecosphere* **4** (2013), no. 3, 1–13, [doi](#).
- [39] L. M. Rogers, R. Delahay, C. L. Cheeseman, S. Langton, G. C. Smith, and R. S. Clifton-Hadley, *Movement of badgers (*Meles meles*) in a high-density population: individual, population and disease effects.*, *Proc Biol Sci* **265** (1998), no. 1403, 1269–1276 (eng), [doi](#).
- [40] M. Rudemo, *Empirical choice of histograms and kernel density estimators*, *Scandinavian Journal of Statistics* (1982), 65–78, <https://www.jstor.org/stable/4615859>.
- [41] R. S. Schick, S. R. Loarie, F. Colchero, B. D. Best, A. Boustan, D. A. Conde, P. N. Halpin, L. N. Joppa, C. M. McClellan, and J. S. Clark, *Understanding movement data and movement processes: current and emerging directions*, *Ecology Letters* **11** (2008), no. 12, 1338–1350, [doi](#).
- [42] I. Silva, M. Crane, B. M. Marshall, and C. T. Strine, *Reptiles on the wrong track? moving beyond traditional estimators with dynamic Brownian bridge movement models*, *Movement ecology* **8** (2020), no. 1, 1–13, [doi](#).
- [43] B. W. Silverman, *Density estimation for statistics and data analysis*, CRC Press, 1986.
- [44] D. Siniff and C. Jessen, *A simulation model of animal movement patterns*, *Advances in Ecological Research* **6** (1969), 185–219, [doi](#).
- [45] G. C. Smith and R. Budgey, *Simulating the next steps in badger control for bovine tuberculosis in england*, *PLOS ONE* **16** (2021), no. 3, 1–8, [doi](#).
- [46] K. Stahl, R. Moore, J. Floyer, M. Asplin, and I. McKendry, *Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density*, *Agricultural and Forest Meteorology* **139** (2006), no. 3, 224–236, [doi](#).
- [47] The Wildlife Trusts, *European badger*, <https://www.wildlifetrusts.org/wildlife-explorer/mammals/european-badger> [cited 5 August 2022].
- [48] W. D. Walter, J. Fischer, S. Baruch-Mordo, and K. Vercauteren, *What is the proper method to delineate home range of an animal using today's advanced gps telemetry systems: The initial step*, pp. 249 – 68, IntechOpen, 2011, [doi](#).
- [49] D. Wilkinson, R. Bennett, I. McFarlane, S. Rushton, M. Shirley, and G. Smith, *Cost-benefit analysis model of badger (*Meles meles*) culling to reduce cattle herd tuberculosis breakdowns in britain, with particular reference to badger perturbation*, *Journal of Wildlife Diseases* **45** (2009), no. 4, 1062–1088, [doi](#).

- [50] R. Woodroffe, C. A. Donnelly, C. Ham, S. Y. B. Jackson, K. Moyes, K. Chapman, N. G. Stratton, and S. J. Cartwright, *Badgers prefer cattle pasture but avoid cattle: implications for bovine tuberculosis control.*, Ecol Lett **19** (2016), no. 10, 1201–1208, [doi](#).
- [51] N. Yamaguchi, H. L. Dugdale, and D. W. Macdonald, *Female receptivity, embryonic diapause, and superfetation in the European badger (*Meles meles*): implications for the reproductive tactics of males and females*, The Quarterly Review of Biology **81** (2006), no. 1, 33–48, [doi](#).

