Data-driven Analysis of Fine-Scale Badger Movement in the UK

Jessica Furber[1], Richard J. Delahay[2], Ruth Cox[2], Rosie Woodroffe[3], Maria O'Hagan[4], Stefan Klus[5], Giovanni Lo Iacono[6,7,8], Mark A. Chambers[6,9], and David J. B. Lloyd[1]

[1]School of Mathematics and Physics, University of Surrey, Guildford, UK
[2]National Wildlife Management Centre, Animal and Plant Health Agency, Sand Hutton, York, YO41 1LZ, UK
[3]Zoological Society of London, Institute of Zoology, London, UK
[4]Department of Agriculture, Environment and Rural Affairs (DAERA), Veterinary Epidemiology Unit, Clare House, Airport Road West, Belfast, BT3 9ED, Northern Ireland
[5]Department of Mathematics, Heriot–Watt University, Edinburgh, UK
[6]School of Veterinary Medicine, University of Surrey, Guildford, UK
[7]The Surrey Institute for People-Centred Artificial Intelligence, Guildford, United Kingdom
[8]Institute for Sustainability, University of Surrey, Guildford, United Kingdom
[9]School of Biosciences, University of Surrey, Guildford, UK

Abstract

Understanding animal movements at different spatial scales presents a significant challenge as their patterns can vary widely from daily foraging behaviours to broader migration or territorial movements. This challenge is of general interest because it impacts the ability to manage wildlife populations effectively. In this study, we conduct diffusion analysis based on European badger (*Meles meles*) movement data obtained from three different regions in the UK (Gloucestershire, Cornwall, and Northern Ireland) and fit a generalised linear mixed-effects model to examine the relationship between variables. We also feature a novel application of *extended dynamic mode decomposition* (EDMD) to uncover patterns relating to badger social organisation. By applying our approach to these different populations, we were able to assess its performance across a range of badger densities. A key result was that in some areas, EDMD clusters matched observed group home ranges, whilst in others, discrepancies likely arose because of population management interventions, such as badger culling. The methods presented offer a promising approach for studying territoriality and the impacts of management strategies on animal movement behaviour.

Keywords: *Meles meles*, EDMD, Koopman operator, GPS tracking data.

1 Introduction

A fundamental challenge in the study of ecological processes such as animal migration, swarming, and disease spread [1–3] is the identification of patterns of movement. Short-term patterns (also referred to as fast temporal scales), such as those related to daily activities like foraging, involve frequent shifts in direction and variable speeds. These short-term fluctuations reflect the dynamic nature of an animal's movement as it navigates its environment, while long-term patterns (slow temporal scales), such as migration or territory establishment, involve more gradual changes. One approach for the modelling of animal movement patterns uses stochastic differential equations to simulate the position of an animal in two dimensions, for example on an energy potential – a spatial map of energy or resource gradients that influences animal movement by attracting or repelling them from certain areas [4,5] – or through an individual-based stochastic simulation model [6]. The standard stochastic differential equation model naturally separates the fast and slow temporal scales by means of a *diffusion* term (fast dynamics) and a *drift* term (slow dynamics). A higher diffusion value indicates more extensive or erratic movements, such as those seen in exploratory behaviours or

during foraging activities. This term helps to model how these movements vary with factors such as sex (e.g., males ranging widely during mating seasons) or season (e.g., increased movement during spring as animals search for food or suitable breeding sites). The drift term provides insight into the deterministic trends in movements, revealing underlying directional biases and tendencies influenced by environmental factors or intrinsic behaviours. For example, it may capture consistent movements along specific routes driven by resource availability or territorial behaviour. Additionally, spatial regions can be identified where animals remain for extended times, which are mathematically referred to as metastable states [7, 8]. Such areas may be of significant ecological or behavioural interest, hence, identifying them is of value for investigating spatial structure in wild animal populations. In this study, a mathematical data-driven approach is demonstrated for the analysis of slow and fast temporal movement scales in a wild mammal.

The movement patterns and social organisation of European badgers (*Meles meles*) have been extensively studied [9–13]. The social behaviour of badgers is of particular interest to mammal ecologists as they are unusual among the mustelids in sometimes forming relatively large mixed sex social groups and occupying mutually defended territories, although elsewhere in their geographical range they may be solitary [9, 14]. Badgers are also of research interest because of their role in the epidemiology of *Mycobacterium bovis* infection (which causes bovine tuberculosis; bTB) in cattle [15–20]. Although cattle are the principal hosts of bTB, many wild mammals are also susceptible [21] and the bacteria can be transmitted in both directions between badgers and cattle [22]. Whilst the principal routes of transmission between badgers and cattle remain unclear, there is growing recognition that badger foraging and defecation behaviour contribute to the contamination of shared environments with infectious bacteria [18, 19, 23]. Consequently, in the UK and Republic of Ireland, badgers have been the subject of management interventions including culling and vaccination [24,25]. The social structure of badger populations can remain stable when population density is high and there is no interference, but several studies have described how disruptions arising from culling may influence badger social behaviour and disease dynamics [26–29]. Recent studies on badger behaviour using GPS collars have investigated movement relative to cattle [19] and road construction [20]. Combining GPS with tri-axial accelerometers and magnetometers allows for detailed movement analysis through so-called 'dead-reckoning', though this method is typically employed over relatively short time periods (days to weeks) due to the high-frequency data logging [30]. In the present study, we aim to use GPS data from field studies of several badger populations to measure short-term behaviour and capture long-term movement patterns using a novel analytical approach.

Here, we carry out for the first time a mathematical analysis of badger GPS data using a framework of fast (diffusion) and slow (drift) spatiotemporal scales to describe movement patterns. We also provide new data-driven methods to determine the effective metastable group home ranges of badgers over time. We apply these approaches to GPS data derived from badger populations from three different regions of the British Isles: Gloucestershire, Cornwall, and County Down. Our sample of study populations represents a range of local badger densities and includes some locations where substantial temporal change in movement behaviour seems likely because of culling. The variations in conditions across our study populations provide an excellent opportunity to test the broad utility of our novel analytical approach.

The paper is structured as follows: Section 2 explores the collection of data, followed by a description of the methods used, including data cleaning and preparation, estimation of the diffusion, and *extended dynamic mode decomposition* (EDMD). In Section 3, we outline the results of the study, beginning with the estimation of the diffusion, followed by the results of the EDMD analysis. Finally, we discuss the results in Section 4 and then conclude in Section 5.

2 Methodology

At time t, the position of a badger is given by $\mathbf{X}_t = (x_t, y_t)^\top \in \mathbb{R}^2$ and can be modelled by a stochastic differential equation defined as

$$\mathrm{d}\mathbf{X}_t = b(\mathbf{X}_t)\,\mathrm{d}t + \sigma(\mathbf{X}_t)\,\mathrm{d}\mathbf{W}_t, \tag{1}$$

where $b\colon \mathbb{R}^2 \to \mathbb{R}^2$ represents the *drift term*, $\sigma\colon \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$ is a real matrix that represents the *diffusion term*, and $\mathbf{W}_t$ is a two-dimensional Wiener process. The Wiener process (alternatively called Brownian motion) is widely used, typically to describe a random, but continuous, motion of atoms or particles, but here has been extended to model the movements of animals. We assume that badgers move isotropically in space, i.e. the diffusion is the same in all directions. This implies that we can write the diffusion matrix as a constant multiplied by the identity matrix, i.e., for some value $c \in \mathbb{R}$ we can represent the diffusion matrix as $\sigma(\mathbf{x}) = cI$ where $I \in \mathbb{R}^{2\times 2}$ is the identity matrix.

In this section we first introduce the data collection, then describe the mathematical methods used, starting with estimation of the diffusion value c from the data to gain further understanding of the dynamics of the everyday movements of badgers. This is followed by an introduction to the Koopman operator and extended dynamic mode decomposition, which make use of the trajectory data to estimate the number of metastable states present in the data. Then we explore the data cleaning and preparation for each method.

2.1 Data Collection

We accessed three sets of data from studies where collars carrying GPS devices had been fitted to badgers. One dataset derived from a study of badgers resident in Woodchester Park, Gloucestershire, England [13, 23], another related to four different sites in Cornwall, England [19], and the third was from a population in County Down, Northern Ireland [20].

The Woodchester Park Study was established in 1976 to investigate bTB in a wild, naturally infected badger population [31]. The study includes analysis of capture-mark-recapture data, which involves trapping and re-trapping individually marked (tattooed) badgers to infer group membership [13], and the analysis of GPS data (Cox & Delahay, unpublished data). The study site itself covers *c.* $7\,\mathrm{km}^2$ of predominantly permanent pasture and woodland [12]. The GPS units used recorded position fixes every 35 minutes from 20:00 to 23:35, and then every 35 minutes between 00:00 and 03:30.

The four sites in Cornwall each included at least two beef farms and two dairy farms [19]. However, they varied in landscape types: sites C2 and C4 in North Cornwall are dominated by pastoral landscapes with steep wooded valleys, site F1 on the North coast of West Cornwall is bounded by granite cliffs and moorland, and site F2 on the South coast of West Cornwall is also dominated by pastoral landscapes with wooded valleys. The GPS units deployed on badgers recorded position fixes every 20 minutes between 18:00 and 06:00 while the badger was active.

The Northern Ireland study area in County Down, near the town of Banbridge, covers *c.* $100\,\mathrm{km}^2$ of primarily pasture and arable land [32]. The resident badger population was the subject of a Test and Vaccinate or Remove (TVR) project whereby captured animals were tested for bTB using the Dual-Path Platform VetTB test (DPP) and test-positives were euthanised, whilst test-negative badgers were vaccinated against bTB with Bacille Calmette Guerin, fitted with GPS collars and released [32–34]. The GPS units recorded position fixes once per hour between 21:00 and 04:00.

The various study areas contained different numbers of badger social groups and individuals fitted with GPS devices over a range of time periods (Table 1). The number of social groups was estimated using field knowledge specific to each site: in Woodchester, bait-marking data were used [35]; in Northern Ireland and Cornwall, capture-recapture and GPS collaring data were applied (Table 1 shows the total number of social groups monitored throughout the total GPS-collar monitoring period). GPS data in the form of locational fixes were not captured at a uniform rate within or among years of study (Figure 1), although most fixes were collected in the latter half of the year. Most collars were deployed in the summer/autumn (the peak season for badger trapping [36]), which means that the battery life waned during the first half of the following year, which overlaps with the closed season when trapping is suspended to prevent the interference with reproductive females and their offspring [37]. Battery longevity is dependent on the frequency with which fixes are recorded.

To account for the impact of culling and other interventions across the different study sites, it is important to consider their varying management histories. At Woodchester, badger culling began on the farmland outside the park in 2018 (Woodchester Park belongs to the National Trust which has a no-cull policy), meaning one might expect change to the social structure in all years from 2018

Table 1: Summary of site and badger monitoring across all study sites and years used in this study.

Site	Woodchester	C2	C4	F1	F2	Northern Ireland
Approximate Size	$c.7\,\mathrm{km}^2$	$c.8\,\mathrm{km}^2$	$c.\ 6\,\mathrm{km}^2$	$c.8\,\mathrm{km}^2$	$c.6\,\mathrm{km}^2$	$c.100\,\mathrm{km}^2$
Years Monitored	2018–2022	2013–2015	2014–2017	2013–2017	2013–2015	2014–2017
Social Groups	18	6	5	7	10	45
Badgers GPS-Collared						
Total	66	12	13	22	20	105
Female	34	3	9	13	12	41
Male	32	9	4	9	8	64

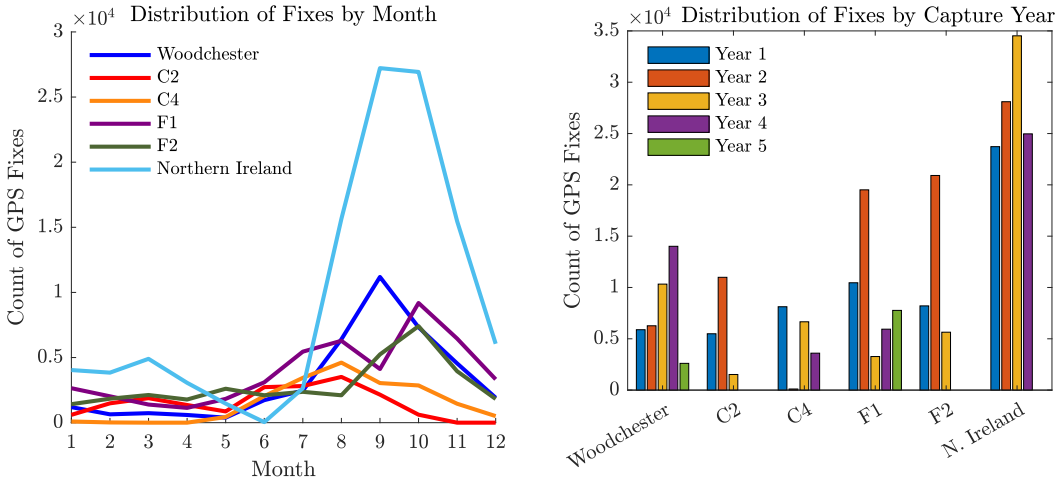


Figure 1: Number of fixes by month (left) and by capture year for respective sites (right).

onwards due to population reductions. In Cornwall, there was no active culling during the study period for sites F1, F2 and C2. Although site F1 was in a former treatment area of the Randomised Badger Culling Trial (RBCT) [38], potentially influencing badger dynamics through residual effects from past culling. The Cornwall site C4, however, was culled from 2016 onwards [18]. In Northern Ireland, the selective removal of bTB-positive badgers after the first year of the study [32] might have disrupted social cohesion, prompting increased ranging behaviour or social group reconfigurations in response to changing population density. These variations in culling and bTB management strategies must be considered when interpreting EDMD results and comparing them with field observations of badger social groups, as they likely played a role in the movement patterns observed.

2.2 Estimating Diffusion of Badgers

The diffusion term in (1) can be estimated using the Kramers–Moyal formula, given by

$$\sigma(\mathbf{x})\sigma(\mathbf{x})^\top := \lim_{t\to 0} \mathbb{E}\left[\frac{1}{t}(\mathbf{X}_t - \mathbf{x})(\mathbf{X}_t - \mathbf{x})^\top \middle| \mathbf{X}_0 = \mathbf{x}\right]. \tag{2}$$

Because the data are collected as a time-series with discrete intervals, it is not possible to consider the limit as $t \to 0$ in Equation (2) (i.e., infinitely small time steps). However, an estimate can still be obtained using finite difference approximations. Since we assume that badgers move isotropically in space, that is the diffusion is the same in all directions, we write the diffusion matrix as $\sigma(x) = cI$, where c is the diffusion value that we wish to calculate, and I is the identity matrix. We estimate c as follows: Given m measurements $\mathbf{x}_i = (x_i, y_i)^\top$ at time t_i, with $i = 1, \ldots, m$, let $\Delta t_i = t_{i+1} - t_i$

denote the time difference between two successive GPS measurements, then

$$c^2 \approx \frac{1}{2(m-1)} \sum_{i=1}^{m-1} \frac{1}{\Delta t_i} \left[(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 \right]. \tag{3}$$

Numerical investigations show that good approximations to c and thus σ can be found despite ignoring the limit and when Δt_i is not constant. Since c is always a positive number (as we take the square root), it is found that c is log-normally distributed, which is confirmed when running a Shapiro–Wilk test for normality [39].

We fitted a generalised linear mixed-effects model (GLMM) to examine the relationship between the diffusion value c and various predictors including site, capture year (i.e. GPS-collar monitoring year), month, and sex. The model also included a random intercept for individual animals to account for repeated measurements on the same individuals. The model used a normal distribution with a log link function and was fit using penalised maximum likelihood estimation (PL). To determine the significance of the predictor variables, likelihood ratio tests were also conducted comparing the full model with all predictors present with the reduced models, where each predictor was excluded in turn. Statistical tests were implemented in Matlab [40].

In the model, the reference levels were set to Northern Ireland for site, year 1 for capture year, January for month, and male for sex. Although selecting a specific reference site was not preferred, the model requires a reference level to serve as a baseline for comparisons. Northern Ireland was chosen due to being the lowest density population out of the six sites, allowing coefficients for other sites to be interpreted as deviations from this baseline. This setup facilitates comparison across sites and years, but the choice of Northern Ireland as the reference is not meant to imply its unique importance beyond providing a necessary baseline.

2.3 Koopman Operator and Extended Dynamic Mode Decomposition

We utilise the Koopman operator framework to analyse the dynamic behaviour of badger populations. We refer to the badger data as a dynamical system, which changes state over time according to a fixed rule, often described by differential equations such as Equation (1). The eigenvalues and eigenfunctions of the Koopman operator provide valuable insights into the behaviour and stability of these systems. We are interested in applying the Koopman operator to badger movement data to investigate metastability, which refers to the tendency of the population to remain in a region for extended periods before transitioning to an alternative region.

We consider discrete trajectories from (1) with a state space $\mathbb{R}^2$ and an evolution operator (also referred to as the flow map) $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ associated with the dynamical system. That is, for some initial condition the evolution operator maps it forward in time, i.e.,

$$\mathbf{x}(0) = \mathbf{x}_0 \;\Rightarrow\; \mathbf{F}_\tau(\mathbf{x}_0) = \mathbf{x}(\tau),$$

where τ is a fixed lag-time. This fixed lag-time can be considered as the time step between GPS observations if the time steps were the same and consistent. Whilst the evolution operator acts on positions in the state space (i.e. $\mathbf{x} \in \mathbb{R}^2$), the Koopman operator $\mathcal{K}$ acts on functions of the state space. We define the Koopman operator $\mathcal{K}_\tau$ for data from a stochastic process as

$$[\mathcal{K}_\tau \psi](\mathbf{x}) = \mathbb{E}[\psi(\mathbf{F}_\tau(\mathbf{x}))], \tag{4}$$

where $\mathbf{x}$ is an arbitrary point, $\mathbb{E}[\cdot]$ denotes the expected value, and $\psi \in L_\infty(\mathbb{R}^2)$ is a measurable quantity of the system such as the speed of a badger. Under certain conditions, the operator can be extended from L_∞ to L_2 (i.e. so that it applies to a broader set of functions), allowing inner products to be used in Galerkin-like methods (see [41] for more details). Since the operator acts on functions, $\mathcal{K}_\tau$ is infinite dimensional and linear even if $\mathbf{F}$ is finite dimensional and nonlinear.

Koopman analysis tries to find key measurement functions. The eigenfunctions $\varphi(\mathbf{x})$ and the corresponding eigenvalues λ of the Koopman operator, defined by

$$[\mathcal{K}_\tau \varphi](\mathbf{x}) = \lambda \varphi(\mathbf{x}),$$

provide detailed insights into the underlying patterns and temporal evolution of the dynamical system. The eigenvalues $\lambda \approx 1$ and the associated eigenfunctions are of particular interest in the badger movement system, as they provide crucial insights into the slow dynamics and the metastable sets. Small eigenvalues govern faster time scales, such as the daily dynamics of the badgers. The number of dominant eigenvalues corresponds to the number of metastable sets within the system. In some cases, it can be challenging to determine the number of metastable states due to no clear spectral gap, which is the barrier between the faster and slower dynamical modes. That is, given N eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_N$, a spectral gap is defined as the first λ_i such that

$$|\lambda_i - \lambda_{i+1}| > \delta > 0,\ i = 1, 2, \ldots, N-1 \tag{5}$$

for some value of $\delta > 0$. In Appendix A.1 we plot the eigenvalues and their successive differences against i and also a different method to find the spectral gap to compare.

Extended dynamic mode decomposition (EDMD) is a data-driven method to estimate the Koopman operator and its eigenvalues, eigenfunctions, and modes. The method requires a data set of snapshot pairs, i.e. $[\mathbf{x}_k, \mathbf{y}_k]_{k=1}^K$, where each pair would consist of observations of the dynamical system such that $\mathbf{y}_k = \mathbf{F}(\mathbf{x}_k)$. In other words, given time series data $\mathbf{x}_k$, then $\mathbf{y}_k = \mathbf{F}(\mathbf{x}_k)$ is the next observation in the time series. Each observation is a GPS-derived location of a badger, numbers of which vary among sites (n), from which a finite number of data points are collected m_i (for $i = 1, \ldots, n$), and the value of m varies from animal to animal. The coordinate data for the ith badger is represented as a vector, $\mathbf{X}_j^i \in \mathbb{R}^2$ at time t_j, where $j = 1, \ldots, m_i$. Ideally, the time intervals between observations in the time series are equal, as EDMD assumes evenly spaced data for accurate analysis. When the time intervals are irregular, interpolation is necessary to construct a uniformly spaced time series before applying the method. We organised the GPS data in the matrices $\mathbf{Y}, \mathbf{Y}' \in \mathbb{R}^{2\times\sum_{k=1}^n (m_k-1)}$,

$$\mathbf{Y} = [\mathbf{X}_1^1, \ldots, \mathbf{X}_{m_1-1}^1, \mathbf{X}_1^2, \ldots, \mathbf{X}_{m_2-1}^2, \ldots, \mathbf{X}_1^n, \ldots, \mathbf{X}_{m_n-1}^n], \tag{6}$$

$$\mathbf{Y}' = [\mathbf{X}_2^1, \ldots, \mathbf{X}_{m_1}^1, \mathbf{X}_2^2, \ldots, \mathbf{X}_{m_2}^2, \ldots, \mathbf{X}_2^n, \ldots, \mathbf{X}_{m_n}^n]. \tag{7}$$

In simpler terms, the $\mathbf{Y}$ matrix is made up of the first pair of coordinates to the penultimate coordinates for each badger, and the $\mathbf{Y}'$ matrix starts from the second entry of coordinates and includes all entries up to the last.

The method also requires a dictionary of observables ψ, $\mathcal{D} = \{\psi_1, \ldots, \psi_{N_K}\}$, where N_K (the number of basis functions) is always finite-dimensional in practice because working with infinite-dimensional dictionaries is not feasible [41]. The function space itself, however, remains infinite-dimensional. Consequently, the operator learned from the data is an approximation, specifically a Galerkin projection. Here, indicator functions are used as the set of *basis functions*—a mathematical term referring to foundational building blocks for approximating complex functions—in EDMD, which is equivalent to using Ulam's method to estimate a Markov state model [42]. Other basis functions such as monomials up to a certain order could be used. However, the resulting matrices might then become numerically ill-conditioned. The optimal choice of dictionary elements depends strongly on the dynamical system and remains an open question. When defining the domain for the EDMD indicator functions, we considered a box discretisation (i.e. we superimposed a grid over the area of coordinates). For consistency, the number of boxes was chosen by splitting the domain into approximately 100 m by 100 m squares.

Let $\mathbf{\Psi} : \mathbb{R}^2 \to \mathbb{C}^{N_K}$ be the vector-valued function $\mathbf{\Psi}(x) = [\psi_1(x), \psi_2(x), \ldots, \psi_{N_K}(x)]^\top$, for an arbitrary value of x. Then, we define the transformed data matrices $\mathbf{\Psi}_Y, \mathbf{\Psi}_{Y'} \in \mathbb{C}^{N_k\times\sum_{k=1}^n (m_k-1)}$ as

$$\mathbf{\Psi}_\mathbf{Y} = [\psi(\mathbf{X}_1^1), \ldots, \psi(\mathbf{X}_{m_2-1}^2), \ldots, \psi(\mathbf{X}_1^n), \ldots, \psi(\mathbf{X}_{m_n-1}^n)], \tag{8}$$

$$\mathbf{\Psi}_{\mathbf{Y}'} = [\psi(\mathbf{X}_2^1), \ldots, \psi(\mathbf{X}_{m_1}^1), \ldots, \psi(\mathbf{X}_2^n), \ldots, \psi(\mathbf{X}_{m_n}^n)]. \tag{9}$$

The minimisation problem for the transformed data matrices becomes,

$$K^\top = \arg\min_{\overline{K}} ||\Psi_{\mathbf{Y}'} - \overline{K}^\top \Psi_\mathbf{Y}|| = \Psi_{\mathbf{Y}'}\Psi_\mathbf{Y}^+, \tag{10}$$

where $^+$ denotes the Moore–Penrose pseudoinverse of a matrix. The matrix $K \in \mathbb{C}^{N_K \times N_K}$ is a representation of the Koopman operator $\mathcal{K}_\tau$ projected onto the subspace spanned by the chosen basis functions, i.e. a Galerkin approximation.

Finally, let ξ_j be the jth eigenvector of K with the eigenvalue λ_j, then the EDMD approximation of an eigenfunction of $\mathcal{K}_\tau$ is

$$\varphi_j(\mathbf{x}) = \xi_j^\top \mathbf{\Psi}(\mathbf{x}). \tag{11}$$

Once the Koopman eigenvalues and eigenfunctions have been computed, the number of dominant eigenvalues is chosen based on the spectral gap defined in (5) and the corresponding eigenfunctions are clustered using k-means clustering [43]. For example, if there are eight dominant eigenvalues, the corresponding first eight eigenfunctions form clusters that define metastable states, representing the spatial areas that badgers rarely leave.

2.4 Data Cleaning and Preparation

To avoid location errors, and to be consistent with the pre-cleaned data obtained from the Cornwall sites [19], we excluded all GPS-collar locations associated with fewer than four satellites, or with *dilution of precision* greater than four (as detailed in Supporting Information [24]) from the Woodchester and Northern Ireland data sets (where the parameters allowed). Moreover, improbable GPS locations (e.g., those far outside the study areas, or in water bodies), were excluded from all data sets. We chose to remove data derived from one Woodchester Park badger and five individuals from the Northern Ireland study because they were clearly isolated from the remaining data. In line with previous studies where badgers are recorded to have been moving at speeds of up to 26.2 m/min [11], we excluded any data points where badgers were found to be moving faster. Additionally, any single data points where the recorded movement within a time interval appeared biologically unrealistic for badgers were removed, such as travelling a distance of greater than 1500 m in two consecutive time steps (e.g. 70 minutes). As a result of these actions, the Woodchester dataset was reduced by 19.0%, the Northern Ireland data by 13.3%, and the data from the Cornwall sites was reduced further by 1.1%.

The coordinates of the cleaned data were converted from the longitude/latitude (Woodchester and Northern Ireland) or British National Grid (Cornwall) into UTM (Universal Transverse Mercator) coordinates, so the distances between coordinates could be measured in meters. To preserve confidentiality of the location of sites (and to be consistent), the coordinates were shifted to a (0,0) base. For the estimation of the diffusion metric for badgers, the data can be used as presented. However, to account for the large distances between the last location of one night-time monitoring period and the first location of the next, during which time badgers were likely inactive in their underground setts, these intervals were excluded from the calculation. Furthermore, since the limit in Equation (2) cannot be taken, once the individual diffusion metric had been estimated, it was decided to exclude any individual's measurements derived from ten or fewer data points, as an insufficient number of points hinders the accurate estimation of the diffusion metric.

Continuous and evenly spaced time series data are essential for EDMD, as the method assumes uniform time intervals between observations. However, GPS data for badgers often features irregular intervals, necessitating interpolation to create consistent time steps. These inconsistencies can arise due to undetected signals (e.g., when badgers are underground or beneath dense vegetation), timing issues with GPS devices (e.g., a 70-second window to connect with satellites), or large gaps caused by GPS collars losing power, which can result in days or months without data until the collar is replaced. To manage these gaps, new ID numbers were assigned to separate series of GPS fixes, ensuring continuity in the data. For instance, at Woodchester, GPS data were available from 66 unique badgers, but this was increased to 108 collar monitoring periods after accounting for interruptions. Recorded times (e.g., 20:01) were adjusted to the nearest pre-programmed time point (e.g., 20:00) to standardise the timing of locations. Notably, the time discrepancy reflects the duration between the start of GPS location calculation and the time of recording, which can vary across sites. At Woodchester Park, likely influenced by its high proportion of woodland, 13.6% of badgers did not have their positions recorded at the designated time, with most locations logged within one minute of the intended capture time, but some taking up to three minutes. In contrast, only 5% of the

badgers at the Cornwall sites failed to register their location within one minute. Most badgers in the Northern Ireland study area registered a location fix within sixty seconds of the appointed hourly time. Those that did not (see [44]) were excluded (2.5%) to facilitate interpolation. Although badgers are known to travel along complex paths (see [30]), linear interpolation was used to create evenly spaced intervals, closely matching the original time steps. This approach minimises the impact of interpolation on the data matrices used in EDMD.

For example, when locations were recorded every 35 minutes, such as in Woodchester, and a factor of 35 minutes is not chosen, we observe cutting corners and missing the actual recorded data point, resulting in calculating a shorter distance travelled by the badger. This underestimation is expected in any sampling of locations from continuous time, as linear interpolation inherently misses finer-scale movements and underestimates the true distance travelled. We found choosing a two-minute interpolation yielded differences in the distances travelled that were too small for EDMD to detect the badger dynamics, whilst also assuming too much about the badger's movement by generating multiple interpolated points between observed locations. Using a 35-minute interval caused EDMD to fail in identifying non-trivial dynamics because the large time gap between points missed finer-scale transitions, oversimplified the system's behaviour, and reduced the temporal resolution needed to capture complex, non-linear dynamics. For the Woodchester data, we found the optimal interval to be 17.5 minutes because it only assumes one location of the badger between observations, and EDMD was able to identify meaningful dynamics. For the Cornwall datasets, which comprised GPS fixes at 20-minute intervals, this was also found to be optimal. In contrast, the Northern Ireland dataset contained fixes recorded at hourly intervals, necessitating interpolation to create data points every 20 minutes for consistency with the Cornwall datasets. It is important to note that this approach means approximately two-thirds of the Northern Ireland dataset consists of interpolated values, which should be considered when interpreting the results.

3 Results

3.1 Badger Diffusion

The results of the GLMM analysis examining the diffusion of badgers are presented in Table 2. The model included several predictors, including the site, the capture year, the month, and the sex, with their respective estimates, standard errors, and confidence intervals reported. Analysis showed a good fit to the data (AIC = 845.89, BIC = 972.16) and revealed several notable trends in the predictors.

It was not surprising to observe that all sites deviated from Northern Ireland, where diffusion rates were significantly lower across all sites ($p < .001$). The site with the largest deviation from Northern Ireland was site C2 in Cornwall, which exhibited a 42% lower diffusion rate ($\exp(-0.551) = 0.576$, i.e. a 42% decrease). In contrast, the site with the smallest deviation from Northern Ireland was site F1 in Cornwall, which showed a 33% lower diffusion value (($\exp(-0.402) = 0.669$, i.e. a 33% decrease).

We also found that capture year impacted the diffusion rate, where in years 3, 4, and 5, the diffusion rates c were significantly lower than in year 1 (all $p < .05$).

Additionally, the diffusion rate had a notable variation between months. Movement was lower from September to December than in January, indicating a decrease in movement during late autumn and early winter. In February, however, the diffusion rate increased by approximately 10% compared to January ($\exp(0.095) = 1.09965$, i.e. a 9.97% increase; $p = 0.011$).

The final fixed effect, sex, revealed to be significantly lower in the estimate for females compared to males by 16% ($\exp(-0.171) = 0.84282$, i.e. a 16% decrease; $p < .001$).

Random effects for the individual animals were close to zero, suggesting minimal between-animal variability in baseline values after accounting for fixed effects. The residual standard deviation (sqrt(Dispersion) = 9.06) suggests that there was still considerable variability in the response variable that was not accounted for by the fixed and random effects in the model.

The likelihood ratio tests assessed the contribution of individual predictors to the model fit. Excluding capture year did not result in a notable change in model fit (LR statistic = 9.14, df = 4,

$p = .058$). In contrast, excluding month resulted in a significantly worse fit (LR statistic = 341.29, df = 11, $p =< .001$), indicating that this is an important predictor. Similarly, excluding sex led to a significant reduction in model fit (LR statistic = 92.83, df = 1, $p < .001$). Yet, excluding site improved model performance, as indicated by a lower AIC (143.22) for the model without site compared to the full model's AIC (845.89). This suggests that site was a weak predictor of the diffusion metric c.

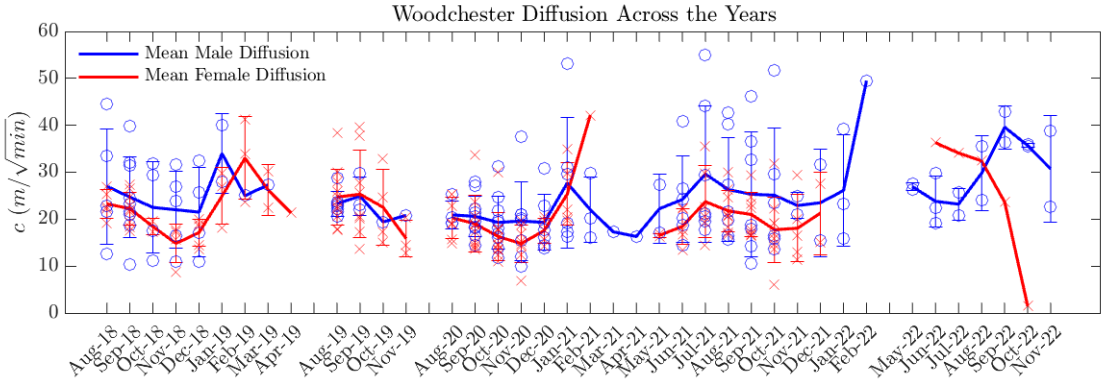
Table 2: Summary of statistical analysis for the diffusion of badgers with a GLMM.

Predictor	**Estimate**	**Standard Error**	**t-Statistic**	**95% Confidence Interval**	**p-Value**
Intercept	3.890	0.048	80.96	[3.79, 3.98]	< 0.001
Site (Reference: Northern Ireland)					
Woodchester	-0.445	0.026	-17.29	[-0.50, -0.39]	< 0.001
C2	-0.551	0.051	-10.74	[-0.65,-0.45]	< 0.001
C4	-0.468	0.056	-8.14	[-0.58,-0.36]	< 0.001
F1	-0.402	0.036	-11.10	[-0.47,-0.33]	< 0.001
F2	-0.473	0.042	-11.36	[-0.56,-0.39]	< 0.001
Capture Year (Ref: Year 1)					
Year 2	-0.035	0.038	-0.91	[-0.11, 0.04]	0.362
Year 3	-0.081	0.040	-2.03	[-0.16, -0.003]	0.043
Year 4	-0.084	0.040	-2.10	[-0.16, -0.006]	0.036
Year 5	-0.098	0.042	-2.35	[-0.18, -0.02]	0.019
Month (Ref: January)					
February	0.095	0.037	2.53	[0.02, 0.17]	0.011
March	0.052	0.042	1.25	[-0.03, 0.13]	0.212
April	0.045	0.047	0.96	[-0.05, 0.14]	0.337
May	0.024	0.050	0.47	[-0.08, 0.12]	0.64
June	-0.074	0.058	-1.27	[-0.19, 0.04]	0.203
July	-0.029	0.040	-0.71	[-0.11, 0.05]	0.475
August	-0.097	0.033	-2.96	[-0.16, -0.03]	0.003
September	-0.231	0.033	-7.06	[-0.30, -0.17]	< 0.001
October	-0.373	0.035	-10.62	[-0.44, -0.30]	< 0.001
November	-0.443	0.038	-11.75	[-0.52, -0.37]	< 0.001
December	-0.283	0.038	-7.37	[-0.36, -0.21]	< 0.001
Sex (Ref: Male)					
Female	-0.171	0.017	-10.13	[-0.20, -0.14]	< 0.001

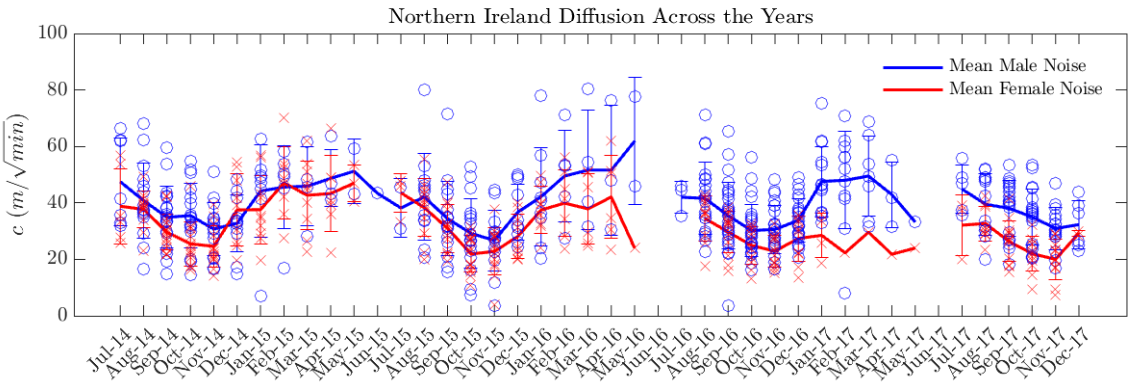
3.2 Extended Dynamic Mode Decomposition

We employed EDMD to analyse the drift term in the stochastic differential equation describing badger movement behaviour. By extracting dominant modes and eigenvalues from observed trajectories, EDMD effectively approximated the drift component, which represents the average or expected rate of change in the system's state. Table 3 summarises the number of social groups identified at each site against the groupings calculated via EDMD. The amalgamated multi-year clustering provides a broad overview of the long-term trends in movement behaviour, where it can be observed that EDMD predicted fewer clusters than those observed via capture–mark–recapture. This pattern was most evident in Northern Ireland, where EDMD predicted 15 metastable clusters where 45 social groups had been identified by field observations, capture/recapture and GPS tracking.

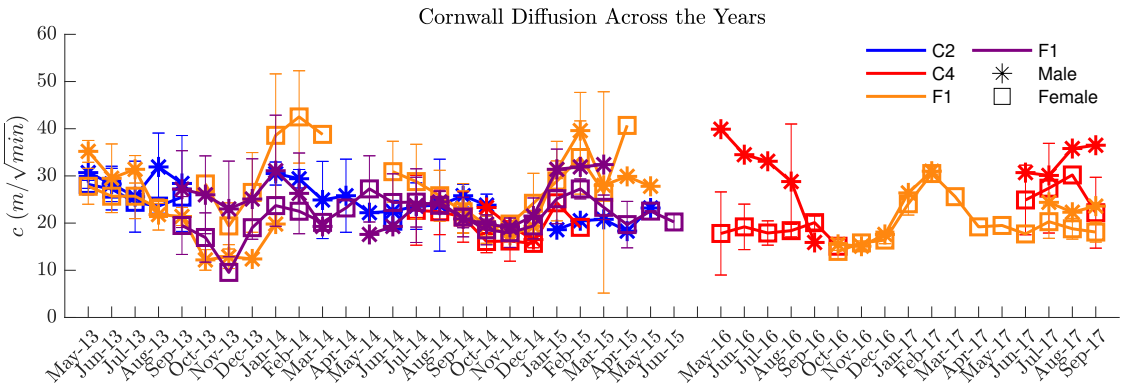
Although some clusters appeared to be stable for periods of time, the number of metastable states can be seen to vary per year (Figure 3). For example, in years two to five at Woodchester, a cluster persisted in the central north of the study area (covering 1.05 km^2, 1.00 km^2, 1.23 km^2, and



(a)



(b)



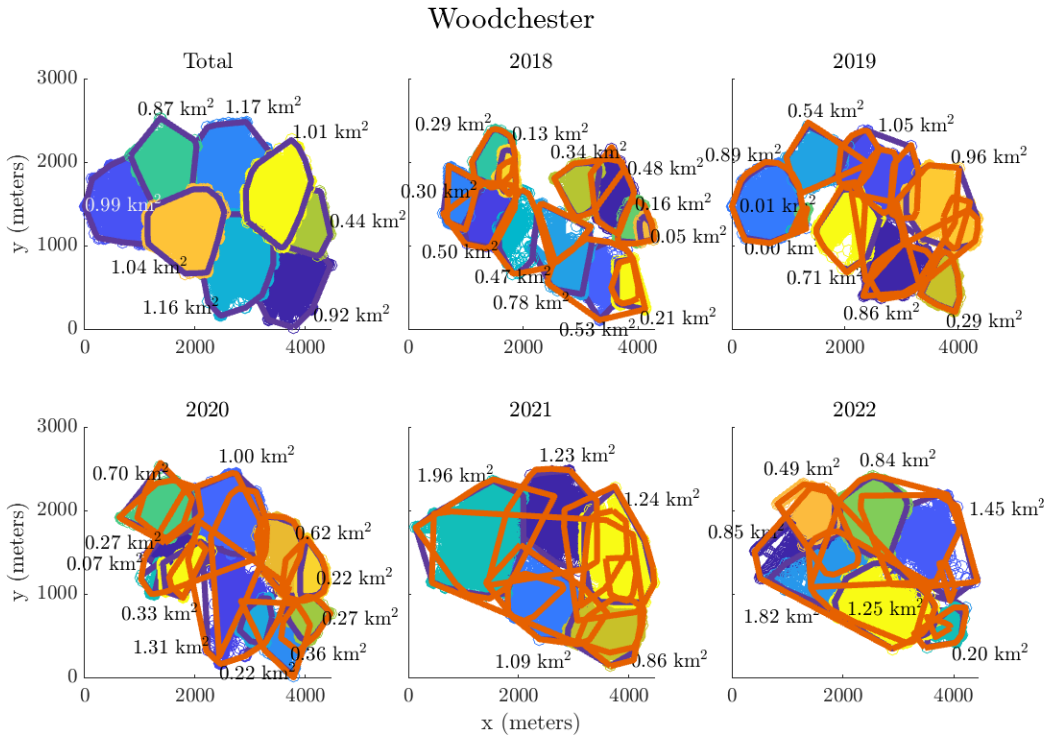
(c)

Figure 2: Mean diffusion (c) values (and standard deviation) for male and female badgers during the study period. Individual diffusion values are shown by month and sex for Woodchester Park and Northern Ireland. Note, that values for October 2022 come from 12 data points originating from a single female badger.

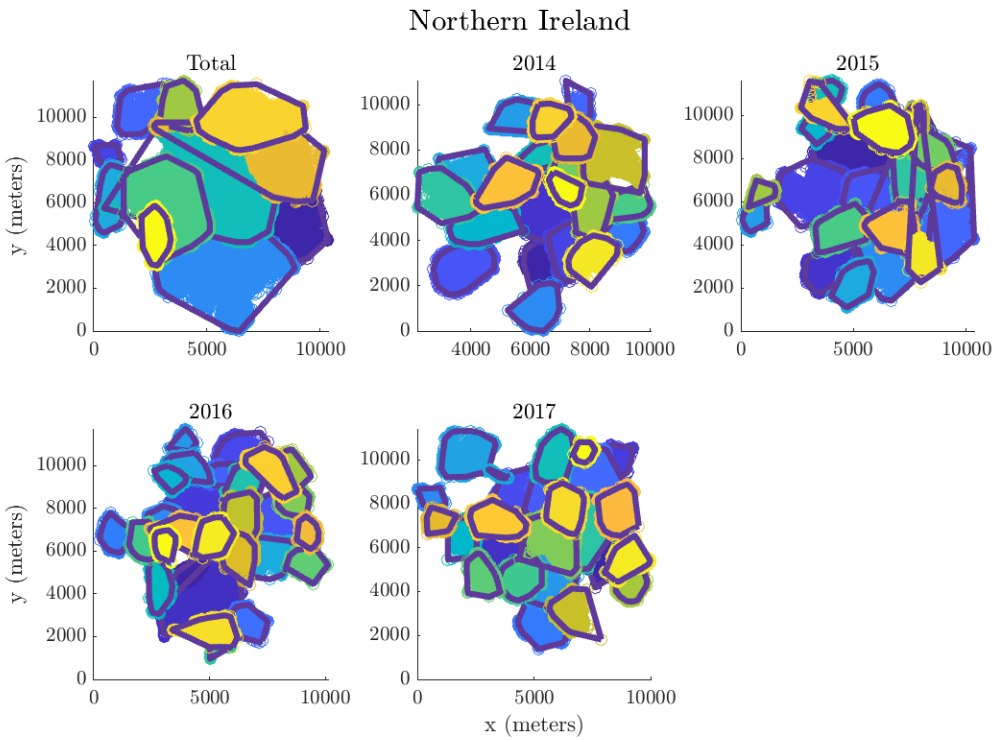
$0.84\,\mathrm{km}^2$, respectively). Yet, during those same years, the estimated number of clusters fluctuated between one and two in the northeast. Similarly, EDMD identified four metastable clusters in the west of the study area throughout the capture period, but this reduced to one large cluster in the fourth year. Similar observations can be made in other sites, for example, in the Cornwall site C4, both the first and second capture year present similar clustering. When there is only one social group present in the data for the year (C2 and C4 during capture year three and two, respectively) then EDMD rightly picks up only one metastable state.

Table 3: Summary of badger social groups for each site estimated from records of capture for each year in comparison with the number of clusters calculated via extended dynamic mode decomposition (EDMD), with the average area of the group home ranges and the area of the clusters for each capture year [mean (standard deviation)].

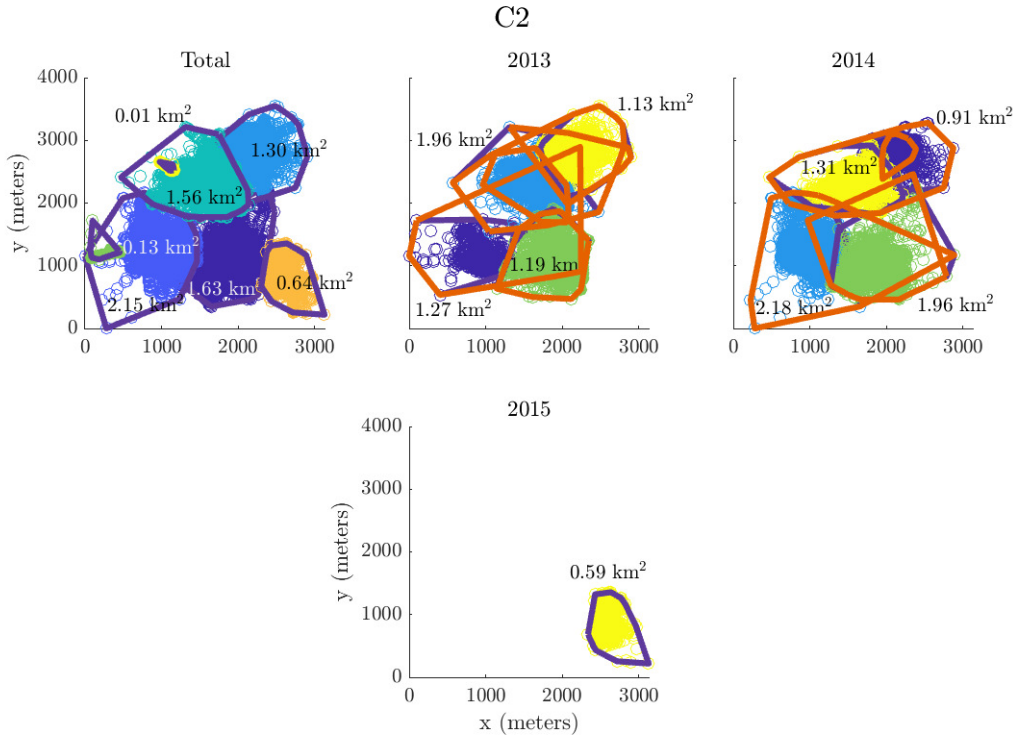
	Woodchester		Northern Ireland	
Capture Year	Social Groups	EDMD	Social Groups	EDMD
Total	18 [1.71 (1.32)]	8 [0.95 (0.23)]	45 [5.13 (3.74)]	11 [9.97 (8.55)]
Year 1	11 [0.44 (0.29)]	12 [0.35 (0.21)]	25 [2.98 (1.94)]	21 [2.37 (1.08)]
Year 2	13 [0.52 (0.30)]	9 [0.59 (0.40)]	33 [3.51 (3.51)]	21 [5.26 (5.79)]
Year 3	10 [0.85 (0.63)]	11 [0.49 (0.38)]	31 [3.53 (2.10)]	28 [2.24 (1.43)]
Year 4	9 [1.56 (1.40)]	5 [1.28 (0.41)]	33 [2.67 (1.71)]	25 [2.59 (1.25)]
Year 5	8 [1.34 (1.23)]	7 [0.99 (0.56)]		
	C2		C4	
Capture Year	Social Groups	EDMD	Social Groups	EDMD
Total	6 [2.20 (1.77)]	7 [1.06 (0.81)]	6 [1.65 (1.14)]	6 [1.24 (0.84)]
Year 1	4 [1.86 (1.17)]	4 [1.39 (0.38)]	4 [0.53 (0.21)]	6 [0.40 (0.19)]
Year 2	3 [2.08 (1.35)]	4 [1.59 (0.59)]	1 [0.20 (0)]	1 [0.20 (0)]
Year 3	1 [0.59 (0)]	1 [0.59 (0)]	3 [1.23 (0.50)]	4 [1.34 (1.02)]
Year 4			2 [2.56 (1.53)]	2 [1.97 (0.68)]
Year 5				
	F1		F2	
Capture Year	Social Groups	EDMD	Social Groups	EDMD
Total	9 [1.71 (1.13)]	5 [2.14 (0.93)]	11 [1.70 (0.78)]	8 [1.30 (0.69)]
Year 1	6 [1.24 (0.67)]	5 [1.51 (0.45)]	7 [1.08 (0.76)]	7 [0.95 (0.71)]
Year 2	7 [1.34 (1.28)]	5 [1.56 (0.90)]	10 [1.13 (0.65)]	9 [0.96 (0.61)]
Year 3	4 [1.18 (0.84)]	4 [1.15 (0.87)]	6 [0.96 (0.67)]	5 [1.13 (0.90)]
Year 4	4 [0.35 (0.08)]	2 [0.57 (0.43)]		
Year 5	3 [0.78 (0.36)]	1 [1.55 (0)]		


Woodchester
Total
2018
2019
2020
2021
2022
y (meters)
x (meters)
0.87 km²
1.17 km²
1.01 km²
0.99 km²
0.44 km²
1.04 km²
1.16 km²
0.92 km²
0.29 km²
0.13 km²
0.34 km²
0.48 km²
0.30 km²
0.16 km²
0.05 km²
0.50 km²
0.47 km²
0.78 km²
0.21 km²
0.53 km²
0.54 km²
1.05 km²
0.89 km²
0.96 km²
0.01 km²
0.00 km²
0.71 km²
0.86 km²
0.29 km²
1.00 km²
0.70 km²
0.62 km²
0.27 km²
0.07 km²
0.22 km²
0.33 km²
0.27 km²
1.31 km²
0.36 km²
0.22 km²
1.23 km²
1.96 km²
1.24 km²
1.09 km²
0.86 km²
0.49 km²
0.84 km²
1.45 km²
0.85 km²
1.82 km²
1.25 km²
0.20 km²

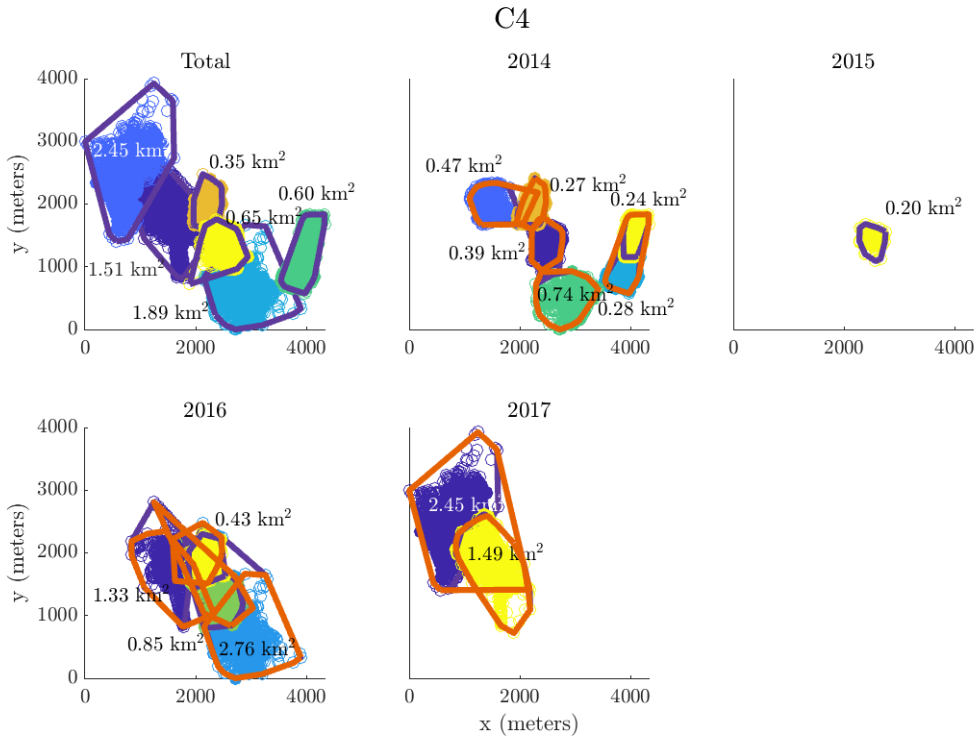
(a)


Northern Ireland
Total
2014
2015
2016
2017
y (meters)
x (meters)

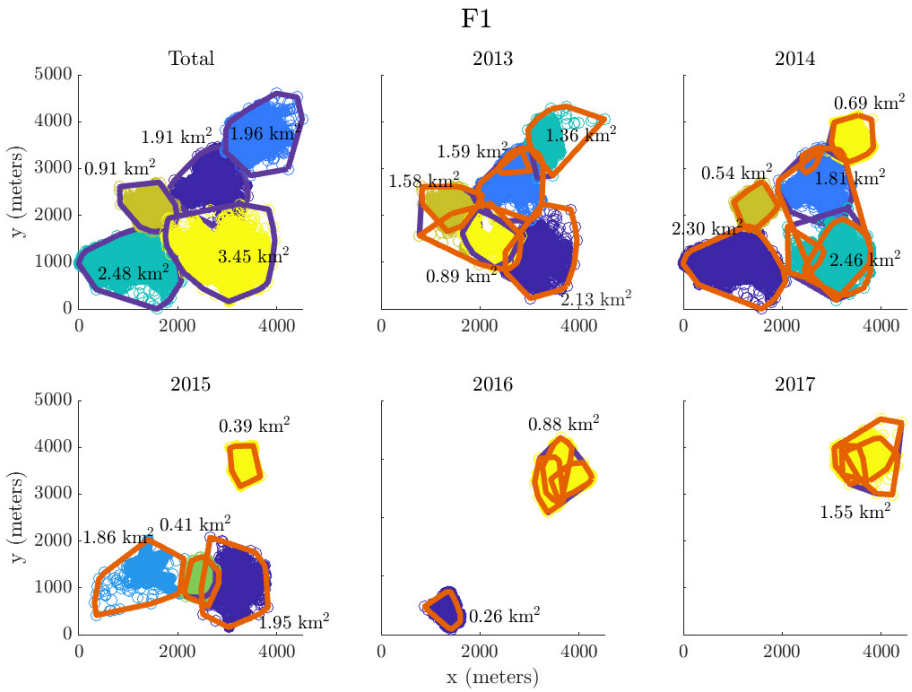
(b)


C2
Total
2013
2014
2015
0.01 km²
1.30 km²
1.56 km²
0.13 km²
1.63 km²
0.64 km²
2.15 km²
1.96 km²
1.13 km²
1.19 km
1.27 km²
0.91 km²
1.31 km²
2.18 km²
1.96 km²
0.59 km²
y (meters)
x (meters)

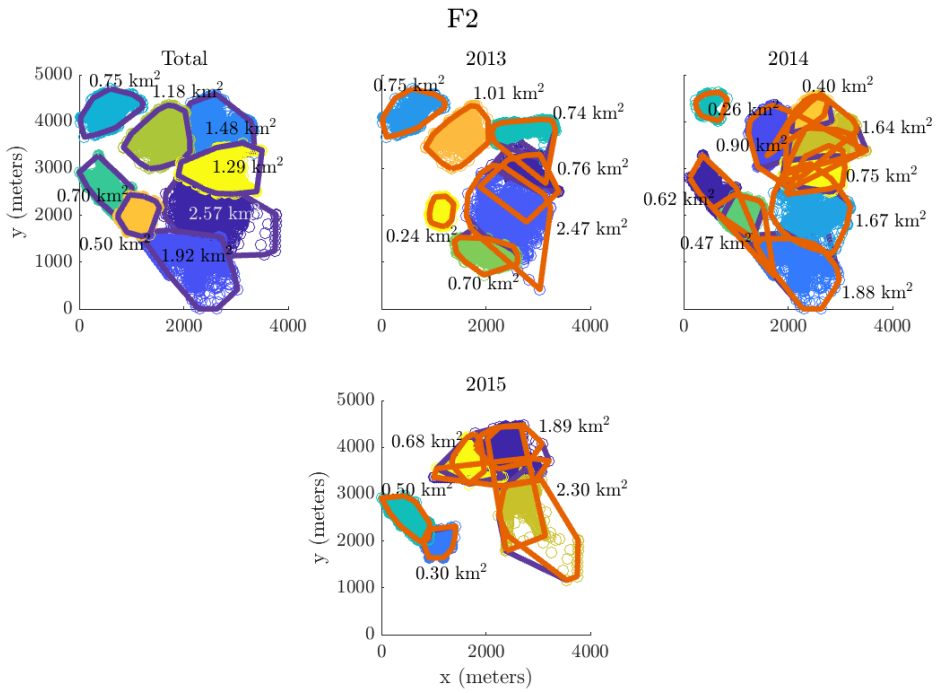
(c)


C4
Total
2014
2015
2016
2017
2.45 km²
0.35 km²
0.60 km²
0.65 km²
1.51 km²
1.89 km²
0.47 km²
0.27 km²
0.24 km²
0.39 km²
0.74 km²
0.28 km²
0.20 km²
0.43 km²
1.33 km²
0.85 km²
2.76 km²
2.45 km²
1.49 km²
y (meters)
x (meters)

(d)



(e)



(f)

Figure 3: The calculated metastable clusters with areas of clusters. The purple lines are the convex hull of the metastable clusters, the orange lines are the convex hull of the social group movement (home ranges) based on the data. Due to the size of the Northern Ireland dataset, the areas of the clusters are not shown.

4 Discussion

Badgers are social animals that often live in groups occupying mutually defended territories across much of their geographic range [9], although extra-territorial movements are not uncommon [13]. Traditional ecological methods, such as bait-marking and capture-mark-recapture studies, have been invaluable for identifying group territories and changes in social structure [35, 45], while radio-tracking has provided further insights into individual movement patterns despite its labour-intensive nature and limited resolution [46]. The advent of GPS technology has since revolutionised the study of badger movements, offering finer-scale tracking that generates rich datasets suited to novel mathematical analyses.

Building on these advancements, our study applied a combination of diffusion analysis and EDMD to investigate the intricate movement patterns and social behaviours of badgers. This approach allowed us to characterise movement patterns across fast (diffusion) and slow (drift) spatiotemporal scales, while identifying effective metastable group home ranges over time. The analysis spanned GPS data from badger populations in three distinct regions of the British Isles, representing a range of local badger densities. Importantly, the study has included populations exposed to varying degrees of culling, providing a unique opportunity to examine the impact of external interventions on movement behaviour. The diversity in conditions across these populations underscores the utility of our novel analytical approach and highlights its potential to yield new insights into badger spatial organisation and movement dynamics.

We first examined the fast temporal dynamics from badger GPS data to gain an understanding of regular badger movements. Diffusion, a metric that captures both spatial extent (distance) and frequency of movement, was used to quantify these dynamics. Our results indicate that males exhibit a higher diffusion value than females, suggesting they have greater spatial mobility. This could mean they are covering larger distances, visiting more locations, or roaming across a broader area more frequently. In contrast, the movements of females appear more localised. Research on the regular, individual movements of male and female badgers has been limited, as the use of GPS tracking technology has only become widespread in the past decade. Prior to the introduction of GPS tracking, many studies focused on movement data collected between capture events, which provided only intermittent snapshots of badger activity [13]. These studies typically infer movement patterns based on locations obtained at specific intervals, which can miss finer details of regular movement dynamics.

Monthly variations in badger movement are also evident in the estimated diffusion value, especially during the latter half of the year (Table 2). In February, we observe that the diffusion rate for females peaks, surpassing that of males, which could be associated with mating behaviours. The behaviour that there is an increased amount of movement was also observed by Woodroffe et al. [47], whereby using the same data they reported clear peaks in the frequency of trespassing occurred in February and September. This trend is most apparent in Figure 2a, where increased movement during this time may reflect heightened activity related to reproductive strategies [48]. This observation is consistent with broader trends in wildlife movement patterns, where mating strategies and the need for reproduction can drive differences in spatial behaviours between the sexes. Furthermore, when compared to January (the reference month in the GLMM), the model suggests that movement patterns in the latter half of the year differ significantly, indicating seasonal shifts in behaviour. This reinforces the idea already seen (see [47]) that environmental and reproductive factors influence movement dynamics across different times of the year.

In our analysis, the site was initially a significant predictor of diffusion values, with animals in comparison to Northern Ireland exhibiting significantly lower diffusion values ($p < .001$). Existing research indicates that badger movement can vary with population density [49]. For instance, [49] compares two populations, noting higher movement frequency in a lower-density population with a more fluid social structure, and lower movement frequency in a higher-density population. Badger populations in Woodchester and Cornwall have historically exhibited higher densities compared to those in Northern Ireland [20, 46, 47]. However, for the Woodchester years included in this study, no density estimates have been published. These differences in density could be driving the variation in diffusion metrics observed here, as badgers in less dense populations may range

over larger territories searching for resources or mates, leading to higher diffusion values. However, culling is also a critical factor that likely influences badger movement behaviour. Studies have shown that culling disrupts spatial organisation and can significantly increase ranging behaviour, potentially amplifying diffusion metrics [18, 28]. For example, in this dataset, Woodchester was exposed to culling on neighbouring land, with the impact likely increasing over time. Similarly, site C4 experienced culling from September 2016 onward, while F1, F2, and C2 had no culling during the study period, and Northern Ireland was subject to prolonged small-scale culling under TVR. These varying culling contexts likely contribute to the observed diffusion patterns, highlighting the need to evaluate both population density and culling effects when interpreting badger movement dynamics. Despite the apparent influence of density and culling, model comparison using AIC indicated that the reduced model without the site variable (AIC = 143.2201) provided a much better fit than the full model (AIC = 845.8872). This result suggests that the site variable does not meaningfully improve the model's overall explanatory power. However, it is important to recognise that 'site' encompasses multiple characteristics beyond just population density, such as differences in culling history or habitat composition. Consequently, the potential effect of density or culling could be obscured by other site-specific factors. By focusing on the reduced model, which emphasises variables like sex and month, we capture more directly relevant influences on diffusion, with geographic location adding unnecessary complexity.

The second part of the study examined slow temporal dynamics to gain understanding of the long-term behaviour of badgers. In this context, 'long-term' refers to patterns observed over several months, as analysed using EDMD to capture and characterise these dynamics. The main advantage of using EDMD to generate these clusters is that it makes use of the temporal dimension of the data, which considers that a badger can leave its territory and then return again to the same territory. Our results identify spatial regions where metastable clusters remain consistent over time at some study sites. For instance, in the Woodchester data, some clusters exhibit long-term stability throughout the study period. However, in other areas, such as North-East of Woodchester park, substantial changes in the estimated areas were observed, for example, with clusters increasing in size over the period 2018 to 2021 from $0.29\,\mathrm{km}^2$ to $1.96\,\mathrm{km}^2$ then decreasing in 2022 to $0.49\,\mathrm{km}^2$. The merging and splitting of badger group home ranges may suggest a breakdown in territorial behaviour, with increased overlap in the areas used by individuals which formerly had exclusive territories. Such changes may be behind the movement patterns observed in populations that are subject to culling [18]. We observed a general decrease in the number of clusters for Woodchester (Table 3), accompanied by an increase in cluster area, suggesting that individual animals are ranging over larger areas. This pattern indicates an increase in roaming behaviour and appears to have intensified during the cull, initiated in 2018 (R. Delahay, R.Cox, unpublished data). The observed variations in metastable clusters in Woodchester may therefore be linked to the onset of culling around the study area. These findings highlight the potential of metastable states as a valuable tool for tracking changes in movement dynamics, particularly in response to external interventions such as culling. Unlike simpler methods, such as bait marking or basic measures of ranging behaviour, EDMD allows for the identification of subtle, higher-order patterns in movement dynamics and the transitions between behavioural states over time. This capability provides a more profound understanding of how interventions influence movement behaviours at a system-wide level.

Furthermore, it has been observed that the obtained metastable states often overlap multiple known group ranges, suggesting that badgers were using larger areas than advertised by territorial marking. In the Northern Ireland data set, for example, the average difference in the number of social groups and EDMD clusters over the four years is 3.5, with fewer EDMD clusters on average, and a gradual increase in the number of clusters and group home ranges. Considering that social groups typically range from 25 to 33 and EDMD clusters from 21 to 28, this represents a moderate difference. This difference in groupings may be due to the removal of bTB-positive badgers after the first year of the study, which might have changed the social structure and encouraged movement among the remaining badgers. On the other hand, Cornwall sites C2 and C4 had an average difference of -0.3 and -0.75, respectively, showing there is very little difference between the field observation and this method. For both sites, the number of social groups ranged from 1 to 4, while EDMD clusters ranged from 1 to 4 for C2 and 1 to 6 for C4. These narrow ranges underscore

that the differences are minor, reflecting a close agreement between the two approaches. In a few cases, we observed a match between the number of social groups identified through traditional field methods and the clusters detected by EDMD. This alignment was especially clear in instances where a single social group dominates, such as in Cornwall site C4 in the second capture year, where both methods detected just one group. Similarly, in the third capture year for Cornwall site C2, only one badger was collared, as work at that site ceased after that year, making this result less meaningful. Nevertheless, in these instances, EDMD consistently captures the spatial boundaries and movement patterns that correspond to the social ranges of badgers. This occasional alignment between EDMD and field observations suggests that while EDMD has the potential to offer an alternative approach to understanding badger social groups and space use, it may also be capturing different aspects of badger movement dynamics compared to traditional methods. Additionally, discrepancies between the methods could reflect external influences, such as the disruptions caused by culling, which complicate the detection of stable patterns. These findings highlight the complementary nature of the methods and the need to carefully interpret results within the context of external pressures.

A limitation of our approach stems from the uneven temporal distribution of GPS data, particularly the scarcity of fixes during the first half of the year due to restrictions on trapping during the breeding season to avoid disturbing reproductive females and their offspring [37]. Additionally, the varying lifespan of GPS device batteries contributed to inconsistent data collection, with those in Cornwall averaging 110 days of recording and those in Northern Ireland lasting up to 273 days [19, 20]. Since many collars were deployed in July and August, most devices ceased functioning by the following May, resulting in a lack of data for June and a limited temporal window of data collection. This inconsistency mainly impacted our diffusion analysis, making it challenging to compare movement patterns across different months. Furthermore, the non-uniform timing of GPS fixes introduced additional inconsistency, impacting our EDMD analysis, which requires equal time intervals between locational fixes. To address this inconsistency, we used linear interpolation. Cornwall's data required minimal manipulation, as locations were recorded every 20 minutes, whereas Woodchester's 35-minute intervals and Northern Ireland's hour intervals necessitated interpolation. While increasing the frequency of fixes could enhance data accuracy, it would also reduce battery life. Additionally, we recognise that a significant number of data points were removed due to the low number of satellites detected (of the overall 19.0% removed for Woodchester, 55.6% of that data was for low number of satellites), which could be attributed to the high proportion of woodland (which can obscure GPS satellite connections) in the study areas. This raises the possibility of introducing a systematic bias, as these data exclusions may have disproportionately affected locations nearer the centre of each badger's range, potentially closer to their main sett. This bias could skew our understanding of badger spatial behaviour, particularly in relation to their core areas of activity. However, Woodroffe et al. [19] (supplementary material) showed that woodland did not affect collar precision in Cornwall, where the same qualitative results were produced when no filters were applied to the data. Despite these limitations and the need to manipulate the data to use some methods, the diffusion results offer a broader understanding of the regular movements of badgers. Furthermore, the EDMD analysis of metastable clusters offers novel insights into long-term badger movement patterns, adding value to the findings even with these constraints on the dataset.

In summary, this study demonstrates the value of EDMD in identifying spatial clusters associated with badger group home ranges. In Northern Ireland and Woodchester Park, the discrepancy between EDMD clusters and known social groups is likely driven by the removal of animals, reflecting how population management interventions may disrupt social cohesion and influence movement patterns. However, discrepancies with field-derived methods highlight the need for further investigation into both the EDMD analysis approach and the accuracy and completeness of the field data. Additionally, the diffusion analysis revealed distinct patterns of movement across temporal scales. These analyses present a novel, mathematically driven perspective for interpreting badger spatial behaviour. Future research should aim to refine EDMD, addressing limitations in GPS data (e.g., variable fix intervals), and explore its potential for monitoring wildlife responses to environmental changes and management interventions.

5 Conclusions

The objective of this study was to ascertain the movements of badgers from GPS tracking data and to investigate spatial structure in wild animal populations using mathematical tools. Whilst badger territories have been defined for decades using ecological methods (e.g. bait marking [9, 50]), the metastability of movement behaviour of the animals in the population has not been estimated. Our results suggest that these combined analyses of diffusion and metastability offer a different interpretation of badger spatial behaviour. The diffusion metrics reveal dynamic movement patterns, while the EDMD offers a useful method that captures the social structures that underlie these movements.

Acknowledgement

This work was partly co-funded by the European Union's Horizon Europe Project 101136346 EU-PAHW. For the purpose of Open Access, the author has applied a Creative Commons Attribution (CC BY) public copyright licence to any Author Accepted Manuscript version arising from this submission

Data Statement

The code used to analyse the data is available at the Github repository:
https://github.com/JFurber/Analysis_of_Badger_Movement.git. The code includes all scripts necessary to reproduce the analysis and results presented in the paper. The analysis was performed using a mixture of R, Python and MATLAB, with specific packages or libraries, which are documented in the repository.

The datasets analysed during the current study are available from the corresponding author on reasonable request due to the sensitivity of the data. The GPS tracking data of badger movements, collected from Cornwall, are lodged on Movebank (www.movebank.org; Movebank Project 158275131). The GPS tracking data of badger movements, collected from Woodchester, are currently unpublished (R.Delahay, R.Cox). The GPS tracking data of badger movements, collected from Northern Ireland, are lodged on Movebank (www.movebank.org; Movebank Project 5019193945).

References

[1] 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*, p. 111, 2022.

[2] 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*, vol. 116, no. 29, pp. 14424–14433, 2019.

[3] R. S. Schick, S. R. Loarie, F. Colchero, B. D. Best, A. Boustany, 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*, vol. 11, no. 12, pp. 1338–1350, 2008.

[4] 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. New York, NY: Springer New York, 2012.

[5] H. K. Preisler, A. A. Ager, and M. J. Wisdom, "Analyzing animal movement patterns using potential functions," *Ecosphere*, vol. 4, no. 3, pp. 1–13, 2013.

[6] G. C. Smith, C. L. Cheeseman, D. Wilkinson, and R. S. Clifton-Hadley, "A model of bovine tuberculosis in the badger *Meles meles*: the inclusion of cattle and the use of a live test," *Journal of Applied Ecology*, vol. 38, no. 3, pp. 520–535, 2001.

[7] C. Schütte and M. Sarich, *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches.* No. 24 in Courant Lecture Notes, American Mathematical Society, 2013.

[8] A. Bovier and F. den Hollander, *Metastability: A Potential-Theoretic Approach.* Grundlehren der mathematischen Wissenschaften, Springer International Publishing, 2016.

[9] H. Kruuk, "Spatial organization and territorial behaviour of the european badger meles meles," *Journal of Zoology*, vol. 184, no. 1, pp. 1–19, 1978.

[10] M. J. Noonan, H. V. Tinnesand, C. T. Müller, F. Rosell, D. W. Macdonald, and C. D. Buesching, "Knowing me, knowing you: Anal gland secretion of european badgers (meles meles) codes for individuality, sex and social group membership.," *J Chem Ecol*, vol. 45, pp. 823–837, Oct 2019.

[11] 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*, vol. 53, no. 4, pp. 265–275, 2007.

[12] 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*, vol. 270, no. 3, pp. 437–447, 2006.

[13] 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*, vol. 265, pp. 1269–1276, Jul 1998.

[14] J. da Silva, R. Woodroffe, and D. W. Macdonald, "Habitat, food availability and group territoriality in the european badger, meles meles," *Oecologia*, vol. 95, no. 4, pp. 558–564, 1993.

[15] R. Woodroffe, S. D. W. Frost, and R. S. Clifton-Hadley, "Attempts to control tuberculosis in cattle by removing infected badgers: Constraints imposed by live test sensitivity," *Journal of Applied Ecology*, vol. 36, no. 4, pp. 494–501, 1999.

[16] 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*, vol. 45, no. 4, pp. 1062–1088, 2009.

[17] D. Ní Bhuachalla, L. A. L. 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)*, vol. 6, pp. 27–38, 2015.

[18] 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*, vol. 56, no. 11, pp. 2390–2399, 2019.

[19] 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*, vol. 19, pp. 1201–1208, Oct 2016.

[20] S. H. A. Redpath, N. J. Marks, F. D. Menzies, M. J. H. O'Hagan, R. P. Wilson, S. Smith, E. A. Magowan, D. W. McClune, S. F. Collins, C. M. McCormick, and D. M. Scantlebury, "Impact of test, vaccinate and remove protocol on home ranges and nightly movements of badgers in a medium density population," *Scientific Reports*, vol. 13, no. 1, p. 2592, 2023.

[21] J. Crispell, S. Cassidy, K. Kenny, G. McGrath, S. Warde, H. Cameron, G. Rossi, T. MacWhite, P. C. L. White, S. J. Lycett, R. R. Kao, J. Moriarty, and S. V. Gordon, "Mycobacterium bovis genomics reveals transmission of infection between cattle and deer in ireland.," *Microbial Genomics*, 2020.

[22] A. J. van Tonder, M. J. Thornton, A. J. K. Conlan, K. A. Jolley, L. Goolding, A. P. Mitchell, J. Dale, E. Palkopoulou, P. J. Hogarth, R. G. Hewinson, J. L. N. Wood, and J. Parkhill, "Inferring mycobacterium bovis transmission between cattle and badgers using isolates from the randomised badger culling trial.," *PLoS Pathog*, vol. 17, p. e1010075, Nov 2021.

[23] 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*, vol. 87, no. 1, pp. 101–112, 2018.

[24] R. Woodroffe, C. A. Donnelly, G. Wei, D. Cox, F. J. Bourne, T. Burke, R. K. Butlin, C. Cheeseman, G. Gettinby, P. Gilks, S. Hedges, H. E. Jenkins, W. T. Johnston, J. P. McInerney, W. I. Morrison, and L. C. Pope, "Social group size affects *Mycobacterium bovis* infection in European badgers (*Meles meles*)," *Journal of Animal Ecology*, vol. 78, no. 4, pp. 818–827, 2009.

[25] M. Abdou, K. Frankena, J. O'Keeffe, and A. W. Byrne, "Effect of culling and vaccination on bovine tuberculosis infection in a european badger (meles meles) population by spatial simulation modelling," *Preventive Veterinary Medicine*, vol. 125, pp. 19–30, 2016.

[26] F. A. Tuyttens, R. J. Delahay, D. W. Macdonald, C. L. Cheeseman, B. Long, C. A. Donnelly, and C. Newman, "The impact of repeated widespread badger culling on cattle tuberculosis in four areas in ireland," *Preventive Veterinary Medicine*, vol. 45, no. 3-4, pp. 123–134, 2000.

[27] L. A. L. Corner, D. Murphy, and E. Gormley, "Temporal dynamics of mycobacterium bovis infection in a high-density badger population: implications for management," *Journal of Applied Ecology*, vol. 40, no. 4, pp. 710–722, 2003.

[28] R. Woodroffe, C. A. Donnelly, D. R. Cox, J. Bourne, C. L. Cheeseman, R. J. Delahay, G. Gettinby, J. P. McInerney, and W. I. Morrison, "Effects of culling on badger meles meles spatial organization: implications for the control of bovine tuberculosis," *Journal of Applied Ecology*, vol. 43, no. 1, pp. 1–10, 2006.

[29] S. P. Carter, R. J. Delahay, G. C. Smith, D. W. Macdonald, P. Riordan, T. R. Etherington, E. R. Pimley, N. J. Walker, and C. L. Cheeseman, "Culling-induced social perturbation in eurasian badgers meles meles and its implications for the control of bovine tuberculosis," *Journal of Applied Ecology*, vol. 44, no. 4, pp. 692–701, 2007.

[30] 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*, vol. 10, no. 1, p. 10, 2022.

[31] C. L. Cheeseman, G. W. Jones, J. Gallagher, and P. J. Mallinson, "The population structure, density and prevalence of tuberculosis (*Mycobacterium bovis*) in badgers (*Meles meles*) from four areas in south-west England," *Journal of Applied Ecology*, vol. 18, no. 3, pp. 795–804, 1981.

[32] F. D. Menzies, C. M. McCormick, M. J. H. O'Hagan, S. F. Collins, J. McEwan, C. F. McGeown, G. E. McHugh, C. D. Hart, L. A. Stringer, C. Molloy, G. Burns, S. J. McBride, L. P. Doyle, E. A. Courcier, K. R. McBride, J. McNair, S. Thompson, D. M. Corbett, R. G. Harwood, and N. A. Trimble, "Test and vaccinate or remove: Methodology and preliminary results from a badger intervention research project," *Veterinary Record*, vol. 189, no. 5, p. e248, 2021.

[33] E. Courcier, A. Pascual-Linaza, M. Arnold, C. McCormick, D. Corbett, M. O'Hagan, S. Collins, N. Trimble, C. McGeown, G. McHugh, K. McBride, J. McNair, S. Thompson, I. Patterson, and F. Menzies, "Evaluating the application of the dual path platform VetTB test for badgers (*Meles meles*) in the test and vaccinate or remove (TVR) wildlife research intervention project in Northern Ireland," *Research in Veterinary Science*, vol. 130, pp. 170–178, 2020.

[34] M. J. H. O'Hagan, A. W. Gordon, C. M. McCormick, S. F. Collins, N. A. Trimble, C. F. McGeown, G. E. McHugh, K. R. McBride, and F. D. Menzies, "Effect of selective removal of badgers

(*Meles meles*) on ranging behaviour during a 'Test and Vaccinate or Remove' intervention in Northern Ireland," *Epidemiology & Infection*, vol. 149, p. e125, 2021.

[35] R. J. Delahay, J. A. Brown, P. J. Mallinson, P. D. Spyvee, D. Handoll, L. M. Rogers, and C. L. Cheeseman, "The use of marked bait in studies of the territorial organization of the european badger (*Meles meles*)," *Mammal Review*, vol. 30, no. 2, pp. 73–87, 2000.

[36] UK Govenment, "How to cage trap, vaccinate and mark badgers to control bovine tb," 2024. https://www.gov.uk/government/publications/cage-trap-and-mark-badgers-to-enable-vaccination-to-prevent-the-spread-of-bovine-tb/how-to-cage-trap-vaccinate-and-mark-badgers-to-control-bovine-tb. Accessed: 6 December 2024.

[37] Acts of Parliament (UK Statutes), "Protection of Badgers Act 1992," 1992 (c. 51). Available at: https://www.legislation.gov.uk/ukpga/1992/51. (Accessed 7 March 2024).

[38] J. Bourne, C. Donnelly, D. Cox, G. Gettinby, J. McInerney, W. Morrison, *et al.*, *Bovine TB: The Scientific Evidence*. London: Defra, 2007. https://www.bovinetb.info/docs/final_report.pdf.

[39] P. Mishra, C. M. Pandey, U. Singh, A. Gupta, C. Sahu, and A. Keshri, "Descriptive statistics and normality tests for statistical data," *Ann Card Anaesth*, vol. 22, pp. 67–72, Jan-Mar 2019.

[40] The MathWorks Inc., *MATLAB version: 24.1.0.2578822 (R2024a)*, 2024. https://www.mathworks.com.

[41] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A data–driven approximation of the Koopman operator: extending dynamic mode decomposition," *Journal of Nonlinear Science*, vol. 25, no. 6, pp. 1307–1346, 2015.

[42] S. Klus, P. Koltai, and C. Schütte, "On the numerical approximation of the Perron–Frobenius and Koopman operator," *Journal of Computational Dynamics*, vol. 3, no. 1, pp. 51–79, 2016.

[43] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *Advances in neural information processing systems*, vol. 14, pp. 849–856, 2001.

[44] M. O'Hagan, C. McCormick, S. Collins, K. McBride, and F. Menzies, "Are major roads effective barriers for badger (meles meles) movements?," *Research in Veterinary Science*, vol. 138, pp. 49–52, 2021.

[45] 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*, vol. 265, pp. 1269–1276, Jul 1998.

[46] B. Garnett, R. Delahay, and T. Roper, "Ranging behaviour of European badgers (*Meles meles*) in relation to bovine tuberculosis (*Mycobacterium bovis*) infection," *Applied Animal Behaviour Science*, vol. 94, no. 3, pp. 331–340, 2005.

[47] R. Woodroffe, C. A. Donnelly, C. Ham, S. Y. B. Jackson, K. Moyes, K. Chapman, N. G. Stratton, and S. J. Cartwright, "Ranging behaviour of badgers *Meles meles* vaccinated with Bacillus Calmette Guerin," *Journal of Applied Ecology*, vol. 54, no. 3, pp. 718–725, 2017.

[48] R. Woodroffe, D. W. Macdonald, and J. da Silva, "Dispersal and philopatry in the european badger, meles meles," *Journal of Zoology*, vol. 237, no. 2, pp. 227–239, 1995.

[49] C. L. Cheeseman, W. J. Cresswell, S. Harris, and P. J. Mallinson, "Comparison of dispersal and other movements in two badger (*Meles meles*) populations," *Mammal Review*, vol. 18, no. 1, pp. 51–59, 1988.

[50] P. Skoog, *The food of the Swedish badger, Meles meles L.* PhD thesis, Stockholm University, 1970.

A Appendix

A.1 Extended Dynamic Mode Decomposition

A.1.1 Choice of Clusters for EDMD

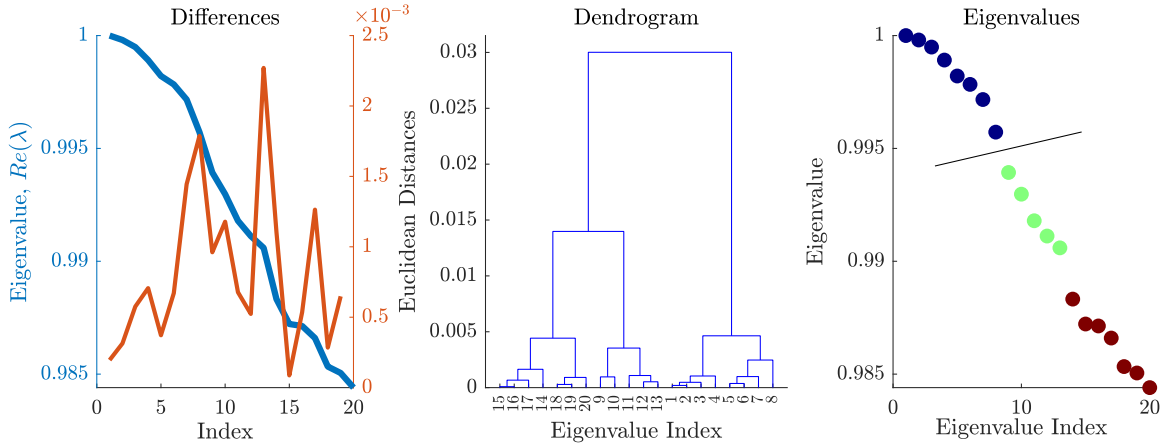
Occasionally, it is not clear where the spectral gap in the spectrum lies. This does not necessarily exclude the usage of spectral methods for identifying the metastable states of a system. So, for an order of N eigenvalues, $\lambda_1 > \lambda_2 > \ldots > \lambda_N$, then a spectral gap is defined as the first λ_i such that (5) is satisfied.

The following figures firstly contain the eigenvalues plotted against index i, along with the successive difference. The key eigenvalue is chosen for the index with the first largest difference, followed by a significant reduction in difference. Then, we use a dendrogram for clustering the eigenvalues based on similarities, using the Ward method. The optimal numbering of clusters based on the dendrogram (and maximum Euclidean distance between clusters 0.01) are seen in the colouring of the third figure, with the line representing the break in eigenvalues based on the first figure. Both offer similar results, bar site C2, which estimated there are four key eigenvalues when we have chosen seven. The overall choice for this process is heuristic, and there is no clear set method for determining the key number of eigenvalues/eigenfunctions to be chosen.

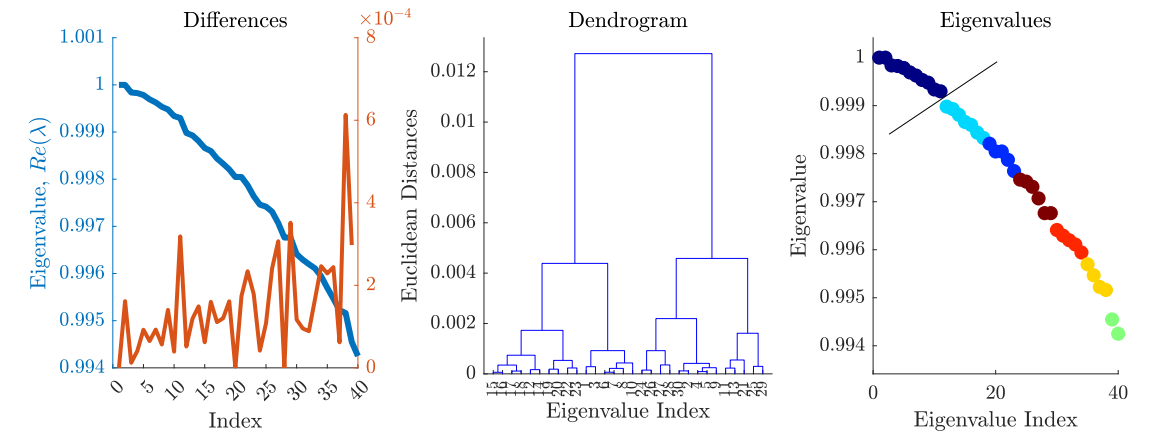
Table 4: Comparison of methods to choose the spectral gap for all sites and years: dendrogram (De) and using the differences in eigenvalues and choosing the most considerable difference (Di). Figure 4 highlights the methods for the total data sets for each site.

	Total		Year 1		Year 2		Year 3		Year 4		Year 5	
	De	Di	De	Di	De	Di	De	Di	De	Di	De	Di
Woodchester	8	8	11	12	9	9	11	11	5	5	10	7
Northern Ireland	11	11	14	21	21	21	14	28	12	25		
C2	4	7	3	4	2	4	1	1				
C4	6	6	5	6	1	1	4	4	2	2		
F1	5	5	5	5	5	5	4	4	2	2	1	1
F2	8	8	6	7	5	9	3	5				

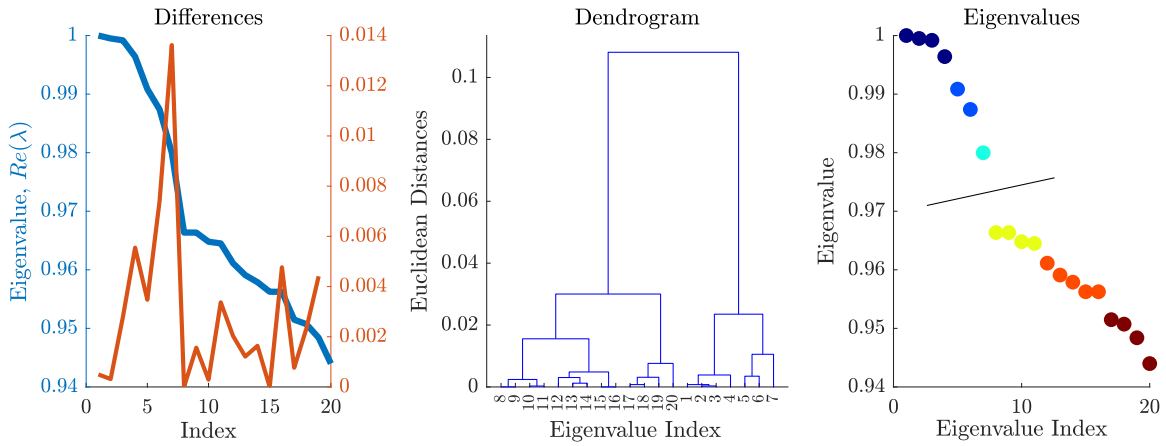
Woodchester



Northern Ireland



C2



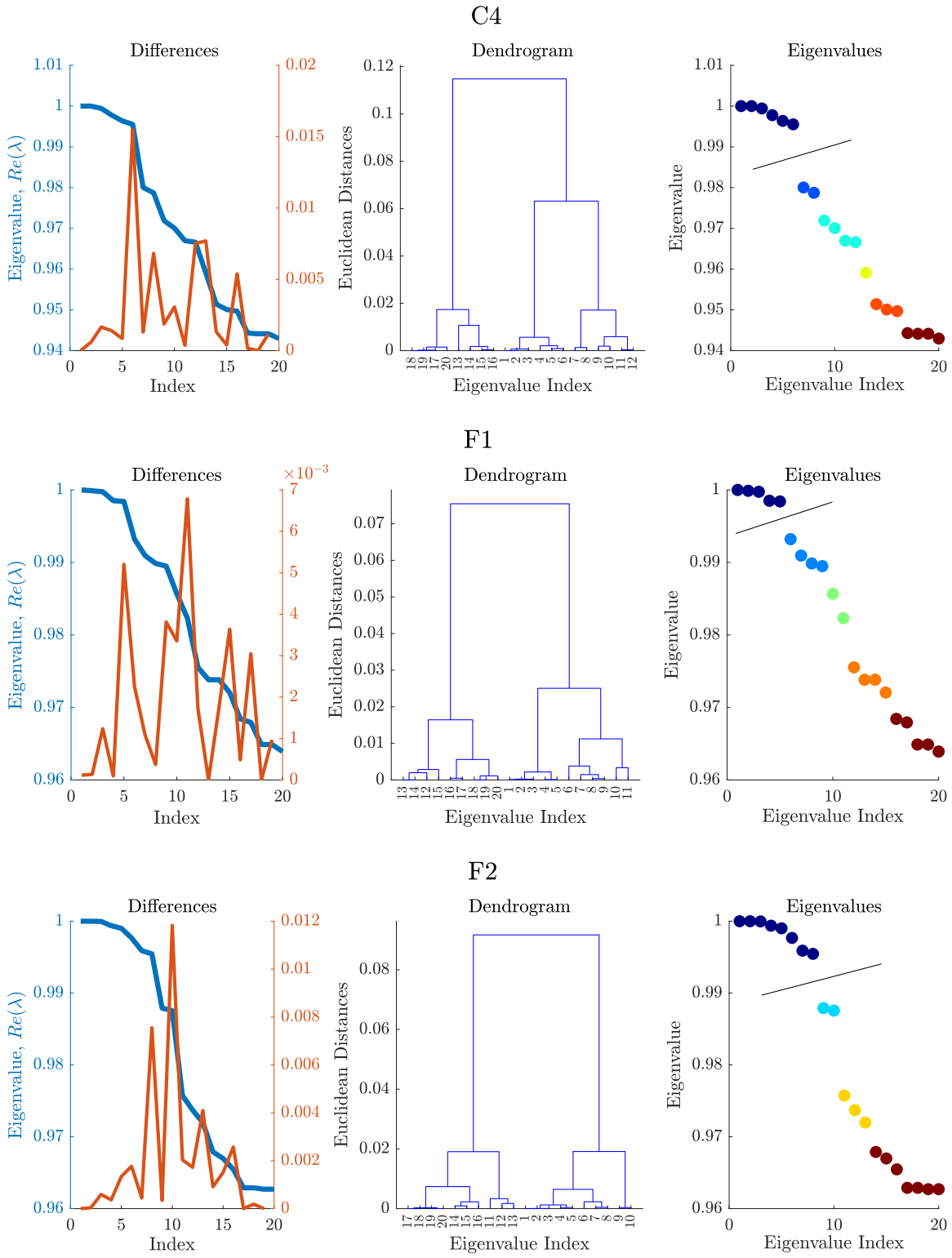


Figure 4: Comparing the methods to choose the spectral gap based on the differences in the eigenvalues and through the use of a dendrogram. The result of the optimal choice seen in the third figure, with the line relating to the calculated differences and the colours relating to the dendrogram.