

## RESEARCH METHODS GUIDE

# How to scale up from animal movement decisions to spatiotemporal patterns: An approach via step selection

Jonathan R. Potts<sup>1</sup>  | Luca Börger<sup>2,3</sup> <sup>1</sup>School of Mathematics and Statistics,  
University of Sheffield, Sheffield, UK<sup>2</sup>Department of Biosciences, College of  
Science, Swansea University, Swansea, UK<sup>3</sup>Centre for Biomathematics, College of  
Science, Swansea University, Swansea, UK**Correspondence**

Jonathan R. Potts

Email: [j.potts@sheffield.ac.uk](mailto:j.potts@sheffield.ac.uk)**Funding information**Engineering and Physical Sciences  
Research Council, Grant/Award Number:  
EP/V002988/1**Handling Editor:** Greta Bocedi**Abstract**

1. Uncovering the mechanisms behind animal space use patterns is of vital importance for predictive ecology, thus conservation and management of ecosystems. Movement is a core driver of those patterns so understanding how movement mechanisms give rise to space use patterns has become an increasingly active area of research.
2. This study focuses on a particular strand of research in this area, based around step selection analysis (SSA). SSA is a popular way of inferring drivers of movement decisions, but, perhaps less well appreciated, it also parametrises a model of animal movement. Of key interest is that this model can be propagated forwards in time to predict the space use patterns over broader spatial and temporal scales than those that pertain to the proximate movement decisions of animals.
3. Here, we provide a guide for understanding and using the various existing techniques for scaling up step selection models to predict broad-scale space use patterns. We give practical guidance on when to use which technique, as well as specific examples together with code in R and Python.
4. By pulling together various disparate techniques into one place, and providing code and instructions in simple examples, we hope to highlight the importance of these techniques and make them accessible to a wider range of ecologists, ultimately helping expand the usefulness of SSA.

**KEYWORDS**

home range, individual-based models, integro-difference equations, movement ecology, partial differential equations, space use, step selection, utilisation distribution

## 1 | INTRODUCTION

Understanding and predicting the spatiotemporal distribution of individuals, populations and communities is a fundamental aim of ecological research. Key drivers of these spatial dynamics are the

movement decisions of individuals in response to individual and local environmental conditions and resources (Fryxell et al., 2008; Nathan et al., 2008), the distribution of conspecific and heterospecific individuals (Osborne et al., 2022) and environmental change (Tuomainen & Candolin, 2011) and disturbance (Courbin et al., 2022). Animal

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Journal of Animal Ecology* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

movements fundamentally affect other ecological processes, including population (Hamilton & May, 1977) and community dynamics (Costa-Pereira et al., 2022), transport processes (Abbas et al., 2012), disease spread (Merkle et al., 2018) and ecosystem processes (Doughty et al., 2016). Under current global change, it is becoming increasingly important to robustly predict changes in individual movement decisions and how these scale up to emergent spatial patterns of animal distributions.

Focusing on environmental conditions and resources, habitat selection methods aim to identify and quantify the link between animal movements and distributions and the environment (Fieberg et al., 2021) and are fundamentally based on a comparison between the distribution of environmental resources and the use of these by the individuals (Manly et al., 2002). A key limitation of such approaches lies in the definition of what is available to each individual (Buskirk & Millspaugh, 2006), as animals are fundamentally limited by their movement capacities and cannot reach every point in the landscape at every single movement step, an issue which has led to long-standing discussions and methodological debates in the literature (McClean et al., 1998; Northrup et al., 2013). This crucial limitation can be solved by integrated step selection analysis (iSSA), which allows simultaneous modelling of movement and habitat selection decisions by animals (Avgar et al., 2016), building upon the earlier technique of step selection analysis (SSA) (Duchesne et al., 2015; Forester et al., 2009; Fortin et al., 2005; Rhodes et al., 2005; Thurfjell et al., 2014). Not only has this fundamental methodological advancement led to an explosion of the use of SSA and iSSA in recent years (Huggler et al., 2022; Northrup et al., 2022; Viana et al., 2018), and methodological extensions (Klappstein et al., 2021; Munden et al., 2021), but researchers have increasingly shown how the movement kernels parameterised during SSA can be 'scaled up' to predict broader-scale space use patterns (Avgar et al., 2016; Fieberg et al., 2021; Potts, Mokross, et al., 2014; Potts & Schlägel, 2020; Signer et al., 2017). Even though this markedly increases the level of understanding and the quality of predictions which can be obtained from animal movement analyses, such upscaling of SSA is seldom done by many studies using SSA, perhaps due to a lack of knowledge or perceived methodological difficulties.

The focus of this methods guide is to elucidate the various methods for scaling up from step selection to utilisation distributions and other broader space use patterns. We aim to guide the reader through various techniques used for scaling up, when they can be applied and when not, and summarising the pros and cons when more than one technique is potentially applicable. We also give example code, in both R and Python, of some simple case studies, to give the reader a practical way of getting started with these techniques.

Throughout, we will distinguish between two different types of movement models, which each require slightly different methods of analysis. The first type models the movement decisions animals make due to spatial variables that remain essentially unaffected by the animals' presence (e.g. terrain, weather). In these situations, one can use a correlative model, whereby the animal movement decision is the response variable and the unaffected spatial variable is the

explanatory variable, and propagate that model forwards through time, as exemplified by studies such as Potts, Bastille-Rousseau, et al. (2014) and Signer et al. (2017).

In the second type of movement model, there are variables that both affect an animal's movement and are affected by the same animal's presence. Examples include between-animal interactions, whereby the presence of individual A (either in the present or in the recent past) affects the movement of individual B, but in turn the movements of individual A are affected by the presence individual B (Couzin et al., 2002; Giuggioli et al., 2013). Another example is where animal movement is affected by the presence or absence of some resource that they then consume and deplete. Thus, these animals also affect the resource landscape by their presence, creating a feedback loop between animal location and landscape variable (Riotte-Lambert & Matthiopoulos, 2020), which may also be mediated by memory (Lewis et al., 2021). In all these situations, while it is possible to use correlative models to make inference about the effect of variable U on variable V (e.g. the presence of individual A on the movement of individual B), these correlative models cannot be reliably propagated forwards in time to predict broad space use patterns, since the reality is that both variables affect each other. There is no a priori sense in which one is the explanatory variable and the other the response variable. Instead, scaling up to broader spatiotemporal scales requires a dynamic modelling approach, for example via individual-based models (IBMs) of interacting individuals (Avgar et al., 2013) or systems of partial differential equations (PDEs) (Moorcroft & Lewis, 2006; Potts & Lewis, 2019).

Analysing these dynamic models is much more complicated than the case of purely correlative models, and we cannot give a complete hands-on guide here. Our approach will therefore be to provide a gateway into these techniques by explaining how to construct such IBMs and PDEs from the output of step selection, and pointing to the possible methods of analysis that might be used. In the case of IBMs, we do provide some code to help the reader get started. However, in the case of PDE models, there is a world of analytic techniques, which are standard tools for many applied mathematicians, but may not be familiar to those without an applied mathematics background. Our guide on PDE techniques will not cover that vast background itself, but rather will *guide collaborations* between ecologists and applied mathematicians by showing how to interface questions about spatial arrangements of ecological systems with mathematical tools for deriving emergent spatial phenomena. We contend that such interdisciplinary collaborations are perhaps the only way forwards for answering many important questions in spatial ecology.

The 'scaling up' procedure we will describe is outlined as follows. The first step is to parametrise a step selection function from empirical data. This step is typically called step selection analysis. A recent 'How to' paper already exists that covers the practicalities and interpretation of SSA (Fieberg et al., 2021), so we will be relatively brief here, referring the interested reader to that paper, also noting the recent review by Northrup et al. (2022). The second step involves using the parametrised step selection function to infer broad-scale

space use patterns. We will describe various existing techniques for this, which use different mathematical formalisms. Some of these techniques give exact results and some are approximate, some stochastic and some deterministic. Moreover, not all of these techniques can be used in every situation. Therefore, we will give a guide as to which technique can (or should) be used in which situations.

Finally, we will explain the sort of information one can gain from this approach. This includes (a) predicting whether or not one might expect steady state space use distributions to emerge (the alternative being those that are in perpetual flux), (b) predicting the shape of such emergent steady state distributions (e.g. stable home ranges) where they exist, and (c) using the emergent spatial patterns as a goodness-of-fit test, which can help detect missing features in step selection models. The methods presented throughout necessarily involve quite a bit of mathematical formalism. To aid the reader, we include various supplementary appendices that give specific examples of our mathematical formalisms, as well as code for implementing the methods in these simple examples, in both R and Python.

## 2 | STEP SELECTION MODELS OF ANIMAL MOVEMENT

The principle aim of SSA is to understand the drivers of animal movement. Such movements can be affected by a very wide range of possible phenomena, including food distribution, predator avoidance, social interactions, corridors to ease passage, physical barriers, topography and many more, together called *movement covariates*. All of these can potentially be revealed through a step selection approach, as long as one has the right data. SSA is a type of resource selection analysis (RSA), that is, it is a way of estimating the probability that an animal will use a given spatial area, as a function of that area's 'resource value' (e.g. access to food, mates, resting area, thermal refuge). SSA specifically focuses on the movement between two locations. It compares the observed movement step to all other locations the animal could have reached during that same time step, given the movement capacity of the animal, aiming to determine why the animal took the particular observed step instead of the many available alternatives.

With the increasingly detailed data on animal movement and its covariates that has become available over the past years, SSA has become a very popular approach for inferring drivers of movement. There have also been various variants of SSA introduced, such as iSSA (Avgar et al., 2016) and time-varying iSSA (Munden et al., 2021). However, for simplicity, we will refer to all of these as SSA unless there is a good reason to be specific. To understand how to use SSA for *inference*, there is a recent 'How to' paper, which also deals with habitat selection more broadly (Fieberg et al., 2021). Those unfamiliar with SSA may want to read this paper before continuing. Other useful papers related to inference in SSA are Thurjell et al. (2014) and Northrup et al. (2022).

Here, however, our focus is different. Instead of focusing on inference, our aim is to give a guide for how to use the output of SSA

for predicting space use patterns, given the information provided by SSA on both an animal's preferences for resources (i.e. movement covariates) and its movement behaviour. In more formal terms, we start by observing that a by-product of SSA (or iSSA) is the parametrisation of a *movement kernel*, describing the probability density,  $p_\tau(\mathbf{z}|\mathbf{x}, \alpha_x, t)$  of moving from one location,  $\mathbf{x}$ , to another location,  $\mathbf{z}$ , between times  $t$  and  $t + \tau$ , given all the information we have on the movement covariates, and given that the animal travelled to  $\mathbf{x}$  on a bearing of  $\alpha_x$ . We want to examine how to derive a *utilisation distribution (UD)* from this movement kernel, which is the probability distribution of finding the animal at any given location, and closely related to the concept of 'home range' (Börger et al., 2008). This requires setting up quite a bit of mathematical notation, but a picture of what is going on underneath these equations is given in Figure 1, and we give some foundational examples in Appendix A to help the reader become comfortable with the mathematical formalisms. We will also focus on location data that are recorded at fixed intervals at a relatively low frequency (e.g. one every few minutes or hours), but see Appendix B for modifications away from this case (following Munden et al., 2021; Potts et al., 2018).

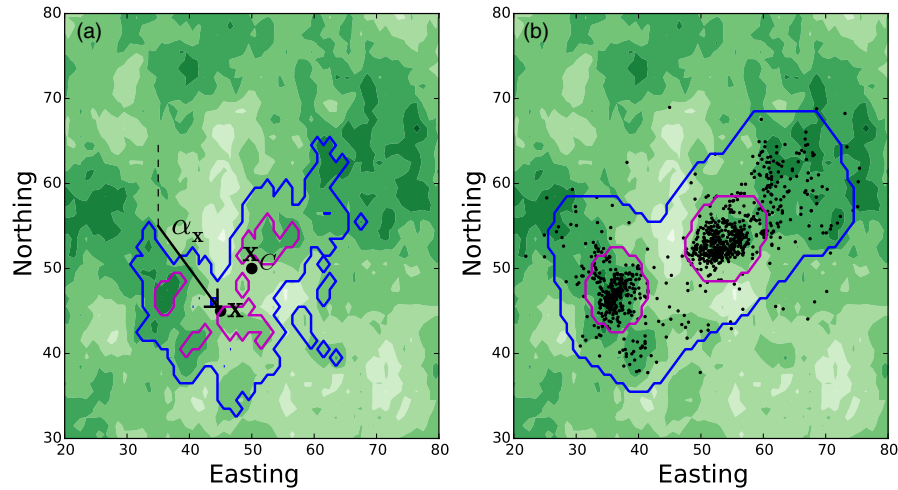
### 2.1 | The movement kernel from step selection

Right away, let us write down the general form for our movement kernel, which combines two sets of equations, one to estimate the intrinsic movement behaviour of the animals and one for estimating the selection for resources, and is written as

$$p_\tau(\mathbf{z}|\mathbf{x}, \alpha_x, t) = \frac{\phi_\tau(\mathbf{z}, \mathbf{x}, \alpha_x, t) \exp[\boldsymbol{\beta}_\tau \cdot \mathbf{Z}(\mathbf{x}, \mathbf{z}, t)]}{\int_{\Omega} \phi_\tau(\mathbf{z}, \mathbf{x}, \alpha_x, t) \exp[\boldsymbol{\beta}_\tau \cdot \mathbf{Z}(\mathbf{x}, \mathbf{z}, t)] d\mathbf{z}} \quad (1)$$

We now explain all the terms from Equation (1) in detail. First,  $\mathbf{Z}(\mathbf{x}, \mathbf{z}, t)$  is a vector of movement covariates. We write  $\mathbf{Z} = (Z_1, \dots, Z_n)$  where, for each  $i = 1, \dots, n$ , the function  $Z_i = Z_i(\mathbf{x}, \mathbf{z}, t)$  may depend upon any combination of the start location  $\mathbf{x}$ , the end location  $\mathbf{z}$ , or the current time  $t$ . However, not every  $Z_i$  need depend upon all of these aspects. For example, if the  $i$ th covariate is something constant in time over the duration of the study, for example, height above sea level, then  $Z_i$  would not depend upon  $t$ . Also,  $Z_i$  may depend upon the start location (at time  $t$ ), the end location (at time  $t + \tau$ ) or both. If  $Z_i$  depends on both  $\mathbf{x}$  and  $\mathbf{z}$ , this implicitly means it could depend upon any of the points between  $\mathbf{x}$  and  $\mathbf{z}$  (note that this does not imply that the animal has actually moved in a straight line between  $\mathbf{x}$  and  $\mathbf{z}$ , merely that it is possible to let  $Z_i$  depend on any points on this line).

Second,  $\boldsymbol{\beta}_\tau$  is a vector denoting the strength of the effect of each covariate,  $Z_i$ . We use the subscript  $\tau$  in Equation (1) to emphasise that  $\boldsymbol{\beta}_\tau$  may depend upon  $\tau$  (Fieberg et al., 2021). However, for notational convenience, we will usually drop this subscript and write  $\boldsymbol{\beta}_\tau = \boldsymbol{\beta} = (\beta_1, \dots, \beta_n)$ . The function  $\exp[\boldsymbol{\beta}_\tau \cdot \mathbf{Z}(\mathbf{x}, \mathbf{z}, t)]$  is sometimes called a *step selection function* (Fortin et al., 2005), but this term has also been used synonymously with the movement kernel (Forester et al., 2009; Potts, Bastille-Rousseau, et al., 2014). Note that we can



**FIGURE 1** From a movement kernel to a utilisation distribution. Panel (a) shows the movement kernel for a simulated animal that has a biases towards (i) better resources, denoted by darker green, (ii) a central place at  $\mathbf{x}_C$  and (iii) to continuing in the same direction (i.e. correlated movement). It has arrived at location  $\mathbf{x}$  at time  $t$  on a bearing of  $\alpha_x$  (measured clockwise from north). The subsequent location,  $\mathbf{z}$ , is sampled from the probability distribution given by the contours. Specifically, the animal has a 95% (respectively 50%) probability of being located within the blue (respectively magenta) contour at time  $t + \tau$  (i.e. after one time step). Panel (b) shows the utilisation distribution after 2000 time steps, where the blue (respectively magenta) curves show 95% (respectively 50%) kernel density estimator, mimicking how home ranges are often calculated from field data. This paper aims to explain various techniques for mathematically deriving, hence predicting, the utilisation distribution (panel b) from knowledge of the movement kernel (panel a).

expand the scalar product of  $\beta_\tau$  and  $\mathbf{Z}(\mathbf{x}, \mathbf{z}, t)$  in Equation (1) to give  $\beta_\tau \cdot \mathbf{Z}(\mathbf{x}, \mathbf{z}, t) = \beta_1 Z_1(\mathbf{x}, \mathbf{z}, t) + \dots + \beta_n Z_n(\mathbf{x}, \mathbf{z}, t)$ .

Third,  $\phi_\tau(\mathbf{z}, \mathbf{x}, \alpha_x, t)$  is a *selection-free movement kernel*, and has historically taken various functional forms, discussed in Fieberg et al. (2021, Section 3.1). The function  $\phi_\tau$  is sometimes referred to as a resource-free or resource-independent movement kernel. However, here we will use the term ‘selection-free’, as in principle  $\phi_\tau$  could itself depend upon resources and other environmental features (Avgar et al., 2016). This kernel can be thought of as describing the intrinsic ability of an animal to move in a given environment, disregarding any decisions the animal makes about where to move. In the simplest examples,  $\phi_\tau$  is just a decaying function of the speed  $|\mathbf{z} - \mathbf{x}| / \tau$ , for example,  $\phi_\tau(\mathbf{z}, \mathbf{x}, \alpha_x, t) = \exp(-|\mathbf{z} - \mathbf{x}| / \tau)$ , but it can have a more complicated dependence on the landscape and/or time (Avgar et al., 2016). Note that the possible dependence on  $\alpha_x$  allows for  $\phi_\tau$  to incorporate a distribution of turning angles, allowing for correlated movement.

Finally,  $\Omega$  is the study area and the denominator in Equation (1) ensures that the movement kernel integrates to 1 with respect to  $\mathbf{z}$ , making  $p_\tau(\mathbf{z} | \mathbf{x}, \alpha_x, t)$  a genuine probability density function. It is sometimes convenient to write

$$K(\mathbf{x}, t) = \int_{\Omega} \phi_\tau(\mathbf{z}, \mathbf{x}, \alpha_x, t) \exp[\beta_\tau \cdot \mathbf{Z}(\mathbf{x}, \mathbf{z}, t)] d\mathbf{z}. \quad (2)$$

Implicit in Equations (1) and (2) is the fact that the movement kernel is truncated so that it is not possible for an animal to move outside of  $\Omega$ , which leads to a boundary condition on any simulation of Equation (1), sometimes called a ‘no go’ condition (Potts, Bastille-Rousseau, et al., 2014). One could also choose other boundary conditions (e.g. reflective or periodic) but we will stick with this boundary

condition throughout, for simplicity. Equation (1) appears as equation (13) in Fieberg et al. (2021) in a slightly different form. We explain in Appendix A how to relate the two formulae precisely, to enable smooth linkage between both papers, alongside some examples of movement kernels with code.

## 2.2 | Incorporating animal interactions

While Equation (1) models the effect of covariates on an animal's movement, some covariates are also affected by the animal's movement. This means the two features of interest, animal locations and covariate values, feedback on one another. A key example is when an animal's movement is affected by the space use of a second animal, whose movement is, in turn, affected by the space use of the first animal. Such two-way interactions may be social, competitive, mutualistic or predator-prey. In any such case, SSA will lead to a different movement kernel for each animal, which interact with each other. This situation leads to a *system of coupled movement kernels*, sometimes called ‘coupled step selection functions’ (Lewis et al., 2021; Potts, Mokross, & Lewis, 2014). If we have  $N$  interacting animals, then the general form of such a system is

$$p_{j,\tau}(\mathbf{z} | \mathbf{x}, \alpha_x, t) = [K_j(\mathbf{x}, t)]^{-1} \phi_{j,\tau}(\mathbf{z}, \mathbf{x}, \alpha_x, t) \exp[\beta_{j,\tau} \cdot \mathbf{Z}_j(\mathbf{x}, \mathbf{z}, t)], \quad (3)$$

where  $j = 1, \dots, N$  indexes the animals.

An example system of coupled movement kernels is where each animal  $j$  has at least one covariate  $Z_{i,j}$  (for some  $i$  between 1 and  $n$ ) that denotes the recent space use of a different animal, say animal  $j'$  where  $j' \neq j$  (perhaps mediated through terrain marking or memory

[Moorcroft & Lewis, 2006; Riotte-Lambert et al., 2015]). This means that the movement of animal  $j$  is affected by the locations of animal  $j'$ . But if, in turn, animal  $j'$  has a covariate, say  $Z_{i,j'}$ , that denotes the space use of animal  $j'$ , then the two animal's movements depend on one another, generating a feedback loop between them. This is what we mean by saying the equations are 'coupled'.

In principle, Equation (3) can be parametrised using SSA in exactly the same way as Equation (1), for each individual  $j = 1, \dots, N$ . However, more needs to be said to put this into practice, as it is not a trivial task to determine the 'recent space use' of each animal. Indeed, even conceptually the idea of 'recent space use' begs questions. How recent? What constitutes 'space use'? The precise one-dimensional path the animal has travelled or some broader area demarcated by that path? How do we infer this area from data?

None of these questions have a single catch-all answer. However, an important step forward was made by Schlägel et al. (2019). Their method starts using the notion of an *occurrence distribution* (OD) to describe the 'recent space use' of an animal (Fleming et al., 2016). The OD is constructed by taking consecutive measured locations of an animal, over a user-defined time-frame, and building a theoretically optimal estimation of the distribution of actual locations. The R package `ctmm` enables users to construct this OD with just a few lines of code, as explained in Calabrese et al. (2016). However, it can take a few minutes for the package to fit the model. Furthermore, the output can be quite finely resolved, and in practice animals may respond to a more smoothed version of the OD. Therefore, it is worth users also experimenting with using either a spatial averaging of the OD, or something more intrinsically smoothed-out, like kernel density estimation (KDE; Worton, 1989) or autocorrelated KDE (AKDE; Fleming et al., 2015).

Whichever method is chosen, to use the OD for SSA simply involves considering the OD in exactly the same way as any other environmental layer that varies through time. To put this in mathematical notation, let us denote the value of the OD of animal  $j$  at time  $t$  and location  $\xi$  by  $O_{j,t}(\xi)$  (where  $\xi$  stands for either  $\mathbf{x}$  or  $\mathbf{z}$ ). If we hypothesise that this OD covaries with the decision of an animal  $j'$  to move to a location then we set  $Z_{i,j'}(\mathbf{x}, \mathbf{z}, t) = O_{j,t}(\mathbf{z})$  for some  $i$  (Schlägel et al., 2019). Alternatively, if we hypothesise that  $O_{j,t}(\xi)$  covaries with the decision of  $j'$  to move from a location then we set  $Z_{i,j'}(\mathbf{x}, \mathbf{z}, t) = O_{j,t}(\mathbf{x})$  for some  $i$ . Following this procedure for each pair of animals  $j, j' = 1, \dots, N$  leads to a collection of covariates in a system of coupled movement kernels (Equation 3), which can be parametrised using SSA (Avagar et al., 2016; Fieberg et al., 2021).

### 3 | EMERGENT PATTERNS FROM SCALING UP

We now turn to the main topic of this paper, which is how to scale up from the movement kernel of Equation (1) to a description of broad-scale space use patterns. For this, we use the concept of a *utilisation distribution* (UD), which measures the probability of finding an animal at a location  $\mathbf{x}$  at time  $t$ . We denote the UD by  $u(\mathbf{x}, t)$ .

We will sometimes be interested in the UD as it changes over time, but we will also examine situations where it is possible to derive a steady state UD,  $u_*(\mathbf{x})$ , which denotes the limit as  $t \rightarrow \infty$  of  $u(\mathbf{x}, t)$ . Mathematically, it is not always the case that  $u_*(\mathbf{x})$  exists, as  $u(\mathbf{x}, t)$  may exhibit oscillatory behaviour at long times, or more complicated spatiotemporal fluctuations (Potts & Lewis, 2019). However, where it does exist, and where this limit is approached in an ecologically relevant time frame, the UD corresponds to what is usually called a *home range*.

In this section, we will examine how to derive both exact and approximate expressions for  $u(\mathbf{x}, t)$  from Equation (1). We will explain the situations in which one can use each of these expressions, and the various benefits and drawbacks of each. In general, while exact expressions are theoretically ideal, they may be either difficult to compute in practice, not amenable to mathematical analysis, or only apply to certain subcases of Equation (1). Approximate expressions are therefore also valuable. We will explain how to calculate each of the expressions in practice and also provide a gateway into mathematical analysis. Finally, we will explain how to determine whether  $u_*(\mathbf{x})$  exists or not, and how to compute it where possible.

#### 3.1 | Mathematical formalisms for scaling up: Integro-difference equations, PDEs and IBMs

Here we describe how to use mathematical methods to derive the space use distribution of an animal, such as a home range, starting from the SSA-estimated movement kernel (combining the movement capacity of the animal and the selection for resources in the external environment). We then show how it is then possible to predict the space use distribution of an animal in a landscape with a certain distribution of resources, based on knowledge of its step-to-step movement decisions.

Perhaps the most general form linking a movement kernel to a UD is the so-called *master equation* (Merkle et al., 2017; van Kampen, 1981), which is an example of an integro-difference equation (IDE). We will begin by assuming that  $p_\tau$  is independent of  $\alpha_{\mathbf{x}}$  (i.e. no correlation in movement), so that  $p_\tau(\mathbf{z} | \mathbf{x}, \alpha_{\mathbf{x}}, t) = p_\tau(\mathbf{z} | \mathbf{x}, t)$ . The master equation in this case is

$$u(\mathbf{z}, t + \tau) = \int_{\Omega} p_\tau(\mathbf{z} | \mathbf{x}, t) u(\mathbf{x}, t) d\mathbf{x}. \quad (4)$$

Intuitively, this equation says

1. start with a UD at time  $t$ , given by  $u(\mathbf{x}, t)$ ,
2. then multiply this by the probability density of moving from  $\mathbf{z}$  to  $\mathbf{x}$  in a timestep of length  $\tau$ , given by  $p_\tau(\mathbf{z} | \mathbf{x}, t)$ ,
3. do this for all  $\mathbf{x}$  and integrate,
4. and then the result is the probability distribution at time  $t + \tau$ .

It is also possible to incorporate the aspect of correlated movement (i.e. dependence on  $\alpha_{\mathbf{x}}$ ) with some extra notational baggage. We



explain this in Appendix C, but focus here on uncorrelated movement for ease of explanation.

One approach to calculating  $u(\mathbf{x}, t)$  is to solve Equation (4) numerically. In practice, this involves discretising the study region,  $\Omega$ , and turning the integral into a sum. Let us denote by  $S$  a set of points obtained from discretising  $\Omega$ . This discretisation can be chosen by the user, but in practice it makes sense to use a grid that is related to the underlying environmental covariates, as these themselves often arrive as a discrete-space raster. Then, for any pair of grid cells,  $s$  and  $s'$ , write the probability of moving from  $s'$  to  $s$  as  $P_\tau(s|s', t)$ .

Next let  $U(s, t)$  denote the probability of finding an animal at grid point  $s$  at time  $t$ . Then, given this discretisation, Equation (4) becomes

$$U(s, t + \tau) = \sum_{s' \in S} P_\tau(s|s', t) U(s', t), \quad (5)$$

An example of calculating Equation (5) over time is given in Appendix D, together with code.

Equation (4) gives an exact solution for the time-evolution of the UD, given a movement kernel. Therefore, in principle, it gives a complete description of how to scale up from a movement kernel to a UD. Furthermore, Equation (5) gives a way of calculating Equation (4), with only some minor approximations due to the discretisation. So it might feel like the job is done. However, there are two key downsides to these equations. The first is that they are not particularly amenable to exact mathematical analysis, so there is not much exact theory that one can draw on (but see Barnett & Moorcroft, 2008, which we discuss in Section 3.2). The second is that a numerical solution can be very time-consuming. Calculating Equation (5) requires calculation of  $P_\tau(s|s', t)$  for every  $s, s' \in S$ , and  $P_\tau(s|s', t)$  itself requires computational of a numerical integral, the denominator in Equation (1). We now deal with each drawback in turn and how to mitigate against them.

First, to make use of mathematical theory, it is beneficial to derive a PDE from Equation (4), as PDEs are in general far more amenable to mathematical analysis than IDEs. However, it does require approximations to be made. Potts and Schlägel (2020) examined the PDE limit in a situation where the master equation is of the form in Equation (4). Additionally, they made two assumptions about the movement kernel in Equation (1). First,  $\phi_\tau$  must be function of only  $|\mathbf{z} - \mathbf{x}|$ , so that  $\phi_\tau(\mathbf{x}, \mathbf{z}, \alpha_x, t) = \psi_\tau(|\mathbf{x} - \mathbf{z}|)$ . Second,  $\mathbf{Z}$  must be function of only  $\mathbf{z}$  (the end of the step) and  $t$ . They also take the limit as  $\tau \rightarrow 0$ . Given these assumptions, Potts and Schlägel (2020) showed that the UD is approximately governed by the following PDE

$$\frac{\partial u}{\partial t} = \underbrace{D_\tau \nabla^2 u}_{\text{diffusive movement}} - \underbrace{2D_\tau \nabla \cdot [u \nabla (\beta_1 Z_1 + \dots + \beta_n Z_n)]}_{\text{drift up the gradient of } \beta_1 Z_1 + \dots + \beta_n Z_n}, \quad (6)$$

where  $D_\tau$  is the diffusion constant, calculated as

$$D_\tau = \frac{1}{4\tau} \int_{\mathbb{R}^2} |\mathbf{x}'|^2 \psi_\tau(|\mathbf{x}'|) d\mathbf{x}', \quad (7)$$

$\mathbf{x}'$  is a dummy variable, and  $|\mathbf{x}'|$  is the length of the vector  $\mathbf{x}'$ . The  $\tau \rightarrow 0$  approximation means that any directional autocorrelation in the data is ignored. Often, this is a perfectly adequate assumption for predicting broad-scale space use patterns (e.g. the probability density function of a correlated random walk is asymptotically the same as a Brownian motion). However, if a user wants to use Equation (6) to make quantitative predictions for a particular study animal, we recommend they truth-test the PDE approximation by running an IBM that includes any directional autocorrelation detected in their data. Finally, from the perspective of calculating  $u(\mathbf{x}, t)$  numerically, there is in our experience no advantage to using this PDE over the master equation. However, the analytic tools for studying PDEs are vast and can give crucial qualitative information about space use, which we will discuss more in Section 4.

As an alternative to numerical solutions of either the PDE in Equation (6) or the IDE in Equation (4), a conceptually simple approach is to use a stochastic IBM (Avagar et al., 2016; Potts, Börger, et al., 2022; Signer et al., 2017). This simply involves simulating stochastic realisations of the movement kernel in Equation (1) and can be done either directly (Appendix A) or via the amt package in R. The amt approach is shown, with examples, in Signer et al. (2019). Here, we only want to add one word of caution, which is that the UD described in Signer et al. (2019), and calculated in amt, is in fact

$$\text{amt - UD} = \int_0^t u(\mathbf{x}, t') dt', \quad (8)$$

which is the cumulative UD up to time  $t$ . If the UD reaches a steady state,  $u_*(\mathbf{x})$ , then the two definitions of UD coincide, but this is not true for transient UD's. Therefore, anyone comparing transient UD's from amt with those calculated from the IDE and PDE methods described here needs to take this discrepancy into account. In the context of home range calculations, the amt method is closer to what is done in practice when measuring home ranges from data, as typically people will measure home ranges using highly autocorrelated movement data. However, conceptually, a home range is usually thought of as reflecting the probability density function of an animal's locations, which is more accurately described by  $u(\mathbf{x}, t)$ .

While IBMs are conceptually simple, it can be time-consuming to use them for calculating  $u(\mathbf{x}, t)$ . For this, it is necessary to simulate the movement kernel up to time  $t$  sufficiently many times to obtain a good measure of the probability distribution, which is the distribution of locations at time  $t$  across all simulation runs. The number of simulations scales linearly with the number of lattice sites in the study area (which could be cells of a rasterised landscape, for example), and usually one would need dozens or hundreds of simulations per lattice site to obtain a reasonable estimate of the probability distribution. To circumvent this, amt estimates Equation (8) instead of  $u(\mathbf{x}, t)$ , and in so doing makes use of every point the simulated animal visits, not just the point at time  $t$  for each simulation. However, information is lost by estimating Equation (8) rather than  $u(\mathbf{x}, t)$ . Particularly, if  $u(\mathbf{x}, t)$  fluctuates over time then these fluctuations are averaged-out in the calculation of Equation (8), so are not properly captured by the amt method.

If you do want to calculate  $u(\mathbf{x}, t)$  using IBMs but without requiring prohibitively intensive simulations, it is instead possible to combine simulations with a smoothing method, like KDE (Potts, Mokross, et al., 2014). To do this, simulate the movement kernel perhaps a few hundred times, each time starting at the same location and running the simulation to time  $t$  (this much can be done in `amt`). Then take the end-point of each simulation to give a dataset of genuinely independent samples of  $u(\mathbf{x}, t)$ . Finally, construct the KDE of this dataset, which is an estimation of  $u(\mathbf{x}, t)$ .

In all these methods, if comparing  $u(\mathbf{x}, t)$  to data, it is advisable to ensure that the value of  $u(\mathbf{x}, t)$  is as small as possible at the boundary of the study area,  $\Omega$ , if at all possible. This avoids any bias that might be caused by the effect of the 'no go' boundary conditions (see sentences after Equation (2)).

### 3.2 | The steady state UD in the absence of animal interactions

One of the most oft-studied emergent spatial patterns from animal movement data is the home range, that is, the space use distribution often observed in nature, where animals restrict their movements over time to a certain area in space, and do not roam over the entire available landscape (Börger et al., 2008). Mathematically, this can be thought of as the *steady state* of a UD, defined to be a configuration that does not change over time. As such, if  $u(\mathbf{x}, t)$  satisfies Equation (4), then the steady state of  $u(\mathbf{x}, t)$  is a function,  $u_*(\mathbf{x})$ , satisfying the following equation

$$u_*(\mathbf{z}) = \int_{\Omega} p_{\tau}(\mathbf{z}|\mathbf{x})u_*(\mathbf{x})d\mathbf{x}, \quad (9)$$

if such a function exists. Here we have to assume that  $p_{\tau}(\mathbf{z}|\mathbf{x}) = p_{\tau}(\mathbf{z}|\mathbf{x}, t)$ , that is, the movement kernel does not change over time, otherwise no steady state can exist. That said, ecological systems can exhibit multiple time-scales; for example a movement kernel may give rise to a UD that becomes close to a steady state over a season, but next season the animal's movement may change, causing the UD to change (Börger et al., 2008). In this case, we can assume that  $p_{\tau}(\mathbf{z}|\mathbf{x}) = p_{\tau}(\mathbf{z}|\mathbf{x}, t)$  for the duration of a season and use the above techniques to calculate the seasonal steady state; but over the course of a year, the UD fluctuates.

In this section, we will assume that  $p_{\tau}(\mathbf{z}|\mathbf{x})$  is independent of  $u(\mathbf{x}, t)$ . That is, the probability of moving to a specific location does not depend upon the past or present UD. This means that Equation (9) is *linear* in  $u_*(\mathbf{x})$ , which is a requirement for the techniques presented in this section. Notice that this linearity requirement precludes the case where we have multiple coupled movement kernels, like in Equation (3) (whereby the movement of one animal depends on the UD of another, whose movement depends on the UD of the first animal). We will return to this case in Section 3.3. The linearity requirement also precludes memory effects whereby animals are attracted to their own UD.

Let us now write the discrete space version of Equation (9), following the notation of Equation (5), as

$$U_*(s) = \sum_{s' \in S} P_{\tau}(s|s')U_*(s'). \quad (10)$$

As described in Section 3.1, Equation (10) is what we tend to calculate in practice. This equation can, in theory, be calculated exactly using matrix inversion (Appendix E) but to our knowledge this has never been done in the context of step selection, perhaps due to computational intensity.

Another exact method, which is also relatively computationally efficient, is that of Barnett and Moorcroft (2008). This, however, relies on two assumptions. The first is that  $\phi_{\tau}(\mathbf{z}, \mathbf{x}, \alpha_{\mathbf{x}}, t)$  can be written as a function of  $|\mathbf{x} - \mathbf{z}|$ , that is,  $\phi_{\tau}(\mathbf{z}, \mathbf{x}, \alpha_{\mathbf{x}}, t) = \psi_{\tau}(|\mathbf{x} - \mathbf{z}|)$ . The second is that the functions  $Z_i(\mathbf{x}, \mathbf{z}, t)$  only depend upon the end-point of the step, that is,  $Z_i(\mathbf{x}, \mathbf{z}, t) = \tilde{Z}_i(\mathbf{z})$  for some function  $\tilde{Z}_i(\mathbf{z})$ . With these assumptions in place, Barnett and Moorcroft (2008) show that the following exact expression for  $u_*(\mathbf{x})$  holds

$$u_*(\mathbf{x}) = \frac{\exp(\beta \cdot \tilde{\mathbf{Z}}(\mathbf{x})) \int_{\Omega} \exp(\beta \cdot \tilde{\mathbf{Z}}(\mathbf{z})) \psi_{\tau}(|\mathbf{x} - \mathbf{z}|) d\mathbf{z}}{\int_{\Omega} [\exp(\beta \cdot \tilde{\mathbf{Z}}(\mathbf{x})) \int_{\Omega} \exp(\beta \cdot \tilde{\mathbf{Z}}(\mathbf{z})) \psi_{\tau}(|\mathbf{x} - \mathbf{z}|) d\mathbf{z}] d\mathbf{x}}. \quad (11)$$

We show how to compute examples of this, with code, in Appendix F.

It is interesting to look at two limiting cases. First, if  $\psi_{\tau}$  is a uniform distribution, meaning that animals can move over the whole of  $\Omega$  in a single timestep, then (Appendix F)

$$u_*(\mathbf{x}) = \frac{\exp(\beta \cdot \tilde{\mathbf{Z}}(\mathbf{x}))}{\int_{\Omega} \exp(\beta \cdot \tilde{\mathbf{Z}}(\mathbf{x}))}, \quad (12)$$

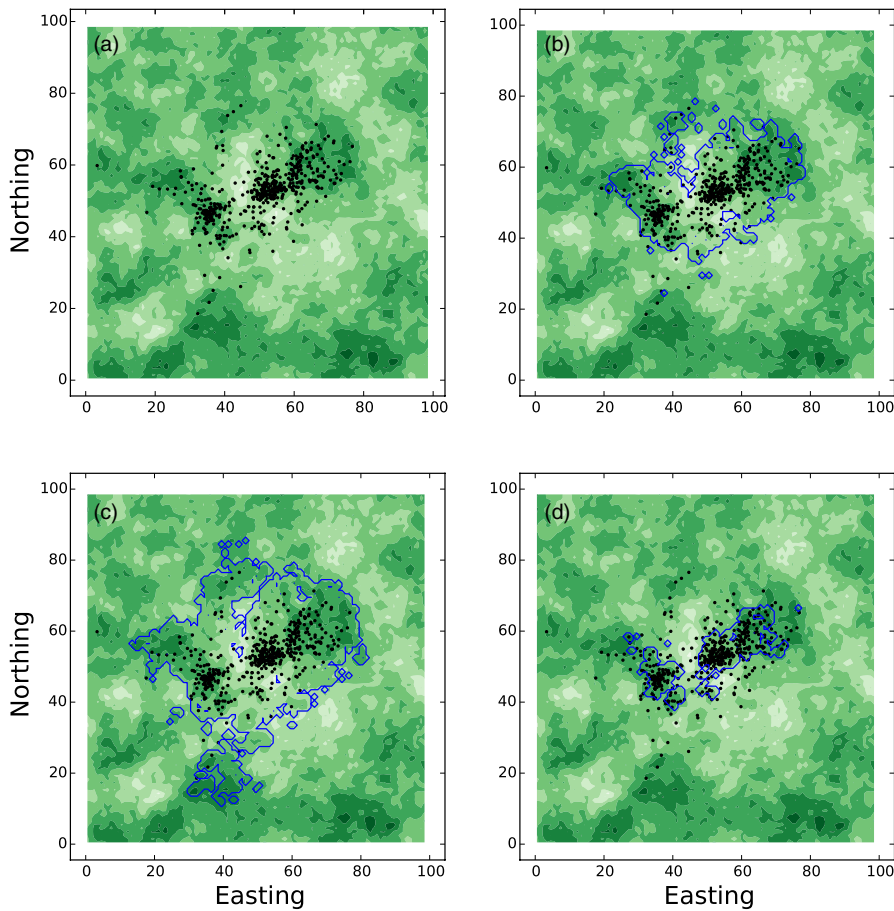
which is just the usual expression for a resource selection function (Manly et al., 2002). Although this assumption on  $\psi_{\tau}(l)$  is quite restrictive, there are real examples. For example, an urban fox can often traverse its whole territory in just a few minutes (Potts et al., 2013), so if  $\Omega$  were the territory of an urban fox then it makes sense to use a uniform distribution for  $\psi_{\tau}$ .

The other extreme is where  $\psi_{\tau}$  is arbitrarily narrow, so the animal is making distinct movement choices over much smaller spatial scales than  $\Omega$ , as is often the case with animals with very large home ranges. In this case, we have (Appendix F)

$$u_*(\mathbf{x}) = \frac{\exp(2\beta \cdot \tilde{\mathbf{Z}}(\mathbf{x}))}{\int_{\Omega} \exp(2\beta \cdot \tilde{\mathbf{Z}}(\mathbf{x}))}, \quad (13)$$

which is identical to Equation (12) but with  $\beta$  switched for  $2\beta$ . In other words, the effect of selection on space use doubles as one moves from selection on a very broad spatial scale to a very narrow spatial scale (Moorcroft & Barnett, 2008). Notice that this factor of 2 also appears in the PDE from Equation (6) (before  $D_{\tau}$ ); indeed, the steady state of Equation (6) is precisely Equation (13) (Potts & Schlägel, 2020).

Figure 2 gives an example of the steady state UD estimations from Equations (11)–(13) for a movement kernel of the following type



**FIGURE 2** Steady state utilisation distributions (UDs). Panel (a) shows locations of a simulated animal with movement process corresponding to Equation (14) with the resource layer shown in the background (darker green means higher quality resources). Panels (b–d) overlay this with predicted UD from Equations (11)–(13), respectively, shown as a contour line surrounding the 95% kernel of the UD estimation. Panel (b) is an exact steady state solution to Equation (8), so captures the space use well. Panels (c), proportional to  $\exp(\beta \cdot \tilde{Z}(\mathbf{x}))$ , and panel (d), proportional to  $\exp(2\beta \cdot \tilde{Z}(\mathbf{x}))$ , respectively over- and under-estimate space use. Here,  $\lambda = 0.2$ ,  $\beta_C = 0.2$ , and  $\beta_R = 1.5$ ,  $\mathbf{x}_C = (50, 50)$ .

$$p_z(\mathbf{z}, \mathbf{x}, \alpha_x | t) = K_j(\mathbf{x}, t)^{-1} \exp(-\lambda |\mathbf{z} - \mathbf{x}|) \exp(\beta_R R(\mathbf{z}) - \beta_C |\mathbf{x} - \mathbf{x}_C|), \quad (14)$$

where  $R(\mathbf{z})$  is a resource layer and  $\mathbf{x}_C$  is a localising point (e.g. a den or nest site). This models movement in a heterogeneous environment where there is additionally some localising tendency towards a single point, such as a den or nest site. Observe from Figure 2 that Equation (12) overestimates the UD size but Equation (13) is an underestimation. We give instructions and example code for reproducing Figure 2 and calculating Equations (11)–(13) in Appendix F.

### 3.3 | UD for interacting animals

In Section 3.2, we examined how to find the steady state distribution in situations where the covariates  $Z_i$  are not affected by the locations of the animal or animals. In other words, the causality goes one way: covariates affect animal locations, but are not affected by those locations. This works fine in many classical examples of step selection, where the focus is on things such as presence of food, ease of motion on the terrain, proximity to locations of interest and so forth. However, in reality, there are many situations where movement covariates are affected by present or past animal locations. Examples include memory effects (Merkle et al., 2017), resource depletion (Riotte-Lambert et al., 2015), social interactions (Moorcroft & Lewis, 2006), competition (Vanak et al., 2013), prey-taxis (Kareiva &

Odell, 1987) and predator avoidance (Bastille-Rousseau et al., 2015), which are all well-established ecological phenomena.

In such situations, the techniques of Section 3.2 do not directly apply. Indeed, there is no guarantee that a steady state emerges at all, and one may instead find UD that fluctuate in perpetuity, never settling (Potts, Giunta, et al., 2022). To understand these patterns, there are two broad approaches: numerical and analytic. For the numerical approach, one could use the IDE formalism from Equation (1) or the PDE of Equation (6). However, we recommend using an IBM instead. The principal reason for this is that IDEs and PDEs only keep track of the probability distribution of animal locations, but IBMs keep track of the actual location of animals (Wang & Grimm, 2007). This has two advantages. First, if performing numerical simulations, one might as well keep as much realism in them as possible (i.e. why not use an IBM?). Second, animals will respond to the actual (past and present) locations of themselves and other animals, not a distribution that reflects the probability of all possible locations that each animal could have taken. For analytic approaches, PDEs are the best tool and we will discuss this more in Section 4.

Writing code for an IBM depends a lot on the specific situation that is being modelled, especially for interacting objects. We give a basic example in Appendix G, of animals that have a mutual avoidance tendency and attraction to a single static resource layer, to help the uninitiated get started. However, we caution the reader that construction of an IBM for their specific situation is likely to require



quite significant thought and modification/re-writing of this simple model, which will depend on the particular structure of Equation (1).

### 3.4 | Goodness-of-fit analysis from emergent patterns

Having described tools for ascertaining emergent spatial patterns from models parametrised at a finer spatiotemporal scale, we now turn our attention to what we can learn from this process. The principle aim is to examine the extent to which these fine-scale processes capture the observed broad-scale patterns. This process can help reveal missing features from the model and generate new hypotheses (Potts, Börger, et al., 2022).

An example of this is shown in Figure 3 using simulated data. In this example, we assume the following movement kernel describes the movement of animal  $j$ , for  $j \in \{1, 2, 3, 4\}$

$$p_{j,\tau}(\mathbf{z}, \mathbf{x}, \alpha_{\mathbf{x}}|t) = K_j(\mathbf{x}, t)^{-1} \exp(-\lambda|\mathbf{z} - \mathbf{x}|) \exp\left(\beta_{j,R}R(\mathbf{z}) - \beta_{j,C}|\mathbf{x} - \mathbf{x}_{j,C}| - \sum_{j' \neq j} \beta_{j,j'} O_{j'}(\mathbf{z}, t)\right), \quad (15)$$

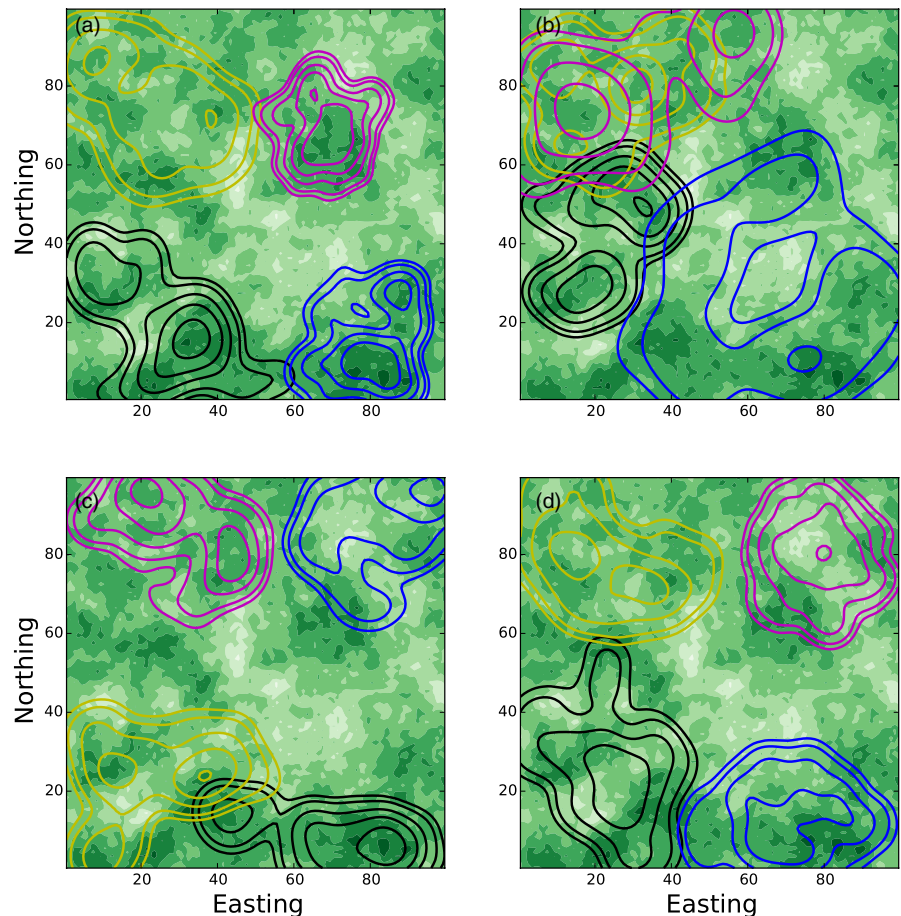
where  $R(\mathbf{z})$  is a resource layer,  $\mathbf{x}_{j,C}$  is a localising point for animal  $j$  and  $O_{j'}(\mathbf{z}, t)$  is the OD of animal  $j'$  at time  $t$ .

Figure 3a shows the KDE distributions of each simulated animal, moving according to an IBM based on the movement kernel

in Equation (15). Figure 3b–d shows the distributions of simulated animals where one of the covariates is missing. There are some technical considerations when constructing an IBM of animals that interact through their OD (Potts, Börger, et al., 2022). In short, one needs to think about how to construct  $O_{j'}(\mathbf{z}, t)$  at each step of the simulation, which involves not just the locations where the animal makes a turn but also locations in between turns. One way to deal with this is to simulate a stepping-stone process (Avgar et al., 2013, 2016). Details of how we have constructed such a stepping-stone process from an example system of coupled movement kernels are given in Appendix G.

Comparing panels in Figure 3 reveals that a failure to incorporate social interactions (i.e.  $\beta_{j,j'} = 0$ ) leads to more overlap between home ranges than is actually the case (compare panels (a) and (b)), a failure to incorporate localising tendency (i.e.  $\beta_{j,C} = 0$ ) leads to UDs emerging in the wrong place (panels (a) and (c)), and a failure to include the resource layer (i.e.  $\beta_{j,R} = 0$ ) leads to UDs that fail to grasp the environmental heterogeneity in  $R(\mathbf{z})$  (panels (a) and (d)). This is perhaps quite obvious in the omniscient situation of simulated data. However, if in a real situation using empirical data, a researcher has not realised about one of these features, and then has parametrised a model that does not include that feature, then comparing empirical data on space use to emergent patterns from simulations of that model can help reveal this missing feature (Potts, Börger, et al., 2022).

**FIGURE 3** The effect of missing covariates. Empirically parametrised individual-based models (IBMs) can be used to detect missing features from a step selection model. Panel (a) shows some simulated data from four animals in a  $100 \times 100$  box, whose movement is governed by three things: Mutual avoidance, attraction to resources and central place attraction. The green shades represent the resource layer (darker green implies better resources), the contours give utilisation distribution (UD) of animal locations, one colour for each animal (contours at heights 0.0001, 0.0002, 0.0005, 0.001, 0.002 from the outside in). Panel (b) shows the UD when mutual avoidance is removed from the model. Panel (c) has central place attraction removed. Panel (d) does not include resources. If researchers have parametrised an SSF model but missed a key covariate governing space use then they may be able to gain insight into what that missing covariate is by comparing simulated and empirical UDs, in the same way as we might compare panel (a) with panels (b–d).



As well as visual examination of discrepancies between data and IBM output, various metrics can be calculated to assess goodness-of-fit. Two possible metrics are to (i) compare the UD or OD sizes between the data and the IBM output, and (ii) to measure the UD or OD overlap between data and IBM. UD size can be measured using any number of metrics, but the locational variance is perhaps the simplest, as it is proportional to standard measures, like 95% KDE, but does not require interpolation or smoothing. Following Fieberg & Kochanny (2005), we recommend using Bhattacharyya's Affinity to measure UD overlap. Details of all these methods are given in Potts, Börger, et al. (2022), which also mentions relations to existing goodness-of-fit test for SSA that do not examine emergent spatial patterns (Fieberg et al., 2018).

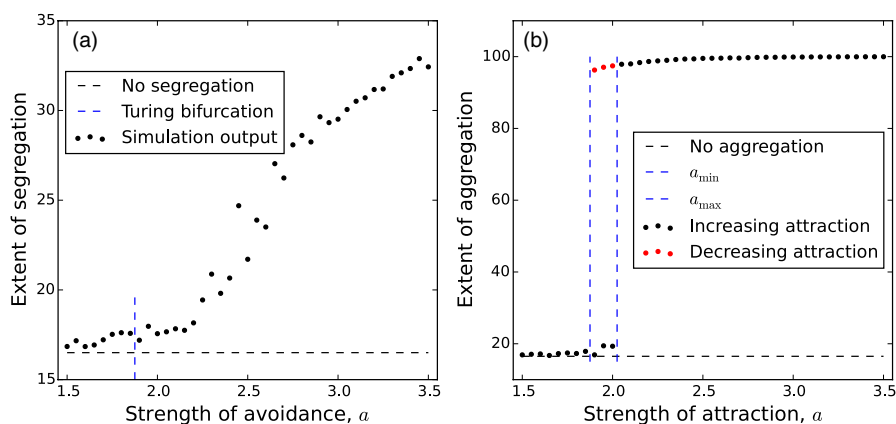
#### 4 | EXPLORING EMERGENT PATTERNS

While IBMs are valuable for comparing model output with data (Section 3.4), they are not always so amenable to mathematical analysis. This is where PDEs come into their own. There is a wealth of techniques for analysing PDEs in the applied mathematics literature (Buttenschön & Hillen, 2021; Evans, 2022; Murray, 2003; Robinson & Pierre, 2003). Many of these are quite technically demanding, and so our best 'how to' suggestion for those who do not have a deep background in applied mathematics is to form collaborations with those who do. The trick for forming such collaborations is to know broadly the sort of questions that can be answered by mathematical techniques and how to phrase them in the language of applied maths in a way that might entice collaborators while keeping firmly grounded in ecological and natural history knowledge. Our philosophy here will be to try to explain how to do this, with the ultimate aim of helping readers form useful collaborations, rather than doing the mathematics themselves.

Perhaps the most elementary technique in pattern formation analysis of PDEs is linear stability analysis (LSA; also sometimes called Turing pattern analysis, after Turing, 1952). This technique asks the following question: if a system is homogeneous (in our case, this means each animal is equally likely to be found anywhere on the whole landscape) and is then perturbed slightly (which will happen naturally as animals move and interact), will those perturbations grow in time? In practice, this means that UDs will segregate or aggregate spontaneously. Therefore, it can be used to answer questions such as whether avoidance processes are sufficiently strong to cause territorial segregation (Potts & Lewis, 2016), or whether attraction processes are sufficient to enable aggregations to emerge spontaneously (Briscoe et al., 2002). Such analysis may also help researchers to separate-out the effect of social interactions on spatial distribution patterns from environmental interactions.

A second question that can be answered by LSA is: as the patterns grow from a homogeneous state, will they oscillate? This means that any segregations or aggregations that emerge will not be stationary but move around in space. This is of key importance in measuring UDs from data, because if a collection of animals have movement processes that lead to perpetually fluctuating space use patterns, then this has to be taken into account when measuring UDs from data. For this, one has to consider a set of locations across a time window. The size of this window should be determined by the natural period of any emergent oscillatory patterns.

The next question, which requires tools beyond LSA, is whether patterns are likely to form suddenly as parameters change. This means that a small environmental perturbation might give rise to a dramatic change in the structure of UDs (i.e. a 'tipping point'). Figure 4b gives an example of this (in the context of IBMs) whereby an increasing tendency for attraction to conspecifics leads to a sudden switch in spatial distribution from homogeneous to highly aggregated (which could, e.g., be driven by increased fear of predation).



**FIGURE 4** Transition from homogeneity to heterogeneous patterns. Panel (a) (respectively panel (b)) shows simulation output from an stochastic IBM consisting of two mutually avoiding (respectively mutually attracting) populations. When the strength of avoidance (respectively attraction) is low, the populations are well mixed, indistinguishable from non-interacting populations. As this strength is increased past the Turing bifurcation point, the populations begin to segregate (respectively aggregate). While the segregation patterns in panel (a) emerge in a continuous fashion, the aggregation patterns in panel (b) appear quite suddenly, and a hysteresis effect is observed. The precise definitions of the quantities in these plots are given in Appendix H.

The existence of these sorts of tipping points can be ascertained by a variety of techniques, perhaps the most well used of which is *weakly nonlinear analysis* (others include *Crandall–Rabinowitz bifurcation theory* and *centre manifold theory*, and different mathematicians have different tastes regarding which to use, so it is valuable to be aware of the nomenclature). These techniques can determine whether the point at which patterns start to form (known as the *bifurcation point*) is *supercritical*, meaning that the size of the patterns is continuously dependent on the underlying process (Figure 4a), or *subcritical*, meaning there is a discontinuous switch from no patterns to patterns (Figure 4b).

The subcritical case is often accompanied by a hysteresis phenomenon, whereby the existence of patterns depends upon the history of the system. For example, in Figure 4b, in the range  $a_{\min} < a < a_{\max}$ , it is possible to see either aggregations or homogeneous UD depending on the history of the system. A real life example might be of a population of herbivores in the presence of a predation risk that is initially mild, but grows steadily, causing them to increase their tendency to move towards one another for safety (given by the parameter  $a$ ). At some point, the predation risk becomes sufficiently high that  $a > a_{\max}$  and an aggregation forms. Suppose that after some time the predation risk stops increasing and instead starts to decrease (perhaps due to human intervention or predator disease). Then  $a$  decreases. However, when  $a$  decreases past  $a_{\max}$ , the aggregations do not yet collapse. Indeed, not until  $a < a_{\min}$  does this collapse happen and the herbivores return to their original, homogeneously spread state.

While the mathematical tools of PDE analysis enable rigorous quantification of UD pattern formation properties, the downside is that various approximations are made when moving from the movement kernel of Equation (1) to the PDE of Equation (6). It is therefore valuable to check that the pattern formation properties observed in PDE analysis are also observed in the more realistic case of an IBM. Recent research has begun to develop techniques for doing this, tailored to the specific case of understanding emergent space use patterns from animal movement processes (Potts, Giunta, et al., 2022). This research shows both how to relate IBMs to PDEs in a rigorous fashion, and gives methods for determining whether patterns emerge, whether they are stationary or fluctuating, aggregative or segregative. Indeed, Figure 4 gives an example of the output of such techniques. In particular, Figure 4a shows the analytically computed Turing bifurcation point of the PDE system, which is close to where the IBMs bifurcate from homogeneous to heterogeneous patterns.

## 5 | DISCUSSION

### 5.1 | Why scale up?

We have described various existing techniques for scaling up from step selection to broader-scale space use patterns, some of which require some relatively involved mathematical and/or numerical

analysis. So what is the value in learning and using these techniques? We highlight two key points.

1. *Prediction in systems with feedbacks*. As shown in Sections 2.2 and 3.3, if there are two or more variables that each affect one another so that there is no a priori demarcation into explanatory and response variables, then correlative models alone are insufficient for making predictions (including RSA and many species distribution models [SDMs]). Instead one needs a dynamic model. Step selection provides a technique for parametrising such models and the methods described here provide techniques for analysing their emergent features. An example of this might be predicting the effect of rewilding strategies on ecosystem restoration. For example, domestic herbivores may be allowed to roam more freely, leading to a more heterogeneous vegetation layer, which, in turn, affects their movements and distributions.
2. *Testing for missing covariates in movement models*. SSA can demonstrate which of a predetermined set of variables covary with movement. However, it cannot ascertain whether the user has focused on the correct set of variables for describing the animal's movement. By propagating the resulting movement kernel forwards in time, we can discover the extent to which it can predict longer-term patterns. Any discrepancy between predictions and data can be used to inform further model development, as in Section 3.4, and also where best to concentrate data gathering efforts.

Underlying both of these is a conceptual move from uncovering predictors to building predictive models. Correlative models, such as SSA, RSA and SDMs, are well developed for uncovering predictors, but they are less developed regarding making actual predictions. Accurate model predictions require that the underlying models both capture the dynamics correctly (Point 1 above) and contain all the necessary mechanisms required for accurate predictions (Point 2 above). Prediction in spatial animal ecology is notoriously tricky (Hao et al., 2019) and requires significant future development. However, by plugging these two conceptual gaps, the methods described here should provide an important step in improving the predictive power of animal space use models.

### 5.2 | Future directions

Building models of animal movement that are able to predict broad-scale space use patterns is a fundamental goal of movement ecology, with a huge range of potential applications to conservation and management (see Section 1). However, there is still a long way to go before accurate predictions from fine-grained movement models become widely possible. We end with a few ideas for where we would like to see these methods going.

1. *Application across taxa and geography*. Many of the methods described here are relatively new and have yet to be applied

in earnest to a wide range of datasets. Broad application across different data and research groups is perhaps the best way to ascertain the value of these methods and discover practical ideas for improvement. Ideally, these would include different taxa, populations in contrasting environments, and data sampled over different time-scale and intensities.

2. *Biologging data.* Alongside movement, modern datasets often contain a wealth of biologging data, such as heart rate, acceleration, body temperature and neurological sensors (Williams et al., 2020). These can help inform the behavioural mode of animals, which, in turn, affects how they move and use space. Methods to incorporate these into movement models, ascertaining the extent to which they feed up to broad space use patterns, is a key research frontier (Klappstein et al., 2021).
3. *Continuous time formulations.* Animals move in continuous time, they may make decisions at any point in time, data may be gathered at completely different points in time, all of which makes a continuous time framework appealing (Parton et al., 2016). In Appendix I, we discuss some current efforts to this end, and some possible ways forward. It is also worth mentioning that methods from Sections 3.4 and 4 are not implicitly tied to a discrete time framework, so it would be valuable to examine how to use these in the context of continuous time models.
4. *Mathematical analysis of emergent phenomena.* Our mathematical understanding of emergent phenomena from moving, interacting populations is still in relative infancy (Eftimie, 2018; Potts & Lewis, 2019). Specifically, efforts are required that explicitly tie these into empirically measured movement processes, for example using methods like those in Moorcroft and Lewis (2006) or Potts and Schlägel (2020). This will require strong collaborations between ecologists and applied mathematicians.

#### AUTHOR CONTRIBUTIONS

Jonathan R. Potts and Luca Börger developed the idea for this Research Methods Guide. Jonathan R. Potts led the writing of the manuscript. Both authors contributed critically to the drafts and gave final approval for publication.

#### ACKNOWLEDGEMENTS

JRP acknowledges support of Engineering and Physical Sciences Research Council (EPSRC) grant EP/V002988/1. We thank Roberto Salguero-Gomez for encouraging us to write this guide and for valuable comments during the process of writing. We also thank Brian J. Smith and an anonymous reviewer for helpful comments.

#### CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

#### DATA AVAILABILITY STATEMENT

Code available from the Zenodo repository <https://doi.org/10.5281/zenodo.7225910> (Potts & Börger, 2022).

#### ORCID

Jonathan R. Potts  <https://orcid.org/0000-0002-8564-2904>

Luca Börger  <https://orcid.org/0000-0001-8763-5997>

#### REFERENCES

- Abbas, F., Merlet, J., Morellet, N., Verheyden, H., Hewison, A. M., Cargnelutti, B., Angibault, J., Picot, D., Rames, J., Lourtet, B., Aulagnier, S., & Daufresne, T. (2012). Roe deer may markedly alter forest nitrogen and phosphorus budgets across Europe. *Oikos*, *121*, 1271–1278.
- Avgar, T., Deardon, R., & Fryxell, J. M. (2013). An empirically parameterized individual based model of animal movement, perception, and memory. *Ecological Modelling*, *251*, 158–172.
- Avgar, T., Potts, J. R., Lewis, M. A., & Boyce, M. S. (2016). Integrated step selection analysis: Bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution*, *7*, 619–630.
- Barnett, A., & Moorcroft, P. (2008). Analytic steady-state space use patterns and rapid computations in mechanistic home range analysis. *Journal of Mathematical Biology*, *57*, 139–159.
- Bastille-Rousseau, G., Potts, J. R., Schaefer, J. A., Lewis, M. A., Ellington, E. H., Rayl, N. D., Mahoney, S. P., & Murray, D. L. (2015). Unveiling trade-offs in resource selection of migratory caribou using a mechanistic movement model of availability. *Ecography*, *38*, 1049–1059.
- Börger, L., Dalziel, B. D., & Fryxell, J. M. (2008). Are there general mechanisms of animal home range behaviour? A review and prospects for future research. *Ecology Letters*, *11*, 637–650.
- Briscoe, B., Lewis, M., & Parrish, S. (2002). Home range formation in wolves due to scent marking. *Bulletin of Mathematical Biology*, *64*, 261–284.
- Buskirk, S. W., & Millspaugh, J. J. (2006). Metrics for studies of resource selection. *The Journal of Wildlife Management*, *70*, 358–366.
- Buttenschön, A., & Hillen, T. (2021). *Non-local cell adhesion models*. Springer.
- Calabrese, J. M., Fleming, C. H., & Gurarie, E. (2016). Ctm: An R package for analyzing animal relocation data as a continuous-time stochastic process. *Methods in Ecology and Evolution*, *7*, 1124–1132.
- Costa-Pereira, R., Moll, R. J., Jesmer, B. R., & Jetz, W. (2022). Animal tracking moves community ecology: Opportunities and challenges. *Journal of Animal Ecology*, *91*, 1334–1344.
- Courbin, N., Pichegru, L., Seakamela, M., Makhado, A., Meyer, M., Kotze, P. G., Mc Cue, S. A., Péron, C., & Grémillet, D. (2022). Seascapes of fear and competition shape regional seabird movement ecology. *Communications Biology*, *5*, 1–7.
- Couzin, I. D., Krause, J., James, R., Ruxton, G. D., & Franks, N. R. (2002). Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, *218*, 1–11.
- Doughty, C. E., Roman, J., Faurby, S., Wolf, A., Haque, A., Bakker, E. S., Malhi, Y., Dunning, J. B., Jr., & Svenning, J. C. (2016). Global nutrient transport in a world of giants. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 868–873.
- Duchesne, T., Fortin, D., & Rivest, L. P. (2015). Equivalence between step selection functions and biased correlated random walks for statistical inference on animal movement. *PLoS ONE*, *10*, e0122947.
- Eftimie, R. (2018). *Hyperbolic and kinetic models for self-organised biological aggregations: A modelling and pattern formation approach*. Springer.
- Evans, L. C. (2022). *Partial differential equations* (Vol. 19). American Mathematical Society.
- Fieberg, J., & Kochanny, C. O. (2005). Quantifying home-range overlap: The importance of the utilization distribution. *The Journal of Wildlife Management*, *69*, 1346–1359.
- Fieberg, J., Signer, J., Smith, B., & Avgar, T. (2021). A 'how to' guide for interpreting parameters in habitat-selection analyses. *Journal of Animal Ecology*, *90*, 1027–1043.



- Fieberg, J. R., Forester, J. D., Street, G. M., Johnson, D. H., ArchMiller, A. A., & Matthiopoulos, J. (2018). Used-habitat calibration plots: A new procedure for validating species distribution, resource selection, and step-selection models. *Ecography*, *41*, 737–752.
- Fleming, C. H., Fagan, W. F., Mueller, T., Olson, K. A., Leimgruber, P., & Calabrese, J. M. (2015). Rigorous home range estimation with movement data: A new autocorrelated kernel density estimator. *Ecology*, *96*, 1182–1188.
- Fleming, C. H., Fagan, W. F., Mueller, T., Olson, K. A., Leimgruber, P., & Calabrese, J. M. (2016). Estimating where and how animals travel: An optimal framework for path reconstruction from autocorrelated tracking data. *Ecology*, *97*, 576–582.
- Forester, J., Im, H., & Rathouz, P. (2009). Accounting for animal movement in estimation of resource selection functions: Sampling and data analysis. *Ecology*, *90*, 3554–3565.
- Fortin, D., Beyer, H., Boyce, M., Smith, D., Duchesne, T., & Mao, J. (2005). Wolves influence elk movements: Behavior shapes a trophic cascade in yellowstone national park. *Ecology*, *86*, 1320–1330.
- Fryxell, J. M., Hazell, M., Börger, L., Dalziel, B. D., Haydon, D. T., Morales, J. M., McIntosh, T., & Rosatte, R. C. (2008). Multiple movement modes by large herbivores at multiple spatiotemporal scales. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 19114–19119.
- Giuggioli, L., Potts, J. R., Rubenstein, D. I., & Levin, S. A. (2013). Stigmergy, collective actions, and animal social spacing. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 16904–16909.
- Hamilton, W. D., & May, R. M. (1977). Dispersal in stable habitats. *Nature*, *269*, 578–581.
- Hao, T., Elith, J., Guillera-Aroita, G., & Lahoz-Monfort, J. J. (2019). A review of evidence about use and performance of species distribution modelling ensembles like biomod. *Diversity and Distributions*, *25*, 839–852.
- Huggler, K. S., Holbrook, J. D., Hayes, M. M., Burke, P. W., Zornes, M., Thompson, D. J., Clapp, J. G., Lionberger, P., Valdez, M., & Monteith, K. L. (2022). Risky business: How an herbivore navigates spatio-temporal aspects of risk from competitors and predators. *Ecological Applications*, *32*, e2648.
- Kareiva, P., & Odell, G. (1987). Swarms of predators exhibit 'preytaxis' if individual predators use area-restricted search. *The American Naturalist*, *130*, 233–270.
- Klappstein, N. J., Potts, J. R., Michelot, T., Börger, L., Pilfold, N. W., Lewis, M. A., & Derocher, A. E. (2021). Energy selection functions: Modelling the energetic drivers of animal movement and habitat use. *Journal of Animal Ecology*, *91*(5), 946–957.
- Lewis, M. A., Fagan, W. F., Auger-Méthé, M., Frair, J., Fryxell, J. M., Gros, C., Gurarie, E., Healy, S. D., & Merkle, J. A. (2021). Learning and animal movement. *Frontiers in Ecology and Evolution*, *9*, 681704.
- Manly, B., McDonald, L., Thomas, D., McDonald, T., & Erikson, W. (2002). *Resource selection by animals: Statistical design and analysis for field studies*. Elsevier Academic Press.
- McClellan, S. A., Rumble, M. A., King, R. M., & Baker, W. L. (1998). Evaluation of resource selection methods with different definitions of availability. *The Journal of Wildlife Management*, *62*, 793–801.
- Merkle, J. A., Cross, P. C., Scurlock, B. M., Cole, E. K., Courtemanch, A. B., Dewey, S. R., & Kauffman, M. J. (2018). Linking spring phenology with mechanistic models of host movement to predict disease transmission risk. *Journal of Applied Ecology*, *55*, 810–819.
- Merkle, J. A., Potts, J. R., & Fortin, D. (2017). Energy benefits and emergent space use patterns of an empirically parameterized model of memory-based patch selection. *Oikos*, *126*(2), 18–5.
- Moorcroft, P., & Barnett, A. (2008). Mechanistic home range models and resource selection analysis: A reconciliation and unification. *Ecology*, *89*, 1112–1119.
- Moorcroft, P., & Lewis, M. (2006). *Mechanistic home range analysis*. Princeton University Press.
- Munden, R., Börger, L., Wilson, R. P., Redcliffe, J., Brown, R., Garel, M., & Potts, J. R. (2021). Why did the animal turn? Time-varying step selection analysis for inference between observed turning-points in high frequency data. *Methods in Ecology and Evolution*, *12*, 921–932.
- Murray, J. D. (2003). *Mathematical biology II: Spatial models and biomedical applications*. Springer-Verlag.
- Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 19052–19059.
- Northrup, J. M., Hooten, M. B., Anderson, C. R., Jr., & Wittemyer, G. (2013). Practical guidance on characterizing availability in resource selection functions under a use-availability design. *Ecology*, *94*, 1456–1463.
- Northrup, J. M., Vander Wal, E., Bonar, M., Fieberg, J., Laforge, M. P., Leclerc, M., Prokopenko, C. M., & Gerber, B. D. (2022). Conceptual and methodological advances in habitat-selection modeling: Guidelines for ecology and evolution. *Ecological Applications*, *32*, e02470.
- Osborne, O. G., Fell, H. G., Atkins, H., van Tol, J., Phillips, D., Herrera-Alsina, L., Mynard, P., Bocedi, G., Gubry-Rangin, C., Lancaster, L. T., Creer, S., Nangoy, M., Fahri, F., Lupiyaningdyah, P., Sudiana, I. M., Julianti, B., Travis, J. M. J., Papadopoulos, A. S. T., & Algar, A. C. (2022). Fauxcurrence: Simulating multi-species occurrences for null models in species distribution modelling and biogeography. *Ecography*, *2022*, e05880.
- Parton, A., Blackwell, P. G., & Skarin, A. (2016). Bayesian inference for continuous time animal movement based on steps and turns. In R. Argiento, E. Lanzarone, I. Antoniano Villalobos, & A. Mattei (Eds.), *International conference on Bayesian statistics in action* (pp. 223–230). Springer.
- Potts, J., Bastille-Rousseau, G., Murray, D., Schaefer, J., & Lewis, M. (2014). Predicting local and non-local effects of resources on animal space use using a mechanistic step-selection model. *Methods in Ecology and Evolution*, *5*, 253–262.
- Potts, J., Mokross, K., & Lewis, M. (2014). A unifying framework for quantifying the nature of animal interactions. *Journal of the Royal Society Interface*, *11*, 20140333.
- Potts, J. R., & Börger, L. (2022). Code from: How to scale up from animal movement decisions to spatio-temporal patterns: An approach via step selection. *Zenodo*, <https://doi.org/10.5281/zenodo.7225910>
- Potts, J. R., Börger, L., Scantlebury, D. M., Bennett, N. C., Alagaili, A., & Wilson, R. P. (2018). Finding turning-points in ultra-high-resolution animal movement data. *Methods in Ecology and Evolution*, *9*, 2091–2101.
- Potts, J. R., Börger, L., Strickland, B. K., & Street, G. M. (2022). Assessing the predictive power of step selection functions: How social and environmental interactions affect animal space use. *Methods in Ecology and Evolution*, *13*, 1805–1818.
- Potts, J. R., Giunta, V., & Lewis, M. A. (2022). Beyond resource selection: Emergent spatio-temporal distributions from animal movements and stigmergent interactions. *Oikos*, *2022*(6), e09188.
- Potts, J. R., Harris, S., & Giuggioli, L. (2013). Quantifying behavioral changes in territorial animals caused by sudden population declines. *The American Naturalist*, *182*, E73–E82.
- Potts, J. R., & Lewis, M. A. (2016). How memory of direct animal interactions can lead to territorial pattern formation. *Journal of the Royal Society Interface*, *13*, 20160059.
- Potts, J. R., & Lewis, M. A. (2019). Spatial memory and taxis-driven pattern formation in model ecosystems. *Bulletin of Mathematical Biology*, *81*, 2725–2747.
- Potts, J. R., & Schlägel, U. E. (2020). Parametrizing diffusion-taxis equations from animal movement trajectories using step selection analysis. *Methods in Ecology and Evolution*, *11*, 1092–1105.
- Rhodes, J., McAlpine, C., Lunney, D., & Possingham, H. (2005). A spatially explicit habitat selection model incorporating home range behavior. *Ecology*, *86*, 1199–1205.



- Riotte-Lambert, L., Benhamou, S., & Chamaillé-Jammes, S. (2015). How memory-based movement leads to nonterritorial spatial segregation. *The American Naturalist*, *185*, E103–E116.
- Riotte-Lambert, L., & Matthiopoulos, J. (2020). Environmental predictability as a cause and consequence of animal movement. *Trends in Ecology & Evolution*, *35*, 163–174.
- Robinson, J. C., & Pierre, C. (2003). Infinite-dimensional dynamical systems: An introduction to dissipative parabolic pdes and the theory of global attractors. Cambridge texts in applied mathematics. *Applied Mechanics Reviews*, *56*, B54–B55.
- Schlägel, U. E., Signer, J., Herde, A., Eden, S., Jeltsch, F., Eccard, J. A., & Dammhahn, M. (2019). Estimating interactions between individuals from concurrent animal movements. *Methods in Ecology and Evolution*, *10*, 1234–1245.
- Signer, J., Fieberg, J., & Avgar, T. (2017). Estimating utilization distributions from fitted step-selection functions. *Ecosphere*, *8*, e01771.
- Signer, J., Fieberg, J., & Avgar, T. (2019). Animal movement tools (AMT): R package for managing tracking data and conducting habitat selection analyses. *Ecology and Evolution*, *9*, 880–890.
- Thurfjell, H., Ciuti, S., & Boyce, M. (2014). Applications of step-selection functions in ecology and conservation. *Movement Ecology*, *2*, 4.
- Tuomainen, U., & Candolin, U. (2011). Behavioural responses to human-induced environmental change. *Biological Reviews*, *86*, 640–657.
- Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society B*, *237*, 37–72.
- van Kampen, N. G. (1981). *Stochastic processes in physics and chemistry*. North-Holland.
- Vanak, A., Fortin, D., Thakera, M., Ogdene, M., Owena, C., Greatwood, S., & Slotow, R. (2013). Moving to stay in place—Behavioral mechanisms for coexistence of African large carnivores. *Ecology*, *94*, 2619–2631.
- Viana, D. S., Granados, J. E., Fandos, P., Pérez, J. M., Cano-Manuel, F. J., Burón, D., Fandos, G., Aguado, M. Á. P., Figuerola, J., & Soriguer, R. C. (2018). Linking seasonal home range size with habitat selection and movement in a mountain ungulate. *Movement Ecology*, *6*, 1–11.
- Wang, M., & Grimm, V. (2007). Home range dynamics and population regulation: An individual-based model of the common shrew *Sorex araneus*. *Ecological Modelling*, *205*, 397–409.
- Williams, H. J., Taylor, L. A., Benhamou, S., Bijleveld, A. I., Clay, T. A., de Grissac, S., Demšar, U., English, H. M., Franconi, N., Gómez-Laich, A., Griffiths, R. C., Kay, W. P., Morales, J. M., Potts, J. R., Rogerson, K. F., Rutz, C., Spelt, A., Trevail, A. M., Wilson, R. P., & Börger, L. (2020). Optimizing the use of biologgers for movement ecology research. *Journal of Animal Ecology*, *89*, 186–206.
- Worton, B. J. (1989). Kernel methods for estimating the utilization distribution in home-range studies. *Ecology*, *70*, 164–168.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Potts, J. R., & Börger, L. (2022). How to scale up from animal movement decisions to spatiotemporal patterns: An approach via step selection. *Journal of Animal Ecology*, *00*, 1–14. <https://doi.org/10.1111/1365-2656.13832>