Received: 30 November 2021 | Accepted: 10 May 2022

DOI: 10.1111/2041-210X.13904

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Assessing the predictive power of step selection functions: How social and environmental interactions affect animal space use

Jonathan R. Potts[1] | Luca Börger[2,3] | Bronson K. Strickland[4] | Garrett M. Street[4,5]

[1]School of Mathematics and Statistics, University of Sheffield, Sheffield, UK

[2]Department of Biosciences, College of Science, Swansea University, Wales, UK

[3]Centre for Biomathematics, College of Science, Swansea University, Wales, UK

[4]Department of Wildlife, Fisheries, and Aquaculture, Mississippi State University, Mississippi State, Mississippi, USA

[5]Quantitative Ecology and Spatial Technologies Laboratory, Mississippi State University, Mississippi State, Mississippi, USA

Correspondence
Jonathan R. Potts
Email: j.potts@sheffield.ac.uk

Funding information
Agricultural Research Service, Grant/Award Number: #58-0200-0-002

Handling Editor: Matthew Schofield

Abstract

1. The ability to predict animal space use patterns is a fundamental concern in changing environments. Such predictions require a detailed understanding of the movement mechanisms from which spatial distributions emerge. However, these are typically complex, multifaceted, and therefore difficult to uncover.
2. Here, we provide a methodological framework for uncovering the movement mechanisms necessary for building predictive models of animal space use. Our procedure begins by parametrising a movement model of each individual in a population using step selection analysis, from which we build an individual-based model (IBM) of interacting individuals, derive predicted broad-scale space use patterns from the IBM and then compare the predicted and empirical patterns. Importantly, discrepancies between these predicted and empirical patterns are used to formulate new hypotheses about the drivers of animal movement decisions and thus iteratively improve the model's predictive power. We demonstrate our method on a population of feral pigs in Mississippi, USA.
3. Our technique incorporates both social interactions between individuals and environmental drivers of movement. At each iteration of model construction, we were able to identify missing features to improve model prediction by analysing the IBM output. These include overuse-avoidance effects of self-attractive mechanisms (i.e. attraction to previously visited sites becomes repulsion if there have been multiple visits in quick succession), which were vital for ensuring predicted occurrence distributions do not become vanishingly small.
4. Overall, we have provided a general method for iteratively improving the predictive power of step selection models. This will enable future researchers to maximise the information obtained from step selection analyses and to highlight potentially missing data for uncovering the drivers of movement decisions and emergent space use patterns. Ultimately, this provides a fundamental step towards the general aim of constructing predictive models of animal space use.

KEYWORDS
animal movement, home range, individual-based model, movement ecology, resource selection, spatial ecology, step selection, utilisation distribution

1 | INTRODUCTION

Understanding the space use of animals is a primary concern of ecological research (Franklin, 2010), giving insights into both behavioural features of individuals (Vázquez Diosdado et al., 2018) and the demographic dynamics of populations (Pagel & Schurr, 2012). From an applied perspective, space use underpins important management actions such as the design of conservation areas (Macdonald & Rushton, 2003) and mitigating the effects of biological invasions (Lewis et al., 2016). Indeed, such applications require more than just an understanding of space use but also tools for predicting spatial distributions under hypothesised future scenarios (Wood et al., 2018).

Often, such predictions make use of statistical modelling tools such as species distribution models (SDMs; Zimmermann et al., 2010) and Resource Selection Functions (RSFs; Manly et al., 2002), which aim to uncover the environmental features that co-vary with animal space use. One can then use these correlations to predict future space use either in a different part of the globe (e.g. when assessing the possibility of a future biological invasion) or in the same location as the environmental features change (e.g. when making conservation decisions). A great deal of recent research effort has gone into improving both the quality of inference and predictive power of SDMs and RSFs, but despite these efforts, the predictive power of such statistical models still tends to be of variable quality (Hao et al., 2019).

A shortcoming of SDMs and RSFs is that they do not usually explicitly model the requirement of animals to move from their past distributions, which are being used to parametrise the model, to their predicted future distributions. Whilst there has been increasing effort in the SDM literature to account for dynamic aspects of landscape (Graham et al., 2010), SDMs that include animal movement are rarer (Holloway & Miller, 2017). Furthermore, they usually only model movement implicitly through summary statistics such as dispersal limitations or migration rates (Miller & Holloway, 2015). In reality, the detailed movement mechanisms that give rise to broader-scale features, such as home ranges, dispersal, or migration, may have a large effect on the ultimate relocation of animals adapting to environmental change (Nathan et al., 2008). For example, there may be physical barriers to movement between past and potential future locations (Beyer et al., 2016). Moreover, even without physical barriers, there may be a time-lag due to the animal having to adapt and re-position themselves in space to make best use of their new environment. In other words, anthropogenic changes can cause ecosystems to be in a transient state (Hastings et al., 2018). Indeed, there is growing evidence that transients are the norm in the Anthropocene (Morozov et al., 2020).

Accounting for ecological transience thus requires a dynamic approach via mechanistic modelling (Francis et al., 2021). Mechanistic models are those that operate with (at least) two levels of description. The first is the *mechanism* (or *process*), from which the model is built. In our case, these are the movement decisions of animals. The second is the *pattern*, which describes the emergent features of the model. In our case, this is any summary statistic that describes animal space use over some extended period of time (e.g. a week, month, season, or year). Since such mechanistic models are dynamic, they can account for transient space use patterns (Morozov et al., 2020). Moreover, their dynamism enables the modeller to incorporate effects of non-linear feedbacks between locations of different individuals and populations (e.g. due to social interactions), which may have a non-trivial effect on the distribution of species (Potts & Lewis, 2019). Consequently, mechanistic modelling of animal space use, that is using movement models to understand spatial distributions, is becoming increasingly popular (Merkle et al., 2017; Michelot et al., 2019; Potts & Lewis, 2014; Signer et al., 2017). Some of the earliest efforts involved modelling home range and territory formation using partial differential equations (PDEs) built from movement processes (Lewis & Moorcroft, 2006), which were shown to have good predictive power for ascertaining changes in territorial structure (Moorcroft et al., 2006). Whilst initial efforts focused on canids, these models have since been extended for use with other territorial (Bateman et al., 2015) and non-territorial species (Ellison et al., 2020). These models have proven useful for uncovering a variety of drivers of space use patterns, including social interactions (Moorcroft et al., 1999), kin-relatedness (Ellison et al., 2020), food distribution (Bateman et al., 2015), and topographical effects (Moorcroft et al., 2006).

One way of parametrising such mechanistic models is to fit the emergent space use distributions to relocation data (Lewis & Moorcroft, 2006). However, there is an implicit assumption that the locations are independent samples of a steady-state distribution. This can be problematic for two reasons: (a) locations are never actually independent and ascertaining the effect of this assumption can be tricky (Ellison et al., 2020), and (b) space use distributions may not be in a steady state (Bateman et al., 2015). Consequently, many recent attempts at mechanistic modelling of space use have focused on parametrising models from the movements between successive locations, often known as the 'step' between one measured location and the next (Potts, Mokross, & Lewis, 2014; Potts & Schlägel, 2020; Signer et al., 2017). This replaces the independence assumption with a Markov assumption such that each location is assumed to be dependent on either the previous location (to incorporate locational autocorrelation) or the previous two locations (incorporating both locational and directional autocorrelation). Such parametrisation can be performed using the technique of step selection analysis (SSA) or one of its variants (Avgar et al., 2016; Forester et al., 2009; Fortin et al., 2005; Munden et al., 2021). This technique is both well-established and relatively user-friendly, especially owing to `R` packages such as `amt`, which

simplify work-flow (Fieberg et al., 2021; Signer et al., 2019). Therefore, it seems a promising way forward for parametrising mechanistic models of space use (Potts & Schlägel, 2020).

As well as parametrising PDE models, many studies have found it favourable to use either integro-difference equations (IDEs) or stochastic individual based models (IBMs) instead (Potts, Bastille-Rousseau, et al., 2014; Signer et al., 2017). The main attraction of PDEs arises from the wealth of analytic tools to understand their pattern formation properties, which are not present with IDEs or IBMs. These analytic tools enable one to ascertain qualitative features of space use without recourse to extensive simulation analysis (Potts & Lewis, 2019). However, constructing PDE models of movement often requires one to take continuum limits, moment closure assumptions, and/or mean field assumptions, which may not be reasonable in all circumstances (Wang & Potts, 2017). Furthermore, solving PDEs numerically can be technically challenging, being a research field in its own right (Ames, 2014). Therefore, approaches via IDEs and IBMs are attractive, especially for constructing mechanistic models of space use that have strong predictive power.

Here, we build on the general approach of using SSA to parametrise mechanistic space use models, but we focus on maximising the ability for our models to capture broad-scale spatial patterns and use an IBM approach for simulation. Our framework uses an iterative procedure of model improvement whereby the emergent space use patterns from an empirically parametrised IBM are compared with data to uncover missing aspects of the model (Figure 1). We focus initially on modelling social interactions between individuals, but also show how to incorporate environmental interactions. We demonstrate our framework via application to a population of feral pigs *Sus scrofa* in northwest Mississippi, USA. Pigs are a strongly invasive, gregarious and omnivorous generalist occurring in a wide variety of ecosystems and landscape types worldwide (McClure et al., 2015). Their diets are flexible and can include a variety of naturally occurring herbaceous vegetation and hard and soft mast (*Quercus* spp.; Ballari & Barrios-García, 2014). They also opportunistically consume vertebrate and non-vertebrate fauna when available (Ballari & Barrios-García, 2014). Feral pigs also consume, trample and damage agricultural crops (e.g. corn, soybean, potato and rice; Paolini et al., 2018), and predate deer and livestock neonates when available, causing significant financial loss and motivation for feral pig management. These pulsed resources of naturally occurring and anthropogenic sources of food lead to seasonal effects in feral pig resource selection (Paolini et al., 2018), with consequent downstream effects on seasonal home range size and structure (Paolini et al., 2019). This flexibility across seasons and landscape structures makes pigs an ideal candidate to evaluate the performance of our mechanistic framework.

Our broad aim is to make inroads into providing a general method for building predictive, mechanistic models of animal space use from a combination of animal tracking and environmental data. There are two steps to this. First, one needs to ascertain whether models parametrised from point-to-point movement data of animals can capture broader-scale space use patterns in the same dataset. This shows how good the model is at predicting the occurrence distribution at future points in time from data on the occurrence distribution at a previous time. Second, one needs to ascertain the extent to which these models can capture space use patterns in novel situations, using data different from those used to parameterise the model. Here, the analysis of our dataset will focus purely on the first question, for simplicity. However, the same techniques that we develop here could also be used to address the second question as well, given the right data. In Section 4, we will give some explanation of how this second step might be done.

Overall, our framework both gives a measure of the predictive capability of models fitted via step selection techniques and methods for ascertaining what might need to be included in these models to improve their predictive power. Thus, as well as having the potential to predict future space-use patterns, our methods can also be used to generate new hypotheses about the drivers of animal movement decisions, thereby enabling researchers to gather more information from their existing datasets.

2 | MATERIALS AND METHODS

2.1 | Data

From November 2015 to May 2016, we captured 16 adult feral pigs using baited corral traps in the Mississippi Alluvial Valley (approximately N33°31′34.32″ W90°40′38.28″). We equipped captured animals with Iridium GPS collars [LOTEK Engineering Ltd. ($n = 3$); Vectronic Aerospace GmbH ($n = 13$)] and collected data at 2-hr refresh rates through January 2017. We removed all animal relocations with horizontal dilution of precision (HDOP) < 10 for a location accuracy of 3–7 m (Rempel & Rodgers, 1997), as well as all relocations

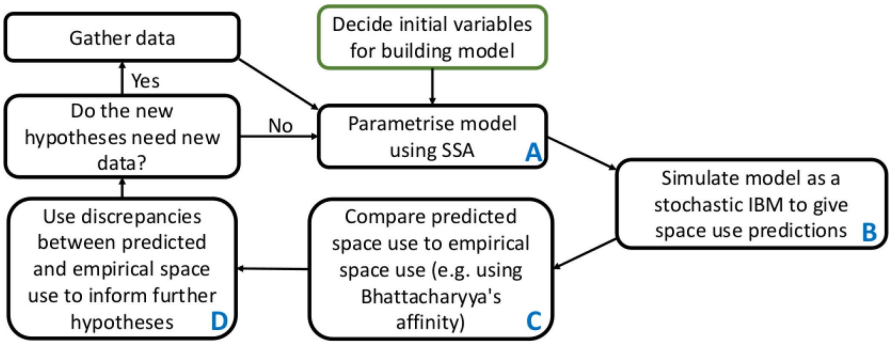

FIGURE 1 Framework. Flow-chart demonstrating our framework for building predictive models of animal movement. The blue letters refer to the steps labelled in Section 2.2. The process starts from the green box, in which one determines which variables to use in building the initial model.

within 24 hr of collar deployment or mortality. Additionally, we used an algorithmic cleaning procedure to remove aberrant relocations based on the known movement capabilities of feral pigs. Feral pigs typically travel at 5 km/hr but are known to sprint up to 50 km/hr (Mayer & Brisbin, 2009), thus if the average velocity between consecutive fixes exceeded 20 km/hr, we removed the later location from the dataset on the basis that such speeds cannot be realistically maintained over that period. This procedure was repeated until each set of consecutive fixes returned an average velocity ≤ 20 km/hr (Paolini et al., 2018). Of the 16 tagged pigs, we selected 12 animals for this analysis based on whether they came within proximity of another collared pig (i.e. their ranges overlapped). Since no other home ranges overlap these, we refer to this area henceforth as the 'study area' and the pigs therein as the 'study population'. We obtained landscape data from the 2016 version of the USDA National Agricultural Statistics Service Cropland Data Layer (CDL) with a 30 m resolution (accessed May 1, 2021; https://nassgeodata.gmu.edu/CropScape/). For a full description of animal capture and data cleaning protocols, see Paolini et al. (2018). All animal procedures and protocols were strictly followed and approved by the Mississippi State University's Institutional Animal Care and Use Committee (Protocol Number: 14–100), and followed the guidelines set forth by the American Society of Mammalogists.

2.2 | Methodological framework

Our methodological framework involves a work process of iterative model improvement (Figure 2) whereby we first (A) use step selection analysis (SSA) to parametrise an individual-based model (IBM) of animal movements and social interactions from data on co-moving animals, then (B) simulate this model to produce predicted spatial occurrence distributions (ODs), which estimate where an animal is located over a particular time interval (Fleming et al., 2015). Next, we (C) compare these ODs to the data to identify movement processes that may be missing from the IBM, then (D) modify the model based on the hypothesised missing movement processes identified in step (C). Finally, we return to step (A) to test whether there is significant evidence for the presence of these new hypothesised movement processes, and the entire process repeats. This broadly follows the 'Strong Inference' philosophy, dating back to Platt (1964), and still highly relevant in the context of ecological modelling (Burnham & Anderson, 2001; Ganusov, 2016). We give full details of each of these steps below, together with how we applied them to the pig data.

Step A: Estimate movement parameters from data using SSA. Here, we parametrise a movement kernel describing the probability density function of animal $i \in \{1, \ldots, N\}$ moving from location x to location z during a single time-step of length τ (where τ is constant; for the pig data, $\tau = 2$ hr) with the following functional form:

$$p_{i,\tau}(\mathbf{z}|\mathbf{x},t) = \frac{\phi_{i,\tau}(|\mathbf{z}-\mathbf{x}|)\exp\left[\beta_{i,1}Z_{i,1}(\mathbf{x},\mathbf{z},t) + \ \ldots \ + \beta_{i,n}Z_{i,n}(\mathbf{x},\mathbf{z},t) + \beta_{i,n+1}|\mathbf{z}-\mathbf{x}|\right]}{K_i(\mathbf{x},t)}. \quad (1)$$

Here, $\phi_{i,\tau}(|\mathbf{z}-\mathbf{x}|)$ is the 'empirical step length distribution' (Forester et al., 2009), a step length distribution obtained by fitting a predetermined functional form (e.g. exponential or gamma) to the empirical step lengths. Also, $|\mathbf{z}-\mathbf{x}|$ is the Euclidean distance between $\mathbf{z}$ and $\mathbf{x}$. The vector of functions $\mathbf{Z}_i(\mathbf{x},\mathbf{z},t) = \left(Z_{i,1}(\mathbf{x},\mathbf{z},t), \ \ldots, Z_{i,n}(\mathbf{x},\mathbf{z},t)\right)$ are spatial features that are hypothesised to co-vary with the animal's choice of next location and the vector $\boldsymbol{\beta}_i = \left(\beta_{i,1}, \ldots, \beta_{i,n}\right)$ denotes the strength of the effect of each $Z_{i,j}(\mathbf{x},\mathbf{z},t)\,(j \in \{1, \ldots, n\})$ on the movement of animal i. Also,

$$K_i(\mathbf{x},t) = \int_\Omega \phi_{i,\tau}(|\mathbf{x}'-\mathbf{x}|)\exp\left[\beta_{i,1}Z_{i,1}(\mathbf{x},\mathbf{x}',t) + \ \ldots \ + \beta_{i,n}Z_{i,n}(\mathbf{x},\mathbf{x}',t) + \beta_{i,n+1}|\mathbf{z}-\mathbf{x}|\right]\mathrm{d}\mathbf{x}', \quad (2)$$

is a normalising function, ensuring $p_{i,\tau}(\mathbf{z}|\mathbf{x},t)$ integrates to 1 (so is a genuine probability density function) and Ω is the study area. Finally, we will choose an exponential step length distribution

$$\phi_{i,\tau}(|\mathbf{z}-\mathbf{x}|) = \lambda_i \exp\left(-\lambda_i|\mathbf{z}-\mathbf{x}|\right), \quad (3)$$

where $1/\lambda_i$ is the mean step length of animal i, although one could pick any λ_i in theory (Avgar et al., 2016; Fieberg et al., 2021). Then, the $\beta_{n+1}|\mathbf{z}-\mathbf{x}|$ term in Equation (1) corrects for any discrepancy between the empirical step length distribution and the resource independent step length distribution, as detailed in Forester et al. (2009).

So far, this is a relatively standard description of the movement kernel for SSA (Avgar et al., 2016; Forester et al., 2009; Fortin et al., 2005) and so the next task is to specify the functions $Z_{i,j}(\mathbf{z},t)$. Since our aim is to understand the effect of the occurrence distribution (OD) of each animal on both itself and each other animal, we set $Z_{i,j}(\mathbf{z},t)$ to be the OD of individual $j \in \{1, \ldots, N\}$ at time t (so that $n = N$), measured in units of km^{-2}. However, this requires specifying both a time-scale over which we measure the OD and a method for constructing an OD from locational data. Helpfully, the method of Schlägel et al. (2019) provides a way of constructing an OD for this very purpose, using a kriging technique. This technique finds the best-fit movement model from the suite of continuous-time movement models defined in the `ctmm` R package (Calabrese et al., 2016), via the `occurrence` function from that package. For the purpose of demonstrating our methods, we set $T = 30$ days to be the time interval over which each occurrence distribution is calculated. (Note that this assumes a very simple memory process whereby the cognitive map of an animal simply consists of their OD over the last 30 days. For better realism one could include more complex processes, e.g. Merkle et al. (2019), but this is a simple starting point for demonstrating our methods.)

To fit the movement kernel in Equation (1) to our pig data, we follow a standard case–control design used in most SSA studies (Avgar et al., 2016; Forester et al., 2009; Fortin et al., 2005). For this, we sample from the kernel $\phi_{i,\tau}$ 10 times for each step (i.e. 1:10 use: availability) to give control locations and compare these to the measured locations to test for a significant effect of the variables $Z_{i,j}$ for $i,j \in \{1, \ldots, N\}$. This procedure of constructing the ODs using kriging and testing for their effect on animal movement using SSA is identical to that of Schlägel et al. (2019).

Step B. Implement IBM using parameter values from Step A. Since the movement kernel, $p_{i,\tau}$, for animal i depends upon the

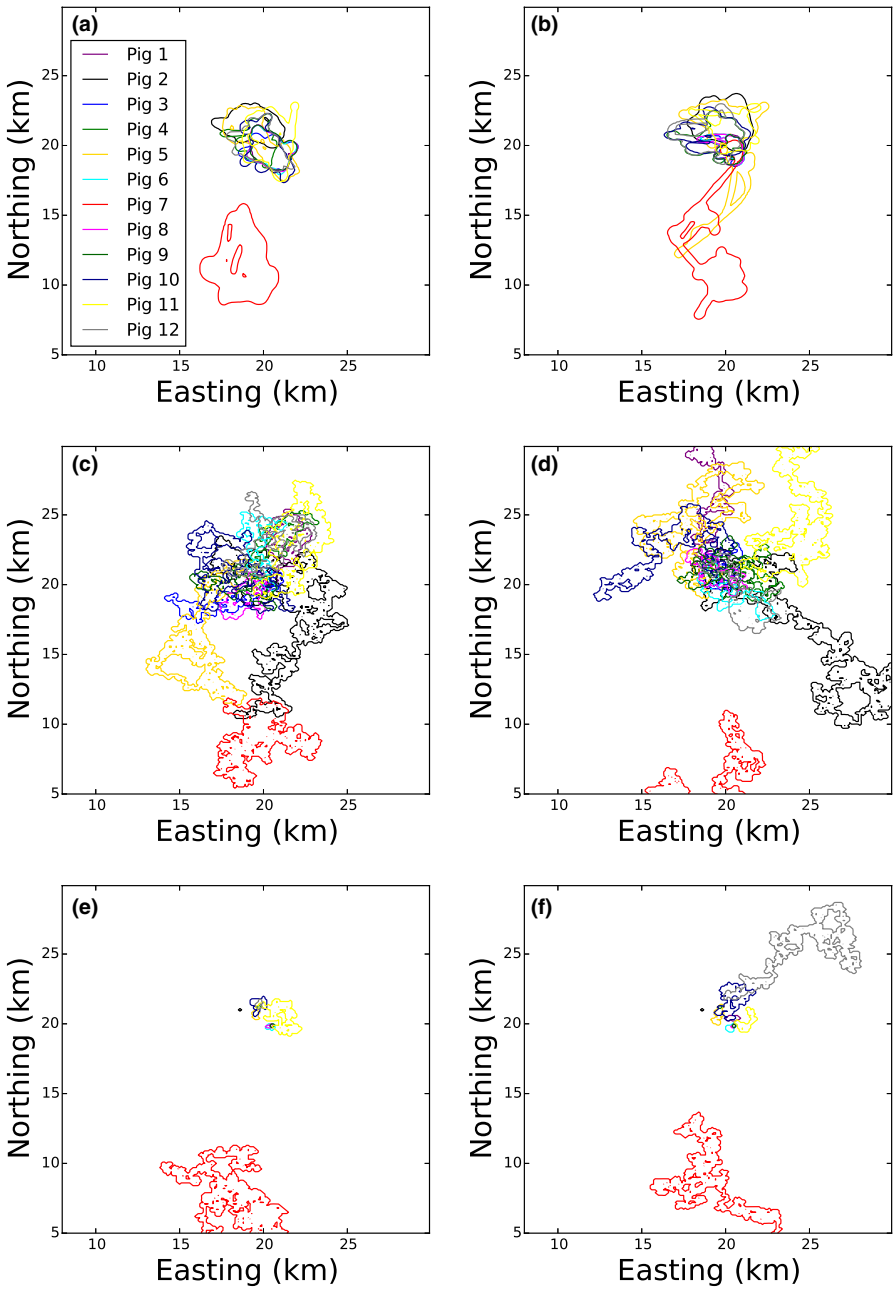

FIGURE 2 Comparing simulations of Null and Interaction models to data. Panel (a) shows the empirical occurrence distribution (OD) of each of the pigs for the period from 5th April to 4th May 2016, used as an initial condition for the simulations. Panel (b) shows the empirical OD for the period from 5th May to 3rd June 2016. Panels (c and d) [respectively, (e) and (f)] show example ODs for simulations of the Null Model (respectively, Interaction Model), which give predictions of the ODs for the period from 5th May to 3rd June 2016. The contour for each OD is taken at a height of 0.001 km^{-2}.

prior locations of all the animals (both animal i and the others), Equation (1) describes a system of *coupled* step selection functions (Potts, Mokross, & Lewis, 2014). These can be analysed using a stochastic IBM of interacting agents, one agent for each animal. Naïvely, one could simply simulate the model in Equation (1) for each animal simultaneously. However, this would require constructing the kriged OD for each animal at each time-step, since each animal's movement at each time step is determined by the ODs of every individual. Such a method would be highly computationally intensive and would make analysis prohibitively time-consuming, so we need a more efficient simulation framework.

For this, we use a nearest-neighbour movement process on a square grid with a relatively-large lattice spacing of $\Delta x = 100\,\text{m}$. The advantage of this is that the OD at time t can be defined as simply the distribution of lattice points that have been visited in the time-period from $t - T$ to t, similar to the model of Giuggioli et al. (2011). This requires no repeated calculation of a kriged OD, and so no requirement to fit a continuous-time movement model for each individual at each time step. Moreover, we will construct such a nearest-neighbour model that is formally related to Equation (1), inasmuch as the drift and diffusion functions are identical in the limit as $\tau \to 0$ (Supplementary Appendix A).

In this model, the probability of animal i moving from a lattice site at location $\mathbf{x}$ to another at location z, during a time-step Δt, is given by

$$f_{i,\Delta t}(\mathbf{z} \mid \mathbf{x}, t) = \begin{cases} \dfrac{d_i}{4C_i(\mathbf{x},t)} \exp\left(\boldsymbol{\beta}_i \cdot \mathbf{Z}_i(\mathbf{x},\mathbf{z},t)\right), & \text{if } \mathbf{z}-\mathbf{x} \in S, \\ \dfrac{1-d_i}{C_i(\mathbf{x},t)} \exp\left(\boldsymbol{\beta}_i \cdot \mathbf{Z}_i(\mathbf{x},\mathbf{z},t)\right), & \text{if } \mathbf{x}=\mathbf{z}, \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

where $S = \{\pm(0,\Delta x), \pm(\Delta x,0)\}$ and

$$C_i(\mathbf{x},t) = \left(1-d_i\right)\exp\left(\boldsymbol{\beta}_i \cdot \mathbf{Z}_i(\mathbf{x},\mathbf{x},t)\right) + \sum_{\mathbf{z}-\mathbf{x}\in S}\frac{d_i}{4}\exp\left(\boldsymbol{\beta}_i \cdot \mathbf{Z}_i(\mathbf{z},t)\right), \quad (5)$$

$$d_i = \frac{\Delta t\int_\Omega |\mathbf{x}|^2\exp\left(\left(\beta_{i,n+1}-\lambda_i\right)|\mathbf{x}|\right)d\mathbf{x}}{\tau(\Delta x)^2\int_\Omega \exp\left(\left(\beta_{i,n+1}-\lambda_i\right)|\mathbf{x}|\right)d\mathbf{x}}. \quad (6)$$

Here, $C_i(\mathbf{x},t)$ can be thought of as a normalising function, to ensure probabilities add to 1, and d encodes the diffusivity of the animal. For Equation (4) to describe genuine probabilities, $f_{i,\Delta t}(\mathbf{z}|\mathbf{x},t)$ must always be positive, so we need to choose Δt such that $d_i < 1$. Notice that there are two time scales in operation here: τ, defined prior to Equation (1), representing the time step between consecutive locations in the data, used for step selection inference, and Δt representing the time step of the IBM.

To parametrise the IBM in Equation (4), we discard any of the $Z_{i,j}$ variables whose corresponding $\beta_{i,j}$-value was not significantly different from 0 ($p > 0.01$) in our step selection analysis from Step A, to avoid over-fitting. We then set $\beta_{i,j} = \widehat{\beta}_{i,j}$ whenever $\beta_{i,j}$ is significantly different from 0, where $\widehat{\beta}_{i,j}$ is the unbiased estimate of $\beta_{i,j}$. Otherwise, we set $\beta_{i,j} = 0$. (However, to check whether our results are sensitive to the effect of discarding parameters, we also repeat our analysis but with $\beta_{i,j} = \widehat{\beta}_{i,j}$ for all $i,j \in \{1,\ldots,N\}$; see Supplementary Table ST2.)

We call this model the *interaction model*, as it encodes all the interactions between individuals for which we have found a significant effect. For comparison, we also constructed a model whereby $\beta_{i,j} = 0$ for every $i,j \in \{1,\ldots,N\}$, which we called the *null model*. The null model is simply a nearest-neighbour random walk with the same diffusion constants as the Interaction Model.

We then perform stochastic simulations of both the interaction model and the null model and compare the space use patterns predicted by these models to those from the data. For each stochastic simulation, we use initial conditions corresponding to the ODs of the 12 pigs during the period from 5th April 2016 to 4th May 2016 inclusive (the first 30 day period for which we have data on all the pigs). We simulate each of the interaction model and the null model for a period corresponding to 30 days of real time (i.e. up to 3rd June 2016).

Step C. Assess model fit and identify gaps. Next, we compare the resulting OD from each simulated pig (from 5th May to 3rd June inclusive) with the corresponding OD from the data for the same time period (inferred using the ctmm procedure described above), using Bhattacharyya's Affinity (BA). This is generally considered the best way to assess overlap of space use in an ecological setting (Fieberg & Kochanny, 2005). The BA gives a value between 0 and 1, where 1 corresponds to perfect agreement and 0 to no overlap. Since the interaction model is stochastic (as with all our IBMs), it will return a different set of ODs each time it is run. Thus, we simulate our stochastic model 1,000 times and calculate the average BA for each pig. We perform the same stochastic simulation procedure on the null model, and compare the resulting BA-values of the null and interaction models, to check whether the interaction model is better at predicting space use than a random walk. When running repeated stochastic simulations, we always ensured that we averaged over sufficiently many stochastic realisations so that any claim we make of an increase (respectively, decrease) in mean BA has a <0.01 probability of being incorrect [i.e. of actually being a decrease (respectively, increase) in mean BA].

As well as the BA, we can gain insight into discrepancies between each model and the data by comparing the sizes of each OD. Since an OD is a probability density function, the variance of the OD provides a standard way to measure its size (Wasserman, 2004). We denote by $V_{i,\mathrm{E}}$ the variance of the empirical OD from 5th May to 3rd June (inclusive) for pig i, by $V_{i,\mathrm{I}}$ the corresponding variance for the interaction model, and by $V_{i,\mathrm{N}}$ the corresponding variance for the null model. To compare magnitudes, we take the ratio of the variances, $V_{i,\mathrm{M}} / V_{i,\mathrm{E}}$. In our examples, it turns out that these ratios can vary by multiple orders of magnitude. Therefore, we take the logarithm of $V_{i,\mathrm{M}} / V_{i,\mathrm{E}}$, and define $Q_{i,\mathrm{M}} = \log_{10}\left(V_{i,\mathrm{M}} / V_{i,\mathrm{E}}\right)$, for a model M (so here $M \in \{I,N\}$), as a measure of the discrepancy between the OD sizes of the model and data for pig i.

Step D. Update the model and iterate. Examining both $Q_{i,\mathrm{M}}$ and the BA (alongside visual examination of simulation output) enables one to construct further models of the general form given in Equation (1), which are then tested for significant improvements on the original model using the SSA procedure from Step A. A positive value of $Q_{i,\mathrm{M}}$ suggests that the model is over-estimating the spatial scale over which pig i roams, and therefore, one might consider incorporating into future models a movement process that is likely to lead to more confined movement. Conversely, a negative value of $Q_{i,\mathrm{M}}$ suggests that the existing confinement processes in the model are too strong; therefore, they needs to be counter-balanced by an parameter that causes wider exploration of space by individuals. It is also valuable to examine the mean of $Q_{i,\mathrm{M}}$ over all i, which we denote by $\overline{Q}_{\mathrm{M}}$.

Repeating the process. In Sections 2.3 and 2.4, we will give details for our specific example of how we constructed new models that we hypothesised might improve on the interaction model in terms of space use prediction. The general procedure is to use the IBM techniques of Steps B and C to ascertain whether or not these new models are better at predicting space use patterns (using BA as a yard-stick). This then enables us to generate further hypotheses about what might still be missing from all of the models so far constructed, so going back through the whole process (steps A–D) once more (Platt, 1964). In principle, one can continue this iterative process until the models being produced are no longer giving better predictions. However, for the purposes of demonstrating our method, we only go through three iterations: first analysis of the initial models (the interaction model and null model), and then two stages of subsequent modelling, each stage aiming to construct models that improve predictive power over any previously constructed model.

2.3 | Assessing different interaction mechanisms

So far, our interaction model assumes that an animal responds to an OD by either moving up or down the ODs gradient. This can be problematic in the case of self-attraction, since it can cause a positive

feedback mechanism whereby simulated individuals immediately move back to the location they have just visited, and therefore end up pinned in an arbitrarily small area.

To mitigate this, we propose three alternative models. In the first, we simply remove self-attraction; that is, we set $\beta_{i,i} = 0$ and called this the *no self-attraction model*. In the second, we replace self-attraction with attraction to a central place, $\mathbf{x}_{i,\mathrm{cp}}$ (for $i \in \{1, \ldots, 12\}$), defined as the centre of mass of the OD of the initial condition for each pig (i.e. the OD from 5th April 2016 to 4th May 2016 inclusive). This requires incorporating an additional variable, $Z_{i,\mathrm{C}}(\mathbf{x}, \mathbf{z}, t) = \left|\mathbf{z} - \mathbf{x}_{i,\mathrm{cp}}\right|$, into the step selection function (Equation 1). We call this model the *central-place attraction model*. Note that this is qualitatively similar to OrnsteinUhlenbeck models (Wang et al., 2019), as the strength of attraction increases with distance from the central point, but it is not mathematically identical.

The third model incorporates a quadratic self-attraction term alongside all the parameters from the interaction model, denoted by $Z_{i,\mathrm{Q}}(\mathbf{x}, \mathbf{z}, t) = \left[Z_{i,I}(\mathbf{x}, \mathbf{z}, t)\right]^2$ (full functional forms of all models are given in Supplementary Appendix B). The rationale behind this is that if the quadratic terms turned out to be negative, this would act as a counter-balance to the pinning effect by making extremely highly used lattice sites repellent rather than attractive. Biologically, this would mean that pigs have a tendency to be attracted to familiar areas up to a point, but will avoid areas that have been overused in the recent past. We call this model the *overuse avoidance model*.

2.4 | Incorporating environmental features

As well as social interactions and self-attraction, animal space use is shaped by its environment. The presence and juxtaposition of food, cover and water ultimately determine the suitability and occupancy of feral pigs on a landscape, which leads to a complex web of environmental features that may ultimately determine pig movement decisions. However, for the purposes of giving a simple demonstration of how to incorporate environmental features into our framework, we focus here on just one feature. Since pigs cannot sweat, they rely on shade and water to thermoregulate, both of which are associated to forested habitat, which pigs are known to favour (Paolini et al., 2018). Therefore, we focus here simply on the presence or absence of forest cover as a means of demonstrating how to incorporate environmental features into our model. We also provide a further demonstration of incorporating environmental effects (specifically the presence of corn crops) in Supplementary Appendix C.

To account for forest cover, for each pig $i = 1, \ldots, N$, we define $Z_{\mathrm{F}}(\mathbf{z}, \mathbf{x}, t) = 1$ if location $\mathbf{z}$ (at the end of the step) is in forest and $Z_{\mathrm{F}}(\mathbf{z}) = 0$ if not (using the notation from Equation 1). Since $Z_{\mathrm{F}}(\mathbf{z}, \mathbf{x}, t)$ is defined to be independent of both the start of the step, $\mathbf{x}$, and time, t, we will write $Z_{\mathrm{F}}(\mathbf{z}) = Z_{\mathrm{F}}(\mathbf{z}, \mathbf{x}, t)$ to ease notation. Figure 4a shows the distribution of forest in our study area and Figure 4b superimposes on this the empirical OD for the period from 5th May to 3rd June 2016. We see that there is an area of forest around 20 km east and 20 km north in which all pigs (except Pig 7; in red) spent most of their time during this period, so it is valuable to keep this in mind when assessing the predictive capability of models with forests.

We construct two models that incorporate forest data. The first (termed the *forest model*) uses both $Z_{\mathrm{F}}(\mathbf{z})$ and all the variables from the overuse avoidance model. The second (termed the *just forest model*) has $Z_{\mathrm{F}}(\mathbf{z})$ as the sole covariate (see Supplementary Appendix B for detailed functional forms). This second model is provided to test whether or not the presence of forest alone is better at predicting space use than models of social interactions.

3 | RESULTS

Using SSA to parametrise the interaction model resulted in 144 values of $\hat{\beta}_{i,j}$, one for each pair $i, j \in \{1, \ldots, 12\}$. Of these, 43 were significantly different from 0 ($p < 0.01$; Table 1). From these 43, two were negative and the other 41 were positive. Recall that $\hat{\beta}_{i,j}$ represents the tendency for pig i to be attracted to (respectively, avoid) pig j if $\hat{\beta}_{i,j} > 0$ (respectively, $\hat{\beta}_{i,j} < 0$). Therefore, 41 of these 43 interactions represented attractive tendencies to pigs' recent ODs. The two that did not were pigs 10 and 11, who had a slight tendency to avoid the ODs of pig 3 (Table 1).

Simulations of the interaction model revealed that this model often predicted ODs that were far smaller than those seen in the data (compare Figure 2b with Figure 2e,f). Indeed $\overline{Q}_{\mathrm{I}} \approx -3.1$, suggesting that predicted ODs we around three orders of magnitude smaller (in area) than the empirical distributions. This contrasted with the null model ($\beta_{i,j} = 0$ for all i, j) where, as one might expect, the predicted ODs were much larger than the empirical distributions (compare Figure 2b with Figure 2c,d). Here, $\overline{Q}_{\mathrm{I}} \approx 0.90$, suggesting the ODs predicted by the null model were about eight times larger than the empirical ODs.

Bhattacharyya's affinity (BA) between the interaction model and the empirical distribution tended to be slightly higher (average across individuals of $BA = 0.170$) than those of the null model (average of $BA = 0.154$; Table 2; results for each individual are in Supplementary Table ST1). However, this still does not represent a very good agreement between model and data, likely due to the very small ODs often predicted by the Interaction Model (Figure 2e,f).

The no self-attraction model was marginally better than the interaction model at predicting space use (Table 2), with an average BA of 0.188 between the model and the data as compared with 0.170. Regarding the predicted OD sizes, we found that the $\overline{Q}$-value for the no self-attraction model was $\overline{Q}_{\mathrm{NSA}} \approx -1.5$. This is an improvement on the interaction model ($\overline{Q}_{\mathrm{I}} \approx -3.1$) but still suggests that the predicted ODs tend to be 1.5 orders of magnitude (about 30 times) smaller in area than the measured ODs. That said, visual output reveals that some predicted ODs are quite a bit bigger than the empirical ODs (Figure 2a,b), and for some pigs, $\overline{Q}_{i,\mathrm{NSA}}$ is positive (e.g. $\overline{Q}_{11,\mathrm{NSA}} \approx 1.3$).

TABLE 1 Best fit parameter values, $\widehat{\beta}_{i,j}$, for the interaction model, where $i, j \in \{1, \ldots, 12\}$. Columns (respectively, rows) correspond to the j-values (respectively, i-values). The labels P1–P12 refer to pigs 1–12, respectively. The numbers in bold are significantly different from 0 ($p < 0.01$), whereas other numbers are not significant.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
P1	**0.28**	0.00	0.02	0.06	0.09	0.04	**0.38**	**0.17**	−0.08	0.02	0.18	**0.50**
P2	−3.90	**0.98**	**2.34**	0.44	0.34	0.27	−1.35	−0.80	0.55	0.08	**0.61**	−0.03
P3	0.10	**1.56**	**0.18**	−1.13	**0.35**	**0.13**	0.47	0.06	−0.06	−0.03	0.18	0.38
P4	0.25	0.32	0.15	**0.48**	0.44	−0.82	**0.39**	0.16	0.30	0.19	0.39	**0.22**
P5	0.00	**0.42**	−0.11	0.17	**0.76**	0.07	0.48	−0.65	**0.84**	0.29	0.36	**0.29**
P6	0.10	0.45	0.05	0.26	**0.15**	**0.21**	**0.44**	0.06	0.04	−0.13	0.05	0.26
P7	0.12	2.02	−0.24	−0.09	0.20	0.07	**0.42**	0.12	0.23	**0.60**	0.28	**0.69**
P8	**0.15**	0.06	−0.01	0.04	0.18	**0.13**	0.24	**0.24**	0.05	−0.09	−0.16	0.40
P9	−0.01	−0.13	0.03	0.21	**0.38**	0.01	0.31	**0.20**	**0.16**	0.06	0.05	**0.33**
P10	0.07	0.19	**−0.26**	−0.14	−0.16	0.20	**0.89**	0.13	0.07	**0.66**	0.37	**0.36**
P11	−0.13	0.15	**−0.78**	0.18	0.09	0.66	**1.09**	−0.25	**0.45**	**0.58**	**0.43**	0.19
P12	**0.57**	0.40	0.14	**0.16**	−0.07	−0.36	**0.32**	−0.27	0.17	**0.34**	0.15	**0.43**

TABLE 2 Summary description of each model, together with average (across individuals) Bhattacharyya's affinity between simulation and data. Each row corresponds to a different model: Null, interaction, no self-attraction, central-place attraction, overuse avoidance (from top to bottom). Supplementary table ST1 contains per-individual values.

Model	Summary description	Ave BA
Null	No social, self, or environmental interactions	0.154
Interact	Social interactions and self-attraction	0.170
No SA	Social interactions but no self-attraction	0.188
CPA	Social interactions and central-place attraction	0.235
OAM	Social interactions, self-attraction, but overuse avoidance	0.297
Forest	OAM plus attraction to forest	0.303
Just forest	Attraction to forest but no social/self interactions	0.162

On fitting the central-place attraction model, we found that 8 of the 12 pigs showed a significant attractive tendency towards the central place (Table 3). Simulations revealed that this model tended to be better at predicting space use than the previous three, with an average BA of 0.235 between the model and the data. However, for this model, $\overline{Q}_{CPA} \approx -1.6$, marginally worse than the no self-attraction model but better than the interaction model. This decrease in OD size can also be seen visually: compare Figure 3c,d to the no self-attraction model (Figure 3a,b) and the interaction model (Figure 2e,f).

The quadratic self-attraction term, $Z_{i,Q}(\mathbf{x}, \mathbf{z}, t)$, turned out to be highly significant for all pigs (Table 3). Furthermore, the resulting overuse avoidance model was better at predicting space use than the previous four, having an average BA of 0.297 between the model and the data. The predicted OD sizes are also closer to the empirical ODs than the other models, with $\overline{Q}_{OA} \approx -0.63$. Nonetheless, this still means that predicted ODs tend to underestimate the OD area by a factor of about four.

The forest model was very marginally better at predicting space use than the overuse avoidance model, with an average BA between model and data of 0.303 compared to 0.297 (Table 2). Whilst we did sufficient simulations to ensure that this increase is significant ($p < 0.01$) the effect size is clearly quite small. This suggests that the effect of forests on space use is minimal compared to the effect of social interactions for our study population. Indeed, for the just forest model, the BA between model and data is only 0.162 (Table 2), which implies worse predictive power than all of the models except the null model (which, recall, is just a random walk). This suggests that social interactions are more important factor in shaping space use than the presence of forest for our study population (though an alternative explanation may be there is another environmental feature we have not tested here).

Indeed, for the forest model, only four of the pigs showed a significant effect of forest presence on movement, when controlling for social interactions (Table 3). All of these four show a positive (i.e. attractive) effect. When we neglect to control for social interactions (the Just Forest Model), however, we find that eight of the pigs show a significant preference for forest, and also that the coefficients for attraction to forest are larger. This suggests that apparent preference for forest may sometimes be an artefact of social interactions with pigs who prefer forest themselves, or vice versa.

Regarding predicted OD size, the $\overline{Q}$ – value for the forest model was $\overline{Q}_F \approx -0.69$, which is actually slightly worse (further from zero) than the overuse avoidance model, $\overline{Q}_{OA} \approx -0.63$. For the just forest model, $\overline{Q}_{JF} \approx 0.98$, suggesting it is over-estimating space use by nearly one order of magnitude, adding further evidence to suggest that the effect of forests alone is insufficient for predicting space use patterns. Figure 4c,d give example simulation output for the forest model and the just forest model, respectively. In the forest model example, all the pigs that began roughly within the area of forest at around 20 km east and 20 km north remain roughly in that area.

TABLE 3 Best fit parameter values (plus/minus standard error) for the central-place attraction parameter, $\widehat{\beta}_{i,C}$, in the central-place attraction model (CPA; second column), the quadratic parameter, $\widehat{\beta}_{i,Q}$, from the overuse avoidance model (third column), the forest parameter, $\widehat{\beta}_{i,F}$ from both the forest model (fourth column) and the just Forest model (fifth column). The numbers in bold are significantly different from 0 ($p < 0.01$), whereas other numbers are not significant.

Pig	CPA	Overuse	Forest	Just Forest
1	**− 1.01 ± 0.22**	**− 0.012 ± 0.0013**	0.113 ± 0.105	**0.296 ± 0.092**
2	**− 0.62 ± 0.10**	**− 0.027 ± 0.0027**	**0.378 ± 0.115**	**0.556 ± 0.102**
3	**− 1.34 ± 0.35**	**− 0.015 ± 0.0031**	− 0.042 ± 0.125	**0.310 ± 0.106**
4	− 0.38 ± 0.18	**− 0.044 ± 0.0053**	− 0.090 ± 0.113	0.026 ± 0.102
5	− 0.17 ± 0.12	**− 0.056 ± 0.0048**	**0.495 ± 0.113**	**0.707 ± 0.097**
6	− 0.48 ± 0.23	**− 0.022 ± 0.0032**	0.144 ± 0.096	0.209 ± 0.083
7	**− 0.39 ± 0.14**	**− 0.014 ± 0.0018**	**0.455 ± 0.135**	**0.821 ± 0.116**
8	**− 0.77 ± 0.25**	**− 0.025 ± 0.0034**	0.182 ± 0.108	**0.287 ± 0.093**
9	**− 1.21 ± 0.21**	**− 0.032 ± 0.0033**	0.217 ± 0.097	**0.347 ± 0.085**
10	− 0.34 ± 0.14	**− 0.039 ± 0.0056**	− 0.204 ± 0.128	− 0.091 ± 0.110
11	**− 0.42 ± 0.09**	**− 0.006 ± 0.0004**	**0.373 ± 0.085**	**0.575 ± 0.074**
12	**− 1.04 ± 0.13**	**− 0.023 ± 0.0020**	0.065 ± 0.090	− 0.088 ± 0.080

This is a feature that cannot be seen in any of the previous models. However, the just forest model displays quite poor predictions of space use patterns, confirming the inflated OD sizes measured by $\overline{Q}_{JF}$.

4 | DISCUSSION

We have described a methodological framework for assessing the power of parametrised step selection models for predicting space use patterns. We have shown how this framework can be used to uncover missing mechanisms in step selection models. It is possible to incorporate into these models both social interactions between animals and the effects of environmental features on movement. Our work provides a generic framework for iterative construction of predictive, mechanistic models of animal space use.

To demonstrate our framework empirically, we used data on concurrent movements of feral pigs in Mississippi, USA. Our approach was to parametrise our step selection model from the whole dataset then ascertain the extent to which the model could reproduce the broad-scale space use patterns in the same dataset. To do this, we set the initial conditions of our model to be the empirical ODs at one point in time, then ascertained the extent to which the model predicted the empirical ODs at future points in time. This procedure of parametrising and testing our model using the same data (but at different spatio-temporal scales) means that any discrepancy between prediction and observation suggests a missing feature in the model.

Alternatively, one could also test the model using a different dataset to the one used to parametrise the model. However, then discrepancies between prediction and data would arise from a combination of differences between datasets and missing features of the model. Therefore we suggest using the same dataset for both constructing the model and testing its emergent features in the first instance. This does not, in and of itself, demonstrate that the model is good at predicting outside the particular study dataset, though. Once one has a model that works well on one dataset, to make the model truly predictive, the next step is to test how it fares when faced with different data that were not used for model parametrisation. For this, one can follow the same iterative procedure of model improvement as described here, but using different data to parametrise from those used to examine ODs. However, when extrapolating in this way, it is crucial to note that a large set of additional considerations may come into play, such as changes in the internal state (and hence movement mode) of the animal, environmental changes (e.g. change of season), etc. Ultimately, this would lead to a train/test type procedure whereby models are successively improved as more data emerges to train them.

Whilst the main purpose of our application to pig data was as a tool for introducing the method, rather than studying pig biology per se, our results nevertheless suggest some interesting features of pig movement, at least in relation to our particular study population. First, having parametrised a model of social interactions, we found that there was a significant signal of self-attraction for each pig (i.e. attraction to its own OD, a.k.a. philopatry). Yet, when we simulated this mechanism, the predicted future ODs were typically several orders of magnitude smaller than the empirical ODs. This shows that there must be some counter-balancing mechanism that is not in the model that allows for more realistically sized ODs to emerge. We found that incorporating a quadratic self-repulsion term, so that pigs avoid over-used areas, was both statistically significant for all pigs and caused slightly larger ODs to emerge. This is in keeping with the marginal value theorem (Charnov, 1976), whereby it is optimal for animals to leave areas that have been over-used. However, the predicted ODs still tend to be smaller than the empirical ODs, suggesting that it may be beneficial to search for other mechanisms that allow the simulated ODs to grow to a size more fitting with the empirical ODs (one hypothesis might be to incorporate quadratic terms into the between-pig interactions).

A second interesting outcome from the application to our study population is that, whilst attraction to forest is statistically significant for most of the pigs, this attraction both becomes less apparent when controlling for social interactions and also has very little effect on the predictive power of the resulting models. This gives evidence to suggest that social interactions and self-attraction are driving space use to a much greater extent than attraction to forest in our

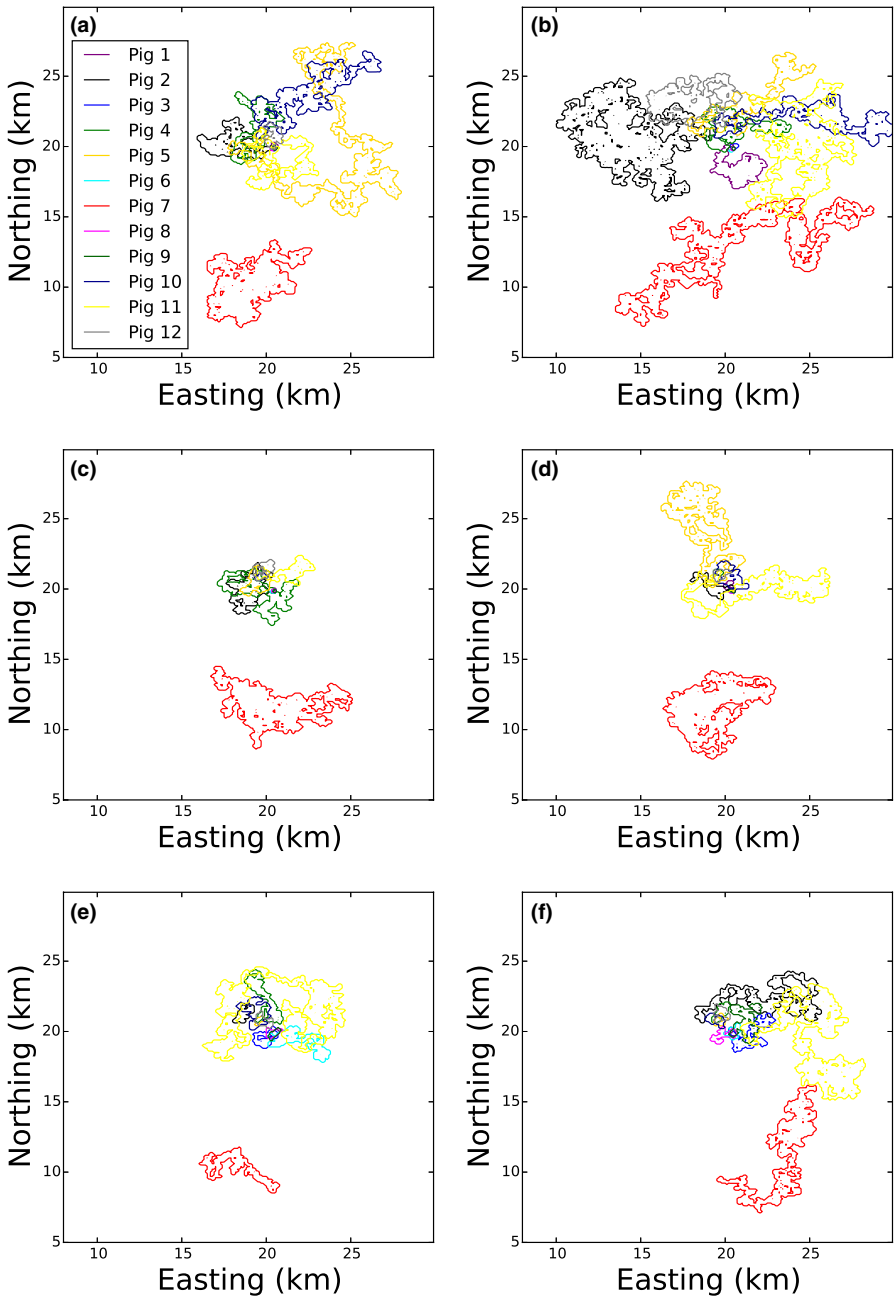

FIGURE 3 Example simulations of occurrence distributions. Panels (a and b) were constructed using the No Self-Attraction Model; panels (c and d) used the Central-Place Attraction Model; panels (e and f) display output from the Overuse Avoidance Model. As in Figure 3, the contours were drawn at a height of $0.001\,km^{-2}$.

particular study population, a similar feature to those found in other species (Cozzi et al., 2018). The results were similar for the effect of corn presence (Supplementary Appendix C). However, we have only tested these two environmental variables, for the purposes of demonstrating our method, and only on a relatively small population of coexistent animals. More detailed environmental information and a wider study may reveal a different picture, including the possibility that the apparent importance of social interactions is (partially or wholly) explicable by other factors (Calabrese et al., 2018).

Indeed, the possibility that untested covariates are driving space use is always an issue in resource and step selection analyses, as there will inevitably be drivers of movement on which researchers do not have data (Fieberg et al., 2018). The methods here help identify such missing drivers of space use, which in turn may improve both inference and predictive power. To this end, we are currently using our methodological advancements to build a more detailed predictive model of the pig system, which will be the subject of future work.

Providing techniques for building models requires a measure of goodness-of-fit between model and data. However, in studies of movement and habitat selection, the goodness-of-fit question is rarely addressed (Potts, Auger-Méthé, et al., 2014), partly due to a paucity of techniques (Fieberg et al., 2018). The latter study proposed use-calibration plots to address this question in the context of resource selection. This examines the distribution of explanatory variables; that is, the environmental variables used to explain locations of animals. Whilst this is a valuable technique in situations where there is a key demarcation between explanatory and response variables, in the situation of social interactions, there are feedbacks

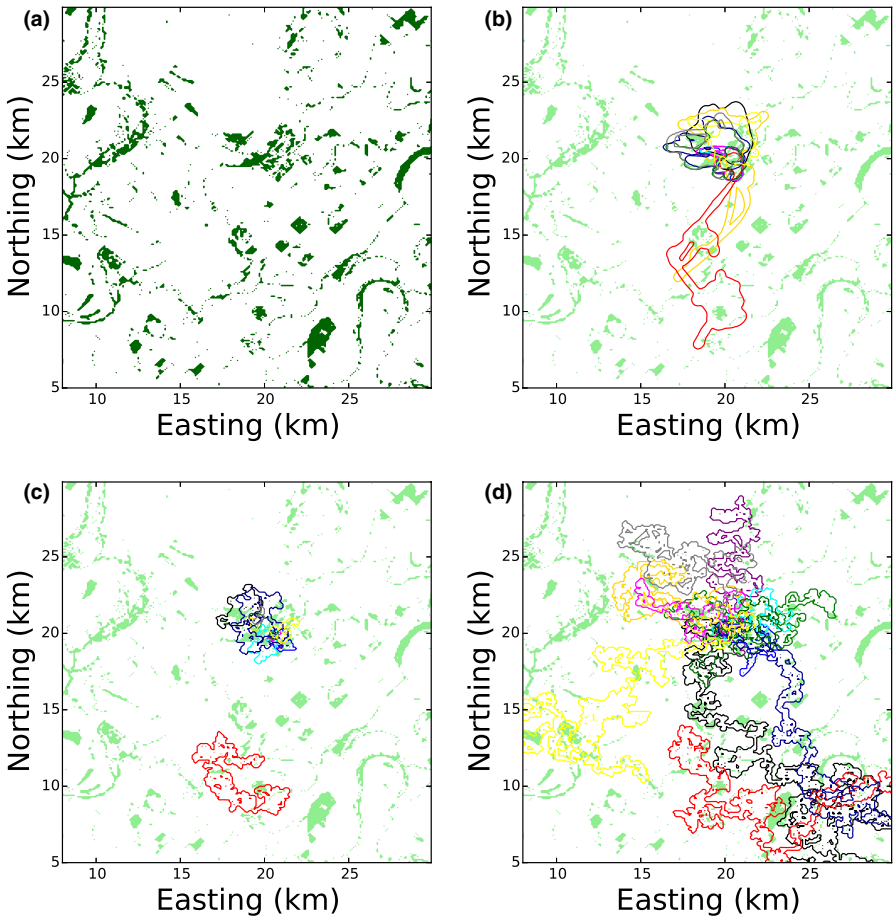

FIGURE 4 Incorporating attraction to forest. Panel (a) shows the distribution of forests in and around the study area. Panel (b) super-imposes (onto the forest distribution) the empirical OD for the period from 5 May 2016 to 3 June 2016. Panels (c and d) show the predicted space use for the same period from the Forest Model and the Just Forest Model, respectively.

between the ODs of animals, meaning one cannot say that one OD is always the explanatory variable and the other the response variable. This is a key difference between statistical and mechanistic modelling. Thus, we require different techniques for assessing goodness-of-fit. Here, we have proposed two: comparing OD overlaps using BA, and OD size via $\overline{Q}$. Such a comparison is similar in flavour to pattern oriented modelling, where one tunes an IBM so that specific summary statistics match the data (Grimm & Railsback, 2012) [which in turn has similarities to approximate Bayesian computation (Sunnåker et al., 2013)]. Here, however, we parametrise our models directly from movement data, rather than by parametrising models by comparing patterns. Instead, we use the patterns (BA and OD size) to assess both goodness-of-fit and predictive power.

A lot of the technical advancements in our approach arose through dealing with social interactions, particularly when finding a way to move from SSF to IBM. If one is purely interested in interactions with more-or-less static environmental covariates, there are existing techniques that may be employed instead (Fieberg et al., 2021; Signer et al., 2017), and these can be combined with assessment of BA and $\overline{Q}$ for assessing predictive power. However, whilst environmental predictors occupy the bulk of SSA studies (as they are simpler to characterise), an increasing number are uncovering the effect of social interactions (Latombe et al., 2014; Schlägel et al., 2019; Swanson et al., 2016; Vanak et al., 2013). Therefore, it is an important research frontier to understand how such interactions affect space use patterns. Our methods should help researchers progress this frontier.

A key challenge when modelling social interactions is that often one does not have every interacting individual tagged contemporaneously. Such questions have been addressed in contexts away from SSA (Calabrese et al., 2018; Niu et al., 2016) and these methods may provide valuable fodder for future advancements in this direction. If one has several animals with roughly similar home ranges then it may not matter greatly if a small subset are missing from movement data (one could check this by subsampling the data, but the effort of redoing the IBM analysis for every subsample would not be trivial). A more critical situation might be where animals are territorial, and a missing animal might mean that there is a gap in the terrain that appears (in the data) not to be occupied but in fact is occupied. Here, one possible way forward might be to take advantage of the fact that we have an IBM, by using this to simulate data for the missing individual (perhaps by using parameter values averaged across the other individuals). After all, an advantage of a mechanistic model is that you can simulate things for which you have no data. However, we expect that exactly how to account for missing individuals will vary depending on the study system, and cracking this problem will require significant future work.

In conclusion, we have proposed an iterative model development cycle for (a) building models that are as predictive as possible given a dataset and (b) using the concept of prediction to inform model

improvement. Whilst our case study was kept deliberately simple, to elucidate the methodology, it provides a fundamental tool for determining the requisite building blocks of detailed, predictive models of animal space use.

AUTHORS' CONTRIBUTIONS

J.R.P., G.M.S. and L.B. conceived the research idea; J.R.P. designed and performed the data analysis and modelling; G.M.S. and B.K.S. provided the data; J.R.P. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

We thank C. D. Gibson, D. H. Foster and K. E. Paolini for the capturing and collaring efforts. Funding for this work was provided through a non-assistance cooperative agreement with the United States Department of Agriculture—Agricultural Research Service (award #58-0200-0-002). This work was additionally supported by the Mississippi Agricultural and Forestry Experiment Station, the Mississippi State University Extension Service and the National Feral Swine Damage Management Program of the United States Department of Agriculture—Animal and Plant Health Inspection Service/Wildlife Services. We thank two anonymous reviewers and an associate editor for constructive comments that helped improve our manuscript.

CONFLICT OF INTEREST

None of the authors have a conflict of interest to declare.

DATA AVAILABILITY STATEMENT

The Cropland Data Layer is publicly available (https://nassgeodata.gmu.edu/CropScape/). The pig relocation data is available on Figshare (Potts et al., 2022).

ORCID

Jonathan R. Potts https://orcid.org/0000-0002-8564-2904
Luca Börger https://orcid.org/0000-0001-8763-5997
Garrett M. Street https://orcid.org/0000-0002-1260-9214

REFERENCES

Ames, W. F. (2014). *Numerical methods for partial differential equations*. Academic Press.

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.

Ballari, S. A., & Barrios-García, M. N. (2014). A review of wild boar *Sus scrofa* diet and factors affecting food selection in native and introduced ranges. *Mammal Review*, *44*, 124–134.

Bateman, A. W., Lewis, M. A., Gall, G., Manser, M. B., & Clutton-Brock, T. H. (2015). Territoriality and home-range dynamics in meerkats, *Suricata suricatta*: A mechanistic modelling approach. *Journal of Animal Ecology*, *84*, 260–271.

Beyer, H. L., Gurarie, E., Börger, L., Panzacchi, M., Basille, M., Herfindal, I., Van Moorter, B., Lele, S. R., & Matthiopoulos, J. (2016). 'You shall not pass!': Quantifying barrier permeability and proximity avoidance by animals. *Journal of Animal Ecology*, *85*, 43–53.

Burnham, K. P., & Anderson, D. R. (2001). Kullback-leibler information as a basis for strong inference in ecological studies. *Wildlife Research*, *28*, 111–119.

Calabrese, J. M., Fleming, C. H., Fagan, W. F., Rimmler, M., Kaczensky, P., Bewick, S., Leimgruber, P., & Mueller, T. (2018). Disentangling social interactions and environmental drivers in multi-individual wildlife tracking data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*, 20170007.

Calabrese, J. M., Fleming, C. H., & Gurarie, E. (2016). ctmm: An R package for analyzing animal relocation data as a continuous-time stochastic process. *Methods in Ecology and Evolution*, *7*, 1124–1132.

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, *9*, 129–136.

Cozzi, G., Maag, N., Börger, L., Clutton-Brock, T. H., & Ozgul, A. (2018). Socially informed dispersal in a territorial cooperative breeder. *Journal of Animal Ecology*, *87*, 838–849.

Ellison, N., Hatchwell, B. J., Biddiscombe, S. J., Napper, C. J., & Potts, J. R. (2020). Mechanistic home range analysis reveals drivers of space use patterns for a non-territorial passerine. *Journal of Animal Ecology*, *89*, 2763–2776.

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.

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.

Francis, T. B., Abbott, K. C., Cuddington, K., Gellner, G., Hastings, A., Lai, Y. C., Morozov, A., Petrovskii, S., & Zeeman, M. L. (2021). Management implications of long transients in ecological systems. *Nature Ecology & Evolution*, *5*, 285–294.

Franklin, J. (2010). *Mapping species distributions: Spatial inference and prediction*. Cambridge University Press.

Ganusov, V. V. (2016). Strong inference in mathematical modeling: A method for robust science in the twenty-first century. *Frontiers in Microbiology*, *7*, 1131.

Giuggioli, L., Potts, J. R., & Harris, S. (2011). Animal interactions and the emergence of territoriality. *PLoS Computational Biology*, *7*, e1002008.

Graham, C. H., VanDerWal, J., Phillips, S. J., Moritz, C., & Williams, S. E. (2010). Dynamic refugia and species persistence: Tracking spatial shifts in habitat through time. *Ecography*, *33*, 1062–1069.

Grimm, V., & Railsback, S. F. (2012). Pattern-oriented modelling: A 'multi-scope' for predictive systems ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 298–310.

Hao, T., Elith, J., Guillera-Arroita, 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.

Hastings, A., Abbott, K. C., Cuddington, K., Francis, T., Gellner, G., Lai, Y. C., Morozov, A., Petrovskii, S., Scranton, K., & Zeeman, M. L. (2018). Transient phenomena in ecology. *Science*, *361*(6406). https://doi.org/10.1126/science.aat6412

Holloway, P., & Miller, J. A. (2017). A quantitative synthesis of the movement concepts used within species distribution modelling. *Ecological Modelling*, *356*, 91–103.

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

Lewis, M., & Moorcroft, P. (2006). *Mechanistic home range analysis*. Princeton University Press.

Lewis, M. A., Petrovskii, S. V., & Potts, J. R. (2016). *The mathematics behind biological invasions*. Springer.

Macdonald, D. W., & Rushton, S. (2003). Modelling space use and dispersal of mammals in real landscapes: A tool for conservation. *Journal of Biogeography*, *30*, 607–620.

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.

Mayer, J., & Brisbin, I. L. (2009). *Wild pigs: Biology, damage, control techniques, and management*. Tech. rep., Savannah River Ecology Laboratory.

McClure, M. L., Burdett, C. L., Farnsworth, M. L., Lutman, M. W., Theobald, D. M., Riggs, P. D., Grear, D. A., & Miller, R. S. (2015). Modeling and mapping the probability of occurrence of invasive wild pigs across the contiguous United States. *PLoS One*, *10*, 1–17.

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). https://doi.org/10.1111/oik.03356

Merkle, J. A., Sawyer, H., Monteith, K. L., Dwinnell, S. P., Fralick, G. L., & Kauffman, M. J. (2019). Spatial memory shapes migration and its benefits: Evidence from a large herbivore. *Ecology Letters*, *22*, 1797–1805.

Michelot, T., Gloaguen, P., Blackwell, P. G., & Étienne, M. P. (2019). The langevin diffusion as a continuous-time model of animal movement and habitat selection. *Methods in Ecology and Evolution*, *10*, 1894–1907.

Miller, J. A., & Holloway, P. (2015). Incorporating movement in species distribution models. *Progress in Physical Geography*, *39*, 837–849.

Moorcroft, P., Lewis, M., & Crabtree, R. (1999). Home range analysis using a mechanistic home range model. *Ecology*, *80*, 1656–1665.

Moorcroft, P. R., Lewis, M. A., & Crabtree, R. L. (2006). Mechanistic home range models capture spatial patterns and dynamics of coyote territories in yellowstone. *Proceedings of the Royal Society. Biological Sciences*, *273*, 1651–1659.

Morozov, A., Abbott, K., Cuddington, K., Francis, T., Gellner, G., Hastings, A., Lai, Y. C., Petrovskii, S., Scranton, K., & Zeeman, M. L. (2020). Long transients in ecology: Theory and applications. *Physics of Life Reviews*, *32*, 1–40.

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.

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.

Niu, M., Blackwell, P. G., & Skarin, A. (2016). Modeling interdependent animal movement in continuous time. *Biometrics*, *72*, 315–324.

Pagel, J., & Schurr, F. M. (2012). Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, *21*, 293–304.

Paolini, K. E., Strickland, B. K., Tegt, J. L., VerCauteren, K. C., & Street, G. M. (2018). Seasonal variation in preference dictates space use in an invasive generalist. *PLoS ONE*, *13*, 1–18.

Paolini, K. E., Strickland, B. K., Tegt, J. L., VerCauteren, K. C., & Street, G. M. (2019). The habitat functional response links seasonal third-order selection to second-order landscape characteristics. *Ecology and Evolution*, *9*, 4683–4691.

Platt, J. R. (1964). Strong inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science*, *146*, 347–353.

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*(96), 20140333.

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

Potts, J. R., Börger, L., Strickland, B. K., & Street, G. M. (2022). Data from: Assessing the predictive power of step selection functions: How social and environmental interactions affect animal space use. *Figshare*, https://figshare.com/s/c79fcfa7ae2739807793

Potts, J. R., & Lewis, M. A. (2014). How do animal territories form and change? Lessons from 20 years of mechanistic modelling. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20140231.

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.

Rempel, R., & Rodgers, A. (1997). Effects of differential correction on accuracy of a GPS animal location system. *Journal of Wildlife Management*, *61*, 525–530.

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.

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., & Dessimoz, C. (2013). Approximate Bayesian computation. *PLoS Computational Biology*, *9*, e1002803.

Swanson, A., Arnold, T., Kosmala, M., Forester, J., & Packer, C. (2016). In the absence of a 'landscape of fear': How lions, hyenas, and cheetahs coexist. *Ecology and Evolution*, *6*, 8534–8545.

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.

Vázquez Diosdado, J. A., Barker, Z. E., Hodges, H. R., Amory, J. R., Croft, D. P., Bell, N. J., & Codling, E. A. (2018). Space-use patterns highlight behavioural differences linked to lameness, parity, and days in milk in barn-housed dairy cows. *PLoS ONE*, *13*, e0208424.

Wang, Y., Blackwell, P., Merkle, J., & Potts, J. (2019). Continuous time resource selection analysis for moving animals. *Methods in Ecology and Evolution*, *10*, 1664–1678.

Wang, Y. S., & Potts, J. R. (2017). Partial differential equation techniques for analysing animal movement: A comparison of different methods. *Journal of Theoretical Biology*, *416*, 52–67.

Wasserman, L. (2004). *All of statistics: A concise course in statistical inference* (Vol. *26*). Springer.

Wood, K. A., Stillman, R. A., & Hilton, G. M. (2018). Conservation in a changing world needs predictive models. *Animal Conservation*, *21*, 87–88.

Zimmermann, N. E., Edwards, T. C., Jr., Graham, C. H., Pearman, P. B., & Svenning, J. C. (2010). New trends in species distribution modelling. *Ecography*, *33*, 985–989.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: 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*, *00*, 1–14. https://doi.org/10.1111/2041-210X.13904