

# Transferring decision boundaries onto a geographic space: Agent rules extracted from movement data using classification trees

Jugal Patel<sup>1</sup>  | Jeffrey Katan<sup>2</sup> | Liliana Perez<sup>2</sup>  | Raja Sengupta<sup>1</sup> 

<sup>1</sup>Department of Geography, McGill University, Montréal, QC, Canada

<sup>2</sup>Department de géographie, Université de Montréal, Montréal, QC, Canada

## Correspondence

Jugal Patel, Department of Geography, McGill University, Montreal, QC H3A 0G4, Canada.

Email: jugal.patel@mail.mcgill.ca

## Abstract

We leverage applied machine learning to determine which environmental features are best associated with the “moving” behaviour(s) of a troop of olive baboons (*Papio anubis*; collared with GPS trackers at Mpala Research Centre, Kenya). Specifically, we develop a behaviour-selection surface informed by classification trees trained using movement trajectories and remotely sensed environmental features. Atop this surface, we simulate agent movement towards set destinations, constrained by the relative extent to which sets of features are associated with behaviour(s). To achieve our goal, we perform: (a) path segmentation using thresholding to label training data; (b) agent-rule extraction using classification trees to associate the relative Euclidean distance of a point from environmental features with behaviour; and (c) implementation of this information into an agent-based model to provide a data-driven simulation of troop movement. We believe this framework can accommodate intensifications in data velocity, veracity, volume, and variety expected from increasingly sophisticated biologists and data-fusion techniques.

## 1 | INTRODUCTION

Animal movement is increasingly being logged and availed using various technologies, ranging across both ecological systems and spatiotemporal scale. For example, as of January 2021, 2.4 billion locations across 1,025 taxa are recorded on Movebank.org—a popular open access repository for sharing animal trajectory data (Kranstauber

et al., 2011). The availability of these data (both in terms of collection and dissemination) represents potential for new methods to complement traditional techniques for modelling animal movement. Calls for an “integrated science of movement”, as well as an “integrated biologging framework”, have made clear the continued contribution of geographers to animal movement ecology research. These include data-oriented methods for movement ecology with explicit considerations for geospatial processes, and a broader overlap and methods exchange amongst applied movement domains of human mobility and animal movement ecology (Demšar, Long, & Siła-Nowicka, 2020; Miller, Dodge, Miller, & Bohrer, 2019; Williams et al., 2020).

Four key components are thought to underlie the movement of individual organisms: (a) internal state; (b) motion capacity; (c) navigation capacity; and (d) external (often contextualized as environmental) factors (Holyoak, Casagrandi, Nathan, Revilla, & Spiegel, 2008). Motion, navigation, and environmental factors are identifiable via either trajectories, remotely sensed capture of the scene, or a combination of the two. For example, motion manifests as the direction, magnitude, and periodicity of movement, all of which can be extracted from time-series location information (Demšar et al., 2015; Long & Nelson, 2013; Long, Nelson, & Wulder, 2010). Additionally, it is common practice to integrate telemetry-based movement data with spatial datasets (e.g., percentage canopy cover or elevation) to identify external factors affecting movement across environments. Using fine-grained spatiotemporal data, the field has evolved to develop understanding of individual mechanisms (e.g., spatio-cognitive memory and internal time measures) that result in emergent movement patterns and behaviour (Nathan et al., 2008).

In parallel, agent (or individual)-based models (ABMs/IBMs) have been developed and designed to model dynamics in complex systems, and are well equipped to explore variation in animal movement patterns, contextualize their linkage to ecological processes, and simulate expected outcomes (Anderson, Downs, Loraamm, & Reader, 2017; Bonnell, Chapman, & Sengupta, 2016; Bonnell, Ghai, Goldberg, Sengupta, & Chapman, 2016; DeAngelis & Diaz, 2019; Holloway, 2018; Long & Nelson, 2012; Pérez & Dragičević, 2011). Representing dynamics as they relate to individual behaviour and interactions within an environment (including non-linear interactions of the environment and interactions amongst individuals) can be explicit with an agent-based modelling framework. Modelling movement ecology using object-oriented data structures to represent and simulate outcomes resulting from interactions of individuals, conspecifics, sympatric species, and humans is demonstrated by “TIGMOD”: an early example of an agent-based movement model with explicit considerations for: (a) space as the primary medium for interaction; and (b) individual heterogeneity (Ahearn, Smith, Joshi, & Ding, 2001). More recently, agent-based models have been developed with the intent of considering contextual information in new ways, and providing alternative simulation methods for animal trajectory data (Anderson et al., 2017; Diaz et al., 2021). Specifically, an African elephant (*Loxodonta africanus*) spatially explicit agent-based model showcased that simulating high-fidelity trajectories is possible using “a resource-driven model with relatively simple decision rules ...”. As well, a Muscovy duck agent-based model was also developed to offer an alternative simulation technique to correlated random walks. Collective decision-making, whereby individual actions lead to the emergence of coordinated movement(s), is an ideal area of research for agent-based modelling frameworks to aid in uncovering how individual interactions and variations in behaviour lead to emergent characteristics (Cook et al., 2020; Hawkes, 2009; Hertel, Niemelä, Dingemanse, & Mueller, 2020; Kennedy, Zerbini, Rone, & Clapham, 2014; King & Sueur, 2011).

Agent-based models can be intuitive and readily interpretable simulations of dynamics involved in animal movement (Tang & Bennett, 2010). However, when agent-based simulations of animal movement are implemented, they rely on insights provided by biologists or obtained from the literature. Relying exclusively on expert insight results in two issues: (a) underlying factors cannot be considered separately, as institutionalized ethological and ecological knowledge is interconnected with countless unperceived or understudied phenomena as well as the researcher's worldview; and (b) animal ecology as a field has a historic focus on explicitly natural (as opposed to coupled) systems (Martin, Blossey, & Ellis, 2012). Tight coupling between what are commonly referred to as “human” (or social) and “natural” (or ecological) systems has been documented across fields. Geographic Information Science (GIScience) researchers working in social and ecological science domains have clarified systems coupling using common terms rooted in complex systems science (Bennett & McGinnis, 2008;

Cenek & Franklin, 2017; Liu et al., 2007). Fundamentally, animal movement (research) is not divorced from human interpretation, human behaviour or interaction, perception of the environment, and/or environmental policy (Cresswell, 2011, 2012, 2014; Semeniuk, Haider, Cooper, & Rothley, 2010). The importance of explicitly considering tight coupling of natural and social systems is reinforced as well-informed wildlife policy often inadvertently regulates human livelihoods (Beaumier, Ford, & Tagalik, 2015; Bennett & McGinnis, 2008; Liu et al., 2007; Perez, Kim, & Sengupta, 2016; Tyrrell, 2007). Agent-based models can provide a framework for understanding animal movement as a key emergent characteristic of socio-ecological complex adaptive systems. Finally, as agent-based models are often spatially explicit, integration is generally possible with any information that can be spatialized and digitized into a Geographic Information System (GIS). Our method aims to provide a core component of an adaptive agent-based movement ecology model that can appropriately consider and integrate various forms of spatial information.

Algorithmic methods can offer considerable advantages when extracting movement rationales from environmental or trajectory data. For example tree-based algorithms, like classification and regression trees, can be used to enable modelling frameworks that make no assumptions about how embedded processes become observable data; nor do they require extensive pre-processing steps or data transformations (Breiman, 2001). Classification trees are recursive algorithms ideally suited to explore data structures as well as analyse complex ecological data (Loh, 2011). Interesting work using such technologies to identify potential environmental factors underlying animal movement, determine population distributions, or predict zoonotic disease transmission risk is already well documented (Ahearn, Dodge, Simcharoen, Xavier, & Smith, 2017; Elith, Leathwick, & Hastie, 2008; Han, Schmidt, Bowden, & Drake, 2015; Leathwick, Elith, Francis, Hastie, & Taylor, 2006; Oloo & Wallentin, 2017; Torrens, Li, & Griffin, 2011; Ward, Evans, & Malleson, 2016). With respect to using these technologies to inform agent-based models, notable examples include: context-sensitive random walks that incorporate local external factors to simulate the movement of tigers at Royal Chitwan National Park in Nepal (Ahearn et al., 2001); genetic algorithms to simulate representative relative-turn angles and step-distance of homing pigeons (Oloo & Wallentin, 2017); reinforcement learning to contextualize the risk and reward of agent behaviour (Sutton & Barto, 1999; Tang & Bennett, 2010); and artificial neural networks to assign weights to link environmental features to an agent's internal spatially explicit map of its surroundings (Huse, Strand, & Giske, 1999; Strand, Huse, & Giske, 2002).

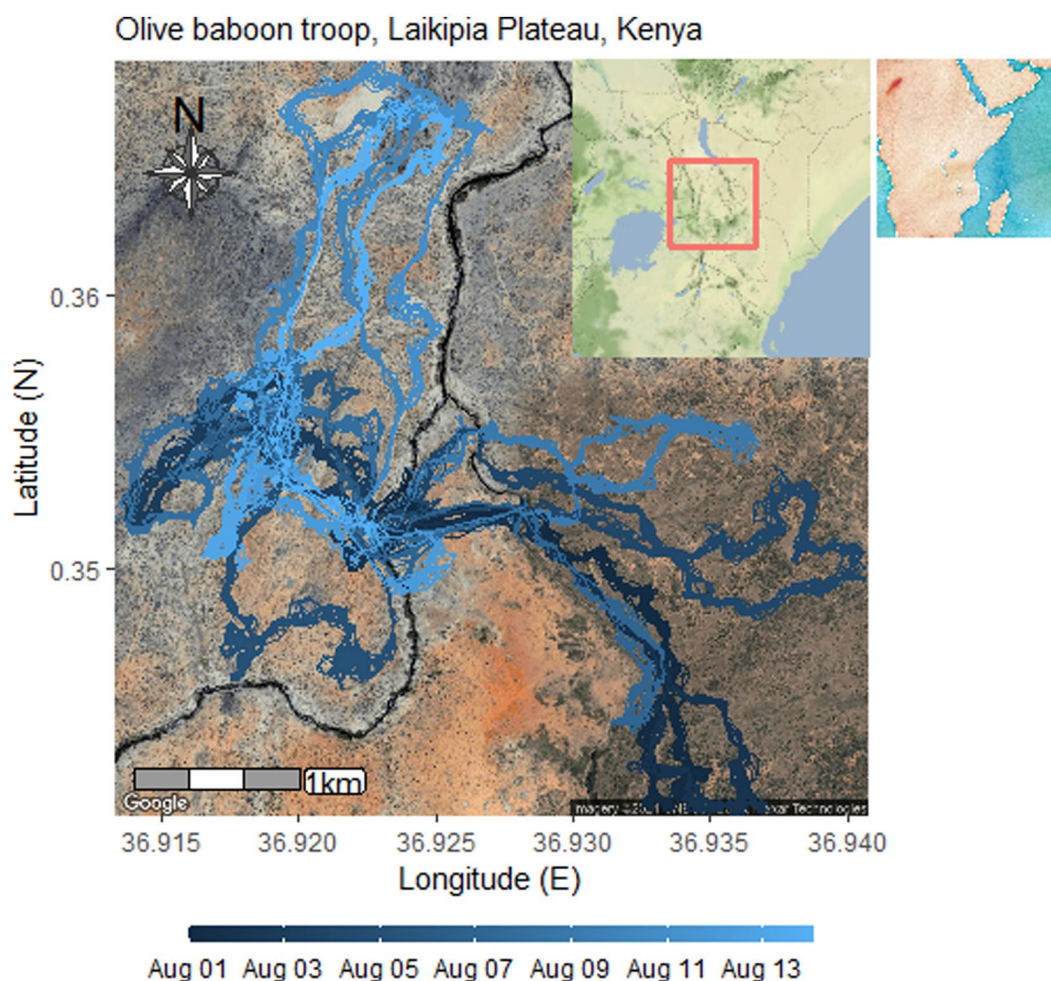
To demonstrate how algorithmic methods can uncover and explicitly consider environmental features as external factors underlying animal movement, we use two simple artificial intelligence technologies: (a) classification trees to extract rules that associate behaviours with environmental features; and (b) agent-based modelling to build a bottom-up simulation of movement based on extracted rules and environmental heterogeneity. Showcasing how high spatiotemporal trajectory data allow for decision boundaries in classification trees to be transferred onto a continuous surface for agent simulation is the core objective of this article.

## 2 | METHODS

Strandburg-Peshkin, Farine, Couzin, and Crofoot (2015) obtained data from 26 olive baboons (14 adults, 10 sub-adults, and 2 large juveniles) with GPS collars (e-Obs Digital Telemetry, Gruenwald, German; with a reported average positional error of 0.26 m) between 1 and 14 August 2012 in collaboration with the Mpala Research Centre, Laikipia Plateau, Kenya. The locations of these olive baboons were logged with a relatively high temporal frequency (1 Hz: 1 record per second) continuously during daytime hours, from 0600 to 1,800 hr local time. We obtained this data from Movebank (ID: 7023252) (Strandburg-Peshkin et al., 2015). The study description noted that not all GPS collars transmitted location data successfully throughout each day, resulting in ~10 million observations of individual baboon locations. We cleaned and spatially projected the collected trajectories with tools provided in the *tidyverse*, *lubridate*, *sp*, and *rgdal* packages for R, removing 5,758 missing observations (of 10,402,385 total observations; ~0.054%) and projecting trajectories to Universal Transverse Mercator (UTM)

zone 37 North for Mpala, Kenya (Bivand, Keitt, Rowlingson, & Pebesma, 2016; Bivand, Pebesma, & Gomez-Rubio, 2013; R Core Team, 2020; Wickham et al., 2019). We set aside days 3, 6, 9, and 11 and do not include these trajectories in building our classification tree or parameterizing the agent-based model. Keeping entire days as validation (as opposed to a random sample) enables more meaningful tests of our simulation. Specifically, as the troop visit different destination sites throughout the 2-week period, keeping multiple days aside provides validation across different destinations. Figure 1 shows the observed trajectories of the olive baboon troop at Mpala Research Centre in Laikipia, Kenya.

We extract environmental features from the scene (using a normalized difference vegetation index to define vegetation and clearings, and on-screen digitization for human trails and the river network), which we contextualize as external factors underlying behaviour selection (and therefore animal movement). Using Operational Land Imager imagery aboard Landsat 8 (dated 26 October 2018) and Google Earth/Digital Globe imagery (dated 25 June 2012), we identified four environmental features: clearings (or open areas); the river; trees; and a regional network of human-made trails. Strandburg-Peshkin et al. (2015) identified road-following and short-range avoidance of dense vegetation as possible environmental factors underlying path selection of the troop. Although



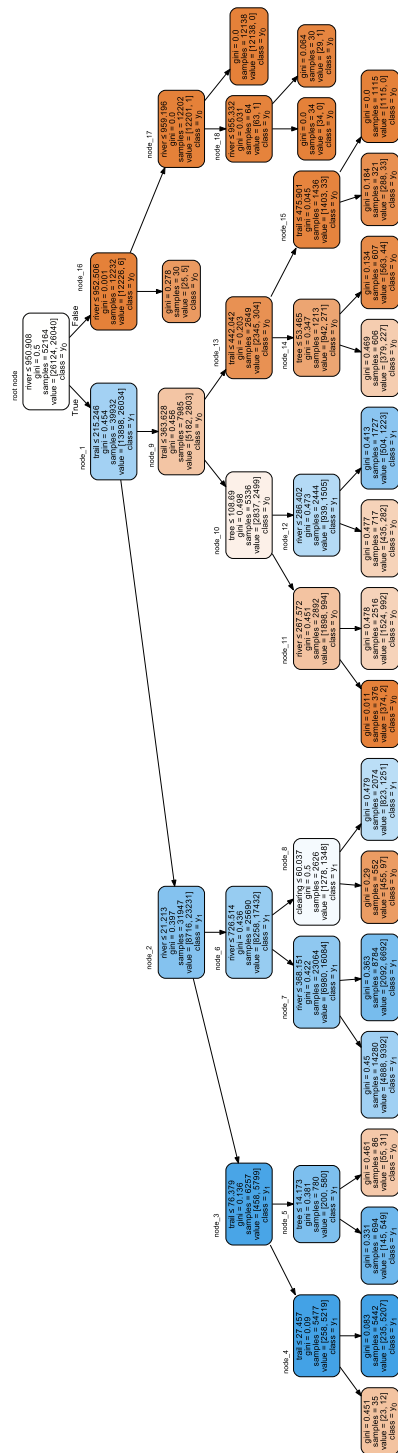
**FIGURE 1** Olive baboon (*Papio anubis*) trajectories collected at the Mpala Research Centre in Kenya are displayed with a blue gradient denoting time. Google imagery used for satellite and Kenya inset maps. Stamen watercolour imagery used for regional inset map of Africa

innovative remote sensing products exist to further identify features, we use just these four readily identifiable features to help demonstrate and highlight the key contribution of this work: a framework for agent-based modelling that explicitly considers environmental heterogeneity as well as uncertainty in unambiguous ways to model path selection. To link trajectory and environmental information, we calculate the Euclidean distance between environmental features and each GPS logged position. This expression of environmental heterogeneity (at each point) was then linked with different (theoretical) behavioural states of the animal using path segmentation (Edelhoff, Signer, & Balkenhol, 2016). The statistical signal provided with this linkage is reduced to the spatial resolution of environmental features (i.e., 3 m spatial resolution of remote sensing products used to collect features).

We threshold our trajectories into two broad behaviour states: moving and sedentary. To perform this segmentation, we use the *adehabitat* package for R to collect physical characteristics at each logged position (Calenge, 2006). Specifically, we consider that any positions with velocity greater than 3 m/s indicate moving behaviours. Sedentary points, those recorded with velocity below 3 m/s, are in a category of behaviours which would require improved spatial resolution of environmental features to extract decision boundaries (relevant to sedentary behaviours). We are only able to consider velocity with such ease due to GPS positions having been logged at a relatively high temporal resolution (1/s). To clarify, our environmental raster layers' spatial resolution is 3 m. And when individual baboons are moving at less than 3 m/s, we cannot make clear associations with environmental data. Of the logged positions in the training dataset, 32,707 observations occurred while baboons in the troop were travelling at greater than 3 m/s. To maintain class balance, we generate 32,707 new random points within the bounding box of troop movements—the equivalent of a “null” model. Figure 2 shows the resultant classification tree.

We use this subset (both null and moving points, ~64,000 observations) to build a classification tree using *scikit learn* (Pedregosa, Weiss, & Brucher, 2011). Our classification tree is parameterized using the Gini impurity as the splitting criterion, a minimum sample size of 20 observations, a maximum depth of five layers, and a minimum impurity decrease of  $1e^{-7}$ . Each split (of the tree) or partition (in data) only occurs if the Gini impurity decreases by at least  $1e^{-7}$  units. It stops splitting once nodes have fewer than 20 observations (“samples”), or once the tree reaches a depth of five layers. These stoppage conditions are common “pruning” techniques for classification and regression trees (Fournier & Cremilleux, 2013). Without these parameters, default values allow the tree to keep building until it has overfit to training data (e.g., default for minimum impurity split is 0, meaning subset data would not need to become more “pure” to continue splitting). The classification tree queries data with partition rules, then selects partitions that result in a maximum decrease in impurity. Subset impurity must decrease iteratively for the classification tree to be working. That is, the perceived homogeneity of subsets must increase as they move down the classification tree. This is due to the underlying algorithmic structure of (any type of) binary tree(s). The tree only continues if its understanding of how to best classify data is improving. Here, the quantifiable metric being minimized is the Gini impurity (as opposed to entropy or other measures in the applied machine learning literature).

Each node of the tree represents a classification made as part of iterative partitions of data. As well, nodes provide a relative expression of uncertainty (as the complement of the nodal Gini impurity). Blue nodes suggest the classification tree interprets its input as a signal for movement. Orange nodes suggest the tree interprets, and would classify, points as null if further iterations are not possible. Local decision boundaries expressed by the tree are unidirectional and unidimensional; nodal partition rules utilize only one logical operator ( $\leq$ ) and apply to only one feature per iteration. Collectively, the classification tree is an algorithmic expression of how to optimally label observed points, based on environmental heterogeneity, as either “moving” or not. This algorithmic expression is often visualized as a binary tree, since querying a single feature using “ $\leq$ ” organizes data into at most two child subsets. For example, in Figure 2, node\_1 categorizes each point as either being  $\leq 215.246$  m or not from the edge of the trail feature. The tree selects this specific partition as its search found this decision boundary to maximize the homogeneity of its child subsets (node\_2 and node\_9). In our implementation, classification functions iteratively organize as many “moving” points together as possible. A core relation to behaviour is being made on the basis



**FIGURE 2** Classification tree with depth of 5; minimum sample size of 20 observations; minimum Gini impurity decrease by  $1e-7$  units



that training data provided to the classification tree are reflective of ecological reality. And that “moving”-type behaviours are a meaningful conceptualization with respect to olive baboon ethology and ecology.

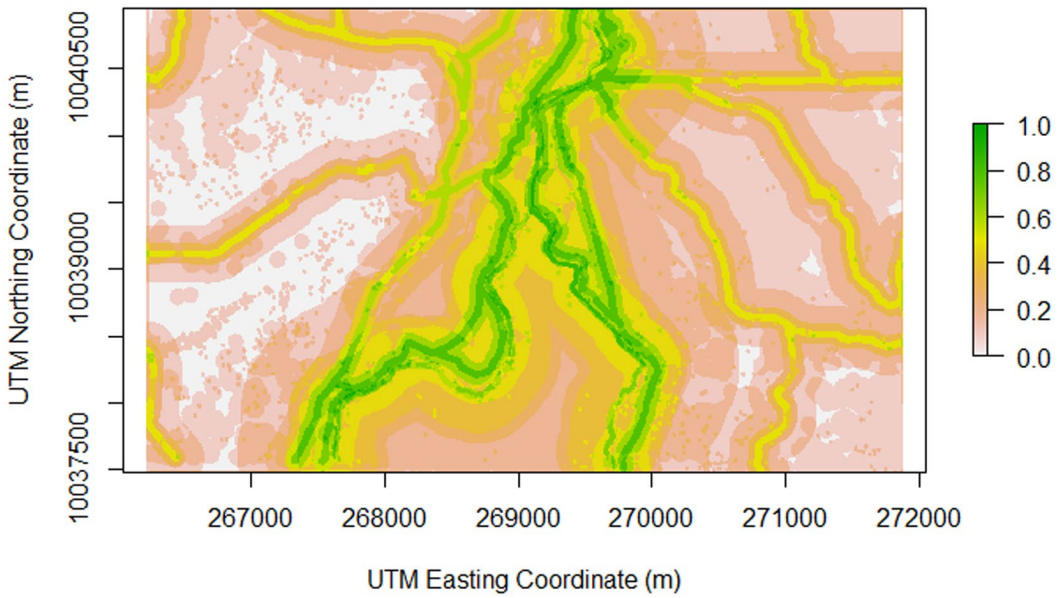
To consider these nodal rules and their associated information in aggregate, we rasterize each local decision boundary. We take the complement of the Gini impurity,  $(1 - G)$ , to quantify the relative homogeneity result of the partition. This complement is a measure of certainty being expressed at the node. The more impure subset data-sets are, the less certainty the rule can offer in classifying points. Aggregated, this becomes an understanding of certainty of classifying data as “moving” based on the relative distance a point is from key environmental features at the regional (in terms of geography) and global (in terms of computation) scale. To arrive at a spatially explicit surface, we rasterize each partition rule based on nodal information. Specifically, the distance to an environmental feature ( $dE$ ), nodal Gini impurity ( $G$ ), and partition threshold ( $pt$ ) are considered to transfer decision boundaries onto a geographic space in the following way:

$$\text{Desirability (movement selection)} = (1 - G), \text{ if } dE \leq pt \quad (1)$$

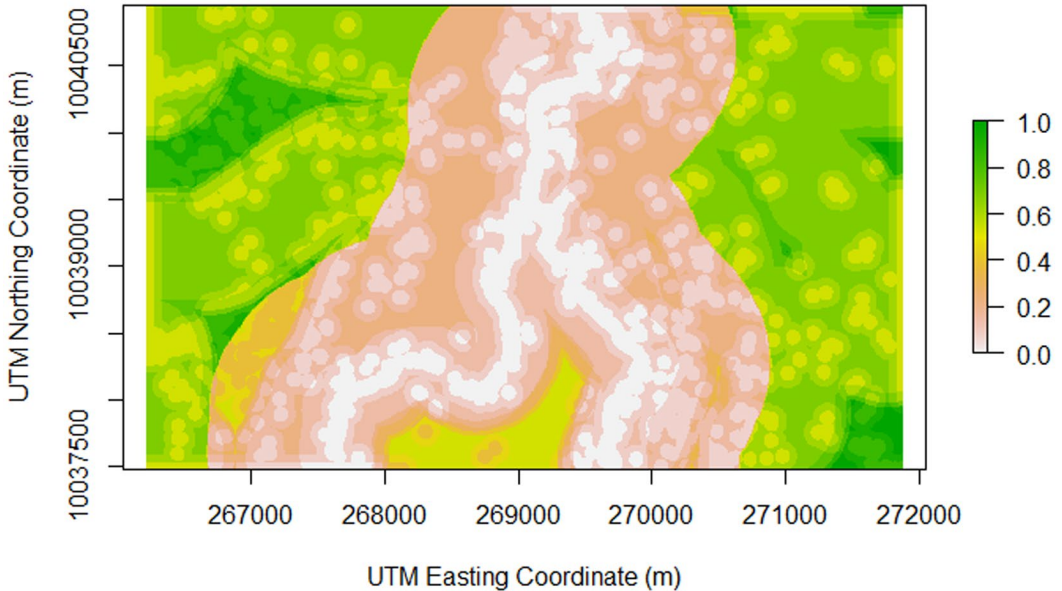
$$\text{Uncertainty (null selection)} = -(1 - G), \text{ if } dE > pt \quad (2)$$

When the partition rule is “True”, that is, when points are within the specified distance of the environmental feature that is being partitioned, we ascribe a weighting based on the complement of the Gini impurity at the specified node. If the signal provided to the training data reflects ecological reality, this surface expression is a spatially explicit consideration of environmental heterogeneity and its influence on behaviour selection. In the simplest terms, our derived behaviour selection is an understanding of where the troop would engage in any “moving” or “>3 m/s” behaviours. Uncertainty selection, expressed as the additive inverse of the complement of the Gini impurity,  $-(1 - G)$ , is used when the classification tree organizes data as not belonging to “moving” behaviours. Since both local movement and null selection are communicated via nodal information, we aggregate the complement  $(1 - G)$  to reflect movement selection, and the additive inverse of the complement  $-(1 - G)$  to reflect uncertainty. Figures 3 and 4 reflect standardized aggregations of these rasters, which we term “Desirability” and “Uncertainty” selection surfaces. Sedentary points are reflected in “patches” of uncertainty in Figure 4, as the expression of environmental heterogeneity here suggests null selection (and sedentary behaviours are likely to belong to null selection within the classification tree). As a careful note, the classification tree is not able to classify points as sedentary; it is only able to classify when environmental heterogeneity is thought to be capable of facilitating moving behaviours. For example, if a known sedentary point is provided to our classification tree, the outputs will express a relatively high measure of how uncertain it is in classifying the observation. The aggregated spatially explicit decision boundaries related to “moving” classification are expressed as a movement behaviour selection surface. For “null” classification, the surface reflects uncertain behaviour selection. Both surfaces are then standardized to maintain their distribution, while ranging from 0 to 1 (Figures 3 and 4). As such, the negative directionality of uncertain behaviour selection is recontextualized such that higher uncertainty values indicate points where the classification tree is relatively unsure of its classification of observations as belonging to “moving” behaviours. Atop these behaviour selection surfaces, we simulate agents representing baboons (“baboon-agents”) to navigate between presumed sleeping sites and destinations.

We develop our agent-based model using NetLogo (and provide its internal logic in Table 1) as it is an easy-to-use multi-agent modelling environment that can incorporate raster datasets. The model consists of one agent representing an individual baboon moving between the origin and user-set destinations. The origin and destinations were selected using kernel density estimation of sedentary segments of observed trajectories. After facing a user-set destination, baboon-agents query the movement selection surface within a  $230^\circ$  vision cone and depth of four cells. This angle and depth are also user-set but are then modulated by uncertainty expressed by the null selection surface. We used  $230^\circ$  as the relative angle within moving segments of observed trajectories fell within  $\pm 2$  rad ( $4$  rad  $\sim 230^\circ$ ). This could be an artifact of moving behaviours as they invariably have some component of



**FIGURE 3** Movement selection surface for troop of olive baboons observed during 1–14 August 2012 at Mpala Research Centre, Kenya. Green indicates areas that express environmental heterogeneity; the classification tree interprets these as spaces likely to facilitate “moving” behaviours. This surface is queried by baboon-agents for movement-related decision-making



**FIGURE 4** Uncertain or null selection surface for troop of olive baboons observed during 1–14 August 2012 at Mpala Research Centre, Kenya. Green indicates areas that express environmental heterogeneity; the classification tree interprets these spaces as likely to facilitate sedentary or “other” behaviours. This surface is queried by baboon-agents for vision cone modulation



TABLE 1 Internal movement logic of baboon-agents

I	Face destination
II	Check movement and uncertain selection surface values within the agent vision cone
III	Take one step towards the cell with greatest movement selection value
IV	Modulate agent vision cone by uncertainty
V	If location has not changed over previous timesteps, take one step directly towards the destination

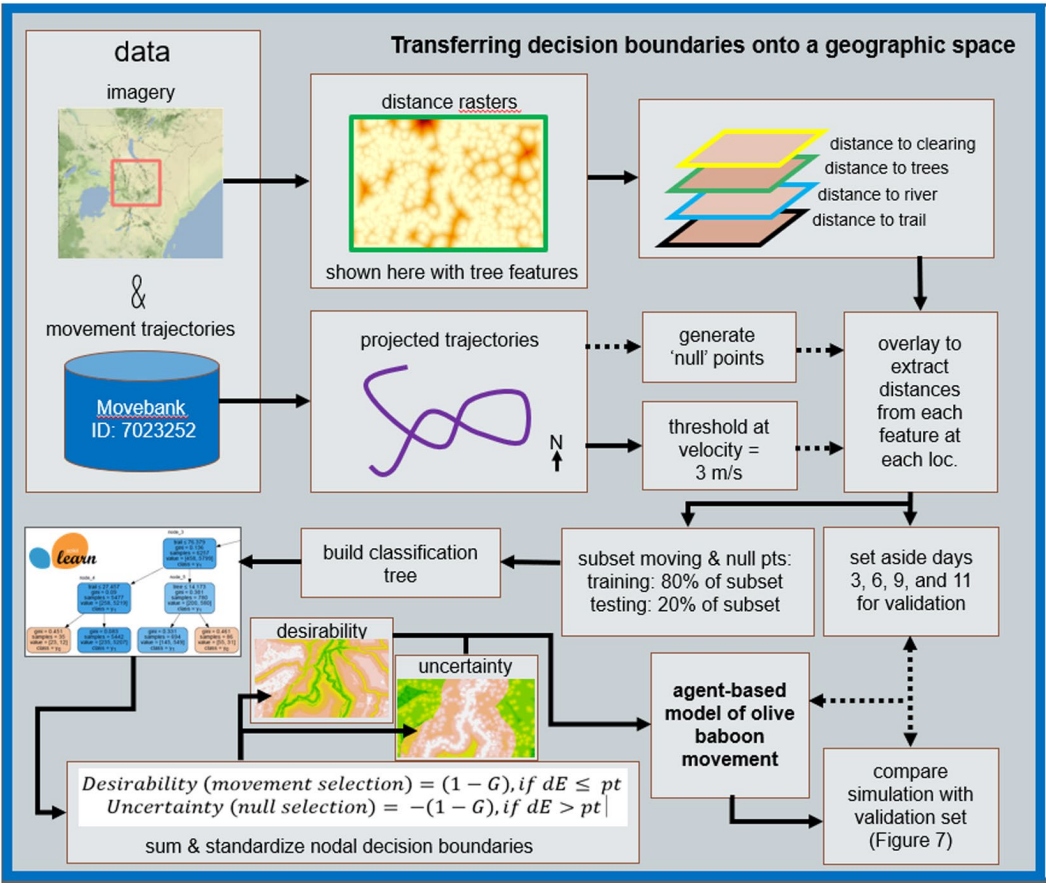
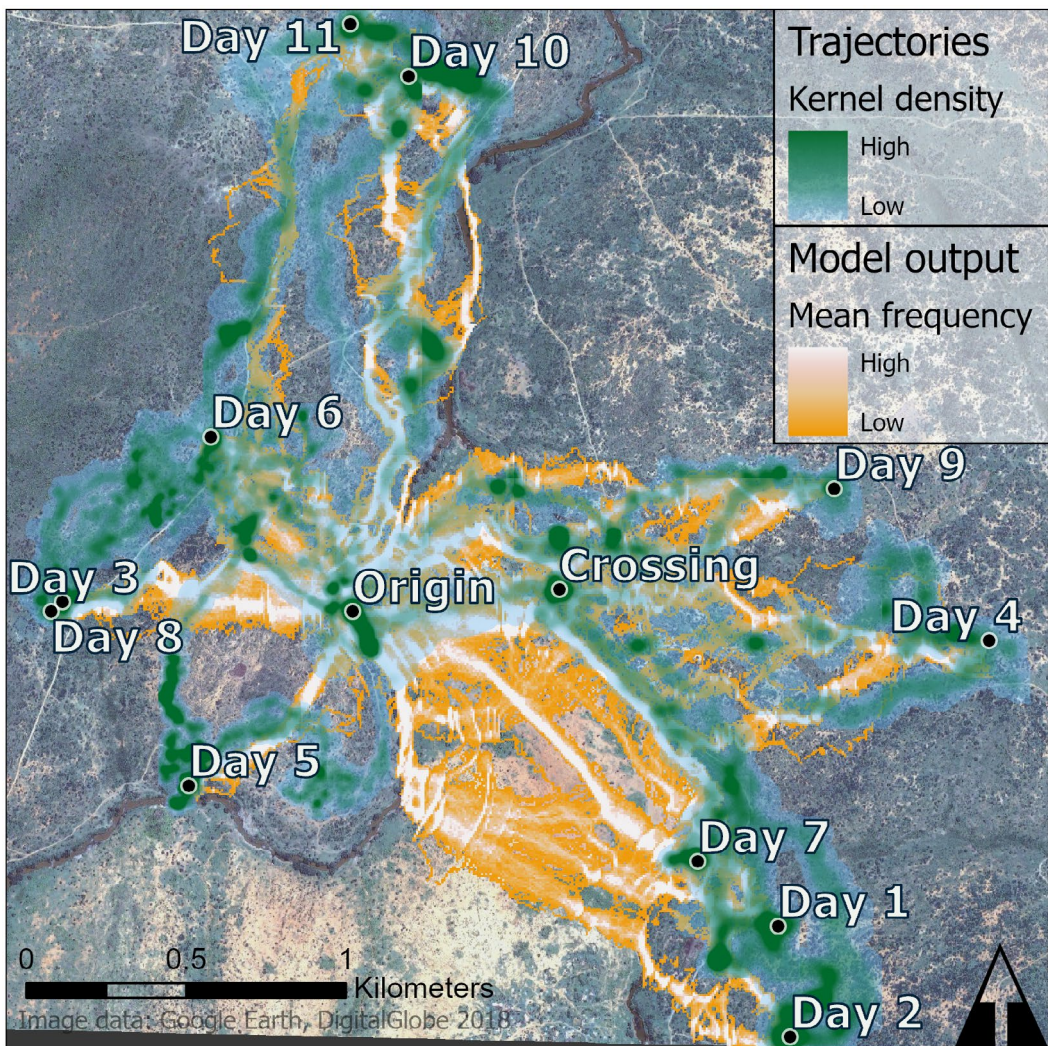


FIGURE 5 Framework for transferring decision boundaries found in classification trees onto a geographic space for spatially explicit agent-based modelling movement

directionality, or high-temporal resolution. We maintain a relatively crude understanding of directionality in our model to test the utility of selection surfaces informed by classification trees. To initiate agent movement from a starting position to an intended destination, baboon-agents query the movement selection surface within their vision cone and select the cell with the highest “movement selection”. Based on uncertainty expressed at that cell, the baboon-agent vision cone angle is multiplied by a factor of  $(1 - \text{Uncertainty})$ , and the vision cone depth by a factor of  $(1 + \text{Uncertainty})$ . Uncertainty, or null selection, here is connected to “not moving” and presumably low directionality behaviours. We also set two optional conditions which direct baboon-agents to resolve movement decisions by taking one step directly towards the destination when too many paths lead to “dead ends” (baboon-agent has moved <11 units in last 25 ticks); or when too few options force baboon-agents to continuously retrace their paths (baboon-agent has moved <1.4 units in last 5 ticks). The spatial resolution of the agent model is 6.5 m,



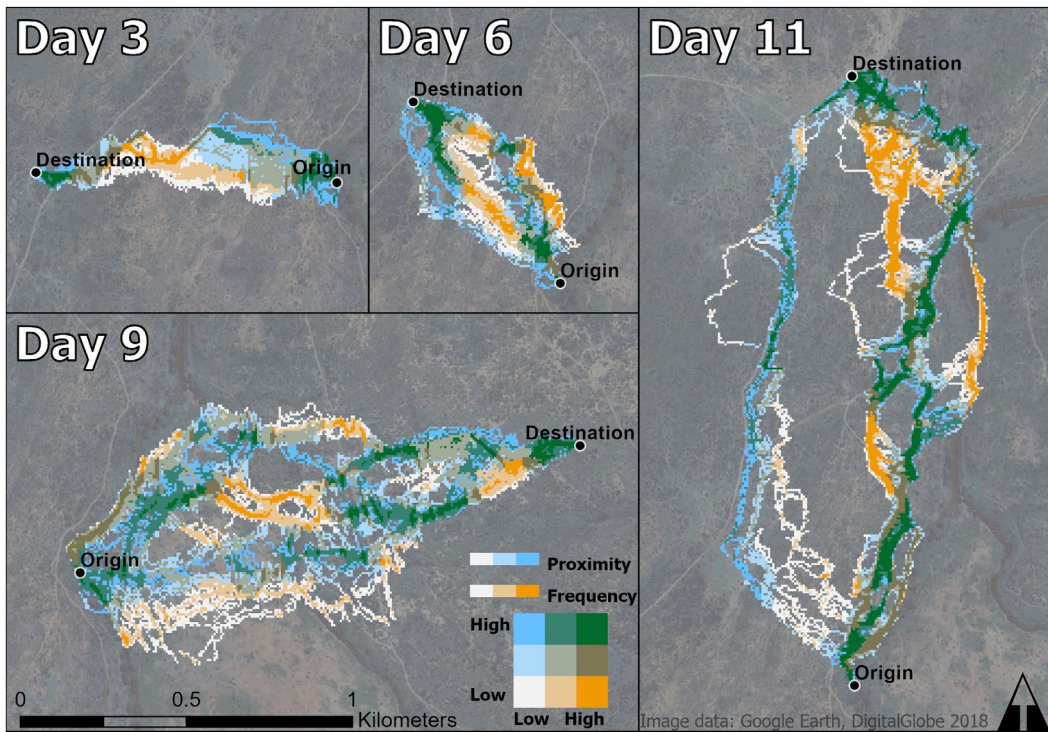
**FIGURE 6** Agent-based model simulation outputs visualized by frequency (in orange) along with kernel density estimations of observed olive baboon trajectories (in green). Note river crossing feature, which was apparent as a unique point for movement in both observed trajectories and simulation output. Basemap imagery provided by Digital Globe, Google Earth™

with temporal resolution 20 s; thus baboon-agents move at an average velocity of 0.325 m/s. To provide an ensemble of movement behaviour, we run our model 100 times for each day. As an important note, our model does not incorporate time beyond using velocity to discern training moving behaviours. Our framework for “transferring decision boundaries onto a geographic space” is visualized in Figure 5.

### 3 | RESULTS

Our procedure demonstrates how large spatiotemporal datasets with high temporal granularity can aid in developing simulations of dynamics underlying animal movement. Specifically, we show how environmental features can be explicitly incorporated into the internal movement logic of agent-based models. There are two key steps to our procedure: first, the extraction of agent rules from trajectory and environmental feature data; then, incorporation





**FIGURE 7** Agent-based model simulation outputs for validation days (3, 6, 9, and 11) visualized by frequency (in orange) and proximity to observed trajectories (in blue). Path selection highlighted by green indicates agreement between observed olive baboon movements and baboon-agent trajectories. Basemap imagery is provided by Digital Globe, Google Earth™

of this extracted information into an agent model. Extracted information for agent rules (or movement rationales) is in the form of the internal data partitions classification trees use to discern between “moving” and uncertain (or null selection) behaviours—commonly referred to as decision boundaries. These decision boundaries are aggregated into a surface atop which agents are simulated to move towards set destinations.

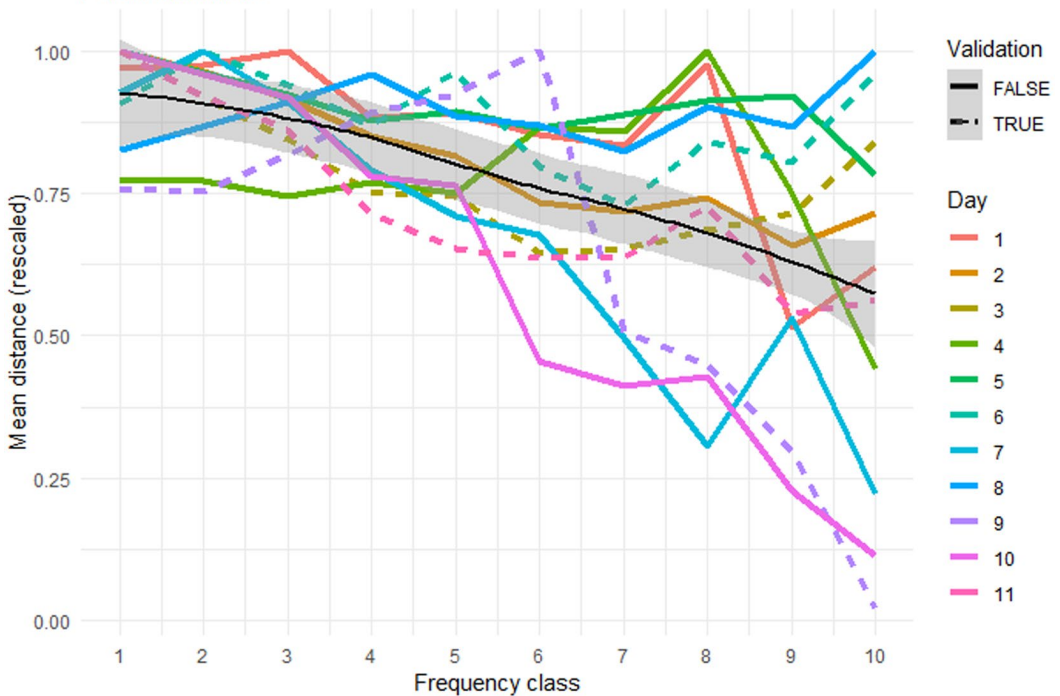
With respect to the first (information extraction) step, our classification tree achieved a classification accuracy of 80%, classification error of 20%, and sensitivity of 0.937. We interpret these results as an indication that the classification tree can discern and correctly identify moving points 80% of the time when provided with unseen point data with linkages to environmental features. Figure 2 is a visualization of the classification tree. Figure 5 describes how the output from the tree is handled, as well as how this technology is used in our framework.

With respect to the second (information implementation) step, our agent-based model is visualized in Figure 6 along with a kernel density estimate of observed trajectories. The origin, daily destinations, and a river crossing site are labelled. This river crossing site is a visibly important site in animations of troop trajectories, and interestingly is also picked up as a unique site for movement in simulations. The proximity of simulated runs to observed trajectories, as well as the frequency with which the simulation preferred a certain path, are shown in Figure 7.

For each day, we calculated the Euclidean distance from the trajectory data and superimposed it with the frequency output from our agent-based model. The distribution of movement selection values in the frequency output map has a very strong right skewness. So, we used a natural log transformation, followed by a linear rescaling, so that all days had the same range, resulting in 10 frequency classes. We were then able to compare, by location, the relation between the frequency map and its proximity to trajectory data. Figure 7 shows this validation effort, reasonably replicating real trajectories for days 9 and 11. The observed troop frequents multiple destinations each

## Distance by Frequency Class

For all simulations



**FIGURE 8** Deviation of each frequency class from observed trajectories. Lower mean distance values (y-axis) indicate overlap between observed trajectories and simulations. These values are rescaled and are a relative measure of distance. Frequency classes denote categories of cells visited by simulated trajectories, and how often trajectories use the cell. Validation days are highlighted using dashed lines. Validation days 9 and 11 deviate less than the average of all days

day throughout the 2-week data period, but especially on days 3 and 6 the troop stays in a relatively small area for most of the day. For such days, because of the way destinations are handled in our agent model (i.e., a single static destination is user-set for each day), outputs show simulated trajectories taking a “shortcut” to final destinations, instead of the more circuitous route real olive baboons took in the region.

While visually comparing the simulated and observed trajectories is helpful, they do not provide a measure of the overall performance of our model. An alternative is to quantify the deviation (as distance) between simulated and observed trajectories. The variation of distance between observed and simulated trajectory is dependent on the distance between destinations and sleeping sites, which varied greatly between each day; the further the destination, the farther “off-track” an agent could go. To better compare performance between days, the mean distance between simulation and observation of each day was rescaled by the maximum distance for that day, such that the maximum for all days is 1. Figure 8 plots these deviations for each simulated day. Here, high-frequency classes represent the cells that simulated trajectories often pass through. Low-frequency classes were relatively unfrequented cells. The trend-line plotted with  $\pm 1$  standard deviations represents an average, which improves as we consider where there are more simulated trajectories. Validation days are represented with dashed lines and confirm what Figure 6 suggested—that days 9 and 11 performed best. In other words, our framework produced simulations where baboon-agents travelled the same spaces as observed trajectories.

## 4 | DISCUSSION

Agent-based models have been used extensively to explore the impact of animal movement patterns across space and time and predict environmental outcomes. Movement rationales expressed in such models have often been based on expert knowledge about the behaviour of the species of interest (Bonnell, Chapman, et al., 2016). While neglecting input from behavioural ecologists would be disingenuous, considerable surveillance bias in animal ecology provides impetus for methods that can accommodate local knowledge as well as intensifications in data velocity, veracity, volume, and variety expected from increasingly sophisticated biologgers and data-fusion techniques (Boyce, 2006; Hebblewhite & Haydon, 2010; Martin et al., 2012; Stallknecht, 2007).

Here we have demonstrated that a large spatiotemporal dataset with high temporal granularity can be used to: (a) develop a surface to highlight environment-related selection of movement; and (b) programme an agent-based model that uses the behaviour selection surface in combination with simple rules of motion to simulate troop movements. This fits well within the broader area of work GIScientists are engaging in towards an integrated science of movement (Miller et al., 2019; Williams et al., 2020). With evermore sophisticated biologgers and data-fusion techniques there will continue to be a rapid rise in volumes of valuable movement data. Our work indicates that there is an opportunity to develop widely applicable methods to use such datasets, extract location-specific movement behaviours, and convert and explicitly consider this information in agent-based modelling frameworks.

Simulations using the described procedure could be tested against expert knowledge rationales regarding preferred habitats, avoidance of high-risk predator or disease areas, territorial defence, social behaviour, or weighted combinations of these rationales. For example, in *Where the animals go*, this troop's description suggests trajectories were collected over a 4-week period (Cheshire & Uberti, 2016). Additionally, in these unseen data, there was report of a leopard which affected troop sleeping site selection after 14 August. This unseen data could serve as further validation; our framework could be adapted to test how simulations perform under these unique conditions. Such an approach could augment expert interpretation to decipher an understanding of components underlying movement behaviour selection.

While the current procedure is demonstrated with a single classification tree, it could be scaled using ensemble techniques to consider and weight the different spaces where and times when moving groups of animals might engage in specific behaviour (e.g., seasonal drought and changes in vegetation characteristics influencing foraging behaviour). Our framework could accommodate other established representations of animal movement (like context-specific random walks) in conjunction with information from probability and uncertainty surfaces to determine path selection. The core contribution of our methodological framework is the transfer of decision boundaries explicitly expressed by classification trees onto a geographic space. Inherent aspects of this space include socio-cultural relations people have with animals, as well as embeddedness of socio-ecological complex adaptive systems. Exciting work on collecting and spatializing local knowledge suggests a wide variety of features (e.g., constructed fences, known migration routes or hunting sites) could be incorporated into our approach (Aswani & Lauer, 2006)—as long as they are ascribed spatiality in a GIS. Agent-based models, when informed by data-driven methods, are unique tools that can accommodate voluminous data with unknown or inconsistent veracity and velocity, as well as explicitly incorporate spatiality of local knowledge. There are two reasons for the overall flexibility of our approach: (a) classification trees' capacity to handle explanatory variables without data transformations; and (b) the spatially explicit nature of our agent-based modelling platform. For example, gradient descent using boosted classification trees could be used to reduce the relative “sharpness” or abruptness visible in Figures 3 and 4, and new environmental features could be identified and characterized using innovative remote sensing and hyperspectral processing technologies. The agent's logic could also be improved in future work. For example, incorporating viewshed analysis into the vision cone modulation would add towards recreating simulations of how troops determine where to go. Cost weighting, informed by either landscape genetics or geographic phenomena (such as urbanization or elevation), could be incorporated into classification tree(s) as features to influence selection surfaces; or explicitly in an agent's internal logic (e.g., for elevation, additional cost surfaces



could be fed into what information an agent is considering before moving). Finally, to reduce the determinism of our model, we could programme agents to select from a range of high-valued movement cells at random. Instead, our efforts focused on testing the informativeness of selection surfaces.

While our model adequately replicates movement behaviour selection, there has been other research in movement ecology that looks to identify behaviours within the environments in which they occur. For example, “path segmentation” refers to changes recorded in an animal's movement behaviour based on observed trajectories (Edelhoff et al., 2016). Associated methods quantitatively describe the geometric properties or physical characteristics of trajectories, which combined with time-series analysis methods can indicate changes in behaviour state. Further, innovative methods exist that could be considered for segmenting animal movement trajectories, including pattern mining and behavioural change point analysis (Zhang, Lee, & Lee, 2019). The research questions commonly addressed using path segmentation methods include: (a) quantitative description of movement patterns; (b) detection of significant change points; and (c) identification of underlying processes/hidden states. Of the latter, state-space models (e.g., hidden Markov models [HMMs]) can be built to estimate transition probabilities and test whether switching between states depends on certain habitat characteristics (Whoriskey et al., 2017). Other models, such as advanced first-difference random correlated walk (DCRW), go further by determining the complexity underlying behavioural states, and also account for temporally irregularly spaced observations and non-Gaussian errors (Jonsen, Flemming, & Myers, 2005). Admittedly, such frameworks necessarily incorporate linkages due to time—something our model does not do. However, an ensemble approach that weights information from trees based on when environmental features and trajectories are collected could function within this framework as an adequate representation of time. Interesting work using conceptualizations surrounding space-time prisms has also been demonstrated in a spatially explicit agent-based movement model (Loraamm, 2020). Methods demonstrated here are a foundation to understand the influence of environmental heterogeneity on behaviour selection. And could be understood as a tool within a broader toolkit for agent-based modelling movement ecology.

## DATA AVAILABILITY STATEMENT

Trajectory data is available on movebank.org under study ID 7023252 (Strandburg-Peshkin et al., 2015).

## ORCID

Jugal Patel  <https://orcid.org/0000-0002-1009-283X>

Liliana Perez  <https://orcid.org/0000-0002-6599-9893>

Raja Sengupta  <https://orcid.org/0000-0003-4914-5844>

## REFERENCES

- Ahearn, S., Dodge, S., Simcharoen, A., Xavier, G., & Smith, J. L. D. (2017). A context-sensitive correlated random walk: A new simulation model for movement. *International Journal of Geographical Information Science*, 31, 867–883. <https://doi.org/10.1080/13658816.2016.1224887>
- Ahearn, S., Smith, J. L. D., Joshi, A. R., & Ding, J. (2001). TIGMOD: An individual-based spatially explicit model for simulating tiger/human interaction in multiple use forests. *Ecological Modelling*, 140(1–2), 81–97. [https://doi.org/10.1016/S0304-3800\(01\)00258-7](https://doi.org/10.1016/S0304-3800(01)00258-7)
- Anderson, J. H., Downs, J. A., Loraamm, R., & Reader, S. (2017). Agent-based simulation of Muscovy duck movements using observed habitat transition and distance frequencies. *Computers, Environment and Urban Systems*, 61, 49–55. <https://doi.org/10.1016/j.compenvurbsys.2016.09.002>
- Aswani, S., & Lauer, M. (2006). Incorporating fishermen's local knowledge and behavior into geographical information systems (GIS) for designing marine protected areas in Oceania. *Human Organization*, 65, 81–102. <https://doi.org/10.17730/humo.65.1.4y2q0vhe4l30n0uj>
- Beaumier, M. C., Ford, J. D., & Tagalik, S. (2015). The food security of Inuit women in Arviat, Nunavut: The role of socio-economic factors and climate change. *Polar Record*, 51, 550–559. <https://doi.org/10.1017/S0032247414000618>
- Bennett, D., & McGinnis, D. (2008). Coupled and complex: Human–environment interaction in the Greater Yellowstone ecosystem, USA. *Geoforum*, 39, 833–845. <https://doi.org/10.1016/j.geoforum.2007.05.009>

- Bivand, R., Keitt, T., Rowlingson, B., & Pebesma, E. (2016). *rgdal: Bindings for the geospatial data abstraction library*. Retrieved from <https://rgdal.r-forge.r-project.org/>
- Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R* (2nd ed.). New York, NY: Springer.
- Bonnell, T. R., Chapman, C. A., & Sengupta, R. (2016). Interaction between scale and scheduling choices in simulations of spatial agents. *International Journal of Geographical Information Science*, 30, 2075–2088. <https://doi.org/10.1080/13658816.2016.1158822>
- Bonnell, T. R., Ghai, R. R., Goldberg, T. L., Sengupta, R., & Chapman, C. A. (2016). Spatial patterns of persistence for environmentally transmitted parasites: Effects of regional climate and local landscape. *Ecological Modelling*, 338, 78–89. <https://doi.org/10.1016/j.ecolmodel.2016.07.018>
- Boyce, M. S. (2006). Scale for resource selection functions. *Diversity and Distributions*, 12, 269–276. <https://doi.org/10.1111/j.1366-9516.2006.00243.x>
- Breiman, L. (2001). Statistical modelling: The two cultures. *Statistical Science*, 16, 199–231. <https://doi.org/10.1214/ss/1009213726>
- Calenge, C. (2006). The package “adehabitat” for the R software: A tool for the analysis of space and habitat use by animals. *Ecological Modelling*, 197, 516–519. <https://doi.org/10.1016/j.ecolmodel.2006.03.017>
- Cenek, M., & Franklin, M. (2017). An adaptable agent-based model for guiding multi-species Pacific salmon fisheries management within a SES framework. *Ecological Modelling*, 360, 132–149. <https://doi.org/10.1016/j.ecolmodel.2017.06.024>
- Cheshire, J., & Uberti, O. (2016). *Where the animals go: Tracking wildlife with technology in 50 maps and graphics*. London, UK: Penguin.
- Cook, C. N., Lemanski, N. J., Mosqueiro, T., Ozturk, C., Gadau, J., Pinter-Wollman, N., & Smith, B. H. (2020). Individual learning phenotypes drive collective cognition. *Proceedings of the National Academy of Sciences of the United States of America*, 117(30), 17949–17956. <https://doi.org/10.1073/pnas.1920554117>
- Cresswell, T. (2011). Mobilities I: Catching up. *Progress in Human Geography*, 35, 550–558. <https://doi.org/10.1177/0309132510383348>
- Cresswell, T. (2012). Mobilities II: Still. *Progress in Human Geography*, 36, 645–653. <https://doi.org/10.1177/0309132511423349>
- Cresswell, T. (2014). Mobilities III: Moving on. *Progress in Human Geography*, 38, 712–721. <https://doi.org/10.1177/0309132514530316>
- DeAngelis, D. L., & Diaz, S. G. (2019). Decision-making in agent-based modeling: A current review and future prospectus. *Frontiers in Ecology and Evolution*, 6, 1–15. <https://doi.org/10.3389/fevo.2018.00237>
- Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., ... Weibel, R. (2015). Analysis and visualisation of movement: An interdisciplinary review. *Movement Ecology*, 3, 1–24. <https://doi.org/10.1186/s40462-015-0032-y>
- Demšar, U., Long, J. A., & Siła-Nowicka, K. (2020). Integrated science of movement. *Journal of Spatial Information Science*, 21, 25–31. <https://doi.org/10.5311/JOSIS.2020.21.721>
- Diaz, S. G., DeAngelis, D. L., Gaines, M. S., Purdon, A., Mole, M. A., & van Aarde, R. J. (2021). Development and validation of a spatially-explicit agent-based model for space utilization by African savanna elephants (*Loxodonta africana*) based on determinants of movement. *Ecological Modelling*, 447, 109499. <https://doi.org/10.1016/j.ecolmodel.2021.109499>
- Edelhoff, H., Signer, J., & Balkenhol, N. (2016). Path segmentation for beginners: An overview of current methods for detecting changes in animal movement patterns. *Movement Ecology*, 4, 21. <https://doi.org/10.1186/s40462-016-0086-5>
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77, 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Fournier, D., & Cremilleux, B. (2013). *Using impurity and depth for decision trees pruning*. Retrieved from <https://www.lri.fr/~antoine/Courses/Master-ISI/eis00.pdf>
- Han, B. A., Schmidt, J. P., Bowden, S. E., & Drake, J. M. (2015). Rodent reservoirs of future zoonotic diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 7039–7044. <https://doi.org/10.1073/pnas.1501598112>
- Hawkes, C. (2009). Linking movement behaviour, dispersal and population processes: Is individual variation a key? *Journal of Animal Ecology*, 78, 894–906. <https://doi.org/10.1111/j.1365-2656.2009.01534.x>
- Hebblewhite, M., & Haydon, D. T. (2010). Distinguishing technology from biology: A critical review of the use of GPS telemetry data in ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 2303–2312. <https://doi.org/10.1098/rstb.2010.0087>
- Hertel, A. G., Niemelä, P. T., Dingemanse, N. J., & Mueller, T. (2020). A guide for studying among-individual behavioral variation from movement data in the wild. *Movement Ecology*, 8, 30. <https://doi.org/10.1186/s40462-020-00216-8>
- Holloway, P. (2018). Simulating movement-related resource dynamics to improve species distribution models: A case study with oilbirds in northern South America. *The Professional Geographer*, 70, 528–540. <https://doi.org/10.1080/00330124.2018.1479972>

- Holyoak, M., Casagrandi, R., Nathan, R., Revilla, E., & Spiegel, O. (2008). Trends and missing parts in the study of movement ecology. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 19060–19065. <https://doi.org/10.1073/pnas.0800483105>
- Huse, G., Strand, E., & Giske, J. (1999). Implementing behaviour in individual-based models using neural networks and genetic algorithms. *Evolutionary Ecology*, 13, 469–483. <https://doi.org/10.1023/A:1006746727151>
- Jonsen, I. D., Flemming, J. M., & Myers, R. A. (2005). Robust state-space modeling of animal movement data. *Ecology*, 86, 2874–2880. <https://doi.org/10.1890/04-1852>
- Kennedy, A. S., Zerbini, A. N., Rone, B. K., & Clapham, P. J. (2014). Individual variation in movements of satellite-tracked humpback whales *Megaptera novaeangliae* in the Eastern Aleutian Islands and Bering Sea. *Endangered Species Research*, 23, 187–195. <https://doi.org/10.3354/esr00570>
- King, A. J., & Sueur, C. (2011). Where next? Group coordination and collective decision making by primates. *International Journal of Primatology*, 32, 1245–1267. <https://doi.org/10.1007/s10764-011-9526-7>
- Kranstauber, B., Cameron, A., Weinzerl, R., Fountain, T., Tilak, S., Wikelski, M., & Kays, R. (2011). The Movebank data model for animal tracking. *Environmental Modelling and Software*, 26, 834–835. <https://doi.org/10.1016/j.envsoft.2010.12.005>
- Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: An analysis using boosted regression trees. *Marine Ecology Progress Series*, 321, 267–281. <https://doi.org/10.3354/meps321267>
- Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., ... Taylor, W. W. (2007). Complexity of coupled human and natural systems. *Science*, 317, 1513–1516. <https://doi.org/10.1126/science.1144004>
- Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1, 14–23. <https://doi.org/10.1002/widm.8>
- Long, J. A., & Nelson, T. (2012). A review of quantitative methods for movement data. *International Journal of Geographical Information Science*, 27, 292–318. <https://doi.org/10.1080/13658816.2012.682578>
- Long, J. A., & Nelson, T. (2013). Measuring dynamic interaction in movement data. *Transactions in GIS*, 17, 62–77. <https://doi.org/10.1111/j.1467-9671.2012.01353.x>
- Long, J. A., Nelson, T., & Wulder, M. (2010). Regionalization of landscape pattern indices using multivariate cluster analysis. *Environmental Management*, 46, 134–142. <https://doi.org/10.1007/s00267-010-9510-6>
- Loraamm, R. W. (2020). Incorporating behavior into animal movement modeling: A constrained agent-based model for estimating visit probabilities in space-time prisms. *International Journal of Geographical Information Science*, 34, 1607–1627. <https://doi.org/10.1080/13658816.2019.1658875>
- Martin, L. J., Blossey, B., & Ellis, E. (2012). Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment*, 10, 195–201. <https://doi.org/10.1890/110154>
- Miller, H. J., Dodge, S., Miller, J., & Bohrer, G. (2019). Towards an integrated science of movement: Converging research on animal movement ecology and human mobility science. *International Journal of Geographical Information Science*, 33, 855–876. <https://doi.org/10.1080/13658816.2018.1564317>
- 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(49), 19052–19059. <https://doi.org/10.1073/pnas.0800375105>
- Oloo, F., & Wallentin, G. (2017). An adaptive agent-based model of homing pigeons: A genetic algorithm approach. *ISPRS International Journal of Geo-Information*, 6, 27. <https://doi.org/10.3390/ijgi6010027>
- Pedregosa, F., Weiss, R., & Brucher, M. (2011). Scikitlearn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
- Pérez, L., & Dragičević, S. (2011). ForestSimMPB: A swarming intelligence and agent-based modeling approach for mountain pine beetle outbreaks. *Ecological Informatics*, 6, 62–72. <https://doi.org/10.1016/j.ecoinf.2010.09.003>
- Perez, L., Kim, E.-K., & Sengupta, R. (Eds.). (2016). *Agent-based models and complexity science in the age of geospatial big data*. Cham, Switzerland: Springer.
- R Core Team. (2020). *R: A language and environment for statistical computing*. Retrieved from <https://www.R-project.org>
- Semeniuk, C. A. D., Haider, W., Cooper, A., & Rothley, K. D. (2010). A linked model of animal ecology and human behavior for the management of wildlife tourism. *Ecological Modelling*, 221, 2699–2713. <https://doi.org/10.1016/j.ecolmodel.2010.07.018>
- Stallknecht, D. E. (2007). Impediments to wildlife disease surveillance, research, and diagnostics. *Current Topics in Microbiology and Immunology*, 315, 445–461. [https://doi.org/10.1007/978-3-540-70962-6\\_17](https://doi.org/10.1007/978-3-540-70962-6_17)
- Strand, E., Huse, G., & Giske, J. (2002). Artificial evolution of life history and behavior. *American Naturalist*, 159, 624–644. <https://doi.org/10.1086/339997>
- Strandburg-Peshkin, A., Farine, D. R., Couzin, I. D., & Crofoot, M. C. (2015). Shared decision-making drives collective movement in wild baboons. *Science*, 348, 1358–1361. <https://doi.org/10.1126/science.aaa5099>
- Sutton, R., & Barto, A. G. (1999). Reinforcement learning. *Journal of Cognitive Neuroscience*, 11, 126–134.

- Tang, W., & Bennett, D. A. (2010). Agent-based modeling of animal movement: A review. *Geography Compass*, 4, 682–700. <https://doi.org/10.1111/j.1749-8198.2010.00337.x>
- Torrens, P., Li, X., & Griffin, W. A. (2011). Building agent-based walking models by machine learning on diverse databases of space-time trajectory samples. *Transactions in GIS*, 15, 67–94. <https://doi.org/10.1111/j.1467-9671.2011.01261.x>
- Tyrrell, M. (2007). Sentient beings and wildlife resources: Inuit, beluga whales and management regimes in the Canadian arctic. *Human Ecology*, 35, 575–586. <https://doi.org/10.1007/s10745-006-9105-2>
- Ward, J. A., Evans, A. J., & Malleson, N. S. (2016). Dynamic calibration of agent-based models using data assimilation. *Royal Society Open Science*, 3, 150703. <https://doi.org/10.1098/rsos.150703>
- Whoriskey, K., Auger-Méthé, M., Albertsen, C. M., Whoriskey, F. G., Binder, T. R., Krueger, C. C., & Mills Flemming, J. (2017). A hidden Markov movement model for rapidly identifying behavioral states from animal tracks. *Ecology and Evolution*, 7, 2112–2121. <https://doi.org/10.1002/ece3.2795>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4, 1686. <https://doi.org/10.21105/joss.01686>
- Williams, H. J., Taylor, L. A., Benhamou, S., Bijleveld, A. I., Clay, T. A., Grissac, S., ... Börger, L. (2020). Optimizing the use of biologgers for movement ecology research. *Journal of Animal Ecology*, 89(1), 186–206. <https://doi.org/10.1111/1365-2656.13094>
- Zhang, D., Lee, K., & Lee, I. (2019). Semantic periodic pattern mining from spatio-temporal trajectories. *Information Sciences*, 502, 164–189. <https://doi.org/10.1016/j.ins.2019.06.035>

**How to cite this article:** Patel J, Katan J, Perez L, Sengupta R. Transferring decision boundaries onto a geographic space: Agent rules extracted from movement data using classification trees. *Transactions in GIS*. 2021;00:1–17. <https://doi.org/10.1111/tgis.12770>