Received: 29 September 2024 | Accepted: 24 July 2025

DOI: 10.1111/2041-210X.70136

Methods in Ecology and Evolution

RESEARCH ARTICLE

Conservation, Ecology and Artificial Intelligence – Advances and Symbiotic Solutions

Predicting animal movement with deepSSF: A deep learning step selection framework

Scott W. Forrest[1,2,3] | **Dan Pagendam**[4] | **Conor Hassan**[1,5] | **Jonathan R. Potts**[6] | **Christopher Drovandi**[1,2] | **Michael Bode**[1,2] | **Andrew J. Hoskins**[7,8]

[1]Centre for Data Science, Queensland University of Technology, Brisbane, Queensland, Australia; [2]School of Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia; [3]Environment, CSIRO, Dutton Park, Queensland, Australia; [4]Data61, CSIRO, Dutton Park, Queensland, Australia; [5]Department of Computer Science, Aalto University, Aalto, Finland; [6]School of Mathematical and Physical Sciences, University of Sheffield, Sheffield, UK; [7]Environment, CSIRO, Townsville, Queensland, Australia and [8]Northern Australian Indigenous Land and Sea Management Alliance, Townsville, Queensland, Australia

Correspondence
Scott W. Forrest
Email: scottwforrest@gmail.com

Funding information
Department of Agriculture, Fisheries and Forestry, Australian Government, Grant/Award Number: Smart Farming Partnerships Program (round 2); Australian Research Council, Grant/Award Number: FT210100260; Australian Government Research Training Program

Handling Editor: Nicolas Lecomte

Abstract

1. Predictions of animal movement are vital for understanding and managing wild populations. However, the fine-scale, complex decision-making of animals can pose challenges for the accurate prediction of trajectories. Integrated step selection functions (iSSFs), a common tool for inferring relationships between animal movement and the environment, are also increasingly used to simulate animal trajectories for prediction. Although admitting a lot of flexibility, the iSSF framework is limited to its reliance on pre-defined functional forms for fitting to data, and iSSFs that involve complex functional forms to model detailed processes can be prohibitively difficult to fit and interpret.
2. Here, we present deepSSF, an approach to fit and predict animal movement data using deep learning. The deepSSF approach replaces the log-linear model of an iSSF with a neural network architecture that receives multiple environmental layers and scalar values as inputs and outputs a single layer representing the next-step probability. We demonstrate an example deepSSF model, built in `PyTorch`, consisting of distinct but interacting habitat selection and movement subnetworks. This allows for explicit representation of both selection and movement processes, thus giving interpretable intermediate outputs. We apply our model to GPS data of introduced water buffalo (*Bubalus bubalis*) in the tropical savannas of Northern Australia.
3. Our deepSSF model was able to learn features that are present in the habitat covariate layers, such as linear features (rivers, forest edges) and the composition of certain habitat areas, without having to specify them pre-emptively within the model framework. It was able to capture complex interactions between the habitat covariates as well as temporal dynamics across time of day and year.

Finally, our deepSSF model generally had better in- and out-of-sample predictive accuracy than the analogous iSSF model.
4. We expect that the deepSSF approach will generate accurate and informative predictions about animal movement, which can be used for deepening our understanding of animal–environment systems and for the practical management of species. We discuss how the wide range of existing deep learning tools could enable the deepSSF approach to be extended to represent memory and social dynamic processes, with the potential for integrating non-spatial data sources such as accelerometers and physiological sensors.

KEYWORDS
conservation, deep learning, habitat selection, movement ecology, predictive ecology, simulations, step selection function, temporal dynamics

1 | INTRODUCTION

The movement of animals through space drives the structure and functioning of ecosystems through processes such as seed dispersal, herbivory and predator–prey relationships (Bello et al., 2024; Côrtes & Uriarte, 2013; Fortin et al., 2005). An ability to predict animal locations is also vital for conservation efforts. Accurate predictions of animal movement can be used to mitigate impacts to wildlife when designing infrastructure such as roads, wind turbines, or mines (Cowan et al., 2024; Mayer et al., 2021; Pay et al., 2022), to reduce the risk to threatened species (Finnegan et al., 2021; Forrest, Rodríguez-Recio, & Seddon, 2024), or to manage invasive or problematic species (Lustig et al., 2019; Patterson et al., 2024; Pili et al., 2022).

Recently, there has been a growing prevalence of predicting animal movement by simulating it from fitted integrated step selection functions (iSSFs), a flexible way of incorporating measured processes such as movement, habitat selection, memory, and social interactions (Avgar et al., 2016; Ellison, Potts, Boudreau, et al., 2024; Ellison, Potts, Strickland, et al., 2024; Potts & Börger, 2023; Signer et al., 2017; Signer et al., 2023). iSSFs have been used for quantifying connectivity (Hofmann et al., 2023; Hooker et al., 2021; Sells et al., 2023; Whittington et al., 2022), expected spatial distributions (Forrest, Pagendam, et al., 2024), and the effect of different landscapes on movement behaviour (MA Cowan, unpublished data). Simulating from an iSSF requires proposing potential steps from a starting location, determining the relative probability of choosing each of these steps (using the fitted movement and habitat selection parameters), and then choosing one of them using these relative probabilities. The end location of this step then becomes the next step's starting point, and the process is repeated. Recent developments aim to increase the flexibility of iSSFs through the addition of temporally dynamic and flexible responses (Forrest, Pagendam, et al., 2024; Klappstein et al., 2024), which can increase the realism of the trajectories. However, animal movement is notoriously difficult to predict at fine scales, and the decisions that animals make depend on a range of factors that may interact in ways that are difficult to predict a priori (Couzin, 2009; Nathan et al., 2008; Strandburg-Peshkin et al., 2015). Therefore, to make accurate predictions, we require a framework that can handle complex interactions and does not require pre-emptive specification of functional forms and responses for animal movement. One such approach is deep learning, a highly flexible data-driven method that can accommodate a large number of inputs and represent an impressively wide range of complex processes (Drori, 2022; LeCun et al., 2015).

Deep learning uses multiple processing layers to construct an internal representation of data that can have multiple layers of abstraction (LeCun et al., 2015). It has led to many advances in both our understanding of the world around us and our ability to generate predictions of complex and abstract phenomena (Raghu & Schmidt, 2020). In recent years, deep learning has advanced rapidly, with progress in both modelling architectures that enable more flexible model fitting and hardware such as graphics processing units (GPUs) that have enabled models to be fitted with much greater speed. Increasingly, deep learning is being applied to complex problems in science, often with impressive results (Jumper et al., 2021; Raghu & Schmidt, 2020). The flexibility of deep learning models and their ability to take an array of different inputs and produce almost any output means that, with some thought and consideration, they can help uncover interesting features in almost any dataset. There are many existing applications of deep learning that are similar to the next-step-ahead predictions of animal movement, such as the movement of a robot, game playing, text generation and the generation of human and vehicle trajectories (Drori, 2022). Although the application of deep learning has yet to be fully utilised or adopted for modelling animal movement, there have been some promising examples that consider next-step ahead predictions using a range of machine learning architectures (Chen et al., 2024; Cífka et al., 2023; Dalziel et al., 2008; Einarson et al., 2024; Kazama et al., 2024; Shenk et al., 2021; Wijeyakulasuriya et al., 2020).

One way to apply deep learning to animal movement is by representing the process as comprised of distinct but interacting

behavioural processes, such as movement and habitat selection, analogous to a step selection function. Some benefits of using a deep learning approach include: (1) an ability to learn about spatial features of the landscape that may not be explicitly captured by the pixel values of the covariates, as we can use deep learning tools that were developed for image data; (2) covariate interactions can also be represented by flexible and non-linear processing layers, which can also interact with temporal dynamics over multiple time scales; (3) given the modular capability of deep learning, there is additional potential for extending the model by using the wide array of deep learning tools, such as recurrent layers (Drori, 2022) or transformers (Vaswani et al., 2017) to incorporate a history of previous locations, which may approximate memory processes; (4) the flexibility of deep learning also allows for other data sources such as accelerometers and physiological sensors to be integrated, which could allow for a more holistic representation of animal movement behaviour (English et al., 2024; Williams et al., 2020).

In this paper we present an approach that formulates a deep learning model to replicate an iSSF, with an explicit representation of movement and habitat selection processes. We term this approach deepSSF, which refers to the general approach, rather that any particular network architecture. There are numerous ways to formulate a deep learning model that could be used for modelling animal movement, and we hope to highlight one possible approach that may be readily extended to use other neural network architectures and incorporate more realistic processes such as memory and social dynamics. We show how the deepSSF approach may be used to learn about hitherto unknown behavioural processes underpinning the movement data, as well as for realistic predictions.

2 | MATERIALS AND METHODS

Technical terms related to deep learning are summarised in the Glossary, found in the Supporting Information, as well as a Supporting Information section of deep learning concepts in the context of the deepSSF approach.

2.1 | From iSSF to deepSSF: Mathematical formalisms

An iSSF denotes the probability, $p_\tau(\mathbf{x}|\mathbf{y},t)$, of moving from one location, $\mathbf{y}$ at time t, to another, $\mathbf{x}$ at time $t+\tau$. We will assume here that $\mathbf{y}$ and $\mathbf{x}$ are locations of vertices in a rectangular lattice S (corresponding to pixels in a rectangular raster), so we are working in discrete space (although iSSA can also be formulated in continuous space). Usually, this probability is given in an exponential functional form, as follows

$$p_\tau(\mathbf{x}|\mathbf{y},t) = K^{-1}(\mathbf{y},t)\underbrace{\exp\left(\lambda_1 M_1(\mathbf{x},\mathbf{y},t) + \ \dots \ + \lambda_m M_m(\mathbf{x},\mathbf{y},t)\right)}_{\text{selection - free movement}}\underbrace{\exp\left(\beta_1 H_1(\mathbf{x},\mathbf{y},t) + \ \dots \ + \beta_n H_n(\mathbf{x},\mathbf{y},t)\right)}_{\text{habitat selection}}. \quad (1)$$

Here, $M_i(\mathbf{x},\mathbf{y},t)$ (for $i = 1, \dots, m$) are movement covariates, which could be as simple as $M_i(\mathbf{x},\mathbf{y},t) = |\mathbf{x}-\mathbf{y}|$ to model the effect of step length on the next location, or could depend upon time or habitat features that affect movement (e.g. snow depth or slope of terrain) (Avgar et al., 2016). The functions $H_j(\mathbf{x},\mathbf{y},t)$ (for $j = 1, \dots, n$) denote aspects of the habitat that animals might select for (if $\beta_i > 0$) or avoid (if $\beta_i < 0$). The normalising function $K(\mathbf{y},t)$ ensures that $\sum_{\mathbf{x}\in S} p_\tau(\mathbf{x}|\mathbf{y},t) = 1$, so that $p_\tau(\mathbf{x}|\mathbf{y},t)$ is a probability distribution on S.

Importantly, in 'ordinary' step selection, $\ln\left[p_\tau(\mathbf{x}|\mathbf{y},t)\right]$ is a *linear* function of the $M_i(\mathbf{x},\mathbf{y},t)$ and $H_j(\mathbf{x},\mathbf{y},t)$ functions (where M is often defined at the start of the step and H at the end of the step). It is also defined *locally*, so that $\ln\left[p_\tau(\mathbf{x}|\mathbf{y},t)\right]$ only depends upon values of the $M_i(\mathbf{x},\mathbf{y},t)$ and $H_j(\mathbf{x},\mathbf{y},t)$ functions at $\mathbf{x}$, and ignores pixels in the vicinity. The key generalisation that deepSSF makes is to replace this linear function by a sum of two compositions of functions, $f_k \circ \dots \circ f_1 + g_l \circ \dots \circ g_1$, where each individual function (i.e. each of $f_1, \dots, f_k, g_1, \dots, g_l$) may be nonlocal and/or non-linear. [Note that $f_{j+1} \circ f_j(x) := f_{j+1}\left(f_j(x)\right)$ for $j = 1, \dots, k-1$ but the former notation avoids nested brackets]. We write $M_i(\cdot,\mathbf{y},t)$ and $H_j(\cdot,\mathbf{y},t)$ to denote, respectively, the matrices whose entries are $M_i(\mathbf{x}',\mathbf{y},t)$ and $H_j(\mathbf{x}',\mathbf{y},t)$ for all $\mathbf{x}' \in S$. Then deepSSF has the following generic movement model formulation

$$p_\tau(\mathbf{x}|\mathbf{y},t) = K^{-1}(\mathbf{y})\exp\left(f_k \circ \dots \circ f_1\left[M_1(\cdot,\mathbf{y},t), \dots, M_m(\cdot,\mathbf{y},t)\right] + g_l \circ \dots \circ g_1\left[H_1(\cdot,\mathbf{y},t), \dots, H_n(\cdot,\mathbf{y},t)\right]\right). \quad (2)$$

In the context of deep learning, the compositions $f_k \circ \dots \circ f_1$ and $g_l \circ \dots \circ g_1$ are usually known as *neural networks*.

The great advantage of deep learning is that, over the past few decades, researchers have discovered specific functional forms for $f_1, \dots, f_k$ and $g_1, \dots, g_l$ that are both able to capture arbitrary non-linear and nonlocal functions to an arbitrarily high degree of accuracy and can be fitted to data rapidly to give high predictive power. The particular functions forms (also known as *layers*) that we will use here are fully connected layers, convolutional layers, rectified linear units (ReLU), and max pooling, which will be explained in greater detail below. In the deepSSF framework presented here, we also insist on a particular form of f_k that allows us to frame $f_k \circ \dots \circ f_1$ as a product of step lengths and turning angle distributions (as gamma and von Mises distributions respectively), but this is not necessary a priori.

2.2 | deepSSF model architecture

To retain the separable movement and habitat selection components described in Equation (1), we developed a model that had two distinct but interacting structures that we call subnetworks, which explicitly represented the movement and habitat selection processes.

These were denoted by $f_k \circ \ldots \circ f_1$ and $g_l \circ \ldots \circ g_1$ respectively in the previous section. As with iSSF (Avgar et al., 2016), the outputs of the deepSSF subnetworks were combined additively into a next-step probability surface. This is similar to the 'log-additive neural model' used by Pagendam et al. (2023). We describe each of the subnetworks in more detail below, and provide a conceptual overview of our deepSSF model in Figure 1.

2.2.1 | Habitat selection subnetwork

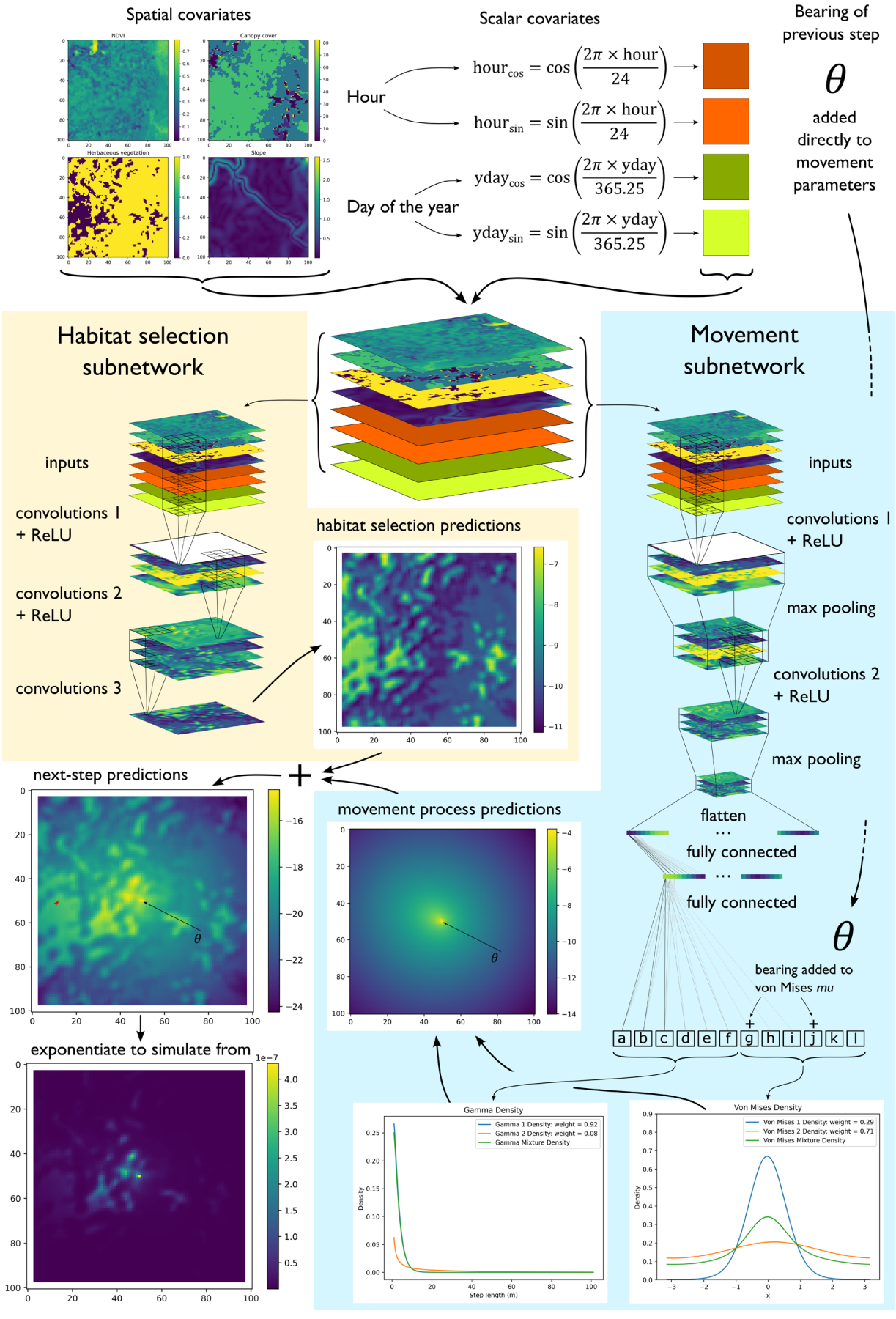
The purpose of the habitat selection subnetwork is to transform the input covariates into a probability surface that describes the selection preferences of the animal, disregarding the effect of movement (as in typical step selection analysis). To achieve this we used convolutional layers. These are often described as 'feature-extractors' (Hertel et al., 2015), as they determine salient features of the input rasters that influence the prediction outcome. In our case they help determine the features of the surrounding landscape that influenced the animal's choice of the next step, and importantly can capture non-local aspects of habitat selection. Convolutional layers are characterised by the use of convolution filters (also known as kernels) that sweep over the spatial inputs and apply pixel-wise operations, resulting in a set of 'feature-maps'. When making predictions, the particular values of the filters do not change, meaning that they apply the same transformation equally across the spatial extent of the inputs (i.e. they are 'spatially equivariant'). Therefore they do not conflate with the movement process, which is biased toward the centre cell, $s(t)$. These filters evaluate the local neighbourhood of grid cells, and with several successive layers they can represent features of the landscape such as forest edges, differing patch sizes, or riparian vegetation. The filters used by convolutional layers process all input layers simultaneously, meaning that interactions between all input covariates are considered, which can also include variables such as the time of the day and day of the year, as well as information about the animal such as its age or sex. For a conceptual overview of convolutional layers see the Supporting Information.

Our habitat selection subnetwork had three consecutive convolutional layers. The first two of these had rectified linear unit (ReLU) activation functions, which enabled the model to capture non-linear effects of the input rasters on selection. These two convolutional layers each had four different $3 \times 3 \times n$ cell convolution filters that can each extract different features of the inputs, where n is the number of spatial inputs. Each convolutional layer resulted in four output layers (one for each filter), which are typically referred to as feature maps. The final convolutional layer had a single $3 \times 3 \times n$ filter, where n is now the number of feature maps generated by the previous convolutional layer. This layer acted to process the feature maps into a single output layer (without an activation function), which is the habitat selection probability surface. The inputs for each convolutional layer were padded (see padding) with zeros and the filters had a stride of 1, which is critical to ensure that the spatial extent (i.e. number of cells) of the output of each convolution operation is the same as the input. This formulation allows for any number of convolution layers to be applied whilst retaining the same spatial extent for the predicted habitat selection surface. The convolutional layers can therefore be considered to apply spatially-nonlocal (as the filters consider neighbouring cells) and non-linear (due to the ReLUs) transformations of the input data.

The number of convolutional layers and filters in our deepSSF model was chosen to be relatively low whilst still having enough complexity to represent the habitat selection process. Determining the number of convolutional layers and filters and the number of nodes in a feedforward network will vary by study due to the volume of data and differing predictability of the data and the spatial layers. We suggest starting with a small model for faster model fitting

FIGURE 1 Conceptual overview of the deepSSF neural network used to predict animal movement. There are two subnetworks: A habitat selection and a movement process subnetwork. Both receive the same inputs, which are spatial layers such as environmental covariates, scalar covariates such as the hour, the day of the year (yday, also called 'ordinal' or 'Julian' day), and the bearing of the previous step. The periodic components (i.e. hour and yday) are decomposed into sine and cosine components to wrap continuously as a period, before being converted into spatial layers with constant values so they can be processed by the convolutional layers, each of which are followed by a rectified linear unit ReLU activation layer. To ensure that the turning angles are relative to the previous step, the bearing of the previous step is added directly to the predicted mean (μ) parameters of von Mises distributions. The habitat selection subnetwork uses convolutional layers that have parameters set to ensure that the output has the same spatial extent as the input, resulting in spatial, non-linear transformations of the input covariates, where all inputs can interact, to produce a probability surface (on the log-scale) describing the likelihood of moving to any cell based on the surrounding environment. The movement process subnetwork uses convolutional layers with max pooling to extract features from the input covariates that are salient to movement, and fully connected layers to process the convolutional layer outputs whilst reducing the dimensionality. The predicted output of the movement subnetwork can be any number of parameters that govern a movement distribution, so we used finite mixtures of two gamma distributions for the step lengths and two von Mises distributions for the turning angles. This results in a total of 12 predicted parameters—a shape, scale and weight for each gamma distribution and a μ, κ and weight for each von Mises distribution. The parameters are then converted to a two-dimensional movement surface (also on the log-scale) which is added to the habitat selection predictions, resulting in a next-step log-probability surface. To generate trajectories, the next-step log-probability surface is exponentiated and normalised such that the sum of all cells is equal to one, and a step is sampled according to these probability values. To highlight the directional persistence, the arrow and θ in the movement and next-step predictions denotes the bearing of the previous step, and the red star to the left of the next-step predictions is the location of the observed next step for those inputs. When fitting the model we are trying to maximise the probability at each observed next-step.

BRITISH ECOLOGICAL SOCIETY Methods in Ecology and Evolution



and increasing it gradually until there are negligible improvements in the loss function on out-of-sample test data. See the Supporting Information for consideration of model and training hyperparameters when fitting deepSSF models.

2.2.2 | Movement process subnetwork

The aim of the movement subnetwork is to predict parameters of step-length and turn-angle distributions that govern a movement kernel. These parameters are, in turn, functions of input rasters, such as the surrounding habitat, and other covariates such as the hour or day of year. The resulting movement parameters are converted to a two-dimensional surface, using the density functions relevant to the predicted parameters (i.e. for gamma and von Mises distributions), which becomes the output of the movement process subnetwork. As we are converting one-dimensional distributions of step lengths and turning angles into a two-dimensional movement surface, we account for the increasing two-dimensional area at larger step lengths by dividing the probability density by the distance from the centre cell (Michelot et al., 2024; Rhodes et al., 2005; Schlägel & Lewis, 2016). This change of variables is illustrated and described in the Supporting Information.

The movement process subnetwork that we used had three components. Two of these include processing layers with parameters that are estimated, and the other converts the predicted movement parameters into a two-dimensional probability surface. Similar to the habitat selection subnetwork, the first component of the movement subnetwork was comprised of convolutional layers, such that it can take spatial and temporal covariates as grids where all inputs can interact, allowing for the movement parameters to depending on the surrounding habitat and the time of day and year. It is not strictly necessary to include the spatial covariates as inputs to the movement process, as the temporal covariates can be inputted directly to the fully connected layers which would obviate any convolution layers, but we illustrate a movement process here that can be influenced by the surrounding habitat. In contrast to the habitat selection subnetwork, the movement subnetwork included max pooling layers after each of the convolutional and ReLU layers, which reduce the number of parameters whilst retaining important features. We used a max pooling kernel size of $2 \times 2 \times n$ cells that had a stride of two, which means that each max pooling layer will reduce the total number of cells in the output grid by a factor of four. We used two convolutional layers with four convolution filters each, which were followed by ReLU activation functions and max pooling layers. As there were two max pooling operations, the final feature maps had 16 times fewer cells than the input layers. The size of the convolution filters were the same as in the habitat selection subnetwork ($3 \times 3 \times n$), although these are independent from those of the habitat selection subnetwork so that the filters will learn different features that are relevant to each process. In the movement subnetwork, the filters will extract features from the covariates that influence an animal's general movement tendencies, thereby influencing the parameters of the movement kernel.

For the movement subnetwork to predict the correct number of movement parameters, the dimension must be reduced so that the output is a vector of length equal to the number of movement parameters. To achieve this we used a flatten operation on the outputs of the convolutional layers, which turns them into a single long vector which has a length equal to the total number of cells in the final feature maps. We processed this vector using fully connected layers, where every cell in the flattened vector connects to every cell in the next 'hidden' layer, which we gave a length of 128. These 128 cells are then connected to every cell in a final vector, which has a length equal to the number of parameters that define the movement kernel that we want to estimate. As all cells in the fully connected layers connect, every cell from the initial covariate inputs can influence the final movement parameters, where the convolutional layers extract the most important features from the spatial covariates. The final component in the movement subnetwork takes the predicted movement parameters and converts them to a two-dimensional probability surface using the appropriate density function for each distribution, whilst dividing by the distance from the centre (i.e. the step length to each cell) to account for the change of movement variables (Supporting Information; Michelot et al., 2024; Rhodes et al., 2005; Schlägel & Lewis, 2016). As the centre cell has a distance of 0, for numerical stability we assigned this cell a positive value, for which we used the average distance from the centre of the cell to any location within the cell. For a 25 m × 25 m cell is $\int_{-12.5}^{12.5}\int_{-12.5}^{12.5}\sqrt{x^2+y^2}\,dx\,dy$, which is equal to ~9.56 m. To prevent overfitting, we used dropout at a rate of $p = 0.1$ (Srivastava et al., 2014) within the fully connected layers.

As there can be any number of outputs from the fully connected layers, any form of step-length and turn-angle distributions can be used, whether parametric distributions or distributions described by other methods such as basic functions, providing that these parameters can be converted to a two-dimensional surface. This formulation then allows for finite mixtures of probability distributions, where there are multiple distributions that are combined together using a weighting. To provide more flexibility to the model to allow it to more easily capture the temporal dynamics, we fitted a mixture of two gamma distributions for the step lengths and a mixture of two von Mises distributions for the turning angles.

2.2.3 | Next-step probability and loss function

To construct the next-step probability surface, the habitat selection and movement process probability surfaces are combined by adding them together when they are on the log-scale, in exactly the same way as for iSSFs (Avgar et al., 2016). Prior to being combined, they are normalised whilst on the log-scale (using the log-sum-exp trick for numerical stability) to be valid probability surfaces (i.e. they sum to 1 after being exponentiated), so each subnetwork contributes equally to the next-step probability surface. We combined

the normalised predictions of the habitat selection and movement processes prior to evaluating the loss function, which means that the final prediction accuracy depends on both subnetworks, and although they have separate sets of parameters and can learn their own representations, the subnetworks are trained simultaneously and implicitly depend on one another.

The loss function quantifies how accurately a model is predicting, which is used to update the parameters of the network. We used a negative log-likelihood (NLL) loss function, where the target that the model is trying to predict is the observed location of the next step. For a 101×101 grid there are $N = 10{,}201$ cells, each with a probability value, $\hat{p}_i$ from the next-step probability surface.

We seek to maximise the probability of selecting the cell where the step was observed to land. This is equivalent to minimising the negative log-probability, $\text{NLL} = -ln(\hat{p}_{i^*})$, where i^* is the grid cell where the step ends. The next-step log-probabilities are given by:

$$\ln(\hat{p}_i) = \ln(\hat{p}_{mi}) + \ln(\hat{p}_{hi}) - K, \quad (3)$$

where $\ln(\hat{p}_{hi})$ are the predicted log-probability values from the habitat selection subnetwork, and $\ln(\hat{p}_{mi})$ are the predicted log-probability values from the movement subnetwork. K includes the normalisation constants for the habitat selection and movement surfaces, denoted by $\ln \sum_{i=1}^{N} \hat{p}_{mi} + \ln \sum_{i=1}^{N} \hat{p}_{hi}$, which ensures that they each sum to 1 after being exponentiated. We found that training was more stable when the probability surfaces were normalised individually prior to being combined rather than normalising the next-step probability surface after combining. This is likely due to the difference in scale between the habitat selection and movement probabilities. We used the log-sum-exp trick to calculate the normalisation constants, but we do not denote it here for clarity (a description can be found in the Glossary). Each set of inputs only has a single observed next-step, resulting in a single loss function evaluation, so we take the average loss $\frac{1}{n} \sum_{i=1}^{n} -ln(\hat{p}_{i^*})$ for a set of n samples (i.e., a batch) when updating the model parameters.

2.3 | Data preparation for deepSSF

To take advantage of deep learning tools that consider the spatial structure of the data, such as convolutional layers, it is necessary to format the input data as grids (i.e. images or rasters). As we were generating next-step probability surfaces, we extracted the environmental information surrounding the animal's current location, within the range of its typical step lengths. This resulted in a set of local rasters for each observed step. We chose a grid size of 101×101 cells, which is 2525 m×2525 m, or 1262.5 m to the nearest edge of the local landscape. We chose this distance as 1262.5 m comprises more than 96% of the observed step lengths. It is important to have an odd number of cells so that there is a central cell, as that is the location of the animal at the start of the step, $s(t)$.

For the periodic inputs such as hour of the day and day of the year, to maintain their periodicity (so hour 24 is adjacent to hour 1, and yday 1 is adjacent to yday 365), we decomposed them into sine and cosine terms using $\tau_{\sin} = \sin(2\pi\tau / T)$ and $\tau_{\cos} = \cos(2\pi\tau / T)$, where τ is the periodic covariate (the hour of the day or day of the year), and T is the largest value in the period, that is 24 or 365.25 (to account for leap years). To use convolutional layers which require images (i.e. grids or rasters) as inputs, we converted any scalar (single value) data to grids of the same spatial extent as the environmental covariates, and set the value of every cell to the value of the scalar.

To assess how well the model predicted the next step (i.e. to evaluate the loss function), we created an additional spatial layer corresponding to the 'target', which is the observed location of the next step. This spatial layer is comprised of zeros at all entries except the entry corresponding to the location of the next step, where it has a value of 1.

2.4 | Study area and data collection

Data were collected from the Djelk Indigenous Protected Area in Western Arnhem Land, Northern Territory, Australia. The area is a culturally significant landscape comprised of tropical savanna with areas of open woodland, rainforest, a varied river and wetland system, and open floodplains. To understand their movement and habitat selection behaviours, 17 female water buffalo (*Bubalus bubalis*) were GPS-tracked in collaboration with Bawinanga Aboriginal Corporation rangers between July 2018 and November 2019. For further details on the data collection see Forrest, Pagendam, et al. (2024). All animal handling procedures were performed by suitably qualified personnel trained in the appropriate capture and handling techniques under the approval of CSIRO Wildlife and Large Animal Ethics Committee (permit no. 2017-27). To exemplify the deepSSF approach, we randomly selected a single individual's data, which had 10,103 hourly GPS locations that were suitable for training the model. This individual's data ranged from the 25th of July 2018 until the 31st of October 2019, with a fix-success rate of 90.9%.

2.5 | Landscape covariates

Buffalo movement decisions are driven by factors such as vegetation composition and density for resource acquisition and shade, access to water, and the terrain (Campbell et al., 2020). In monsoonal ecosystems of Northern Australia, vegetation and the distribution of water change dramatically throughout the year. To represent the seasonal changes in vegetation, we used monthly normalised difference vegetation index (NDVI), which measures photosynthetic activity and approximates the density and health of vegetation (Myneni et al., 1995; Reed et al., 1994). NDVI is an informative covariate in this landscape (Campbell et al., 2020) as it distinguishes between the broad vegetation classes, identifies wet and flooded areas, and can quantify buffalo's forage resources as they are typically under open canopy. Monthly NDVI layers were generated from Sentinel-2 spectral imagery at 10 m×10 m resolution using

Google Earth Engine by taking the clearest pixels from a range of images for that month to alleviate the effects of obstruction from clouds. We selected the highest quality-band-score, which is based on cloud and shadow probability for each pixel, resulting in a single obstruction-free image of the NDVI values for each month of each year. We also used temporally static layers for canopy cover and herbaceous vegetation, which are derivatives of Landsat-7 imagery and were sourced from Geoscience Australia at 25 m × 25 m resolution (Source: Geoscience Australia; Landsat-7 image courtesy of the U.S. Geological Survey). We represented broad-scale topographic features that affect buffalo movement by including a slope covariate, which was summarised from a 25 m × 25 m digital elevation model using the `terra` R package (Hijmans, 2024; R Core Team, 2024), and was calculated using the methodology of Horn (1981), which uses the elevation difference in the x and y directions of the eight neighbouring cells to estimate slope. The canopy cover layer was a proportion from 0 (completely open) to 1 (completely closed), and the herbaceous vegetation layer was binary, with 1 representing grasses and forbs, and 0 representing other (which is predominately woody growth). All spatial variables were discretised into grids (i.e. rasters) and resampled to be 25 m × 25 m resolution.

2.6 | Model fitting and predictions

The data were sequentially separated into training, validation, and test datasets, with the first 80% of the data used for model fitting (training), the next 10% for validation, and the final 10% used for testing the model and calculating out-of-sample predictive accuracy. As is typical in deep learning, the training data were split into batches, and the loss function is evaluated for all samples in the batch. We used a batch size of 32 samples; the model was robust to varying batch sizes. Smaller batches lead to more stochastic (noisier) gradients, whereas larger batches are more stable. Often, the batch size that leads to the best trained neural network is some intermediate batch size (32 is a common choice), which provides enough information to calculate gradients such that parameter updates are taken in a beneficial direction, but also enough noise that it is able to escape from local minima. Smaller batches can also be computationally beneficial in that less data needs to be processed to obtain a gradient estimate when updating the parameters, but there is a higher CPU-GPU data transfer overhead, and smaller batches can underutilise GPU capacity.

Using the average loss of a batch, the weights of the model are updated by performing backpropagation (Drori, 2022; Murphy, 2022), and then the next batch is put in. A complete iteration through all batches of the training data is termed an epoch. At the end of each epoch, the loss function is evaluated on all of the validation data, and if the average validation loss decreases, the model is said to have improved (no backpropagation is performed after evaluating the validation data, to ensure it does not contribute to updating the model's parameters). Fitting the model typically involves many epochs. The test data are reserved until training has completed. They are used to test the overall performance of the model on unseen data, which is typically used to compare between models.

The model was initialised with random weights, and we used the stochastic optimiser 'Adaptive Moment Estimation', Adam, (Kingma & Ba, 2014) with an adaptive learning rate (the amount that the weights and bias parameters are updated) for *each of the processes*, that decreased when the contribution to the loss function for each process plateaued. Using a separate optimisers for each process also allowed us to use different starting learning rates, and we started the learning rate at 10^{-4} for the habitat selection process and 10^{-5} for the movement process. The learning rates were decreased when their respective contribution to the loss had not decreased for five epochs, and training was terminated when the average combined validation loss had not decreased for 15 epochs. To prevent saving an overfitted model, the model weights are only saved when the validation loss decreased, so the final model weights are those that had the best performance on the validation data.

2.6.1 | Simulating from the deepSSF model

Simulating from the deepSSF model requires selecting a starting location, hour of the day, and day of the year (the bearing was randomly initialised). The local environmental covariates are cropped out as grids with 101 × 101 cells centred on the starting location, and the time inputs are decomposed into their respective sine and cosine components and converted to spatial layers. The trained model then processes the stack of spatial inputs, resulting in a next-step probability surface, which is exponentiated, normalised to sum to 1, and sampled with respect to the probability weights. The sampled location is then the next step, and the process is repeated until the desired number of locations is reached. Due to having discrete cells that are sampled from, and selecting the centre cell would result in a step length of 0, we add jitter to each simulated location. We sample the jitter from a normal distribution centred on the middle of the cell with a standard deviation of 6.5 m, as 95% of this density is within the cell. If the sampled jitter would put the simulated location outside of the cell, it was resampled. We tested the sensitivity of this choice by varying the SD of the jitter distribution, as well as sampling uniformly from within the cell; none of which meaningfully changed the results.

We simulated 50 deepSSF trajectories from the starting location of the observed trajectory that covered the late-dry season 2018, resulting in 3000 steps per trajectory. We assessed whether the simulated trajectories displayed realistic characteristics of animal movement trajectories by visual comparison and by comparing summary statistics of the movement and habitat selection with observed data. For the movement behaviour, we compared the simulated and observed step lengths by plotting the distribution of step lengths and turning angles (Supporting Information) and also the mean step length for each hour of the day, which is a similar approach to Forrest, Pagendam, et al. (2024) (Figure 4). For the habitat selection, we binned the simulated and observed locations into hourly bins and calculated the mean values of each of the covariates, which was also similar to Forrest, Pagendam, et al. (2024) (Figure 4).

2.6.2 | Landscape-scale habitat selection

As trained convolution filters can be applied to images (i.e. rasters) of any size, it is possible to generate habitat selection surfaces of any size by using the trained filters from the habitat selection subnetwork. These filters have been trained to extract the salient features of the input data and produce higher probability values where an animal is likely to take its next step, which interact with the time of day and the day of the year. Additionally, only local covariates that are 'available' to the animal are used as inputs, which prompted the initial development of step selection functions (Fortin et al., 2005; Rhodes et al., 2005).

The resulting probability surface denotes only the next-step habitat selection, and not the expected distribution of that next step (nor the expected distribution of animal locations per se). This is because this surface ignores movement constraints, effectively assuming that all locations on the landscape are equally accessible at the next time step. That said, it does provide a visual representation of what the habitat selection subnetwork has learned about the animal's selection process from the environmental covariates, and how this changes with respect to the other covariates such as the hour and day of the year. This can be used as a diagnostic tool for further model development, or as an end in itself, as it highlights the features present in the environmental covariates that the model considered to be influential in determining the next step's location. From these maps, we can also plot the correlation between the covariate values and the habitat selection predictions, illustrating which covariate values were associated with higher habitat selection (Supporting Information). Quantifying the relationship between the input covariate values and the predicted probabilities in this way is similar to partial dependence plots (Friedman, 2001).

2.6.3 | Next-step ahead validation and comparison to iSSF

Whilst the realism and emergent properties of simulated trajectories are difficult to assess, we can validate the deepSSF models on their predictive performance at the next step, for each of the predicted habitat selection, movement and next-step surfaces. Ensuring that each of the predicted surfaces are normalised to sum to one, they can be compared to the predictions of typical step selection functions when the same probability surfaces are generated for the same local covariates. We can also compare both of these to completely random movement and habitat selection by calculating uniform probabilities over the movement, habitat selection and next-step probability surfaces, which is calculated as equal probability for every cell ($p = 1/n$ cells). This approach not only allows for comparison between models, but can be informative as to when in the observed trajectory the model performs well or poorly, which can be analysed across the entire tracking period or for each hour of the day, and can lead to critical evaluation of the covariates that are used by the models and thus allow for model refinement.

We compared the deepSSF models to the predictive accuracy of two typical iSSFs fitted using conditional logistic regression to the same covariates as the deepSSF model described above (NDVI, canopy cover, herbaceous vegetation and slope). One of these models was fitted without temporal dynamics, denoted 'iSSF' (Model 1), and the other iSSF was fitted with temporal dynamics on the movement and habitat selection processes over a daily period using two pairs of harmonics, as presented by Forrest, Pagendam, et al. (2024), which we refer to as the 'dynamic iSSF' (Model 2). We fitted deepSSF models to two sets of covariates, the first of which was fitted to the same covariates as the iSSF models, which we refer to simply as the 'deepSSF' model (Model 3), as well as a deepSSF model fitted to 12 Sentinel-2 bands and slope at the same 25 m × 25 m cell resolution, which we refer to as 'deepSSF S2' (Model 4), which is described in detail in the Supporting Information. These models are outlined in Table 1. To compare against random movement and habitat selection we also include a comparison to uniform probabilities (1/n cells in probability surface), where the uniform probabilities are referred to as Model 0, as it is not necessary to fit a model to calculate them.

We fitted each of these four models to the first 80% of a randomly selected individual buffalo's data (ID 2005, $n = 8082$ steps), which had more than a year of high-quality data. We assessed the 'in-sample' performance of the models by generating and evaluating next-step-ahead predictions over the data the model was fitted to (typically referred to as the 'training data'), and the 'out-of-sample' performance by generating next-step predictions the final 10% of the data, which was not used in model fitting ($n = 1010$ steps). The

TABLE 1 The models that were fitted to an individual buffalo's GPS tracking data to compare predictive accuracy, as well as uniform probabilities to assess performance against random movement and habitat selection.

Model	Name	Description	Spatial covariates
0	uniform	(1/n cells in probability surface)	—
1	iSSF	baseline iSSF	NDVI, canopy cover, herbaceous vegetation, slope
2	dynamic iSSF	dynamic iSSF with daily harmonics	
3	deepSSF	deepSSF model described above	
4	deepSSF S2	deepSSF model (Supporting Information)	12 Sentinel-2 bands, slope

Note: Both deepSSF models also included variables relating to time, which was a harmonic representation of the hour of the day and the day of the year.

penultimate 10% of data was used for the deepSSF validation process, which involves assessing the predictive performance through the loss function to determine whether to reduce the learning rate and when to terminate model fitting to prevent overfitting. Although it is not used to update the parameters of the model, it is still considered to be used in the training process, hence the use of the test dataset, which is completely novel to the model.

2.7 | Software packages

We used a variety of software packages to perform data exploration and processing, statistical analysis and model fitting, as well as model interpretation and plotting. In R (R Core Team, 2024) the primary packages that we used were `tidyverse` (Wickham et al., 2019) and `amt` (Signer et al., 2019) for data processing, `terra` (Hijmans, 2024) for processing raster data, and `ggplot2` (Wickham, 2016) and `ggpubr` (Kassambara, 2023) for plotting. In Python, the primary packages we used were `NumPy` (Harris et al., 2020) and `pandas` (McKinney, 2010) for data processing, `rasterio` (Gillies, 2013) for processing raster data, `PyTorch` (Paszke et al., 2019) for building and training the deepSSF models, and `matplotlib` (Hunter, 2007) for plotting.

3 | RESULTS

On a central processing unit (CPU—Dell Latitude 7420 with Intel Core i7 11th Gen 3.0 GHz and 16 GB RAM) the deepSSF model took 1–2 h to train, which typically comprised between 50 and 80 epochs (complete iterations through the training data), depending on the initial conditions. When trained on a graphics processing unit (GPU) using Google Colaboratory (2024) or a Macbook Pro with M4 chip, the training time reduced to 10–20 min.

The deepSSF model was able to capture the habitat selection and movement dynamics of the buffalo's data (Figure 2), resulting in simulated trajectories that represented important features of the data (Figure 3). The deepSSF approach generated trajectories that were visually similar to the observed data (Fieberg et al., 2024); although as the simulations did not have a memory or centralising tendency, they wandered more widely than the observed data, and there was less obvious revisitation structure to the trajectories.

The model clearly captured temporal dynamics in both the movement and habitat selection processes, which interacted over both hourly (Figure 4) and seasonal (Figure 5) time scales. The model captured the stationary distributions of the step lengths and turning angles well (Supporting Information). However, when log-transformed, the evidence of discretising the movement surface to fit and simulate from the model is clear (Supporting Information). As the discrepancy is for step lengths less than 10 m, and the overall distribution and temporal dynamics are well captured, this is unlikely to cause problems when generating predictions. The observed turn-angle distribution also did not follow a standard von Mises distribution due to the plateauing of turn angles away from the peak, but this was also captured reasonably well by the flexibility of the mixture of the von Mises distributions, despite flattening the peak of turning angles close to 0. The deepSSF model was able to capture the hourly temporal dynamics in habitat selection (Figure 4).

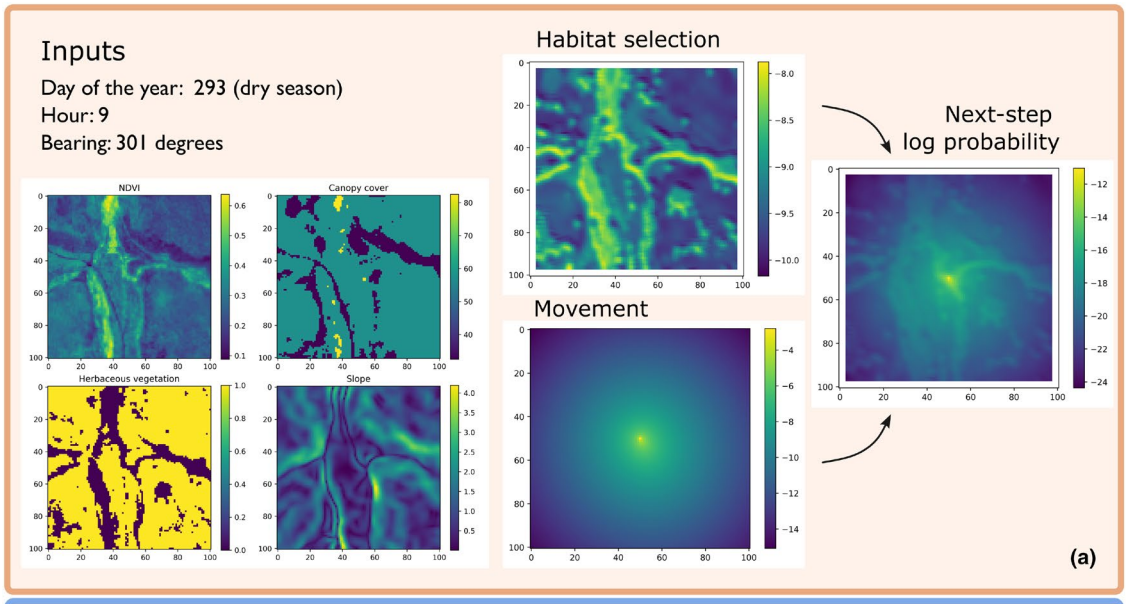
The landscape-scale habitat selection maps clearly show the representation of hourly and seasonal dynamics of habitat selection (Figure 5). The habitat selection varies significantly throughout each day, which for buffalo often has opposing trends at certain times of the day (Supporting Information). Across the year, the habitat selection also changes, which in this ecosystem is likely driven by the distribution of water (Figure 5).

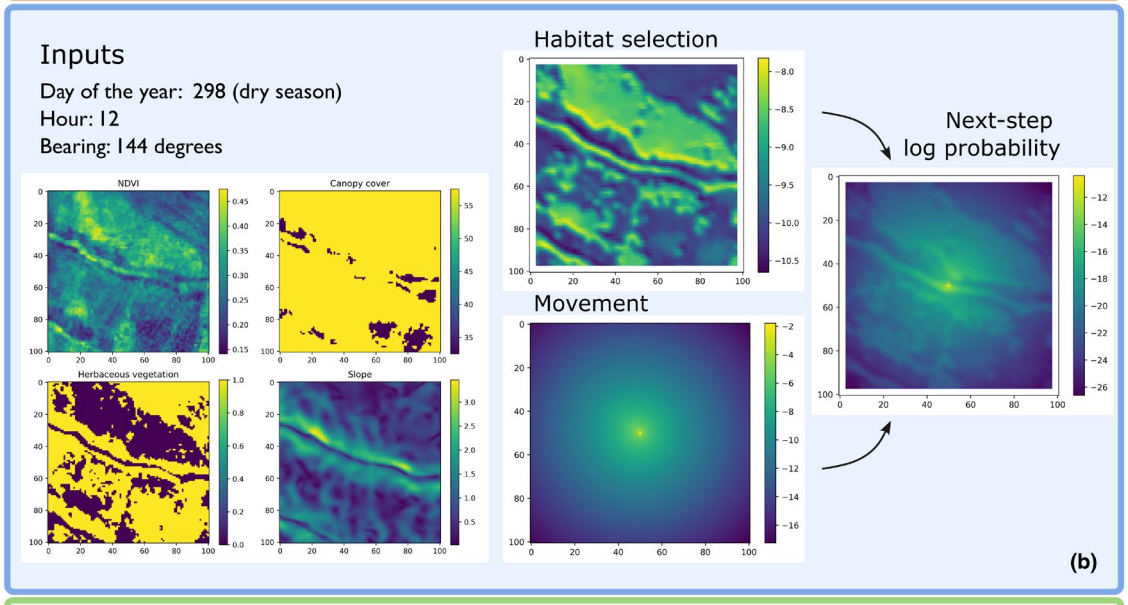
When compared to typical iSSF models, the deepSSF models had higher average prediction values for habitat selection, movement and next-step probabilities at the location of the next step, particularly for the deepSSF model fitted to 12 Sentinel-2 spectral bands and slope (Figure 6; Table 2). The higher performance of the deepSSF S2 model suggests that these raw layers contain more information that is relevant to buffalo than oft-used derived measures like NDVI, particularly in the dry season. This was a similar result to when the probabilities were binned into the hours of the day, where the night-time performance of the deepSSF model was high, but it was lower during the day, again in contrast to the deepSSF S2 model.

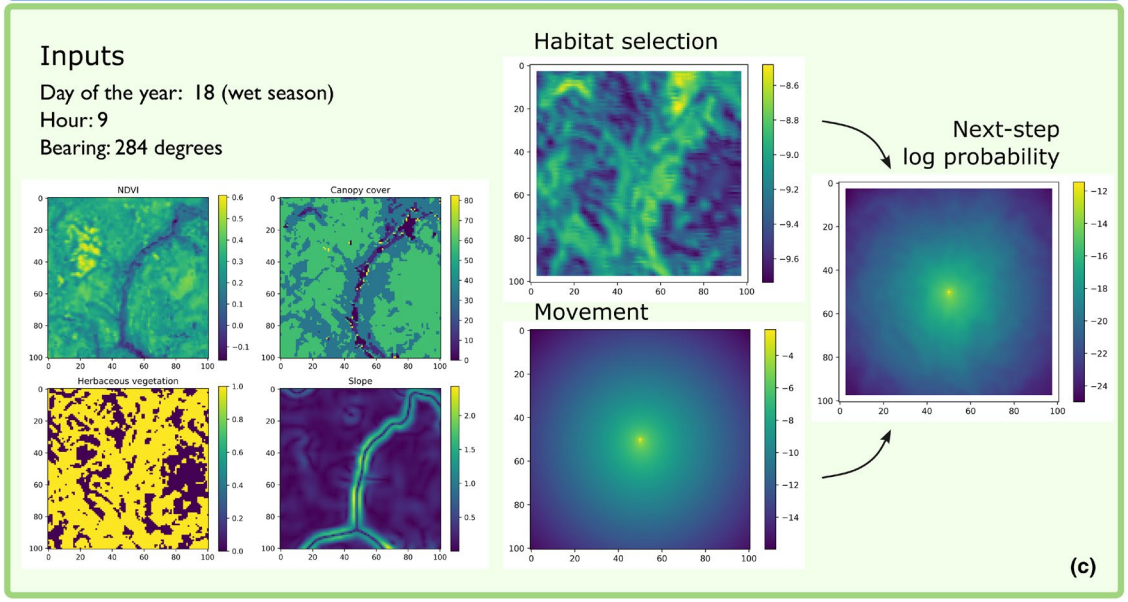
4 | DISCUSSION

We have presented an example of how deep learning can be applied to animal movement data, with promising results for generating realistic and accurate trajectories of animal movement, and potential for revealing subtle properties of animal movement behaviour.

FIGURE 2 Here we show three example steps, with the inputs, intermediate habitat selection and movement predictions, and the resulting next-step log probabilities outputted from a deepSSF model (Model 3 in Table 1). The day of the year and hour inputs listed above the spatial covariates were decomposed into sine and cosine components and these values were converted to grids of the same spatial extent prior to inputting to the model. The bearing was added directly to the μ parameters of the von Mises distributions to deviate the turning angle from the previous step. The intermediate outputs provide an indication of what the model considers to be important for predicting the location of the next step. In panel (a), there is higher selection for the areas that have woody vegetation (when bottom left covariate is 0), with higher values of normalised difference vegetation index (NDVI), and low values of slope along a watercourse feature. In panel (b), there is high selection for the low slope feature running laterally (which is a watercourse), but also high selection for the southern edge of the woody area. In panel (c), the covariate values are of similar magnitude, but the model identifies several areas in the north of the covariates as a high probability of selection, which are localised areas of floodplain that are known to be used by buffalo, particularly during the wet season.







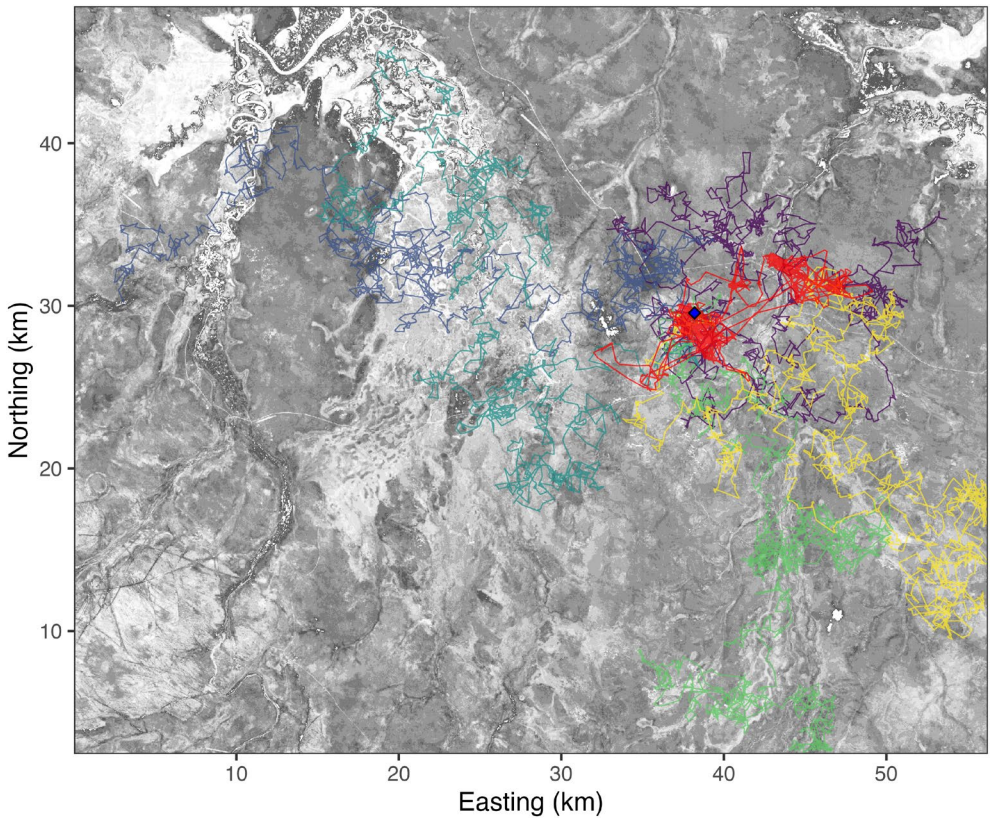


FIGURE 3 Example trajectories of hourly locations that were simulated from the starting location of the observed trajectory using the deepSSF model, with normalised difference vegetation index (NDVI) as the background (with higher values as darker grey). The red trajectory is the observed buffalo trajectory, and the other colours are five simulated trajectories. Both the observed and simulated trajectories have 3000 steps each. Without a memory process, the trajectories of the simulated data do not have any home ranging behaviour and wander more widely than the observed data. There are several instances of directed and high speed movement by the buffalo which were not replicated by the deepSSF model. This is due to the size of the predicted probability surfaces, which in our case we chose to include 96% of the observed step lengths, but which place an upper limit on possible step lengths and will prevent very large movements.

Deep learning is a highly flexible and data-driven approach, requiring minimal pre-emptive specification about the functional form of the model. This allows the model to learn its own representation about the nature of the movement and habitat selection processes represented by the deepSSF model, which lends a robustness to the model's outputs. We have also shown that we do not necessarily require 'big data' for deep learning, here using only a single individual's data with ~10,000 GPS locations (and of those the model was trained on only ~8000, with the remainder used for validation and testing).

As we expected, the deepSSF models outperformed the iSSF models for the in-sample and out-of-sample data, which was particularly the case when the model was trained with Sentinel-2 spectral bands and slope as the spatial covariates (deepSSF S2). The habitat selection performance dropped slightly on the out-of-sample data for all models (including iSSFs), although there was minimal evidence for overfitting. The performance of the deepSSF S2 model trained directly on Sentinel-2 layers suggests that these inputs contain more information that is relevant to buffalo movement and habitat selection than derived quantities like NDVI and canopy cover. The value of using raw layers was particularly clear for predictions in the dry season and for the hours during the middle of the day, where the deepSSF model had a lower performance compared to the deepSSF S2 model. In general, this suggests that, when using deep learning as a model for predicting animal movement, users should make as few a priori derivations as possible, instead providing the model with the rawest data possible and letting it find the relevant covarying derived quantities itself. However, the explainability of the model outputs when fitted directly to the Sentinel-2 data was lower as it is not as meaningful to correlate the input values with the predicted outputs, suggesting an explainability-predictability trade-off (Dwivedi et al., 2023; Henriques et al., 2024).

Assessing how the predictive performance of the models and covariates varies across different time periods is informative not only for the models' overall performance but also for the relevance of the spatial covariates, as lower predictive performance suggests that the buffalo were responding to environmental features not captured in the inputs, which can be used to guide model refinement.

As we only fitted the deepSSF model to the data from a single individual, it is likely that it would benefit from being fitted to more data to represent the broader environmental space that the buffalo population in this area is likely to encounter. A model with a partial-pooling or hierarchical representation (i.e. a 'mixed model') may be

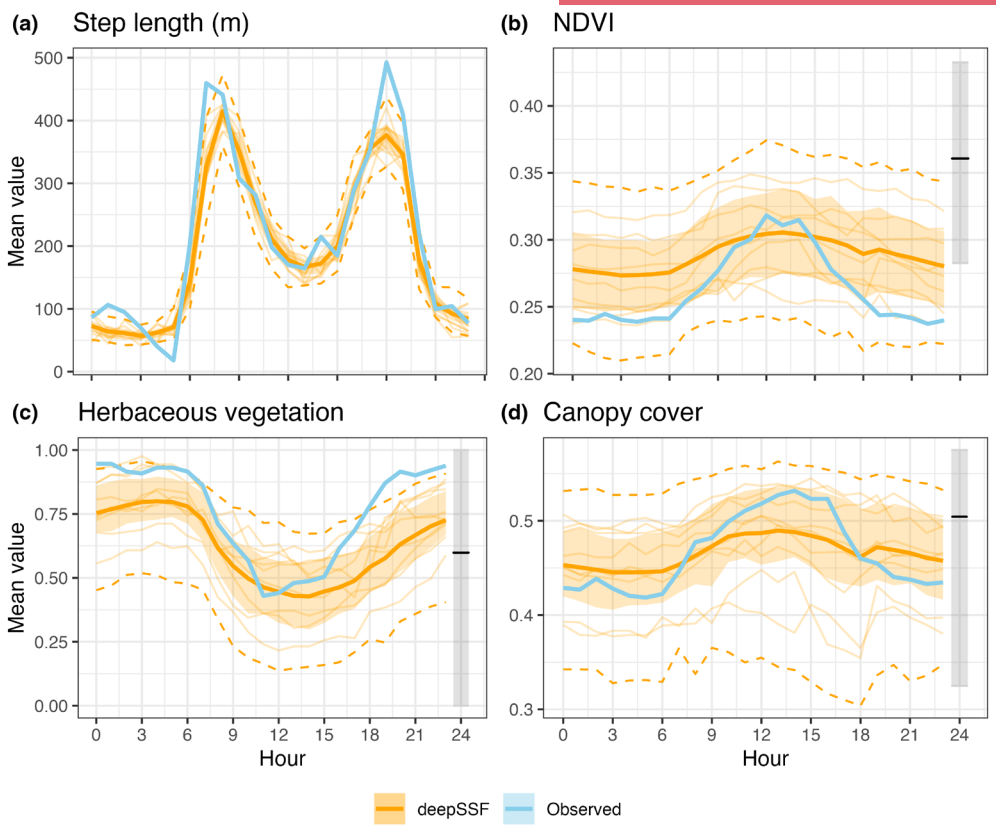


FIGURE 4 To assess the temporally dynamic habitat use of the simulated trajectories, we binned all simulated (orange) and observed (blue) steps into the hours of the day, and compared the mean values of the step lengths and each covariate for each hour of the day. We simulated 50 trajectories with 3000 steps each (as above) that covered the late-dry season 2018. We took a subset of the observed data that covered the same period. The deepSSF model was able to capture the hourly temporal dynamics in animal movement, similar to the results of temporally dynamic step selection functions in Forrest, Pagendam, et al. (2024). The shaded ribbons enclose the 25% and 75% quantiles, and the dashed lines are the 2.5% and 97.5% quantiles. The solid coloured line is the mean for that hour across all trajectories or for the buffalo data, and we show the hourly mean for 10 simulated individuals as thin lines. The grey bar on the right encloses the 25% and 75% quantiles of the values of all cells in the landscape in Figure 3, thereby quantifying the available background.

particularly useful here, as the data from each individual would be considered as more dependent than the data between individuals (Muff et al., 2020). Until a solution in this area is developed, a deepSSF model can be trained on all of the observed data, without representation in the model that the data come from different individuals (complete pooling); or separate models can be trained on the data from different individuals (such as we have here), creating a 'population' of individuals to simulate from.

An additional consideration when fitting deep learning models is that model training is stochastic, and due to the flexibility of the model, there are likely to be multiple minima that can be found by the model. This means that fitting the model with the same parameters may result in varying performance and slightly different predictions. Although the model is still likely to create a useful representation of the data and generate accurate predictions, exactly what the model has learned will differ. This observation may also relate to the amount and complexity of the data, which we did not explore here.

Deep learning models are not interpretable in the same way as a simpler statistical model (i.e. we cannot understand the influence of particular parameters such as through coefficients), but in our case we have model explainability through visualising the intermediate (movement and habitat selection predictions) and final outputs (next-step predictions) from the model, summarising the trajectories, and from correlating the covariate values with the habitat selection probability values. From these model outputs we were able to determine that there is substantial daily variation, and some yearly variation, in buffalo movement and habitat selection behaviour, which corroborates with fine-scale dynamic iSSFs (Forrest, Pagendam, et al., 2024) and with the observed data. The model was also able to represent multi-scale temporal dynamics, with daily movement and habitat selection behaviour that also changed throughout the year. Whilst we acknowledge that understanding what the deepSSF model has learned requires some more investigation by the user, we also believe that because animal movement behaviour is a sophisticated process, the outputs of typical statistical models such as iSSFs can oversimplify the process, potentially obscuring subtle insights into the behaviour of the species and the nature of the movement process.

There have been several other papers which have provided a conceptual foundation for this study. Dalziel et al. (2008) offered

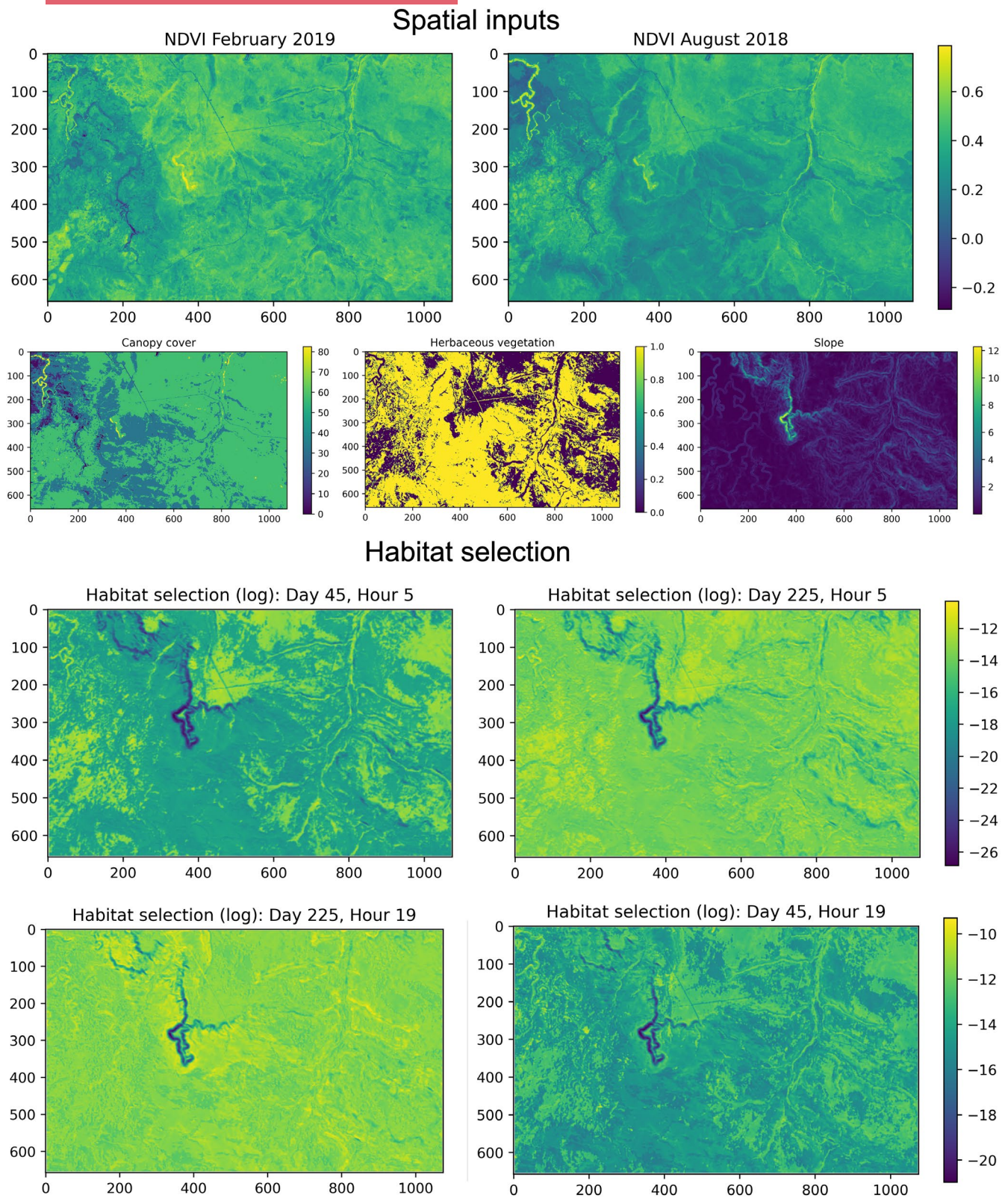


FIGURE 5 As convolution filters are fixed in size and apply transformations uniformly across grids, they can be applied to any landscape extent to approximate habitat selection (without accounting for movement dynamics). Here we show habitat selection at two times, 5 AM and 7 PM, in the middle of the wet season (yday = 45) and the dry season (yday = 225). The spatial covariates are shown in the top panels, and the habitat predictions for each day and hour are shown below. The units of the x- and y-axes are the cell indices, and the landscape is therefore ~25 km × ~15 km with 25 m cells. These plots illustrate that the model was able to represent the multi-scale temporal dynamics of habitat selection that was present in the observed data. It should be noted that these maps should not be interpreted as long-term spatial distributions, but only as habitat selection surfaces that ignore the animal's movement dynamics.

a similar approach to what we have presented here, although the study was not widely adopted; since then, deep learning has advanced rapidly. Cífka et al. (2023) used transformers to train a general model (MoveFormer) for animal movement, which also replicated a step selection function. This approach shows very promising results for representing memory by incorporating a history of locations and using a transformer architecture, although applying the model to a wide range of species demanded reducing the location frequency, ignoring fine-scale processes. This approach also required proposing random steps which are compared to the observed steps (as in iSSFs), rather than directly inputting spatial information which can allow for spatial features to influence movement. Given the modular nature of deep learning, it is possible that components of the deepSSF approach and MoveFormer may be combined to represent fine-scale behavioural processes while using the more flexible (but data-hungry) transformer architecture, and may reveal insights about the spatial memory process. Another promising approach for representing sequences is via (deep) reinforcement learning (Villeneuve et al., 2021), which may also be combined with other processing layers such as we have here (Chen et al., 2024), and can provide accurate movement predictions for several steps into the future.

Whilst we present one possible deepSSF model here, there are a number of exciting possible developments that would lead to a more realistic representation of animal movement. Firstly, a straightforward extension is to retain the deepSSF architecture and processes that we presented, but to add more information to the inputs of the movement process (and/or into the habitat selection process), which may also include a representation of past locations (Ellison, Potts, Strickland, et al., 2024; Oliveira-Santos et al., 2016; Rheault et al., 2021; Schlägel & Lewis, 2014). In our configuration, the model only received the bearing of the previous step (to confer turn-angle persistence) as well as the surrounding spatial covariates. It would be straightforward to incorporate a greater number of previous turning angles and previous step lengths, which would confer a persistence in speed and approximate a momentum process, and may be particularly helpful in cases of highly correlated movement kernels, such as high-frequency tracking data or for animal migrations. An example of such a movement process is Wijeyakulasuriya et al. (2020), where 39 variables, such as lagged (x, y) coordinates and lagged x, y velocities, were used at each time step to predict the movement of ants.

Secondly, there are a number of tools in the deep learning toolbox that may be applied to extend the modelling approach we have presented here. Potential processing architectures may include recurrent layers or transformer layers, which can accommodate long sequences as inputs, such as a history of locations, allowing the model to represent a flexible and data-driven memory process (Chen et al., 2024; Cífka et al., 2023; Shenk et al., 2021). An additional architecture that may be able to represent social dynamics may be graph neural networks (GNNs) (Zhou et al., 2020), where individuals in a population may be represented by nodes and their relationship to each other by edges, which can be parameterised from movement data (Scharf et al., 2016). A graphical approach may allow for predictive modelling at a population level with a socially informed model across many individuals. Both of these processes (memory and social dynamics) may allow for more accurate predictions but may also represent and reveal some of the more sophisticated and abstract mechanisms of these processes.

Thirdly, integrating other data sources, such as those from other biologging devices, is likely to allow for a more nuanced and comprehensive representation of animal movement behaviour that may also reduce the dependence on any particular temporal scale (Adam et al., 2019; DeRuiter et al., 2017; Leos-Barajas et al., 2017; Munden et al., 2021; Saldanha et al., 2023). This integration can be achieved due to the modular nature of deep learning, where a subnetwork may process sensor data such as from high-frequency accelerometers or magnetometers, which could be summarised to the same temporal scale as the locational data and may, for example, become an input to the movement subnetwork, where quantities of the observed activity would inform the movement kernel at the next step.

Finally, it may also be possible to train a 'foundational' or 'general' deepSSF model for animal movement from the data of many species, taking advantage of open-source databases such as Movebank (Kranstauber et al., 2011), similar to the approach of MoveFormer (Cífka et al., 2023). A model such as this could learn the general principles of animal movement dynamics, which could then be used as a pre-trained model and 'transferred' to a species of interest (Pan & Yang, 2010). Besides representing general animal movement dynamics, a purpose of a model such as this would be that less data would be required for new species, enabling accurate predictions with few data.

5 | CONCLUSIONS

Here we presented the deepSSF approach for modelling and predicting animal movement using deep learning. We developed an approach that represented animal movement behaviour as distinct but interacting processes, each represented by a subnetwork. This approach allows for greater modelling flexibility and modularity, and provides intermediate outputs that can be interpreted in direct reference to the input covariates. The deepSSF model we presented generated realistic trajectories, as well as provided insights into water buffalo's daily and seasonal movement and habitat selection behaviour in northern Australia. The deepSSF model fitted to derived covariates, such as NDVI and canopy cover, and the deepSSF model fitted to Sentinel-2 satellite imagery showed high predictive performance both in- and out-of-sample, across the full range of the data. The results for both models will likely improve when more data is used for training; and be better able to generalise under different environmental scenarios.

We kept the processes in our deepSSF model straightforward to exemplify the approach, but there are numerous possibilities for extensions. There are a wide array of computational architectures that can represent different features of the data; processes such as memory and social dynamics may be represented, and almost any

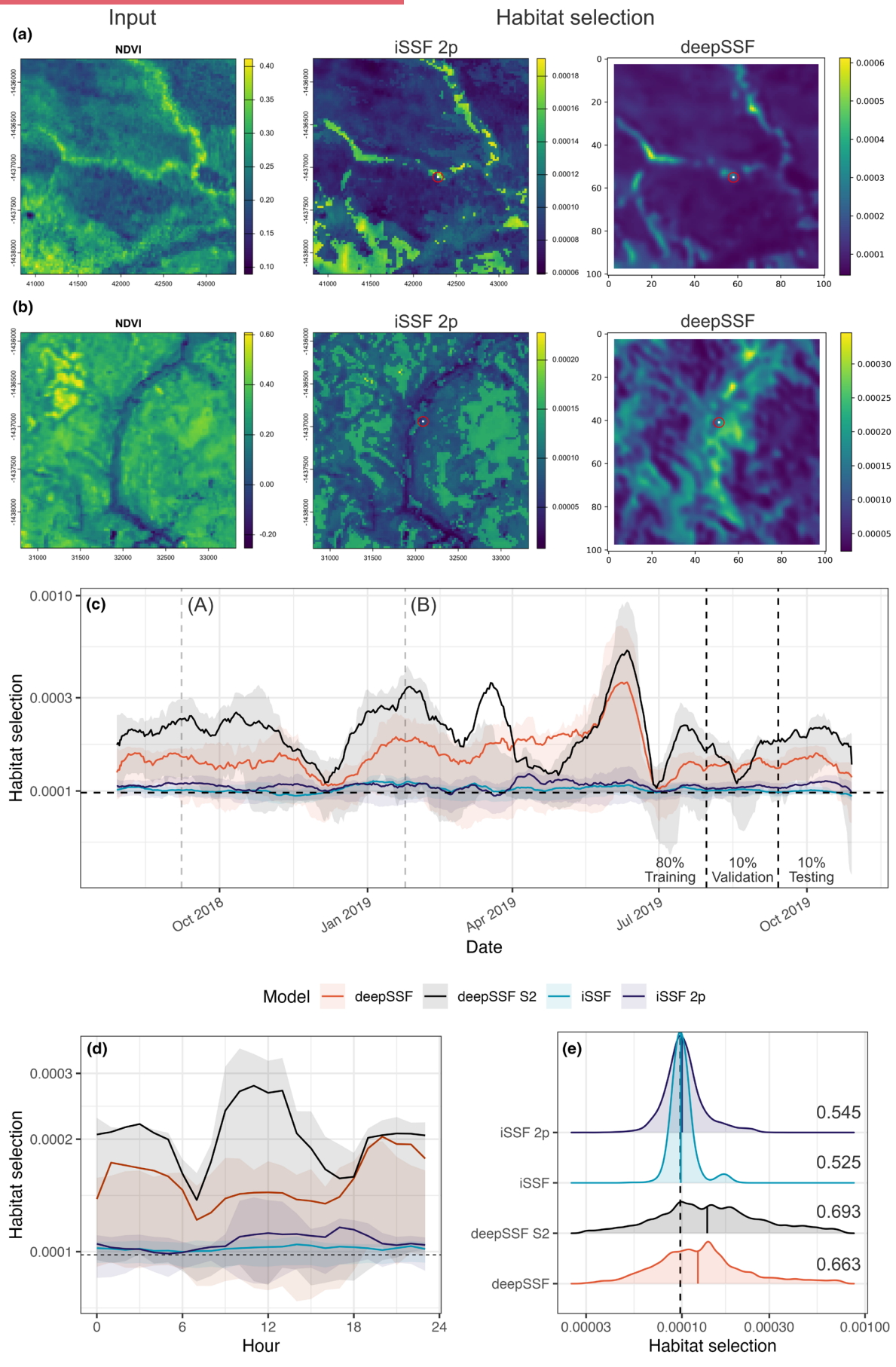


FIGURE 6 Probability values for the normalised habitat selection process at the observed location of the next step. We compare between the four models described in Table 1. Panels (a) and (b) show two example steps, with one of the inputs (NDVI), and the predicted habitat selection for the iSSF and deepSSF models, which we have normalised to sum to 1. We estimate the predictive accuracy by extracting the probability value at the observed location of the next step, which is shown as the white cell surrounding by a red circle for visibility. We follow this procedure to extract the habitat selection, movement and next-step probability for every step in the entire trajectory. In panel (c) we show the habitat selection probabilities, which have been smoothed with a rolling window of 15 days, with the mean (solid line) and 25% and 75% quantiles enclosed by the ribbon. The vertical lines at (a) and (b) show where in the trajectory the example steps were taken, and the dashed lines to the right indicate the 80/10/10 split of the data into training (model fitting), validation and test datasets. The *y*-axis is on the log-scale. We show a similar plot in panel (d), except here we take the average and 25% and 75% quantiles for each hour of the day. In panel (e) we show the distribution of predicted habitat selection probabilities for every step. The number to the right is the proportion of steps that had a predicted probability that was greater than the uniform probability (i.e., random selection). The deepSSF models often had higher probability values that were more 'confident', which led to more accurate predictions of average, but also more variability than the iSSF models.

TABLE 2 Average probability values (±SD) for the habitat selection, movement and next-step probability surfaces at the location of the next step, which we used to compare between iSSF and deepSSF models.

All probability values are 1×10^{-4}		In-sample ($n = 8080$ steps)		
Model	**Probability**	**Habitat**	**Movement**	**Next step**
0	*uniform*	*0.980*	*0.980*	*0.980*
1	iSSF	1.026 ± 0.194	177.4 ± 243.2	180.3 ± 248.0
2	iSSF 2p	1.063 ± 0.330	211.8 ± 312.5	218.2 ± 248.0
3	deepSSF	1.974 ± 2.481	702.0 ± 1411	752.5 ± 1488
4	deepSSF Sentinel-2	**2.158 ± 2.448**	**1221 ± 2426**	**1277 ± 2545**
All probability values are 1×10^{-4}		**Out-of-sample ($n = 1010$ steps)**		
Model	**Probability**	**Habitat**	**Movement**	**Next step**
0	*uniform*	*0.980*	*0.980*	*0.980*
1	iSSF	0.990 ± 0.108	235.3 ± 267.8	236.8 ± 271.1
2	iSSF 2p	1.075 ± 0.232	283.7 ± 267.8	301.9 ± 379.7
3	deepSSF	1.610 ± 0.943	1111 ± 1732	1258 ± 1975
4	deepSSF Sentinel-2	**1.918 ± 1.605**	**1896 ± 2960**	**1970 ± 3058**

Note: We calculated the habitat selection, movement and next step (the combination of habitat selection and movement) probability values for each step of the data that the models were fitted to. Each of the prediction surfaces was normalised to sum to one such that they were valid probability surfaces, which makes the probabilities at the observed next step comparable between models. The models are described in Table 1. The uniform values are shown in italics, and the bold values show the most accurate model for each dataset and process. The increased out-of-sample performance of the movement and next-step predictions are likely due to more predictable movement behaviour in that part of the trajectory.

input data can be integrated by adding subnetworks, which may process different inputs and produce outputs but can be combined in later stages of the model. Further developments are promising for high-frequency telemetry data and migration pathways, where a more complex movement process may better capture the highly autocorrelated structure of these data.

We consider the deepSSF approach to be a valuable addition to the toolbox for modelling animal movement, and we provide a combination of R and Python code to process data, train deepSSF models, and produce all outputs in this article (see Supporting Information). As deep learning is very flexible and requires few a priori assumptions, a deepSSF model can realistically model animal decision-making on fine scales, lending the potential to reveal subtle insights into animal movement behaviour that may be difficult to ascertain from parametric statistical models.

AUTHOR CONTRIBUTIONS

Scott Forrest, Dan Pagendam, Conor Hassan and Andrew Hoskins conceived and developed the deepSSF ideas; Andrew Hoskins designed the data collection; Scott Forrest and Dan Pagendam developed the modelling methodology; Scott Forrest analysed the data and led the writing of the manuscript. All authors contributed critically to discussions throughout the design and modelling process, to manuscript drafts, and gave final approval for publication.

ACKNOWLEDGEMENTS

SWF was supported by an Australian Government Research Training Program Scholarship and a CSIRO top-up scholarship. CD was supported by an Australian Research Council Future Fellowship (FT210100260). This work was funded by the federal government's Department of Agriculture, Fisheries and Forestry

under the Control tools and technologies for established pest animals and weeds competitive grants program and the Smart Farming Partnerships Program (round 2). Computational resources and services used in this work were provided by the eResearch Office, Queensland University of Technology, Brisbane, Australia. SWF thanks Charlotte Patterson and members of the Applied Mathematical Ecology (AMEG) and Bayesian Research and Applications Group (BRAG) for helpful discussions and feedback on the work, and Maryam Goldchin of CSIRO for help setting up a Python environment.

CONFLICT OF INTEREST STATEMENT

Michael Bode is an associate editor on *Methods in Ecology and Evolution*, but played no part in the reviewing or handling of this manuscript. All authors declare no other conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.70136.

DATA AVAILABILITY STATEMENT

Code to replicate all analyses, as well as information to help understand deep learning in the context of the deepSSF approach can be found at https://swforrest.github.io/deepSSF/, which is associated with the GitHub repository at https://github.com/swforrest/deepSSF. The GitHub repository is archived in a Zenodo repository at https://doi.org/10.5281/zenodo.16557091 (Forrest & Pagendam, 2025).

GENERATIVE AI STATEMENT

ChatGPT was used to generate a starting point for glossary entries, which were edited by the authors. GitHub Copilot was used during coding, predominately to generate straightforward and repetitive code. Generative AI was not used in the generation of the ideas or at any stage of writing of the manuscript.

ORCID

Scott W. Forrest https://orcid.org/0000-0001-9529-0108
Dan Pagendam https://orcid.org/0000-0002-8347-4767
Conor Hassan https://orcid.org/0000-0002-6200-2795
Jonathan R. Potts https://orcid.org/0000-0002-8564-2904
Christopher Drovandi https://orcid.org/0000-0001-9222-8763
Michael Bode https://orcid.org/0000-0002-5886-4421
Andrew J. Hoskins https://orcid.org/0000-0001-8907-6682

REFERENCES

Adam, T., Griffiths, C. A., Leos-Barajas, V., Meese, E. N., Lowe, C. G., Blackwell, P. G., Righton, D., & Langrock, R. (2019). Joint modelling of multi-scale animal movement data using hierarchical hidden Markov models. *Methods in Ecology and Evolution*, *10*(9), 1536–1550. https://doi.org/10.1111/2041-210x.13241

Avgar, T., Potts, J. R., Lewis, M. A., & Boyce, M. (2016). Integrated step selection analysis: Bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution*, *7*(5), 619–630. https://doi.org/10.1111/2041-210x.12528

Bello, C., Crowther, T., Ramos, D. L., Morán-López, T., Pizo, M. A., & Dent, D. H. (2024). Frugivores enhance potential carbon recovery in fragmented landscapes. *Nature Climate Change*, *14*, 636–643. https://doi.org/10.1038/s41558-024-01989-1

Campbell, H. A., Loewensteiner, D. A., Murphy, B. P., Pittard, S., & McMahon, C. R. (2020). Seasonal movements and site utilisation by Asian water buffalo (*Bubalus bubalis*) in tropical Savannas and floodplains of northern Australia. *Wildlife Research*, *48*(3), 230–239. https://doi.org/10.1071/WR20070

Chen, Z., Wang, D., Zhao, F., Dai, L., Zhao, X., Jiang, X., & Zhang, H. (2024). AnimalEnvNet: A deep reinforcement learning method for constructing animal agents using multimodal data fusion. *Applied Sciences*, *14*(14), 6382.

Cífka, O., Chamaillé-Jammes, S., & Liutkus, A. (2023). MoveFormer: A transformer-based model for step-selection animal movement modelling. *bioRxiv*, 2023.03.05.531080. https://doi.org/10.1101/2023.03.05.531080

Côrtes, M. C., & Uriarte, M. (2013). Integrating frugivory and animal movement: A review of the evidence and implications for scaling seed dispersal: Frugivory, animal movement, and seed dispersal. *Biological Reviews of the Cambridge Philosophical Society*, *88*(2), 255–272. https://doi.org/10.1111/j.1469-185X.2012.00250.x

Couzin, I. D. (2009). Collective cognition in animal groups. *Trends in Cognitive Sciences*, *13*(1), 36–43.

Cowan, M. A., Dunlop, J. A., Gibson, L. A., Moore, H. A., Setterfield, S. A., & Nimmo, D. G. (2024). Movement ecology of an endangered mesopredator in a mining landscape. *Movement Ecology*, *12*(1), 5. https://doi.org/10.1186/s40462-023-00439-5

Dalziel, B. D., Morales, J. M., & Fryxell, J. M. (2008). Fitting probability distributions to animal movement trajectories: Using artificial neural networks to link distance, resources, and memory. *The American Naturalist*, *172*(2), 248–258. https://doi.org/10.1086/589448

DeRuiter, S. L., Langrock, R., Skirbutas, T., Goldbogen, J. A., Calambokidis, J., Friedlaender, A. S., & Southall, B. L. (2017). A multivariate mixed hidden markov model for blue whale behaviour and responses to sound exposure. *The Annals of Applied Statistics*, *11*(1), 362–392. https://doi.org/10.1214/16-aoas1008

Drori, I. (2022). *The science of deep learning*. Cambridge University Press. https://doi.org/10.1017/9781108891530

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., & Ranjan, R. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, *55*(9), 1–33. https://doi.org/10.1145/3561048

Einarson, D., Frisk, F., Klonowska, K., & Sennersten, C. (2024). A machine learning approach to simulation of mallard movements. *Applied Sciences*, *14*(3), 1280. https://doi.org/10.3390/app14031280

Ellison, N., Potts, J. R., Boudreau, M. R., Börger, L., Strickland, B. K., & Street, G. M. (2024). Social interactions and habitat structure in understanding the dynamic space use of invasive wild pigs. *Wildlife Biology*, *2024*, e01247. https://doi.org/10.1002/wlb3.01247

Ellison, N., Potts, J. R., Strickland, B. K., Demarais, S., & Street, G. M. (2024). Combining animal interactions and habitat selection into models of space use: A case study with white-tailed deer. *Wildlife Biology*, *2024*, e01211. https://doi.org/10.1002/wlb3.01211

English, H. M., Börger, L., Kane, A., & Ciuti, S. (2024). Advances in biologging can identify nuanced energetic costs and gains in predators. *Movement Ecology*, *12*(1), 7. https://doi.org/10.1186/s40462-024-00448-y

Fieberg, J., Freeman, S., & Signer, J. (2024). Using lineups to evaluate goodness of fit of animal movement models. *Methods in Ecology and Evolution*, *15*, 1048–1059. https://doi.org/10.1111/2041-210x.14336

Finnegan, L., Viejou, R., MacNearney, D., Pigeon, K. E., & Stenhouse, G. B. (2021). Unravelling the impacts of disturbance type and

regeneration on movement of threatened species. *Landscape Ecology*, *36*(9), 2619–2635. https://doi.org/10.1007/s10980-021-01259-x

Forrest, S., & Pagendam, D. (2025). Swforrest/deepSSF: deepSSF MEE v1.0.1. *Zenodo*, https://doi.org/10.5281/ZENODO.16557091

Forrest, S. W., Pagendam, D., Bode, M., Drovandi, C., Potts, J. R., Perry, J., Vanderduys, E., & Hoskins, A. J. (2024). Predicting fine-scale distributions and emergent spatiotemporal patterns from temporally dynamic step selection simulations. *Ecography*, *2025*, e07421. https://doi.org/10.1111/ecog.07421

Forrest, S. W., Rodríguez-Recio, M., & Seddon, P. J. (2024). Home range and dynamic space use reveals age-related differences in risk exposure for reintroduced parrots. *Conservation Science and Practice*, *6*, e13119. https://doi.org/10.1111/csp2.13119

Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T., & Mao, J. S. (2005). Wolves influence elk movements: Behavior shapes a trophic cascade in yellowstone national park. *Ecology*, *86*(5), 1320–1330. https://doi.org/10.1890/04-0953

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/AOS/1013203451

Gillies, S. (2013). *Rasterio: Geospatial raster I/O for python programmers*.

Google Colaboratory. (2024). Google Colaboratory. [Accessed: 2024-9-23].

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Henriques, J., Rocha, T., de Carvalho, P., Silva, C., & Paredes, S. (2024). Interpretability and explainability of machine learning models: Achievements and challenges. In *Ifmbe proceedings* (pp. 81–94). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-59216-4_9

Hertel, L., Barth, E., Kaster, T., & Martinetz, T. (2015). Deep convolutional neural networks as generic feature extractors. In *2015 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–4). IJCNN. https://doi.org/10.1109/ijcnn.2015.7280683

Hijmans, R. J. (2024). *Spatial data analysis* [r package terra version 1.7–71].

Hofmann, D. D., Cozzi, G., McNutt, J. W., Ozgul, A., & Behr, D. M. (2023). A three-step approach for assessing landscape connectivity via simulated dispersal: African wild dog case study. *Landscape Ecology*, *38*(4), 981–998. https://doi.org/10.1007/s10980-023-01602-4

Hooker, M. J., Clark, J., Bond, B. T., & Chamberlain, M. J. (2021). Evaluation of connectivity among American black bear populations in Georgia. *The Journal of Wildlife Management*, *85*(5), 979–988. https://doi.org/10.1002/jwmg.22041

Horn, B. K. P. (1981). Hill shading and the reflectance map. *Proceedings of the IEEE*, *69*(1), 14–47. https://doi.org/10.1109/PROC.1981.11918

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kassambara, A. (2023). *Ggpubr: 'ggplot2' based publication ready plots*.

Kazama, K., Fujita, K., Shinoda, Y., & Koike, S. (2024). Sika deer trajectory prediction considering environmental factors by timeseries transformer-based architecture. *Expert Systems with Applications*, *250*, 123630. https://doi.org/10.1016/j.eswa.2024.123630

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv [cs.LG]*.

Klappstein, N. J., Michelot, T., Fieberg, J., Pedersen, E. J., & Mills Flemming, J. (2024). Step selection functions with non-linear and random effects. *Methods in Ecology and Evolution*, *15*, 1332–1346. https://doi.org/10.1111/2041-210x.14367

Kranstauber, B., Cameron, A., Weinzerl, R., Fountain, T., Tilak, S., Wikelski, M., & Kays, R. (2011). The movebank data model for animal tracking. *Environmental Modelling & Software*, *26*(6), 834–835. https://doi.org/10.1016/j.envsoft.2010.12.005

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Leos-Barajas, V., Gangloff, E. J., Adam, T., Langrock, R., van Beest, F. M., Nabe-Nielsen, J., & Morales, J. M. (2017). Multi-scale modeling of animal movement and general behavior data using hidden markov models with hierarchical structures. *Journal of Agricultural, Biological, and Environmental Statistics*, *22*(3), 232–248. https://doi.org/10.1007/S13253-017-0282-9

Lustig, A., James, A., Anderson, D., & Plank, M. (2019). Pest control at a regional scale: Identifying key criteria using a spatially explicit, agent-based model. *The Journal of Applied Ecology*, *56*(7), 1515–1527. https://doi.org/10.1111/1365-2664.13387

Mayer, M., Coleman Nielsen, J., Elmeros, M., & Sunde, P. (2021). Understanding spatio-temporal patterns of deer-vehicle collisions to improve roadkill mitigation. *Journal of Environmental Management*, *295*, 113148. https://doi.org/10.1016/j.jenvman.2021.113148

McKinney, W. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, *445*, 51–56.

Michelot, T., Klappstein, N. J., Potts, J. R., & Fieberg, J. (2024). Understanding step selection analysis through numerical integration. *Methods in Ecology and Evolution*, *15*(1), 24–35. https://doi.org/10.1111/2041-210x.14248

Muff, S., Signer, J., & Fieberg, J. (2020). Accounting for individual-specific variation in habitatselection studies: Efficient estimation of mixed-effects models using bayesian or frequentist computation. *The Journal of Animal Ecology*, *89*(1), 80–92. https://doi.org/10.1111/1365-2656.13087

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*, 2041-210X.13574. https://doi.org/10.1111/2041-210x.13574

Murphy, K. P. (2022). *Probabilistic machine learning: An introduction*. MIT Press.

Myneni, R. B., Hall, F. G., Sellers, P. J., & Marshak, A. L. (1995). The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*, *33*(2), 481–486. https://doi.org/10.1109/TGRS.1995.8746029

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

Oliveira-Santos, L. G. R., Forester, J. D., Piovezan, U., Tomas, W. M., & Fernandez, F. A. S. (2016). Incorporating animal spatial memory in step selection functions. *The Journal of Animal Ecology*, *85*(2), 516–524. https://doi.org/10.1111/1365-2656.12485

Pagendam, D., Janardhanan, S., Dabrowski, J., & MacKinlay, D. (2023). A log-additive neural model for spatio-temporal prediction of groundwater levels. *Spatial Statistics*, *55*, 100740. https://doi.org/10.1016/j.spasta.2023.100740

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/tkde.2009.191

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B.,

Fang, L., … Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. *arXiv [cs.LG]*.

Patterson, C. R., Lustig, A., Seddon, P. J., Wilson, D. J., & van Heezik, Y. (2024). Eradicating an invasive mammal requires local elimination and reduced reinvasion from an urban source population. *Ecological Applications*, *34*, e2949. https://doi.org/10.1002/eap.2949

Pay, J. M., Patterson, T. A., Proft, K. M., Cameron, E. Z., Hawkins, C. E., Koch, A. J., Wiersma, J. M., & Katzner, T. E. (2022). Considering behavioral state when predicting habitat use: Behavior-specific spatial models for the endangered tasmanian wedge-tailed eagle. *Biological Conservation*, *274*, 109743. https://doi.org/10.1016/j.biocon.2022.109743

Pili, A. N., Tingley, R., Chapple, D. G., & Schumaker, N. H. (2022). virToad: Simulating the spatiotemporal population dynamics and management of a global invader. *Landscape Ecology*, *37*(9), 2273–2292. https://doi.org/10.1007/s10980-022-01468-y

Potts, J. R., & Börger, L. (2023). How to scale up from animal movement decisions to spatiotemporal patterns: An approach via step selection. *The Journal of Animal Ecology*, *92*(1), 16–29. https://doi.org/10.1111/1365-2656.13832

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raghu, M., & Schmidt, E. (2020). A survey of deep learning for scientific discovery. *arXiv [cs.LG]*.

Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., & Ohlen, D. O. (1994). Measuring phenological variability from satellite imagery. *Journal of Vegetation Science*, *5*(5), 703–714. https://doi.org/10.2307/3235884

Rheault, H., Anderson, C. R., Jr., Bonar, M., Marrotte, R. R., Ross, T. R., Wittemyer, G., & Northrup, J. M. (2021). Some memories never fade: Inferring multi-scale memory effects on habitat selection of a migratory ungulate using step-selection functions. *Frontiers in Ecology and Evolution*, *9*, 702818. https://doi.org/10.3389/fevo.2021.702818

Rhodes, J. R., McAlpine, C. A., Lunney, D., & Possingham, H. P. (2005). A spatially explicit habitat selection model incorporating home range behavior. *Ecology*, *86*(5), 1199–1205. https://doi.org/10.1890/04-0912

Saldanha, S., Cox, S. L., Militão, T., & González-Solís, J. (2023). Animal behaviour on the move: The use of auxiliary information and semi-supervision to improve behavioural inferences from hidden markov models applied to GPS tracking datasets. *Movement Ecology*, *11*(1), 41. https://doi.org/10.1186/s40462-023-00401-5

Scharf, H. R., Hooten, M. B., Fosdick, B. K., Johnson, D. S., London, J. M., & Durban, J. W. (2016). Dynamic social networks based on movement. *The Annals of Applied Statistics*, *10*(4), 2182–2202. https://doi.org/10.1214/16-AOAS970

Schlägel, U. E., & Lewis, M. A. (2014). Detecting effects of spatial memory and dynamic information on animal movement decisions. *Methods in Ecology and Evolution*, *5*(11), 1236–1246. https://doi.org/10.1111/2041-210X.12284

Schlägel, U. E., & Lewis, M. A. (2016). Robustness of movement models: Can models bridge the gap between temporal scales of data sets and behavioural processes? *Journal of Mathematical Biology*, *73*(6–7), 1691–1726. https://doi.org/10.1007/s00285-016-1005-5

Sells, S. N., Costello, C. M., Lukacs, P. M., Roberts, L. L., & Vinks, M. A. (2023). Predicted connectivity pathways between grizzly bear ecosystems in western montana. *Biological Conservation*, *284*, 110199. https://doi.org/10.1016/j.biocon.2023.110199

Shenk, J., Byttner, W., Nambusubramaniyan, S., & Zoeller, A. (2021). Traja: A python toolbox for animal trajectory analysis. *Journal of Open Source Software*, *6*(63), 3202. https://doi.org/10.21105/joss.03202

Signer, J., Fieberg, J., & Avgar, T. (2017). Estimating utilization distributions from fitted stepselection functions. *Ecosphere*, *8*(4), e01771. https://doi.org/10.1002/ecs2.1771

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*(2), 880–890. https://doi.org/10.1002/ece3.4823

Signer, J., Fieberg, J., Reineking, B., Schlägel, U., Smith, B., Balkenhol, N., & Avgar, T. (2023). Simulating animal space use from fitted integrated step-selection functions (iSSF). *Methods in Ecology and Evolution*, *15*, 43–50. https://doi.org/10.1111/2041-210x.14263

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research: JMLR*, *15*(56), 1929–1958. https://doi.org/10.5555/2627435.2670313

Strandburg-Peshkin, A., Farine, D. R., Couzin, I. D., & Crofoot, M. C. (2015). Shared decisionmaking drives collective movement in wild baboons. *Science*, *348*(6241), 1358–1361. https://doi.org/10.1126/science.aaa5099

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 2017-December.

Villeneuve, C., Dulude-De Broin, F., Legagneux, P., Berteaux, D., & Durand, A. (2021). Preserving the integrity of the canadian northern ecosystems through insights provided by reinforcement learning-based arctic fox movement models. *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*.

Whittington, J., Hebblewhite, M., Baron, R. W., Ford, A. T., & Paczkowski, J. (2022). Towns and trails drive carnivore movement behaviour, resource selection, and connectivity. *Movement Ecology*, *10*(1), 17. https://doi.org/10.1186/s40462-022-00318-5

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., Fran¸cois, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wijeyakulasuriya, D. A., Eisenhauer, E. W., Shaby, B. A., & Hanks, E. M. (2020). Machine learning for modeling animal movement. *PLoS One*, *15*(7), e0235750. https://doi.org/10.1371/journal.pone.0235750

Williams, H. J., Taylor, L. A., Benhamou, S., Bijleveld, A. I., Clay, T. A., de Grissac, S., Demˇsar, U., English, H. M., Franconi, N., Gómez-Laich, A., Griffiths, R. C., Kay, W. P., Morales, J. M., Potts, J. R., Rogerson, K. F., Rutz, C., Spelt, A., Trevail, A. M., Wilson, R. P., & Börger, L. (2020). *Optimizing the use of biologgers for movement ecology research*. https://doi.org/10.1111/1365-2656.13094

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, *1*, 57–81. https://doi.org/10.1016/J.AIOPEN.2021.01.001

SUPPORTING INFORMATION

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

Supplementary Material. Glossary.

Appendix A. Deep learning concepts.

Appendix B. Training a deepSSF model with Sentinel-2 imagery.

Appendix C. Additional deepSSF model outputs.

Appendix D. Accounting for change of movement variables.

Appendix E. Consideration of deepSSF 'hyperparameters'.

Appendix F. Issues arising from discretising the movement probability surface.

Supplementary Animation. Habitat selection log-probability surfaces for the first 50 hourly steps of an observed trajectory, for the deepSSF model (centre panel) and the iSSF model with two pairs of harmonics (right panel - denoted Model 2p).

How to cite this article: Forrest, S. W., Pagendam, D., Hassan, C., Potts, J. R., Drovandi, C., Bode, M., & Hoskins, A. J. (2025). Predicting animal movement with deepSSF: A deep learning step selection framework. *Methods in Ecology and Evolution*, *00*, 1–21. https://doi.org/10.1111/2041-210X.70136