**Title:** Choices to landscapes: Mechanisms of animal movement scale to landscape patterns

**Authors and affiliations:**
Will Rogers[1], will.rogers@yale.edu https://orcid.org/0000-0001-9899-8064
Scott Yanco[1,2] scott.yanco@yale.edu https://orcid.org/0000-0003-4717-9370
Walter Jetz[1,2], walter.jetz@yale.edu https://orcid.org/0000-0002-1971-7277

[1] Department of Ecology & Evolutionary Biology, Yale University, P. O. Box 208106, New Haven, CT 06520-8106
[2] Yale Center for Biodiversity and Global Change

**Running title:** Movement processes resolve spatial patterns of occurrence

**Keywords:** Species distribution, Habitat Selection, Occurrence, Movement, Step-selection, Ergodicity

**Statement of authorship:** WR and SY conceptualized the paper. WR wrote the code and carried out the analyses supporting the manuscript. WR wrote the initial draft with contributions from SY. All authors provided feedback on the concept and contributed to the final draft.

**Corresponding author:**
Name: Will Rogers
Email: will.rogers@yale.edu, rogerswill47@gmail.com
Mail: Department of Ecology & Evolutionary Biology, Yale University, P. O. Box 208106, New Haven, CT 06520-8106
Address: 165 Prospect Street, New Haven, CT, US 06511
Phone: (203) 432-6138
Fax: (203) 432-5176

**Abstract:**

Understanding the geographic distributions of animals is central to ecological inquiry and conservation planning. Movement-based habitat selection models, like discrete-time step-selection functions, are useful for identifying key environmental attributes which animals select upon and can nearly perfectly capture preference patterns between environmental conditions. However, predictions based only on selection can often fail to accurately describe true geographic distributions. We show that by integrating local selection patterns, movement models, and explicit landscape constraints within the same framework, we can better predict simulated distributions than occurrence-based frameworks of selection alone. Using three case studies, we show that this framework can better predict distributions of organisms across increasing scales of out-of-sample prediction: within individuals, between individuals, and even between regional contexts. Movement modeling is a powerful tool to describe both the selection paradigm that organisms apply to the environment and the movement patterns that enable them to apply those selection paradigms. By understanding movement and selection at the same time, habitats that species "prefer" may not be ultimately occupied, as many "preferred" patches may be unreachable given movement strategies and landscape context, often leading to smaller geographic distributions than estimated by selection or occurrence alone.

**Introduction:**

Interactions between organisms and their environment shape fundamental ecological phenomena such as species ranges (Kearney & Porter 2009), population dynamics (Roughgarden 1975; Kareiva *et al.* 1990), community composition (Menge & Olson 1990), and evolutionary feedbacks (Post & Palkovacs 2009). These relationships also provide critical applied insights about, for example, spatiotemporal patterns of biological diversity (Buckley *et al.* 2011; Woodin *et al.* 2013; Evans *et al.* 2015), population trends (Bozinovic *et al.* 2011; Madin *et al.* 2012), or genetic connectivity (Cushman &

Lewis 2010; Creel *et al.* 2019; Fletcher Jr *et al.* 2022) – insights necessary to address the twin planetary-scale ecological crises of climate change and biodiversity loss (Powers & Jetz 2019). Therefore, there is significant value in frameworks that can link the relationship between organisms and their environments to their niches (Kearney 2006; Matthiopoulos 2022), and make predictions from such relationships about the distribution of individuals and species geographically

Inferring relationships between individuals and their environment has traditionally relied on observations of occurrence patterns over some set of environmental contexts which vary geographically (Boyce & McDonald 1999; Boyce *et al.* 2002), typically by using environmental dimensions to contrast observations from locations where organisms are putatively absent (Mackey & Lindenmayer 2001; Johnson *et al.* 2006; Moorcroft & Barnett 2008). For example, species distribution models (SDMs) compare the patterns of presences as compared to possible null distributions over large geographic extents (Elith *et al.* 2011; Merow *et al.* 2014). Resource selection functions (RSFs) - an methodologically similar framework to SDMs (Fieberg *et al.* 2021; Northrup *et al.* 2022) - use logistic regression to model the probability of individual occurrences within some domain of predefined availability, like home ranges (Boyce & McDonald 1999; Manly *et al.* 2007). Both SDMs and RSFs are occurrence-based frameworks, capture information about multiple biological mechanisms that lead to species or individual occurrences while accounting for elements of the observation process itself (Keating & Cherry 2004; Phillips *et al.* 2009; Sastre & Lobo 2009; Wisz & Guisan 2009; Newbold *et al.* 2010). Notably, SDMs and RSFs assume that all environmental conditions within the landscape, an animal's home range, etc. are available for that animal to have selected upon, irrespective of the geographic configuration of these environments.

In reality, not all environmental conditions within some geographic extent may be available to an individual, either because of physiological constraints on animal movement or the spatial arrangement of habitats (Box 1). For example, geographically isolated patches of preferred

environmental conditions may be unreachable by an organism and, therefore, go unused (Soberon & Peterson 2005; Matthiopoulos *et al.* 2020; Van Moorter *et al.* 2023). Thus, patterns of animal movement interacting with underlying environmental preferences may strongly influence realized patterns of spatial occurrence if movement constraints or geographic heterogeneity (or both) are severe (Fig. 1). This is an important problem, as both SDMs and RSFs are unequipped to describe the movement as a force driving patterns of environmental selection. Unlike these occurrence-based frameworks, Step-selection functions (SSFs) use regularly-spaced GPS locations to estimate patterns of selection between environmental conditions that are available based on the movement constraints of the organism, comparing patterns of selection with each "step" in a GPS track (Avgar *et al.* 2016). As a result, SSFs may capture more realistic relationships between organisms and their environment by more accurately describing availability in local geographic extents.

While SSFs incorporate movement in selection processes and can even integrate selection with movement characteristics (Avgar *et al.* 2016), there is currently not analytical method to use SSFs for predicting geographic distributions of occurrence. SSFs are a form of conditional logistic regression, meaning that model inference about preference for or against environments at the end of a step is conditioned on the environments at the beginning of the step. Therefore, spatial predictions can only be made in comparison to some baseline conditions. In practice, researchers typically select some arbitrary set of environmental conditions (frequently the averager or median conditions within the prediction domain) to use as the basis for prediction. This projection of SSFs ignores the movement constraints inherent in model parametrization and instead generates patterns of "selection" across a landscape which may not be equivalent to actual relative use (Avgar *et al.* 2017). The inability of SSFs to directly return geographical predictions remains a hurdle for the applicability of movement models for ecological inference and conservation planning. To overcome this, some authors have suggested agent-based simulations based on parameterized SSF Field (Potts & Borger 2023), which provide the ability to

102 truly interrogate SSF fit but may invoke computational constraints with no guarantee of ergodicity.
103 Similarly, multi-stage frameworks have used separate resource selection (to assign quality to locations)
104 and step selection components (to assign connectivity between locations) to estimate geographic
105 distributions based (Van Moorter *et al.* 2023), which provides scalability but discretizes mechanisms of
106 movement and selection. Effectively capturing both the patterns of selection in environmental space
107 and then reprojecting those environmental conditions onto geographic space while considering
108 movement constraints is critical for our ability to link relationships between individuals and
109 environments to geographic distributions useful for ecological inferences (Schick *et al.* 2008;
110 Matthiopoulos 2022).

111 Here, we outline a conceptual approach to use SSFs as a model of concurrent movement and
112 selection processes to return geographical output about the probability of use that should (i)
113 parametrize organismal-environmental interactions more mechanistically than occurrence-based
114 frameworks and (ii) generate more accurate predictions of habitat distribution than occurrence or
115 selection frameworks that ignore movement when defining habitat over geographic domains.
116 Specifically, we use fitted SSFs to generate stochastic matrices of movement at local scales informed by
117 concurrent selection and movement limitations, and we retrieve stable-state distributions (SSDs) within
118 an ergodic framework. Because SSFs measure organismal-environment interactions at finer scales than
119 occurrence-based frameworks, we sought to test how occurrence-based (RSF), selection alone (SSF), and
120 concurrent selection and movement (SSD) predicted the distributions of organisms across the
121 landscape. We hypothesized that our framework incorporating both selection and movement would be
122 more accurate than frameworks based on occurrence or selection alone. We tested this hypothesis in a
123 simulation with known movement and selection constraints, as well as three biological datasets.
124

**Box 1: Individual fundamental preferences are not equivalent to realized use**

The suite of various ecological and environmental conditions necessary for species to persist is often conceptualized as a species' niche. This concept of the niche is divorced from spatial variation such that positions in niche space may correspond to multiple geographic locations (or none), a concept known as Hutchinson's duality (see Colwell & Rangel 2009). Because of this duality, it is possible to project a niche on a map, charting the areas that meet the conditions to support a species. While a tantalizing prospect, biogeographers were quick to note alarming inconsistencies between observed distributions of species and those expected from geographic niche projections (e.g., Pearson & Dawson 2003). Among a myriad of factors affecting the actual distributions of species, one nearly universal phenomenon is that the conditions species need to persist are discontinuous over real landscapes. To access these patches, species must disperse and cross regions of relatively low quality. Therefore, factors like dispersal and landscape heterogeneity are core constraints on a species' distribution (Pulliam 2000; Soberon & Peterson 2005; Soberón 2007), alongside factors like species interaction (e.g., competition, predation, parasitism) identified decades prior (Hutchinson 1978).

The role of dispersal in the geographic projection of species' niches is extremely relevant to individual- and population-level spatial inferences in movement ecology. While the individual niche concept is hotly contested, individuals within populations certainly vary physiologically and behaviorally. This variation should affect the distributions of individuals in space, potentially fueled by some latent difference in underlying needs or constraints. For example, male ungulates commonly occupy different environments than females, often with larger home ranges and occasionally more predation risk. This spatial segregation is hypothesized to arise from sex-based differences in reproductive constraints or resource requirements (reviewed in Main *et al.* 1996). Clearly, individual differences go far beyond sex,

with highly individual traits like personality likely factoring into space use (Spiegel *et al.* 2017). While quantifying individual variation in space use is an exciting endeavor, these differences should have real-world consequences for where individuals and their populations can persist, where they will go, and how they will get there – objectives central to our ability to inform conservation action.

Habitat selection modeling is a promising avenue to understand individual- and population-level space use and directly parallels niche models occurring at a species level. It is common to map the patterns of habitat selection of an individual, disregarding the spatial context of the habitat selection. For species niche modeling, an understanding of dispersal and heterogeneity was necessary to reconcile measured niches with the geographic distributions of species. Just as dispersal and landscape heterogeneity limit the ability of species to occupy all suitable conditions, movement traits, and landscape heterogeneity may limit the ability of individuals to occupy all suitable conditions in a landscape. Consequently, the underlying environmental preferences of an individual should not equal its real distribution geographically under movement constraints in heterogeneous landscapes (Fig. 1).

**Methods:**

*Landscapes as transition matrices*

We first considered a landscape, $G$, made up of $n$ cells along with environmental data (Fig. 2A). We treated the occupancy of these cells as states, between which individuals might transition according to rates described by a stochastic transition matrix, $A$, of $n$ rows and columns. We defined each value in this matrix, $A_{i,j}$, as the probability of moving from one cell ($G_i$) to another ($G_j$) in the landscape. For an animal in $G_i$, the sum of all transition probabilities to all other cells in $G$ (including $G_i$) must sum to one (we isolate the probability of transitioning from $G_i$ to $G_i$ for clarity later):

$$P(G_i|G_i) + \sum_{k=1,k\neq i}^{n} P(G_{k=j}|G_i) = 1. \qquad \text{Eq. 1}$$

In this way, Matrix *A* contains the directional probabilities of movement between all cells in *G* based on the occupancy of any given cell $G_i$, parametrized by later movement modeling, creating a row-stochastic matrix. The inverse of this matrix, $A^{-1}$, defines a column-stochastic matrix with a first principal eigenvector (associated with the unit eigenvalue in a stochastic matrix) defining the probabilities of the stable-state distribution across *G*.

*SSFs can parametrize transition matrices*

To gather these transition probabilities, we used predicted relative risk informed by fitted SSFs, a quantity derived by comparing baseline conditions (at $G_i$) to a set of available conditions (*G*) and often described as a relative strength of selection (Avgar et al. 2017). Relative risks are not equivalent to probabilities but can be kernelized when the number of available choices are finite. To do this, we divide Eq. 1 by the probability of moving from $G_i$ to $G_i$:

$$\frac{P(G_i|G_i)}{P(G_i|G_i)} + \sum_{k=1,k\neq i}^{n} \frac{P(G_{k=j}|G_i)}{P(G_i|G_i)} = \frac{1}{P(G_i|G_i)}. \qquad \text{Eq. 2A}$$

Therefore, Eq. 2A shows that the sum of relative risks for moving to all cells in *G* is equal to the inverse probability of remaining in the current cell. Or, simplified, the inverse sum of all relative selection strengths is equal to the probability of remaining in $G_i$ (Eq. 2B):

$$\frac{1}{1 + \sum_{k=1,k\neq i}^{n} \frac{P(G_{k=j}|G_i)}{P(G_i|G_i)}} = P(G_i|G_i). \qquad \text{Eq. 2B}$$

More importantly, Eq. 2B now describes how to link a limited set of relative selection strengths to the probabilities of remaining in place. Therefore, we could solve for the probabilities of moving to all comparison cells by multiplying the relative selection strength returned by the SSF predictions by the probability of staying in place.

*Extension to Integrated Step Selection Functions*

SSFs can also explicitly incorporate behavioral elements of the movement model independently or (more commonly) interactively with the habitat selection model. For example, integrated SSFs (iSSFs) frequently include a term to describe step distances, resulting in a distributional description of the likely movement distances around a focal position (Avgar *et al.* 2016). Functionally, this forces the probability of selection for very long-distance movements to zero even when habitat conditions are otherwise suitable. In this way, iSSF models can define local neighborhoods for animal movement. By incorporating distance into SSF models, we can confidently compare $G_i$ to cells in $G$ within a biologically-informed limit on distance (herein the 95$^{th}$ percentile of observed step distances, Fig. 2G). This substantially reduces the information burden carried by the transition matrix $A$ by treating it as a sparse matrix containing only transitions to cells within a reasonable biological limit of movement (Fig. 2H). This dramatically increases the speed of computing stable-state distributions (Fig. 2K), though users should explore the tradeoff between the question at hand and computational demand.

We tested the performance of our method in describing animal distributions using one simulated dataset and three real datasets. We compared the predictive performance of these mechanistic models to models lacking individual behavioral mechanisms: RSFs and SSFs. While SSF predictions capture patterns of selection, movement is ignored by assuming that local patterns of selection rules can be generalized across entire landscapes (Fig. 2F). Similarly, RSF predictions can capture patterns of occurrence; however, patterns are only observable over narrow spatial extents and do not capture patterns of movement in predictions to landscapes. SSFs and RSFs therefore are juxtaposed to the more mechanistic SSD approach.

*Simulation*

We defined a simulated animal mover based on an SSF framework (Potts & Borger 2023). We defined "true" environmental suitability by combining four rasters of random, spatially-autocorrelated variables using four random adjustment parameters sampled from a uniform distribution (Fig. 3A, SI Fig. 1, SI Fig. 2). The movement model also included negative selection for longer step lengths and positive selection for similar turn angles across steps. We simulated 10,000 initial steps (in-sample) and then a set of 500,000 steps with random restarts every 1,000 locations (out-of-sample), with initial locations sampled based on the environmental suitability surface. The dataset of 500,000 locations generated a map of spatial intensity of use (Fig. 3B). We converted the dataset of 10,000 steps into a movement track, jittered locations very slightly to create a continuous distribution of step lengths, and sampled 40 random steps per observed step using the "amt" package (Fig. 3C, SI Fig. 3) (Signer *et al.* 2019). We converted the observed gamma distribution of step lengths into a wider gamma distribution with a larger mean; specifically, we increased the shape and decreased the scale of the gamma distribution by a factor of two. This increased random sampling of step distances that were both very close and very far based on the animal mover (SI Fig. 4). Accordingly, we fit an SSF treating the step-terminus values of the four input rasters as predictors in the step selection function, and we included a polynomial for step length as a predictor using the "amt" package (SI Table 2) (Signer *et al.* 2019). For simplicity, in all analyses below, we did not model turn angle in all SSF models. We used the fitted SSF model to parametrize $A$ and generate a vector of stable-state distributions as described above (SI Fig. 5, SI Fig. 6).

We used "amt" package to fit an RSF using package defaults and the four input rasters as covariates (SI Table 2) (Signer *et al.* 2019). We also used the mean step length to predict the selection probability for all cells in the landscape using a random cell in the landscape as a baseline from the fitted SSF above. We then compared the error in predictions in two ways. First, we calculated the root-mean-squared error (RMSE) between the input environmental suitability and the simulated out-of-sample intensity of use, the SSD probability, the RSF kernelized probability of use, and the SSF kernelized

relative selection strength. Second, we calculated the RMSE between the simulated out-of-sample intensity of use and the SSD, RSF, and SSF predictions.

*Case Study #1 - Intra-individual prediction*

To test predictive accuracy within individuals, we used a red deer dataset for a single individual in the "amt" package to generate SSD, RSF, and SSF predictions across a raster of distance to forest (calculated from a binary forest raster provided in the "amt" package, aggregated to a resolution of 100m by 100m). We separated the first (chronologically) 75% of observed data as an in-sample training dataset and retained the latter 25% of observed data as an out-of-sample testing dataset. As above, we sampled random steps by widening the observed gamma distribution of step lengths by a factor of two with 40 random points per observed step. We included distance to forest as a covariate in models along with a polynomial of step length in SSF models (SI Table 2, SI Fig. 7). We generated the stable-state distribution as above using a neighborhood defined by the 95$^{th}$ percentile of observed step lengths. We then generated the RSF and SSF probabilities as above and aggregated each of the three predicted surfaces into centiles (100 bins). We tallied the proportion of out-of-sample testing locations that fell in each centile of the predicted surfaces, and calculated Spearman's rank correlation coefficient, rho, for the relationship between stable-state distribution, RSF, and SSF predictions vs. real use intensity. This is a common model check for RSF predictions (Morris *et al.* 2016), and it can be interpreted as the stronger the monotonic relationship between the percentiles of predicted use and use intensity, the better a given prediction captures the biological intensity of use (even if not linear). To ensure that comparisons were not sensitive to the underlying out-of-sample datasets, we bootstrapped the observed set of GPS points (sampling with replacement) 100 times, generating 100 estimates of Spearman's rho per predicted surface.

*Case Study #2 - Inter-individual predictions*

To demonstrate the predictive utility of SSD across individuals, we used a previously published roe deer dataset for five individuals in Northern Italy (Cagnacci *et al.* 2011) to assemble the same SSD, SSF, and RSF predictions. We resampled tracks to 4 hours (the median interval in GPS transmissions), and as above, we isolated the first 75% of individual movements as in-sample, and the next 25% as out-of-sample testing datasets. Because the dataset contained migratory movements, we decided to model only summer movements of deer (defined as Julian day 120-290, SI Fig. 11). We included distance to forest (calculated from 2006 CORINE landcover classes of conifer, mixed, and broadleaf forests) and slope (calculated from STRM elevation models, aggregated at a scale of 129m by 185.2m) as covariates in our models and predictions (SI Fig. 12) with a polynomial of step length, while the RSF only included the square-root of distance to forest and slope (SI Table 2). We performed the same out-of-sample evaluations as above with 100 bootstrapped Spearman rank correlations between surface centiles and observed use intensity in 25% out-of-sample for individuals. Additionally, we used individual-level models (fit using 75% of individual data) to predict the distributions of all other individuals' spatial use intensity in the dataset and assessed the quality of using the same bootstrapping method (referred to as "between individual" predictions).

*Case Study #3 - Prediction to novel contexts*

Lastly, to explore the predictive capacity of the SSD method to novel landscapes, we used a dataset of eight fisher from Northern New York, USA (SI Fig. 16) (LaPoint *et al.* 2013) to generate stable-state distribution, RSF, and SSF probabilities considering a landscape with square-root distance to forests or wetlands and square-root human population density (Falcone 2016)(SI Table 1). Seven fisher inhabited a similar environmental extent with a mix of forested and human landscapes (approximately 10km by 15km, SI Fig. 17), whereas one fisher occupied a completely different landscape approximately 30km

away from the others and with far less human infrastructure and population density (SI Fig. 22). We regarded the seven local fisher as focal individuals for modeling, and we retained the non-local fisher as a unique opportunity to evaluate individual-level predictive performance among environmental contexts. We calculated distance to forest using the National Landcover Database from 2006 to identify land classes which were wetland and deciduous, mixed, and evergreen forest. We dissolved the resulting raster to generate a distance to forest or wetland using QGIS (v. 3.22) and GDAL proximity analysis for computational efficiency before aggregating spatially to 91.2m by 123.2m resolution. Because individuals in the fisher dataset had different GPS fix-rates, we resampled tracks to the median fix rate of each individual (range: 2-15 minutes). We performed the same out-of-sample evaluations for within individual (75/25 split) and between individual comparisons (one individual predicting the distribution of all others) for all seven focal fishers using RSF, SSF, and SSD surfaces. We then performed a novel comparison of each of the seven focal fisher movement models to predict the spatial distribution of the non-local fisher in the novel environmental context, again using 100 bootstraps of the test dataset to assess sensitivity of surface predictions.

The code used to generate this paper is available online (https://github.com/will-rogers/StableStateSSF.git) and is provided as compiled document in an online supplement to the main text. All code will be archived on Zenodo upon acceptance.

**Results:**

*Simulation*

We found that the kernelized surface of the relative strength of selection was almost perfectly linearly related to the input environmental suitability surface (Pearson's $r$ = 0.997), a pattern supported by the fact that the SSF surface most closely matched (by an order of magnitude) the input spatial patterns of environmental suitability (RMSE = $7.57 \times 10^{-06}$) as compared to SSD (RMSE = $9.57 \times 10^{-05}$, $r$ =

0.934) and RSF surfaces (RMSE = $6.33 \times 10^{-05}$, $r$ = 0.931, Fig. 3G). Observed patterns of out-of-sample space use intensity had a relatively large error in predictions of input environmental suitability (RMSE $=7.84 \times 10^{-05}$, $r$ = 0.950). These patterns of error were also non-linear (Fig. 3G), with the SSD surface and the simulated intensity of use strongly departing from a 1:1 relationship. The SSD predictions (RMSE = $4.15 \times 10^{-05}$, Fig. 3H) were almost twice as accurate as the SSF surface (RMSE = $8.13 \times 10^{-05}$) in predicting the true simulated distribution of use intensity. The RSF surface lagged behind the best predicting model (RMSE = $8.28 \times 10^{-05}$). We found that the SSF predictions strongly departed from a 1:1 relationship, while the SSD predictions were now linear with the simulated intensity of use surface (Fig. 3H).

*Case Study #1 - Intra-individual prediction*

As in the simulation, the general patterns between surfaces appeared similar, but there was far greater variation in spatial intensity of predicted use in the SSD surface (Figs. 4A-C, SI Figs. 8-10). We found that the SSD surface for the red deer case study was best correlated with the observed use of intensity for modeled data and out-of-sample predictions for the last 25% of the individual's movement (Fig. 5A-B). Using only a single predictor in the model (distance to forest), the SSF and RSF surfaces were similar after kernelizing with weaker correlations with modeled and the out-of-sample locations (Fig. 5A-B).

*Case Study #2 - Inter-individual predictions*

We compared the RSF, SSF, and SSD surfaces in their ability to predict the future distributions of individual roe deer as well as the distributions of other individuals in the dataset (Figs. 4D-F, SI Figs. 13-15). For the predictions on modeled and out-of-sample predictions on the next 25% of chronological observations, we found that two individuals were best predicted by SSF surfaces, while the majority (three of five) were best predicted by SSD surfaces (Fig. 5C-D). This pattern was even more stark when

comparing the ability for individuals to predict the distribution of others in the dataset. Only one individual was better predicted by the RSF surface, while four of five were best predicted by the SSD surface (Fig. 5E).

*Case Study #3 - Prediction to novel contexts*

We used the same frameworks as above to assess the ability for RSF, SSF, and SSD surfaces to predict modeled data, the next 25% of individual data, and other individuals in the dataset (SI Figs. 18-20). We also compared the ability for individual models to make predictions in a novel environmental context (SI Figs. 23-25). We found that, as in previous examples, most individual movements of modeled data and out-of-sample predictions within individuals were best predicted by SSD surfaces (six of seven) as compared to RSF (zero of seven) and SSF surfaces (one of seven, Fig. 5F-G). We also found that the SSD surface of all seven individuals was better able to predict both local (Fig. 5H) and non-local space use by other individuals in the dataset, as compared to the RSF and SSF surfaces (Fig. 5I). In fact, the RSF and SSF surfaces were uniformly *negatively* correlated with the non-focal individual's space use in the non-local context, while the SSD surface was positively correlated with observed space use in a novel environmental context (Fig. 5I).

**Discussion**

Predicting animal distribution at landscape scales on the basis of organism-environment interactions is an active area of research that informs both basic ecological inquiry as well as applied work informing conservation, management, and global policy. We highlight the simultaneous integration of environmental preference, movement capacity, and spatial arrangement of habitats as an important outstanding challenge for current frameworks. In simulations and analyses of three biological datasets, we found that the SSD approach was able to recover patterns in animal distribution that were not

321 captured by occurrence (RSF) or SSF frameworks. We found that occurrence-based methods, though
322 commonly used, can generate poor inferences about how animals are distributed in landscapes, despite
323 accurately capturing then nature of environment-organism relationships. Conversely, movement models
324 like SSFs have been proposed as a more mechanistic approach to understand these interactions, yet we
325 found that SSFs can indeed fail to capture patterns of organism distribution in landscapes when used
326 improperly (i.e., making unconditional or arbitrarily conditioned predictions across landscapes). While
327 SSFs can accurately define fine-scale selection paradigms (e.g., SSFs can nearly perfectly capture habitat
328 preference parameters, Fig. 3G), we found that the incorporation of movement mechanisms was
329 necessary to generate effective predictions of animal distribution (Fig. 3H). We propose that by
330 combining the fine-scale applicability of SSFs with ergodic frameworks, such as SSD, to capture scale-
331 dependent movement, we can generate high-resolution and accurate depictions of organismal
332 distributions in space (Fig. 1). In sum, these findings suggest that local mechanisms of organismal
333 interactions with their environment can be better generalized to novel spatial extents and unobserved
334 organisms than occurrence-based frameworks or frameworks that rely on the application of local
335 movement rules divorced of movement mechanisms.

336 Our approach builds upon previous frameworks and methods, but we introduce a new way of
337 using SSFs to parametrize transition matrices and incorporate distance-weighted movement
338 probabilities. For example, transition matrices are the currency underlying frameworks like *circuitscape*
339 (McRae *et al.* 2008), and are fundamental to more recent movement frameworks (Fletcher *et al.* 2019).
340 Similarly, it is recognized that SSF output can capitulate animal distributions through simulations from
341 SSF models, not from direct predictions of SSF models over an entire landscape (Potts & Borger 2023).
342 The advance presented here is that ergodic frameworks provide a convenient way to generate
343 simulations of SSF output assuming random starts in all possible landscape cells. Similarly, by using SSFs
344 to parametrize transition matrices and incorporate distance-weighted movement probabilities, animal

movement models can generate sparse transition matrices that are less assumptive than previous methods used to generate transition matrices (McRae *et al.* 2008; Fletcher *et al.* 2019), and are quantitatively more specific, directional, and biologically reasonable.

Mechanistic models can produce predictions of patterns that are possibly less affected by bias than inferences and predictions based on patterns. Animal movement and step-selection can be effective mechanistic models of animal-environment interactions, and we demonstrate how these mechanisms can generate strong predictions of emergent animal distributions in landscapes. Conceptually, we propose that SSF and movement modeling in general is one of the best available ways to use remotely-sensed behavior to understand organism-centric mechanisms - like organism-environment interactions - that empower predictions of emergent phenomena.

Movement ecology may serve as an effective link between mechanism-focused subfields of behavior and physiology with pattern-focused subfields like biogeography. Harnessing the mechanisms implicit in remotely-sensed information offers a tantalizing opportunity to merge within-organism processes with their manifestation at grander scales in space. Importantly, this framework also removes potentially fraught questions about how long to run SSF simulations and from where in landscapes to initiate simulations.

Still, there remain limitations to the framework proposed above that would be fruitful avenues for future work. One that may be immediately apparent to movement modelers is the lack of turn angle inclusion in our ergodic framework. This is principally a computational issue. The transition matrix specifies movement between states (cells) in the environment. If we were to also include how the turn angle of entrance into a cell as a state dependency, it is not clear how to incorporate such a continuous variable in state definitions. This is because incorporating turn angles (no matter how biologically relevant) breaks the Markovian character of SSFs and, in doing so, offers conceptual and computational complications. We suggest that SSFs be fit so that each step is only dependent on the current conditions

of the step in the above framework. We note that step distance is often heavily correlated with turn angle (it is less possible to move large distances in consecutive steps with a high turn angle), and step distance may effectively capture much of the turn angle signal. However, this is certainly an over-generalization and this phenomenon would benefit from a more quantitative investigation. We recognized that the landscape prediction from SSFs was an effective diagnostic to test the predictions from an SSF in comparison to our biological expectations of animal distribution. We argue that this framework may offer an important ability to assess SSF fit, a currently nebulous endeavor. There are also many extensions of stochastic matrices for time-dependent predictions, connectivity modeling, and use with environmentally variable survival (Fletcher 2019). Treating SSFs as ways to parametrize transition matrices offers exciting opportunities for extensions in movement ecology and applications of organismal processes to larger scales of predictions.

Acknowledgments:

We thank Johannes Signer, Francesca Cagnacci, and Scott LaPoint for their generosity in providing the red deer, roe deer, and fisher datasets (respectively) as open-access resources for the scientific community, and for their permission to use those datasets in our study. We also thank John Fieberg and members of the Jetz and Ezenwa labs for their generous feedback on early versions of the project.

## Figures

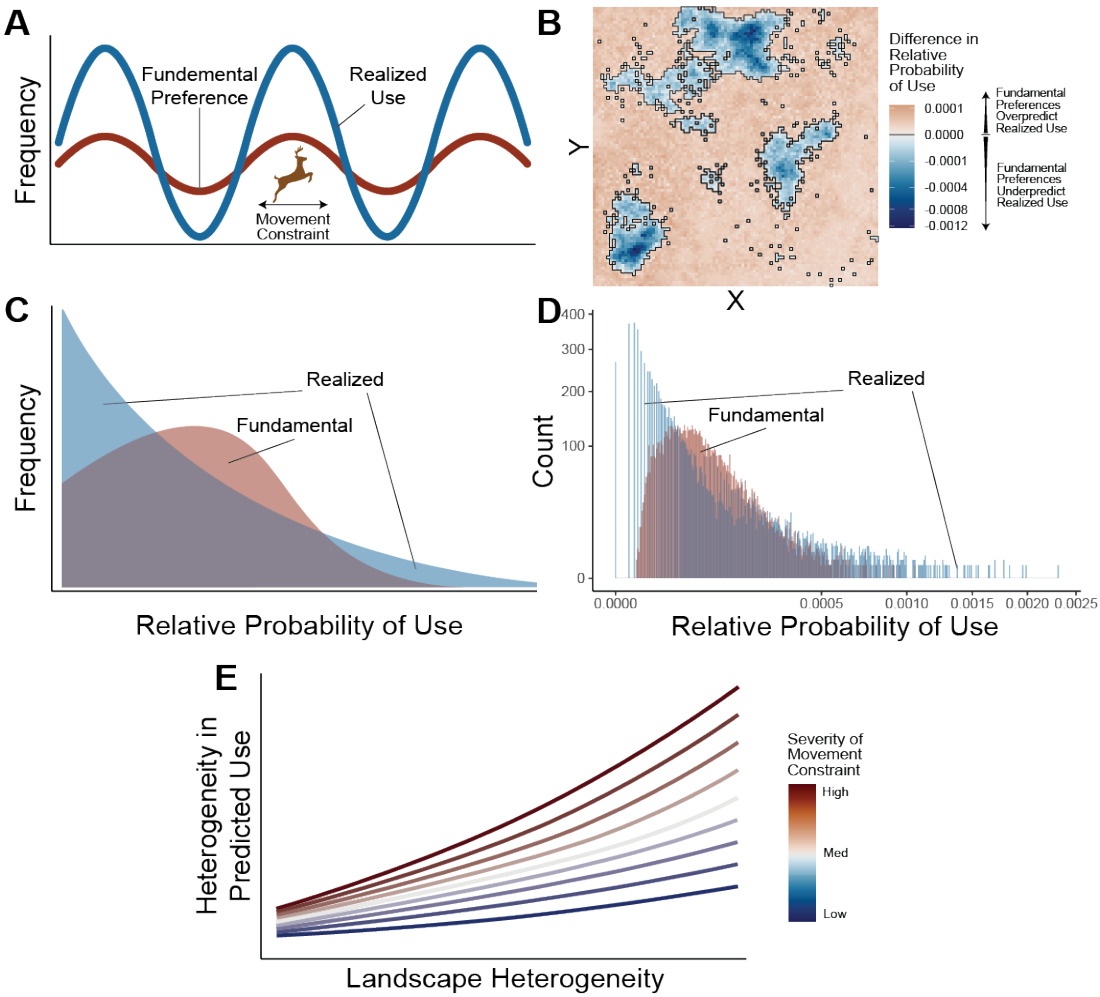
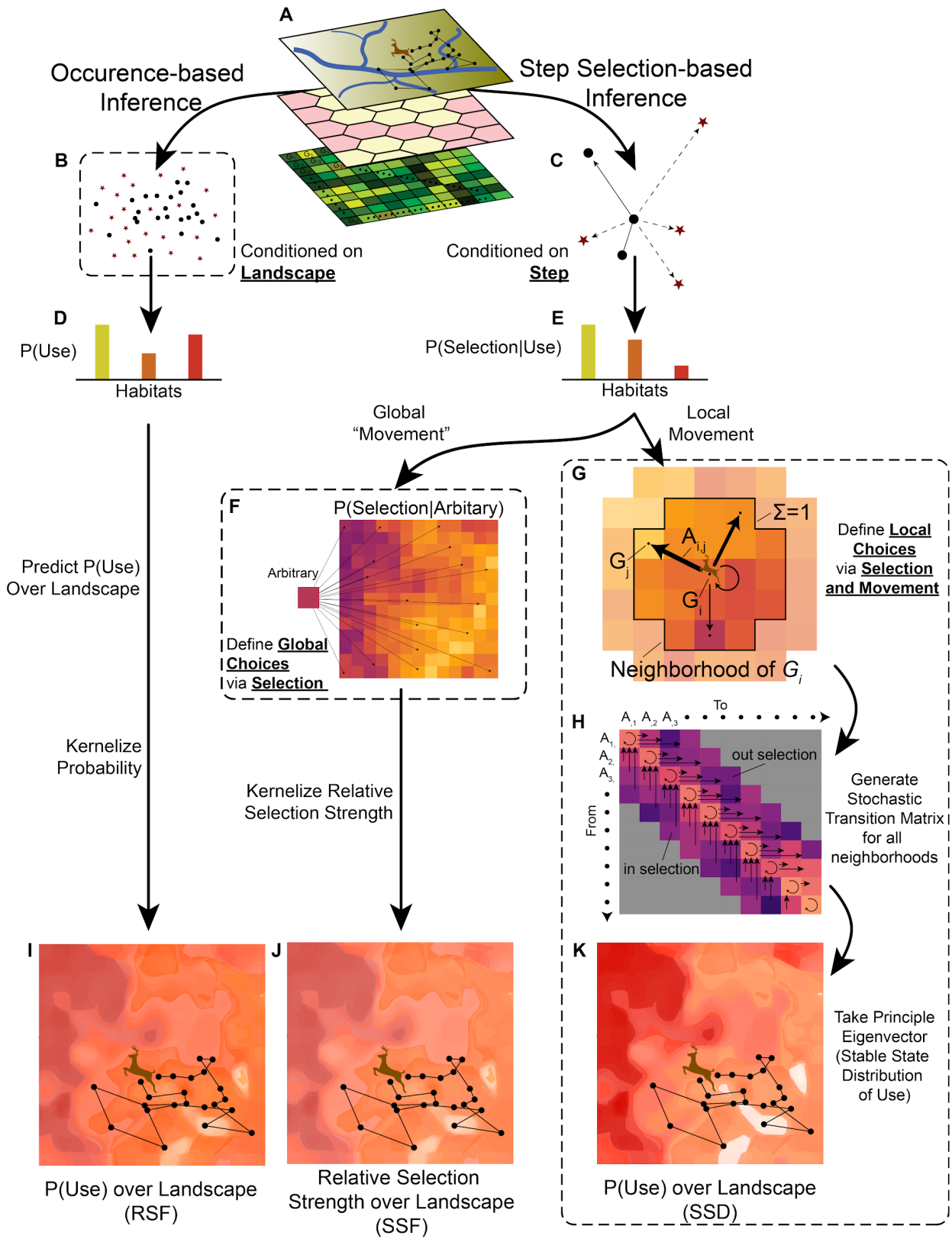



**Figure 1: Fundamental preferences do not equal realized use**

(**A**) The theoretical frequency of use based on the preference of a single-dimensional environment (red) vs. the realized frequency of use given some movement constraint (blue). (**B**) The visualization of the underlying preference surface used in simulations minus the realized frequency of use from simulations with selection against longer steps. (**C**) Theoretical and (**D**) empirical depictions of heterogeneity for the frequency of use of individual patches or locations within a landscape with (blue) and without movement constraints (red). (**E**) In effect, movement constraints should be predicted to amplify underlying patterns of landscape heterogeneity.


A
Occurence-based Inference
Step Selection-based Inference
B
Conditioned on Landscape
C
Conditioned on Step
D
P(Use)
Habitats
E
P(Selection|Use)
Habitats
Global "Movement"
Local Movement
F
P(Selection|Arbitary)
Arbitrary
Define Global Choices via Selection
G
Σ=1
Define Local Choices via Selection and Movement
Neighborhood of $G_i$
H
To
From
out selection
in selection
Generate Stochastic Transition Matrix for all neighborhoods
Predict P(Use) Over Landscape
Kernelize Probability
Kernelize Relative Selection Strength
I
P(Use) over Landscape (RSF)
J
Relative Selection Strength over Landscape (SSF)
K
Take Principle Eigenvector (Stable State Distribution of Use)
P(Use) over Landscape (SSD)

398

**Figure 2:** (**A**) Animal movements accompanied by environmental contexts, distilled into a landscape made of cells, *G*. (**B**) Animal locations can be distilled into use vs. available frameworks conditional on the landscape, suitable for resource selection analyses (RSFs). (**C**) Alternatively, tracking data can be converted into use vs. available frameworks conditional on steps, suitable for step selection analyses (SSFs). (**D**) RSFs predict the probability of use over a landscape, which can be kernelized to return (**I**) probabilities of use over *G*. (**E**) SSFs predict the relative probability between comparisons within a context, a distinct metric from the probability of use. (**F**) It is common to make an arbitrary comparison between some either observed or unobserved environmental contexts in *G* to all possible contexts in *G*. (**J**) These spatial relative risks are then kernelized, representing something akin to RSF output, though mathematically inequivalent. (**G**) Harnessing the ability of SSFs to parametrize local movement patterns, we can algebraically solve the probability of moving within a set of local environmental contexts ($G_j$) by defining that the relative selection strength for a focal cell ($G_i$) compared to itself must equal 1. (**H**) Aggregating patterns of in- and out-selection over all neighborhoods of all $G_i$ resolves a sparse stochastic matrix for all reasonable movements. (**K**) Solving for the principal eigenvector of the stochastic matrix returns the stable state distribution - the probability of ending in some $G_i$ given all possible starts over the landscape, *G*.

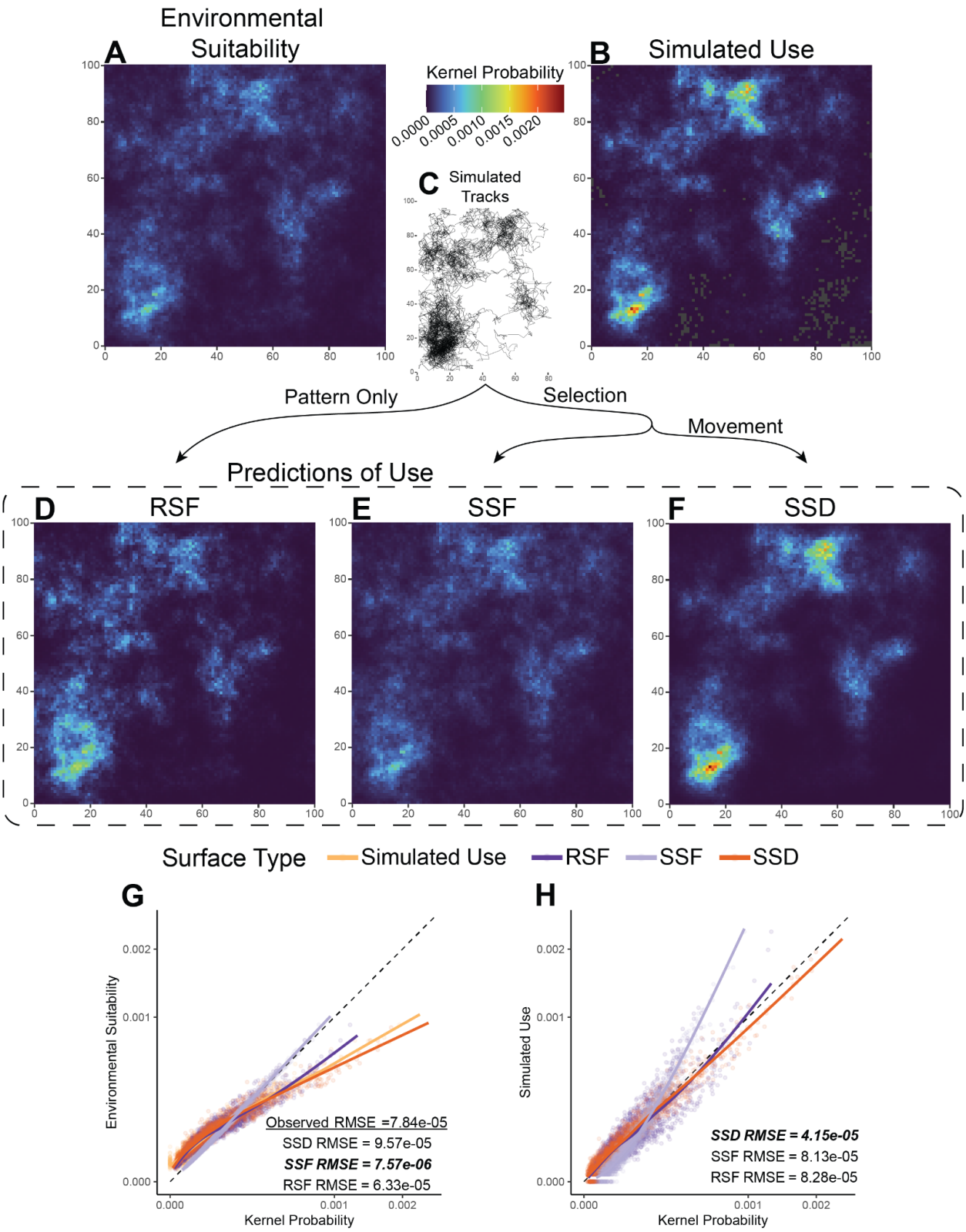



**Figure 3:** (**A**) Combining four random, spatially-autocorrelated rasters to generate a landscape of "suitable" environmental contexts ($G_i$). (**B**) Simulating 500 random animal tracks (with

constraints on step length and turn angle) over this landscape for 1,000 steps returns intensity of use with similar patterns as the input “suitability”, but with different intensities of use. (**C**) We simulated 10,000 steps for a single animal using the same input surface to fit a resource selection function (RSF) and a step-selection function (SSF). (**D**) Kernelized RSF probability of use and (**E**) kernelized SSF relative selection strength mapped over the landscape. (**F**) Using the SSF model, we estimated local movements, compiled those probabilities into a stochastic matrix, and solved the matrix for the principal eigenvector, resolving the stable state distribution of use over the landscape (SSD). (**G**) Comparison of the input “suitability” vs. observed use intensity from simulations, SSD, SSF, and RSF outputs, with a sampling of observed points (for plotting efficiency), a generalized additive model fit (GAM, with all data), and the root-mean square error (RMSE) between the input and four other surfaces. (**H**) Comparison of the observed use intensity from simulations vs. SSD, SSF, and RSF outputs, with a sampling of observed points (for plotting efficiency), GAM fit (with all data), and RMSE between the input and three other surfaces. We highlighted the minimum error surface in each scenario and underlined the error in the observed surface to highlight how divergent expectations are from simulated use. Spatial predictions are visualized untransformed and with the same color scale to highlight differences in intensity.

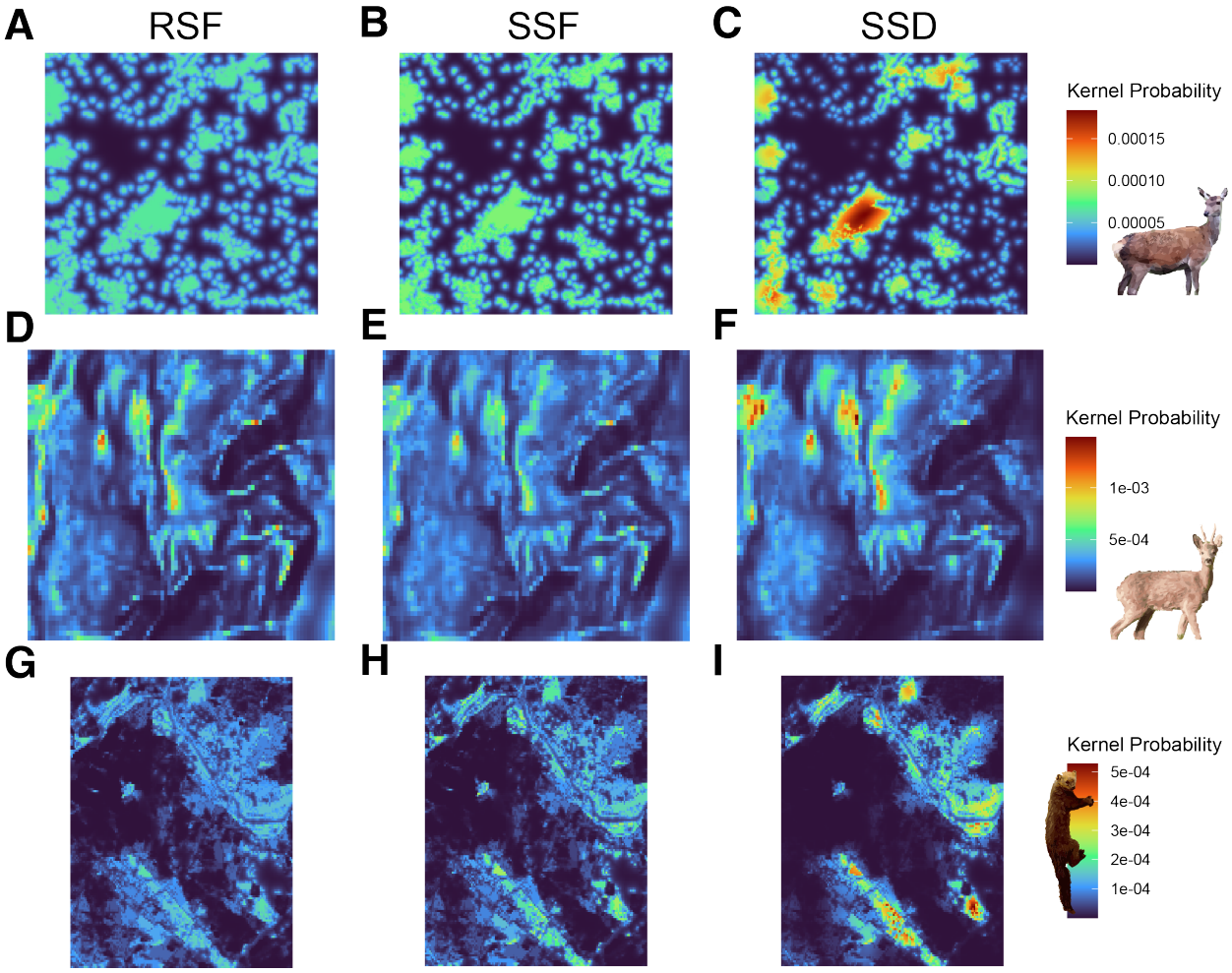



**Figure 4:** (**A-C**) Spatial predictions of "use" (as a kernel over the landscape) for red deer, (**D-F**) summer use of roe deer, (**G-I**) and fisher datasets based on (**A,D,G**) output from a resource selection function (RSF), (**B,E,H**) output from a step-selection function (SSF), and (**C,F,I**) the stable state distribution (SSD) from local movement patterns estimated from the SSF. Spatial predictions for each species are visualized untransformed and with the same color scale to highlight differences in intensity.

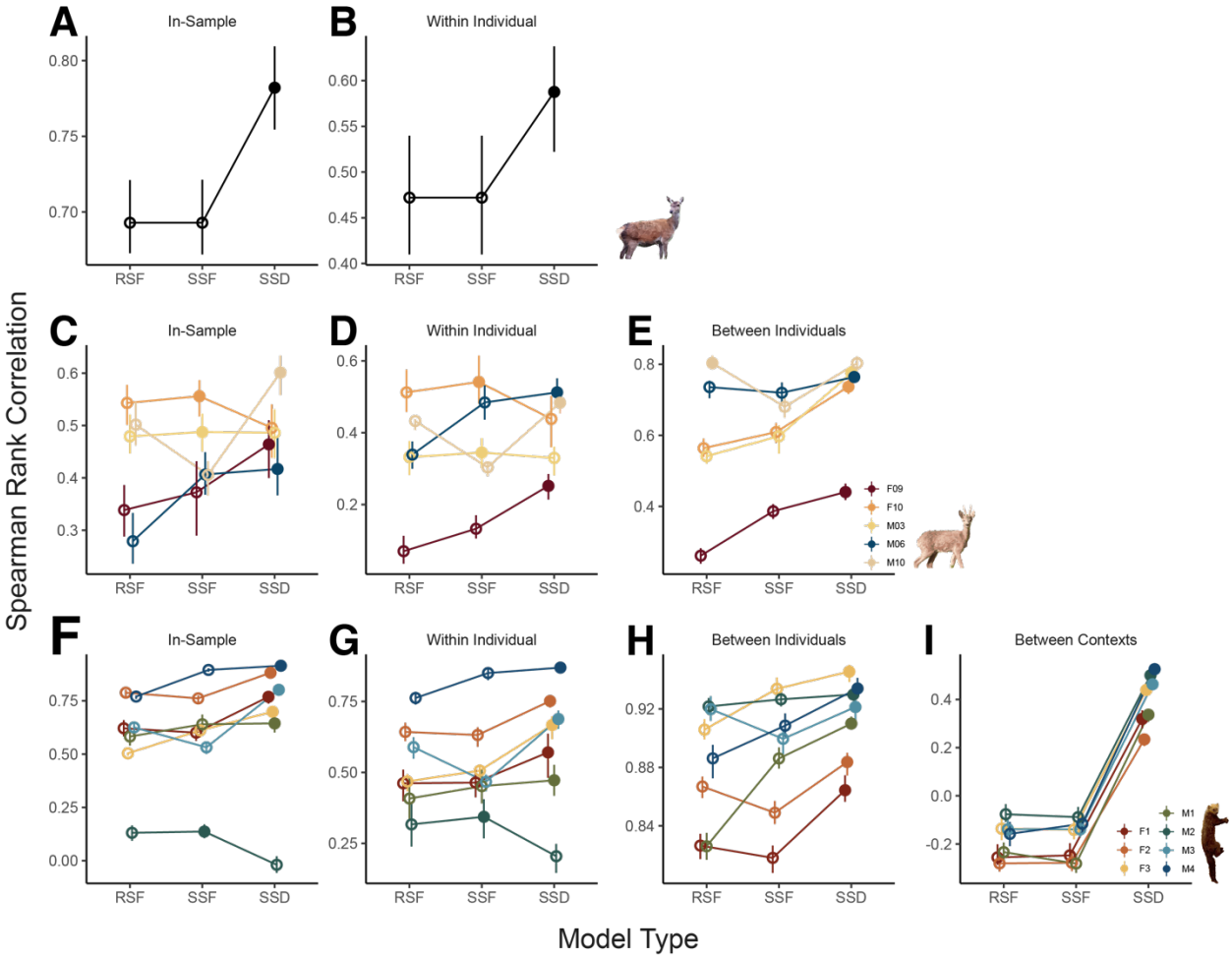



**Figure 5:** (**A,B**) Median (point) and 95% mass distributions (ranges) for 100 bootstraps of Spearman's rho coefficients between centiles of predicted use surfaces and proportion of true GPS fixes, where bootstraps are performed by resampling modeled and out-of-sample data with replacement for red deer (*Cervus elaphus*), (**C-E**) roe deer (*Capreolus capreolus*), and (**F-I**) fisher (*Pekania pennanti*) datasets. In four set of tests, we used the first (chronologically) 75% of observed locations for individuals to predict (**A,C,F**) the modeled data ("In-Sample"), (**B,D,G**) the next (chronologically) 25% of observed locations per individual ("Within Individual"), (**E,H**) all observed locations for the other individuals in a dataset within the same geographic context ("Between Individuals"), and (**I**) to predict all observed locations for an individual in a completely different environmental context than the individuals modeled ("Between Contexts").

References

Alston, J. M., C. H. Fleming, R. Kays, J. P. Streicher, C. T. Downs, T. Ramesh, B. Reineking, and J. M. Calabrese. 2023. Mitigating pseudoreplication and bias in resource selection functions with autocorrelation-informed weighting. Methods in Ecology and Evolution **14**:643-654.

Avgar, T., S. R. Lele, J. L. Keim, and M. S. Boyce. 2017. Relative Selection Strength: Quantifying effect size in habitat- and step-selection inference. Ecol Evol **7**:5322-5330.

Avgar, T., J. R. Potts, M. A. Lewis, and M. S. Boyce. 2016. Integrated step selection analysis: Bridging the gap between resource selection and animal movement. Methods in Ecology and Evolution **7**:619-630.

Boyce, M. S., and L. L. McDonald. 1999. Relating populations to habitats using resource selection functions. Trends in ecology & evolution **14**:268-272.

Boyce, M. S., P. R. Vernier, S. E. Nielsen, and F. K. A. Schmiegelow. 2002. Evaluating resource selection functions. Ecological Modelling **157**:281-300.

Bozinovic, F., D. A. Bastías, F. Boher, S. Clavijo-Baquet, S. A. Estay, and M. J. A. Jr. 2011. The Mean and Variance of Environmental Temperature Interact to Determine Physiological Tolerance and Fitness. Physiological and Biochemical Zoology **84**:543-552.

Buckley, L. B., S. A. Waaser, H. J. MacLean, and R. Fox. 2011. Does including physiology improve species distribution model predictions of responses to recent climate change? Ecology **92**:2214-2221.

Calabrese, J. M., G. Certain, C. Kraan, and C. F. Dormann. 2014. Stacking species distribution models and adjusting bias by linking them to macroecological models. Global Ecology and Biogeography **23**:99-112.

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

Creel, S., G. Spong, M. Becker, C. Simukonda, A. Norman, B. Schiffthaler, and C. Chifunte. 2019. Carnivores, competition and genetic connectivity in the Anthropocene. Pages 1-8 Scientific reports.

Cushman, S. A., and J. S. Lewis. 2010. Movement behavior explains genetic differentiation in American black bears. Landscape Ecology **25**:1613-1625.

Dingemanse, N. J., A. J. N. Kazem, D. Réale, and J. Wright. 2010. Behavioural reaction norms: animal personality meets individual plasticity. Trends in ecology & evolution **25**:81-89.

Evans, T. G., S. E. Diamond, and M. W. Kelly. 2015. Mechanistic species distribution modelling as a link between physiology and conservation. Conservation Physiology **3**.

Fieberg, J., J. Signer, B. Smith, and T. Avgar. 2021. A 'How to' guide for interpreting parameters in habitat-selection analyses. J Anim Ecol **90**:1027-1043.

Fletcher Jr, R. J., J. A. Sefair, N. Kortessis, R. Jaffe, R. D. Holt, E. P. Robertson, S. I. Duncan, A. J. Marx, and J. D. Austin. 2022. Extending isolation by resistance to predict genetic connectivity. Methods in Ecology and Evolution **13**:2463-2477.

Grimm, V., and S. F. Railsback. 2005. Princeton University Press, Princeton.

JOHNSON, C. J., S. E. NIELSEN, E. H. MERRILL, T. L. McDONALD, and M. S. BOYCE. 2006. Resource Selection Functions Based on Use-Availability Data: Theoretical Motivation and Evaluation Methods. The Journal of Wildlife Management **70**:347-357.

Kareiva, P., A. Mullen, S. R. Southwood, M. P. Hassell, and R. M. May. 1990. Population dynamics in spatially complex environments: theory and data. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **330**:175-190.

Kays, R., M. C. Crofoot, W. Jetz, and M. Wikelski. 2015. Terrestrial Animal Tracking as an Eye on Life and Planet. Science.

Kearney, M., and W. Porter. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. Ecology Letters **12**:334-350.

Keating, K. A., and S. Cherry. 2004. USE AND INTERPRETATION OF LOGISTIC REGRESSION IN HABITAT-SELECTION STUDIES. The Journal of Wildlife Management **68**:774-789.

Mackey, B. G., and D. B. Lindenmayer. 2001. Towards a hierarchical framework for modelling the spatial distribution of animals. Journal of Biogeography **28**:1147-1166.

Madin, J. S., M. O. Hoogenboom, and S. R. Connolly. 2012. Integrating physiological and biomechanical drivers of population growth over environmental gradients on coral reefs. Journal of Experimental Biology **215**:968-976.

Manly, B., L. McDonald, D. L. Thomas, T. L. McDonald, and W. P. Erickson. 2007. Resource selection by animals: statistical design and analysis for field studies. Springer Science & Business Media.

McRae, B. H., B. G. Dickson, T. H. Keitt, and V. B. Shah. 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. Ecology **89**:2712-2724.

Menge, B. A., and A. M. Olson. 1990. Role of scale and environmental factors in regulation of community structure. Trends in ecology & evolution **5**:52-57.

Merow, C., M. J. Smith, T. C. Edwards Jr, A. Guisan, S. M. McMahon, S. Normand, W. Thuiller, R. O. Wüest, N. E. Zimmermann, and J. Elith. 2014. What do we gain from simplicity versus complexity in species distribution models? Ecography **37**:1267-1281.

Moorcroft, P. R., and A. Barnett. 2008. Mechanistic home range models and resource selection analysis: A reconciliation and unification. Ecology **89**:1112-1119.

Morris, L. R., K. M. Proffitt, and J. K. Blackburn. 2016. Mapping Resource Selection Functions in Wildlife Studies: Concerns and Recommendations. Appl Geogr **76**:173-183.

Mueller, T., K. A. Olson, G. Dressler, P. Leimgruber, T. K. Fuller, C. Nicolson, A. J. Novaro, M. J. Bolgeri, D. Wattles, S. DeStefano, J. M. Calabrese, and W. F. Fagan. 2011. How landscape dynamics link

individual- to population-level movement patterns: a multispecies comparison of ungulate relocation data. Global Ecology and Biogeography **20**:683-694.

Newbold, T., T. Reader, A. El-Gabbas, W. Berg, W. M. Shohdi, S. Zalat, S. B. El Din, and F. Gilbert. 2010. Testing the accuracy of species distribution models using species records from a new field survey. Oikos **119**:1326-1334.

Northrup, J. M., E. Vander Wal, M. Bonar, J. Fieberg, M. P. Laforge, M. Leclerc, C. M. Prokopenko, and B. D. Gerber. 2022. Conceptual and methodological advances in habitat-selection modeling: guidelines for ecology and evolution. Ecological Applications **32**.

Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications **19**:181-197.

Post, D. M., and E. P. Palkovacs. 2009. Eco-evolutionary feedbacks in community and ecosystem ecology: interactions between the ecological theatre and the evolutionary play. Philosophical Transactions of the Royal Society B: Biological Sciences **364**:1629-1640.

Potts, J. R., and L. Borger. 2023. How to scale up from animal movement decisions to spatiotemporal patterns: An approach via step selection. J Anim Ecol **92**:16-29.

Powers, R. P., and W. Jetz. 2019. Global habitat loss and extinction risk of terrestrial vertebrates under future land-use-change scenarios. Nature Climate Change **9**:323-329.

Roughgarden, J. 1975. A Simple Model for Population Dynamics in Stochastic Environments. The American Naturalist **109**:713-736.

Sastre, P., and J. M. Lobo. 2009. Taxonomist survey biases and the unveiling of biodiversity patterns. Biological Conservation **142**:462-467.

Scharf, H. R., M. B. Hooten, D. S. Johnson, and J. W. Durban. 2018. Process convolution approaches for modeling interacting trajectories. Environmetrics **29**.

Schick, R. S., S. R. Loarie, F. Colchero, B. D. Best, A. Boustany, D. A. Conde, P. N. Halpin, L. N. Joppa, C. M. McClellan, and J. S. Clark. 2008. Understanding movement data and movement processes: Current and emerging directions. Ecology Letters **11**:1338-1350.

Signer, J., J. Fieberg, and T. Avgar. 2019. Animal movement tools (amt): R package for managing tracking data and conducting habitat selection analyses. Ecology and evolution **9**:880-890.

Sih, A. 2013. Understanding variation in behavioural responses to human-induced rapid environmental change: a conceptual overview. Animal Behaviour **85**:1077-1088.

Spiegel, O., S. T. Leu, C. M. Bull, and A. Sih. 2017. What's your move? Movement as a link between personality and spatial dynamics in animal populations. Ecology Letters **20**:3-18.

Stillman, R. A., S. F. Railsback, J. Giske, U. Berger, and V. Grimm. 2014. Making Predictions in a Changing World: The Benefits of Individual-Based Ecology. BioScience **65**:140-150.

van der Vaart, E., A. S. A. Johnston, and R. M. Sibly. 2016. Predicting how many animals will be where: How to build, calibrate and evaluate individual-based models. Ecological Modelling **326**:113-123.

Van Moorter, B., C. M. Rolandsen, M. Basille, and J.-M. Gaillard. 2016. Movement is the glue connecting home ranges and habitat selection. Journal of Animal Ecology **85**:21-31.

Wisz, M. S., and A. Guisan. 2009. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC Ecology **9**:8.

Woodin, S. A., T. J. Hilbish, B. Helmuth, S. J. Jones, and D. S. Wethey. 2013. Climate change, species distribution models, and physiological performance metrics: predicting when biogeographic models are likely to fail. Ecology and evolution **3**:3334-3346.

1.

Avgar, T., Lele, S.R., Keim, J.L. & Boyce, M.S. (2017). Relative Selection Strength: Quantifying effect size in habitat- and step-selection inference. *Ecol Evol*, 7, 5322-5330.

2.

Avgar, T., Potts, J.R., Lewis, M.A. & Boyce, M.S. (2016). Integrated step selection analysis: Bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution*, 7, 619-630.

3.

Boyce, M.S. & McDonald, L.L. (1999). Relating populations to habitats using resource selection functions. *Trends in Ecology & Evolution*, 14, 268-272.

4.

Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K.A. (2002). Evaluating resource selection functions. *Ecological Modelling*, 157, 281-300.

5.

Bozinovic, F., Bastías, D.A., Boher, F., Clavijo-Baquet, S., Estay, S.A. & Jr., M.J.A. (2011). The Mean and Variance of Environmental Temperature Interact to Determine Physiological Tolerance and Fitness. *Physiological and Biochemical Zoology*, 84, 543-552.

6.

Buckley, L.B., Waaser, S.A., MacLean, H.J. & Fox, R. (2011). Does including physiology improve species distribution model predictions of responses to recent climate change? *Ecology*, 92, 2214-2221.

7.

Cagnacci, F., Focardi, S., Heurich, M., Stache, A., Hewison, A.J.M., Morellet, N. *et al.* (2011). Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors. *Oikos*, 120, 1790-1802.

8.

Colwell, R.K. & Rangel, T.F. (2009). Hutchinson's duality: The once and future niche. *Proceedings of the National Academy of Sciences*, 106, 19651-19658.

9.

Creel, S., Spong, G., Becker, M., Simukonda, C., Norman, A., Schiffthaler, B. & Chifunte, C. (2019). Carnivores, competition and genetic connectivity in the Anthropocene. *Scientific Reports*, 9, 1-8.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19, 181-197.

38.

Post, D.M. & Palkovacs, E.P. (2009). Eco-evolutionary feedbacks in community and ecosystem ecology: interactions between the ecological theatre and the evolutionary play. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1629-1640.

39.

Potts, J.R. & Borger, L. (2023). How to scale up from animal movement decisions to spatiotemporal patterns: An approach via step selection. *J Anim Ecol*, 92, 16-29.

40.

Powers, R.P. & Jetz, W. (2019). Global habitat loss and extinction risk of terrestrial vertebrates under future land-use-change scenarios. *Nature Climate Change*, 9, 323-329.

41.

Pulliam, H.R. (2000). On the relationship between niche and distribution. *Ecology Letters*, 3, 349-361.

42.

Roughgarden, J. (1975). A Simple Model for Population Dynamics in Stochastic Environments. *The American Naturalist*, 109, 713-736.

43.

Sastre, P. & Lobo, J.M. (2009). Taxonomist survey biases and the unveiling of biodiversity patterns. *Biological Conservation*, 142, 462-467.

44.

Schick, R.S., Loarie, S.R., Colchero, F., Best, B.D., Boustany, A., Conde, D.A. *et al.* (2008). Understanding movement data and movement processes: Current and emerging directions. *Ecology Letters*, 11, 1338-1350.

45.

Signer, J., Fieberg, J. & Avgar, T. (2019). Animal movement tools (amt): R package for managing tracking data and conducting habitat selection analyses. *Ecology and Evolution*, 9, 880-890.

46.

Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species

572