Appendix S1 – from step selection to utilisation distribution

Step selection models are mechanistic depictions of the movement behavior of individual animals in discrete time. Diffusion approximations allow shifting from this Lagrangian perspective to an Eulerian one. Partial differential equations (Moorcroft & Barnett 2008), or integral equations (Barnett & Moorcroft 2008; Potts *et al.* 2014b), can be used, for example, to approximate some step-selection processes, thus formally linking resource selection and steady-state utilisation distributions (see below). Alternatively, transport equations may provide powerful tools to derive macroscopic patterns from RW-based microscopic movement processes (Kareiva & Odell 1987; Hillen & Painter 2013).

For step-selection processes, the steady-state utilisation distribution is given by the steady state of the following master equation (Okubo & Levin 2001; Potts *et al.* 2014b):

$$UD_{t+\tau}(x') = \int_{\Omega} R(x'|x)\, UD_t(x) dx, \qquad \text{Eq. 1.1}$$

where $R(x'|x)$ is the redistribution kernel (given by kernel-generating functions such as Eq. 1 or Eq. 4 in the main text) over time interval τ, and $UD_t(x)$ is the utilisation distribution across the spatial domain Ω at time t. The steady-state UD is the limiting function $UD_{\infty}(x) = \lim_{t\to\infty} UD_t(x)$, which arises as the solution of the following integral equation

$$UD_{\infty}(x') = \int_{\Omega} R(x'|x)\, UD_{\infty}(x) dx. \qquad \text{Eq. 1.2}$$

Eq. 1.2 is usually impossible to solve exactly (but see (Barnett & Moorcroft 2008)). Therefore approximate and/or numerical methods are required. One approximate method involves constructing a partial differential equation (PDE) limit of Eq. 1.1, and performing steady-state analysis (Moorcroft *et al.* 2006; Moorcroft & Barnett 2008). However, such PDEs might not be solvable and/or the approximation might give qualitatively different results to the exact system (Potts & Lewis, in review). Another method involves solving Eq. 1.1 numerically through time until $UD_{t+\tau}(x)$ is sufficiently close to $UD_t(x)$ for all x (Potts *et al.* 2014b). This approach might prove to be particularly computationally demanding due to the computational time required in calculating the integral. To speed this up, one could use Monte-Carlo methods to solve Eq. 1.1, which is equivalent to performing stochastic simulations of the redistribution kernel (as was done in the current study; Appendix S5). After a sufficient burn-in period, the positions of the simulated animals approximately represent samples from the steady-state distribution $UD_{\infty}(x)$. Regardless of the methods of choice, understanding the relationship between mechanistic kernel-generating functions and resulting UD's is crucial if we wish to translate spatial animal behaviour to population distribution and redistribution patterns.

Appendix S2 – Inferring step-length distributions

Our aim here is to demonstrate the use of a conditional logistic regression analysis of step-length ($l > 0$) data to obtain maximum likelihood estimates of the parameter/s of one of the following step-length distributions: exponential (with rate λ), half-normal (with standard deviation σ), gamma (with shape k and scale q), or log-normal (with shape σ and scale μ). Taking the natural logarithm of each of these four probability density functions, we obtain:

$$\ln\big(f_1(l;\lambda)\big) = \ln(\lambda) - \lambda l, \qquad \text{Eq. 2.1}$$

$$\ln\big(f_2(l;\sigma)\big) = -\ln\left(\frac{\sqrt{2}}{\sigma\sqrt{\pi}}\right) - \frac{1}{2\sigma^2} l^2, \qquad \text{Eq. 2.2}$$

$$\ln\big(f_3(l;k,q)\big) = -\ln(\Gamma(k)q^{\kappa}) - \frac{1}{q} l + (k-1)\ln(l), \qquad \text{Eq. 2.3}$$

and

$$\ln(f_4(l;\sigma,\mu)) = \left(-\ln\big(\sigma\sqrt{2\pi}\big) - \frac{\mu^2}{2\sigma^2}\right) - \left(\frac{\mu}{2\sigma^2} - 1\right)\ln(l) - \frac{1}{2\sigma^2}\ln(l)^2. \qquad \text{Eq. 2.4}$$

Hence, for all four distributions, the probability of observing a step of length l can be expressed as an exponential function of some linear combination of l, $ln(l)$, l^2, and/or $ln(l)^2$.

Let us now assume we have obtained a set of T spatial positions, x_t, sampled at a unit temporal interval along an animal's path ($t = 1, 2, \ldots, T$), with the t'th step-lengths given by l_t ($= \|x_t - x_{t-1}\|$). For simplicity, we shall assume the animal is traversing a homogeneous landscape, that its movement behavior is temporally invariant, and that it lacks any velocity autocorrelation (i.e., the process is first-order Markovian). A step-selection analysis is based on matching each observed point along the path (each but the first), with a set of s control points, $x'_{t,i}$ ($i = 1, 2, \ldots, s$), where the distance between each point and the previous used location is the step-length, $l'_{t,i}$ ($= \left\|x'_{t,i} - x_{t-1}\right\|$). We shall start by sampling the set s with equal probability within some maximal distance from x_{t-1} (e.g., a distance corresponding to the maximum movement capacity of the focal species). Using matched case-control logistic regression we now can estimate the value/s of the parameter/s governing the assumed step-length distribution. Assuming an exponential step-length distribution (Eq. 2.1), the likelihood of the observed data is:

$$\prod_{t=2}^{T} \frac{\exp(\beta_1 \cdot l_t)}{\sum_{i=0}^{s} \exp(\beta_1 \cdot l'_{t,i})}, \qquad \text{Eq. 2.1.1}$$

where β_1 is the statistical estimator of λ. Similarly, we can formulate the likelihood assuming a half-normal distribution (Eq. 2.2):

$$\prod_{t=2}^{T} \frac{\exp\left(\beta_1 \cdot {l_t}^2\right)}{\sum_{i=0}^{s} \exp\left(\beta_1 \cdot {l'_{t,i}}^2\right)}, \qquad \text{Eq. 2.2.1}$$

where β_1 is the statistical estimator of σ ($\beta_1 = -1/2\sigma^2$); the likelihood assuming a gamma distribution (Eq. 2.3):

$$\prod_{t=2}^{T} \frac{\exp\left(\beta_1 \cdot l_t + \beta_2 \cdot \ln(l_t)\right)}{\sum_{i=0}^{s} \exp\left(\beta_1 \cdot l'_{t,i} + \beta_2 \cdot \ln(l'_{t,i})\right)}, \qquad \text{Eq. 2.3.1}$$

where β_1 is the statistical estimator of θ ($\beta_1 = -1/q$) and β_2 is the statistical estimator of k ($\beta_2 = k-1$); and the likelihood assuming a log-normal distribution (Eq. 2.4):

$$\prod_{t=2}^{T} \frac{\exp\left(\beta_1 \cdot \ln(l_t) + \beta_2 \cdot \ln(l_t)^2\right)}{\sum_{i=0}^{s} \exp\left(\beta_1 \cdot \ln(l'_{t,i}) + \beta_2 \cdot \ln(l'_{t,i})^2\right)}, \qquad \text{Eq. 2.4.1}$$

where β_2 is the statistical estimator of σ ($\beta_2 = -1/2\sigma^2$) and β_1 being the statistical estimator of μ ($\beta_1 = \beta_2\mu+1$).

The error in these estimators is inversely proportional to s. Increased efficiency (i.e., smaller error for the same s) can be obtained by sampling the control set under the assumed step-length distribution. To exemplify, assuming an exponential step-length distribution, we first estimate λ based on the observed steps: $\lambda_1 = \frac{T-1}{\sum_{t=2}^{T} l_t}$. We then sample $T-1$ sets of s control points, $x'_{t,i}$ ($i = 1, 2, \ldots, s$), so that the probability of obtaining a sample at some distance, $l'_{t,i}$, from the previous observed point, is given by the exponential PDF: $\lambda_1 \cdot \exp(-\lambda_1 \cdot l'_{t,i})$. Finally, we get an MLE for β_1 using Eq. 2.1.1. Note that, if indeed no other process is considered in the analysis, the expectancy of the resulting β_1 is 0. Otherwise, $\lambda = \lambda_1 + \beta_1$. In the main text ('A hypothetical example') and in Appendix S3 we demonstrate this point in the case where sampling is performed under a gamma distribution.

Appendix S3 – Deriving an iSSA likelihood function

Here we derive Eq. 3 in the main text. We start by deriving the probability density $f(x_t|x_{t-1},x_{t-2})$ of moving to a point x_t at time t, given that the previous two positions (at time-steps t-1 and t-2) were x_{t-1} and x_{t-2} respectively. This probability density is proportional to the product of three expressions, corresponding to propositions A, B and C in the main text (section "A hypothetical example"). Proposition A says that the animal is exponentially selecting for high values of $h(x)$, meaning that $f(x_t|x_{t-1},x_{t-2})$ is proportional to $\exp[\omega \cdot h(x_t)]$ for some $\omega > 0$. Proposition B says that $f(x_t|x_{t-1},x_{t-2})$ is proportional to a von Mises distribution with mean 0 and concentration parameter proportional to $\theta_1+\theta_2 \cdot y(x_{t-1})$. That is, $f(x_t|x_{t-1},x_{t-2})$ is proportional to $\exp([\theta_1+\theta_2 \cdot y(x_{t-1})] \cdot \cos[\alpha_{t-1}-\alpha_t])$, where α_{t-1} and α_t are the headings from x_{t-2} to x_{t-1}, and from x_{t-1} to x_t, respectively. Finally, proposition C says that $f(x_t|x_{t-1},x_{t-2})$ is proportional to a gamma distribution of the step length, l_t (the Euclidian distance from x_{t-1} to x_t). Furthermore, it says that the shape of the distribution depends on the time of day D_t, so that $f(x_t|x_{t-1},x_{t-2})$ is proportional to $\exp[\theta_3 \cdot l_t + \ln(l_t) \cdot (\theta_4+\theta_5 \cdot D_t)]$ (see Appendix S2). Taking the product of these three expressions and normalising, we find the following probability density

$$f(x_t|x_{t-1},x_{t-2}) = \frac{\exp[\omega \cdot h(x_t)+[\theta_1+\theta_2 \cdot y(x_{t-1})] \cdot \cos(\alpha_{t-1}-\alpha_t)+\theta_3 \cdot l_t+(\theta_4+\theta_5 \cdot D_t) \cdot \ln(l_t)]}{\int_\Omega \exp[\omega \cdot h(x')+[\theta_1+\theta_2 \cdot y(x_{t-1})] \cdot \cos(\alpha_{t-1}-\alpha_{t'})+\theta_3 \cdot l_t+(\theta_4+\theta_5 \cdot D_t) \cdot \ln(l_{t'})]dx'}. \quad \text{(Eq 3.1)}$$

Here, α_t' is the direction from x_{t-1} to x' (any location in the spatial domain), and l_t' is the distance between x_{t-1} and x'.

The integral in the denominator ensures that $f(x_t|x_{t-1},x_{t-2})$ integrates to 1, so is a probability distribution. This integral can be found numerically, but the calculation is typically very computationally intensive. Therefore, as is standard in SSA, and as explained in the Main Text, we sample s points from an "availability" distribution. In our case, this distribution is given by sampling a direction uniformly at random, and a distance from gamma distribution, $g(l|b_1,b_2)$ (given in the main text in Eq. 2), where b_1 and b_2 are estimated shape and scale parameters. As explained in Forester *et al.* (2009, Eqs. 4, 5 and surrounding text), this leads to the following discrete-choice approximation for the conditional probability of the animal being at x at time t, given its previous two locations and a set of s control points sampled under g:

$$\frac{\exp[\omega \cdot h(x_t)+[\theta_1+\theta_2 \cdot y(x_{t-1})] \cdot \cos(\alpha_{t-1}-\alpha_t)+\theta_3 \cdot l_t+(\theta_4+\theta_5 \cdot D_t) \cdot \ln(l_t)]/g(l_t|b_1,b_2)}{\sum_{i=0}^{s} \exp[\omega \cdot h(x'_{t,i})+[\theta_1+\theta_2 \cdot y(x_{t-1})] \cdot \cos(\alpha_{t-1}-\alpha'_{t,i})+\theta_3 \cdot l'_{t,i}+(\theta_4+\theta_5 \cdot D_t) \cdot \ln(l'_{t,i})]/g(l'_{t,i}|b_1,b_2)}, \quad \text{(Eq. 3.2)}$$

where $l'_{t,i}$ are the sampled step lengths, $\alpha'_{t,i}$ are the sampled directions, and $x'_{t,i}$ are the resulting positions found by moving a distance of $l'_{t,i}$ in the direction $\alpha'_{t,i}$.

Substituting g in Eq. 3.2 by its explicit form (Eq. 2 in the main text) yields the following expression for the conditional probability:

$$\frac{\exp\left[\omega\cdot h(x_t)+[\theta_1+\theta_2\cdot y(x_{t-1})]\cdot\cos(\alpha_{t-1}-\alpha_t)+(\theta_3+b_2^{-1})\cdot l_t+(\theta_4-b_1+\theta_5\cdot D_t)\cdot\ln(l_t)\right]}{\sum_{i=0}^{S}\exp\left[\omega\cdot h(x'_{t,i})+[\theta_1+\theta_2\cdot y(x_{t-1})]\cdot\cos(\alpha_{t-1}-\alpha'_{t,i})+(\theta_3+b_2^{-1})\cdot l'_{t,i}+(\theta_4-b_1+\theta_5\cdot D_t)\cdot\ln(l'_{t,i})\right]} \qquad \text{(Eq. 3.3)}$$

from whence Eq. 3 in the main text follows.

Appendix S4 – iSSA practical user guide

Here we provide tips and guidelines for conducting a fruitful integrated step selection analysis (iSSA). We refer the reader to Thurfjell *et al.* (2014) for a more general review of applications of SSA.

1. Collect animal positional data
    - To maximize the usefulness of the data in an iSSA, positional data should be collected at a constant fix rate (equal time steps).
    - Generally speaking, high fix rate (short time steps and hence short step-lengths) is expected to increase the reliability of the analysis. This is particularly true when aiming to capture continuous use of small spatial units, such as roads. Note however that fix rate should be adjusted to the typical displacement rate of the study species. Fix rate should be considered too high (and hence wasteful) if during a single step the animal is expected to travel less than the positional error (~ 20-30 m for GPS tags), or the spatial resolution of the focal habitat map (~ 20-250 m for most satellite derived maps).
    - Both habitat selection and movement behavior may depend on time of day and season. Sampling should attempt to capture as much temporal variability as possible during the time of the study. For example, fix schedule that is out of synch with time of day (e.g., every 5 hours) can help capture more (temporal variability) with less (fixes).
2. Tabulate observed (case) steps
    - Clean the data - even good GPS datasets contain erroneous positions. Exclude fixes taken before and shortly after tag deployment and after mortality/drop-off events. Plot the observed trajectories and visually look for potential positional errors. Run a code/script that scans the data for extreme values such as unreasonably long (or fast) steps or return-trips (relocations starting and ending at approximately the same point).
    - Because fixes are not always taken at their designated time, it is useful to define a reasonable temporal tolerance range for the step duration (e.g., 1 hr ± 10 min).
    - We recommend including the following fields in the 'used-step' table: individual ID, unique step ID, step start-point (time, easting and northing), step endpoint (time, easting and northing), step-duration, step-length, and step-heading (relative to the true north).
    - If velocity autocorrelation is to be included in the analysis, one must make sure to only calculate velocity deviations (e.g., turn-angles or step-length differences) between successive valid steps. Any step that does not have a valid step leading to it (e.g., because the previous fix is missing), cannot be characterized by valid velocity deviations and hence must be

excluded from the analysis (note that such a step should still be used to calculate velocity deviations for the proceeding step).

3. Sample lengths of available (control) steps
    - Available step lengths should be sampled based on one of the following probability density distributions (see Appendix S2): a uniform distribution within some maximal distance (e.g., the longest observed step), exponential, normal, gamma, and log-normal.
    - If there is no a priori reason to use a particular distribution, we would recommend using the gamma because it is flexible and includes the exponential as a special case.
    - Whatever the theoretical distribution of choice is, the observed step-lengths should be used to estimate its parameters (using either the method of moments or maximum likelihood). If there is strong a priori biological reason to think that these estimates should differ between distinct portions of the data (individual ID, sex, season, study area, etc.), more efficient model fitting can be gained by partitioning the data accordingly.
    - If the user is interested in determining which theoretical distribution best fits the data, we recommend using the uniform distribution to sample available step-lengths. The step-length, its natural logarithm, its square, and the square of its natural logarithm should then be included as predictors in a set of four competing models (see Appendix S2 for details), and AIC can be used to choose the best one.
4. Sample available (control) step headings
    - In the simplest case, available step headings are sampled from a uniform distribution between 0 and 2π.
    - If directional correlations or bias are evident in the data, increased efficiency may be gained by sampling available step headings from a von Mises distribution where the concentration parameter is estimated from the observed directional persistence/bias distribution.
    - If observed directional persistence/bias is correlated with step-length (e.g., the animal tends to turn less when making short steps), increased efficiency can be gained by accounting for the correlation structure when sampling available step headings.
5. Generate available (control) steps
    - Combine sampled step lengths and headings (with appropriate cross-correlation structure) to generate available steps starting at a used step start-point and ending in random endpoints.

- Identify each cluster, consisting of a single used step and its matched set of available steps, with a unique cluster ID. Code all used steps as '1' (case) and all available steps as '0' (control).

6. Attach step attributes
    - Characterize each step (cases and controls) with the following: step-length, step turn-angle and angular deviation from a preferred direction (if relevant), temporal covariates (e.g., time of day, season), and spatial covariates (e.g., elevation, NDVI, cover, temperature, etc.).
    - Covariates could be matched (in space and/or time) to the step's start-point, to its endpoint, and/or based on some interpolation between the two (e.g., average along the step, within an ellipse bounded by the start and end positions, or along a Brownian bridge). Note that, if covariates are purely temporal or are measured at the step start-point, their value would be identical for all steps belonging to the same cluster and their independent effects are thus statistically unidentifiable. The effects of such covariates are identifiable when interacting with other variables (e.g., an interaction between step-length and season).
7. Fit a conditional logistic regression
    - As long as sample sizes are sufficient (see Appendix S6), we recommend fitting iSSA for each individual independently (rather than using a mixed effects approach). This allows for a straightforward and unbiased evaluation of both inter- and intra-individual variability (Fieberg *et al*. 2010). Population level inference can then be gained by averaging individual model fits.
    - Function *clogit* in R is often used to fit conditional logistic regression. Note that this function (as many other conditional logistic regressions routines) rely on a Cox proportional hazard model to obtain MLEs and hence its output is a *coxph* output.
8. Adjust movement coefficients
    - Once step-length and/or turn-angle coefficient estimates are obtained, those must be combined with the tentative parameter estimates used for sampling available steps (see Appendix 1 and the main text for details). If steps were sampled from a uniform distribution, no adjustments are needed.
9. Simulate space-use
    - Once the integrated step selection function has been fully parametrized, it can be used to simulate space-use across any discrete map of its spatial covariates.

- This requires a simulation model that iteratively calculates the redistribution kernel at each simulated position and then samples from this kernel to select the next position. Note that the parametrized step-length distribution is one-dimensional, it describes the probability density of any particular displacement over the prescribed time step. When calculating the full two-dimensional redistribution kernel, rather than just drawing from such kernel, the basal probability density for any distance, r, from the center of the kernel is proportional to $2\pi r$. One must thus correct for dimensionality by dividing by $2\pi r$. This requires care when the step-length approaches 0, possibly by setting a minimal value for r (e.g., $1/2\pi$), so as to prevent the kernel from collapsing onto a Dirac delta function.
- A Monte Carlo approximation of the utilization distribution can be gained by simulating long and/or multiple trajectories (starting from the same or different positions) and then normalizing space-use across the map (see also Appendix S5).
- Such simulations may be used for cross validation (e.g., using ROC AUC to quantify the predictive power of the model over a validation dataset), as well as for predicting the ecological consequences of habitat loss, fragmentation, and other environmental changes that might affect animal space-use.

Appendix S5 – simulation experiments

Simulating movement and fitting statistical models: Fine-scale space-use dynamics were stochastically simulated, in discrete time and space, using a 'stepping-stone' movement process. The simulation operates within a discretized spatial domain consisting of 11,600 hexagonal cells (corresponding to 10,000 squared spatial units, so that the distance between adjacent cells is one spatial unit) and wrapped around a torus to eliminate any edge effects. Each cell in the domain, x, is characterised by some values of a continuous, normally distributed habitat quality, $h(x)$. Spatial autocorrelation in habitat quality was generated by first assigning a pseudorandom number [~ $U(0,1)$] to each spatial cell in the domain and then locally averaging those within a fixed autocorrelation range, ρ. The ranks of the resulting 'smoothed' values were then used to assign $h(x)$ values drawn independently from a standard normal distribution. Five different ρ values were used (0, 1, 5, 10, and 50 spatial units) to generate five levels of spatial autocorrelation (see Fig 5.1 for an illustration, and Appendix S9 for a discussion on the interpretation of ρ). For each spatial autocorrelation level, 1,000 different landscape realisations were generated, differing from each other by the spatial configuration of h (but not by its frequency distribution or spatial autocorrelation).

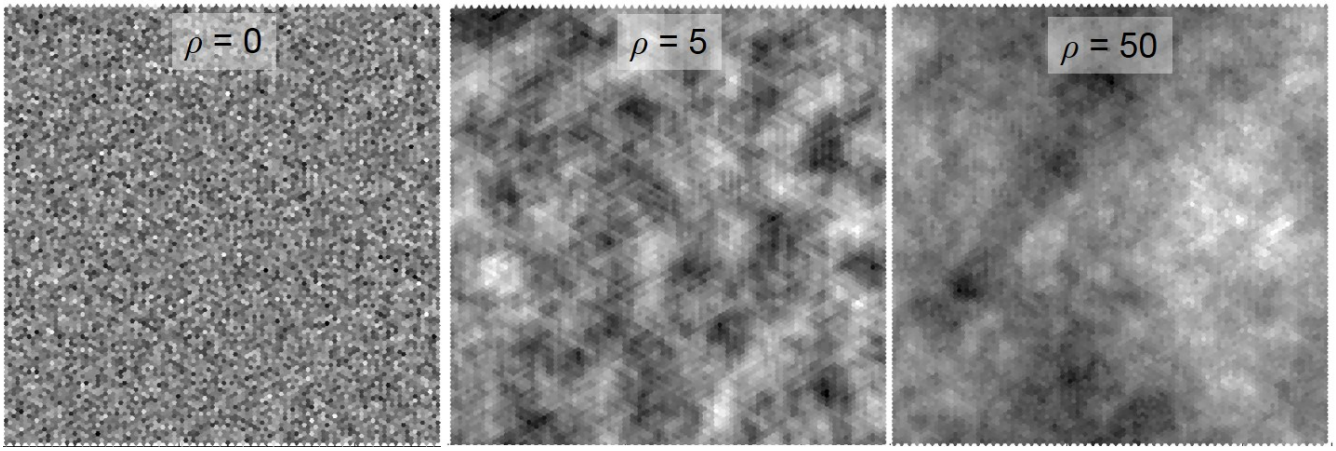


Figure 5.1 - Simulated landscape maps with different levels of habitat (grey scale) spatial autocorrelation.

At each simulation time-step, *t*, the position of the animal at the next time-step, $x_{t+\tau}$ (where τ is the duration of a single simulation step), is stochastically determined according a truncated redistribution kernel, including the current location and the six adjacent cells. The attractiveness of each of the seven available cells is calculated as an exponential function of a basal movement cost, μ, and a habitat induced attraction (or repulsion), a product of local habitat quality, $h(x)$, and the habitat selection intensity, ω. The truncated redistribution kernel is thus given by:

$$p(x_{t+\tau} = x) = \frac{I(\|x - x_t\| \leq 1) \cdot \exp[\omega \cdot h(x) - \mu \cdot \|x - x_t\|]}{\sum I(\|x - x_t\| \leq 1) \cdot \exp[\omega \cdot h(x) - \mu \cdot \|x - x_t\|]}, \quad \text{Eq. 5.1}$$

where I is an indicator function (valued either 1 or 0, depending on the validity of the expression in parenthesis), and the denominator is a sum over all landscape cells so that the kernel sums to one. Note that here we chose to focus on a stepping-stone process as we aim to model animal movement behaviour at its most fundamental scale, that of a single step. Accordingly, we see the spatial unit of our simulation as approximately equivalent to the body-length of the moving animal, and the temporal unit, τ, as approximately equivalent to the time required to move one such spatial unit. Also note that, for simplicity, we did not include any directionality effects (neither persistence nor bias) in the simulation nor in the subsequent analysis (we refer the reader to Duchesne *et al.* 2015 for further details on this topic).

Equation 5.1 was used to generate 1,000 'observed' positional time series for each of the five spatial autocorrelation levels – one trajectory was simulated across each of the 1,000 landscape realisations for each of the five ρ values. Unless otherwise specified, all simulations were run with ω = 1 and μ = 1.7918 (corresponding to a habitat-independent movement probability of 0.5). Simulated trajectories were always initiated in the best cell in each landscape (the one with the highest *h* value). The first 20,000 (approximately twice the domain size) time steps of each trajectory were omitted as a 'burn-in period', whereas the remaining 10^5 time steps were rarefied by retaining every 100th position, resulting in an 'observed' time-series of T = 1,000 positions (with observed step-duration, $\Delta t = 1 = 100 \cdot \tau$). Rarefication was performed here so as to emulate typical telemetry data where most animals likely adjust their spatial position at a substantially higher rate (say, every few minutes or seconds) than the rate at which we sample their position (typically, every few hours).

Each of the resulting 5,000 rarefied trajectories were then separately analysed using RSA and SSA formulations of various levels of complexity (including several iSSAs), and maximum-likelihood coefficient estimates were obtained for each realisation. For the RSA, occurrences were coded as 1 ('used') whereas all cells in the landscape were coded as 0 ('available'; see Discussion for implications). The resulting binary response variable was then statistically modeled as function of the local habitat value, $h(x)$, using logistic regression (function *glm* in R with a Binomial eror distribution and a logit

link). This yielded, for each observed trajectory, two RSA-based coefficients estimates, an intercept (a meaningless scaling parameter in the case of our 'used vs available' design), and a selection coefficient, β_{RSA}.

For the SSAs, each observed point along each rarefied trajectory (but the first; x_t, $t = 2, 3, \ldots, T$) was matched with an availability set of $s = 10$ random spatial positions ($x'_{t,i}$, $i = 1, 2, \ldots, 10$). The probability of sampling an available point was a sole function of its distance from the previously observed point (i.e., the length of the potential step ending at that point, $l'_{t,i}$), and was given by a gamma distribution with ρ-specific shape and scale parameters (β_1 and β_2) estimated from the joint distribution of observed step-lengths across all 1,000 realisations. To enable log-transformations and for consistency with kernel generating (see Appendix S4, section 9), the length of all steps (used and available) that resulted in 0 displacement (staying in the same cell) was set to $1/2\pi$. As in the RSA, used (i.e., observed) points were coded as 1 ('case') whereas available points were coded as 0 ('control'). The resulting binomial response variable was then statistically linked to various covariates using conditional (case-control) logistic regression (function *clogit* in R, with point ID as the strata), fitting an independent model to each trajectory.

Ten different SSA formulations were fitted using one or more of the following six covariates (see Table 1): habitat values at the end of each step, $h(x_t)$, the average habitat value along each step, $h(x_{t-1},x_t)$, the step-length, l_t ($= ||x_{t-1} - x_t||$), its natural-log transformation, $ln(l_t)$, and the statistical interactions between l_t, $ln(l_t)$, and $h(x_{t-1},x_t)$. Note that 'steps' are treated here as straight line segments along which the habitat is averaged. This is the most common formulation found in the literature but is certainly not the only one. Steps may be instead defined, for example, as ellipses bounded by two consecutive positions, or even as a Brownian bridge (a spatial probability density kernel derived from an explicit diffusion process). Models that included only $h(x_t)$ and/or $h(x_{t-1},x_t)$ correspond to traditionally used SSA (models *a*, *b*, and *c* in Table 1), whereas models that additionally included l_t and $ln(l_t)$ correspond to iSSA. The likelihood of a single observed trajectory given the full model (including all six covariates; model *j* in Table 1) is exactly proportional to:

$$\prod_{t=2}^{1000} \frac{\exp[\beta_3 \cdot h(x_t) + \beta_5 \cdot h(x_{t-1},x_t) + [\beta_5 + \beta_7 \cdot h(x_{t-1},x_t)] \cdot l_t + [\beta_6 + \beta_8 \cdot h(x_{t-1},x_t)] \cdot \ln(l_t)]}{\sum_{i=0}^{10} \exp[\beta_3 \cdot h(x'_{t,i}) + \beta_4 \cdot h(x_{t-1},x'_{t,i}) + [\beta_5 + \beta_7 \cdot h(x_{t-1},x'_{t,i})] \cdot l'_{t,i} + [\beta_6 + \beta_8 \cdot h(x_{t-1},x'_{t,i})] \cdot \ln(l'_{t,i})]}, \quad \text{Eq. 5.2}$$

where the 0'th available step correspond to the used step ($x'_{t,i=0} = x_t$; note the similarity to Eq. 3 in the main text). The derivation of Eq. 5.2 is identical in form to that of Eq. 3 (see Appendix S3) so is omitted here. This formulation allows for statistically inferring a hypothetical movement process with gamma-distributed step-lengths (with shape and scale that could be governed by the traversed habitats), a habitat-mediated step selection, and a habitat-mediated destination (i.e., endpoint) selection.

Quantifying and comparing utilisation distributions: The predictive capacity of the models was estimated based on the agreement between their predicted utilisation distributions (UD) and the 'true' UD, generated by the true underlying movement process (i.e., Eq. 5.1). UDs were generated across a single landscape realisation for each of the five ρ values ('validation landscapes', independent of those used to generate the movement data), where the same validation landscape was used to generate true-UDs, RSA-based UDs, and SSA-based UDs. The RSA-based UD value at each landscape cell, x, was calculated as $\frac{exp(\widetilde{\beta_{\text{RSA}}} \cdot h(x))}{\sum_{x=1}^{11600} exp(\widetilde{\beta_{\text{RSA}}} \cdot h(x))}$, with $\widetilde{\beta_{\text{RSA}}}$ being the median of the 1,000 β_{RSA} estimates obtained for any particular ρ value (the median was used here as a measure of centrality that is relatively insensitive to outliers).

For the movement models (the true process, given by Eq. 5.1, and the step-selection process parameterized based on the various SSAs), steady-state utilisation distributions were generated by simulating 11,600 trajectories, each starting from a different cell across each of the five validation landscapes (see Appendix S1 for alternative approaches). For the true UDs, these simulations were carried out using Eq. 5.1, with the same parameter values used to generate the original simulated trajectories ($\omega = 1$ and $\mu = 1.7918$). As before, the first 20,000 time-steps of each trajectory were omitted as a 'burn-in period' and the remaining 10^5 time-steps were retained. The number of time steps spent in each cell in the domain (across all 11,600 simulated trajectories) was then divided by $1.16 \cdot 10^9 \, \tau$ (the total time spent in the domain by all simulated individuals) to yield the true UD for a single landscape realisation of a given ρ value. SSA-based UDs were generated in a similar fashion to the true UDs, but using the parameterized step-selection models instead of Eq. 5.1, with parameter values corresponding to the median of the 1,000 β_1 estimates obtained for any particular ρ value. To match the temporal extent of true UD simulations ($\Delta t = 100 \cdot \tau$), SSA simulations were run for 1,200 steps and the initial 200 steps where omitted as a burn-in period.

Thus far we have described the generation of (approximately) steady-state UDs, pertaining to the population's spatial distribution as measured over extended time periods, and assumed to be temporally stable. However, for our movement models it also is possible to calculate and compare transient UDs by simulating movement over shorter time spans. The shorter the time span, the more sensitive the UD is to initial conditions (i.e., the path's starting point). Transient UDs were simulated similarly to the 'steady-state' UDs, except for three differences. First, simulated trajectories were always initiated in the best cell in the landscape (the one with the highest h value). Second, no burn-in period was omitted for the SSA-based simulations, and only the first 99 steps were omitted from the true simulations (to match the SSA

starting conditions). Third, each path lasted only $10^4 \cdot \tau$ (or $100 \cdot \Delta t$ in the SSA case). Other details are as above.

To compare the true-UDs (*truth*) to the statistically predicted ones (*model*) we use the Kullback-Leibler Divergence, *KLD*(*model*,*truth*), a measure of the information lost when the latter is used to approximate the former (see Potts *et al.* 2014c for similar usage). Note that the 'earth mover's distance' (Rubner *et al.* 2000), also may be a useful tool if one wishes to analyse the detailed movement dynamics rather than the resulting UD, as it can give various aspects of information regarding model bias and predictive power (Potts *et al.* 2014a). To facilitate intuitive interpretation of the resulting values, we further define a *KLD*-based performance measure (i.e., a pseudo R^2), G_{KLD}, quantifying goodness-of-fit relative to an information-free null model – a uniform UD (*null*):

$$G_{KLD}(model, null, truth) = e^{-KLD(model,truth)/KLD(null,truth)} \quad \text{Eq. 5.3}$$

Hence, the better the focal model performs compared to a null expectation, the closer its G_{KLD} value would be to 1, whereas if it performs worse than the null, its G_{KLD} value would drop below ~0.368 and asymptotically approach 0.

Appendix S6 – Evaluating the iSSA parameter identifiability and estimability

To evaluate the identifiability and estimability of the iSSA parameters we simulated three, 10^6 steps-long, trajectories across the intermediate spatial autocorrelation (ρ = 5) validation landscape (see Appendix S5 for details about landscape configurations). Trajectories were generated by stochastic sampling out of a redistribution kernel of the general form of Eq. 1 in the main text.

Methods: The movement kernel was defined by a uniform-random turn distribution (i.e., no directional persistence or bias) and gamma distributed step-lengths. The shape, k, and scale, q, of the step length distribution varied across the landscape as functions of the habitat value at the step's start point, $h(x)$:

$$k(x) = 2 \cdot \left[1 + \frac{h_{max} - h(x)}{h_{max} - h_{min}}\right], \quad \text{Eq. 6.1}$$

and

$$q(x) = 5 \cdot \left[1 + \frac{h(x) - h_{min}}{h_{max} - h_{min}}\right]^{-1}, \quad \text{Eq. 6.2}$$

where h_{max} and h_{min} are the maximal and minimal habitat values occurring on the landscape. Hence, k is a linear function of h with intercept = $2\left(1 + \frac{h_{max}}{h_{max} - h_{min}}\right)$ and slope = $\frac{-2}{h_{max} - h_{min}}$; as h increases, k decreases from 4 to 2. Similarly, q^{-1} is a linear function of h with intercept = $0.2\left(1 - \frac{h_{min}}{h_{max} - h_{min}}\right)$ and slope = $\frac{0.2}{h_{max} - h_{min}}$; as h increases, q decreases from 5 to 2.5. The habitat selection function was given by $exp[\omega \cdot h(x)]$ with one of three values of the selection coefficient, ω = 0, 1, or 2 (hence the three different trajectories). Biologically, this scenario may correspond, for example, to habitat-induced foraging behavior where animals tend to move less as their local habitat value increases (area restricted search), while simultaneously selecting locations with high habitat value.

As described in Appendix S5, each used point (but the first; x_t, t = 2, 3, …, 10^6) along each of the three trajectories was matched with an availability set of s = 10 random spatial positions ($x'_{t,i}$, i = 1, 2, …, 10). The probability of sampling an available point was a sole function of its distance from the previously observed point (i.e., the length of the potential step ending at that point, $l'_{t,i}$), and was given by a gamma distribution with shape and scale parameters (β_1 and β_2) estimated based on all observed step-lengths along each trajectory. Used points were coded as 1 ('case') whereas available points were coded as 0 ('control'). The resulting binomial response variable was then statistically linked to the following covariates using conditional (case-control) logistic regression (function *clogit* in R with point ID as the strata): the habitat values at the end of each step, $h(x_t)$, the step-length, l_t, its natural-log

transformation, $ln(l_t)$, and the products of the habitat value at the step's start point, $h(x_{t-1})$ and the step length and its natural-log transformation, $l_t \cdot h(x_{t-1})$ and $ln(l_t) \cdot h(x_{t-1})$.

To evaluate identifiability, parameters were estimated based on the entire trajectory (999,999 steps; this takes 5-10 CPU minutes). To evaluate estimability under constrained sample size, the model was fitted 1,000 times, each time with a different segment of the full trajectory, where segment length varied from 100 to 1,000 observed points (and hence sample size varied from 99 to 999 steps).

Results and discussion: All model fits converged successfully and in a timely manner. Parameter estimates obtained based on the full trajectory were unbiased (i.e., accurate) compared to the true values. Moreover, parameter estimates obtained based on shorter segments of the trajectory were also, on average, unbiased compared to the true values.

A model may be considered unidentifiable if the maximum of the likelihood surface occurs at more than just a single point in the parameter space, often along a 'ridge' (a line or a line segment through the parameter space along which the likelihood is exactly the maximum likelihood). Where the likelihood function is analytically tractable, detecting model unidentifiability is simply a question of whether the maximum of the likelihood function is unique. Most often however, numerical approximations are used to obtain MLE's, approximations that may fail to detect a ridge in the likelihood surface, and hence an unidentifiable model. We thus rely here on two indicators of model identifiability. First, we rerun the model fit 10 times for each of the three trajectories, making sure that parameter estimates did not change between fits (to the 8^{th} significant figure). Second, we used the standard errors of the parameter estimates to assess the steepness of the likelihood profile around the MLEs, where large standard errors (despite the very large sample size) would have been indicative of a flat likelihood profile and hence a potential ridge. All standard errors were very small (relative to the magnitude of the parameter estimates) and we thus conclude all five parameters in this scenario are identifiable, regardless of the strength of habitat selection (Figs 6.1-6.3).

Identifiability is a necessary but not sufficient condition for estimability. To evaluate estimability we focus here on the precision of parameter estimates obtained from multiple independent realizations of the same process. Low precision means that different realizations of the same process yield markedly different parameter estimates, indicating an estimability problem. As can be expected, precision (and hence estimability) increased with sample size, but with diminishing returns beyond ~400 observed positions (Figs 6.1 – 6.3). For the smallest sample size used here (99 steps) precision was very poor, particularly for the parameters related to the scale of the step-length distribution. Estimates of the intercept of the shape function, and both the intercept and the slope (with regards to h) of the inverse

scale function, show decreased precision as selections strength increases, a pattern not observed for the other two parameters.

Low precision, particularly when sample size is low, was likely driven by high correlations between some of the statistical coefficients (Table 6.1). The movement components of the iSSA are inherently 'correlation-prone' and are hence vulnerable to estimability issues. Whereas the current analysis indicates satisfactory inferential performance, the results may differ under different scenarios. It is thus important to note that estimability analysis should be tailored to the specific system and model that are being evaluated. This is particularly true when additional movement components are included in the model, such as directional persistence and/or bias, where strong collinearity may result in loss of estimability. We highly recommend users to take advantage of the mechanistic nature of the iSSA and simulate potential space-use patterns under the desired model structure so that estimability can be adequately assessed.

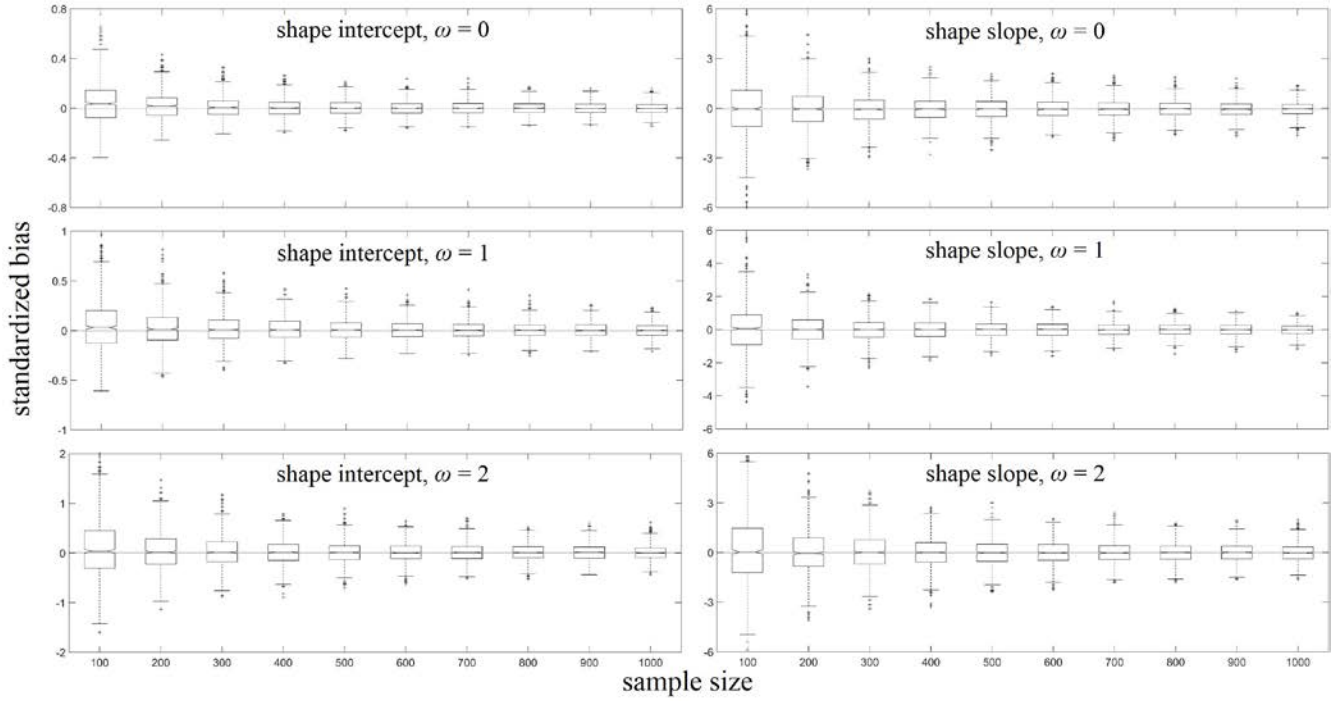


Figure 6.1 – the standardized bias (the difference from the true value divided by that true value) in the estimated linear intercept and slope (the effect of *h*) of the shape parameter of the step-length distribution as function of sample size. Grey lines reflect the 95% confidence region of the estimates obtained based on the full trajectory (999999 steps; based on the estimated standard errors) and is hence a measure of identifiability.

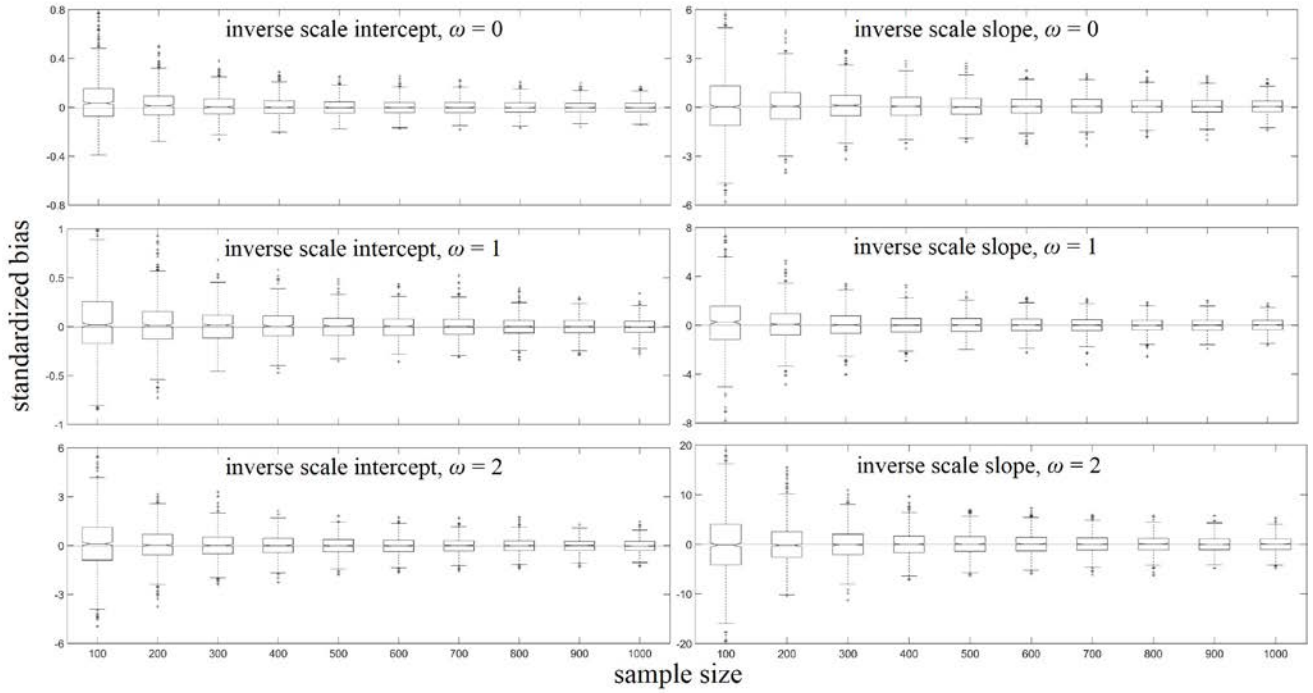


Figure 6.2 – the standardized bias in the estimated linear intercept and slope of the inverse scale parameter of the step-length distribution as function of sample size. Grey lines reflect the 95% confidence region of the estimates obtained based on the full trajectory.

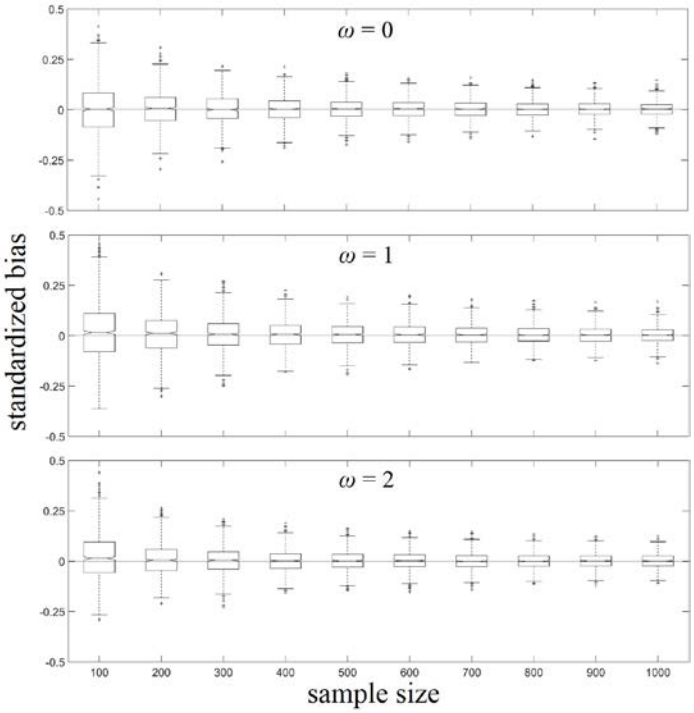


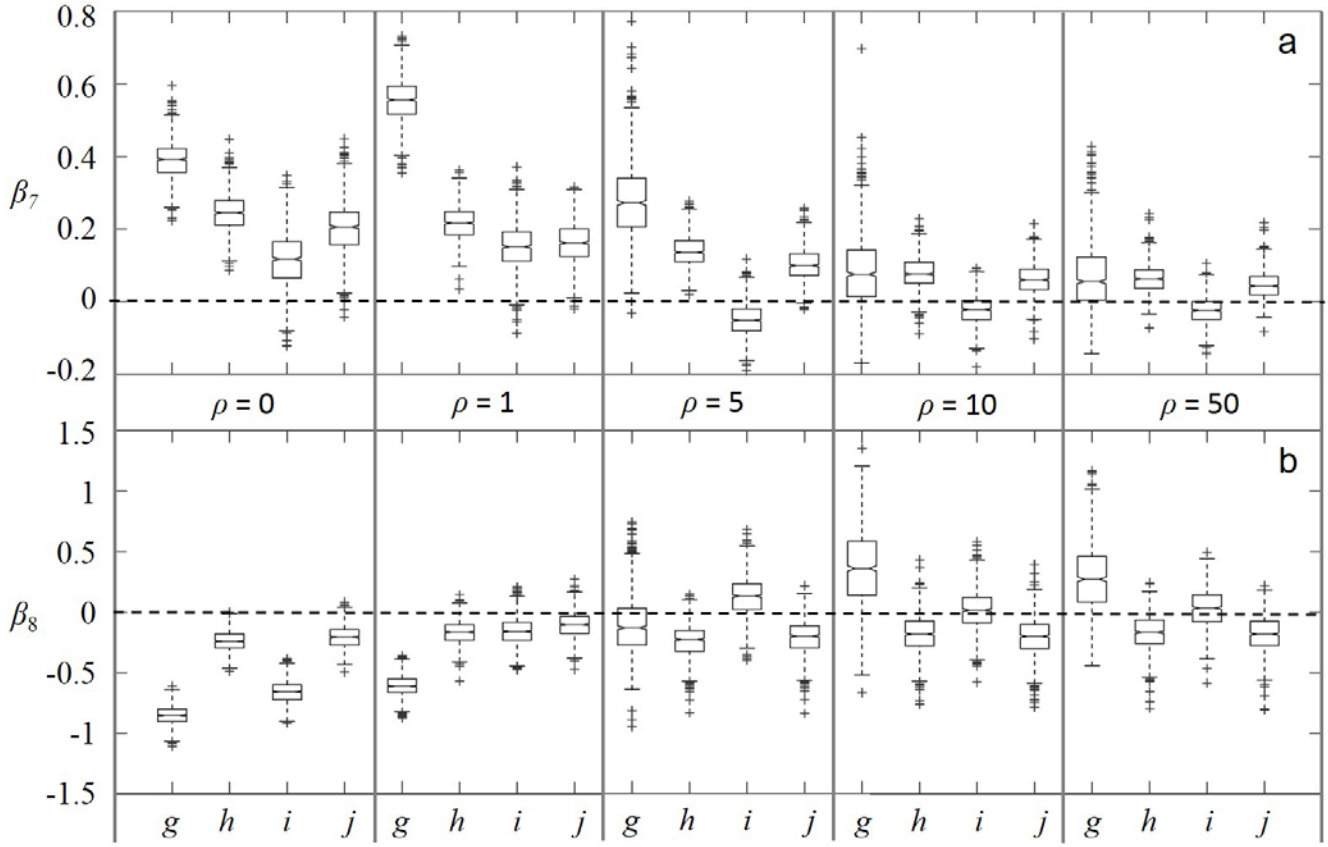
Figure 6.3 – the standardized bias in the estimated selection coefficient as function of sample size. Grey lines reflect the 95% confidence region of the estimates obtained based on the full trajectory.

covariate 1	covariate 2	Pearson's *r*		
		$\omega = 0$	$\omega = 1$	$\omega = 2$
$h(x_t)$	l_t	0.067	0.001	0.111
	$ln(l_t)$	-0.074	0.016	-0.128
	$l_t \cdot h(x_{t-1})$	-0.120	0.021	-0.049
	$ln(l_t) \cdot h(x_{t-1})$	0.191	0.063	0.164
l_t	$ln(l_t)$	-0.914	-0.874	-0.833
	$l_t \cdot h(x_{t-1})$	0.071	-0.715	-0.959
	$ln(l_t) \cdot h(x_{t-1})$	0.133	0.782	0.840
$ln(l_t)$	$l_t \cdot h(x_{t-1})$	0.054	0.543	0.739
	$ln(l_t) \cdot h(x_{t-1})$	-0.270	-0.865	-0.981
$l_t \cdot h(x_{t-1})$	$ln(l_t) \cdot h(x_{t-1})$	-0.867	-0.778	-0.789

Table 6.1 – pairwise Pearson's correlations among model coefficients. Correlations were calculated based on the 1000 repetitions of the model fit with a sample size of 999.

Appendix S7 - β_7 and β_8

Statistically inferred interactions between the mean habitat along the step and the step-length (a) and the natural logarithm of the step-length (b). The dashed line represents no effect. Other details are as in Fig 2 in the main text.



Appendix S8 – predictive performance

The 11 different models evaluated here (10 SSAs + 1 RSA) and their Kullback-Leibler divergence from the true UDs (steady-state and transient) at five different levels of habitat spatial autocorrelation. SSA formulations including an endpoint effect (β_3) are highlighted in grey. Models *d* to *j* include step-length effects (iSSA). The lowest (best-performing) SSA Kullback-Leibler divergence values at each level of habitat spatial autocorrelation are bolded to allow comparison with the corresponding RSA values. For reference, the first row lists the KLD values for null model - a uniform UD. Note the superior performance of the RSA projections in predicting the steady-state UDs. Interestingly, a similar result was obtained by Fieberg (2007), who found that ignoring the autocorrelation in the positional time-series (as is done in RSA) often led to slightly better estimates of the UD.

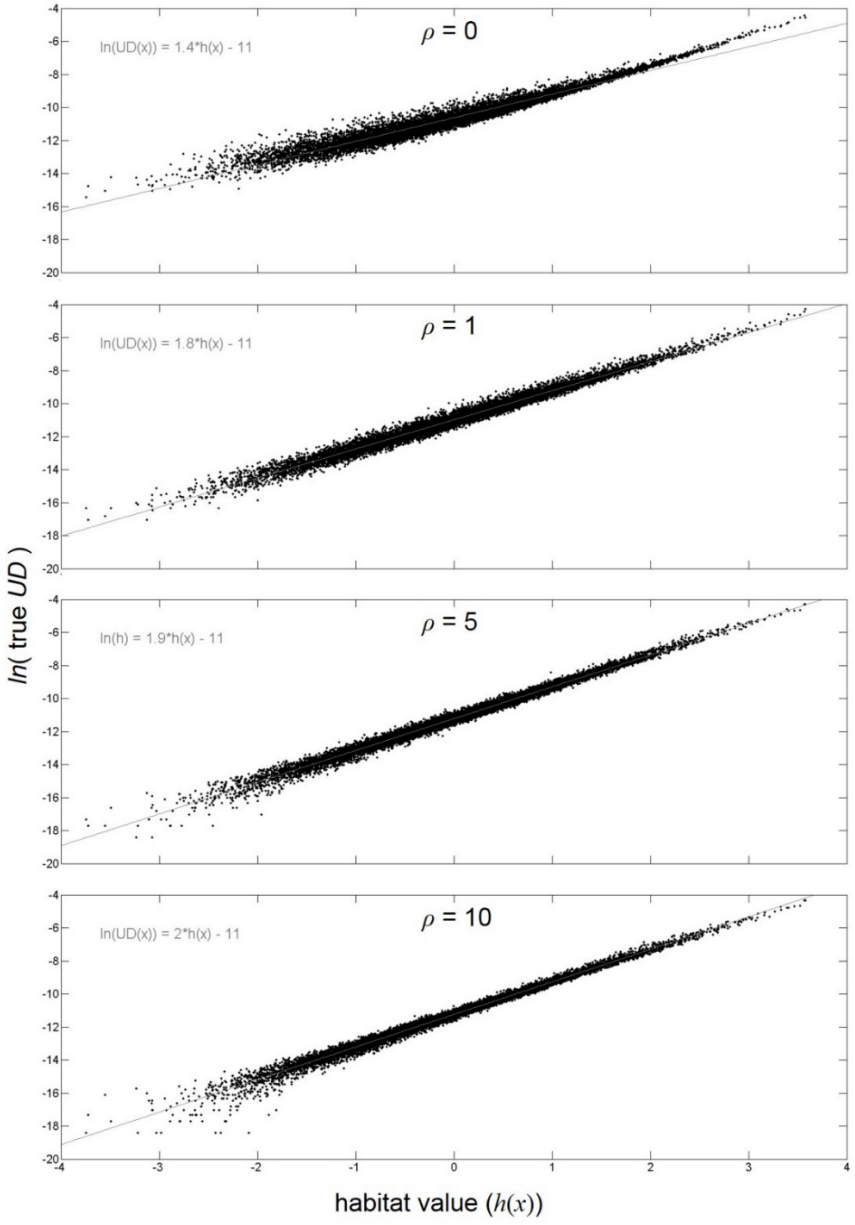
model	KLD (steady state)					KLD (transient)				
	$\rho = 0$	$\rho = 1$	$\rho = 5$	$\rho = 10$	$\rho = 50$	$\rho = 0$	$\rho = 1$	$\rho = 5$	$\rho = 10$	$\rho = 50$
null	1.48	1.66	1.80	1.78	1.84	2.78	3.62	3.00	2.66	2.27
RSA	0.02	0.02	0.01	0.02	0.01	1.02	1.28	0.65	0.34	0.07
a	0.28	0.19	**0.08**	**0.08**	**0.12**	2.52	1.05	0.10	**0.06**	0.09
b	1.14	0.80	0.26	0.24	0.18	1.61	0.76	0.28	0.26	0.21
c	0.32	0.19	0.10	0.11	0.19	2.09	0.73	0.14	0.08	0.13
d	**0.02**	**0.09**	0.14	0.13	0.16	**0.08**	**0.07**	**0.09**	0.08	**0.09**
e	1.05	0.92	0.28	0.29	0.21	0.95	1.17	0.29	0.29	0.21
f	0.03	0.15	0.31	0.19	0.33	0.12	0.13	0.14	0.11	0.18
g	0.52	0.36	0.95	1.42	1.55	4.42	3.03	0.82	1.25	0.92
h	0.19	0.12	0.18	0.11	0.22	0.32	0.09	0.10	0.08	0.12
i	0.78	0.79	0.30	0.27	0.21	4.49	0.93	0.28	0.27	0.21
j	0.05	0.15	0.22	0.18	0.19	0.13	0.13	0.11	0.12	0.11

Appendix S9 – interpreting ρ

Statistical inference of animal movement behaviour is highly scale-dependent, varying with both the temporal resolution of the observed positional time series and the spatial resolution (or grain) of the landscape. For example, Avgar *et al.* (2013) investigated the sensitivity of parameter estimates in their cognitive-movement model to coarsening (or rarefying) both temporal and spatial resolutions of their simulated data. Here we evaluate these scale-sensitivities by varying a single parameter, ρ, the landscape's spatial autocorrelation range. Note however that ρ be interpreted in various ways. The most straightforward interpretation is to think of ρ as a measure of habitat patchiness, ranging from no patches ($\rho = 0$), through multiple small patches, to a single patch covering (on average) half of the spatial domain ($\rho = 50$). Alternatively, ρ may be thought of as a being inversely related to the gap between the scale of the behavioural process and the scale of the observation. In between observed positions (in our case, 99% of the time), a moving animal experiences substantially more environmental variability as ρ declines from 50 to 0, variability that is not observed in the rarefied data. Hence, increasing ρ can be interpreted as decreasing rarefication, and its effect on parameter estimates and predictive capacity can be viewed in this light. Lastly, ρ can be interpreted as a measure of the animal's movement (or patch departure) rate. During the observation period, an animal may visit many different habitat patches ($\rho = 0$), or stay in a single one ($\rho = 50$), depending on a variety of factors (e.g., the ratio between the animal's metabolic requirements and the patch's productivity). Instead of explicitly modeling these factors, ρ can be thought of as crudely representing the resulting spatiotemporal relationships between the animal and the grain of the landscape. For an in-depth discussion of spatial scale dependency we refer the reader to Sandel (2015).

Appendix S10 – habitat selection and utilisation distribution

The natural logarithm of utilisation distributions emerging from a simple stepping-stone movement process with exponential habitat selection at four levels of spatial autocorrelation, plotted against the underlying habitat value at each cell (the fifth level, ρ = 50, was omitted as it is indistinguishable from ρ = 10). Grey lines represent best fit linear regression, with their parameter estimates presented on the top left corner of each panel. Note that our spatially discrete simulation model does not allow the scale of the movement to exceed the scale of habitat variation, only to equal it (ρ = 0), and as a result, the observed log-linear slope does not approach 1 but is rather minimised at an intermediate value.



Appendix S11 – appendices reference list

Avgar, T., Deardon, R. & Fryxell, J.M. (2013). An empirically parameterized individual based model of animal movement, perception and memory. *Ecological Modelling*, **251**, 158–172.

Barnett, A.H. & Moorcroft, P.R. (2008). Analytic steady-state space use patterns and rapid computations in mechanistic home range analysis. *Journal of Mathematical Biology*, **57**, 139–59.

Duchesne, T., Fortin, D. & Rivest, L. (2015). Equivalence between step selection functions and biased correlated random walks for statistical inference on animal movement. *PLoS ONE*, **in press**.

Fieberg, J. (2007). Kernel density estimators of home range: smoothing and the autocorrelation red herring. *Ecology*, **88**, 1059-1066.

Fieberg, J., Matthiopoulos, J., Hebblewhite, M., Boyce, M.S. & Frair, J.L. (2010). Correlation and studies of habitat selection: problem, red herring or opportunity? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **365**, 2233-2244.

Forester, J.D., Im, H.K. & Rathouz, P.J. (2009). Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. *Ecology*, **90**, 3554–65.

Hillen, T. & Painter, K. (2013). Transport and anisotropic diffusion models for movement in oriented habitats. *Dispersal, individual movement and spatial ecology: A mathematical perspective* (eds M.A. Lewis, P.K. Maini & S. V. Petrovskii), pp. 177–222. Lecture Notes in Mathematics. Springer Berlin Heidelberg, Berlin, Heidelberg.

Kareiva, P.M. & Odell, G. (1987). Swarms of predators exhibit 'preytaxis' if individual predators use area-restricted search. *American Naturalist*, **130**, 233–270.

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

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

Okubo, A. & Levin, S.A. (2001). *Diffusion and Ecological Problems, Modern Perspectives*, 2nd edn. Springer.

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

Potts, J.R., Bastille-Rousseau, G., Murray, D.L., Schaefer, J. a. & Lewis, M. a. (2014b). Predicting local and non-local effects of resources on animal space use using a mechanistic step-selection model. *Methods in Ecology and Evolution*, **5**, 253–262.

Potts, J.R., Mokross, K. & Lewis, M.A. (2014c). A unifying framework for quantifying the nature of animal interactions. *Journal of the Royal Society Interface*, **11**.

Rubner, Y., Tomasi, C. & Guibas, L. (2000). The Earth Mover's Distance as a Metric for Image Retrieval. *International Journal of Computer Vision*, **40**, 99–121.

Sandel, B. (2015). Towards a taxonomy of spatial scale-dependence. *Ecography*, **38**, 358–369.

Thurfjell, H., Ciuti, S. & Boyce, M.S. (2014). Applications of step-selection functions in ecology and conservation. *Movement Ecology*, **doi:10.118**.