REVIEW

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

Understanding step selection analysis through numerical integration

Théo Michelot[1] ⓘ | Natasha J. Klappstein[1] ⓘ | Jonathan R. Potts[2] ⓘ | John Fieberg[3] ⓘ

[1]Dalhousie University, Halifax, Nova Scotia, Canada

[2]University of Sheffield, Sheffield, UK

[3]University of Minnesota, Saint Paul, Minnesota, USA

Correspondence
Théo Michelot
Email: theo.michelot@dal.ca

Abstract

1. Step selection functions (SSFs) are flexible statistical models used to jointly describe animals' movement and habitat preferences. The popularity of SSFs has grown rapidly, and various extensions have been developed to increase their utility, including the ability to use multiple statistical distributions to describe movement constraints, interactions to allow movements to depend on local environmental features, and random effects and latent states to account for within- and among-individual variability. Although the SSF is a relatively simple statistical model, its presentation has not been consistent in the literature, leading to confusion about model flexibility and interpretation.

2. We believe that part of the confusion has arisen from the conflation of the SSF model with the methods used for statistical inference, and in particular, parameter estimation. Notably, conditional logistic regression (CLR) can be used to fit SSFs in exponential form, and this model fitting approach is often presented interchangeably with the actual model (the SSF itself). However, reliance on CLR reduces model flexibility, and suggests a misleading interpretation of step selection analysis as being equivalent to a case–control study.

3. In this review, we explicitly distinguish between model formulation and inference technique, presenting a coherent framework to fit SSFs based on numerical integration and maximum likelihood estimation. We provide an overview of common numerical integration techniques (including Monte Carlo integration, importance sampling and quadrature), and explain how they relate to popular methods used in step selection analyses.

4. This general framework unifies different model fitting techniques for SSFs, and opens the way for improved inferential methods. In this approach, it is straightforward to model movement with distributions outside the exponential family, and to apply different SSF model formulations to the same data set and compare them with AIC. By separating the model formulation from the inference technique, we hope to clarify many important concepts in step selection analysis.

KEYWORDS
animal movement, habitat selection, Monte Carlo integration, step selection function

1 | INTRODUCTION

The increased availability of animal tracking data has led to the widespread use of statistical methods to estimate habitat selection at the scale of the observed movement step. Perhaps the most common model is the step selection function (SSF; Rhodes et al., 2005; Fortin et al., 2005), whereby the likelihood of moving to the spatial location x_{t+1} given previous locations $x_{1:t} = \{x_1, x_2, \ldots, x_t\}$ is in the following form,

$$p(x_{t+1} \mid x_{1:t}) = \frac{w(x_t, x_{t+1})\phi(x_{t+1} \mid x_{1:t})}{\int_\Omega w(x_t, z)\phi(z \mid x_{1:t})dz}, \quad (1)$$

where Ω is the study region. The function w describes the effects of environmental variables (e.g. resources, risks and environmental conditions; Matthiopolous et al., 2023), and ϕ accounts for the effects of movement constraints (e.g. on the range of observed step lengths). The habitat selection function is often assumed to take an exponential (or 'log-linear') form, that is, $w(x_t, x_{t+1}) = \exp\{h(x_t, x_{t+1})\beta_h^\mathsf{T}\}$, where $h(x_t, x_{t+1})$ is a vector of habitat variables for the step, and β_h is the vector of associated selection parameters. The form of ϕ reflects assumptions about movement patterns of the animal, and it is often written as a function of the step length and turning angle to capture movement speed and tortuosity of an animal's movement. We call 'SSF' the numerator of Equation (1), but the terminology is not consistent across the literature, and the term has been used variously to refer to w, to $w \times \phi$, or to the whole right-hand side of Equation (1). We choose to define $w \times \phi$ as the SSF to reflect the fact that an animal's selection of a step is based on both habitat preferences and movement constraints. Figure 1 shows an example SSF with the two model components.

Step selection analysis refers to a wide range of methods for applying SSFs to animal tracking data, with the aim to estimate the parameters of the habitat selection function w and the movement kernel ϕ. Although the data-generating mechanism for this model is described by Equation (1), this is a difficult computational problem due to the presence of the integral in the denominator. This integral is required in Equation (1) to ensure that $p(x_{t+1} \mid x_{1:t})$ integrates to 1, that is, that it is a valid probability density function with respect to

x_{t+1} (Forester et al., 2009; Rhodes et al., 2005). It also has a more intuitive interpretation: to evaluate the likelihood of a given step, we weigh its suitability against the suitability of all other possible steps in the study region. Here, 'suitability' refers both to the habitat quality of a location (as captured by w) and to its accessibility (as captured by ϕ). This integral cannot generally be calculated analytically, because the integrand (i.e. the expression that is integrated) depends on w, which is usually a function of environmental covariates with no mathematically convenient functional forms. That is, the integral cannot be rewritten in terms of simple functions that could be directly implemented with a computer, and so Equation (1) cannot easily be evaluated for a given movement track and set of parameters.

Although the integral in Equation (1) cannot be computed directly, methods have been developed to replace the expression by a tractable approximation. In some cases, such approximations are equivalent to applying conditional logistic regression (CLR) to a case–control data set, where each observed location ('case') is associated with a set of locations from the landscape ('controls'; Forester et al., 2009). This has been a popular framework for step selection analysis, because CLR can be fitted quickly and conveniently using statistical software (e.g. using the survival package in R; Therneau, 2023). Consequently, SSFs are often conflated with CLR, even though the latter is merely a convenient tool to fit the former by approximating the likelihood in some special cases. In our view, this presentation can lead to confusion about model interpretation, and reduces the flexibility of step selection analyses. In particular, we can avoid the need to make strong assumptions about the functional forms of w and ϕ if we are willing to use numerical methods other than CLR for parameter estimation.

In this review, we show that most methods used to estimate parameters in step selection analyses can be viewed as applications of numerical integration. A variety of numerical integration techniques are commonly applied in statistics and mathematics to approximate integrals in cases where there is no known formula to compute them exactly. We will present several numerical tools developed for this purpose, and describe their utility in step selection analyses. This perspective suggests we can contrast existing methods (e.g. to identify those with lower approximation errors), and opens the way for improved inferential techniques in step
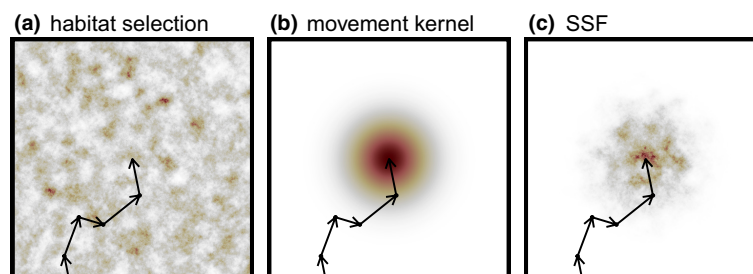
(a) habitat selection **(b)** movement kernel **(c)** SSF

FIGURE 1 Example model components at some time t, where the last point x_t is at the centre of each panel and the previous steps are shown as black segments: (a) habitat selection function $w(x_t, z)$, (b) movement kernel $\phi(z \mid x_t)$ based on distance from x_t and (c) resulting step selection function (SSF) $w(x_t, z)\phi(z \mid x_t)$. The integral on the denominator of Equation (1) is the volume under the SSF, which is required to transform the SSF into a probability distribution (sometimes called the step density).

selection analysis. Lastly, we hope that our review will motivate further exploration of numerical integration and estimation methods, which have broad utility across a wide range of ecological applications.

2 | LIKELIHOOD APPROXIMATION IN STEP SELECTION ANALYSIS

2.1 | Maximum likelihood estimation

We consider n observed locations $x_1, x_2, \ldots, x_n$, recorded without measurement error at regular time intervals. The goal of step selection analysis is to estimate the parameters β_h of the habitat selection function w, which quantify the strength of selection or avoidance of spatial covariates, and the parameters β_m of the movement kernel ϕ, which quantify movement tendencies (e.g. speed and tortuosity). We denote as $\beta = (\beta_h, \beta_m)$ the vector of all parameters. Based on the standard assumptions that each step follows the model of Equation (1), and that successive steps are conditionally independent (i.e. given past locations), the likelihood of the parameters under the step selection model can be written as

$$L(\beta; x_1, \ldots, x_n) = \prod_{t=1}^{n-1} p(x_{t+1}|x_{1:t}) = \prod_{t=1}^{n-1} \frac{w(x_t, x_{t+1})\phi(x_{t+1}|x_{1:t})}{\int_\Omega w(x_t, z)\phi(z|x_{1:t})dz}.$$

Note that both w and ϕ depend on some of the parameters in β (specifically, w depends on β_h and ϕ on β_m), but we do not make this explicit for notational convenience. The likelihood (or log-likelihood) can be optimised numerically with respect to β, for example using the function `optim` or `nlm` in R, to obtain maximum likelihood estimates of β. For an overview of maximum likelihood estimation in an ecological context, see for example Chap. 6–7 of Bolker (2008) or Chap. 10 of Fieberg (2023).

Model fitting requires computing the likelihood function for the observed data, and therefore evaluating the integral

$$I = \int_\Omega w(x_t, z)\phi(z|x_{1:t})dz \tag{2}$$

for each time step $t \in \{1, 2, \ldots, n-1\}$. In this section, we describe several methods to approximate I by some quantity $\hat{I}$, which can be substituted in the likelihood formula to carry out approximate inference. The main approaches are summarised in Table 1.

2.2 | Monte Carlo integration

We use the term 'Monte Carlo integration' to refer to all forms of numerical integration that rely on random sampling (i.e. all methods presented in this review, except for quadrature). Monte Carlo integration is a method for evaluating an integral of the form $\int_\Omega f(z)g(z)dz$, where f is a probability density function (section 3.2 of Robert & Casella, 2010). The general idea is to generate a sample from the distribution f, and use it to find an unbiased estimate of the integral. It can be shown that

$$\int_\Omega f(z)g(z)dz \approx \frac{1}{K}\sum_{k=1}^{K} g(z_k), \quad \text{where } z_k \sim f, \tag{3}$$

with the numerical error of the approximation decreasing as K increases. Throughout this review, we use the notation '$z_k \sim f$' to indicate that z_k follows the distribution with probability density function f.

The most common approaches to fitting SSFs can be viewed as different forms of Monte Carlo integration applied to the integral in Equation (2), which result from different choices of the functions f and g (always with the constraint that $f \times g = w \times \phi$). Generally, this choice may impact the accuracy and precision of the approximation, so choosing it requires thought (see Section 2.2.3 and Rizzo, 2019).

Maximum likelihood estimation based on the numerical approximation of the integral in Equation (3) defines a general framework of approximate inference for SSFs. Methods of inference based on Monte Carlo likelihood approximations are common in econometrics, where they are usually called *simulated* maximum likelihood estimation, to highlight their inherent stochasticity (section 3.1.2 of Gourieroux & Monfort, 1996). Even though the Monte Carlo estimator of the integral given in Equation (3) is not biased, simulated maximum likelihood estimators generally are, due to the log-transformation of the likelihood before it is maximised (Gourieroux & Monfort, 1996). As a result, Monte Carlo-based step selection parameter estimators are biased, but this bias decreases with the number of integration points.

TABLE 1 Summary of most common numerical integration approaches used in step selection analyses: Monte Carlo with known movement kernel (MC), uniform Monte Carlo (UMC), importance sampling (IS) and uniform quadrature (UQ).

	Approximate likelihood	$z_k \sim$?	References
MC	$\dfrac{w(x_t, x_{t+1})}{\sum_{k=0}^{K} w(x_t, z_k)}$	$\phi(\cdot|x_{1:t})$	Fortin et al. (2005)
UMC	$\dfrac{w(x_t, x_{t+1})\phi(x_{t+1}|x_{1:t})}{\sum_{k=0}^{K} w(x_t, z_k)\phi(z_k|x_{1:t})}$	Uniform	Forester et al. (2009)
IS	$\dfrac{w(x_t, x_{t+1})\phi(x_{t+1}|x_{1:t})}{\sum_{k=0}^{K} w(x_t, z_k)\phi(z_k|x_{1:t})/h(z_k|x_{1:t})}$	User-defined h	Forester et al. (2009)
UQ	$\dfrac{w(x_t, x_{t+1})\phi(x_{t+1}|x_{1:t})}{\sum_{k=0}^{K} w(x_t, z_k)\phi(z_k|x_{1:t})}$	Regular grid	Rhodes et al. (2005)

Note: The columns include the approximate likelihood of a step from x_t to x_{t+1}, method for determining the distribution of integration points z_k, and references to key papers presenting each approach.

Using Monte Carlo integration, Equation (2) can be approximated as $I \approx \frac{1}{K} \sum_{k=1}^{K} g(z_k)$, for functions f and g chosen such that $z_k \sim f$ and $f \times g = w \times \phi$. In this review, we include the observed location x_{t+1} as an additional integration point in the random sample $\{z_1, \ldots, z_K\}$, and we denote $z_0 = x_{t+1}$ for convenience. This slight deviation from the formal definition of Monte Carlo integration is justified in this context for three reasons: (1) the resulting formulas have clear links to those presented in the literature, (2) it improves numerical stability and decreases bias for small K (as illustrated with simulations in Appendix A) and (3) the effect of this change vanishes for large values of K. The approximate likelihood of a step under Monte Carlo integration is then

$$p(x_{t+1} \mid x_{1:t}) \approx \frac{w(x_t, x_{t+1})\phi(x_{t+1} \mid x_{1:t})}{\frac{1}{K+1}\sum_{k=0}^{K} g(z_k)}, \quad \text{where } z_k \sim f,$$

for some general f and g. We present several important special cases below, which encompass most existing methods for step selection analysis (e.g. Fortin et al., 2005), including extensions proposed by Forester et al. (2009), Duchesne et al. (2015) and Avgar et al. (2016). A key characteristic of the first approach, described in Section 2.2.1, is that it assumes that the movement kernel ϕ is known prior to the step selection analysis. It has become increasingly common to jointly estimate the movement kernel and habitat selection, and the approaches in Sections 2.2.2 and 2.2.3 focus on that situation.

2.2.1 | Assuming that the movement kernel is known

A popular approach to step selection analysis is to define the movement kernel prior to fitting the SSF, typically from empirical or parametric distributions of step lengths and turning angles (Fortin et al., 2005). Then, since ϕ is assumed to be known, we can apply Monte Carlo integration to the integral of Equation (2) by choosing $f = \phi$ and $g = w$. Equation (2) becomes

$$I \approx \frac{1}{K+1} \sum_{k=0}^{K} w(x_t, z_k), \quad \text{where } z_k \sim \phi(\cdot \mid x_{1:t}).$$

That is, we generate the random Monte Carlo sample from the movement kernel ϕ, and we take the mean of the habitat selection function at those random points to evaluate the integral. This characteristic of the typical step selection analysis workflow has led practitioners to view the z_k as a sample from the 'available' landscape. The key limitation of this approach is that the movement kernel cannot be estimated jointly with habitat selection parameters because the points z_k are sampled from it prior to model fitting. In addition, this approach to estimating ϕ without consideration of w has been shown to result in biased parameter estimators since the observed movements are a function of both processes (Forester et al., 2009).

Using this method, the likelihood of a step (Equation 1) is approximated by

$$p(x_{t+1} \mid x_{1:t}) \approx \frac{w(x_t, x_{t+1})\phi(x_{t+1} \mid x_{1:t})}{\frac{1}{K+1}\sum_{k=0}^{K} w(x_t, z_k)} \propto \frac{w(x_t, x_{t+1})}{\sum_{k=0}^{K} w(x_t, z_k)}, \quad (4)$$

where $z_k \sim \phi(\cdot \mid x_{1:t})$. Note that multiplicative constants, that is, terms that do not depend on the parameters β, can be omitted from the likelihood function with no effect on inference. Here, ϕ can be omitted from the numerator because it is assumed known, and therefore does not depend on any estimated parameter, and $1/(K+1)$ is omitted from the denominator. When w is written in the usual exponential form, this is the likelihood of a CLR model, and parameters can be estimated using standard software such as the `clogit` function in the survival R package (Therneau, 2023). However, the method of Equation (4) has also been used for non-exponential functional forms for w, by using likelihood maximisation procedures other than CLR (Potts et al., 2014).

This approach was initially proposed by Fortin et al. (2005), and has been widely used due to the convenience of implementation using CLR, and the intuitive appeal of interpreting the random locations as a sample of availability. Estimating the movement kernel separately has drawbacks, however: as previously mentioned, this approach leads to biased parameter estimators and does not propagate statistical uncertainty about the movement model to the habitat selection parameters. In addition, it does not allow for interactions between the movement parameters and local environmental features. As a consequence, recent research in step selection analysis has focused on formulations where ϕ and w are estimated simultaneously (Avgar et al., 2016; Forester et al., 2009; Rhodes et al., 2005), and we present those in the next subsections. The interpretation of the random points z_k is different when ϕ is estimated as part of the SSF, and we discuss this in Section 4.

2.2.2 | Uniform Monte Carlo sampling

If the movement kernel ϕ is not known a priori, we must choose a different probability density function f from which to generate random points. One natural choice is to use a uniform distribution over the domain Ω, because it is easy to sample from. In this case, the two functions in Equation (3) are defined as $f(z) = 1/A(\Omega)$ (where $A(\Omega)$ is the area of Ω), and $g(z) = A(\Omega)w(x_t, z)\phi(z \mid x_{1:t})$. Then, Equation (2) can be approximated as

$$I \approx \frac{A(\Omega)}{K+1} \sum_{k=0}^{K} w(x_t, z_k)\phi(z_k \mid x_{1:t}), \quad \text{where } z_k \sim \text{Unif}(\Omega).$$

Intuitively, the random points are used to estimate the mean value of the SSF over Ω, and the integral is approximated by the product of that mean value and $A(\Omega)$ (see Figure 2a for a one-dimensional example).

This method, called 'uniform sampling' by Forester et al. (2009), allows for joint estimation of habitat selection and movement parameters (e.g. Schlägel & Lewis, 2014). However, the uniform sampling approach can be computationally demanding, because good
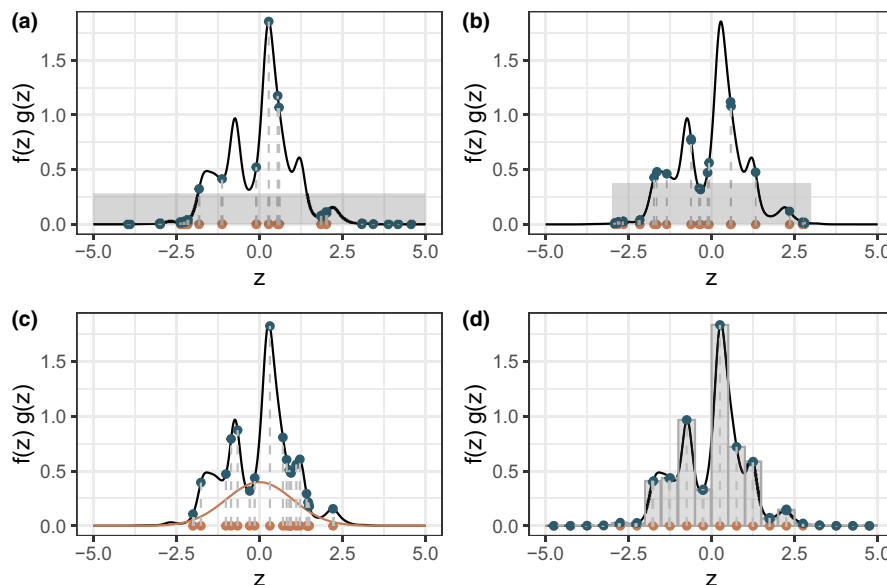
FIGURE 2 Illustration of numerical integration in one dimension, for the function $f \times g$ shown as a black line, over the interval $\Omega = [-5, 5]$. The orange dots are the integration points, and the blue dots are the corresponding function evaluations. (a) Monte Carlo integration with a uniform sample over Ω; the height of the grey rectangle is the mean of function evaluations. (b) Monte Carlo integration with a uniform sample over $[-3, 3]$. (c) Importance sampling with random points generated from a normal distribution that roughly approximates the integrand $f \times g$. (d) Quadrature over a regular grid using a Riemann sum. In (a), (b) and (d), the shaded area approximates the area under the curve; there is no such simple visualisation method for importance sampling. In this small simulated example, the true integral is 2.53, and the approximations are (a) 2.83, (b) 2.28, (c) 2.41 and (d) 2.80.

performance requires that the integration points provide adequate coverage of the study region Ω, and this can often only be achieved for large values of K. In most step selection analyses, the number of points can be greatly reduced based on the observation that, at each step, the SSF decreases sharply with distance from the start point x_t (due to movement constraints of the animal). Points far from x_t therefore contribute a negligible amount to the integral, and the approximation is virtually unchanged if the domain of integration is truncated to a disc around x_t, with radius large enough to encompass any possible step (Boyce et al., 2003; Craiu et al., 2008). Fewer points are needed to ensure good coverage of this disc, which reduces the computational cost of evaluating the integral (Klappstein et al., 2022). Uniform sampling on a truncated interval is illustrated in Figure 2b.

Using this approximation, Equation (1) becomes

$$p(x_{t+1} | x_{1:t}) \approx \frac{w(x_t, x_{t+1})\phi(x_{t+1} | x_{1:t})}{\frac{A(\Omega)}{K+1} \sum_{k=0}^{K} w(x_t, z_k)\phi(z_k | x_{1:t})} \propto \frac{w(x_t, x_{t+1})\phi(x_{t+1} | x_{1:t})}{\sum_{k=0}^{K} w(x_t, z_k)\phi(z_k | x_{1:t})},$$

(5)

where $z_k \sim \text{Unif}(\Omega)$. Like before, we remove $A(\Omega) / (K + 1)$ from the denominator because it is a constant and thus will not affect maximum likelihood estimation. Here, we cannot omit ϕ from the numerator because it is not assumed to be known, and therefore is a function of the parameters β of interest. If both ϕ and w have an exponential form, then Equation (5) is equal to the CLR likelihood, as long as we include the observed location as an integration point.

Note that uniform Monte Carlo sampling refers to *spatially uniform points*, and this should not be confused with sampling points with uniform distances from x_t (Avgar et al., 2016). Generating

uniform distances will not result in a spatially uniform distribution of end points z_k, and so the above formulas do not hold in that case. This is due to the fact that the set of possible long steps is spread over a larger area than the set of possible short steps; therefore if distances are uniform, points will be relatively more concentrated around the origin than far from it (Rhodes et al., 2005). The case where distances are sampled from uniform distributions can in fact be viewed as a special case of importance sampling (Section 2.2.3).

2.2.3 | Importance sampling

The precision of numerical integration depends on the choice of integration points; generally, the variability in the approximation is lower if points are concentrated in areas where the function takes large values. Importance sampling is a method to increase the precision of Monte Carlo integration by generating random points from a user-defined distribution h, called the importance function (section 3.3 of Robert & Casella, 2010, and see an illustration in Figure 2c). To apply importance sampling to an SSF, we choose $f = h$ and $g = (w \times \phi) / h$ (so that $f \times g = w \times \phi$ as required), and Equation (3) gives us the following approximation for the SSF integral,

$$I \approx \frac{1}{K+1} \sum_{k=0}^{K} \frac{w(x_t, z_k)\phi(z_k | x_{1:t})}{h(z_k | x_{1:t})}, \quad \text{where } z_k \sim h(\cdot | x_{1:t}). \quad (6)$$

The only constraint on h is that it should be strictly positive over Ω; when this is not the case, the approximation of the integral is

truncated to the support of h (i.e. the geographical area over which $h > 0$). Note that, in Equation (6), the importance function h is used to weigh the contribution of each sampled point to the approximation; this is required to correct for the preferential sampling of some points over others when generating z_k.

Importance sampling is useful because the function h can be chosen in such a way that the variance of the integral estimator decreases (i.e. its precision increases). The aim is to choose a function h with a shape that is as similar as possible to g (section 6.6 of Rizzo, 2019). This is a convenient framework in the context of SSFs, because it is often possible to determine where the SSF will take large values, based on the movement constraints of the animal. Animals are likely to avoid long steps, and so the SSF often decays rapidly as distance from the start point increases (Figure 1). The speed of this decay can be determined approximately from the data, for example by fitting a distribution to the observed step lengths, and this information can be used to define the importance function h. For example, h could be chosen as a bivariate normal distribution centred on the last location x_t, or the two-dimensional spatial distribution implied by step length and turn angle distributions estimated from the data.

Using importance sampling with function h, the approximate likelihood of a step under the SSF model is

$$p(x_{t+1} | x_{1:t}) \approx \frac{w(x_t, x_{t+1})\phi(x_{t+1} | x_{1:t})}{\frac{1}{K+1}\sum_{k=0}^{K} w(x_t, z_k)\phi(z_k | x_{1:t})/h(z_k | x_{1:t})},$$

$$\propto \frac{w(x_t, x_{t+1})\phi(x_{t+1} | x_{1:t})}{\sum_{k=0}^{K} w(x_t, z_k)\phi(z_k | x_{1:t})/h(z_k | x_{1:t})}, \quad (7)$$

where $z_k \sim h$. The two other Monte Carlo approaches (Sections 2.2.1 and 2.2.2) can be viewed as special cases of importance sampling. If h is the probability density function of a uniform distribution over geographical space, it is constant and can be omitted in Equation (7), and we obtain Equation (5). Alternatively, if ϕ is assumed to be known, and we choose $h = \phi$, then h and ϕ cancel out in the denominator of Equation (7); omitting ϕ from the numerator because it is known, this simplifies to Equation (4).

Although the term 'importance sampling' has rarely been used in the SSF literature, this and equivalent methods have been widely advocated, starting with the recommendation of Forester et al. (2009)

to distinguish between the sampling function h and the movement model ϕ. Forester et al. (2009) derived a formula very similar to Equation (7), but with one small difference: their numerator is divided by $h(x_{t+1} | x_{1:t})$. This difference is inconsequential because $h(x_{t+1} | x_{1:t})$ does not depend on the estimated parameters, so excluding it does not affect inference. The widely used methods of Avgar et al. (2016), often called integrated step selection analysis, are based on Forester et al. (2009) and can also be viewed as importance sampling. They focus on cases where h and ϕ are chosen to be from the same exponential family of distributions, so that the calculations simplify and CLR can be used, but there is no such restriction when the approach is implemented with maximum likelihood estimation. In the approach of Avgar et al. (2016), the parameter estimates are adjusted after model fitting to correct for the sampling design. Here, the bias is corrected directly by including h in Equation (7) when implementing the likelihood function. Another notable example is Johnson et al. (2008), who explicitly suggested importance sampling for a weighted distribution model analogous to the SSF in Equation (1), and used maximum likelihood estimation to fit the model without the need for CLR. More recently, Klappstein et al. (2023) and Pohle et al. (2023) also recognised that the approach proposed by Forester et al. (2009) was a form of importance sampling.

Figure 3 contrasts importance sampling and uniform Monte Carlo sampling (Section 2.2.2) for an example SSF. Similar to the one-dimensional example shown in Figure 2, the intuition is that the precision of the approximation of the integral depends on the coverage of areas where the SSF takes high values. More specifically, the variance of the approximation is minimised when the distribution of integration points (i.e. the importance function h) is proportional to the SSF $w \times \phi$. In practice, w and ϕ are not known, but a good heuristic is to approximate the movement kernel with parametric distributions (which can then be sampled from), and use these to define h.

2.3 | Quadrature

Quadrature is a deterministic (non-random) alternative to Monte Carlo sampling, where the function is evaluated on a user-defined grid of points over the domain of integration. The simplest example
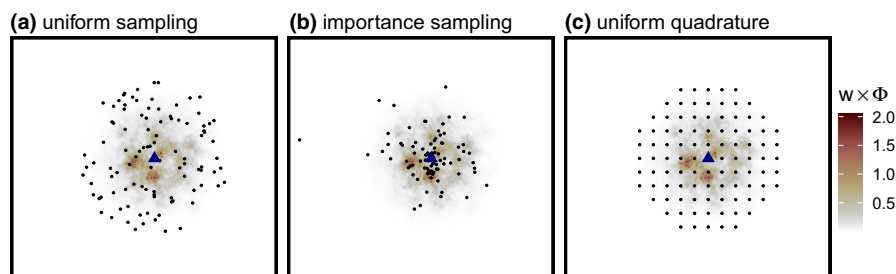

(a) uniform sampling   **(b)** importance sampling   **(c)** uniform quadrature

$w \times \Phi$
2.0
1.5
1.0
0.5

FIGURE 3 Illustration of three integration designs for an example step selection function. The triangle in the centre shows the last location x_t, and the heatmap shows the function that needs to be integrated, that is, $w(x_t, z)\phi(z | x_{1:t})$. This function decreases with distance to x_t due to the animal's movement constraints. The black dots in each panel represent 100 points used for numerical integration: (a) uniform points over a disc, (b) points generated from an importance function based on distance to x_t, and (c) regular quadrature grid over a disc. Importance sampling generates more points in areas where the function is high, and will typically lead to a better approximation of the integral.

of quadrature is the Riemann sum (see Figure 2d); in one dimension, the integral is approximated by the sum of the areas of rectangles with heights determined by function evaluations along a regular grid of points. The two-dimensional extension of this approach consists of evaluating the function at regularly spaced points along a two-dimensional grid, and approximating the integral by the sum of the volumes of cuboids.

In the context of an SSF, quadrature can be applied by evaluating $w(\mathbf{x}_t, \mathbf{z}_k)\phi(\mathbf{z}_k | \mathbf{x}_{1:t})$ at the centres of $K + 1$ grid cells, $\{\mathbf{z}_0, \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_K\}$, and calculating

$$I \approx A \sum_{k=0}^{K} w(\mathbf{x}_t, \mathbf{z}_k)\phi(\mathbf{z}_k | \mathbf{x}_{1:t}),$$

where A is the area of each grid cell. Unlike Monte Carlo approaches, there is no variance in the estimate of the integral obtained from quadrature, because the grid is fixed (for a given K) rather than random. However, there is some numerical error in the estimate (akin to statistical bias), and this error decreases as the spatial resolution of the quadrature grid increases (i.e. as K increases).

Rhodes et al. (2005) proposed this approach, and pointed out that the grid cells need not cover the entire study area, and the approximation can be truncated to the region where the function is non-negligible (within some distance of $\mathbf{x}_t$). This can speed up computations, similar to the observation in Section 2.2.2 that Monte Carlo integration with uniform sampling can be performed over a disc centred on the last location, as long as the radius of the disc is big enough. We call the approach where integration points are on a regular grid 'uniform quadrature', and it is illustrated in Figure 3c for an example SSF. The equation for approximating the integral using uniform quadrature is identical to that obtained for uniform Monte Carlo sampling, and so the expression for the SSF likelihood is given by Equation (5).

In step selection analysis, environmental covariates are often available only over the discrete cells of a raster, making the centroids of the raster cells a natural choice for the grid of quadrature (Rhodes et al., 2005), but this is not the only possible choice. Arce Guillen et al. (2023) described a new method of inference for SSFs, where the function to integrate is evaluated at the nodes of a (deterministic) triangular mesh. This is another form of quadrature, which works well for the general class of spatial point processes implemented in the `inlabru` spatial modelling software that they use (Bachl et al., 2019; Simpson et al., 2016).

Some have argued that deterministic numerical integration is preferable in habitat selection modelling because it has better properties than Monte Carlo integration for low-dimensional integrals like the one in Equation (2) (Arce Guillen et al., 2023; Warton & Shepherd, 2010). However, this comparison assumed uniform Monte Carlo sampling, for which the performance can be improved substantially using importance sampling with a well-chosen function h. An interesting alternative might therefore be adaptive quadrature, where the spatial arrangement of points in a (deterministic) spatial grid is chosen based on the shape of the function. The general idea is to iteratively subdivide the domain of integration, in such a way that regions where the function is more irregular are subdivided further. This is analogous to the idea in importance sampling of generating points in regions where the function is more complex or takes higher values (Pinheiro & Chao, 2006).

3 | ILLUSTRATION

We illustrate some of the key concepts and methods using simulations and a real data analysis. The general approach to fitting SSFs presented in Section 2 requires implementing the (approximate) likelihood function, rather than relying on existing CLR software. Writing custom code greatly increases the flexibility of the model formulation and inference methods. To help readers implement their own step selection analyses, we provide R code that can be used as a starting point. We aimed for a trade-off between simplicity and flexibility, and we provide basic functions that can be tailored to fit a wide range of model formulations. The documented code and detailed examples are provided in Appendix C.

3.1 | Comparing sampling designs in simulations

Different methods of fitting SSFs can be viewed as different numerical integration approaches for the same underlying model. When using Monte Carlo and quadrature methods, the placement of integration points is known to affect the precision of the results (Rizzo, 2019), and this provides a rigorous grounding for the intuition that some methods of generating locations perform better than others. The closer the distribution of integration points is to the true SSF, the more precise the approximation will be. In this section, we use simulations to assess the effect of several design choices on our ability to recover the parameters of an SSF. We do not intend these simulations to be exhaustive, or to provide general guidelines to select the best sampling design in every application, as the choice of method and model is study specific. Rather, our aim is to showcase the qualitative effect of different design choices, encouraging researchers to explore the various options for themselves when they perform step selection analysis.

We considered three designs for the integration points: (1) points sampled uniformly at random over a disc of radius equal to the maximum observed step length, (2) importance sampling where h was based on a gamma distribution of distances and a von Mises distribution of turning angles and (3) uniform quadrature where integration points were defined on a regular spatial grid over a disc of radius equal to the maximum observed step length. In all cases, the parameters of the importance function were chosen based on the empirical distributions of step lengths and/or turning angles.

We first simulated a movement track of length 1000 from an SSF with known parameters. For the habitat selection component, we used $w(\mathbf{x}_t, \mathbf{x}_{t+1}) = \exp(\beta \times c_1(\mathbf{x}_{t+1}))$, where c_1 was a simulated

spatial covariate shown in Appendix B and $\beta = 5$ was the corresponding selection parameter. The movement kernel was chosen as $\phi(\mathbf{x}_{t+1} | \mathbf{x}_{1:t}) = \exp(-4L_t + 3\cos(\alpha_t))$, where L_t and α_t are the step length and turning angle at time t respectively. This formulation implies that step lengths followed a gamma distribution with mean 0.5 and standard deviation 0.35, and turning angles followed a von Mises distribution with mean zero and concentration 3. For each scenario, we ran the following steps for increasing numbers of random points, $K \in \{5, 10, \ldots, 195, 200\}$:

1. Generate K integration points for each observed step, based on a given numerical integration method;
2. Estimate the selection parameter β using maximum likelihood estimation, based on the integration points from step 1.

For Monte Carlo sampling methods, we repeated these steps 50 times to capture estimator variability, yielding 50 estimated parameters $\widehat{\beta}$ for each approach and each number of random points. For deterministic quadrature, there is no sampling variability and so the procedure only needed to be run once for any given K. Note that, for quadrature, the number of integration points was constrained by the design of the grid, and so it does not always exactly match the values of K listed above. For each simulation, we evaluated estimator bias due to the numerical integration, as the difference between the estimated parameter and the asymptotic estimate obtained when K is very large (here, $K = 5000$). We did not compare the estimates to the true parameter values used in simulation, because this would not make it possible to separate bias due to integration error (which we are interested in) and finite-sample bias inherent to maximum likelihood estimation (which does not depend on integration method).

Figure 4 compares the bias in the selection parameter β for the three numerical integration approaches, over a range of numbers of random points. The bias and variance decreased as the number of random points increased for all tested methods. Overall, uniform Monte Carlo sampling had lower precision (i.e. higher variation) than importance sampling; this is to be expected, as importance sampling

uses observed movement patterns to generate integration points efficiently. Uniform quadrature was generally positively biased (i.e. the selection parameter was overestimated) and, although the bias decreased as K increased, both random sampling methods seemed to perform better in this analysis. As described in Section 2.2, the bias is due to the log-transformation of the likelihood for optimisation: the parameter value that maximises the approximate log-likelihood can be biased even if the estimator of the likelihood is unbiased (Gourieroux & Monfort, 1996).

These simulations show that the performance of any method of numerical integration is dependent on the number of integration points. In practice, there is no consensus about how high this number should be, and Thurfjell et al. (2014) reported that different studies have used a wide range of numbers of random points, between $K = 2$ and $K = 200$. The minimum number needed for a given analysis depends on many factors, such as the length of the observed times series, the sampling scheme used, and the complexity of the SSF model formulation. For this reason, we recommend that practitioners try several numbers of random points, until the parameter estimates stabilise, to ensure that the approximation error in the results is small. This is consistent with the results presented in fig. 2 of Fieberg et al. (2021), and similar to the advice of Warton and Shepherd (2010) and Northrup et al. (2013) in the context of resource selection functions. Small samples can also lead to numerical instability and failure to converge during model fitting.

3.2 | Comparing models using AIC and the same set of integration points

To demonstrate how an understanding of numerical integration techniques and their use in step selection analyses can facilitate model comparisons, we considered location data from a red deer (*Cervus elaphus*) in Northern Germany, automatically loaded with the amt R package as the data object `deer` (Signer et al., 2019). The locations are on a regular 6-hour time grid, and the package
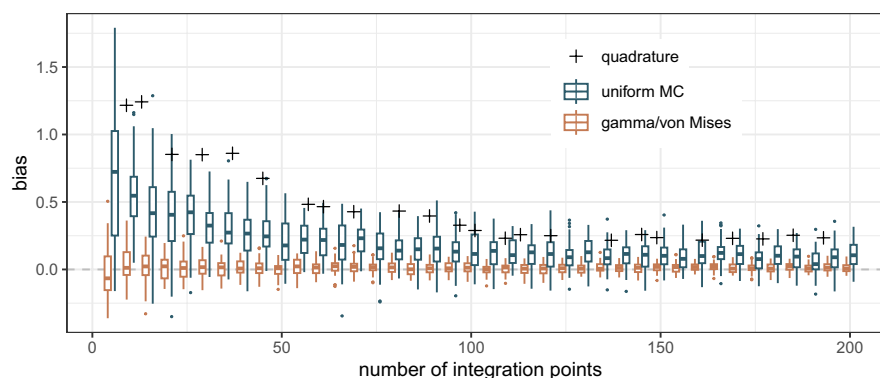

FIGURE 4 Results of simulation study. Bias (estimate − truth) in estimated selection parameter for one spatial covariate, as function of the number of integration points. Results are compared for three methods for performing numerical integration: uniform Monte Carlo (blue boxes), importance sampling with gamma-distributed distances and von Mises-distributed angles (red boxes), and uniform quadrature (black crosses). Each box represents 50 estimated values. For quadrature, the number of integration points was constrained by the design of the grid, so values of K do not exactly match those used for other methods.

also provides a binary raster layer for forest cover (through the `get_sh_forest()` function). We will use importance sampling with a single set of integration points to fit multiple models and compare them using AIC.

We generated integration points using gamma-distributed step lengths and von Mises-distributed turning angles, with parameters estimated from their empirical distributions. We compared four SSF formulations, all with the same habitat selection component w (with forest cover as covariate), but with different families of distributions of step lengths L_t and turning angles α_t in the movement kernel: (i) $L_t \sim \text{Exp}(\theta_1)$ and $\alpha_t \sim \text{uniform}(-\pi, \pi)$, (ii) $L_t \sim \text{gamma}(\theta_1, \theta_2)$ and $\alpha_t \sim \text{uniform}(-\pi, \pi)$, (iii) $L_t \sim \text{gamma}(\theta_1, \theta_2)$ and $\alpha_t \sim \text{von Mises}(0, \kappa)$ and (iv) $L_t \sim \text{Weibull}(\theta_1, \theta_2)$ and $\alpha_t \sim \text{wrapped Cauchy}(0, \kappa)$. Note that we make a distinction between the distributions used to generate random locations (which determine the importance function h) and the distributions used to specify the movement kernel ϕ. This allows us to fit all four SSFs using the same integration points, that is, the exact same data set, such that the models can be compared using AIC.

The results are shown in Table 2. For this data set, AIC favoured the SSF formulation (iv), where step lengths were modelled with a Weibull distribution and turning angles with a wrapped Cauchy distribution. This example illustrates two advantages of implementing maximum likelihood estimation using numerical integration: the flexibility of choosing non-exponential models for the movement kernel (the Weibull and wrapped Cauchy distributions are not in the exponential family), and the ability to compare models with different movement formulations.

4 | RETHINKING STEP SELECTION ANALYSIS

4.1 | Beyond conditional logistic regression

It is common for step selection analysis to be presented as CLR, where each observed step is a 'case' associated with a set of 'control' (random) steps. Indeed, in many important special cases, the SSF likelihood is approximately equivalent to that of a CLR model (Forester et al., 2009; Fortin et al., 2005), which allows parameters to be estimated with CLR software (Signer et al., 2019; Therneau, 2023). To

justify that approach, Forester et al. (2009) described an SSF as a discrete-choice model, which is a popular way to present resource selection functions (Cooper & Millspaugh, 1999). In that framework, the animal is assumed to have access to a finite number of discrete and mutually exclusive resource units, and the model describes how it chooses one unit over the others. In a step selection model, however, the animal has infinitely many movement options (because it moves over a continuous space). To reduce the problem to a discrete choice, Forester et al. (2009) conditioned each movement decision on a set of random ('control') points, and assumed that those represented the animal's options for that time step. This can be viewed as an approximation of the target model, akin to the numerical integration approaches that we presented in Section 2, and the two approaches lead to equivalent formulas.

The equivalence between CLR and the SSF likelihood has allowed ecologists to leverage standard statistical software for quantifying drivers of movement and habitat selection for many years. Recently, Muff et al. (2020) and Chatterjee et al. (2023) demonstrated how a similar equivalence between CLR and Poisson regression with fixed stratum-specific intercepts could be exploited to model individual variability in habitat selection and movement parameters using random effects. Despite these equivalencies, we emphasise that CLR is not the model of interest, but rather a tool to fit the target SSF model, shown in Equation (1). Notably, the equivalence between CLR and the SSF likelihood only holds when both the habitat selection function w and the movement kernel ϕ have an exponential form. CLR is therefore limited to distributions from the exponential family when modelling ϕ, and most step selection modelling research has therefore defaulted to using an exponential or gamma distribution for step lengths and the von Mises distribution for turn angles (Avgar et al., 2016; Duchesne et al., 2015). There is no such restriction in the general model, however, which allows for the use of a much wider range of distributions, such as the Weibull distribution for step length and the wrapped Cauchy distribution for turning angle. As our applied example illustrates (Section 3.2), these alternatives can potentially lead to improved model fit. Likewise, it is possible to model habitat selection using functions other than the exponential (e.g. Potts et al., 2014; Schlägel et al., 2017), although it is a natural choice for continuous variables since selection functions should be positive and unbounded (McDonald, 2013).

4.2 | Separating model and inference

As we have already mentioned, it is important to distinguish between the choice of model formulation and the method of parameter estimation. Here, the model is of the form shown in Equation (1), and the main modelling decision is defining the functional forms of w and ϕ. For a given model formulation, many possible numerical integration methods can be used to approximate the likelihood, and therefore to estimate the model parameters, as described in Section 2. These methods, together with any implementation scheme such as CLR or other likelihood maximisation, constitute the inference procedure,

TABLE 2 Model comparison for deer analysis.

	Step length	Turning angle	AIC	ΔAIC
(iv)	Weibull	Wrapped Cauchy	2500	0
(ii)	Gamma	Uniform	2507	7
(iii)	Gamma	Von Mises	2509	9
(i)	Exponential	Uniform	2521	21

Note: The four SSFs are specified using different distributions of step lengths and turning angles, and they are shown in order of increasing AIC (i.e. the better model is at the top). ΔAIC is the difference in AIC between each model and the better model.

and are not the model itself. Choices of model and inference procedure are both important for the analysis: the model formulation should capture important features of the data-generating process (i.e. animal movement and habitat selection), while the inference procedure should be chosen carefully to reduce the approximation error. We posit that much confusion has arisen in the context of step selection analysis due to the conflation of model and inference, especially when comparing techniques. Figure 5 summarises the distinction between the choice of model formulation and the choice of a method of inference in step selection analysis.

One particular area of confusion has been the interpretation of the integration points $\{z_1, \ldots, z_K\}$ generated as part of model fitting. Due to the historical influence of Fortin et al. (2005), who assumed that the movement model ϕ was known and used it to generate random locations (Section 2.2.1), the z_k are commonly assumed to be a sample of 'available' points (or, equivalently, steps connecting the previous location and z_k are seen as possible movements). However, this is crucially not the correct interpretation in other approaches, such as uniform sampling (Section 2.2.2) and importance sampling (Section 2.2.3). Indeed, with those methods, the sample of random locations does not have any particular biological interpretation, and it merely constitutes a numerical tool to approximate an integral over space. This is the case in most modern step selection analyses, in which availability is estimated jointly with habitat selection through the parameters of the function ϕ, rather than assumed known a priori (Avgar et al., 2016; Forester et al., 2009; Rhodes et al., 2005).

Another important distinction is between the distributions used to model movement, and the distributions used to generate random locations. The modelled distributions are specified through the choice of ϕ, independently of the choice of sampling function (e.g. h in importance sampling). Forester et al. (2009) and Avgar et al. (2016) showed that it is mathematically convenient to generate

random locations from the same family of distributions that is used in ϕ, but this is not a necessity. As we demonstrate in Section 3.2, it is for example possible to use a gamma distribution of distances to generate random locations (i.e. in the importance function h), but specify ϕ so as to model step lengths with a Weibull distribution in the analysis. Furthermore, it is straightforward to fit SSFs with different movement kernel formulations on the exact same data set (including identical random locations). This makes it possible to select the formulation for the movement kernel based on standard model selection criteria such as AIC. This is not possible within the workflow outlined by Avgar et al. (2016) and implemented by Signer et al. (2019), which requires h and ϕ to be from the same family. When the model and method of inference are separate, it is also easier to determine how different functional forms for the movement kernel ϕ can be implemented in practice, and we discuss this in Appendix D.

In fact, the distribution from which the random locations are generated only matters insofar as different distributions might require different numbers of integration points to achieve low numerical error (Section 3.1). For a large enough number of integration points, the choice of distribution is inconsequential. In practice, when the computational cost is high, it might not be an option to increase the size K of the random sample arbitrarily; in this case, it is preferable to choose an importance function h that reduces the estimation variance for a given K. Importantly, the choice of h does not reflect a modelling assumption, and it is used merely as an inferential tool to reduce the approximation error. Another implication is that, for a given model formulation, the interpretation of the movement and habitat selection parameters β does not depend on the model fitting approach. Indeed, different numerical integration methods merely lead to different approximations of the same parameters. Note that the Monte Carlo approach described in Section 2.2.1 (where movement is fixed a priori) is slightly different: in that case, the choice of
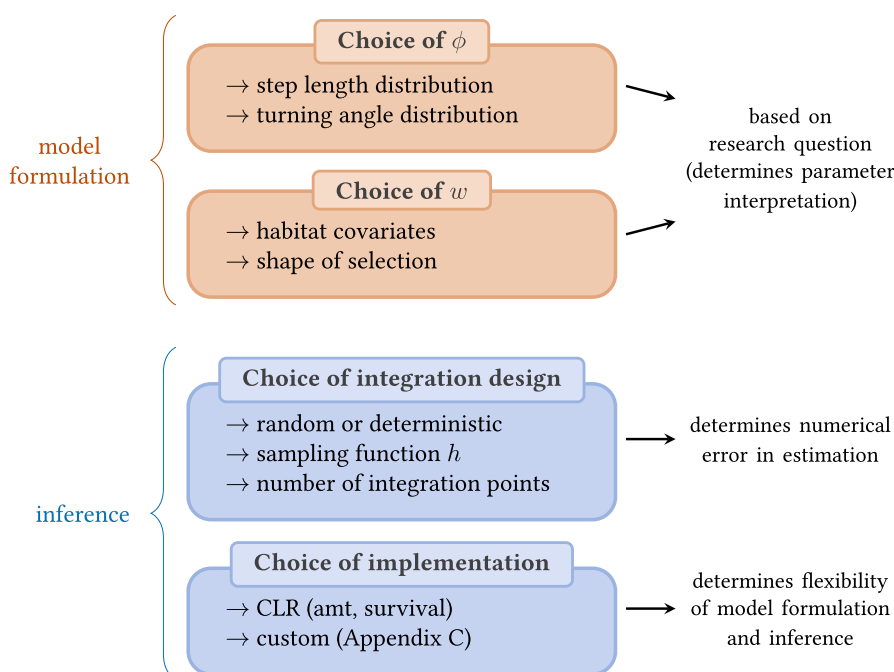

FIGURE 5 Summary of modelling and inference choices for step selection analysis.

integration points reflects a modelling decision (i.e. the form of the movement kernel ϕ), and has a direct impact on the interpretation of the parameters (as discussed by Beumer et al., 2023). This requires the additional assumption that movement is not affected by habitat selection, and the habitat selection parameters obtained from that method therefore only have the same interpretation as the output of other approaches if this assumption holds.

5 | CONCLUSION

The general problem of evaluating complex integrals has been studied extensively, and many different approaches could be adapted in the context of step selection analysis. For example, the Laplace approximation is a versatile method where it is not necessary to evaluate the function at integration points. Instead, it is assumed that the integrand is well approximated by a normal distribution, for which the integral is known. In step selection analyses, the integrand typically combines the movement kernel and the habitat selection function, resulting in a complex (possibly multimodal) function, and it is not clear whether the Gaussian assumption would be reasonable. As suggested in Section 2.3, another promising direction is to combine ideas from quadrature and importance sampling, and choose the nodes of a deterministic grid to improve the approximation in areas where the SSF is irregular. This is similar to adaptive quadrature, which is already widely used in ecology for non-Gaussian generalised linear mixed models (Bolker et al., 2009).

All model fitting approaches are approximately equivalent in the limit where the number of points in the Monte Carlo sample or the quadrature grid is large (i.e. as $K \to \infty$). It might therefore seem unnecessary to concern ourselves with the design choices described in this review, because we can always increase K until the bias and/or variance in the estimation are negligible. Because SSFs are relatively simple models, computation time is often small for modest-sized data sets, and the additional cost of increasing K might be moderate. However, we have shown that using simplistic numerical integration techniques can cause bias to persist even for quite large K (e.g. uniform Monte Carlo and quadrature; Figure 4), so thinking carefully about the numerical integration technique employed may often be preferable to simply increasing K. Furthermore, step selection models are becoming more sophisticated and complex, so computational effort might become the bottleneck of many analyses (e.g. multistate models; Nicosia et al., 2017). We anticipate that the sampling design will become increasingly critical in those cases, as well as in studies with large data sets.

We are certainly not recommending that all biologists stop using CLR software for step selection analyses, as it is a convenient, fast and stable implementation for many purposes. In cases where the habitat selection and movement are modelled in exponential form, the CLR approach outlined by Forester et al. (2009), Duchesne et al. (2015) and Avgar et al. (2016) can safely be applied, for example, with the amt R package (Signer et al., 2019). However, even in that context, understanding the role numerical integration plays in parameter estimation can shed light on several important technical details. In particular, this approach clarifies the role of 'control' points in step selection analysis: although those locations are generally a numerical tool rather than a biologically relevant set of spatial locations, placing them in areas where the animal is most likely to move decreases the required number of points. In addition, importance sampling explains the post-hoc parameter adjustment proposed by Avgar et al. (2016), which is needed to correct for the choice of importance function (see also Appendix S3 of that paper). Overall, we have shown that the interpretation of step selection models ought to be separated from any specific model fitting approach, and that understanding this dichotomy between model construction and model fitting leads to much broader application of SSFs than those confined to a CLR approach.

ORCID

Théo Michelot https://orcid.org/0000-0002-3838-4113
Natasha J. Klappstein https://orcid.org/0000-0001-6527-9629
Jonathan R. Potts https://orcid.org/0000-0002-8564-2904
John Fieberg https://orcid.org/0000-0002-3180-7021

REFERENCES

Arce Guillen, R., Lindgren, F., Muff, S., Glass, T. W., Breed, G. A., & Schlägel, U. E. (2023). Accounting for unobserved spatial variation in step selection analyses of animal movement via spatial random effects. *Methods in Ecology and Evolution*, 14, 2639–2653. https://doi.org/10.1111/2041-210X.14208

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(5), 619–630.

Bachl, F. E., Lindgren, F., Borchers, D. L., & Illian, J. B. (2019). inlabru: An R package for bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, 10, 760–766.

Beumer, L. T., Schmidt, N. M., Pohle, J., Signer, J., Chimienti, M., Desforges, J.-P., Hansen, L. H., Højlund Pedersen, S., Rudd, D. A., Stelvig, M., & van Beest, F. M. (2023). Accounting for behaviour in fine-scale habitat selection: A case study highlighting methodological intricacies. *The Journal of Animal Ecology*, 92, 1937–1953.

Bolker, B. M. (2008). *Ecological models and data in R*. Princeton University Press.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127–135.

Boyce, M. S., Mao, J. S., Merrill, E. H., Fortin, D., Turner, M. G., Fryxell, J., & Turchin, P. (2003). Scale and heterogeneity in habitat selection by elk in Yellowstone National Park. *Ecoscience*, 10(4), 421–431.

Chatterjee, N., Wolfson, D., Kim, D., Gomez, J. V., Freeman, S., Bacheler, N. M., Shertzer, K., Taylor, C., & Fieberg, J. (2023). Modeling individual variability in habitat selection and movement using integrated step-selection analyses. *bioRxiv*, 2023-07. https://doi.org/10.1101/2023.07.31.551213

Cooper, A. B., & Millspaugh, J. J. (1999). The application of discrete choice models to wildlife resource selection studies. *Ecology*, 80(2), 566–575.

Craiu, R. V., Duchesne, T., & Fortin, D. (2008). Inference methods for the conditional logistic regression model with longitudinal data. *Biometrical Journal*, 50(1), 97–109.

Duchesne, T., Fortin, D., & Rivest, L.-P. (2015). Equivalence between step selection functions and biased correlated random walks for statistical inference on animal movement. *PLoS ONE*, 10(4), e0122947.

Fieberg, J. (2023). *Statistics for ecologists: A frequentist and Bayesian treatment of modern regression models*. An open-source online textbook. https://statistics4ecologists-v1.netlify.app/

Fieberg, J., Signer, J., Smith, B., & Avgar, T. (2021). A 'How to' guide for interpreting parameters in habitat-selection analyses. *Journal of Animal Ecology*, 90(5), 1027–1043.

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(12), 3554–3565.

Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T., & Mao, J. S. (2005). Wolves influence elk movements: Behavior shapes a trophic cascade in Yellowstone National Park. *Ecology*, 86(5), 1320–1330.

Gourieroux, C., & Monfort, A. (1996). *Simulation-based econometric methods*. Oxford University Press.

Johnson, D. S., Thomas, D. L., Ver Hoef, J. M., & Christ, A. (2008). A general framework for the analysis of animal resource selection from telemetry data. *Biometrics*, 64(3), 968–976.

Klappstein, N. J., Potts, J. R., Michelot, T., Börger, L., Pilfold, N. W., Lewis, M. A., & Derocher, A. E. (2022). Energy-based step selection analysis: Modelling the energetic drivers of animal movement and habitat use. *Journal of Animal Ecology*, 91(5), 946–957.

Klappstein, N. J., Thomas, L., & Michelot, T. (2023). Flexible hidden Markov models for behaviour-dependent habitat selection. *Movement Ecology*, 11(1), 30.

Matthiopolous, J., Fieberg, J., & Aarts, G. (2023). *Species-habitat associations: Spatial data, predictive models, and ecological insights* (2nd ed.). University of Minnesota Libraries Publishing, University of Minnesota.

McDonald, T. L. (2013). The point process use-availability or presence-only likelihood and comments on analysis. *Journal of Animal Ecology*, 82(6), 1174–1182.

Muff, S., Signer, J., & Fieberg, J. (2020). Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using bayesian or frequentist computation. *Journal of Animal Ecology*, 89(1), 80–92.

Nicosia, A., Duchesne, T., Rivest, L.-P., & Fortin, D. (2017). A multi-state conditional logistic regression model for the analysis of animal movement. *The Annals of Applied Statistics*, 11(3), 1537–1560.

Northrup, J. M., Hooten, M. B., Anderson, C. R., & Wittemyer, G. (2013). Practical guidance on characterizing availability in resource selection functions under a use–availability design. *Ecology*, 94(7), 1456–1463.

Pinheiro, J. C., & Chao, E. C. (2006). Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 15(1), 58–81.

Pohle, J., Signer, J., Eccard, J. A., Dammhahn, M., & Schlägel, U. E. (2023). How to account for behavioural states in step-selection analysis: a model comparison. *arXiv preprint arXiv:2304.12964*.

Potts, J. R., Bastille-Rousseau, G., Murray, D. L., Schaefer, J. A., & Lewis, M. A. (2014). Predicting local and non-local effects of resources on animal space use using a mechanistic step selection model. *Methods in Ecology and Evolution*, 5(3), 253–262.

Rhodes, J. R., McAlpine, C. A., Lunney, D., & Possingham, H. P. (2005). A spatially explicit habitat selection model incorporating home range behavior. *Ecology*, 86(5), 1199–1205.

Rizzo, M. L. (2019). *Statistical computing with R*. CRC Press.

Robert, C. P., & Casella, G. (2010). *Introducing Monte Carlo methods with R* (Vol. 18). Springer.

Schlägel, U. E., & Lewis, M. A. (2014). Detecting effects of spatial memory and dynamic information on animal movement decisions. *Methods in Ecology and Evolution*, 5(11), 1236–1246.

Schlägel, U. E., Merrill, E. H., & Lewis, M. A. (2017). Territory surveillance and prey management: Wolves keep track of space and time. *Ecology and Evolution*, 7(20), 8388–8405.

Signer, J., Fieberg, J., & Avgar, T. (2019). Animal movement tools (amt): R package for managing tracking data and conducting habitat selection analyses. *Ecology and Evolution*, 9(2), 880–890.

Simpson, D., Illian, J. B., Lindgren, F., Sørbye, S. H., & Rue, H. (2016). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *Biometrika*, 103(1), 49–70.

Therneau, T. M. (2023). *A package for survival analysis in R*. R package version 3.5-5.

Thurfjell, H., Ciuti, S., & Boyce, M. S. (2014). Applications of step-selection functions in ecology and conservation. *Movement Ecology*, 2, 1–12.

Warton, D. I., & Shepherd, L. C. (2010). Poisson point process models solve the 'pseudo-absence problem' for presence-only data in ecology. *The Annals of Applied Statistics*, 4, 1383–1402.

SUPPORTING INFORMATION

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

Appendix A: Including the observed location as an integration point.

Appendix B: Simulated covariate.

Appendix C: Implementing step selection analysis.

Appendix D: Implementation of different movement distributions.

Data S1: Code for analyses of Section 3 and Appendix A.

How to cite this article: Michelot, T., Klappstein, N. J., Potts, J. R., & Fieberg, J. (2023). Understanding step selection analysis through numerical integration. *Methods in Ecology and Evolution*, 00, 1–12. https://doi.org/10.1111/2041-210X.14248