12 December 2025

A plain language review and guidance for modeling animal habitat-selection

Brian D. Gerber, U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Biology, Colorado State University, 1484 Campus Delivery, Fort Collins, CO, 80523 USA.
Casey M. Setash, Colorado Parks and Wildlife, Fort Collins, CO, 80526 USA.
Jacob S. Ivan, Colorado Parks and Wildlife, Fort Collins, CO, 80526 USA.
Joseph M. Northrup, Wildlife Research and Monitoring Section, Ontario Ministry of Natural Resources and Environmental and Life Sciences Graduate Program, Trent University, Peterborough, ON, Canada.

Correspondence: Brian D. Gerber, U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Biology, Colorado State University, 1484 Campus Delivery, Fort Collins, CO, 80523 USA. Email: brian.gerber@colostate.edu

ABSTRACT Animal habitat selection is the process of how individual organisms disproportionately use habitat compared to what is available to them. Understanding habitat selection is important for the study of ecology and conservation. However, learning the foundations of making inference or prediction on animal habitat selection can be quite challenging. Foremost, the literature is large and highly technical. We summarize important considerations in getting the basics right, pointing to key papers in the modern literature. We also demonstrate many of these considerations in an online vignette and associated code. We hope this work will help jumpstart student and practitioner learning about habitat selection modeling, provide guidance when reviewing analyses, and lead to rigorous ecological studies that help guide the management and conservation of animal populations.

KEYWORDS: habitat selection, logistic regression, random effect, resource selection function, statistical model, step selection function.

Habitat selection is a central tenet of ecology and a major focus in studies of animal ecology and evolution. The results of these studies provide inference to the behavioral process of how animals choose locations to maximize fitness when confronted with variation in habitat quality (McLoughlin et al. 2010). The corresponding spatial predictions (Morris et al. 2016a; Winter et al. 2024) are equally important as they are often used in managing animal populations (Morris et al. 2016b, Buderman et al. 2023, Colorado Parks and Wildlife, 2024) and making land conservation decisions to protect habitat (Guisan et al. 2013; Dellinger et al. 2020). How individual animals interact with their environment leads to population-level patterns of distribution and abundance (Boyce and McDonald 1999, Matthiopoulos et al. 2015, Boyce et al. 2016), which are often the scales at which management policy and decisions are targeted

(Williams et al. 2002). As such, the study of habitat selection is fundamental to understanding wild animals and to their conservation and management.

The process by which habitat selection is studied varies but is commonly done using a use-availability design, where features at animal locations are contrasted with features at locations deemed available to the animal for selection (Manly et al. 2002, Johnson et al. 2006, Lele et al. 2013). Animal location observations can be collected many ways (e.g., aerial or ground surveys) but are commonly obtained from radio/GPS/cellular/satellite telemetry technologies. These technologies have advanced considerably in the last several decades (Kays et al. 2015) and similarly, so have the conceptual and statistical frameworks to make inference on and prediction about animal habitat selection (Matthiopoulos et al. 2023). Positive outcomes of these combined advances are finer-scale animal location data, multiple sources of data on the same individual or population, and more flexibility in fitting diverse statistical models when studying animal-habitat relationships (Tucker et al. 2018, Northrup et al. 2022). However, a less desirable outcome is confusion by practitioners and students learning about or implementing habitat selection modeling because more data, while helpful in many respects, can create modeling challenges that were absent from these approaches in the past; thus, highly complex statistical approaches have been introduced to address these issues and they are described in often highly technical literature.

The result of the many advances described above is that there is a growing divide between those developing methods for understanding habitat selection and those that most need to use these methods. Our aim is to provide a relatively simple plain language introduction to and review of the literature on the study of animal habitat selection and guidance on important considerations when fitting statistical models to make inference on animal-habitat selection using a use-availability design. We do this by reviewing the modern literature, stating concepts and guidance plainly, and pointing readers to more technical publications on specific topics. **We highlight notable points by bold font**. Our motivation stems from working with students and practitioners learning and applying these methods and finding it difficult to provide material for a simple introduction to the topic that focuses on getting the basics right. Further, getting the basics right is important for all studies because habitat selection results can have consequences for habitat conservation and management, as well as government policy decisions.

Our challenge is twofold. The first is that this review cannot be exhaustive because of the breadth of the material. We have chosen topics that we see as important to highlight or often confusing when learning. Throughout, we provide references for readers to dive deeper into topics not covered or not covered exhaustively. We have also chosen to focus more on recent literature, after the underlying statistical foundations relevant to habitat selection modeling with presence-only data have been reconciled (Warton and and Sheperd 2010; Aarts et al. 2012; Fithian and Hastie 2013; Johnson et al. 2013), and generally preferencing articles with a statistical focus over applications. The focal cited literature is thorough on the topics, aimed to guide learners, but not exhaustive, given the long history of the topic. For more extensive reviews and establishment of the history of habitat selection analyses, please see Northrup et al. (2022) and Matthiopoulos et al. (2023). Second, there is an inherent risk when writing without technical detail (i.e., statistical notation) that the lack of specificity will lead to misinterpretation. Conversely, there is also a risk that highly technical material will lose early learners or practitioners and keep them from following recommended practices. We hope by writing plainly, without technical detail, that we will reach practitioners and students near the beginning of their learning about habitat selection modeling. However, this document may also be a good review for practitioners implementing these models and reviewing manuscripts. Our intention is to communicate fundamental aspects of habitat selection and habitat selection modeling that will encourage the uptake in following recommended guidance from the modern literature.

With the goal of providing a straightforward and digestible introduction to this topic, we include quotes from publications that explain fundamental concepts to draw attention to the publications readers can reference for further reading. We define fundamental terminology in sections 1 through 5, provide guidance in sections 6 through 15, each with one or more highlighted points in bold (summarized in Box 1), and provide recommended literature for further learning in Table 1. We also offer a more detailed checklist with relevant questions for thinking through a habitat selection analysis in Box S1, available in the Supporting Information. In addition, we provide code in the R programming language (v4.4.2; R Core Team, 2024; available on GitHub: https://github.com/bgerber123/Habitat-Selection-Guidance/ and GitLab https://code.usgs.gov/cooperativeresearchunits/colorado/habitat-selection, archived on Zenodo: https://doi.org/10.5281/zenodo.14847204, and viewable online at: https://bgerber123.github.io/hsfguide/) that follows the outline of the manuscript sections and implements much of the points we discuss. We hope readers will use this work as a beginning step and ultimately go beyond to discover the wealth of information written on modeling animal-habitat relationships.

1. What is habitat?

The term habitat has been defined variably throughout its history in the ecological literature. Some authors define it simply as the conditions and resources leading to the occupancy of a species (Hall et al. 1997), while others define it in relation to variation in density of individuals

(Morris 2023). We define habitat below to maximize flexibility and interpretability while staying true to aspects of the foundational literature on the topic. As such, we do not take a hardline stance on any one strict definition but allow for users to apply their own definition within this flexible framework and encourage explicitly including that definition in one's own work.

We define habitat as "a set of conditions, resources and risks for the species of interest." (Northrup et al. 2022; see also Matthiopoulos et al. 2023 section 1.2; Krausman and Morrison, 2016). While colloquially habitat is often equated with land cover type, e.g., "forest habitat", animal habitat encompasses much more than just cover (Krausman and Morrison, 2016), and in the above definition, includes any biotic or abiotic environmental factors that influence the fitness of the animal. Conditions are environmental variables influencing the functioning of the species (e.g., salinity, temperature, diel period, social context). Resources are a substance, object or place required by the species to carry out its life history and positively impact fitness; importantly, resources are limited and can be reduced by the species (e.g., forage or prey). Risks are those environmental variables that are negatively related to a species' fitness (e.g., predator abundance or proximity to anthropogenic features). See Matthiopoulos et al. (2023; section 1.2) and Northrup et al. (2022; Table 1) for additional details.

2. What is habitat selection?

From Matthiopoulos et al. (2023), habitat selection is "the process whereby an organism uses a habitat disproportionately more (or less) than that habitat's availability."

From Northrup et al. (2022), "The process through which individual animals differentially use habitat relative to their availability at a given population density".

The word "selection" is important to note here because it emphasizes the behavioral and availability processes resulting in use or disuse. Habitat use is conceptually not the same as habitat selection. From Beyer et al. 2010, "Habitat use is the proportion of their time that animals spend in a particular habitat". Habitat selection is differentiated from habitat use by considering what is available to the animal. While conceptually different, in practice it is common to assume constant availability within a large geographic area (see section *8. Traditional habitat selection function (HSF) or step-selection function (SSF)?)*, such that habitat selection is proportional to habitat use and directly equivalent to relative habitat use.

3. What is a habitat-selection function?

From Fieberg et al. 2021, "Habitat-selection functions (HSFs; also referred to as 'resource-selection functions'; Boyce and McDonald, 1999) provide a framework for linking locations of individual animals to important features of their environment (i.e., resources, risks and environmental conditions)." Conceptually, the HSF is a quantitative description of how the animal selects habitat relative to its availability (i.e., which components of the habitat and how strong that selection is relative to other components). It is conceived as the mathematical function that translates available habitat into used habitat. **A HSF does not refer to a statistical model but is one component of an underlying statistical model** (see Hooten et al. 2017 Chapter 4 and Michelot et al. 2024, paragraph 1 notes discrepancies in language). Further, while not consistent in the literature, a HSF is also not equivalent to a habitat selection analysis, which generally refers to the whole process of analyzing data (e.g., structuring the response variable and covariates, fitting and evaluating the model, making inference or prediction from the model).

One of the most common approaches focuses on comparing *used* animal locations (along with associated environment covariates at those locations) to a set of locations and their associated environmental covariates that are considered *available* within a defined area (e.g., species' range, home range, the next "step" in a movement path); the used locations are part of the larger available sample. The specific types of statistical models used to estimate HSFs are discussed in section 9. *The data generating model and the model being fit*.

4. 'Habitat selection function' or 'resource selection function'?

The literature interchangeably uses habitat (Northrup et al. 2022; Matthiopoulos et al. 2023) and resource selection function (RSF; Manly et al. 2002) to describe the same functional analysis. Throughout, we choose to use habitat selection function for the reasons articulated by Matthiopoulos et al. (2011): "Although widely used, the term "resource selection" is perhaps inappropriate since the dimensions of environmental space can be nondepletable conditions (e.g., temperature) as well as resources (e.g., forage) and because organisms select combinations of values of environmental variables rather than single resources or conditions (e.g., it makes little sense to say that a particular species selects temperature. It is more likely that it selects a particular range of temperatures combined with types of vegetation, and ranges of moisture, slope, etc.)." Essentially, the joint combination we consider within this framework of resources, risks, and conditions leads to a combined inference on what is habitat, which is more than depletable resources. The shift in language also helps differentiate the potential confusion between a resource unit (now as a habitat unit; e.g., grid cell of a raster) and a resource type (e.g., soil type; Lele et al. 2013). Ultimately, the choice of name is a semantic argument. We find it convenient to think about what comes out of these selection functions as learning about habitat and what goes into the selection function as more than resources, thus our use of HSF rather than RSF. **In writing and presentations, consider how you are defining terminology and using terms (e.g., HSF/RSF, analysis, function, model, selection, resource, etc.), aiming for consistency and clarity.**

5. What about scale?

The scale at which practitioners examine habitat selection and the terminology surrounding it are often points of confusion. When we discuss spatial scale, we are typically referring to the spatial extent and grain of a study, defined as "the overall area encompassed by a study" and "the size of the individual units of observation," respectively (Wiens, 1989; see also McGarigal et al. 2016). Resolution, the cell size at which environmental covariates are quantified in a raster, is also an aspect of scale, but typically is under less control by the user so not discussed further here. The extent and grain of the study fundamentally determine the extent and grain of the resulting inference. This is discussed through the orders of selection defined in Johnson's (1980) seminal paper; these orders are used frequently throughout the literature to denote the scale at which selection is occurring. In summary, selection can refer to habitat a species or population selects from a large landscape (first order), selection of home ranges from within a population or species' range (second order), selection of habitat from within an individual's home range (third order, most common in the literature), or selection of "food items from those available at the site" that has been generalized to "micro-movement and behavior (e.g., acquisition of food, mating, nest building)" (Hooten et al. 2017) within portions of a home range (fourth order). Notably, individual animals within a study may or may not have the same set of available locations, depending on the scale of interest. For example, the nest locations of mallards (*Anas platyrhynchos*) might be considered used locations and available locations might be sampled from the area typically considered the size of a breeding mallard's home range (e.g. Dyson et al. 2019, Setash et al. 2024; third order selection). In this case, each individual would have different sets of available locations. Alternatively, GPS locations from several mule deer (*Odocoileus hemionus*) might be considered used while available locations are sampled from the national forest they inhabit to evaluate home range characteristics (e.g., Peterson et al. 2022). In this case, the available locations would be the same across individuals. This example does not fall exactly into the orders of selection described above (between 2nd and 3rd order), as it does not consider selection of the whole home range within the landscape (2nd order) nor selection of locations within just the home range (3rd order), but selection of locations from the landscape and not the home range. This serves as an important reminder that Johnson's (1980) orders are a good conceptual framework, but not exhaustive ("it is no doubt possible to divide these selection orders more finely", Johnson 1980).

The temporal scale of habitat selection is equally important to consider (Bischof et al 2024; Dejeante et al. 2024). Individuals may change selection through time, across diel, seasonal, or annual periods (Borowik et al. 2024; Forrest et al. 2025; Ganoe et al. 2025). It can be important to consider whether the environment is changing through time and to capture these changes by including dynamic spatial layers (e.g., anthropogenic activity that varies by season or diel time) or discretizing the analysis into different temporal periods. However, a temporally dynamic habitat selection analysis will depend on the temporal extent of sampling and whether there are observed animal locations across different time periods. Ultimately, choosing the spatial and temporal scale of inference should be based on the research goal.

6. Considering objectives and data collection

In the process of studying habitat, it is important to remember that we will never have all the spatial quantities that make up habitat, and, in reality, the set of conditions, resources and risks that maximizes fitness for an individual animal is dynamic. Thus, our goal is one of approximation, as in any modeling exercise. How we go about approximating ecological and behavioral processes depends on what we know of the species, how the species was sampled, and the intended objective of the study. We can consider 'what we know of the species' by making sure to include fundamentally important variables that have clear support from studies of the same species in the same or different populations. We can consider 'how the species was sampled' by recognizing the individual as our main sampling unit and our locations as sub-samples that are potentially correlated. We can consider the 'intended objective of the study' by deciding if the goal is inference, prediction, description, or some mixture (Tredennick et al. 2021). For example, if our goal was strictly prediction, we may consider a high number of correlated covariates that improves predictive performance but may jeopardize inference (Gerber and Northrup, 2020). We can also focus on deriving spatial layers relevant for specific objectives that motivated the study, such as creating maps of selection or avoidance relative to development when we are interested in anthropogenic impacts. **It is each practitioner's responsibility to think deeply about the species, system, sampling design, and objectives ahead of data collection and acknowledge the limitations of the study in terms of its habitat selection related objectives.** Some specific examples of questions to consider before collecting data include: How many individuals of each group of interest (e.g., age/sex class) is it necessary to sample to adequately assess habitat selection at a given degree of landscape heterogeneity? At what scale(s) should available habitat be measured for a given species or question and can those data be collected via remotely-sensed imagery or do they need to be collected in the field? What GPS fix rate is necessary to evaluate habitat selection (e.g., every minute, once a day)? Are the spatial variables at the spatio-temporal scale that is hypothesized to affect habitat selection? What spatial variables are recognized as important but are not being considered because they are highly correlated with other variables? All these choices have implications for how we sample individuals and model the data and emphasize the importance of understanding the anticipated model structure being used before collecting data.

7. What is the scientific goal of implementing a habitat-selection function?

The goal of using most statistical models is either inference ("to evaluate the strength of evidence in a data set for some statement about nature"; Tredennick et al. 2021) or prediction (estimating new response variable values for new observations of the covariates; Tredennick et al. 2021). The inferential goal of a habitat selection analysis is likely to evaluate the different magnitudes of effects (i.e., estimated coefficients, often denoted as β; see Fierberg et al. 2021 for further discussion of how to interpret and compare coefficients) from hypothesized spatial variables of risks, resources, and conditions. In other words, a statistical model is fit and estimates of coefficients/effects for each spatial variable are evaluated in terms of whether they are large, small, or near zero and whether they lead to selection less than available (negative coefficients) or selection more than available (positive coefficients). Subsequently, this information is used to make inference about the habitat selection process itself, or some other, related ecological or evolutionary process. The prediction goal of a habitat selection analysis is likely to apply the fitted model to a raster of variables across a landscape of interest to produce a spatial map depicting where selection or avoidance is expected to occur across that landscape

(Morris et al. 2016b; Gerber and Northrup, 2020) or, depending on the approach and assumptions, an approximation of the utilization distribution (Potts and Schlägel, 2020; Signer et al. 2017; Signer et al. 2024). It is important to recognize that there are tradeoffs when using models for inference versus prediction (Shmueli 2010; Fieberg and Johnson 2015; Tredennick et al. 2021); this decision can impact how complex a model is, which variables are included, and how many models are considered.

In addition to the above, HSFs can also be used to make inference to broader population-level processes. Specifically, habitat selection is a mechanism that drives species distribution and population abundance. But, can we use habitat selection analyses to predict abundance? From Street et al. 2017, predicting density "relies on the notion that high-quality habitats will contain more individuals than low-quality habitats (Fretwell and Lucas 1970), and that the resulting distribution is predictable." Namely, that animals will follow an ideal free distribution (see Northrup et al. 2022, section 'How do selection functions provide inference to habitat selection?') and we have an appropriate approximating model that includes variables that can differentiate areas of low and high abundance. There is reasonable evidence suggesting that predicting population density is possible using habitat selection analyses (Boyce, 2006; Boyce and McDonald, 1999; VanDerWal et al. 2009; Boyce et al. 2016; Weber et al. 2017; Matthiopoulos et al 2019). Some consideration from Boyce et al. 2016 is "Scale and domain of the sampling frame, both in time and space, are crucial considerations limiting application of these models. Ultimately, identifying reliable models for predicting abundance from habitat data requires an understanding of the mechanisms underlying population regulation and limitation." **In other words, it is important to consider the regulating and/or limiting mechanisms of abundance in the population of interest, and the relationship between the sampling scale**

and the scale of the underlying mechanisms. Lastly, beyond methodological challenges in predicting abundance, a species' behavioral plasticity may dampen or even eliminate demographic effects (e.g. on survival, reproduction), such that there is no impact on population density, despite high variation in habitat selection (Northrup et al. 2021). **As such, leaping from habitat selection to inferring variation in population density will be most supported with validation.** For example, comparing independent estimates of population density to habitat selection predictions.

8. Traditional habitat selection function (HSF) or step-selection function (SSF)?

A traditional HSF involves fitting a model to a collection of animal location data, usually at a broad scale (e.g., first through third order), without explicit consideration of the times at which the location data were collected. Availability data are sampled from within the entire spatial extent of interest. A step-selection function (SSF or step selection analysis (SSA)) involves fitting a model to location data in such a way as to evaluate selection relative to what is available at each step along a movement path (Forester et al. 2009); the temporal order of the location guides the data setup. Availability data are sampled only from within an area deemed to be available to an animal over the interval between consecutive locations (e.g., where did the animal choose to go next given where they were and the habitat available for them to select for the next step, which is constrained by their movements; Avgar et al. 2016). Michelot et al. (2019) reconciles the connection between traditional HSFs and SSFs, such that multi-scale inference from a single modeling framework is possible. However, until implementation across scales is more widely available for practitioners, the decision of traditional HSF vs. SSF is an important question and one that should be based on the goals of the study and meeting modeling assumptions.

Fitting a model to estimate parameters of a traditional HSF or SSF will likely depend on two main considerations, 1) the scale of inference that is being sought, and the 2) temporal dependence between consecutive relocations of the same individual being tracked. Traditional HSFs typically provide inference to an animal's selection at a broad extent (Johnson 1980). The researcher decides the scale by assigning the geographic extent across which they sample available locations. From there, the researcher fits a statistical model to estimate parameters of the HSF under two critical assumptions.

8.1. Critical Assumption 1: Accessibility of habitat

It is assumed that all available locations within the selected geographic extent are equally available to the animal (i.e., uniform distribution within the defined geographical extent). This implies that there are no constraints on the animal in accessing all available locations from any given animal location. This may not be true if a dataset consists of relocations with a short time period between consecutive locations. For example, if the interval between locations is 15-minutes and the species' max speed is 1.5 m/s (typical of a 10 kg animal; Cloyed et al. 2021), they can maximally cover 1350 m in this time. Thus, locations that are more than this distance from an animal location are not realistically available to the animal. If this movement constraint is ignored in an HSF, it will lead to incorrect inference on selection. From Northrup et al. (2013): "These coefficients likely represent some mix of a behavioral response to the environmental factors, and noise induced by the distribution of the covariates on the landscape and the movement of the animal (Beyer et al. 2010)".

8.2. Critical Assumption 2: Independence of location data

Apart from the issues of equal availability within a defined geographic area (e.g., a home range), short intervals between consecutive locations also induce the issue of temporal dependence (i.e., autocorrelation) between consecutive animal locations. Ignoring this dependence will lead to an overly inflated sample size and estimated uncertainty (i.e., variance) of effects (e.g., coefficients) will be too precise. The consequence of this will often lead to finding that many or even most estimated effects have very low p-values, which may lead to incorrect interpretation (Alston et al. 2022). From Alston et al. (2022), “Estimates of resource selection are therefore overconfident and often biased when autocorrelated data are modeled as independent and identically distributed (IID). This is one form of pseudoreplication, which has long been acknowledged as a problem in ecology (Hurlbert, 1984)”. Note that the IID assumption is a common one in statistical modeling and can be understood as observations that are not dependent on each other and follow a similar process, often defined by a probability distribution (e.g., the Normal/Gaussian distribution). **In summary, ignoring temporal dependence in the data and assuming there are no movement constraints when there are may lead to poor inference with misleading statistical conclusions.** With that said, if it is reasonable to assume there are no movement constraints and accessible habitat is captured well, traditional HSFs are a powerful framework to learn about animal-habitat relationships.

Step-selection functions were developed to deal with serial dependence (i.e., autocorrelation, like described above) as well as temporally varying availability distributions resulting from movement constraints (Fortin et al., 2005; Thurfjell et al., 2014). Rather than treat locations as independent (with availability that is static through time), SSFs model transitions, or ‘steps’, connecting sequential locations (Fortin et al. 2005; Thurfjell et al. 2014; Fieberg et al. 2021) in geographical space. By design, each used animal location is compared against only those resources, conditions, and risks that are reasonably available to the individual within the interval during which the next location will be observed.

If a researcher prefers broadscale inference with a traditional HSF, but violates one or both critical assumptions above, they have the choice of thinning high frequency data with a short interval to reduce autocorrelation and movement constraints to fit a broadscale HSF (Swihart and Slade, 1997; Northrup et al. 2013). There are however downsides to thinning data (Alston et al. 2022). Alternatively, there are approaches that retain use of the full, auto-correlated dataset, but which “weight each observed location in an animal's movement track according to its level of non-independence, expanding confidence intervals and reducing bias that can arise when there are missing data in the movement track” (Alston et al. 2022).” For additional guidance, see Box 1 and 2 of Florko et al. (2025) for a decision tree and a comparison chart of methods.

9. The data generating model and the model being fit

In any modeling analysis, the analyst must make an assumption about the model that generated the data. As above, this is always an approximation of the true process that produced the data. For traditional HSFs, the accepted data generating model is an Inhomogeneous Poisson Point process model (also understood via the weighted distribution theory; Warton and Shepherd 2010, Aarts et al. 2012, Fithian and Hastie 2013, Johnson et al. 2013). Most people do not fit this model directly to their data but use statistical models with other names that approximate the Inhomogeneous Poisson Point process model. This can create confusion and lead to decisions that seem appropriate for a given model (e.g., logistic regression model), but are in fact, not appropriate because they are not actually fitting this model. Below, we introduce some statistical model names, connections among them, and relevant details; this section may be overwhelming and skipped to be later revisited. For now, take away that the same HSF or SSF can be approximated by different modeling approaches, much the same way different software programs can be used to fit the same model.

In brief, for a Inhomogeneous Poisson Point process model, spatial locations of individuals are considered points, and the average density of those points in space is dictated by the intensity parameter of the Poisson statistical distribution (often denoted with the Greek letter λ) and its variation across that space (Aarts et al. 2012). Each used point is a realization from a weighted probability distribution, whereby a HSF dictates how animals are selecting resources, and that is weighted by what is available to them to produce the distribution of used locations (Gerber and Northrup 2020, see paragraphs 2-4 or Hooten et al. 2017, Chapter 4.2). The SSF is an extension of the point process model (i.e., spatio-temporal point process model), in which the available samples are conditional on the previous time step (Hooten et al. 2017, Chapter 4.7). For a more general overview of point process models and species distribution modeling, see Hefley and Hooten, 2016.

For common implementations of traditional HSFs and SSFs, one of the major areas of confusion is that the model being fit (by way of coding and name of statistical model being implemented) is often an approximation of the model that is intended to be fit (i.e., the point process model). It is common practice to use convenient and widely available statistical models and software to approximate the true underlying statistical model. Traditional HSF models are often indirectly fit using a generalized linear modeling (GLM) framework; specifically, as a logistic or Poisson regression (Fieberg et al. 2021; see affiliated code). This is not the only way though, as equivalent models can be made using maximum entropy modeling, boosted regression trees, generalized additive models, and more (Fithian and Hastie 2013). For SSFs, it is common practice to approximate the spatio-temporal point process model using conditional logistic regression (specific type of discrete choice model; see affiliated code) or Poisson regression with stratum-specific (related to differentiating each step) fixed intercepts (Muff et al. 2020, Michelot et al. 2024). These approaches are only done out of convenience because the point process models are difficult to fit directly as they contain a large, often intractable integral (Hooten et al. 2017 Eqn. 4.6 and 4.30, Michelot et al. 2024 Eqn. 1). This raises significant confusion, however, because the model being fit (via coding) and the model that we actually desire is not the same. **In summary, the habitat selection model being fit is approximating the underlying assumed generating model (i.e., point process model, which includes a large integral) for the sake of simplicity and accessibility.** Practitioners format used and available data to meet the specifications of the approximating model (e.g., logistic or Poisson regression) rather than the underlying point process model. Therefore, appropriate steps need to be taken to ensure the approximation process is done with minimal approximation error (see next section, *10. The model being approximated*). Somewhat contrastingly, and adding to the confusion on this point, it is advisable to use best practices associated with the underlying habitat selection model (i.e., the point process model) rather than those associated with the analysis being coded, such as logistic regression, when conducting model diagnostics or goodness of fit tests. As an example, using the area under the receiver operating characteristic curve is inappropriate for HSFs because this is a diagnostic for logistic regression models, which are only used out of convenience, but are not the true underlying model.

10. The model being approximated

Fitting traditional HSFs using a logistic or Poisson regression framework and SSFs using conditional logistic regression or Poisson regression with stratum-specific fixed intercepts assumes the habitat selection process follows an exponential form (see “Interpreting Coefficients and Predictions” section below; Hooten et al. 2017, Gerber and Northrup 2020; Michelot et al. 2024). The practical implication of this assumption is that inference is on relative selection, rather than absolute selection (which would be the probability of selection; Lele and Keim 2006). In other words, our inference about selection at habitat units with their unique environmental/spatial values is relative to other such units; this is equivalent to how a measure of relative population abundance indicates where there are likely more or less individuals across different areas, but the true number of individuals (absolute population abundance) in different areas is unknown. While it is theoretically possible to estimate absolute selection, it often requires strict assumptions that are unlikely to be met (Hastie and Fithian, 2013). Relative selection is usually perfectly sufficient for most studies given that we are often interested in comparing habitat units and the relative selection of one habitat unit over another.

To make sure we are approximating the model we intend (i.e., point process model) with the model we fit (e.g., Poisson or logistic regression), care needs to be taken when setting up the data for analysis. Most importantly, the number of locations in the available sample (the zeros) of the logistic regression needs to be very large (Northrup et al. 2013; Hooten et al. 2017). Similarly, the number of grid cells being sampled when using Poisson regression needs to be very large. This has a statistical justification that is beyond the scope of this paper but see Gerber and Northrup (2020) for further detail. The natural next question is, what is large? It is common to see in the literature a pre-set ratio of used to available (e.g., 1 used: X available), but this does not guarantee an appropriate approximation. The answer ‘to what is large’ depends on the species and study and can only be determined by conducting a sensitivity analysis. In other words, large should be determined by increasing the ratio of available to used locations (i.e., sample many more available locations than used) until the estimated effects or coefficients and their uncertainty no longer change, or change little (Northrup et al. 2013; Figure 1). To start, create random or systematic locations within the geographic area of interest (e.g., home range). These available locations are used to characterize the spatial variation from the hypothesized covariates. Fit the model, examine the estimated selection coefficients with their measure of uncertainty, and then do the whole exercise again with a larger number of sampled available points. If the selection coefficients and their uncertainty do not change, the original number of available points sampled was sufficient (see affiliated code). If they do change, keep repeating the exercise until they do not.

The greater the spatial complexity of the landscape, the more available locations will be needed. One recommendation from Street et al. 2021 is that the “availability sampling should be conducted in a regular (non-random) fashion at a spatial interval equal to the resolution of the underlying landscape data such that every possible location within the availability boundary is considered (Benson, 2013; Fieberg et al., 2021). This would produce an availability observation for every raster pixel and thus overlap between used and available locations.” If using logistic regression, the available samples are best weighted (Fithian and Hastie 2013) to help ensure the approximation is accurate. If using the ‘glm’ or similar function (e.g., glmmTMB: Brooks et al. 2017) in the R programming language (R Core Team, 2024), this would mean specifying a vector in the argument ‘weights’ (see affiliated code); this vector would be a series of 1s corresponding to the used locations (the ones; indicating no weighting) and a series of large numbers repeated (e.g., 1000) corresponding to the available locations (the zeros).

This same issue for SSFs and fitting a conditional logistic or Poisson model with stratum-specific fixed intercepts has not been discussed in the literature, but the same idea applies. The available locations are used to approximate the spatial surface in selection driven by the environmental covariates being used in the model. Since the scale of SSFs is much smaller, there is likely less landscape heterogeneity within each sampled area. While this likely means less overall area sampled as available per used location, the total number of sampled available points might be much larger than when using a traditional HSF. A sensitivity analysis will still help evaluate when enough available samples matched to each used location leads to converged parameter estimates. **Ultimately, whether fitting a traditional HSF or SSF, the number of available samples needed depends on context, and it is the responsibility of the researcher to investigate parameter sensitivity to the size of the availability sample and demonstrate that estimates have converged (i.e., are no longer changing) in any manuscript.** For either HSF or SSF, a starting consideration could be 10 available locations to each used location, increasing by 10 up until at least 100.

11. Individuals

For many habitat selection studies, multiple individuals are tracked simultaneously. This provides an important opportunity to evaluate how individuals vary in their habitat selection. Generally, across species and populations, individual variation in habitat selection is the norm (Pape and Löffler, 2015; Montgomery, et al. 2018; Muff et al. 2020; Ganoe et al. 2025). This is because individuals vary in their age, experience, reproductive status, dominance, morphology and more. Furthermore, there is considerable evidence that animal personality can be an important driver of individual specialization regarding resource use (Leclerc et al. 2016; Brehm and Mortelliti, 2021). Lastly, from Bastille-Rousseau and Wittemyer (2019), "Understanding heterogeneity in individual responses to ecological properties is critical to investigate the mechanisms underpinning population level spatial processes (Westneat et al. 2019)."

What does this mean for habitat-selection studies? ***A priori*****, it is safe to assume there is individual variation in habitat selection, this variation is important to making inference on, and thus it is important that the individual is our main sampling unit.** Specific to SSFs, it may also be important to account for individual variation in the movement process, as well as the habitat-selection model (Chatterjee et al. 2024). Of course, when individuals are not sampled directly, such as observing animal locations via snow tracking, we cannot directly account for individual variation. However, considerations can be made as to the spatial scale and how individual behavior (e.g., foraging vs seeking cover) differs across space (Hebblewhite et al. 2014).

When individuals are tracked, pooling data across individuals and ignoring individual variation without additional statistical treatment can lead to poor statistical conclusions (Figure 2). Pooling data across individuals leads to a situation where we are obtaining multiple samples from the same sampling unit (i.e., the individual) to estimate effects. From Muff et al. 2020, "treating all observations as though they are independent would result in optimistic standard errors and confidence intervals, leading to what Hurlbert (1984) referred to as pseudoreplication" (Figure 2). Pooled data will likely lead to finding inappropriate statistical support (i.e., have p-values that are too small) for effects only because the appropriate sampling unit (i.e., the individual) is being ignored. Note that, using a random effect on the intercept only, but not the slope(s), pools the data to estimate the slope (more details in section 12. *Population Inference*). This creates a lack of independence on yet another level, separate from what was described earlier when locations were collected at short fix intervals (sections 8.1 and 8.2). Pooling data also can lead to issues with certain individuals contributing more data to the model if, as is common in telemetry studies, individuals are tracked for different durations, which can further induce inappropriate inference at the population level.

Pooling individual data may be required when each individual has few used locations and thus estimating effects can be challenging. If data are pooled across individuals with large variation in the sample size of used locations, the manuscript should explain that inference is heavily weighted to certain individuals. From Gillies et al. 2006, "pooling will lead to bias towards the individuals with the most data". Further, when individual characteristics are known, it is important to consider these in our habitat selection analysis. To do so, we have to recognize that individual characteristics are not spatial variables and as such are dealt with differently (e.g., interactions with spatial variables or separate models for each group; see Northrup et al. (2022), Box 2). Third, if we are sampling individuals from a larger population we are likely interested in a population-level inference, which we should consider along with our individual-level inference. **The population-level inference (see section 12. Population Inference for details) provides information on the average or typical response, which is useful, but is most accurately expressed within the context of individual-level responses rather than ignoring them (Winter et al. 2024).** The variability in responses across individuals is important, for example, on the magnitude of the variation for each variable of interest. But also, only focusing on the population-level response can be misleading in some circumstances. For example, if half the individuals respond negatively to an environmental feature and the other half respond equally positively, the population-level inference will be that they do not select for or against this feature (effect would be zero). By ignoring the individual-level estimates, the population-level inference that there is no selection or avoidance by individuals would be very misleading.

When modeling habitat selection, we want to have the statistical power to estimate selection and have statistical clarity (Dushoff et al. 2019) of when an effect is different than zero (i.e., no support for an effect). Two considerations in determining the number of used animal locations are required: 1) the magnitude of the selection strength (effect size) and 2) the complexity of landscape (Street et al. 2021). Street et al. (2021; Figure 1) demonstrate the relationship between the magnitude of selection strength and the number of used locations required to have a 50% and 95% chance of rejecting the null hypothesis (i.e., no effect) at a significance level of $p = 0.05$. This is done for a reasonably complex, and thus realistic, resource layer; for evaluation of specific landscape layers, see the code available by Street et al (2021) or that associated with this manuscript. The take home is that for a given time period (e.g., season) and a single selection behavior (e.g., foraging), only a moderate number of used locations (e.g., 100) are needed to achieve statistical clarity when there are moderate to large selection effects ($\leq$ - 0.2 or $\geq$ 0.2). Recognizing this provides hope to those studies with modest sample sizes in considering an individual-based analysis and not pooling data from multiple individuals. In many instances of low sample size, it is still worthwhile to fit models with full random effects (intercepts and slopes) structures. However, as more complexity is considered in the model (e.g., different behaviors or time periods), the number of required used locations will increase. The lack of clear and generalizable guidance on sample sizes necessary to fit full models is a considerable gap in the literature. We encourage future research to focus on eliminating that gap, but for now, we encourage collaborating with a statistician to get parameter estimates to converge.

12. Population Inference

Commonly, species are not managed at the individual-level, but at the population-level, and we are interested in how a population generally behaves and interacts with their environment. For habitat selection studies, this means we are often interested in scaling up our inference from an individual's selection to an average individual's selection. One way to do this is to first fit separate models to each individual and estimate parameters. Then, for each shared parameter between individual models, estimate a population-level mean and measure of uncertainty using defined equations (see Sawyer et al. 2006) or bootstrap methods (Prokopenko et al. 2017; see affiliated code). The benefit of this approach is that the statistical distribution from which individual-level effects are realized does not need to be assumed (Fieberg et al. 2010). However, it is necessary that effects for each individual can be estimated, such that sample sizes are adequate for each individual; this concern is relaxed in the next approach. Importantly, each individual's model structure needs to be set up similarly, such that the estimated effects for each individual mean the same thing. For example, if using a land cover classification variable, each individual needs to have the same reference level (i.e., which land cover variable is considered the baseline or model intercept), so the estimate of differences from the reference to the other levels mean the same thing.

An alternative to separate individual models is to jointly estimate individual and population selection parameters in a single model. This can be done using a random effect (Gillies et al. 2006; Duchesne et al. 2010), which is when a statistical model allows parameters to be random variables. The whole model is also known as a hierarchical or multilevel model. It is common to have fixed (i.e., not random) and random parameters together in a single model. In this context, we can think about how some effects on habitat selection might be the same across all individuals (fixed effect) or vary by individual in such a way that they come from a shared distribution of possible values (random effect). This shared distribution is often assumed to be a Normal distribution, which can be fully described by two model parameters, the 1) mean and 2) standard deviation. The mean of the random effect represents the population-level parameter effect and the standard deviation a measure of the variability across the sampled individuals. Fortunately, this modeling approach also retains estimates of individual-level selection parameters and their uncertainty. Depending on how the model is fit (i.e., means parameterization vs. effects parameterization), the individual selection parameters may represent the individual difference from the population-level mean.

Before implementing random effects, we should understand some background about them and implications when using them. Thinking of two extremes of using data, we could pool all used locations from all individuals (which we do not suggest), ignore the individual as a factor in a model and estimate a single effect for a habitat selection variable of interest (referred to as complete pooling). Or, we could treat all individuals as different and unrelated and estimate an effect for each individual separately (no pooling; e.g., separate models for each individual; Gelman and Hill 2007, Chapter 12). Random effects represent the middle ground and are referred to as partial-pooling where an estimated population-level distribution (i.e., the random effect) provides a “soft constraint” (Gelman and Hill, 2007 Chapter 12, pg 257) on the individual parameter estimates. While each individual-level effect is estimated from the data for that individual, the soft constraint has the effect of pulling or shrinking the individual effect towards the population-level mean (across all individuals). The effect of this shrinkage is that the estimate of the population-level parameters from across the individuals is helping inform the estimates of the individual, i.e., there is sharing of information across individuals.

Why is sharing information good? The primary reason is that the amount of sharing will be mediated by the sample size for each individual. Therefore, the shrinkage can help estimate individual coefficients that otherwise would be difficult to estimate if data were modeled completely separately (no pooling). For example, if an individual rarely interacts with a land cover category, it might be impossible to estimate a coefficient for that category in an individual model. When sharing information, the coefficient for this individual would be estimated close to the population level mean, allowing inclusion of this animal in the analysis. Furthermore, for individuals with small sample sizes, coefficients may be estimated at the extremes (highly negative or positive) due to low replication of used locations (referred to as 'Type M' errors; Lemoine et al. 2016). A nice feature of the random effect is that if there is little to no individual variation, the estimated population standard deviation will be near zero and will represent a model that has pooled data. "Conversely, when group-level coefficients vary greatly multilevel modeling reduces to classical regression with group indicators" (Gelman and Hill 2007, Chapter 11, page 247). As such, from Gelman and Hill (2007, Chapter 11, pg 247), "there is little risk from applying a multilevel model, assuming we are willing to put in the effort to set up the model and interpret the resulting inferences." Lastly, taking a random effect approach leads to allowing a joint inference to individual- and population-level effects with proper measures of uncertainty (Gilbert et al. 2023). This makes using a random effect a convenient statistical option, but one that can certainly lead to computational challenges (see text at the end of this section).

Several considerations are needed when implementing random effects in a habitat selection analysis. The foremost is whether to include a random effect for the intercept only, or the intercept and slope parameters (Muff et al. 2020). For HSFs, the intercept is largely the ratio of used to available, which is partly controlled by the analyst (Warton and Shepherd, 2010, Theorem 3.2). As such, individuals will likely have different intercepts simply because each individual is unlikely to have the same number of used locations (i.e., an unbalanced design). Random intercepts account for this unbalanced design. From Gillies et al. 2006, "Without a random intercept for individuals with unbalanced data, sample size differences may influence model coefficients." Furthermore, the intercept is also influenced by the variable values within the model, which may not be the same for all individuals. Thus, when jointly modeling individuals in a habit selection model, allowing the intercept to be different by individual is a must. One way of doing this is using a random intercept. **However, we do not want the intercepts to be shrunk towards the population-mean because it would induce bias in the selection function (Muff et al. 2020). Therefore, the best practice is to not estimate the variance or standard deviation of the random intercept, but to fix it to a large value (e.g., variance of 10^6; Muff et al. 2020).** While this is important for traditional HSFs, it is even more important for SSFs being approximated via the Poisson model with stratum-specific fixed intercepts (Muff et al. 2020).

Muff et al. (2020) noted in their literature review that most habitat selection studies that use random effects only use random intercepts and not random slopes. Therefore, most studies are assuming there is no variation in an individual's response (i.e., fixed effect parameter) to their hypothesized variables about habitat selection. From Schielzeth and Forstmeier (2009), "Problematically, random intercept models wrongly treat repeated measurements within individuals as independent data points with respect to the population slope.” The outcome of this is described by Muff et al. (2020) as, “Random intercept-only models tend to underestimate standard errors of (potentially biased) fixed effect parameters.” For studies with many used locations and several to many individuals, this will lead to too many rejections of the null hypothesis (i.e., effect of no difference) for the population-wide mean slope. **Therefore, the recommended approach is to always use random slopes/effects along with random intercepts (with a fixed, large variance on the intercept; Muff et a. 2020).** Doing so will account for unbalanced sampling across individuals, reduce pseudoreplication, recognize that individuals are the main sampling unit, capture the *a priori* assumption that individuals are expected to show variability, estimate appropriate standard errors, and guard against overly optimistic interpretations of statistical clarity (i.e., rejecting the null hypothesis of no difference for habitat selection effects).

The intention of making inference from across a set of sampled individuals is that this inference reflects the larger population, including unsampled individuals. But, how many individuals are required to safely make inference to the larger population? Street et al. (2021) demonstrated that when population-level effects are greater than 1 (assuming covariates have been standardized to a mean 0 and standard deviation of 1), it is possible that fewer than 10 individuals are needed (Street et al. Figure 5). **It is not warranted to assume studies with few individuals sampled are unable to make population-level inference.** Rather, the reliability of making such an inference is best evaluated based on effect sizes, variation among individuals, landscape complexity (i.e., Street et al. 2020, Eqn. 8; see also the affiliated code with this manuscript), and some assessment (even qualitative) of the representativeness of the individuals in the small sample. Generally, the “more variation amongst individuals is likely to mean one has to sample a higher number of animals” (Street et al. 2020).

A common reason for forgoing random intercepts and slopes is they create computational issues when fitting models (Bolker, 2009). These issues can happen for many reasons. The model may be over parameterized for the sample size (i.e., too many variables in the model; see recommendations by Fieberg and Johnson 2015) or there is redundancy among intercepts (can happen when factor levels are included as binary covariates and one level is not dropped). Another common estimation issue is when too many land cover variables are used, such that some types are rarely available and rarely used. In these cases, aggregating land cover types may be necessary. Estimation should generally be feasible when land cover types are highly available but rarely used, but initial computational issues are likely. Generic solutions relating to the optimization process (iterative numerical evaluations happening when fitting a model) is to try changing the optimization function, the maximum number of iterations, and providing starting values for the optimization function (e.g., estimates from models fit separately to individuals). If implementing random effects, a good way to diagnose which individuals are causing the issues is to first fit the same model to each individual separately (i.e., no random effects) and evaluate whether each model converges and provides reasonable estimates. From here, leave out the individuals with suspect model fits and fit the full random effect model. If this model fits, add in individuals until it is clear which individual’s data are causing issues. Once the problem data are identified, it might be necessary to discuss options with statistically minded colleagues.

13. Considering context in habitat selection analyses

Habitat-selection is about behavior, and “animals are likely to select disparate resources for different behaviors” (Northrup et al. 2022). For example, if an animal is actively seeking or guarding a mate, their risk tolerance is likely to be different to when they are not, such that their selection of habitat can be quite different. Considering all animal locations simultaneously without regard to behavior or contextual environments (e.g., diel period) can “mask strong diel period- and behavior-specific differences in habitat selection” (Suraci et al. 2019). **Generally, ignoring behavioral dependence can strongly impact inference on habitat selection (Roever et al 2014; Abrahms et al. 2016).** From Beyer et al. (2010), “We argue that analysing location data outside its behavioural context limits our ability to establish the link between habitat use, preference, selection and, ultimately, fitness. Further, “Collecting behavioural data to complement fine-scale location data is prerequisite to developing a better mechanistic understanding of habitat use.” (Beyer et al. 2010)

Incorporating behavior into habitat selection analyses requires identifying when behaviors are occurring and assigning a behavior type to each animal location (e.g., resting, foraging, hunting). For some species, certain behaviors may be obvious, such as a bird roosting at nighttime versus feeding or loafing during the daytime. When behaviors are not obvious, one can classify locations by movement trajectory types that consider speed and turning angle to sequence locations into ‘behavior states’ (Michelot et al. 2017). Habitat selection analyses can then be done separately for the animal locations in each of the different states (Suraci et al. 2019). This approach is the most straightforward for practitioners, however, it is important to recognize that it ignores uncertainty about the state classifications. Several newer approaches are being developed to jointly model habitat selection and behavioral states (Nicosia et al. 2017; Klappstein et al. 2023; Pohle et al. 2024; McClintock and Lander, 2024). However, “Caution is warranted when interpreting behavioral insights rendered from statistical models applied to location data, particularly when there is no form of validation data.” (Buderman et al. 2021). Ideally, additional field data (e.g., direct observations or accelerometers) could be used to validate the meaning of these behavioral states. Ultimately, we will rarely capture detailed behaviors such that we can expect that our habitat selection inferences are in fact averaging across multitudes of unknown behaviors.

Not only is habitat selection a behavioral process, but the behaviors and resulting selection have fitness consequences (Matthiopoulos et al. 2015; see Matthiopoulos et al. 2022 Figure 1.2). What is available to an individual is not only a function of their current position on the landscape and their ability to move to other habitats, but also the competitive forces incurred by other individuals. By definition, habitat selection is dependent on the population density of the species and is expected to change based on intraspecific competition (Avgar et al. 2020). Specifically, increasing population density will on average lead to more general selection of habitats, i.e., the selection intensity of certain habitats will be reduced (Fortin et al. 2008; van Beest et al. 2014) because their fitness payoff relative to other available habitats has been reduced from intraspecific competition. When estimates of population density are available it is important to consider them in habitat selection analyses (McLoughlin et al. 2010) and when comparing habitat selection results of the same species across regions or years. Differences in habitat selection may be a result of different population density.

Another consideration for habitat selection modeling is to recognize that for many species, selection is highly context dependent (Northrup et al. 2022, section ‘Context dependence in habitat selection; Suraci et al. 2019). When considering spatial variables of resources, risks, and conditions we need to consider how selection regarding one variable may depend on another variable (Ganoe et al. 2025). For example, an animal may have strong selection for areas near a road with low traffic volumes at nighttime to forage within, but differently select for the same areas (i.e., a stronger or weaker magnitude of selection) during the daytime or twilight, and also differently select for areas near roads with higher traffic volumes at nighttime, daytime, or twilight (e.g., Scrafford et al. 2018). To consider this spatial context dependency, we need to consider interactions among the variables of distance to road, traffic volume, and diel time. **The general point being that complex landscapes with spatially varying risks, resources, and conditions require hypothesizing interactions among spatial variables.**

14. Interpreting coefficients and predicting

Interpreting estimated effects requires a clear understanding of linear models, design matrices, and statistical notation. We suggest starting with Edwards and Auger-Méthé (2019) to learn notation, then progress to a regression or linear model textbook to learn about design matrices and estimation (e.g., Fieberg, 2024; Chapter 3), and finally apply this knowledge when reading Fieberg et al. (2021).

When predicting from a HSF, we need to remember that we (probably) have assumed an exponential form for our habitat selection model and that we are approximating the point process model (e.g., by logistic regression). Understanding the implication of this will require some general modeling knowledge; specifically, that parameters are often transformed to another scale (via a ‘link’ function; e.g., logit, log) for estimation but can be back-transformed (via a ‘inverse-link’ function; e.g., logit-inverse, exp) to the original scale of the data. In fitting a HSF via logistic regression, the presumed inverse function is the logit-inverse, but since we are approximating an exponential function the parameters are estimated on the log-scale, thus the correct inverse-function is the exponential function. Therefore, simply back-transforming (e.g., logit-inverse) the combinations of our estimated effects and variable values using common coding functions (e.g., predict.glm, plogis) is inappropriate to predict the relative intensity of selection for a given set of covariates. This holds for any prediction being carried out, including providing a series of new data values to interpret interactions or simply plotting the selection or avoidance behavior across a range of values for a continuous covariate. **To compute predicted values, we need to remove the intercept from the selection model (often denoted as β_0; it is largely a measure of the ratio of used to available locations, which is controlled by the analyst) and exponentiate the summed combinations of effects (β_1 to β_n) and variable values ($e^{(\beta_1 \times X_1 + \beta_2 \times X2 \ldots + \beta_n \times X_n)}$) where e can be implemented in R using the exp() function and the X_1, X_2,...X_n values are the values of the variables of resources, risks and environmental conditions**. The resulting values of these predictions will be between 0 and positive infinity. It is common in the literature to see spatial prediction maps of the relative intensity of selection scaled between zero and one. This is not necessarily inappropriate, as long the predictions are calculated as described above first, but there is a concern that readers may unintentionally interpret zero to one values as absolute probabilities, which they are not. In addition, without explicit language/notation on how the predictions are calculated, readers will not know whether the predictions are being calculated correctly.

Following above in calculating predicted values, we can interpret avoidance as predictions less than one and selection as greater than one. Mapping these predictions is most effective when they highlight the spatial differences in selection and avoidance and is a common goal of habitat selection studies. The predicted values described above are used to map habitat selection. Rasters of predictor variables are sampled across the landscape of interest (likely the same rasters used to sample predictor variables for used and available locations) and, using the above transformation, predicted selection values are assigned to each raster cell. Importantly, predictions are most robust when they focus on the study system where the observed data were collected. Predicting outside of this region (model transferability) can be done but with great caution (Wiens et al. 2008; McLoughlin et al. 2010; Olson et al. 2020; Matthiopoulos et al. 2011; Matthiopoulos et al 2023) and it is important to consider the sampling design to ensure the sampled area is representative of the area to which selection is being predicted. Lastly, predicting from a SSF needs additional care because selection is conditional on each step and thus needs to be considered when interested in scaling up predictions to a larger landscape. These processes can be computationally intensive but are an area of active research (Potts and Börger, 2023; Signer et al. 2024).

15. Model Selection

For some, learning habitat selection modeling will coincide with learning about the role of ‘model selection’. The first question one needs to ask is what is the goal of the model(s)? Tredennick et al. 2019 starts from this point of view and helps frame different ways to think about model selection depending on the goal. We also suggest Bolker (2024) and Fieberg, 2024 (Chapter 8) who separate model selection by goals of inference or prediction.

There are a few points about model selection specific for habitat selection models that are worth considering here. It is not unusual in GPS telemetry studies to have many thousands of animal locations per individual. Given such a sample size, inference is straightforward because you can include all variables of interest in a single model (which also forces you to think *a priori* about which variables are of interest). Model selection may be irrelevant because there is no need for trading off bias for variance in parameter estimates by considering simpler models, which is, in technical terms, what model selection accomplishes. **It is thus reasonable to focus on fitting one model that captures all the variables that represent the complete hypotheses of interest. Do not assume model selection is a requirement.** More generally, considering a single model makes for clearly interpretable inference; from Bolker 2024, "If researchers want to use limited data sets to accurately estimate the strength of multiple competing ecological processes along with reliable confidence intervals, the current best approach is to use full (maximal) statistical models (possibly with Bayesian priors) after making principled, *a priori* decisions about model complexity." This is particularly salient in habitat selection models where there are often extremely large numbers of potential covariates that could be included in models and typically little logical reason for excluding potential habitat covariates (doing so amounts to testing whether a habitat feature has no effect on a species which can more easily be assessed by looking at coefficient magnitudes).

CONCLUSION

The literature landscape pertaining to habitat selection is vast, full of jargon, often technical, confusing, and unfortunately includes numerous examples of poor implementation practices. While important, the technical notation of the statistical processes underlying HSFs can intimidate ecologists and field-hardened practitioners, alike. This review was intended to identify, in no uncertain terms, some important guidelines to conduct rigorous, statistically sound habitat selection studies, and to help readers better understand the reasons behind these guidelines in plain terms. This guide is a first step to gaining the foundation needed to conduct and review habitat selection studies, but extensive resources abound to continue learning. We have referenced many of those resources throughout this paper, but also provide a more consolidated list of recommendations in Table 1. We also offer Box S1 (Supporting Information) to provide support in thinking through a habitat selection analysis.

Learning the principles of any ecological field takes time and effort, and those new to the field of habitat selection have their work cut out for them as the methodology has rapidly advanced and studies have quickly proliferated over the past several decades. We urge anyone conducting habitat selection studies to take their time and focus on getting the basics right (i.e., sampling design, model fitting, and interpretation). Ecological studies are difficult, and researchers often cannot accomplish everything they set out to do. This is entirely okay. They can, however, make an effort to understand the justification for why they are doing what they are, acknowledge the components that cannot be accommodated in sampling or analysis, and ensure their sampling and analysis methodology are appropriate for the questions of interest. The statistical underpinnings of habitat selection analyses are of paramount importance to ensuring proper applied inference for wildlife management decisions. We urge practitioners to note that there is no better way to learn than to simply try these methods on a study system of interest and to seek help when needed.

ACKNOWLEDGEMENTS

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We are grateful to the graduate students and practitioners who motivated this manuscript and the many that provided feedback on earlier drafts. We are also grateful to P. Williams, N. Jaffe, three anonymous reviewers, and the Kohl Lab for reviews on previous drafts.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ETHICS STATEMENT

Not applicable.

DATA AVAILABILITY

Data and code supporting this manuscript is available at GitHub at https://github.com/bgerber123/Habitat-Selection-Guidance/ and Gitlab at https://code.usgs.gov/cooperativeresearchunits/colorado/habitat-selection, archived at Zenodo https://doi.org/10.5281/zenodo.14847204, and viewable online at https://bgerber123.github.io/hsfguide/.

REFERENCES

Aarts, G., J. Fieberg, and J. Matthiopoulos. 2012. Comparative interpretation of count, presence–absence and point methods for species distribution models. Methods in Ecology and Evolution 3:177-187.

Abrahms, B., Jordan, N. R., Golabek, K. A., McNutt, J. W., Wilson, A. M., and J. S. Brashares. 2016. Lessons from integrating behaviour and resource selection: activity-specific responses of African wild dogs to roads. Animal Conservation 19:247-255.

Alston, J.M., C.H. Fleming, R. Kays, J.P. Streicher, C.T. Downs, T. Ramesh, B. Reineking, and

J.M. Calabrese. 2022. Mitigating pseudoreplication and bias in resource selection functions with autocorrelation-informed weighting. Methods in Ecology and Evolution 14:643-654.

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

Avgar, T., Betini, G. S., and J. M. Fryxell. 2020. Habitat selection patterns are density dependent under the ideal free distribution. Journal of Animal Ecology 89:2777-2787.

Bastille-Rousseau, G. and G. Wittemyer. 2019. Leveraging multidimensional heterogeneity in resource selection to define movement tactics of animals. Ecology Letters 22:1417-1427.

Beyer, H. L., D. T. Haydon, J. M. Morales, J. L. Frair, M. Hebblewhite, M. Mitchell, and J. Matthiopoulos. 2010. The interpretation of habitat preference metrics under use-availability designs. Philosophical Transactions of the Royal Society B 365:2245–2254.

Benson, J. F. 2013. Improving rigour and efficiency of use-availability habitat selection analyses with systematic estimation of availability. Methods in Ecology and Evolution 4:244–251.

Bischof, R., Vallejo-Vargas, A. F., Semper-Pascual, A., Schowanek, S.D., Beaudrot, L., Turek, D., Jansen, P.A., Rovero, F., Johnson, S.E., Guimarães Moreira Lima, M. and Santos, F., 2024. The moon’s influence on the activity of tropical forest mammals. Proceedings B 29:20240683.

Bolker, B. 2009. Learning hierarchical models: advice for the rest of us. Ecological

Applications 19:588-592.

Bolker, B. M. Multimodel approaches are not the best way to understand multifactorial systems. 2024. Entropy 26:506.

Borowik, T., Kowalczyk, R., Ratkiewicz, M., Maślanko, W., Duda, N., and Żmihorski, M. 2024. A highly variable habitat selection in moose across diel and seasonal scales. Movement Ecology 12:69.

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

Boyce, M. S. 2006. Scale for resource selection functions. Diversity and Distributions 12:269-276.

Boyce, M. S., Johnson, C. J., Merrill, E. H., Nielsen, S. E., Solberg, E. J., and B. Van Moorter. 2016. Can habitat selection predict abundance?. Journal of Animal Ecology 85:11-20.

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A, Skaug, J. J., Mächler, M, and B. M. Bolker. 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. The R Journal 9:378-400.

Brehm, A. M. and A. Mortelliti. 2021. Land-use change alters associations between personality and microhabitat selection. Ecological Applications 31:e02443.

Buderman, F. E., Gingery, T. M., Diefenbach, D. R., Gigliotti, L. C., Begley-Miller, D., McDill, M. M., Wallingford, B. D., Rosenberry, C. S. and P. J. Drohan. 2021. Caution is warranted when using animal space-use and movement to infer behavioral states. Movement Ecology 9:1-12.

Buderman, F.E., Helm, P. J., Clark, J. D., Williamson, R. H., Yarkovich, J., and Mullinax, J.

M. 2023. A multi-level modeling approach to guide management of female feral hogs in Great Smoky Mountains National Park. Biological Invasions 25:3065-3082.

Chatterjee, N., Wolfson, D., Kim, D., Velez, J., Freeman, S., Bacheler, N. M., Shertzer, K., Taylor, J. C., and J. Fieberg. 2024. Modelling individual variability in habitat selection and movement using integrated step-selection analysis. Methods in Ecology and Evolution 15:1034-1047

Cloyed, C. S., Grady, J. M., Savage, V. M., Uyeda, J. C., and A. I. Dell. 2021. The allometry of locomotion. Ecology 102:e03369.

Colorado Parks and Wildlife. 2024. Colorado East Slope Mountain Lion (Puma concolor) Management Plan. https://cpw.widen.net/s/nhqqffhxhn/east-slope-mountain-lion-management-plan_final_approved.

Dejeante, R., Valeix, M., and Chamaillé-Jammes, S. 2024. Time-varying habitat selection analysis: A model and applications for studying diel, seasonal, and post-release changes. Ecology 105:e4233.

Dellinger, J. A., Cristescu, B., Ewanyk, J., Gammons, D. J., Garcelon, D., Johnston, P., Martins, Q., Thompson, C., Vickers, T. W., Wilmers, C. C., Wittmer, H. U. and S. G. Torres. 2020. Using Mountain Lion Habitat Selection in Management. Journal of Wildlife Management 84:359-371.

Duchesne, T., Fortin, D. and N. Courbin. 2010, Mixed conditional logistic regression for habitat selection studies. Journal of Animal Ecology 79: 548-555.

Dushoff, J., Kain, M. P., and B. M. Bolker. 2019. I can see clearly now: Reinterpreting statistical significance. Methods in Ecology and Evolution 10:756-759.

Dyson, M. E., Slattery, S. M., and B.C. Fedy. 2019. Microhabitat nest-site selection by ducks in the boreal forest. Journal of Field Ornithology 90:348-360.

Edwards, A. M. and M. Auger-Méthé. 2019. Some guidance on using mathematical notation in ecology. Methods in Ecology and Evolution 10:92-99.

Fieberg, John. 2024. Statistics for Ecologists: A Frequentist and Bayesian Treatment of Modern Regression Models. An open-source online textbook. Retrieved from the University Digital Conservancy, https://doi.org/10.24926/9781959870029. Accessed December 27, 2024.

Fieberg, J. and D. H. Johnson, D. H. 2015. MMI: Multimodel inference or models with management implications?. The Journal of Wildlife Management 79:708-718.

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

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

Fithian, W. and T. Hastie. 2013. Finite-sample equivalence in statistical models for presence-only data. The Annals of Applied Statistics 7:1917.

Florko, K., Togunov, R. R., Gryba, R., Sidrow, E., Ferguson, S. H., Yurkowski, D. J., and M. Auger-Méthé. 2025. An introduction to statistical models used to characterize species-habitat associations with animal movement data. Movement Ecology 13:27.

Forester, J. D., Im, H. K. and Rathouz, P. J. 2009. Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. Ecology 90: 3554-3565.

Forrest, S. W., Pagendam, D., Bode, M., Drovandi, C., Potts, J. R., Perry, J., ... & Hoskins, A. J.

(2025). Predicting fine-scale distributions and emergent spatiotemporal patterns from temporally dynamic step selection simulations. Ecography, 2025(2), e07421.

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

Fortin, D., Morris, D. W., and P. D. McLoughlin. 2008. Habitat selection and the evolution of specialists in heterogeneous environments. Israel Journal of Ecology and Evolution 54:311-328.

Fretwell, S. D. and H. L. Lucas. 1970. On territorial behavior and other factors influencing habitat distribution in birds. I. Theoretical development. Acta Biotheoretica 19:16-36.

Ganoe, L. S., Northrup, J. M., Mayer, A. E., Brown, C., and Gerber, B. D. 2025. Individuality, diel time, and landscape context shape space-use of an elusive carnivore in a risky environment. Landscape Ecology 40:81.

Gerber, B. D. and J. M. Northrup. 2020. Improving spatial predictions of animal resource selection to guide conservation decision making. Ecology 101:e02953.

Gilbert, N. A., H. N. Eyster, and E. F. Zipkin. 2023. Propagating uncertainty in ecological models to understand causation. Frontiers in Ecology and the Environment 21:114-115.

Gillies, C. S., Hebblewhite, M., Nielsen, S. E., Krawchuk, M. A., Aldridge, C. L., Frair, J. L., Saher, D. J., Stevens, C. E., and C. L. Jerde. 2006. Application of random effects to the study of resource selection by animals. Journal of Animal Ecology 75:887-898.

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I., Regan, T. J., Brotons, L., McDonald-Madden, E., et al. 2013. Predicting species distributions for conservation decisions. Ecology Letters 16:1424–1435.

Hall, L. S., Krausman, P. R., & Morrison, M. L. (1997). The habitat concept and a plea for standard terminology. Wildlife society bulletin, 173-182.

Hastie, T., and W. Fithian. 2013. Inference from presence-only data; the ongoing controversy. Ecography 36:864–867.

Hebblewhite, M., Miquelle, D. G., Robinson, H., Pikunov, D. G., Dunishenko, Y. M., Aramilev, V. V., Nikolaev, I.G., Salkina, G.P., Seryodkin, I.V., Gaponov, V.V., et al. 2014. Including biotic interactions with ungulate prey and humans improves habitat conservation modeling for endangered Amur tigers in the Russian Far East. Biological Conservation 178:50-64.

Hefley, T. J. and Hooten, M. B. 2016. Hierarchical species distribution models. Current Landscape Ecology Reports 1:87-97.

Hooten, M. B., D. S. Johnson, B. T. McClintock, and J. M. Morales. 2017. Animal movement: statistical models for telemetry data. CRC Press, Boca Rotan, Florida, USA.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:87–211.

Johnson, D. H. 1980. The comparison of usage and availability measurements for evaluating resource preference. Ecology 61:65-71.

Johnson, C. J., Nielsen, S. E., Merrill, E. H., McDonald, T. L., & Boyce, M. S. (2006). Resource selection functions based on use-availability data: theoretical motivation and evaluation methods. The Journal of Wildlife Management, 70(2), 347-357.

Johnson, D. S., Hooten, M. B., and C. E. Kuhn. 2013. Estimating animal resource selection from telemetry data using point process models. Journal of Animal Ecology 82:1155–1164.

Kays, R., Crofoot, M. C., Jetz, W., and M. Wikelski. 2015. Terrestrial animal tracking as an eye on life and planet. Science 348:aaa2478.

Klappstein, N. J., Thomas, L., and T. Michelot. 2023. Flexible hidden Markov models for behaviour-dependent habitat selection. Movement Ecology 11:30.

Krausman, P. R., and Morrison, M. L. 2016. Another plea for standard terminology. The Journal of Wildlife Management 80:1143-1144.

Leclerc, M., Vander Wal, E., Zedrosser, A., Swenson, J. E., Kindberg, J., F. and Pelletier. 2016. Quantifying consistent individual differences in habitat selection. Oecologia 180:697-705.

Lele, S. R. and J. L. Keim. 2006. Weighted distributions and estimation of resource selection probability functions. Ecology 87:3021-3028.

Lele, S. R., Merrill, E. H., Keim, J., and M. S. Boyce. 2013. Selection, use, choice and occupancy: clarifying concepts in resource selection studies. Journal of Animal Ecology, 82:1183-1191.

Lemoine, N. P., Hoffman, A., Felton, A. J., Baur, L., Chaves, F., Gray, J., Yu, Q. , and M. D. Smith. 2016. Underappreciated problems of low replication in ecological field studies. Ecology 97:2554-2561.

Manly, B. F. J., McDonald, L. L., Thomas, D. L., McDonald, T. L., and W. P. Erickson. 2002. Resource selection by animals; statistical design and analysis for field studies. Second edition. Kluwer Academic Publishers, London, United Kingdom.

Matthiopoulos, J., Hebblewhite, M., Aarts, G. and J. Fieberg. 2011. Generalized functional responses for species distributions. Ecology 92:583-589.

Matthiopoulos, J., Fieberg, J., Aarts, G., Beyer, H. L., Morales, J. M., and D. T. Haydon. 2015. Establishing the link between habitat selection and animal population dynamics.

Ecological Monographs 85:413-436.

Matthiopoulos, J., Field, C., and R. MacLeod. 2019. Predicting population change from models based on habitat availability and utilization. Proceedings of the Royal Society B 286:20182911.

Matthiopoulos, J, Fieberg, J, and G. Aarts. 2023. Species-Habitat Associations: Spatial data, predictive models, and ecological insights, 2nd edition. University of Minnesota Libraries Publishing. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/217469. Accessed December 27, 2024.

McClintock, Brett T. and Michelle E. Lander. 2024. A multistate Langevin diffusion for inferring behavior-specific habitat selection and utilization distributions. Ecology 105:e4186.

McGarigal, K. Wan, H.Y., Zeller, K.A., Timm, B.C., and S.A. Cushman. 2016. Multi-scale habitat selection modeling: a review and outlook. Landscape Ecology 31:1161-1175.

McLoughlin, P. D., Morris, D. W., Fortin, D., Vander Wal, E., and A. L. Contasti. 2010. Considering ecological dynamics in resource selection functions. Journal of Animal Ecology 79:4-12.

Michelot T, Langrock R, Bestley S, Jonsen ID, Photopoulou T, and T. A. Patterson. 2017. Estimation and simulation of foraging trips in land-based marine predators. Ecology 98:1932–44.

Michelot, T., Blackwell, P. G., and J. Matthiopoulos. 2019. Linking resource selection and step selection models for habitat preferences in animals. Ecology 100:e02452.

Michelot, T., Klappstein, N. J., Potts, J. R., and J. Fieberg. 2024. Understanding step selection analysis through numerical integration. Methods in Ecology and Evolution 15:24–35.

Montgomery, R. A., K. M. Redilla, W. Ortiz-Calo, T. Smith, B. Keller, and J. J. Millspaugh. 2018. Evaluating the individuality of animal-habitat relationships. Ecology and Evolution 8:10893–10901.

Morris, D. W. 2003. Toward an ecological synthesis: a case for habitat selection. Oecologia 136:1-13.

Morris, L. R., Proffitt, K. M., and J. K. Blackburn. 2016a. Mapping resource selection functions in wildlife studies: concerns and recommendations. Applied Geography 76:173-183.

Morris, L. R., Proffitt, K. M., Asher, V., and J. K. Blackburn. 2016b. Elk resource selection and implications for anthrax management in Montana. The Journal of Wildlife Management 80:235-244.

Muff, S., Signer, and J. Fieberg. 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 89:80-92.

Nicosia A, Duchesne T, Rivest L-P, and D. Fortin 2017. A multi-state conditional logistic regression model for the analysis of animal movement. Annals of Applied Statistics 11:1537-60.

Noonan, M. J., Tucker, M. A., Fleming, C. H., Akre, T. S., Alberts, S. C., Ali, A. H., et al. 2019. A comprehensive analysis of autocorrelation and bias in home range estimation. Ecological Monographs 2:e01344.

Northrup, J. M., Anderson Jr, C. R., Gerber, B. D., and G. Wittemyer. 2021. Behavioral and demographic responses of mule deer to energy development on winter range. Wildlife Monographs 2081-37.

Northrup, J. M., Hooten, M. B., Anderson Jr, C. R., and G. Wittemyer. 2013. Practical guidance on characterizing availability in resource selection functions under a use–availability design. Ecology 94:1456-1463.

Northrup, J.M., E. Vander Wal, M. Bonar, J. Fieberg, M.P. Laforge, M. Leclerc, C.M. Prokopenko, and B.D. Gerber. 2022. Conceptual and methodological advances in habitat-selection modeling: guidelines for ecology and evolution. Ecological Applications 32:e02470.

Olson L. E., Bjornlie N., Hanvey G., Holbrook, J. D., Ivan, J. S., Jackson, S., Kertson, B., King, T., Lucid, M., Murray, D., et al.. 2021. Improved prediction of Canada lynx distribution through regional model transferability and data efficiency. Ecology Evolution 11:1667-1690.

Pape, R. and J. Löffler. 2015. Ecological dynamics in habitat selection of reindeer: an interplay of spatial scale, time, and individual animal's choice. Polar Biology 38:1891-190

Paton, R.S. and J. Matthiopoulos. 2016. Defining the scale of habitat availability for models of habitat selection. Ecology 97:1113-1122.

Peterson, C. J., DeCesare, N.J., Hayes, T.A., Bishop, C.J., and M.S. Mitchell. 2022. Consequences of migratory strategy on habitat selection by mule deer. Journal of Wildlife Management 86:e22135.

Pohle, J., Signer, J., Eccard, J. A., Dammhahn, M., and Schlägel, U. E. 2024. How to account for behavioral states in step-selection analysis: a model comparison. PeerJ 12:e16509.

Potts, J. R., and Schlägel, U. E. 2020. Parametrizing diffusion-taxis equations from animal movement trajectories using step selection analysis. Methods in Ecology and Evolution, 11:109201105.

Potts, J. R. and Börger, L. 2023. How to scale up from animal movement decisions to spatiotemporal patterns: An approach via step selection. Journal of Animal Ecology 92:16-29.

Prokopenko, C. M., Boyce, M. S., and T. Avgar. 2017. Characterizing wildlife behavioural responses to roads using integrated step selection analysis. Journal of Applied Ecology 54:470-479.

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Roever, C. L., Beyer, H. L., Chase, M. J., and R. J. Van Aarde. 2014. The pitfalls of ignoring behaviour when quantifying habitat selection. Diversity and Distributions, 20:322-333.

Sawyer, H., Nielson, R. M., Lindzey, F., and L. L. McDonals. 2006. Winter habitat selection of mule deer before and during development of a natural gas field. The Journal of Wildlife Management 70:396-403.

Scrafford, M. A., Avgar, T., Heeres, R., and M. S. Boyce. 2018. Roads elicit negative movement and habitat-selection responses by wolverines (Gulo gulo luscus). Behavioral Ecology 29:534-542.

Schielzeth, H. and W. Forstmeier. 2009. Conclusions beyond support: overconfident estimates in mixed models. Behavioral Ecology 20:416-420.

Schneider, D. C. 2001. The rise of the concept of scale in ecology: The concept of scale is evolving from verbal expression to quantitative expression. BioScience 51:545-553.

Setash, C. M., Behney, A.C. , Gammonley, J. H., Pejchar, L., Reddy, M.A., and D. N. Koons.

2024. Agricultural mosaics offer nesting habitat to dabbling ducks in the arid Intermountain West of the United States. Ecosphere 15:e70072.

Shmueli, G. To Explain or to Predict?. 2010. Statistical. Science 25:289 – 310.

Signer, J., Fieberg, J., and T. Avgar. 2017. Estimating utilization distributions from fitted step-selection functions. Ecosphere 8:e01771.

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

Signer, J., Fieberg, J., Reineking, B., Schlägel, U., Smith, B., Balkenhol, N., and T. Avgar. 2024. Simulating animal space use from fitted integrated Step-Selection Functions (iSSF). Methods in Ecology and Evolution 15:43-50.

Signer, J., Fieberg, J., Reineking, B., Schlägel, U., Smith, B., Balkenhol, N., and T. Avgar. 2024. Simulating animal space use from fitted integrated Step-Selection Functions (iSSF). Methods in Ecology and Evolution 15:43–50.

Street, G. M., Rodgers, A. R., Avgar, T., Vander Vennen, L. M., and J. M. Fryxell. 2017. Comparing resource selection and demographic models for predicting animal density. The Journal of Wildlife Management 81:16-25.

Street, G. M., Potts, J. R., Börger, L., Beasley, J. C., Demarais, S., Fryxell, J. M., McLoughlin, P. D., Monteith, K. L., Prokopenko, C. M., Ribeiro, et al. 2021. Solving the sample size problem for resource selection functions. Methods in Ecology and Evolution 12:2421-2431.

Suraci, J. P., Frank, L. G., Oriol-Cotterill, A., Ekwanga, S., Williams, T. M., and C. C.

Wilmers. 2019. Behavior-specific habitat selection by African lions may promote their persistence in a human-dominated landscape. Ecology 100:e02644.

Swihart, R. K., and N. A. Slade. 1997. On testing for independence of animal movements. Journal of Agricultural, Biological, and Environmental Statistics 48-63.

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

Tredennick, A. T., Hooker, G., Ellner, S. P., and P. B. Adler. 2021. A practical guide to selecting models for exploration, inference, and prediction in ecology. Ecology 102:e03336.

Tucker, M. A., Böhning-Gaese, K., Fagan, W. F., Fryxell, J. M., Van Moorter, B., Alberts, S. C., et al. 2018. Moving in the Anthropocene: Global reductions in terrestrial mammalian movements. Science 359:466-469.

van Beest, F. M., Uzal, A., Vander Wal, E., Laforge, M. P., Contasti, A. L., Colville, D., and P. D. McLoughlin. 2014. Increasing density leads to generalization in both coarse-grained habitat selection and fine-grained resource selection in a large mammal. Journal of Animal Ecology 83:147-156.

VanDerWal, J., Shoo, L. P., Johnson, C. N., and S. E. Williams. 2009. Abundance and the environmental niche: environmental suitability estimated from niche models predicts the upper limit of local abundance. The American Naturalist 174:282-291.

Warton, D. I., and L. C. Shepherd. 2010. Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology. The Annals of Applied Statistics 4:13831402.

Weber, M. M., Stevens, R. D., Diniz-Filho, J. A. F., and C. E. V. Grelle. 2017. Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. Ecography 40:817-828.

Westneat, D. F., Potts, L. J., Sasser, K. L. and J. D. Shaffer. 2019. Causes and consequences of phenotypic plasticity in complex environments. Trends in Ecology and Evolution 34:555–568.

Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecology 3:385-397.

Winter, V. A., B. J. Smith, D. J. Berger, R. B. Hart, J. Huang, K. Manlove, F. E. Buderman, and T. Avgar. 2024. Forecasting animal distribution through individual habitat selection: insights for population inference and transferable predictions. Ecography e07225.

Williams, B. K., Nichols, J. D., and M. J. Conroy. 2002. Analysis and management of animal populations. Elsevier Academic Press, San Diego, California, USA.

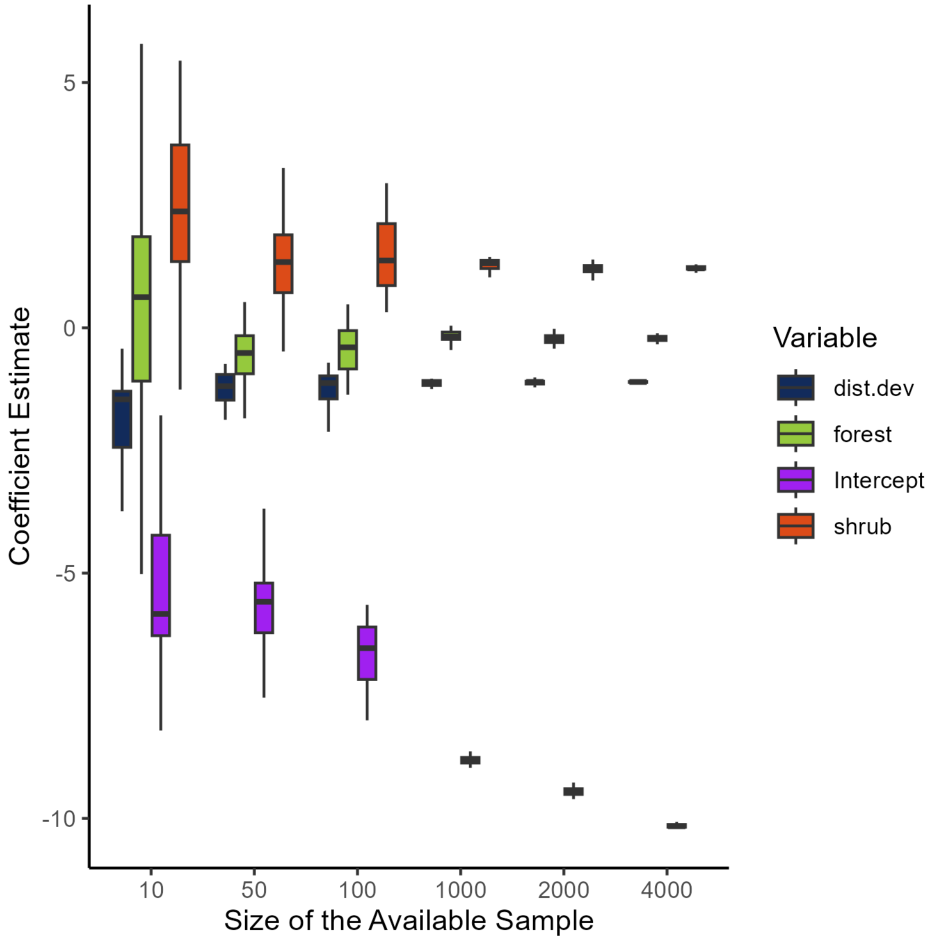

Figure 1. Variation of estimated slope coefficients of individual habitat selection for three variables when approximating the point process model with different sizes of the available sample. Note that as the available sample gets larger, the estimates stabilize and reduce in variation; this is specifically a traditional habitat selection, but the concept also applies to SSFs. The intercept does not stabilize but will decrease as the available sample size increases. Associated code replicates this figure.

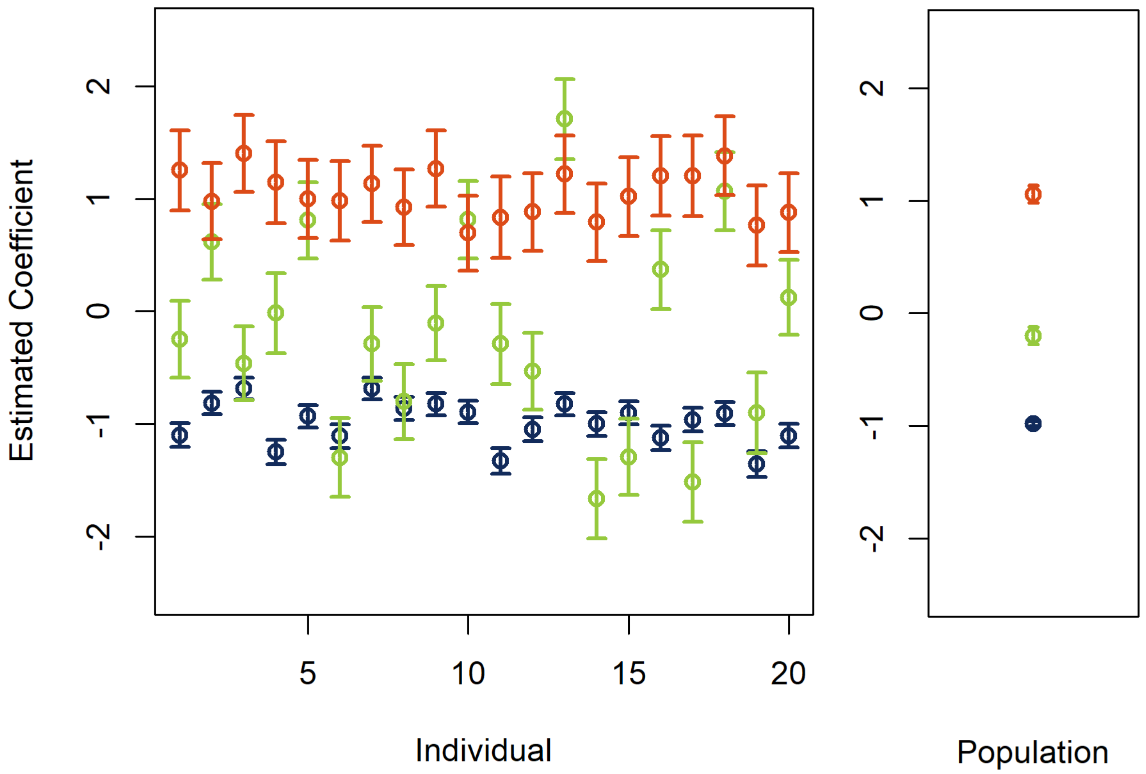

Figure 2. Individual estimates of slope coefficients and 95% confidence intervals for three spatial variables across 20 individuals from simulated data (left panel) compared with population-level estimates and 95% confidence intervals when pooling all individual data and ignoring individual variation in responses. Associated code replicates this figure.

Box 1: Guidelines for getting the basics right in habitat selection analyses

1. Think thoroughly about the question(s) of interest and the appropriate geographic and temporal scales in the specific study system before data collection begins.
    a. Which groups or classes are of interest in your system? Are you planning to break up the dataset by age/sex class/study area? How many individuals of each group must be sampled?
2. Define availability and determine whether you want to fit a HSF or SSF
    a. Are you interested in identifying where individuals select home ranges (second order HSF) or locations within that home range (third order HSF) or along a movement path (SSF)?
3. Decide whether you are interested in inference or prediction and how that impacts model building
    a. Inference is most straightforward when a single model is used with all relevant variables that capture the hypotheses of interest; it is important that the mechanisms motivating variables are given great thought. Prediction will often be improved by considering many models and does not necessarily require clear mechanisms behind including variables.
4. Understand and acknowledge the assumptions underlying your specific analysis
    a. Assumption of an exponential selection function that forces inference to the relative intensity of selection or relative probability of selection.
    b. Logistic regression and conditional logistic regression are used to approximate the underlying spatial or spatio-temporal point process model, respectively.
5. Conduct a sensitivity analysis for availability sample size.

    a. Fit the model, examine the estimated selection coefficients, and then do the whole exercise again with a larger number of sampled available points. If the selection coefficients do not change, the original number of available points sampled was sufficient. If they do change, keep repeating the exercise until they do not.

6. Account for animal behavior and heterogeneity in your analysis
    a. How quickly can the animal of interest move? Available locations from an animal location that are more than this distance within the considered time frame are not realistically available to the animal.
    b. Is it reasonable to consider consecutive points on an individual's track independent or are they temporally autocorrelated? It is important to account for this autocorrelation either by using a SSF, weighting techniques (Alston et al. 2022), or data thinning.
    c. Ecological literature leads us to *a priori* consider individual variation, which is an important source of variation, and thus our main sampling unit is the individual. When individuals are tracked, pooling data across individuals and ignoring individual variation leads to poor statistical conclusions. Accounting for individual level differences can be accounted for using a full random effects (intercept and slopes) structure.
    d. Consider all animal locations with regard to behavior and contextual environments (e.g., diel period, life stage, reproductive behaviors, etc.), which may require interactions among variables.
7. Correctly interpret your model parameters and inference

   a. To compute predicted values, we need to remove the intercept from the selection model (β_0) and exponentiate the summed combinations of effects and variable values ($e^{(\beta_1 \times X_1 + \beta_2 \times X_2 \ldots + \beta_n * X_n)}$) where e can be implemented in R using the exp() function and the X_1, X_2,...X_n values are the new values of predictor variables (i.e., not the original data input into the model).
   b. Inferring population density from habitat selection is most robust when there is auxiliary or additional validation of population density.

8. Decide whether model selection is necessary. It is reasonable to focus on fitting one model that captures all the variables that represent the complete hypotheses of interest. Do not assume model selection is a requirement.

Table 1: Resources organized by topic for further learning about habitat selection studies.

Topic	**Subtopics or Areas of Common Misconceptions**	**Resources**
Commonly-used terminology definitions	Glossary of terms	Northrup et al. (2022)
	Resource selection function	Manly et al. (2007)
	1st-4th order selection	Johnson (1980)
Utilization distributions	Estimating UD from SSF	Signer et al. (2017) Signer et al. (2024) Potts and Schlägel (2020). Potts and Börger (2023)
Predicting population abundance from habitat selection		Matthiopoulos et al. (2015), Boyce et al. (2016)
Scale in habitat selection		McGarigal et al. (2016), Paton and Matthioploulos (2016) Johnson (1980)
Defining availability in HSFs	Sample size of available points	Northrup et al. (2013)
	Movement ability of the	Beyer et al. (2010)

	animal	
	Temporal autocorrelation of consecutive locations	Alston et al. (2022), Noonan et al. (2019)
	Spatial heterogeneity/complexity	Fieberg et al. (2021), Fithian and Hastie (2013), Street et al. (2021)
Random effects and population-level inference	Random effects	Gelman and Hill (2007)
	Examples of hierarchical habitat selection models	Muff et al. (2020), Winter et al. (2024)
Sample size considerations	Number of used locations required per individual and number of individuals required/power analyses	Street et al. (2021)
Model selection		Fieberg, 2024 (Chapter 8), Fieberg and Johnson (2015), Gerber and Northrup (2020)
Interpretation of parameters		Fieberg et al. (2021)