Appendix 8 – Things to consider before using these methods

Modelling habitat selection using tracking data from central place foraging species: A practical guide for quantitative ecologists

***

Model choice should be guided by the ecological question being asked rather than statistical convenience and potential users should consider several key points before selecting a habitat selection model that is fit for their data (see Figure 5 in main text).

A first consideration is the **scale of inference**. In this guide, we focus on models that estimate patterns of space use at a broader behavioural scale, which we refer to as *regional-scale* selection. These approaches describe how animals use space across larger spatial domains (i.e., often corresponding to second-order selection at the home-range level), and are well suited to inform conservation decision-making (Guisan *et al.*, 2013; Porfirio *et al.*, 2014; Sofaer *et al.*, 2019). However, habitat selection can also be examined at finer spatial and temporal resolutions, where inference is explicitly conditioned on local behaviour (i.e., third- or fourth-order selection); this is the realm of step-selection functions (SSFs), which model the choices animals make between successive locations (Florko *et al.*, 2025). In brief, SSFs quantify how animals choose where to go next by comparing observed steps to a set of potential alternatives. In this sense, they define availability at each time step in a similar way as STPPs do, but the two approaches differ fundamentally in how this availability is constructed. In SSFs, availability is approximated by generating candidate steps from empirical distributions of step lengths and turning angles, which are typically estimated from the data. These represent locations that could be plausibly reached from the previous position, and inference is based on comparing observed steps to this finite set only. By contrast, STPPs define availability through a continuous movement kernel derived from an explicit movement model (e.g., a Brownian motion or an Ornstein–Uhlenbeck process; Appendices 3 & 5). This kernel specifies a probability density over all locations that could be reached within a given time interval, rather than a finite number of alternatives. As a result, STPPs treat availability as a continuous surface that can be evaluated across the entire spatial domain (therefore supporting regional-scale inference), with the movement kernel determining how strongly different locations contribute to availability at each time step. In practice, the STPP kernel rapidly down-weights distant locations and is therefore approximated locally for computational reasons (*via* local quadrature; see Appendices 3 & 5). However, the intensity surface is still defined over the full spatial domain and not restricted to a discrete set of points. For this reason, increasing the number of available points in an SSF only improves the numerical approximation of its likelihood, but does not cause the SSF to converge to an STPP. While outside the scope of this paper, SSFs play an important role in ecology and conservation (Thurfjell *et al.*, 2014), and are the subject of a substantial body of literature. We direct interested readers to Forester et al. (2009), Signer et al. (2017), Fieberg *et al.* (2021), Michelot et al. (2024), or Florko et al. (2025) for further details.

Second is the **inferential goal**. Clarifying whether the objective of the analysis lies in process understanding or prediction can help determine which formulation is most appropriate. Models such as IPPs, LGCPs, or mSTPPs operate on a single spatial domain and provide a time-averaged description of space use, which is often more appropriate for spatial prediction and facilitates extrapolation of habitat relationships across a broader range of environmental conditions. In principle, these models can learn from the full range of covariates present in the study region. By contrast, STPPs retain an explicit link to the timing and sequence of observations, making them well suited to understanding how animals respond to their immediate environment, but typically restrict inference to a narrower range of conditions encountered along movement paths. This need not be viewed as a strict trade-off; for example, it may be possible to carry forward the habitat relationships estimated from an STPP into an mSTPP to support spatial prediction, although further work is needed to formalise how this can be achieved. In our experience analysing real-world tracking data from seabirds, where both inference and prediction are of primary interest, such a complementary approach could be particularly valuable.

Closely related to this is **how availability is represented**. Because habitat selection is defined relative to availability (Eq. 1 in main text), different models can yield ostensibly similar results while answering subtly different ecological questions. In practice, availability may be approximated through sampled background points (e.g., as in logistic regression), spatial discretisation (e.g., as in gridded Poisson regression), implicit domain-wide definitions (e.g., as in IPPs or LGCPs), or movement-based constraints (e.g., as in STPPs or mSTPPs) (Figure 2 in main text). Each of these formulations alters the meaning of estimated effects and the scale at which inference applies. In logistic and gridded Poisson regression, coefficients describe relative use with respect to the background sample or the gridded map of the landscape. In IPPs and LGCPs, they quantify how the intensity of use changes across continuous space and can be interpreted as effects on the expected density of locations per unit area. In STPPs and mSTPPs, coefficients describe selection conditional on movement, reflecting how habitat influences space use once accessibility constraints have been taken into account (Table 1 in main text). Critically, models based on sampled or discretised availability are not inherently inferior to point processes but instead represent a simpler way of addressing the same ecological question. In many cases, inference based on relative use is in fact sufficient, particularly when the goal is to rank habitats or estimate covariate effects rather than to produce spatially explicit predictions. However, as demonstrated in our worked example, choosing simplicity carries the risk of understating uncertainty and producing estimates that diverge from those obtained under a full point process formulation.

A further consideration is how different models **handle dependence in the data** (Figure 2 in main text). Traditional RSFs commonly ignore autocorrelation or attempt to reduce it by thinning the data, whereas LGCPs absorb spatial dependence *via* latent random fields and STPPs incorporate mechanistic dependence through an underlying movement model. These approaches are not interchangeable, and each encodes a different ecological assumption about why observations are clustered in space and time. Choosing among them requires deciding whether clustering is best understood as a property of the environment, the movement process, or both.

Finally, practical considerations such as **data volume, computational cost, and software implementation** should not be overlooked. Movement-based models can be computationally demanding, particularly when kernels must be evaluated at each time step (e.g., in an STPP), while spatial models such as LGCPs require careful specification of spatial smooths and associated parameters (e.g., basis dimension, spline type). However, many of the approaches described here can be implemented within a common modelling framework such as *mgcv* (Appendices 3 & 5), allowing users to move between formulations without adopting entirely new software.

REFERENCES

Fieberg J, Signer J, Smith B, Avgar T (2021). A “How to” guide for interpreting parameters in habitat-selection analyses. *Journal of Animal Ecology*, **90**: 1027–1043. DOI: 10.1111/1365-2656.13441

Florko KRN, Togunov RR, Gryba R, Sidrow E, Ferguson SH, Yurkowski DJ, Auger-Méthé M (2025). An introduction to statistical models used to characterize species-habitat associations with animal movement data. *Movement Ecology*, **13**: 27. DOI: 10.1186/s40462-025-00549-2

Forester JD, Im HK, Rathouz PJ (2009). Accounting for animal movement in estimation of resource selection functions: Sampling and data analysis. *Ecology*, **90**: 3554–3565. DOI: 10.1890/08-0874.1

Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C, Martin TG, Rhodes JR, Maggini R, Setterfield SA, Elith J, Schwartz MW, Wintle BA, Broennimann O, Austin M, Ferrier S, Kearney MR, Possingham HP, Buckley YM (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, **16**: 1424–1435. DOI: 10.1111/ele.12189

Michelot T, Klappstein NJ, Potts JR, Fieberg J (2024). Understanding step selection analysis through numerical integration. *Methods in Ecology and Evolution*, **15**: 24–35. DOI: 10.1111/2041-210x.14248

Porfirio LL, Harris RMB, Lefroy EC, Hugh S, Gould SF, Lee G, Bindoff NL, Mackey B (2014). Improving the use of species distribution models in conservation planning and management under climate change. *PloS One*, **9**: e113749. DOI: 10.1371/journal.pone.0113749

Signer J, Fieberg J, Avgar T (2017). Estimating utilization distributions from fitted step-selection functions. *Ecosphere* , **8**: e01771. DOI: 10.1002/ecs2.1771

Sofaer HR, Jarnevich CS, Pearse IS, Smyth RL, Auer S, Cook GL, Edwards TC Jr, Guala GF, Howard TG, Morisette JT, Hamilton H (2019). Development and delivery of species distribution models to inform decision-making. *Bioscience*, **69**: 544–557. DOI: 10.1093/biosci/biz045

Thurfjell H, Ciuti S, Boyce MS (2014). Applications of step-selection functions in ecology and conservation. *Movement Ecology*, **2**: 4. DOI: 10.1186/2051-3933-2-4