



# Applying a Multiverse to Population Habitat Analyses

Benjamin Michael Marshall\*<sup>1</sup> and Alexander Bradley Duthie\*\*<sup>1</sup>

<sup>1</sup> Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling, FK9 4LA, Scotland, UK

\*benjaminmichaelmarshall@gmail.com

\*\*alexander.duthie@stir.ac.uk

---

## Abstract

abc

---

## Keywords

Movement ecology, simulation, compana, resource selection functions, step selection function, habitat preference, habitat selection, animal movement, multiverse, research choice, researcher degrees of freedom,



## 1 Introduction

abc

## 2 Methods

### 2.1 Simulating the Scenarios

- Landscape simulation.
- `abmAnimalMovement` settings

We used the animal movement using `abmAnimalMovement` v.0.1.3.0 (Marshall & Duthie, 2022) to simulate the movement data of an animal with a predefined (i.e., known) habitat preference. The `abmAnimalMovement` package provides an agent-based approach to simulating terrestrial animal movement using raster environmental data to guide the animals decisions. We used the `NLMR` v.1.1.1 package (Sciaini et al., 2018) to generate the three required resource/environmental rasters: movement resistance, foraging quality, and shelter site quality. The `abmAnimalMovement` package has systems for simulating activity cycles, three separate behavioural states (differing in movement characteristics and resource prioritisation), and site fidelity. For the purposes of this study we used one of the pre-created example pseudo-species: Badger, described in the package manuscript. In brief, the badger is a terrestrial species occupying several shelter sites, with a 8-12 hour activity cycle with minor seasonal variation, and is subject to differing movement resistance across the landscape.

For the purposes of the analysis we simplified the landscape information into categories –akin to the sort of land-use information more frequently available to researchers of animal movement. We focused on foraging quality because it influences the greatest amounts of movement compared to sheltering or exploratory movements. We converted the continuous foraging quality raster into a binary, where higher quality areas (greater than 0.5) are classed as 2, and lower quality areas as 0.

We used that simulated landscape and `abmAnimalMovement` to simulate a population of \_\_\_, that later can be sampled from. All individuals of this population had the same simulation settings apart from starting location. Therefore, the variation between individuals is due to stochasticity rather than variation in the predefined habitat preference.

### 2.2 Sampling and Analysis Options

- targets construction

To manage the sampling of the population and the compounding growth of subsequent analysis decisions, we used the `targets` v.0.14.2 and `tarchetypes` v.0.7.4 R packages (Landau, 2021a,b). These packages allowed branching workflow pipeline, while keeping track of object creation thereby optimising the compute time required to explore the multiverse of analysis choice.

#### 2.2.1 Sampling

- tracking regime
- sample size

The first decision in most animal movement studies will be concerning tracking regime. This decision is frequently dictated by more practical considerations such as anatomy and behaviour of the animal, cost of the tracking devices, and environmental factors. Here we aimed to cover a range of tracking regimes that vary in the frequency of location fixes ( ), and the total duration of tracking ( ). We created subsampled datasets based on every combination of tracking frequencies and durations, provided they would result in greater than 30 datapoints per individual.

An important component of assessing population level habitat selection is the number of individuals included in analysis. Therefore, we randomly generated a number of samples from our population of \_\_\_ simulated individuals. We varied these samples sizes from \_\_\_ to \_\_\_ individuals, and ran \_\_\_ repeats for each size. A sample never mixed tracking regimes.



### 2.2.2 Analysis

Building on the decisions concerning tracking regime and population sampling, our multiverse expanded dramatically by exploring four primary analysis routes. These routes included an area based approach using Compana analysis, and three step-based approaches including averaged individual step-selection models, two-step conditional regression models, and a Poisson model.

- area based: compana, area method, contour, available points, space sampling, type II/III, compana test *adehabitatHS* v.0.3.16 (Calenge & Mathieu Basille, 2023), ctmm package v.0.6.1 (Fleming & Calabrese, 2023)
- ssf: Model Formula (SSF or iSSF), Available Points per Step, Distribution of Step Lengths, Distribution of Turn Angles, Model Averaging Method

amt v.0.1.7 (Signer, Fieberg & Avgar, 2019)

- twoStep: Model Formula (SSF or iSSF), Available Points per Step, Distribution of Step Lengths, Distribution of Turn Angles

TwoStepCLogit v.1.2.5 (Craiu et al., 2016)

- poisson: Model Formula (SSF or iSSF), Available Points per Step, Distribution of Step Lengths, Distribution of Turn Angles

INLA v.23.4.24 (Rue, Martino & Chopin, 2009; Lindgren, Rue & Lindström, 2011; Martins et al., 2013; Rue et al., 2017; Kourounis, Fuchs & Schenk, 2018)

Muff, Signer & Fieberg (2020)

## 2.3 Assessing the multiverse

- spec curves
- brm models: one per each analysis method

Specification curves provide an overview of the estimates of a given range of analyses. Tighter more steep curves suggest greater agreement between all the analysis end points. Here we also plotted the estimates against the different decisions that results in the estimates, allowing direct comparison on how the decision impacts the variation in the estimates.

To better detect the impact of decisions, while accounting for the random variation stemming from the differences in individuals/samples, we ran a number of Bayesian Regression Models. The Bayesian Regression Models aimed to describe how much of the deviation from a median answer could be explained by the various sampling and analysis decisions. For each analysis route we ran a model that included tracking frequency, tracking duration, sample size, and all the corresponding analysis choices. All continuous variables were scaled to help determine their relative importance to each other. For the area based Compana approach the population effects included: the continuous variable contourScaled; and the categoric predictors samplingPatternst, testrandomisation, areaMethodMCP, and typeIII. For the step-based approaches they all included: modelFormulamf.ss, stepDistgamma, turnDistvonmises. The step-selection model approach also included: averagingMethodNaiveaverage.

## 3 Results

### 3.1 Specification Curves

(Fig. 1).

(Fig. 2).

(Fig. 3).

(Fig. 4).

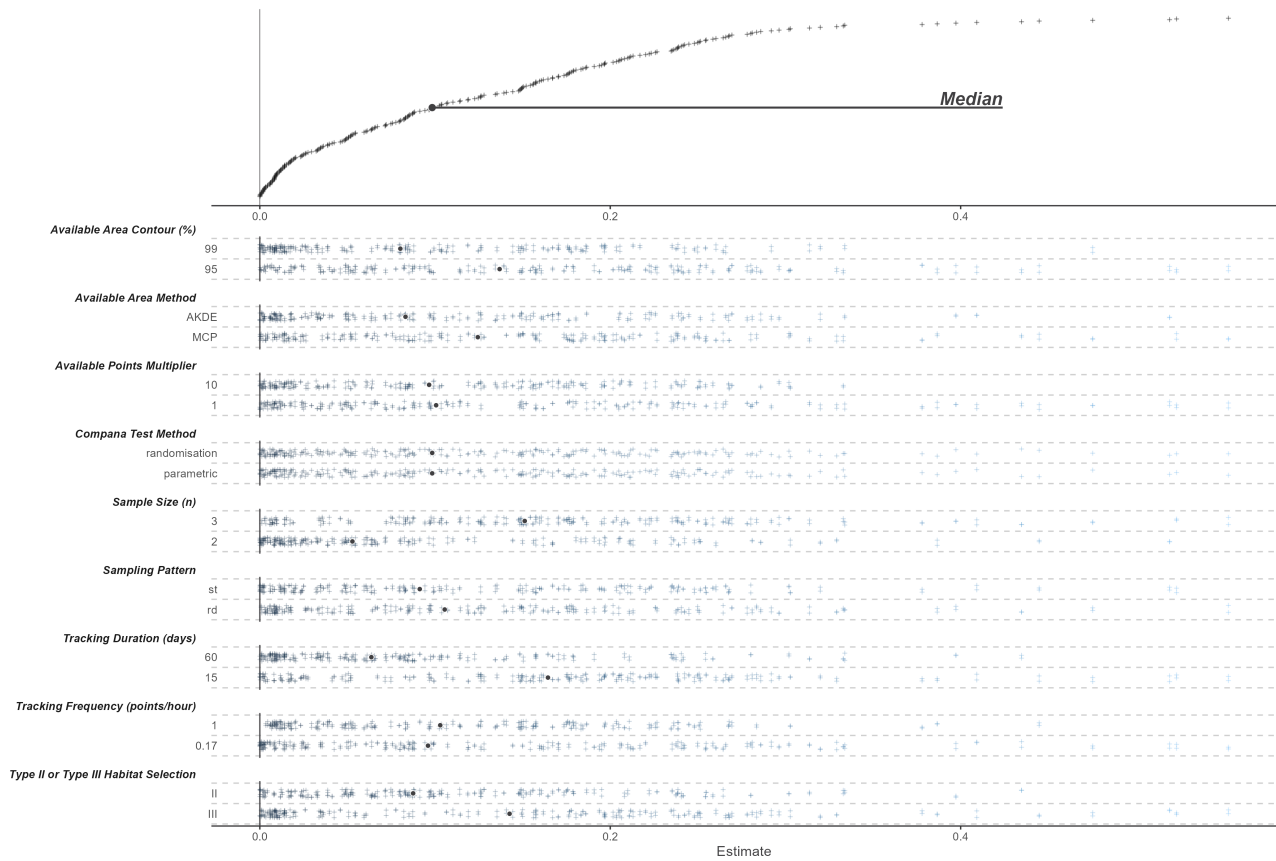


Figure 1. Spec curve

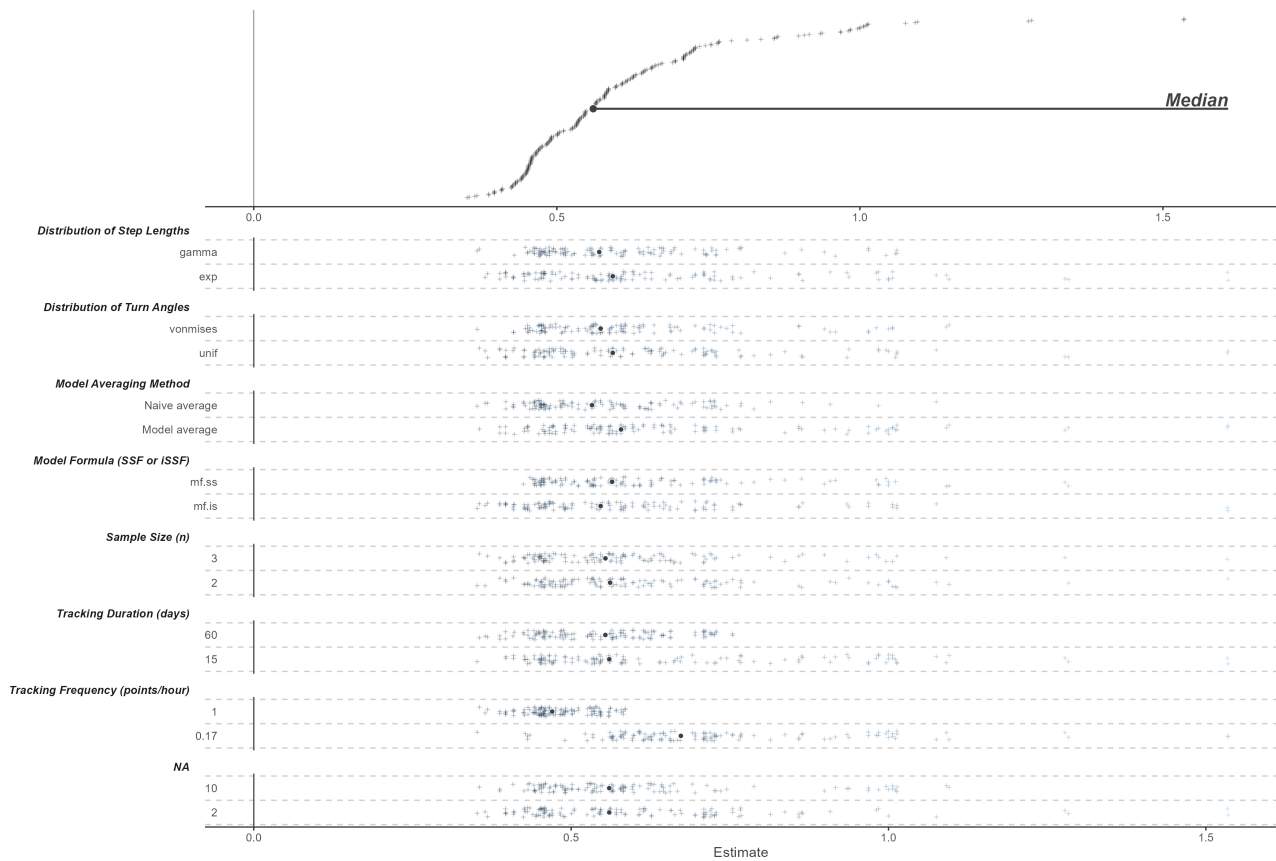


Figure 2. Spec curve

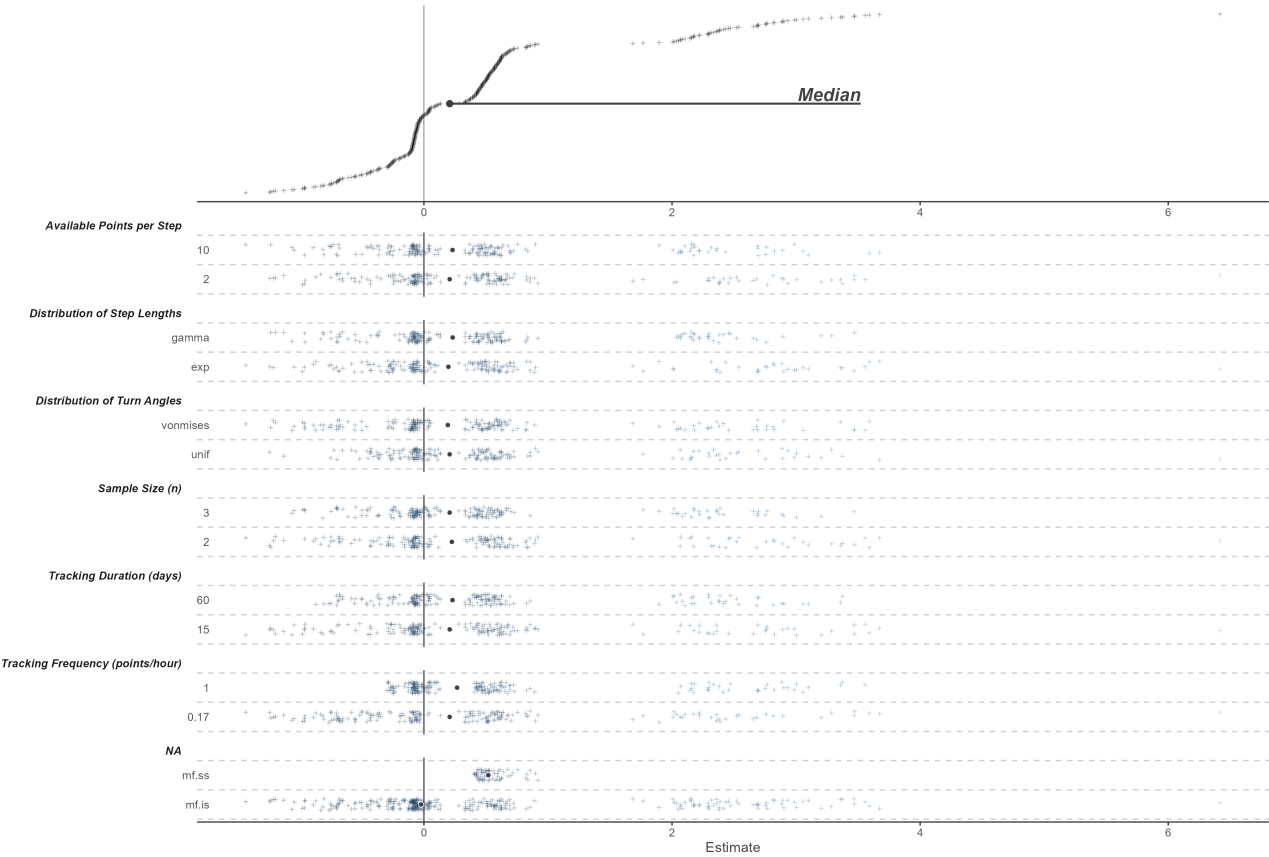


Figure 3. Spec curve

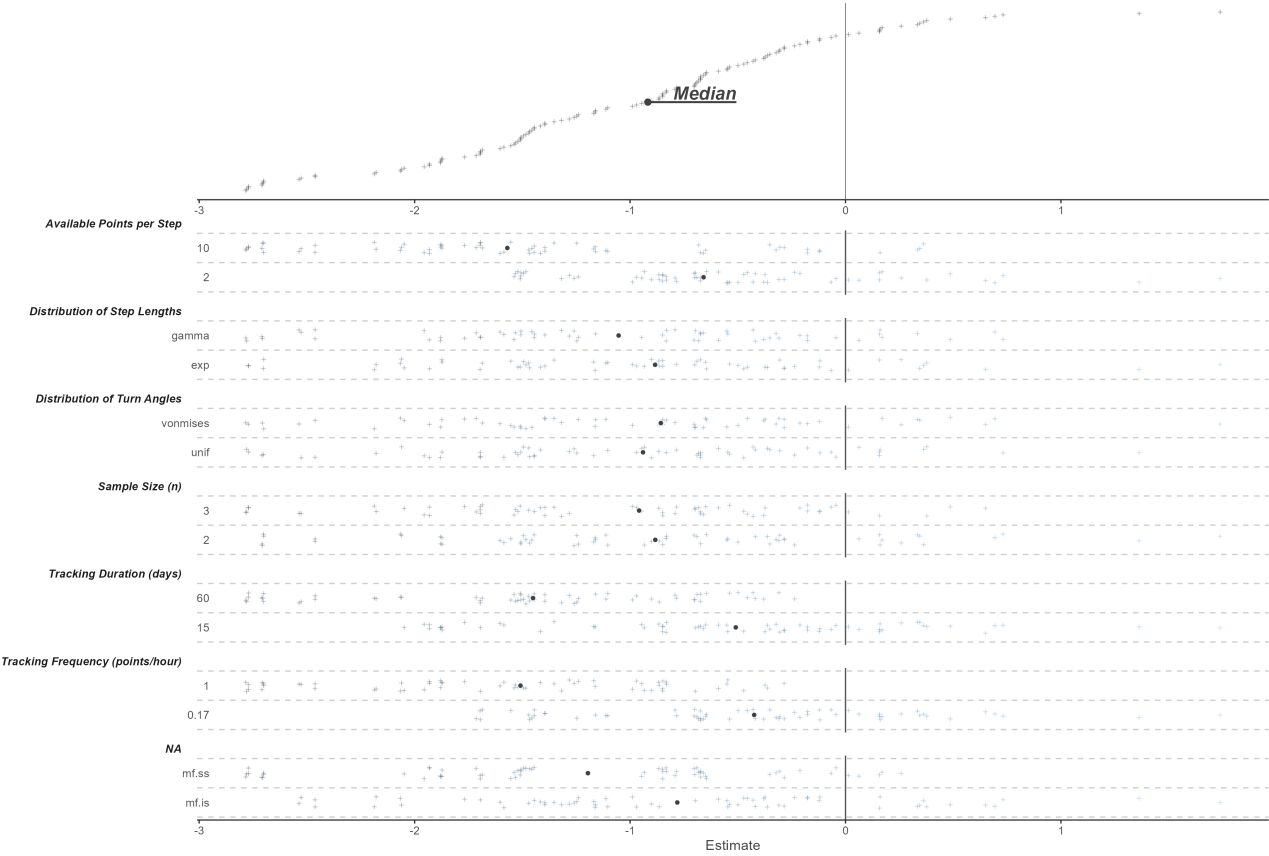


Figure 4. Spec curve

## 3.2 Model Results

The conditional  $R^2$  values differed for the three models. The Compana results model had a conditional  $R^2$  of 0.33; whereas the SSF model returned 0.59, and the Poisson model returned 0.94.

The marginal  $R^2$  represents the bulk of the conditional  $R^2$  suggesting an important role for the fixed/population effects. The Compana results model had a conditional  $R^2$  of 0.48; whereas the SSF model returned 0.51, and the Poisson model returned 0.83.

The sample size was negatively correlated with deviation from the median estimate ( $\beta$  -0.03; 95% HDI -1.15 - 1.8).

(Fig. 5).

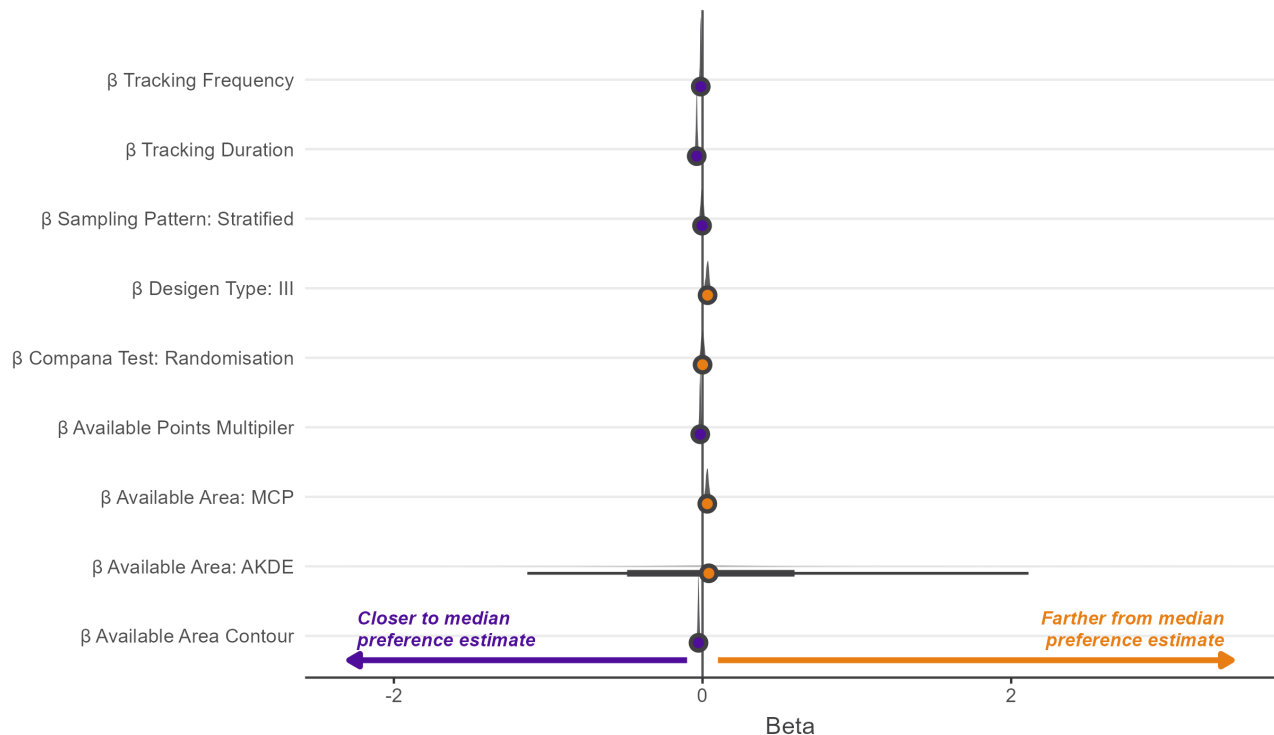


Figure 5. Beta coefs

(Fig. 6).

(Fig. 7).

(Fig. 8).

## 4 Discussion

### 4.1 Limitations

### 4.2 Conclusions

## 5 Acknowledgements

BMM was funded by the Natural Environment Research Council (NERC) via the IAPETUS2 Doctoral Training Partnership.

## 6 Software availability

In addition to packages already mentioned in the methods we also used the following.

We used R v.4.2.2 (R Core Team, 2023) via RStudio v.2023.6.2.561 (RStudio Team, 2022). We used here v.1.0.1 (Müller, 2020) and qs v.0.25.5 (Ching, 2023) to manage directory addresses and saved objects.

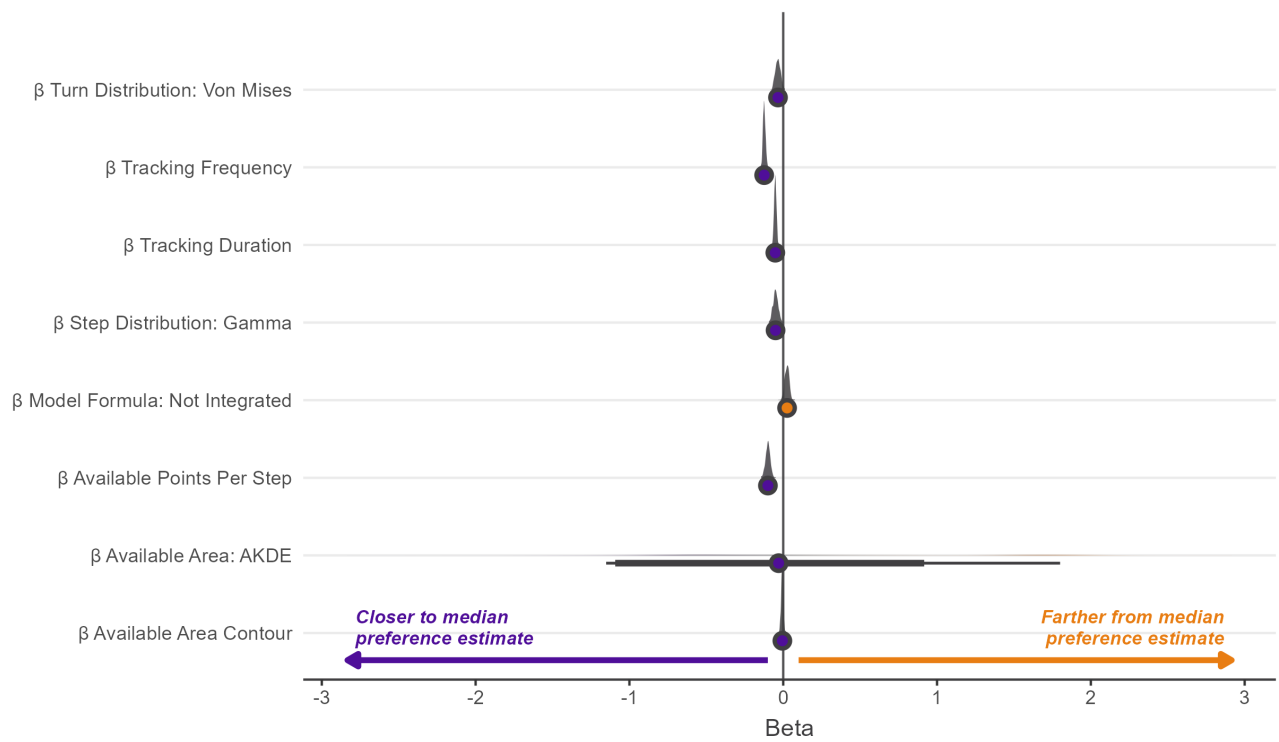


Figure 6. Beta coeffs

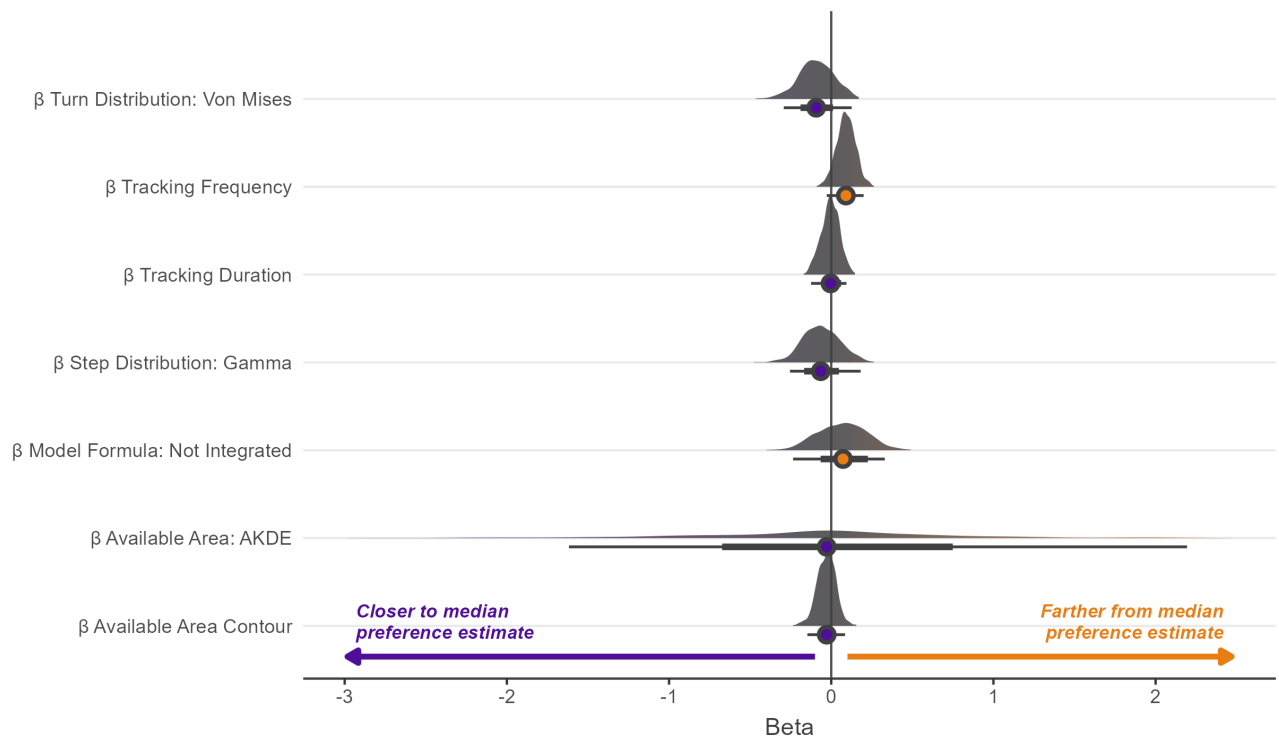
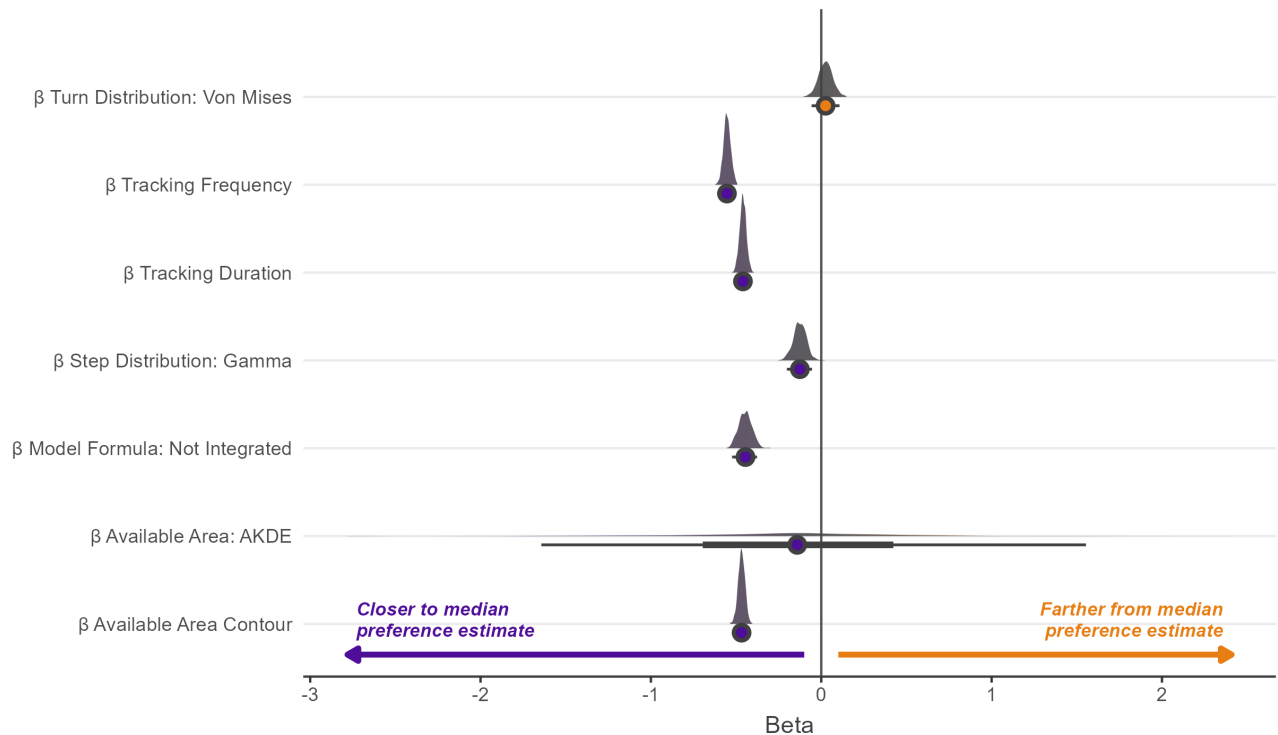


Figure 7. Beta coeffs



**Figure 8.** Beta coefs

We used *raster* v.3.6.14 (Hijmans, 2023) and *RandomFields* v.3.3.14 (Schlather et al., 2015) to aid landscape raster creation alongside *NLMR* v.1.1.1 (Sciaini et al., 2018).

We used *ggplot2* v.3.4.2 for creating figures (Wickham, 2016), with the expansions: *patchwork* v.1.1.2 (Pedersen, 2022), *ggridges* v.0.5.4 (Wilke, 2022), and *ggdist* v.3.2.0 (Kay, 2023a).

We used *brms* v.2.19.0 (Bürkner, 2021) to run Bayesian models, with diagnostics generated using *bayesplot* v.1.10.0 (Gabry et al., 2019), *tidybayes* v.3.0.2 (Kay, 2023b), and *performance* v.0.10.2 (Lüdtke et al., 2021).

We used the *dplyr* v.1.0.10 (Wickham et al., 2023), *tibble* v.3.1.8 (Müller & Wickham, 2023), and *stringr* v.1.5.0 (Wickham, 2022) packages for data manipulation.

We used *sp* v.1.5.1 (Bivand, Pebesma & Gomez-Rubio, 2013), *adehabitatHR* v.0.4.20 (Calenge & Scott Fortmann-Roe, 2023), *move* v.4.1.12 (Kranstauber, Smolla & Scharf, 2023) for manipulation of spatial data and estimation of space use not otherwise mentioned in the methods.

We used *rmarkdown* v.2.19 (Xie, Allaire & Golemund, 2018; Xie, Dervieux & Riederer, 2020; Allaire et al., 2023), *bookdown* v.0.33 (Xie, 2016, 2022), *tinytex* v.0.44 (Xie, 2019, 2023a), and *knitr* v.1.41 (Xie, 2014, 2015, 2023b) packages to generate type-set outputs.

We generated R package citations with the aid of *grateful* v.0.1.13 (Francisco Rodríguez-Sánchez, Connor P Jackson & Shaurita D. Hutchins, 2023).

## 7 Data availability

## 8 Supplementary Material

### References

- Allaire J, Xie Y, Dervieux C, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R. 2023. *rmarkdown: Dynamic documents for r*.
- Bivand RS, Pebesma E, Gomez-Rubio V. 2013. *Applied spatial data analysis with R, second edition*. Springer, NY.
- Bürkner P-C. 2021. Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software* 100:1–54. DOI: [10.18637/jss.v100.i05](https://doi.org/10.18637/jss.v100.i05).
- Calenge C, Mathieu Basille contributions from. 2023. *adehabitatHS: Analysis of habitat selection by animals*.





- Calenge C, Scott Fortmann-Roe contributions from. 2023. *adehabitatHR: Home range estimation*.
- Ching T. 2023. *qs: Quick serialization of r objects*.
- Craiu RV, Duchesne T, Fortin D, Baillargeon S. 2016. *TwoStepCLogit: Conditional logistic regression: A two-step estimation method*.
- Fleming CH, Calabrese JM. 2023. *ctmm: Continuous-time movement modeling*.
- Francisco Rodríguez-Sánchez, Connor P Jackson, Shaurita D. Hutchins. 2023. *grateful: Facilitate citation of r packages*.
- Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. 2019. Visualization in bayesian workflow. *J. R. Stat. Soc. A* 182:389–402. DOI: [10.1111/rssa.12378](https://doi.org/10.1111/rssa.12378).
- Hijmans RJ. 2023. *raster: Geographic data analysis and modeling*.
- Kay M. 2023a. *ggdist: Visualizations of distributions and uncertainty*. DOI: [10.5281/zenodo.3879620](https://doi.org/10.5281/zenodo.3879620).
- Kay M. 2023b. *tidybayes: Tidy data and geoms for Bayesian models*. DOI: [10.5281/zenodo.1308151](https://doi.org/10.5281/zenodo.1308151).
- Kourounis D, Fuchs A, Schenk O. 2018. Towards the next generation of multiperiod optimal power flow solvers. *IEEE Transactions on Power Systems* PP:1–10.
- Kranstauber B, Smolla M, Scharf AK. 2023. *move: Visualizing and analyzing animal track data*.
- Landau WM. 2021b. *Tarchetypes: Archetypes for targets*.
- Landau WM. 2021a. The targets r package: A dynamic make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software* 6:2959.
- Lindgren F, Rue H, Lindström J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach (with discussion). *Journal of the Royal Statistical Society B* 73:423–498.
- Lüdtke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D. 2021. performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software* 6:3139. DOI: [10.21105/joss.03139](https://doi.org/10.21105/joss.03139).
- Marshall BM, Duthie AB. 2022. *abmAnimalMovement: An r package for simulating animal movement using an agent-based model*. *F1000* 0:0.
- Martins TG, Simpson D, Lindgren F, Rue H. 2013. Bayesian computing with INLA: New features. *Computational Statistics and Data Analysis* 67:68–83.
- Muff S, Signer J, Fieberg J. 2020. Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using bayesian or frequentist computation. *Journal of Animal Ecology* 89:80–92. DOI: [10.1111/1365-2656.13087](https://doi.org/10.1111/1365-2656.13087).
- Müller K. 2020. *here: A simpler way to find your files*.
- Müller K, Wickham H. 2023. *tibble: Simple data frames*.
- Pedersen TL. 2022. *Patchwork: The composer of plots*.
- R Core Team. 2023. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- RStudio Team. 2022. *RStudio: Integrated development environment for r*. Boston, MA: RStudio, PBC.
- Rue H, Martino S, Chopin N. 2009. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society B* 71:319–392.
- Rue H, Riebler AI, Sørbye SH, Illian JB, Simpson DP, Lindgren FK. 2017. Bayesian computing with INLA: A review. *Annual Reviews of Statistics and Its Applications* 4:395–421.
- Schlather M, Malinowski A, Menck PJ, Oesting M, Strokorb K. 2015. Analysis, simulation and prediction of multivariate random fields with package RandomFields. *Journal of Statistical Software* 63:1–25.
- Sciaini M, Fritsch M, Scherer C, Simpkins CE. 2018. NLMR and landscapetools: An integrated environment for simulating and modifying neutral landscape models in r. *Methods in Ecology and Evolution* 00:1–9.
- Signer J, Fieberg J, Avgar T. 2019. Animal movement tools (amt): R package for managing tracking data and conducting habitat selection analyses. *Ecology and Evolution* 9:880–890.
- Wickham H. 2016. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
- Wickham H. 2022. *stringr: Simple, consistent wrappers for common string operations*.
- Wickham H, François R, Henry L, Müller K, Vaughan D. 2023. *dplyr: A grammar of data manipulation*.
- Wilke CO. 2022. *Ggridges: Ridgeline plots in 'ggplot2'*.
- Xie Y. 2014. knitr: A comprehensive tool for reproducible research in R. In: Stodden V, Leisch F, Peng RD eds. *Implementing reproducible computational research*. Chapman; Hall/CRC,.
- Xie Y. 2015. *Dynamic documents with R and knitr*. Boca Raton, Florida: Chapman; Hall/CRC.
- Xie Y. 2016. *bookdown: Authoring books and technical documents with R markdown*. Boca Raton, Florida: Chapman; Hall/CRC.
- Xie Y. 2019. *TinyTeX: A lightweight, cross-platform, and easy-to-maintain LaTeX distribution based on TeX live*. *TUGboat* 40:30–32.
- Xie Y. 2022. *Bookdown: Authoring books and technical documents with r markdown*.
- Xie Y. 2023b. *knitr: A general-purpose package for dynamic report generation in r*.
- Xie Y. 2023a. *tinytex: Helper functions to install and maintain TeX live, and compile LaTeX documents*.
- Xie Y, Allaire JJ, Grolemond G. 2018. *R markdown: The definitive guide*. Boca Raton, Florida: Chapman; Hall/CRC.



Xie Y, Dervieux C, Riederer E. 2020. *R markdown cookbook*. Boca Raton, Florida: Chapman; Hall/CRC.