



Applying a Multiverse to Population Habitat Analyses

Benjamin Michael Marshall*¹ and Alexander Bradley Duthie**¹

¹ 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

```
optionsCompleteList <- readRDS(here::here("data", "optionsCompleteList.rds"))
```

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 3, 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 (1 to 0.17 points per hour), and the total duration of tracking (15 to 60 days). We created sub-sampled datasets based on every combination of tracking frequencies and durations, provided they would result in greater than 30 data points 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 3 simulated individuals. We varied these samples sizes from 2 to 3 individuals, and ran 1 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

We ran the Compana analysis (Compositional Analysis of Habitat Use) using the *adehabitatHS* v.0.3.16 (Calenge & Mathieu Basille, 2023). Compana allows for the assessment of habitat selection of multiple animals in defined habitat types; it also allows habitat selection to be estimated at different scales. We explored two scales in the multiverse: selection within the home range (Type III), and selection within an overall population available area (Type II). The former design (Type III) requires availability habitats to be defined on an individual-by-individual basis, while the latter (Type II) design uses a summarised population level availability.

To define availability, we created home range polygons then sampled points within, recovering the corresponding habitat at those points. The home range (aka availability) polygons can be generated via many different processes. We selected to explore MCP and AKDE (Minimum Convex Polygons, Autocorrelated Kernel Density Estimators). We used the *ctmm* v.0.6.1 (Fleming & Calabrese, 2023) package for the creation of AKDEs, and the *adehabitatHR* v.0.4.20 (Calenge & Scott Fortmann-Roe, 2023) package for the MCPs. The AKDEs required the fitting of a movement model, we fit multiple models and selected the top scoring model by AIC. For the best model we followed recommendations of , opting to run AKDEs as weighted and using (PREML???) as both these options tend to be the most robust in scenarios of lower tracking frequencies. There are many options involved in the creation of AKDEs that could shape habitat availability and therefore estimations of selection. We avoided a deeper exploration as the impact of the internal AKDE choices are likely smaller than those between area methods (e.g., compared to MCPs), and the AKDE creation can more guided by the underlying data. By comparison the MCPs have very few options impacting their creation. Mainly MCP size varies based on the percentage of outliers excluded from the polygon area. For all area methods we vary this percentage: 95%, 99%, that correspond to increasingly large areas included in the availability polygon (i.e., covering areas that are less used by the individual). We selected these contour values as they are the most frequently used in spatial studies

Once we had created all areas for each individual, for every combination of tracking regime (duration x frequency, total regimes: 4), we created the sample/population available areas by combining all polygons. We used this combined polygon for the type II design, where all individuals have the same availability.

We used the availability polygons to define the extent within which we sampled habitat characteristics. The habitat characteristics were extracted from the classified landscape raster at various points. We varied both the number of points generated (1 to 10), as well as the pattern of how they were distributed within the polygon (random or stratified).

The final decision in the area based method approach was whether the Compana analysis was tested using randomisation or in a parametric fashion. The former uses randomisation tests to estimate the habitat selection, while the latter uses chi-squared.

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

The other analysis pathways do not rely on an available area, instead they focus on randomly generated steps that mirror the observed movement of the animal. As such the decisions involved in running the step-selection, two-step conditional regression, and Poisson models are largely the same.

The first decision is the number of random steps generated per observed step; we ranged this from 2 to 10. The generation of random steps is guided by two distributions, one describing the step lengths from the last observed location, and another describing the turn angle from the last observed direction of travel. We chose to explore the impacts of different distributions. For the step length we explored gamma and exponential distributions; for the turn angle we explored Von Mises and uniform.

All the step-based methods require the definition of a formula, where use/available is predicted by the habitat characteristics at those locations. Model formulation opens up many possible alternative routes of analysis, but for this



exploration we focus on the impacts of integrating step and turn angle interactions with habitat characteristics. We ran two versions of all models, one with no interactions and one where both step lengths and turn angles interaction with the habitat classification. Previous work has suggested that the integrated (i.e., model with interactions) performs better

A key component of all the population approaches is how they translate a highly structured dataset into a overall summary. We ran individual step selection models using the *amt* v.0.1.7 (Signer, Fieberg & Avgar, 2019) package, to explore rudimentary summary methods. The individual step selection models need to be averaged in some manner to extract a population mean selection. We explore two simple options for achieving this: a naive mean of the final estimates (i.e., a mean of the estimated betas), and a model average using the *MuMIn* v.1.47.5 package (Bartón, 2023). The *MuMIn* model average approach is weighted based on AICc. However, AICc is not directly comparable between models fitted with different datasets. As the model formulas are identically complex and the datasets equal in data quantity, the differences in AIC are purely describing the goodness of fit to the different datasets. Therefore, the model average provided is weighted by model fit.

The difficulties model averaging models based on different data has led to the creation of models that account for the individual variability in the model formula. A Two-Step Conditional Regression is one such solution, where strata (time steps) and clusters (individual animals) are accounted for. We implemented these models using the *TwoStepClogit* v.1.2.5 (Craiu et al., 2016) package.

More recently Muff, Signer & Fieberg (2020) have suggested a reformulation of the step selection models to provide faster reliable estimations of population habitat selection. The suggestion is to reformulate the model as a Poisson model, but critically with fixed . To avoid the slow convergence of MCMC estimation, Muff, Signer & Fieberg (2020) make use of integrated nested Laplace approximation (INLA) to approximate Bayesian inference. We followed the code provided by Muff, Signer & Fieberg (2020) and made use of *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) to run Poisson models to estimate habitat selection.

2.3 Assessing the multiverse

- spec curves

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.

- brm models: one per each analysis method

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: $(x - \text{mean}(x))/\text{sd}(x)$. For the area based Compana approach the population effects included: the continuous variable contour (continuous); and the categoric predictors sampling pattern (random, stratified), test (randomisation, parametric), area method (AKDE, MCP), and design type (II, III). For the step-based approaches they all included: model formula (integrated, non-integrated), step distribution (gamma, exponential), turn distribution (Von Mises, uniform). The step-selection model approach also included: averaging method (naive, MuMIn average).

We ran the Bayesian Regression Models using the *brms* v.2.19.0 (Bürkner, 2021) package. Once complete, we checked model convergence using *rhat* values, *acf*, and trace plots. After assessment of the model convergence measures, we modified the running parameters .

3 Results

3.1 Specification Curves

(Fig. 1).

(Fig. 2).

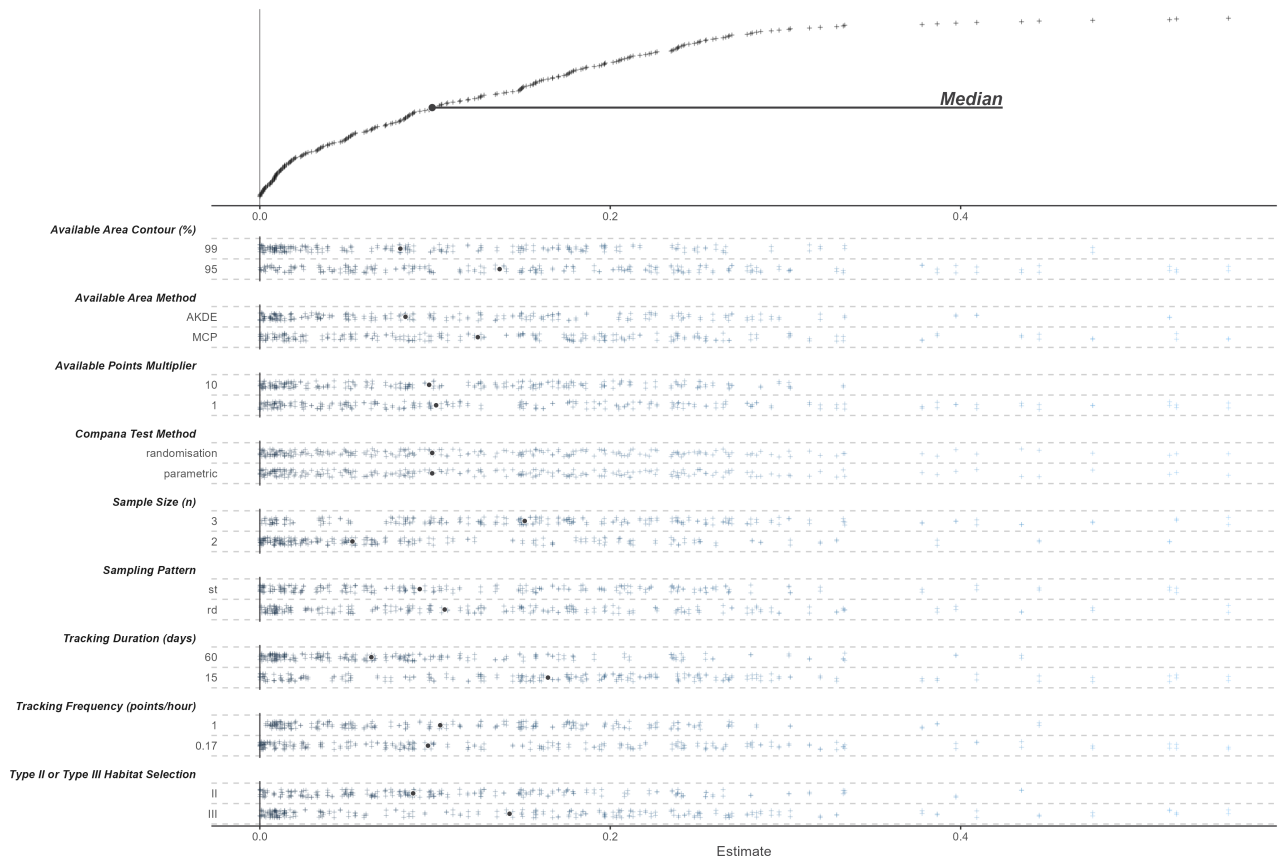


Figure 1. Spec curve

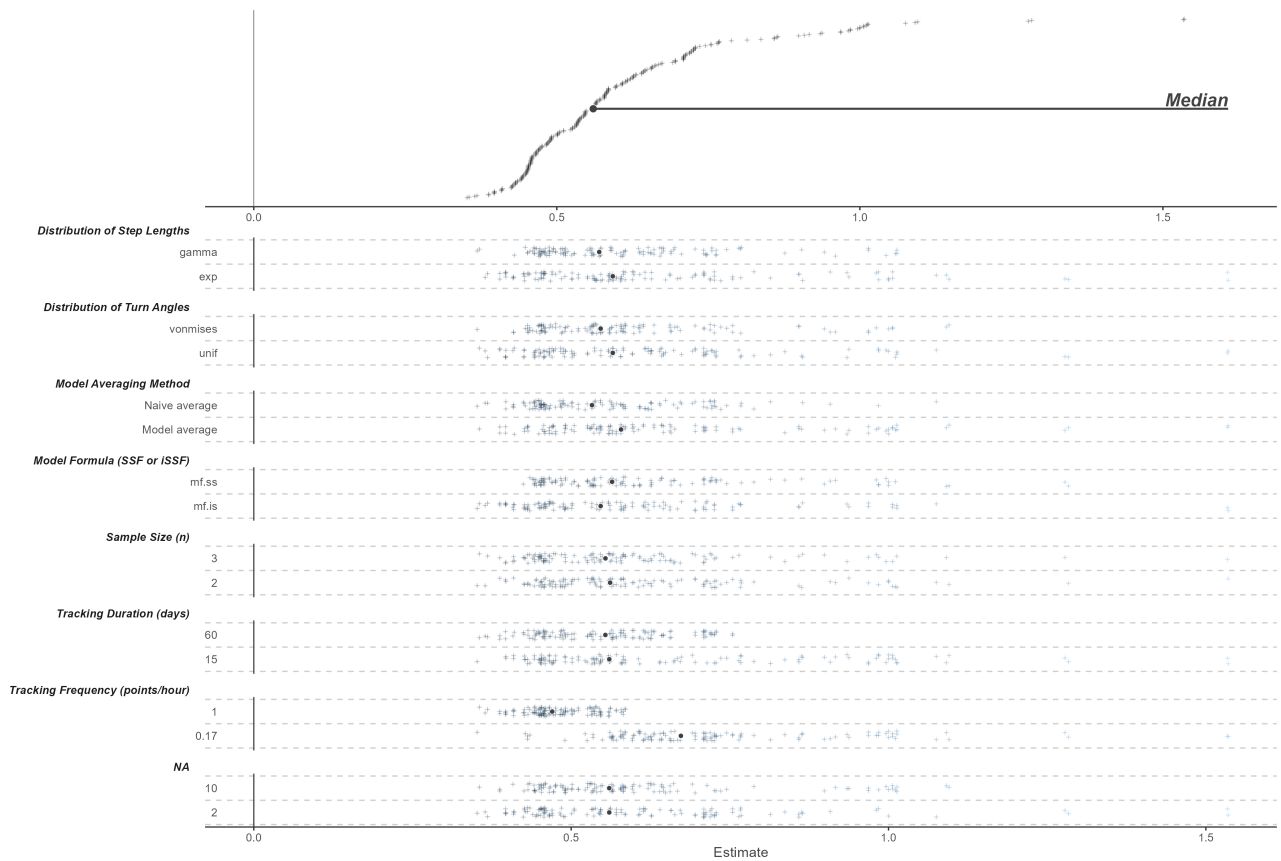


Figure 2. Spec curve



(Fig. 3).

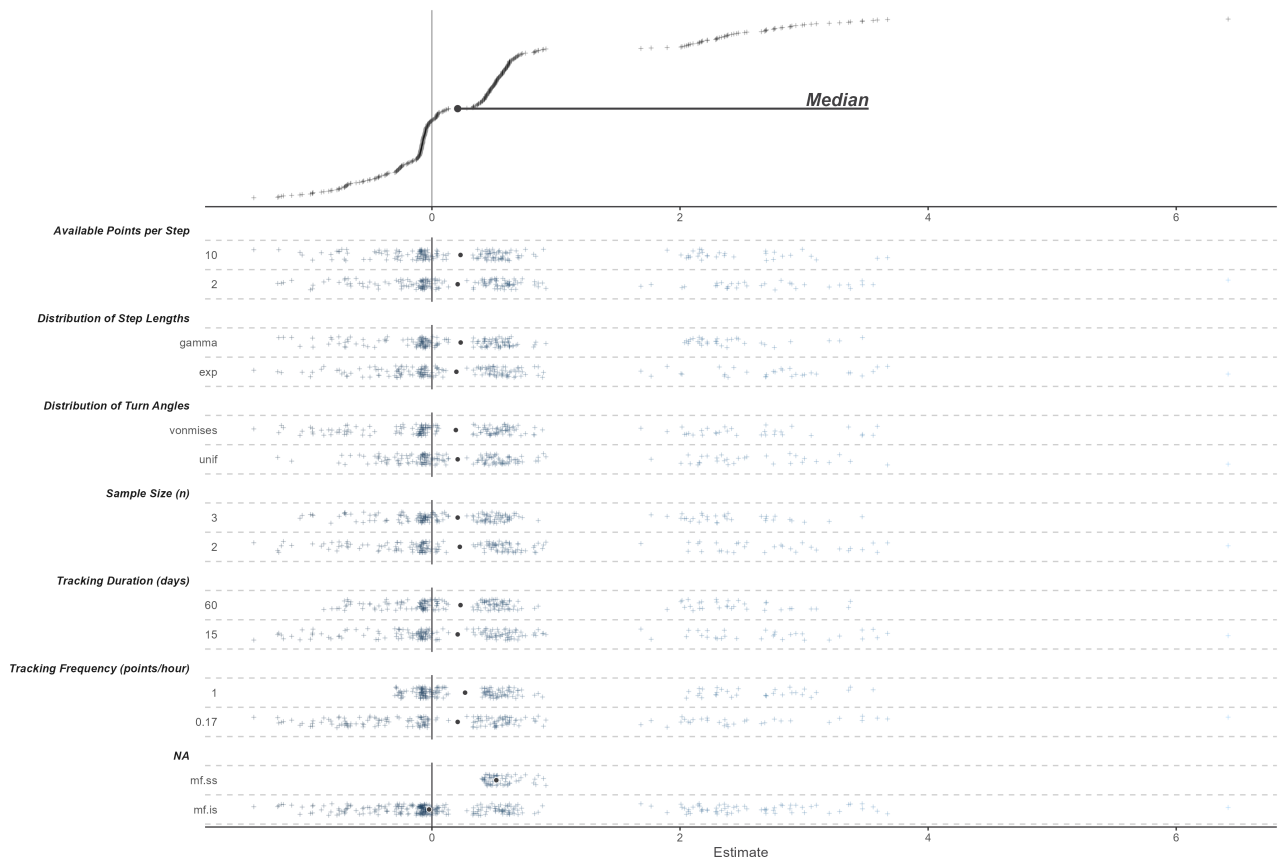


Figure 3. Spec curve

(Fig. 4).

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 (β -0.03; 95% HDI -1.15 - 1.8).

(Fig. 5).

(Fig. 6).

(Fig. 7).

(Fig. 8).

4 Discussion

4.1 Limitations

4.2 Conclusions

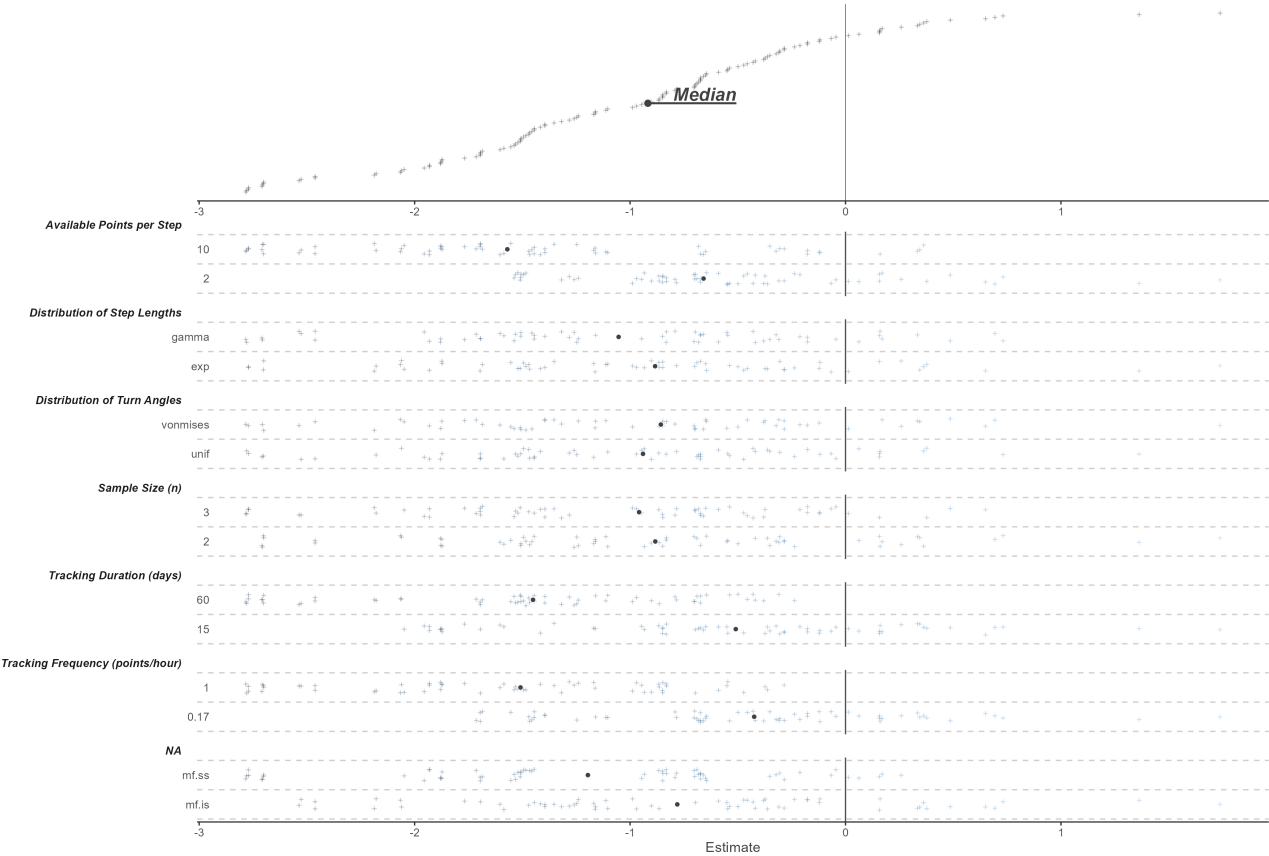


Figure 4. Spec curve

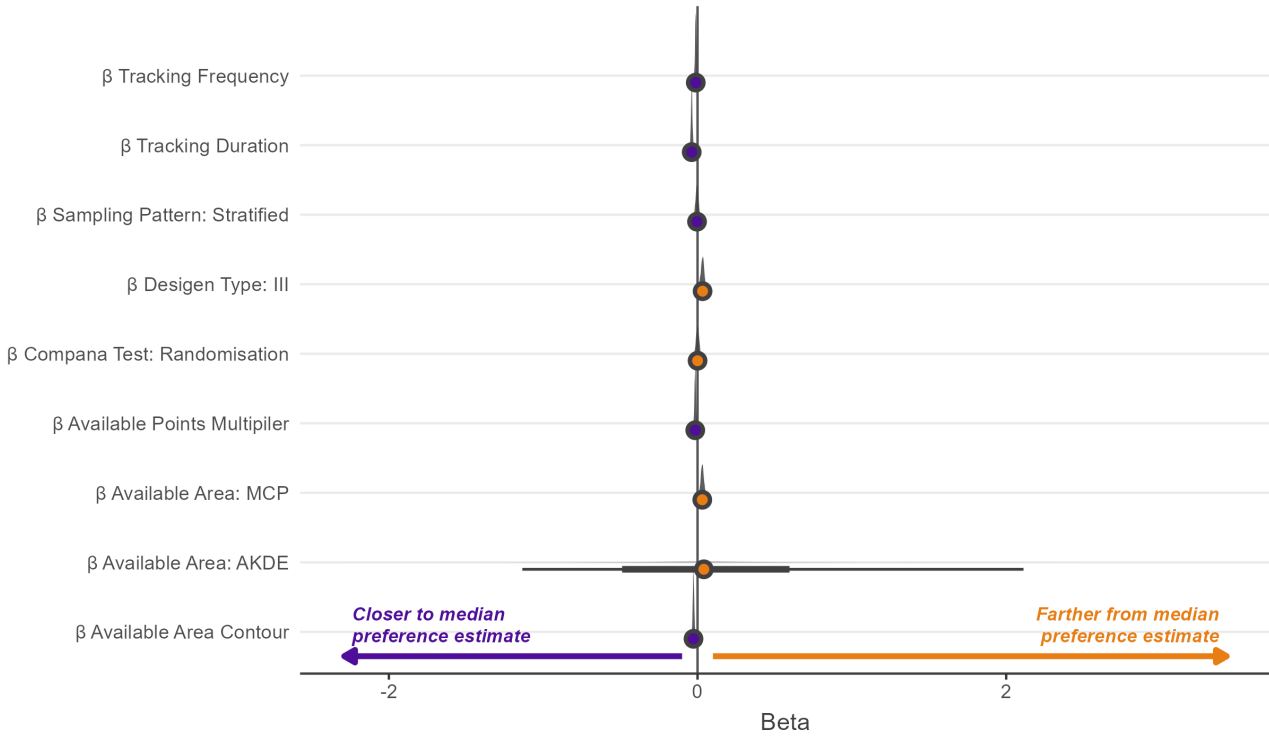


Figure 5. Beta coeffs

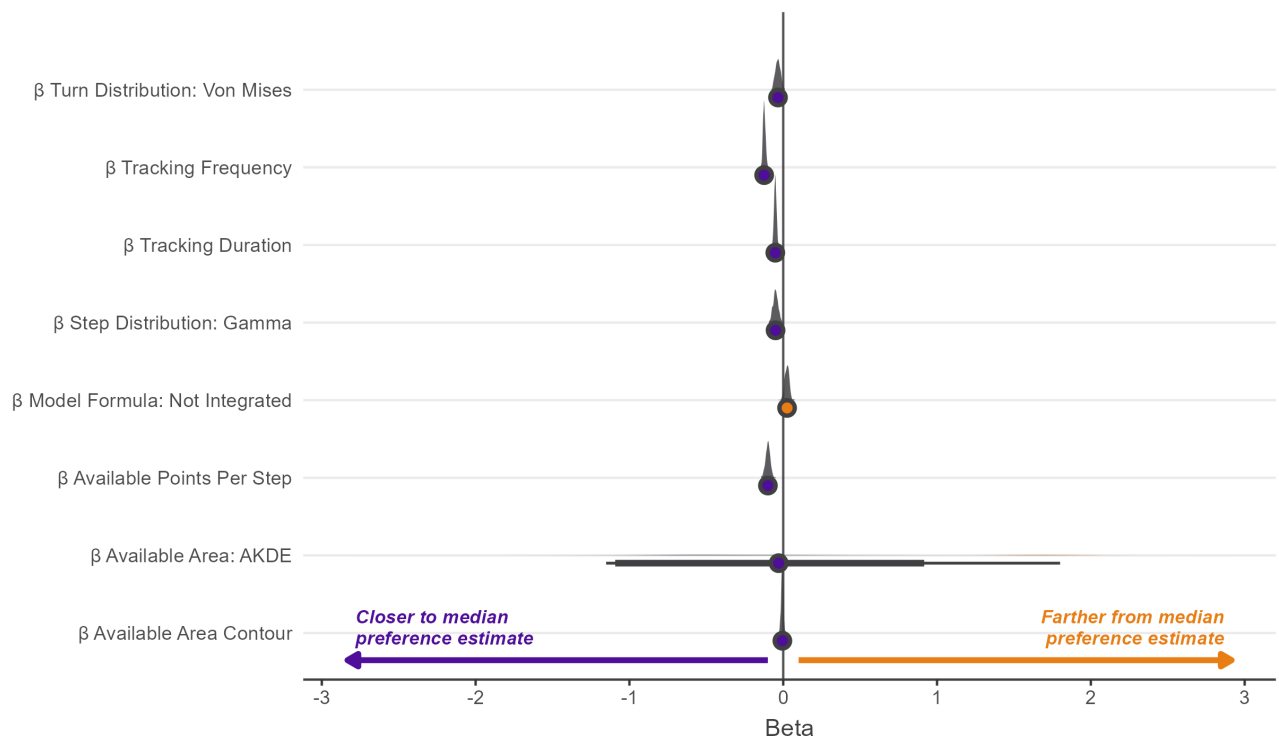


Figure 6. Beta coeffs

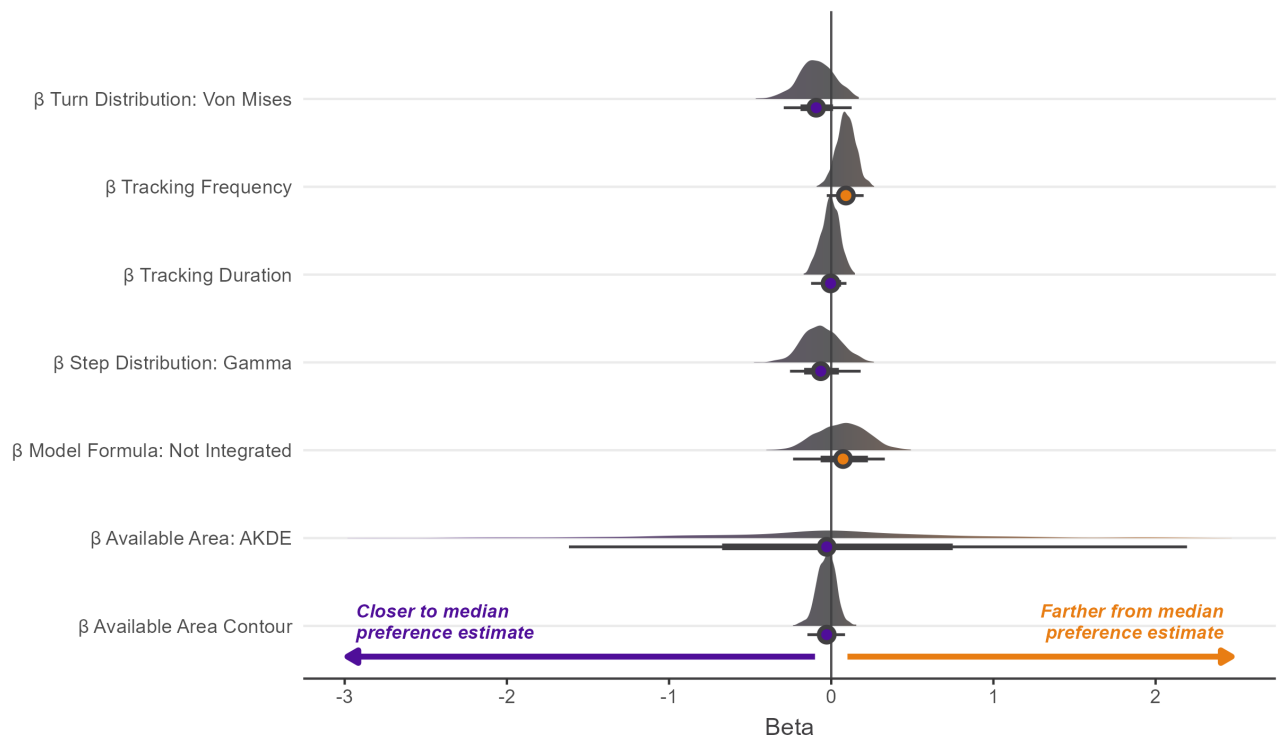


Figure 7. Beta coeffs

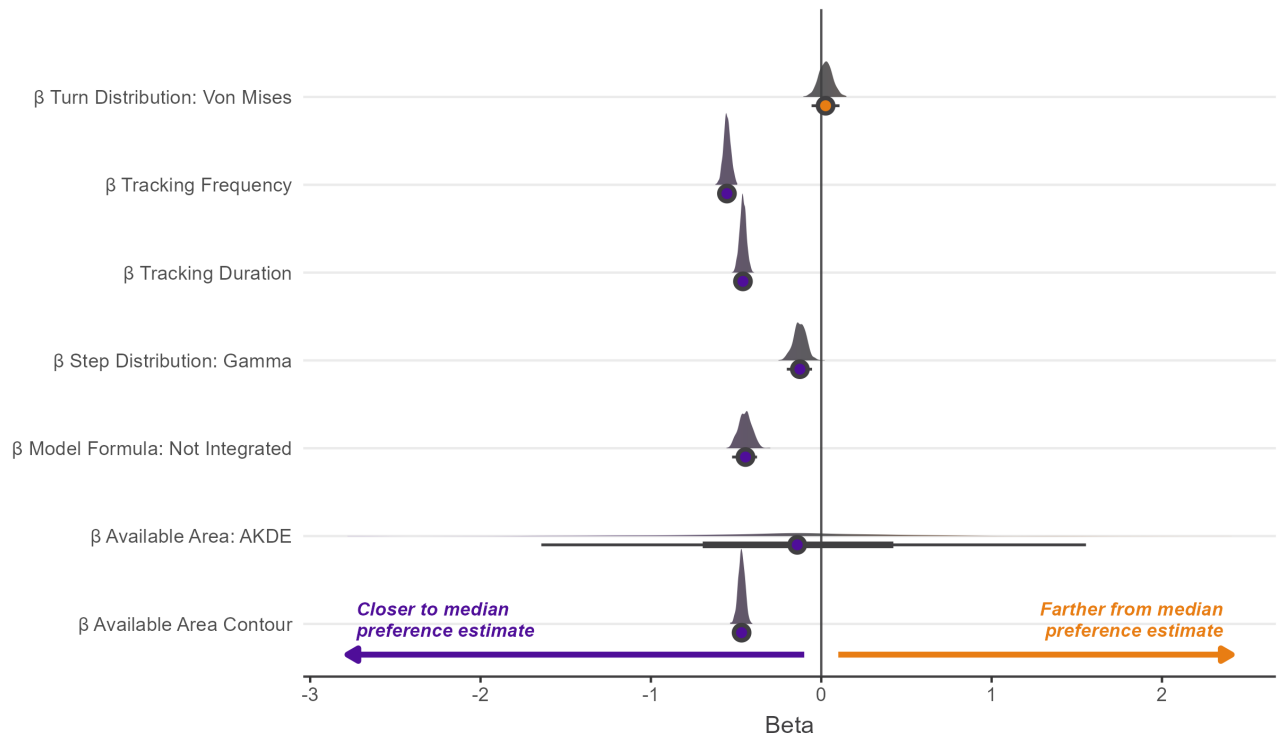


Figure 8. Beta coeffs

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.

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 used 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), 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*.
- Bartoń K. 2023. *MuMIn: Multi-model inference*.
- 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, Stokorb 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, Golemund 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.