Applying Step Selection Functions to Temporally

Irregular GPS Data - A Simulation Study

David D. Hofmann1,2,§  Gabriele Cozzi1,2  John Fieberg 3

July 24, 2023

1. Department of Evolutionary Biology and Environmental Studies, University of Zurich,

Winterthurerstrasse 190, 8057 Zurich, Switzerland.

1. Botswana Predator Conservation Program, Wild Entrust, Private Bag 13, Maun, Botswana.
2. Department of Fisheries, Wildlife, and Conservation Biology, University of Minnesota,

St. Paul, MN, USA.

§ Corresponding author (david.hofmann2@uzh.ch)

**Running Title:** Step Selection Functions and Irregular GPS Data

**Keywords:** step selection functions, gps data, animal movement, irregular data, missing

fixes

**Abstract**

Step selection functions are a versatile and powerful tool to study habitat and movement preferences of a focal species using GPS data. While the method requires GPS fixes that are regularly spaced in time, GPS data in reality are often incomplete, leading to irregular intervals between subsequent fixes. To address this issue, researchers typically only consider bursts of GPS fixes that were successfully collected at comparable intervals, thus discarding large portions of their data. Here, we reassess this practice and conduct a simulation study with known habitat and movement parameters to examine whether inclusion of irregular data improves or impairs model performance. Furthermore, we explore the usefulness of five alternative approaches to account for potential biases arising from temporal irregularity and compare them to a baseline approach. Our results suggest that including irregular data improves model performance, especially when an appropriate method to account for temporal irregularity is chosen.

**Contents**

[**1 Introduction** **1**](#_Toc22001)

[1.1 Animal Space Use and Habitat Preferences 1](#_Toc22002)

[1.2 Step Selection Functions 1](#_Toc22003)

[1.3 Integrated Step Selection Functions 2](#_Toc22004)

[1.4 The Assumption of Regularity 2](#_Toc22005)

[1.5 What We Do 3](#_Toc22006)

[1.6 Either into Discussion or Remove 4](#_Toc22007)

[**2 Methods** **4**](#_Toc22008)

[2.1 Landscape Simulation 4](#_Toc22009)

[2.2 Movement Simulation 5](#_Toc22010)

[2.3 Data Rarefication 6](#_Toc22011)

[2.4 Computing Bursts 6](#_Toc22012)

[2.5 Fitting Distributions 6](#_Toc22013)

[2.6 Step Selection Functions 7](#_Toc22014)

[2.7 Conditional Logistic Regression Model 9](#_Toc22015)

[**3 Results** **9**](#_Toc22016)

[**4 Discussion** **10**](#_Toc22017)

[**5 Authors’ Contributions** **12**](#_Toc22018)

[**6 Data Availability** **12**](#_Toc22019)

[**7 Acknowledgements & Funding** **12**](#_Toc22020)

# Introduction

## Animal Space Use and Habitat Preferences

Understanding how animals move across the landscape, what habitats they prefer, and what resources they select are fundamental questions in movement ecology (Nathan, 2008). Recent advances in the realms of GPS tracking (Cagnacci et al., 2010; Williams et al., 2019) and remote sensing technologies (Toth and J´o´zk´ow, 2016; Rumiano et al., 2020) have brought forward new and exciting opportunities to study how animals move and interact with their environment (Tomkiewicz et al., 2010; Kays et al., 2015). A common approach to study resource selection and movement behavior is by contrasting characteristics at locations where an animal was observed to characteristics at alternative locations that are deemed available to an animal (Fortin et al., 2005; Manly et al., 2007; Cushman and Lewis, 2010). While **GPS data** appears particularly useful for such purposes, (terms in bold at first occurrence are defined in Table 1) GPS data tend to be inherently autocorrelated, thus violating the assumption of independence in resource selection studies (Fieberg et al., 2010). The uprise and increased availability of GPS data has therefore also stimulated the development of innovative techniques to analyze and model animal movement while specifically addressing the issue of autocorrelation (Signer et al., 2017; Seidel et al., 2018; Joo et al., 2020; Fieberg et al., 2021).

## Step Selection Functions

One widely used approach to handle autocorrelation when relating animal movement from

GPS data to environmental conditions is the application of step-selection functions (SSFs; Fortin et al., 2005). SSFs operate by linking consecutive **GPS fixes** into straight-line segments that are defined as **steps** (Turchin, 1998). Each **observed step** is combined with a set of alternative **random steps** that are generated by sampling **step-lengths** and **turning-angles** from observed values or from parametric distributions fitted to observed values (Fortin et al., 2005; Thurfjell et al., 2014). Environmental conditions at observed steps are then contrasted to environmental conditions at random steps in a (mixed effects) conditional logistic regression framework (Fortin et al., 2005; Muff et al., 2020). Model estimates emerging from the regression model then quantify the relative selection or avoidance of different environmental factors and are informative about habitat selection (Avgar et al.,

2017).

## Integrated Step Selection Functions

The original SSF framework proposed by Fortin et al. (2005) has been refined and generalized by Avgar et al. (2016), who introduced the term *integrated* SSF (iSSF) and showed that

including descriptors of step-characteristics in the conditional logistic regression model allows to obtain information on both an animal’s **habitat kernel** and **movement kernel**. In fact, parameter estimates associated with step-characteristics can be used to update tentative distribution parameters for step-lengths and turning-angles, thus reducing biases in the movement kernel that stem from the habitat kernel (Forester et al., 2009; Duchesne et al., 2015; Avgar et al., 2016; Fieberg et al., 2021). Ultimately, a model parametrized using iSSFs resembles a fully mechanistic model from which movement can be simulated (Avgar et al., 2016; Signer et al., 2017), making iSSFs a powerful and versatile tool for both inference and prediction (Hofmann et al., 2023). iSSFs are readily accessible through the R-package amt (Signer et al., 2019) and have proven extremely effective in numerous ecological studies (Thurfjell et al., 2014), providing insights into seasonal space use (Vales et al., 2022; Enns et al., 2023), resource selection during distinct behavioral phases (Elliot et al., 2014; Abrahms et al., 2017; Broekhuis et al., 2019), and to map landscape resistance or movement corridors (Zeller et al., 2020; Buchholtz et al., 2020; Hofmann et al., 2023).

## The Assumption of Regularity

A key assumption of iSSFs is that GPS fixes are collected at regular intervals, thus producing steps with **regular step-durations** (∆*t*; Fortin et al., 2005; Thurfjell et al., 2014). Here, we refer to such data as **regular GPS data** and assume the regular step-duration to be one (i.e. ∆*t* = 1). Regular step-durations are critical to ensure that the distributions from which turning-angles and step-lengths are sampled when generating random steps are representative of an animal’s availability domain. Given that most GPS devices are programmed to record fixes at regular, pre-determined intervals, this assumption appears trivial to satisfy (Hofman et al., 2019). In reality, however, GPS failure often introduces **missingness** thus confronting researchers with **irregular GPS data** and **irregular step-durations**. Difficulties to obtain GPS fixes at the desired intervals can be linked to various factors such as topography (Lewis et al., 2007), vegetation and canopy cover (Phillips et al., 1998; Hansen and Riggs, 2008), animal behavior (Mattisson et al., 2010), and collar orientation (D’eon and Delparte, 2005). In a comprehensive study, Hofman et al. (2019) showed that across 3’000 GPS devices and 160 species the average fix success rate was 78%, highlighting that irregular GPS data is a common problem in ecological studies. It is generally agreed that researchers utilizing such data with iSSFs should adopt a forgiveness of one (i.e. only use data with regular step-durations), and discard any steps exceeding the regular step-duration (Thurfjell et al., 2014). It is therefore common practice to only retain **bursts** of regular GPS data. The main issue with this approach is the non-linearity in which missing fixes reduce the number of **valid steps** that can be used for further analysis (Figure 1 and Figure 2). For example, the hypothetical trajectory presented in Figure 1 highlights a case in which fix No. 4 failed to be collected. The absence of fix No. 4 prevents the computation of a step between fixes 3 and 4, and between fixes 4 and 5. The absence of these steps, in turn, also prohibits computing a turning-angle for the step between fixes 5 and 6. In result, the lack of a single fix reduced the effective sample size, i.e. the number of valid steps, by three. A quick simulation study in which we examined how a gradual reduction from 1’000 to 0 fixes influences the number of valid steps revealed that already at a missingness of 25% the number of valid steps drops below 500 (Figure 2), i.e. by more than 50%. If a modeler increased his forgiveness by one, and was willing to also accept step-durations of two, an additional 250 valid steps could be gained. The ability to capitalize on such “lost” data appears critical for applications where GPS data is limited, such as, for instance, on dispersing individuals (Rudnick et al., 2012; Fattebert et al., 2015; Cozzi et al., 2020).

## What We Do

Here, we reassess the practice of removing irregular GPS data in iSSFs and investigate whether including irregular GPS data could, in fact, be used to improve model performance. Our presumption is that even irregular data contains valuable information on habitat and movement preferences that could be tapped into using appropriate methods. To test this notion, we conduct a comprehensive simulation study where we simulate regular GPS data according to a movement model with known preferneces. Using simulations instead of real data has the benefit that underlying parameters are known, thus allowing us to assess the reliability of different methods under different conditions (K´ery and Royle, 2016). We then introduce different levels of missingness and apply SSFs to estimate simulation parameters. We employ five different iSSF flavors, each aiming towards accounting for irregularity in the data, and compare them to a baseline approach that simply ignores it. To examine the impact of different landscape configurations on our results, we rerun our simulations for different levels of spatial autocorrelation.

## Either into Discussion or Remove

Several methods to deal with temporal irregularity in GPS data have been developed previously. An intuitive solution is provided by McClintock (2017), who suggests to fitting Johnson’s continuous-time correlated random walk movement model (Johnson et al., 2008) to observed data and impute missing fixes from it. This approach is readily available through the R package crawl (Johnson et al., 2022), yet has only been tested for use with hidden Markov movement models but not with iSSFs (McClintock, 2017). Alternatively, Munden et al. (2021) introduced time-varying iSSFs, where step-durations are treated as random variables (similar to step-lengths and turning-angles) and randomly drawn from observed data or parametric distributions for each random step. The step-duration is then also included as step-characteristic in the regression model. Notably, both Munden’s and McClintock’s methods assume that step-durations are entirely random, as is often the case in marine applications, where GPS fixes can only opportunistically be taken when animals surface. Our focus, however, lies on data that was mostly collected at regular intervals, but occasionally failed to do so. A final solution is showcased by Vales et al. (2022), who propose to simultaneously estimate habitat and movement kernels, while also modeling the probability of a GPS fix being successful given environmental circumstances. This framework was originally proposed by Nielson et al. (2009) for resource selection functions and focused on mitigating potential biases emerging from habitats that are undersampled due to GPS failure. Here, we rather focused on utilizing observed fixes that may still contain valuable information on habitat selection but typically get discarded to ensure regularity.

# Methods

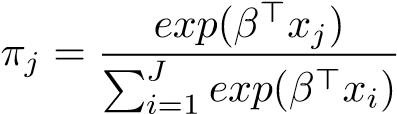
We implemented the simulation study in the programming language R (R Core Team, 2022) and achieved parallelization using the R-package pbmcapply (Kuang, 2022). We generated figures using the ggplot2 (Wickham, 2016), ggpubr (Kassambara, 2023), and ggh4x (van den Brand, 2023) R-packages. All scripts to reproduce this study are available through an online repository.

## Landscape Simulation

We simulated a virtual landscape comprising one categorical and two continuous spatial layers, each with a resolution of 300 x 300 pixels (Figure 3) and spanning across x- and y-coordinates from 0 to 300. The first layer, forest (categorical), represented areas covered by woodland and was simulated using a Gaussian random field neutral landscape model (Schlather et al., 2015) with magnitude of variation in the landscape of one and magnitude of variation in scale of zero. We thresholded the layer into the categories non-forest (scored 0) and forest (scored 1) using the 50% percentile as threshold and ensured smooth forest borders by applying a moving window of 9 pixels within which we computed the modal value (Figure SX). The second layer, elev (continuous), resembled an elevation layer and was also simulated using a Gaussian random field neutral landscape model (Schlather et al., 2015) with magnitude of variation in the landscape of one and magnitude of variation in scale of zero. We simulated both layers using the R-package NLMR (Sciaini et al., 2018) and varied the autocorrelation parameter during simulation from 5 to 10 to 20 to investigate if different levels of autocorrelation affected our results. The third layer, dist (continuous), indicated the distance (in pixels) to the center of the virtual landscape (*x* = 150*,y* = 150), and can be understood as a point of attraction, such as, for instance, the center of an animal’s homerange. We computed spatial distances and handled raster data using the r-package raster (Hijmans et al., 2023). We normalized simulated layers to a range between zero and one and replicated the simulation of each layer 100 times for the different autocorrelation scenarios, resulting in 300 unique landscapes. To mitigate edge effects, we created an artificial 50pixel wide boundary zone within which we randomly resampled covariate values from the simulated layers (Koen et al., 2010).

## Movement Simulation

To simulate movement across the virtual landscape, we employed the iSSF simulation algorithm developed by (Signer et al., 2017) and applied in (Hofmann et al., 2023). In step one, we generated a random starting location by sampling random x- and y-coordinates on the simulated landscape. To prevent starting points near map borders, we restricted sampled locations to x- and y-coordinates between 50 and 250 (white dotted rectangle in Figure 3). In step two, we generated a set of 10 random steps originating from the sampled location, by sampling turning-angles from a von Mises distribution with concentration parameter *κ* = 0*.*5 and step-lengths from a gamma distribution with shape parameter *k* = 3 and scale parameter *θ* = 1. In step three, we computed average covariate values along each random step from the underlying covariate layers. In step four, we assigned to each step *j* a probability *πj* of being selected. This probability was calculated based on Equation 1:

 (Equation 1)

Where the *β*’s represent relative selection strengths and *xj* the average covariate values along the *j*th step. The probability of a step being selected thus depended on its associated covariates, as well as on the covariates of all alternative steps. We defined the simulation parameters as *βdist* = −20, *βelev* = 0*.*5, and *βforest* = −1. That is, simulated individuals were attracted to the landscape’s center, preferred elevated areas, and avoided areas covered by forest. In step five, we sampled one of the random steps based on predicted probabilities and computed the new position of the individual. We then repeated steps two to five until a total of 1’000 steps were simulated. We repeated the simulation for each of the 300 simulated landscapes, producing 300 unique trajectories.

## Data Rarefication

We rarefied the simulated GPS data by randomly removing a fixed fraction of GPS fixes. To assess the impact of different degrees of missingness, we varied the fraction of removed data from 0% (complete dataset) to 50% with increments of 10%. The random removal of GPS fixes introduced temporal irregularity and so the resulting step-durations differed depending on the time elapsed between remaining fixes. We replicated the rarefication of each trajectory 100 times.

## Computing Bursts

We used the rarefied data to compute movement bursts. A movement burst consisted of a sequence of consecutive GPS fixes with step-durations that did not exceed the forgiveness. To test how different levels of forgiveness impacted our results, we varied forgiveness from 1 (maximum allowed step-duration = 1) to 5 (maximum allowed step duration = 5). As an example, if the forgiveness was one, already a single missing GPS fix introduced a new burst. If the forgiveness was two, in contrast, step-durations up to two were allowed before a new burst was introduced. Within each burst, we computed turning-angles and step-lengths. Note that the first step in a burst is always missing a turning angle as its relative orientation to the previous step cannot be assessed.

## Fitting Distributions

Using the bursted data, we parametrized turning-angle and step-length distributions. Specifically, we used the fit distr function from the amt package (Signer et al., 2019) to fit a von Mises distribution to turning-angles, and a gamma distribution to step-lengths. Notably, we employed two differing approaches:

1. *Regular:* In this approach, we fitted parametric distributions considering only steplengths and turning-angles from steps that exhibited a step-duration of one. Any steps with irregular step-durations therefore did not affect distributional parameter estimates. This represents the traditional approach in iSSFs.
2. *Dynamic:* In this approach, we fitted separate parametric distributions to turningangles and step-lengths from steps with different durations. In other words, we parametrized separate turning-angle and step-length distributions representative of steps with durations one, two, three, four, and five (which corresponds to the maximum forgiveness level we tested for). Because some step-durations only rarely occurred at low levels of missingness, parametrization was occasionally impossible. To overcome this issue, we artificially enforced a GPS data missingness of 50% (irrespective of the true missingness scenario), again by randomly removing GPS fixes. This introduced a sufficient amount of longer step-durations and enabled us to estimate dynamic distribution parameters across all scenarios. We replicated the data removal and parameter estimation 10 times and computed average distribution parameters for each step-duration.

## Step Selection Functions

We implemented six different flavors of iSSFs that mainly differed in the way in which random steps were generated, but sometimes also in the model call that was used for the conditional logistic regression model (Figure 4). The first approach , *uncorrected*, resembled a traditional iSSF and did not try to account for the temporal irregularity within the data, whereas all other approaches were targeted towards reducing potential biases arising from the inclusion of irregular GPS data. Irrespective of the approach, we paired each observed step with 10 random steps by sampling step-lengths from a fitted gamma distribution and turning-angles from a fitted von Mises distribution:

*Uncorrected*: In the *uncorrected* approach, we sampled step-lengths and turning-angles from parametric distributions fitted to observed steps with step-durations of one. This approach ignores the fact that observed steps may exhibit different step-durations. At higher forgiveness levels, the generated random steps are thus unlikely to be representative of the true availability domain.

*Na¨ıve*: In the *na¨ıve* approach, we sampled step-lengths and turning-angles from parametric distributions fitted to observed steps with step-durations of one. However, we linearly scaled sampled step-lengths to the step-durations of the observed steps. For instance, for any step with a step-duration of two, we doubled sampled step-lengths. This approach implicitly assumes that step-lengths scale linearly with step-durations. Since it is not clear how turning-angles should scale with step-duration, we did not correct sampled turning-angles.

*Dynamic*: In the *dynamic* approach, we sampled step lengths and turning-angles from distributions that were fit to different step-durations. That is, for observed steps with step-duration of two, we sampled step-lengths and turning-angles from distributions fit to observed steps with a duration of two.

*Multistep*: In the *multistep* approach, we sampled step-lengths and turning-angles from parametric distributions fitted to observed steps with step-durations of one. However, if an observed step had a step-duration larger than one, we extended each random step by yet another random step. We repeated this extension until the number of consecutive random steps matched the observed step-duration. For instance, for an observed step with step-duration of two, we sampled random steps twice and concatenated them into “random paths”. The paths were then simplified to a straight line connecting the first and last coordinate of each path, which represented the final random step.

*Model*: In the *model* approach we sampled step-lengths and turning-angles from parametric distributions fitted to different step-durations, similar to the *dynamic* approach. However, we accounted for potential biases in distribution parameters by including interactions between the step-duration and other step-characteristics (cfr Section 2.7). Notably, we only included steps with durations *>* 1 if the respective duration was represented at least 5 times in the rarefied dataset. This was to avoid numerical instabilities when estimating model parameters.

*Imputed*: In the *imputed* approach, we substituted missing fixes using predictions from a simple movement model. Specifically, we fitted Johnson’s single-state movement model (Johnson et al., 2008) to each observed trajectory and used the parametrized model to predict locations for the missing fixes. For this, we utilized the crawl package, where we fitted Johnson’s model using the crwFit function and made predictions using crwPredict. This resulted in a complete dataset without any missing fixes. Hence, the trajectory again consisted of a single continuous burst. We then sampled step-lengths and turning-angles from distributions fitted to observed steps with step-duration of one.

Together, an observed and its 10 associated random steps formed a stratum that received a unique ID. Along all observed and random steps we computed average covariate values from the underlying covariate layers.

## Conditional Logistic Regression Model

We estimated relative selection strengths (i.e. the *β*s in Equation 1) using conditional logistic regression, implemented using the R-package survival (Therneau, 2023). Our response variable was a binary variable (observed) indicating if a step was an observed (scored 1) or random step (scored 0) and used step-stratum as stratifier variable. Besides the habitat covariates (dist, elev, forest), we also included descriptors of the step length and turningangle (sl, log(sl), cos(ta)) as predictors in our regression model. The model call was therefore as follows:

*case* ∼ *dist* + *elev* + *forest* + *sl* + *log*(*sl*) + *cos*(*ta*)

For the *model* approach, the model call was slightly adjusted (see above) and also included interactions between the factor covariate step-duration and the movement descriptors (sl, log(sl), cos(ta)), thus looking as follows:

*case* ∼ *dist* + *elev* + *forest* + *sl* + *log*(*sl*) + *cos*(*ta*) + *sl* : *duration*

+ *log*(*sl*) : *duration* + *cos*(*ta*) : *duration*

Including descriptors of step characteristics allowed us to later update our tentative parameters for the step-length and turning-angle distributions (for further details and associated formulas, see Fieberg et al., 2021). We thus obtained corrected parameter estimates for *k*, *κ*, and *θ*. We recorded parameter estimates across all replicates, computed their means and bootstrap 95% CIs for each of the different treatment combinations.

# Results

All tested approaches reliably and accurately recovered parameters of the habitat kernel, yet, at elevated levels of missingness and forgiveness, the *uncorrected*, *na¨ıve*, and *imputed* approach produced slightly biased estimates for the movement kernel. These results are qualitatively comparable across different scenarios; hence we here only focus on results from the scenario with landscape autocorrelation of 20 and a missingness of 50% (Figure 5). We provide results for all other scenarios in Appendix SX. Irrespective of the chosen method, increasing the forgiveness from 1 to 5 substantially improved precision in model estimates for the habitat kernel without introducing bias (upper panel in Figure 5). With respect to the movement kernel (lower panel in Figure 5), however, the *uncorrected*, *na¨ıve*, and *imputed* approaches tended to provide biased parameter estimates, especially at high levels of forgiveness. The imputation approach appeared to perform particularly bad at estimating the kappa parameter for the turning-angle distribution, yet this was unrelated to the level of forgiveness and largely owed to missingness per se (see also Figure SX). The *dynamic*, *model*, and *multistep* approaches, on the other hand, performed well, even to very high forgiveness

levels.

# Discussion

In this study, we aimed to examine whether including irregular GPS data in iSSFs improves or impairs parameter estimation. We performed a comprehensive simulation study, using known habitat and movement parameters, to evaluate the reliability of six different iSSF methods in retrieving simulation parameters. Our results demonstrate that the inclusion of irregular GPS data enhances the precision of model estimates, resulting in reduced variability across replicates. However, while all six approaches appeared robust at estimating parameters of the habitat kernel, some provided biased estimates for the movement kernel. Overall, our results highlight that leveraging on irregular GPS data can be fruitful, especially when an appropriate method to handle GPS irregularity is chosen.

Parameter estimates for the movement kernel emerging from the *uncorrected* approach exhibited increasing bias as a modeler’s forgiveness increased. This bias arose due to the generation of random steps by sampling step-lengths and turning angles from distributions fitted to steps with step-durations of one, even though steps with longer step-durations were present in the observed data and included in the model. Consequently, the model attempted to acount for these longer step-durations by pulling parameter estimates for the movement kernel closer towards those of step-durations *>* 1. In our case, the estimates for the shape (*k*) and scale (*θ*) parameters suggested that random steps were too short, while the estimate for the concentration parameter (*κ*) indicated that generated random steps were overly directional. This resulted from the fact that steps with durations *>* 1 were longer and less directional, rather than any inherent bias in our tentative distributions for steps with durations = 1.

Although the *na¨ıve* approach yielded slightly more accurate estimates for the movement kernel compared to the *uncorrected* approach, its underlying assumption of a linear relationship between step-length and step-duration proved to be overly simplistic (see also Figure SX). Consequently, as the modeler’s forgiveness increased, the bias in movement estimates also escalated. Moreover, the *na¨ıve* approach cannot account for differences in the turning-angle distributions of steps with different durations, thus limiting its usability.

The *dynamic* approach performed well, providing precise and unbiased estimates for both kernels. Given that step-length and turning-angle distributions were fitted separately to different step-durations but also considering that our simulation assumed no interactions between the movement and habitat kernels, this was to be expected. While this approach appears well suited for scenarios without interactions between the two kernels, its effectiveness may diminish in cases where interactions exist. This is largely due to the necessity of dynamically fitting separate distributions to different combinations of environmental characteristics, which may quickly become computationally challenging.

In contrast to the *dynamic* approach, the *model* approach offered another level of flexibility by incorporating step-duration in the regression model. This enabled the derivation of distinct correction parameters for the distributions fitted to each step-duration. This approach can be expected to maintain robustness in the presence of interactions between the habitat and movement kernels, as the step-duration can be readily included in three-way interactions. In fact, future studies could investigate impact of such interactions and how they influence the efficacy of the presented approaches. We included the step-duration as a factor covariate, which was sensible given the limited step durations (1 to 5). However, this can hinder parameter estimation in cases where some step-durations are not represented sufficiently. Alternatively, step-durations could be included in the iSSF model as continuous covariate, in which case also higher order polynomial relationships could be modeled. This also appears promising for cases where GPS data is entirely irregular, such as in marine

applications.

The *multistep* approach performed similarly well to the *dynamic* and *model* approaches. It is comparably simple and requires little tweaking to accommodate for GPS irregularity.

Out of all tested approaches, the *imputation* approach performed the worst, as it produced highly biased parameter estimates for the movement kernel. This is probably owed to the imputation approach that utilizes an overly simplistic movement model to augment missing fixes. In result, step-lengths and turning-angles for steps between imputed fixes become biased, which is why the iSSF model fails to retrieve the true simulation parameters. While this approach appears to perform well with hidden Markov movement models (McClintock, 2017), we advise against its use with iSSFs.

Our simulation study assumed that GPS fixes were missing at random, i.e. that failure to obtain a fix was unrelated to the habitat types, time of the day, etc. Multiple studies have shown, however, that fix aqcuisition often depends on factors such as canopy cover

(Hansen and Riggs, 2008; DeCesare et al., 2005), or time of the day (Graves and Waller, 2006). Future studies should strive to model these relationships and to render missingness that is habitat dependent.

In conclusion, our study shows that inclusion of irregular GPS data can aid with model performance, yet only when an appropriate iSSF framework is selected. Here, the *dynamic*, *model*, and *multistep* approaches performed well and reliably retrieved parameters of the habitat and movement kernels, even at high levels of missingness and forgiveness. We hope this article sparked interest in reassessing the common use of discarding large portions of GPS data and promotes leveraging on irregular data using appropriate methods.

# Authors’ Contributions

D.D.H., G.C., and J.F. conceived the study and designed methodology; D.D.H. implemented the analysis, J.F. assisted with modeling; D.D.H. wrote the first draft of the manuscript and all authors contributed to the drafts at several stages and gave final approval for publication.

# Data Availability

Code to reproduce the simulation study will be made available through a github repository upon publication of this article.

# Acknowledgements & Funding

This study was funded by a Swiss National Science Foundation Grant (SOMENUMBER) to G. Cozzi.

**References**

Abrahms, B., Sawyer, S. C., Jordan, N. R., McNutt, J. W., Wilson, A. M., and Brashares, J. S. (2017). Does Wildlife Resource Selection Accurately Inform Corridor Conservation? *Journal of Applied Ecology*, 54(2):412–422.

Avgar, T., Lele, S. R., Keim, J. L., and Boyce, M. S. (2017). Relative Selection Strength: Quantifying Effect Size in Habitat- and Step-Selection Inference. *Ecology and Evolution*, 7(14):5322–5330.

Avgar, T., Potts, J. R., Lewis, M. A., and Boyce, M. S. (2016). Integrated Step Selection Analysis: Bridging the Gap Between Resource Selection and Animal Movement. *Methods in Ecology and Evolution*, 7(5):619–630.

Broekhuis, F., Madsen, E. K., and Klaassen, B. (2019). Predators and Pastoralists: How Anthropogenic Pressures Inside Wildlife Areas Influence Carnivore Space Use and Movement Behaviour. *Animal Conservation*.

Buchholtz, E. K., Stronza, A., Songhurst, A., McCulloch, G., and Fitzgerald, L. A. (2020). Using Landscape Connectivity to Predict Human-Wildlife Conflict. *Biological Conservation*, 248:108677.

Cagnacci, F., Boitani, L., Powell, R. A., and Boyce, M. S. (2010). Animal Ecology Meets GPS-Based Radiotelemetry: A Perfect Storm of Opportunities and Challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1550):2157–2162.

Cozzi, G., Behr, D. M., Webster, H. S., Claase, M., Bryce, C. M., Modise, B., Mcnutt, J. W., and Ozgul, A. (2020). African Wild Dog Dispersal and Implications for Management. *The Journal of Wildlife Management*, pages 614–621.

Cushman, S. A. and Lewis, J. S. (2010). Movement Behavior Explains Genetic Differentiation in American Black Bears. *Landscape Ecology*, 25(10):1613–1625.

DeCesare, N. J., Squires, J. R., and Kolbe, J. A. (2005). Effect of Forest Canopy on GPSBased Movement Data. *Wildlife Society Bulletin*, 33(3):935–941.

D’eon, R. G. and Delparte, D. (2005). Effects of Radio-Collar Position and Orientation on GPS Radio-Collar Performance, and the Implications of PDOP in Data Screening. *Journal of Applied Ecology*, 42(2):383–388.

Duchesne, T., Fortin, D., and Rivest, L.-P. (2015). Equivalence between Step Selection Functions and Biased Correlated Random Walks for Statistical Inference on Animal Movement. *PLoS ONE*, 10(4):e0122947.

Elliot, N. B., Cushman, S. A., Macdonald, D. W., and Loveridge, A. J. (2014). The Devil Is in the Dispersers: Predictions of Landscape Connectivity Change with Demography. *Journal of Applied Ecology*, 51(5):1169–1178.

Enns, G. E., Jex, B., and Boyce, M. S. (2023). Diverse Migration Patterns and Seasonal Habitat Use of Stone’s Sheep (Ovis dalli stonei). *PeerJ*, 11:e15215.

Fattebert, J., Robinson, H. S., Balme, G., Slotow, R., and Hunter, L. (2015). Structural Habitat Predicts Functional Dispersal Habitat of a Large Carnivore: How Leopards Change Spots. *Ecological Applications*, 25(7):1911–1921.

Fieberg, J., Matthiopoulos, J., Hebblewhite, M., Boyce, M. S., and Frair, J. L. (2010). Correlation and Studies of Habitat Selection: Problem, Red Herring or Opportunity? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1550):2233– 2244.

Fieberg, J., Signer, J., Smith, B., and Avgar, T. (2021). A ‘How to’ Guide for Interpreting Parameters in Habitat-Selection Analyses. *Journal of Animal Ecology*, 90(5):1027–1043.

Forester, J. D., Im, H. K., and Rathouz, P. J. (2009). Accounting for Animal Movement in Estimation of Resource Selection Functions: Sampling and Data Analysis. *Ecology*, 90(12):3554–3565.

Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T., and Mao, J. S. (2005). Wolves Influence Elk Movements: Behavior Shapes a Trophic Cascade in Yellowstone National Park. *Ecology*, 86(5):1320–1330.

Graves, T. A. and Waller, J. S. (2006). Understanding the Causes of Missed Global Positioning System Telemetry Fixes. *Journal of Wildlife Management*, 70(3):844–851.

Hansen, M. C. and Riggs, R. A. (2008). Accuracy, Precision, and Observation Rates of Global Positioning System Telemetry Collars. *The Journal of Wildlife Management*, 72(2):518–526.

Hijmans, R. J., Bivand, R., Pebesma, E., and Sumner, M. D. (2023). Terra: Spatial Data Analysis.

Hofman, M. P. G., Hayward, M. W., Heim, M., Marchand, P., Rolandsen, C. M., Mattisson,

J., Urbano, F., Heurich, M., Mysterud, A., Melzheimer, J., Morellet, N., Voigt, U., Allen, B. L., Gehr, B., Rouco, C., Ullmann, W., Holand, Ø., Jørgensen, N. H., Steinheim, G., Cagnacci, F., Kroeschel, M., Kaczensky, P., Buuveibaatar, B., Payne, J. C., Palmegiani, I., Jerina, K., Kjellander, P., Johansson, O., LaPoint, S., Bayrakcismith, R., Linnell,¨

J. D. C., Zaccaroni, M., Jorge, M. L. S., Oshima, J. E. F., Songhurst, A., Fischer, C., Mc Bride, R. T., Thompson, J. J., Streif, S., Sandfort, R., Bonenfant, C., Drouilly, M., Klapproth, M., Zinner, D., Yarnell, R., Stronza, A., Wilmott, L., Meisingset, E., Thaker, M., Vanak, A. T., Nicoloso, S., Graeber, R., Said, S., Boudreau, M. R., Devlin, A., Hoogesteijn, R., May-Junior, J. A., Nifong, J. C., Odden, J., Quigley, H. B., Tortato, F., Parker, D. M., Caso, A., Perrine, J., Tellaeche, C., Zieba, F., Zwijacz-Kozica, T., Appel, C. L., Axsom, I., Bean, W. T., Cristescu, B., P´eriquet, S., Teichman, K. J., Karpanty, S., Licoppe, A., Menges, V., Black, K., Scheppers, T. L., Schai-Braun, S. C., Azevedo, F. C., Lemos, F. G., Payne, A., Swanepoel, L. H., Weckworth, B. V., Berger, A., Bertassoni, A., McCulloch, G., Sustr, P., Athreya, V., Bockmuhl, D., Casaer, J., Ekori, A., Melovski,ˇ D., Richard-Hansen, C., Van De Vyver, D., Reyna-Hurtado, R., Robardet, E., Selva, N., Sergiel, A., Farhadinia, M. S., Sunde, P., Portas, R., Ambarli, H., Berzins, R., Kappeler, P. M., Mann, G. K., Pyritz, L., Bissett, C., Grant, T., Steinmetz, R., Swedell, L., Welch, R. J., Armenteras, D., Bidder, O. R., Gonz´alez, T. M., Rosenblatt, A., Kachel, S., and Balkenhol, N. (2019). Right on Track? Performance of Satellite Telemetry in Terrestrial Wildlife Research. *PLoS ONE*, 14(5):e0216223.

Hofmann, D. D., Cozzi, G., McNutt, J. W., Ozgul, A., and Behr, D. M. (2023). A ThreeStep Approach for Assessing Landscape Connectivity Via Simulated Dispersal: African Wild Dog Case Study. *Landscape Ecology*.

Johnson, D., London, J. M., and McClintock, B. (2022). NMML/crawl: Last CRAN release. Zenodo.

Johnson, D. S., London, J. M., Lea, M.-A., and Durban, J. W. (2008). Continuous-Time Correlated Random Walk Model for Animal Telemetry Data. *Ecology*, 89(5):1208–1215.

Joo, R., Boone, M. E., Clay, T. A., Patrick, S. C., Clusella-Trullas, S., and Basille, M. (2020). Navigating through the r packages for movement. *Journal of Animal Ecology*, 89(1):248–267.

Kassambara, A. (2023). Ggpubr: ’ggplot2’ Based Publication Ready Plots.

Kays, R., Crofoot, M. C., Jetz, W., and Wikelski, M. (2015). Terrestrial Animal Tracking as an Eye on Life and Planet. *Science*, 348(6240):aaa2478–aaa2478.

K´ery, M. and Royle, J. A. (2016). Introduction to Data Simulation. In *Applied Hierarchical Modeling in Ecology*, pages 123–143. Elsevier.

Koen, E. L., Garroway, C. J., Wilson, P. J., and Bowman, J. (2010). The Effect of Map Boundary on Estimates of Landscape Resistance to Animal Movement. *PLoS ONE*, 5(7):e11785.

Kuang, K. (2022). Pbmcapply: Tracking the Progress of Mc\*pply with Progress Bar.

Lewis, J. S., Rachlow, J. L., Garton, E. O., and Vierling, L. A. (2007). Effects of Habitat on GPS Collar Performance: Using Data Screening to Reduce Location Error: GPS Collar Performance. *Journal of Applied Ecology*, 44(3):663–671.

Manly, B. F., McDonald, L., Thomas, D. L., McDonald, T. L., and Erickson, W. P. (2007). *Resource Selection by Animals: Statistical Design and Analysis for Field Studies*. Springer Science & Business Media.

Mattisson, J., Andr´en, H., Persson, J., and Segerstr¨om, P. (2010). Effects of Species Behavior on Global Positioning System Collar Fix Rates. *Journal of Wildlife Management*, 74(3):557–563.

McClintock, B. T. (2017). Incorporating Telemetry Error into Hidden Markov Models of Animal Movement Using Multiple Imputation. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3):249–269.

Muff, S., Signer, J., and Fieberg, J. (2020). Accounting for Individual-Specific Variation in Habitat-Selection Studies: Efficient Estimation of Mixed-Effects Models Using Bayesian or Frequentist Computation. *Journal of Animal Ecology*, 89(1):80–92.

Munden, R., B¨orger, L., Wilson, R. P., Redcliffe, J., Brown, R., Garel, M., and Potts,

J. R. (2021). Why Did the Animal Turn? Time-Varying Step Selection Analysis for Inference Between Observed Turning-Points in High Frequency Data. *Methods in Ecology and Evolution*, 12(5):921–932.

Nathan, R. (2008). An Emerging Movement Ecology Paradigm. *Proceedings of the National Academy of Sciences*, 105(49):19050–19051.

Nielson, R. M., Manly, B. F. J., McDonald, L. L., Sawyer, H., and McDonald, T. L. (2009). Estimating Habitat Selection When Gps Fix Success Is Less Than 100%. *Ecology*, 90(10):2956–2962.

Phillips, K. A., Elvey, C. R., and Abercrombie, C. L. (1998). Applying GPS to the Study of Primate Ecology: A Useful Tool? *American Journal of Primatology*, 46(2):167–172.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rudnick, D. A., Ryan, S. J., Beier, P., Cushman, S. A., Dieffenbach, F., Epps, C. W., Gerber, L. R., Hartter, J., Jenness, J. S., Kintsch, J., Merenlender, A. M., Perkl, R. M., Preziosi, D. V., and Trombulak, S. C. (2012). The Role of Landscape Connectivity in Planning and Implementing Conservation and Restoration Priorities. *Issues in Ecology*, (16):1–23.

Rumiano, F., Wielgus, E., Miguel, E., Chamaill´e-Jammes, S., Valls-Fox, H., Corn´elis, D., Garine-Wichatitsky, M. D., Fritz, H., Caron, A., and Tran, A. (2020). Remote Sensing of Environmental Drivers Influencing the Movement Ecology of Sympatric Wild and Domestic Ungulates in Semi-Arid Savannas, a Review. *Remote Sensing*, 12(19):3218.

Schlather, M., Malinowski, A., Menck, P. J., Oesting, M., and 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., and Simpkins, C. E. (2018). Nlmr and Landscapetools: An Integrated Environment for Simulating and Modifying Neutral Landscape Models in R. *Methods in Ecology and Evolution*, 9(11):2240–2248.

Seidel, D. P., Dougherty, E., Carlson, C., and Getz, W. M. (2018). Ecological Metrics and Methods for GPS Movement Data. *International Journal of Geographical Information Science*, 32(11):2272–2293.

Signer, J., Fieberg, J., and Avgar, T. (2017). Estimating Utilization Distributions from Fitted Step-Selection Functions. *Ecosphere*, 8(4):e01771.

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

Therneau, T. M. (2023). *A Package for Survival Analysis in R*.

Thurfjell, H., Ciuti, S., and Boyce, M. S. (2014). Applications of Step-Selection Functions in Ecology and Conservation. *Movement Ecology*, 2(4).

Tomkiewicz, S. M., Fuller, M. R., Kie, J. G., and Bates, K. K. (2010). Global Positioning System and Associated Technologies in Animal Behaviour and Ecological Research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1550):2163–2176.

Toth, C. and J´o´zk´ow, G. (2016). Remote Sensing Platforms and Sensors: A Survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115:22–36.

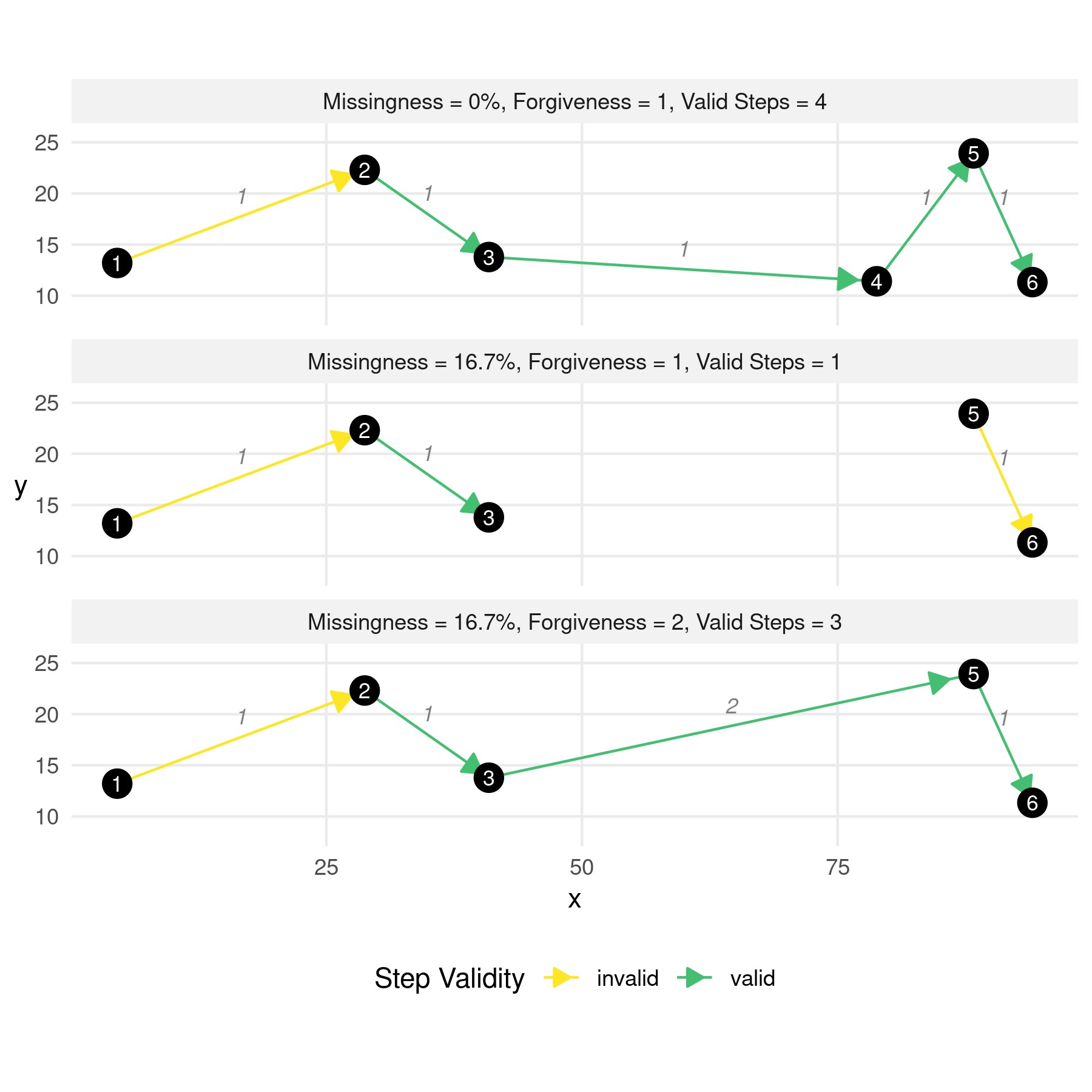
Turchin, P. (1998). *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*. Sinauer Associates, Sunderland, Mass.

Vales, D. J., Nielson, R. M., and Middleton, M. P. (2022). Black-Tailed Deer Seasonal Habitat Selection: Accounting for Missing Global Positioning System Fixes. *The Journal of Wildlife Management*, 86(8). van den Brand, T. (2023). Ggh4x: Hacks for ’ggplot2’.

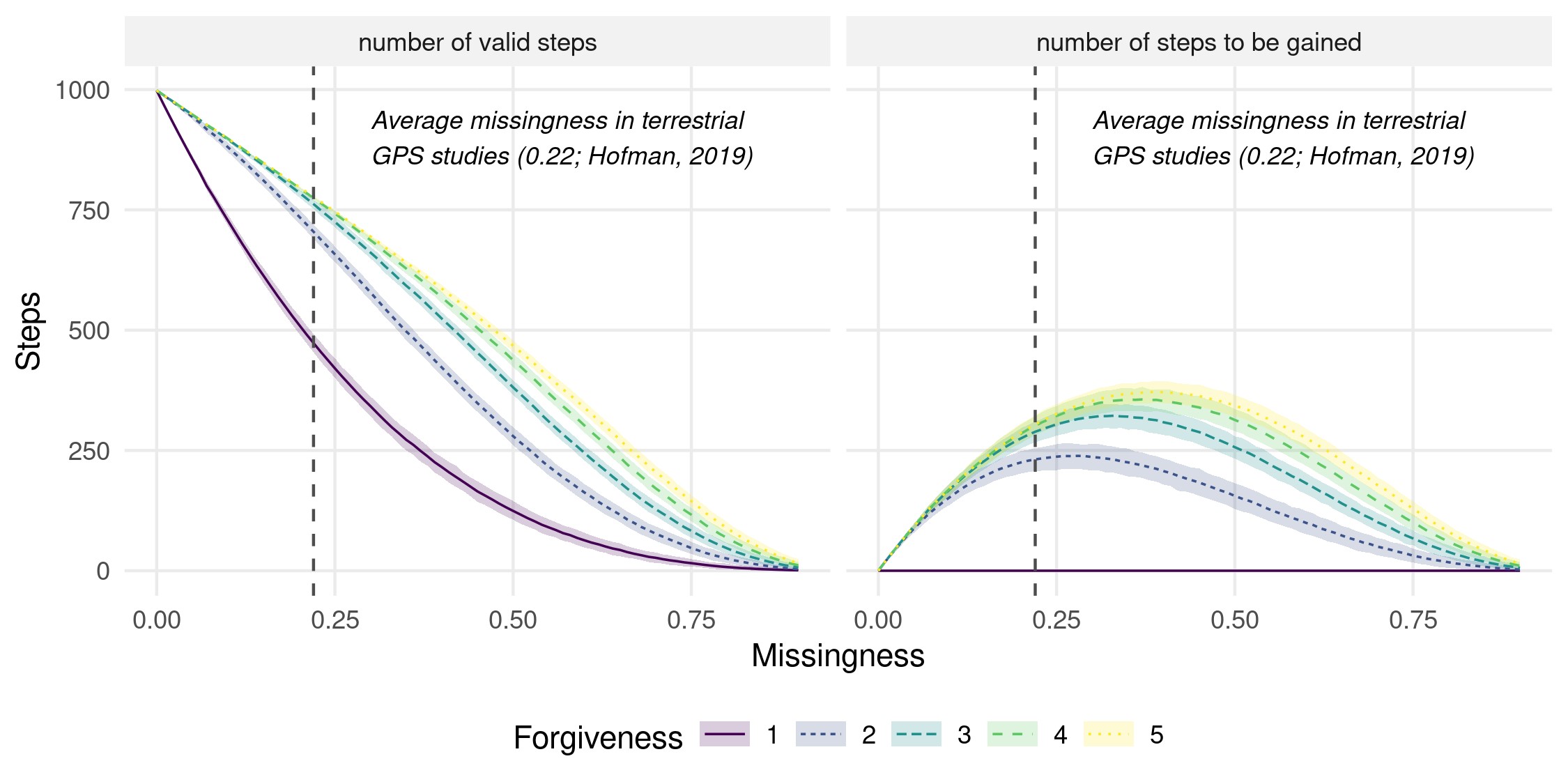
Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Williams, H. J., Taylor, L. A., Benhamou, S., Bijleveld, A. I., Clay, T. A., Grissac, S., Demˇsar, U., English, H. M., Franconi, N., G´omez-Laich, A., Griffiths, R. C., Kay, W. P., Morales, J. M., Potts, J. R., Rogerson, K. F., Rutz, C., Spelt, A., Trevail, A. M., Wilson, R. P., and B¨orger, L. (2019). Optimizing the Use of Biologgers for Movement Ecology Research. *Journal of Animal Ecology*.

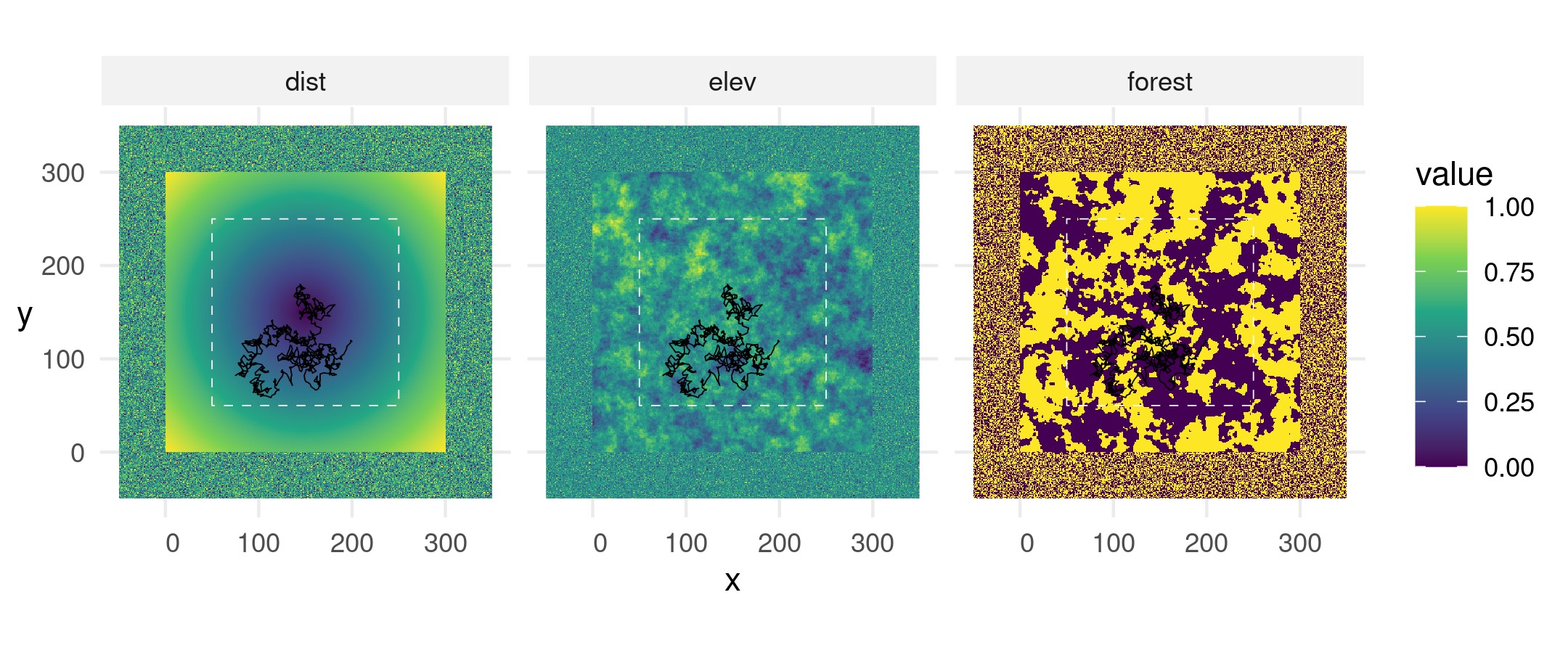
Zeller, K. A., Wattles, D. W., Bauder, J. M., and DeStefano, S. (2020). Forecasting Seasonal Habitat Connectivity in a Developing Landscape. *Land*, 9(7):233.



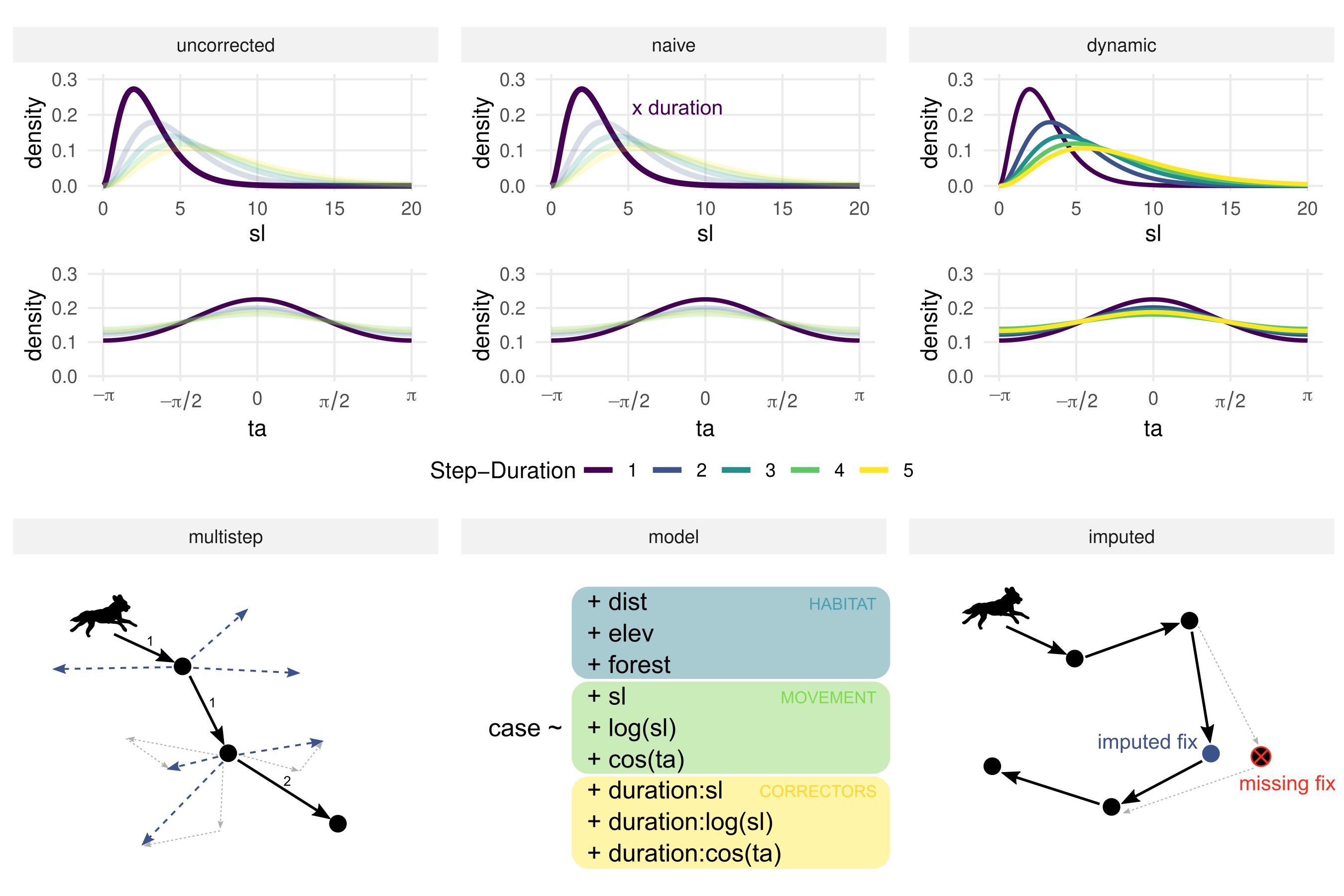
**Figure 1:** Demonstration of how missingness affects the number of valid steps that can be used for step selection functions for different levels of missingness and forgiveness. The upper panel depicts a trajectory with zero missingness. That is, all aspired fixes were successfully collected on a regular interval (yielding a regular step-duration of one). This trajectory produces one invalid (the first step has no turning-angle assigned to it) and four valid steps that can be included in the iSSF model. In the central panel, fix No. 4 was not obtained, introducing a missingness of 16.7%. If the modeler has a forgiveness of one, this implies that only a single step can be included for further analysis, as all other steps are invalid (either because no turning-angle can be computed or because step-durations exceed the forgiveness). If, however, the modeler has a forgiveness of two, such as in the lower panel, a total of three steps can be obtained for further analysis.



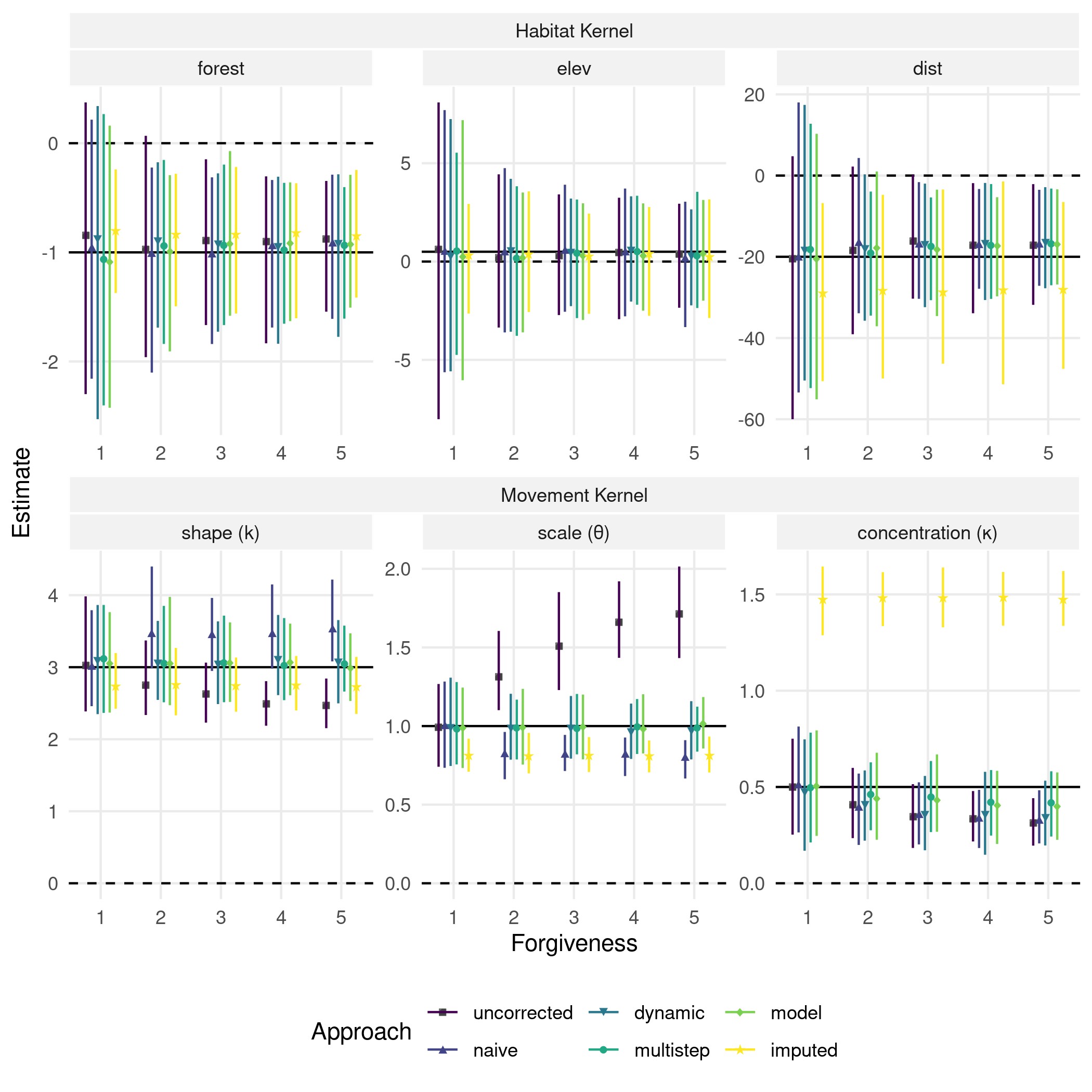
**Figure 2:** Illustration of how missingness in GPS data reduces the number of valid steps that can be used in step selection functions. At a missingness of 0, all 1’000 GPS points can be used to compute steps, thus resulting in 998 valid steps. If missingness increases, step-durations become irregular, which means that the number of valid steps decreases substantially. However, if the modeler is willing to increase his forgiveness, additional steps can be gained. The right panel shows how the number of valid steps that can be gained when increasing the forgiveness from 1 to 2, 3, 4, and 5, respectively. Ribbons indicate the bootstrap 95% CIs.



**Figure 3:** Virtual landscape across which we simulated movement trajectories. All layers have a resolution of 300 x 300 pixels and were generated randomly. Simulated individuals were initiated within the white dashed rectangle, which ensured that they would not be released directly at a map border. We also added a 50-pixel wide buffer-zone within which we randomize covariate values. The black line shows the simulated trajectory on this landscape (cfr. Section 2.2).



**Figure 4:** Graphical representation of the six different approaches we utilized to estimate the true simulation parameters. The first approach, *uncorrected*, resembled a traditional iSSF and did not try to account for the temporal irregularity within the data, whereas all other approaches were targeted towards reducing potential biases arising from the inclusion of irregular GPS data. In the *na¨ıve* approach, sampled step-lengths were multiplied by the step-duration, assuming that steplengths would scale linearly with step-duration. In the *dynamic* approach, step-lengths and turningangles were sampled from distributions fitted to different step-durations. In the *multistep* approach, observed steps were paired with sequences of random steps that matched the step-duration. In the *model* approach, we included the step-duration as corrector covariate in the regression model. In the *imputation* approach, missing fixes were augmented using predictions from Johnson’s single-state movement model.



**Figure 5:** Parameter estimates with regards to the habitat kernel (upper panel) and the movement kernel (lower panel) in the scenario with landscape autocorrelation set to 20 and GPS data missingness to 50%. True simulation parameters are indicated by the solid black lines. Parameter estimates from the different approaches are given by the colored shapes, and their bootstrap 95% CIs across 100 replicates are given by the colored lines. Importantly, parameter estimates for the movement kernel have only been computed for step-durations of one, as this relates to the true simulation parameters.

**Table 1:** Glossary of terms. Terms in the glossary are printed in bold at first occurrence in the main text. Definitions are always given in the context of step selection functions.

|  |  |
| --- | --- |
| Term | Definition |
| **GPS data** | Series of data points on time, longitude, and latitude recorded by a global positioning system (GPS) device |
| **GPS fix** | A single data point from GPS data |
| **Step** | Straight line connecting two consecutive GPS fixes |
| **Observed step** | Steps that are derived from recorded GPS data of an animal |
| **Random step** | Steps that originate from recorded GPS fixes but end in random locations. They are generated by pairing observed GPS fixes with random turning angles and step lengths |
| **Step-length** | Euclidean distance traversed by a step |
| **Turning-angle** | Measure of the change in direction between two consecutive steps in a trajectory |
| **Habitat kernel** | Probabilistic description of an animal’s habitat preferences. Estimated by comparing habitat characteristics at an animals observed steps to habitat characteristics at random steps |
| **Movement kernel** | Probabilistic description of an animal’s movement capacity. Usually estimated based on step-characteristics from observed steps. |
| **Step-duration** | Length of time an animal spends in a particular step. Calculated by determining the time elapsed between consecutive GPS fixes |
| **Regular GPS data** | GPS data were all GPS fixes have been successfully obtained on the aspired GPS schedule (e.g. every two hours). |
| **Irregular GPS data** | GPS data were some GPS locations failed to be obtained on the aspired GPS schedule. |
| **Regular step-duration** | Step duration that emerges when fixes are successfully collected at the aspired GPS schedule (here, we’ll assume the regular step duration to be one) |

**Irregular step-duration** Step duration that emerges when fixes are not successfully

collected at the aspired GPS schedule (here, we’ll assume irregular step durations to be multiples of one)

|  |  |
| --- | --- |
| **Missingness** | Fraction of GPS data that should have been collected, but for some reason was not. For example, if only eight out of ten aspired fixes were successfully collected, the missingness would be 0.2 (i.e. 20%) |

**Forgiveness** How forgiving a step selection modeler is with regards to step-durations. A modeler with forgiveness of one, for instance, only considers regular steps, while a modeler with forgiveness of two would consider irregular steps up to a step-duration of two.

|  |  |
| --- | --- |
| **Burst** | Sequence of consecutive GPS fixes where the step-duration does not exceed the forgiveness of the modeler |
| **Valid step** | A step for which a step-length, turning angle, and step duration can be computed and for which the step-duration does not exceed the forgiveness of the modeler. These steps can be used for step-selection analysis |