Methods for Implementing Integrated Step-Selection Functions with Incomplete Data

David D. Hofmann[1,2,§] iD   Gabriele Cozzi[1,2] iD   John Fieberg [3] iD

February 14, 2024

[1] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.

[2] Botswana Predator Conservation Program, Wild Entrust, Private Bag 13, Maun, Botswana.

[3] Department of Fisheries, Wildlife, and Conservation Biology, University of Minnesota, St. Paul, MN, USA.

[§] Corresponding author: david.hofmann2@uzh.ch

Abstract

Integrated step-selection analyses (iSSAs) are versatile and powerful frameworks for studying habitat and movement preferences of tracked animals. iSSAs utilize integrated step-selection functions (iSSFs) to model movements in discrete time, and thus, require animal location data that are regularly spaced in time. However, many real-world datasets are incomplete due to tracking devices failing to locate an individual at one or more scheduled times, leading to slight irregularities in the duration between consecutive animal locations. To address this issue, researchers typically only consider bursts of regular data (i.e., sequences of locations that are equally spaced in time), thereby reducing the number of observations used to model movement and habitat selection. We reassess this practice and explore four alternative approaches that account for temporal irregularity resulting from missing data. Using a simulation study, we compare these alternatives to a baseline approach where temporal irregularity is ignored and demonstrate the potential improvements in model performance that can be gained by leveraging these additional data. We also showcase these benefits using a case study on a spotted hyena (*Crocuta crocuta*).

1 Introduction

Understanding how animals move across the landscape, what habitats they prefer, and what resources they select are fundamental questions in movement ecology (Nathan, 2008). Thanks to recent advances in animal tracking (Cagnacci et al., 2010; Williams et al., 2019; Beardsworth et al., 2022) and remote sensing technologies (Toth & Jóźków, 2016; Rumiano et al., 2020), new opportunities and analytical tools have emerged for studying how animals move and interact with their environment (Tomkiewicz et al., 2010; Kays et al., 2015; Nathan et al., 2022). Methods commonly used to analyze animal movement data, including step-selection analyses (Fortin et al., 2005; Thurfjell et al., 2014; Fieberg et al., 2021) and hidden Markov models (Michelot et al., 2016), require **animal locations** (terms in bold at first occurrence are defined in Table 1) that are collected at a constant sampling frequency, leading to data that are equally spaced in time. Yet, it is common to encounter missing locations in most telemetry data sets (Frair et al., 2010; Hofman et al., 2019; Vales et al., 2022), which introduces unwanted irregularities in the duration between successive locations. Thus, there is a need for analytical tools that enable the analysis of such data, while mitigating potential biases arising from temporal irregularity introduced through missing animal locations.

Step-selection analyses (SSAs) are widely used to study animals' movement capacities and habitat-selection patterns (Fortin et al., 2005; Thurfjell et al., 2014). Straight-line segments connecting consecutive animal locations, referred to as **steps**, form the basic building blocks of the statistical likelihood in SSAs. Specifically, SSAs model the probability u of finding an individual at location s at time $t + 1$, given the animal's past positions at time t and $t - 1$, s_t and s_{t-1}, respectively:

$$u(s_{t+1}) = \frac{\phi(s_{t+1}, s_t, s_{t-1}; \gamma) w(x(s_{t+1}); \beta)}{\int_{s \in G} \phi(s_{t+1}, st, s_{t-1}, \gamma) w(x(s_{t+1}); \beta) ds} \qquad \text{(Equation 1)}$$

Here, the function ϕ represents an animal's **movement kernel** which is usually expressed in terms of **step-length** and **turning-angle** distributions, with γ representing parameters in these distributions. The function w is the **habitat-selection function** and reflects an animal's preferences β for environmental characteristics x at location s_{t+1}. In most applications, w is modeled as a log-linear function of x, taking the form $w = exp(x^\top \beta)$. The integral in the denominator of Equation 1 ensures that u is a proper probability distribution (i.e., that it integrates to 1). Following Michelot et al. (2024), we call the product $\phi \times w$ the step-selection function (SSF), as it highlights that the probability of finding an animal at a certain location depends on both the animal's movement kernel and its habitat-selection

1

function.

Given a series of observed steps, finding the movement and habitat-selection parameters that maximize the likelihood in Equation 1 requires approximating the integral in the denominator for each observed step. A variety of numerical integration techniques can be used for this purpose (Michelot et al., 2024), but a common approach is to combine **observed steps** with **random steps** generated by sampling **step lengths** and **turning angles** from parametric distributions informed by the data (Fortin et al., 2005; Thurfjell et al., 2014). Environmental conditions at observed steps are then contrasted with environmental conditions at random steps in a (mixed effects) conditional logistic regression framework (Fortin et al., 2005; Muff et al., 2020). To jointly estimate parameters in ϕ and w, movement descriptors (e.g., step length (sl), its natural logarithm (log(sl)), and the cosine of the turning angle (cos(ta))) can be included in the conditional logistic regression model, and their estimated coefficients can be used to update the initial (tentative) step-length and turning-angle distributions (Duchesne et al., 2015; Avgar et al., 2016; Fieberg et al., 2021). The specific descriptors that need to be included depend on the assumed step-length and turning-angle distributions (for more details, see Appendix C of Fieberg et al., 2021). This approach to estimating parameters of the SSF, termed *integrated* SSA (or iSSA) by Avgar et al. (2016), is similar to using importance sampling to approximate the integral in Equation 1 (Michelot et al., 2024) and is readily accessible through the R-package `amt` (Signer et al., 2019).

SSAs 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; Cozzi et al., 2018; Broekhuis et al., 2019), and the effects of landscape familiarity or memory on animal movements (Kim et al., 2023). A model parametrized using iSSA resembles a fully mechanistic movement model that can be used to simulate space use under novel conditions (Avgar et al., 2016; Signer et al., 2017; Hofmann, Cozzi, McNutt, et al., 2023; Signer et al., 2023). This characteristic has made iSSAs a useful tool for quantifying landscape resistance and identifying movement corridors (Buchholtz et al., 2020; Zeller et al., 2020; Hofmann et al., 2021; Hofmann, Cozzi, McNutt, et al., 2023).

A key assumption when conducting an iSSA is that the data have been collected at a constant sampling frequency, thus producing **trajectories** with **regular step durations** (Δt; Fortin et al., 2005; Thurfjell et al., 2014). Here, we refer to such data as **regular animal locations**, and without loss of generality, we assume the regular step duration to be one (i.e., $\Delta t = 1$). Regular step durations ensure that step lengths and turning angles

are independent of the step duration, and therefore, steps can be pooled when estimating movement parameters. Since animal locations are usually obtained using automated tracking devices, such as GPS collars programmed to record data at regular intervals, satisfying this assumption seems straightforward. In reality, however, device limitations often imply that some of the aspired datapoints fail to be collected, thus introducing **missingness** and confronting researchers with **irregular animal locations** and **irregular step durations** (Frair et al., 2010). In a comprehensive study, Hofman et al. (2019) showed that across 3,000 GPS devices and 160 species, the average success rate of obtaining a scheduled animal location was 78% (implying a missingess of 22%), thus highlighting that irregular animal locations are a frequent phenomenon in ecological studies.

It is generally recommended that, in the case of such irregular data, researchers should only retain **bursts** of steps with regular step durations (possibly with some tolerance) and discard the rest (Thurfjell et al., 2014). We will refer to this modeling approach as having a **forgiveness** level of one, indicating that only steps with step durations $\Delta t = 1$ are retained for further analysis. In R, the `amt` package provides the function `track_resample` specifically for identifying bursts of steps with regular step durations (Signer et al., 2019). The main drawback of this approach is that it may result in a substantial amount of data being discarded (Figure 1 and Figure 2). For instance, consider a hypothetical trajectory in which location 4 is missing (Figure 1a). The absence of this location prevents the computation of a step between locations 3 and 4, as well as between locations 4 and 5. Furthermore, without these steps, it becomes impossible to compute a turning angle for the step between locations 5 and 6. Consequently, the lack of a single location reduces the effective sample size, which is the number of **valid steps**, by three. Assuming that animal locations are missing at random, a missingness level of 25% causes the number of valid steps to drop by 58% (Figure 2). A modeler willing to increase their level of forgiveness to two (i.e., allowing for inclusion of steps with $\Delta t \leq 2$) would be able to increase the number of valid steps by 57% (Figure 1 and Figure 2), therefore achieving a substantial gain in effective sample size. The ability to capitalize on irregular data is likely to be particularly important for applications where data are already limited, such as, for instance, when modeling dispersing individuals (Rudnick et al., 2012; Fattebert et al., 2015; Cozzi et al., 2020). However, increasing the forgiveness also implies that step durations of the retained steps become irregular, thus necessitating appropriate tools to account for such irregularity.

Various methods have been employed in the past to address temporal irregularities in animal location data. These may serve as valuable starting points for developing approaches

that enable the integration of irregular data with iSSFs.

1. *Imputation*: An intuitive solution is provided by McClintock (2017), who suggests fitting a continuous-time correlated random walk movement model (Johnson et al., 2008) to the collected data and to use the fitted model to impute missing fixes. By imputing missing locations, the analysed trajectories become entirely regular again and can be analysed using traditional techniques. This approach, which we coin the *imputation* 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 and not with iSSFs (McClintock, 2017).

2. *Naïve*: Another approach is outlined by Munden et al. (2021), who introduced time-varying iSSFs. In this framework, a change-point detection algorithm is applied to the series of observed animal locations to identify distinct decision points where the animal turns (Potts et al., 2018; Munden et al., 2021). Steps are then created to represent straight-line movements in-between these decision points, but because decision points are not regularly spaced in time, the resulting step durations are irregular. Thus, step durations are treated as random variables, and, instead of generating random steps by sampling step lengths and turning angles, the authors generate random steps by sampling step durations, step speeds, and turning angles. The underlying assumption is that step lengths scale linearly with step duration, and can therefore be meaningfully represented by combinations of step speeds and step durations. Although this approach was developed with ultra-high-frequency data in mind, we might *naïvely* apply it more broadly to the case of missing data if we believe the presumed linear relationship to hold true (this assumption may be reasonable when step durations are short, but it is unlikely to hold for longer step durations since an animal's path will deviate from a straight line between successive locations). Hence, we propose, with our *naïve* approach, to scale the generated random steps by the observed step duration.

3. *Dynamic+Model*: Instead of generating random steps by sampling step lengths and turning angles from distributions fitted to a single step duration, one may choose to fit separate distributions to steps of different durations, thus acknowledging potentially non-linear relationships between step duration and step lengths or turning angles. Because random steps in this approach are sampled using different *tentative* distributions, it is necessary to include interactions between step durations and other step descriptors (e.g., sl, log(sl), and cos(ta)) in the conditional logistic regression model to allow updating the distributions to the different step durations. We therefore refer

4

to this approach as the *dynamic+model* approach, highlighting that step-length and turning-angle distributions are dynamically adjusted to observed step durations and that the step duration is included as a modifier of the coefficients of step descriptors in the regression model.

4. *Multistep*: Finally, we propose a *multistep* approach, where random steps of varying step durations are generated by stitching together sequences of random steps from the regular step duration. For example, one can generate a random step of duration $\Delta t = 2$ by combining two random steps of step duration $\Delta t = 1$.

Our goal with this article is to reassess the practice of discarding irregular animal locations in iSSFs and to investigate whether retaining irregular data could, in fact, serve to improve model performance. Our hypothesis is that even irregular data contains valuable information on habitat and movement preferences that could be leveraged if appropriate methods are applied. To test this notion, we conducted a comprehensive simulation study where we simulated regular animal location data with known movement and habitat parameters. We then introduced varying levels of missingness and applied iSSFs to estimate simulation parameters. Specifically, we employed the four alternative iSSF approaches outlined above and compared them to the traditional approach of including only bursts of regular data and to an uncorrected approach that simply ignored irregular step durations when using a forgiveness level > 1. To examine the impact of different landscape configurations on derived estimates, we ran our simulations for different levels of spatial autocorrelation. The use of simulations instead of real data had the benefit that underlying parameters of the movement kernel and habitat-selection function were known, which allowed us to assess the reliability of different methods in retrieving true simulation parameters under different conditions (sensu Kéry and Royle, 2016). We then compare the traditional approach (using only bursts of steps with regular step durations) to the best-performing approach (with irregular data) using GPS locations collected on a spotted hyena (*Crocuta crocuta*).

We anticipated that increasing forgiveness without adjusting for the introduced irregularity would entail a bias-variance trade-off. Specifically, we anticipated that increasing forgiveness would allow improving estimator precision, but at the cost of introducing bias due to failing to account for irregular sampling intervals. We expected this bias to be particularly pronounced at high levels of missingness. Furthermore, we hypothesized that accounting for irregularity in the *naïve*, *dynamic+model*, and *multistep* approaches would improve model accuracy, while alleviating potential bias, thus providing an effective means of incorporating additional data. Because the *imputation* approach relied on an intermedi-

5

ate movement model to predict missing animal locations, we had no prior expectations for
how well it would perform.

2 Methods

We implemented the simulation study in the programming language R version 4.3.2 (R Core
Team, 2023) and achieved parallelization of simulation-runs using the R-package pbmcapply
(Kuang, 2022). We generated figures using the ggplot2 (Wickham et al., 2023), ggpubr
(Kassambara, 2023), and ggh4x (Brand, 2023) R-packages. We manipulated raster data and
computed spatial distances using the R-package raster (Hijmans et al., 2023). An overview
of the simulation design and the different iSSF approaches is presented in Figure 3 and all
codes to reproduce this study are available through an online repository (Hofmann, Cozzi,
& Fieberg, 2023).

2.1 Landscape Simulation

We simulated a virtual landscape comprising two continuous and one categorical (binary)
spatial layers, each with a resolution of 300 x 300 pixels (Figure 4) and spanning across x-
and y-coordinates from 0 to 300. The first layer, Dist (continuous), quantified the distance
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 home-range. The second layer,
Elev (continuous), resembled an elevation layer and was simulated by sampling random pixel-
values from a normal distribution. To achieve spatial autocorrelation, we applied a circular
moving window with radius r within which we tallied pixel-values. We varied r from 5, to
10, to 20, depending on the simulated level of autocorrelation (Figure S1). The third layer,
Forest (categorical), represented areas covered by woodland and was simulated similarly to
the Elev layer, but we binarized the layer by setting all simulated values above the 50%
quantile to forest and all other values to non-forest (our reference class). We normalized
values of all simulated layers to a range between zero and one and replicated the simulation
of each layer 100 times per autocorrelation scenario, thus producing 300 unique landscape
configurations.

2.2 Movement Simulation

To simulate movement across the virtual landscape, we employed the iSSF simulation al-
gorithm developed by Signer et al. (2017) and applied in Hofmann, Cozzi, McNutt, et al.

(2023). This procedure consists of a sequence of five steps that are repeated n times to generate a movement trajectory. 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 4). In step two, we generated a set of 1,000 random steps originating from the current location, by sampling turning angles from a von Mises distribution with concentration parameter $\kappa = 0.5$ and step lengths from a gamma distribution with shape parameter $k = 3$ and scale parameter $\theta = 1$. In step three, we extracted covariate values at the end of each random step from the underlying covariate layers. In step four, we assigned to each step j a probability π_j of being selected using the equation below:

$$\pi_j = \frac{exp(\beta^\top x_j)}{\sum_{i=1}^{J} exp(\beta^\top x_i)} \qquad \text{(Equation 2)}$$

Here, β represents the vector of habitat-selection parameters and x_j the covariate value at the end of the jth step. The probability of a step being selected thus depended on its associated covariate values, the covariate values of all other random steps, and the simulated preferences β. We defined the habitat-selection parameters as $\beta_{dist} = -20$, $\beta_{elev} = 0.5$, and $\beta_{forest} = -1$. 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 simulated individual's new position. We then repeated steps two through five until the trajectory comprised a total of 1,000 steps. Each simulated step was assumed to have a step duration of exactly $\Delta t = 1$. We repeated the simulation for each of the 300 simulated landscapes, producing 300 unique trajectories (example trajectory presented in Figure 4).

2.3 Data Rarefication

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

2.4 Computing Bursts

We used the rarefied data to compute bursts consisting of a sequence of animal locations with step durations that did not exceed the forgiveness value. To test how different levels of forgiveness impacted our results, we varied forgiveness from 1 (maximum allowed step duration was $\Delta t = 1$) to 5 (maximum allowed step duration was $\Delta t = 5$). As an example, if the forgiveness was 1, any step with step duration $\Delta t > 1$ resulted in a new burst. If the forgiveness was 2, in contrast, step durations of up to $\Delta t = 2$ were allowed before a new burst was introduced (Figure 1b). Within each burst, we calculated step lengths and turning angles. However, due to the grouping of steps into bursts, the orientation of the first step within each burst relative to the previous step could not be determined. As a result, this step always lacked a turning angle and was considered invalid (Figure 1b).

2.5 Fitting Distributions

Based on the steps retained within bursts, we parametrized tentative step-length and turning-angle distributions. Specifically, we used the `fit_distr` function from the `amt` package (Signer et al., 2019), which is a wrapper function for the `fitdist` function from the `fitdistrplus` package (Delignette-Muller & Dutang, 2015), and fitted a gamma distribution to step lengths and a von Mises distribution to turning angles. Notably, we employed two different fitting procedures:

1. *Regular Distributions:* In this procedure, we fitted parametric distributions considering only step lengths and turning angles from steps that exhibited a step duration of $\Delta t = 1$ (i.e., the regular step duration). Any steps with irregular step durations ($\Delta t > 1$) were discarded and did not affect distributional parameter estimates. This represents the traditional procedure in iSSFs where only regular bursts of animal locations are considered when estimating tentative movement parameters.

2. *Dynamic Distributions:* In this procedure, we fitted separate parametric distributions to step lengths and turning angles from steps of different step durations. That is, we parametrized separate turning-angle and step-length distributions representative of steps with durations of $\Delta t = 1, 2, 3, 4$ and 5 (which corresponds to the maximum forgiveness level we tested for). Some step durations only rarely occurred at low levels of missingness, thus complicating parametrization of the associated distributions. To facilitate estimation of dynamic distribution parameters across all Δt (Figure S2), we resampled data to different step durations using the `track_resample` function from

8

the `amt` package (Signer et al., 2019) before fitting tentative parameters. This ensured a sufficient number of steps for each step duration to estimate associated parameters. An alternative approach would be to increase missingness in the data even further, thus introducing a larger number of longer step durations.

2.6   Step-Selection Functions

We implemented a baseline *uncorrected* approach and four alternative iSSF approaches that mainly differed in the way in which random steps were generated, but sometimes also in the model call that was used to estimate parameters (Figure 3). In the *uncorrected* approach, we treated data as if they were regular, ignoring potential issues arising from having variable step durations. When forgiveness was one, this approach corresponded to the traditional iSSF approach. All other approaches were targeted towards reducing potential biases arising from the inclusion of steps with irregular step durations. Irrespective of the approach employed, we paired each observed step with a total of 200 random steps:

- *Uncorrected*: In the uncorrected approach, we generated random steps by sampling step lengths and turning angles from statistical distributions fitted to steps with step durations of $\Delta t = 1$, regardless of the forgiveness value or observed step durations. Thus, this approach ignored any potential effect of step duration when generating random step lengths and turning angles.

- *Imputed*: In this approach, we sampled step lengths and turning angles from statistical distributions fitted to observed steps with a step duration of $\Delta t = 1$. However, prior to generating random steps, we imputed missing fixes using predictions from a simple movement model. Specifically, we fitted a single-state movement model (Johnson et al., 2008) to the simulated trajectories and used the parametrized model to predict coordinates for all missing animal locations. For this, we used the functions `crwFit` and `crwPredict` from the `crawl` R-package (Johnson et al., 2022). Although the `crwFit` function provides capabilities to incorporate location measurement error, we assumed animal locations were measured without error. The imputation resulted in a complete dataset without any missing animal locations, such that each trajectory consisted of a single continuous burst of locations equally spaced in time. As such, the imputation approach is not affected by the forgiveness level.

- *Naïve*: In the *naïve* approach, we again sampled step lengths and turning angles from regular distributions fitted to steps with step durations of $\Delta t = 1$. However, we linearly

9

scaled the sampled step lengths depending on the step durations of the observed steps. For instance, we doubled the sampled step lengths for any observed step with a step duration of $\Delta t = 2$. This approach naïvely assumed that step lengths scale linearly with step durations, which is unlikely to be true, as most animals don't move in straight lines between successive observations. Furthermore, the linear approximation is likely to get worse as step duration increases (i.e., as the forgiveness value increases). Since it is not clear how turning angles should scale with step duration, we did not adjust the sampled turning angles.

- *Dynamic+Model*: In the *dynamic+model* approach, we sampled step lengths and turning angles from dynamic distributions that were fit to different step durations. That is, for observed steps with step duration of $\Delta t = 2$, we sampled step lengths and turning angles from distributions fit to observed steps with $\Delta t = 2$. We then included interactions between the step duration and other step descriptors (e.g., sl, log(sl), cos(ta)), allowing us to update movement parameters for each step duration separately. To avoid numerical instabilities with the conditional logistic regression model, we only included steps with durations $\Delta t > 1$ if the respective duration was represented at least 5 times in the rarefied dataset.

- *Multistep*: In the *multistep* approach, we sampled step lengths and turning angles from statistical distributions fitted to observed steps with step durations of $\Delta t = 1$. We then generated a sequence of random steps such that their combined step duration equaled the step duration of each observed step. For instance, for an observed step with step duration of $\Delta t = 2$, we generated sets of two random steps, which we then concatenated into a "random path". The paths were then simplified to straight lines connecting the first and last coordinate of each path, which represented the final random step.

Together, an observed step and its 200 associated random steps formed a *stratum* that received a unique ID. At the end of each step, we extracted covariate values from the underlying covariate layers.

2.7 Conditional Logistic Regression Model

We estimated movement and habitat-selection parameters for the simulation scenarios presented in Figure 3 using conditional logistic regression, implemented using the `clogit` function in the R-package `survival` (Therneau et al., 2023). We defined a binary response

10

variable (observed) indicating if a step was an observed (scored 1) or a random step (scored 0) and used the step's ID as a stratification variable. We included habitat covariates (dist, elev, forest) and step descriptors (sl, log(sl), cos(ta)) as predictors in the regression model. For the *dynamic+model* approach, we also included interactions between the step duration, coded as a factor, and step descriptors. To update tentative movement parameters (denoted by the subscript $_0$) and obtain the selection-free movement kernel (denoted by the ˆ symbol), we employed the formulas provided in (Avgar et al., 2016; Fieberg et al., 2021). Specifically, we updated the shape ($\hat{k}$) and scale ($\hat{\theta}$) parameters of the step-length distribution (gamma) using:

$$\hat{k} = k_0 + \beta_{log(sl)}$$

$$\hat{\theta} = \frac{1}{\frac{1}{\theta_0} - \beta_{sl}}$$

We updated the concentration parameter ($\hat{\kappa}$) of the turning-angle distribution (von Mises) using:

$$\hat{\kappa} = \kappa_0 + \beta_{cos(ta)}$$

We kept track of the estimates of the updated movement ($\hat{k}$, $\hat{\theta}$, and $\hat{\kappa}$) parameters and the habitat-selection ($\hat{\beta}_{dist}$, $\hat{\beta}_{elev}$, $\hat{\beta}_{forest}$) parameters, and compared them to the true simulation parameters. We also quantified model accuracy via the root-mean-square error (RMSE).

3 Results

Results were qualitatively similar for all three landscape autocorrelation scenarios and for different combinations of missingness and forgiveness (Figure S4). Here, we report on results for a landscape with autocorrelation of 20, while either holding constant missingness at a conservative 20% (Figure 5) or the forgiveness level at two (Figure 6) (results for all other combinations are summarized in Figure S4). The *imputation* approach resulted in biased estimators of β_{dist} and β_{forest}, whereas all other approaches were able to recover the parameters of the habitat-selection function with minimal bias (Figure 5). Note, the *imputation* approach always starts with a full trajectory and is therefore unaffected by the forgiveness level. For all other methods, increasing the forgiveness from 1 to 5 improved the precision of the estimators of habitat-selection parameters without introducing noticeable bias, with the biggest gains in precision and RMSE occurring when moving from a forgiveness

11

of one to a forgiveness of two (Figure 5). This highlights the potential benefits of leveraging additional data compared to the traditional approach, which uses only bursts of regular data (represented by the *uncorrected* approach and forgiveness = 1).

The *uncorrected*, *naïve*, and *imputation* approaches resulted in biased estimators of the parameters in the movement kernel, particularly for high values of forgiveness (Figure 5a) and high levels of missingness (Figure 6a). The *imputation* approach appeared to perform particularly poorly at estimating the concentration parameter of the turning-angle distribution (Figure 6). The *multistep* and *dynamic+model* approaches resulted in unbiased estimators of parameters in the step-length distribution, but estimators of the concentration parameter exhibited a slight bias. This bias was, however, much smaller than we observed with the other approaches we considered. Increasing missingness negatively influenced the precision and accuracy of estimates, yet its impact could be dampened using the *dynamic+model* and *multistep* approaches (Figure 6b).

4 Case Study

To showcase the applicability of the *dynamic+model* approach, which appeared to perform best with simulated data, we conducted a case study with real GPS data obtained on "Apollo", a spotted hyena (*Crocuta crocuta*) inhabiting the Okavango Delta ecosystem in northern Botswana. Apollo's data were collected between 2007 and 2011 using GPS radio collars (*GPS Plus*; Vectronic Aerospace GmbH, Berlin, Germany) and comprised 9,316 GPS locations (details in Cozzi, 2013 and Cozzi et al., 2015). Because hyenas are nocturnal (Cozzi et al., 2012), GPS collars were set to record data at two-hourly intervals between 18:00 and 06:00 o'clock, and to only record a single location at noon. For simplicity, we only considered nightly bursts and removed all locations obtained at noon. Missingness in this dataset was low (< 10 %, Cozzi et al., 2015) and to better showcase the usefulness of the *dynamic+model* approach, we thinned the data by randomly removing 25% of the obtained locations. As spatial covariate layers, we used Water, DistanceToWater and Trees (Figure S5). Water was a binary variable representing major rivers and areas inundated by floodwater, whereas DistanceToWater was a continuous variable indicating the distance (in meters) to the nearest pixel categorized as water. Trees was a continuous variable indicating the percent tree cover in each pixel. We resampled all layers to a common resolution of 250 m by 250 m and merged them into a single raster-stack (Figure S5). The derivation of each covariate layer is described in detail in Hofmann et al., 2021. We dynamically fitted step-length and turning-angle distributions to steps with step durations of 2, 4, and 6 hours, respectively,

assuming a gamma distribution for step lengths and a von Mises distribution for turning angles. Instead of resampling the observed track to different step durations when fitting dynamic distributions (like we did in the simulation study), we introduced a larger amount of steps with step durations longer than two hours by thinning the data again (by another 10%). The benefit of this approach was that steps with irregular step durations occurred more randomly and were not limited to the hours specified by the resampling algorithm. Finally, we used iSSFs with the *dynamic+model* approach to estimate the habitat-selection function and movement kernel of Apollo. For this, we considered three cases:

- *F1*: We assumed a forgiveness of one (i.e., only steps with a regular step duration of 2 hours), which is akin to conducting a traditional iSSA.

- *F3-S*: We assumed a forgiveness of three (i.e., considered steps with step durations of up to three times the regular step duration) and included interactions between the step duration (Δt) and step descriptors (sl, log(sl), and cos(ta)) in the regression model.

- *F3-SH*: We assumed a forgiveness of three (i.e., considered steps with step durations of up to three times the regular step duration) and included interactions between the step duration (Δt) and step descriptors (sl, log(sl), and cos(ta)), as well as between the step duration and habitat covariates in the regression model.

Notably, we included F3-SH to investigate if including interactions between the step duration and habitat-covariates would provide insights into scale-dependent habitat selection. In all cases, we generated 200 random steps and extracted spatial covariates at the end of observed and random steps. We then fit the three models using the conditional logistic regression framework as implemented in the `survival` R-package (Therneau et al., 2023). Lastly, we computed updated movement parameters for a regular step duration of 2 hours.

Results from the iSSF models show that increasing the level of forgiveness led to improvements in estimator precision (Figure 7). This was achieved by increasing the effective sample size from 2,179 to 4,505 valid steps (Table S1). The improvement in estimator precision was weaker for F3-SH than for F3-S, as the F3-SH model was more complex due to inclusion of additional interaction terms. Point estimates for the habitat-selection and movement parameters were similar for all 3 models, and evidence for scale dependency in habitat selection was fairly weak. F3-S and F3-SH had similar AIC scores ($\Delta AIC \leq 1$; Table S1), and the interaction terms were statistically significant only for the step duration of 6 hours and only for one of the habitat covariates (Table S1).

13

5 Discussion

We conducted a simulation study with known habitat and movement parameters to investigate if retaining irregular animal locations via increased forgiveness improves or worsens parameter estimation in iSSFs. We also tested the performance of four different approaches that attempt to correct for potential biases introduced by using temporally irregular data, and we compared them to an uncorrected baseline approach. Our results demonstrated that retaining irregular animal locations can improve the precision of estimators of habitat-selection parameters but may lead to biased estimators of the parameters in the movement kernel. Overall, our results highlight the potential benefits of leveraging irregular animal locations, especially if an appropriate method for handling irregular data is chosen.

The *uncorrected* baseline approach ignored the fact that increasing forgiveness introduced irregularity in the data. Consequently, estimators of the parameters in the movement kernel were increasingly biased as forgiveness increased due to the inclusion of steps with varying step durations when fitting the model. Steps with longer step durations tended to have larger step lengths and less directed turning angles, which led to an overestimation of θ and underestimation of k and κ. Yet, estimators of the habitat-selection parameters remained unbiased or nearly so, even at high levels of forgiveness, and they were more precise than the standard estimator (represented by the *uncorrected* approach with forgiveness = 1). These results highlight the potential benefits that can be reaped by including additional data.

Similarly, the *naïve* approach performed well when estimating habitat-selection parameters but resulted in biased estimators of movement parameters, especially for large forgiveness values. This result was unsurprising given that our simulated trajectories were tortuous, and therefore, step lengths were not linearly related to step durations. Indeed, we found that, although there was a near linear relationship between step duration and the (tentative) scale parameter (θ_0), the relationships between step duration and the (tentative) movement parameters κ_0 and k_0 were non-linear (Figure S4). Overall, the usefulness of this approach appears highly limited, as it is often not clear by what factor distributional parameters for step lengths and turning angles should be multiplied to match the observed step duration.

The *dynamic+model* approach provided a flexible, easily implementable, and powerful framework for retrieving precise and unbiased estimators of the step-length and habitat-selection parameters, irrespective of the forgiveness level. The estimator of the concentration parameter exhibited some bias, but less than when using the *uncorrected* and *naïve* approaches. To implement the *dynamic+model* approach, we included interactions between step descriptors (sl, log(sl), and cos(ta)) and step duration in the conditional logistic re-

14

gression model. This allowed the parameters of the movement kernel to depend on the step duration. A complication, however, is that turning angles are influenced by the step duration of both the current and previous step (Figure S3). The bias in the concentration parameter likely arose from only accounting for the step duration associated with the current step and not the previous one. Moreover, fitting tentative distributions for different step durations can be challenging due to some step durations occurring only rarely. However, by resampling observed animal locations to different step durations using the `track_resample` function from the `amt` R-package (Signer et al., 2019) the needed data can easily be generated. We included step duration as a categorical covariate, yet there may be times when it would be advantageous to treat it as a continuous covariate (e.g., with its effect modeled using a low-degree polynomial or regression spline with few degrees of freedom). Treating step duration as a continuous variable may help to alleviate convergence issues in cases where some step durations are rare, and it might allow applying the *dynamic+model* approach to data that are entirely irregular.

The *multistep* approach also performed well and was relatively easy to implement. This approach is somewhat *ad hoc* in that it uses the tentative movement parameters to generate random steps to match observed steps with longer step durations (in multiples of Δt). It is similar to, but slightly less principled, than the approach developed by Vales et al. (2022), which formally constructs the likelihood for multistep durations by integrating out the missing steps. An advantage of this latter approach is that one can also attempt to account for non-random missingness by explicitly modeling factors related to the probability of obtaining a successful location (Vales et al., 2022). Nonetheless, integrating over the missing steps, as in Vales et al. (2022), can be computationally intensive and prohibitive with large data sets. Another downside of both of these approaches (the *multistep* approach and the approach of Vales et al., 2022) is that they can only be applied in cases where step durations are a fixed multiple of the regular step duration; i.e., unlike the *dynamic+model* approach, they cannot be applied when data are highly irregular.

Of the methods we considered, the *imputation* approach performed the worst. It resulted in biased estimators of parameters in both the habitat-selection function and the movement kernel. This bias likely resulted from using an overly simplistic movement model to impute missing fixes. Moreover, the imputation procedure may have led to imputed animal locations that masked important selection properties, therefore leading to inaccurate parameter estimates. While this approach appears to perform well with hidden Markov movement models (McClintock, 2017), we advise against its use with iSSFs.

For the scenarios we considered in our simulation study, the estimators of habitat-selection parameters were insensitive to the inclusion of irregular data and performed well, except for the *imputation* approach. This suggests that accounting for irregular step durations may not be particularly important if one is only interested in the habitat-selection function. When the movement kernel is also of interest, we suggest the *dynamic+model* approach, since it is flexible, easy to implement, and allows one to use more data than the traditional approach that requires bursts of regular data, leading to more precise estimators.

Several authors have emphasized that movement and habitat-selection parameters in an SSF are scale dependent and should be expected to change as the sampling frequency changes (see for example Avgar et al., 2016; Signer et al., 2017; Fieberg et al., 2021). Furthermore, Barnett and Moorcroft (2008) developed an analytical framework for investigating scale dependence and showed that habitat-selection parameters should depend on the relative width of the movement kernel in relation to habitat heterogeneity. Thus, the relative insensitivity of the habitat-selection parameters to the inclusion of steps with varying step duration was somewhat unexpected. It would be interesting to explore the robustness of this result across a wider range of simulation scenarios in the future.

More generally, the spatial scale of a habitat-selection analysis has been recognized as an important factor, which is why Johnson (1980) proposed a hierarchical framework for examining habitat-selection across different orders (e.g., species range, individual home range, within a home range). Johnson's proposed framework acknowledges that habitat-selection may act differently at different scales, and that the interpretation of ecological processes changes depending on the spatial scale at which they are investigated (Wiens, 1989; Levin, 1992). This understanding has encouraged scientists to conduct extensive scaling analyses and to comprehensively examine habitat-selection at multiple scales (DeCesare et al., 2012; McGarigal et al., 2016; Pitman et al., 2017; Zeller et al., 2017). In studies employing SSAs, the issue of scale is often neglected, and data are most frequently analyzed at the spatio-temporal scale at which they were collected. This choice maximizes the number of locations that can be used in the analysis, yet prevents a thorough understanding of scale dependency. The use of irregular data in SSAs poses another challenge, as steps with unequal step durations may reflect selection processes occurring at different scales. The severity of this issue obviously depends on the original sampling frequency, the degree of missingness, and the scale at which animals are making decisions that are relevant in terms of their movement behavior and habitat selection. By including irregular animal locations via increased forgiveness, we may therefore average over selection processes occurring at multiple scales,

which could produce estimates of habitat-selection parameters that are misleading due to contradictory effect signs at different scales. To better account for such scale-dependent processes, it may be beneficial to include interactions between step duration and habitat features (e.g., dist, elev, forest), thus allowing habitat-selection parameters to also vary as a function of step duration. We demonstrated how this could be implemented in the case study.

It is important to note that we considered a limited number of scenarios in our simulation study. For instance, we assumed that animal locations were missing at random, i.e., failure to obtain a fix was unrelated to habitat types, time of the day, etc. However, several studies have shown that missingness is often non-random and related to difficulties with satellite transmission due to topography (Lewis et al., 2007), canopy cover (Phillips et al., 1998; DeCesare et al., 2005; Hansen & Riggs, 2008), time of the day (Graves & Waller, 2006), animal behavior (Mattisson et al., 2010), or collar orientation (D'eon & Delparte, 2005). In fact, Vales et al. (2022) highlighted that missingness and the associated under-representation of certain habitat types may lead to biased estimators of parameters in iSSFs, but that accounting for the probability of obtaining a location in differing environmental conditions may alleviate this bias. Future studies should strive to further investigate these relationships and examine how our proposed approaches perform when missingness is habitat-dependent.

A major benefit of using iSSFs is the ability to allow an individual's movement kernel to depend on local habitat features (Avgar et al., 2016). In our simulation study, we considered simplified scenarios where the movement kernel was unchanging, which simplified the simulation and inference. Nevertheless, such interactions often play a crucial role in real ecosystems. For instance, Dickie et al. (2020) employed iSSFs and revealed that several large mammal species moved faster while on linear features. Similarly, Hofmann, Cozzi, McNutt, et al. (2023) found that African wild dogs moved significantly slower and less directed in areas that were covered by floodwater. Future studies could investigate simulation scenarios in which individuals alter their movement tendencies in response to local environmental features (i.e., models with habitat dependent movement kernels) and examine how this influences the robustness of our proposed approaches.

While our results suggest that irregularity due to missing animal locations can effectively be accounted for in iSSFs and that increasing the forgiveness, thus allowing for inclusion of irregular data, improves estimator precision, we also found a decreasing marginal benefit of increased forgiveness. In fact, increasing the forgiveness beyond a value of two (i.e., allowing for steps of twice the regular step duration) only marginally improved model performance

17

in our case. This can also be seen in Figure 2, which shows that the largest number of steps that can be gained is when increasing the forgiveness level from one to two. Having a higher forgiveness beyond two may thus not even be necessary, therefore limiting the need to correct biases emerging from the inclusion of irregular data.

Although we focused on the case of missing location data, the proposed approaches may also prove useful for situations where sampling is irregular for other reasons. For example, it is not uncommon to adjust sampling regimes after a preliminary phase, following improvements to collar-battery-lifetime, or for sampling rates to vary depending on type and manufacturer of the collar (Brown et al., 2023). Similarly, it is common practice to adjust the GPS regime to the biology of the focal species and only record data during a specific time of the day (e.g., Broekhuis et al., 2013; Cozzi, 2013; Elliot et al., 2014). These irregularities might be addressed using the *dynamic+model* approach, with interactions between step duration and movement descriptors. Interactions between step duration and habitat-selection covariates should also be considered, particularly if the sampling regime is adjusted to coincide with changes in animal behavior. We expect this approach will work fairly well in many cases, but we might expect a slight bias in the estimated concentration parameters, as observed in our simulation study.

Our study contributes to the growing body of literature that extends iSSFs and improves the method's robustness under various conditions. This includes approaches for modeling irregular data (Munden et al., 2021; Eisaguirre et al., 2024), accounting for spatial dependence among residuals (Arce Guillen et al., 2023), methodological frameworks for fitting iSSFs with random slopes (Muff et al., 2020) and random smooths (Klappstein et al., 2024), incorporating the probability of successfully obtaining an animal location in different habitat conditions (Vales et al., 2022), and considering the behavioral states of the tracked animals (Klappstein et al., 2023; Pohle et al., 2023).

In conclusion, our study shows that inclusion of irregular animal locations can improve model performance, yet only when an appropriate approach to account for irregularity is selected. Here, the *dynamic+model* and *multistep* approaches performed well and resulted in improved estimators of habitat-selection and movement parameters, even at elevated levels of missingness and forgiveness. Both methods are easy to implement, and the associated models can readily be fitted using the R-packages `amt` (Signer et al., 2019), `survival` (Therneau et al., 2023), `coxme` (Therneau, 2022), and `mgcv` (Wood, 2011; 2017; Klappstein et al., 2024). To facilitate uptake and encourage use of the proposed approaches among practitioners, we provide all of our codes through an online repository, which includes an example application

18

of the *dynamic+model* approach. With this, we hope practitioners will rethink the common use of discarding large portions of data and instead use methods that can accommodate irregular data.

6 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.

7 Data Availability

Code to reproduce this study is available through the University of Minnesota's Data repository (https://doi.org/10.13020/6wcd-6s43, Hofmann, Cozzi, and Fieberg, 2023). The repository also includes two example analyses that showcase the application of the *dynamic+model* approach to simulated data, as well as to the GPS data of Apollo.

8 Acknowledgements & Funding

19

References

Abrahms, B., Sawyer, S. C., Jordan, N. R., McNutt, J. W., Wilson, A. M., & Brashares, J. S. (2017). Does Wildlife Resource Selection Accurately Inform Corridor Conservation? *Journal of Applied Ecology*, *54*(2), 412–422. https://doi.org/10.1111/1365-2664.12714

Arce Guillen, R., Lindgren, F., Muff, S., Glass, T. W., Breed, G. A., & Schlägel, U. E. (2023). Accounting for Unobserved Spatial Variation in Step Selection Analyses of Animal Movement Via Spatial Random Effects. *Methods in Ecology and Evolution*, 2041–210X.14208. https://doi.org/10.1111/2041-210X.14208

Avgar, T., Potts, J. R., Lewis, M. A., & 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. https://doi.org/10.1111/2041-210X.12528

Barnett, A. H., & Moorcroft, P. R. (2008). Analytic Steady-State Space Use Patterns and Rapid Computations in Mechanistic Home Range Analysis. *Journal of Mathematical Biology*, *57*(1), 139–159. https://doi.org/10.1007/s00285-007-0149-8

Beardsworth, C. E., Gobbens, E., van Maarseveen, F., Denissen, B., Dekinga, A., Nathan, R., Toledo, S., & Bijleveld, A. I. (2022). Validating Atlas: A Regional-Scale High-Throughput Tracking System. *Methods in Ecology and Evolution*, *13*(9), 1990–2004. https://doi.org/10.1111/2041-210X.13913

Brand, T. van den. (2023, August 30). *Ggh4x: Hacks for 'ggplot2'* (Version 0.2.6). https://cran.rstudio.com/web/packages/ggh4x/index.html

Broekhuis, F., Madsen, E. K., & Klaassen, B. (2019). Predators and Pastoralists: How Anthropogenic Pressures Inside Wildlife Areas Influence Carnivore Space Use and Movement Behaviour. *Animal Conservation*, *22*(4), 404–416. https://doi.org/10.1111/acv.12483

Broekhuis, F., Cozzi, G., Valeix, M., McNutt, J. W., & Macdonald, D. W. (2013). Risk Avoidance in Sympatric Large Carnivores: Reactive or Predictive? *Journal of Animal Ecology*, *82*(5), 1098–1105. https://doi.org/10.1111/1365-2656.12077

Brown, M. B., Fennessy, J. T., Crego, R. D., Fleming, C. H., Alves, J., Brandlová, K., Fennessy, S., Ferguson, S., Hauptfleisch, M., Hejcmanova, P., Hoffman, R., Leimgruber, P., Masiaine, S., McQualter, K., Mueller, T., Muller, B., Muneza, A., O'Connor, D., Olivier, A. J., . . . Stabach, J. (2023). Ranging Behaviours Across Ecological and Anthropogenic Disturbance Gradients: A Pan-African Perspective of Giraffe (Giraffa spp.) Space Use. *Proceedings of the Royal Society B: Biological Sciences*, *290*(2001), 20230912. https://doi.org/10.1098/rspb.2023.0912

Buchholtz, E. K., Stronza, A., Songhurst, A., McCulloch, G., & Fitzgerald, L. A. (2020). Using Landscape Connectivity to Predict Human-Wildlife Conflict. *Biological Conservation*, *248*, 108677. https://doi.org/10.1016/j.biocon.2020.108677

Cagnacci, F., Boitani, L., Powell, R. A., & 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. https://doi.org/10.1098/rstb.2010.0107

Cozzi, G. (2013). *Patterns of Habitat Use and Segregation Among African Large Carnivores* (Doctoral dissertation). University of Zurich. Zurich.

Cozzi, G., Behr, D. M., Webster, H. S., Claase, M., Bryce, C. M., Modise, B., Mcnutt, J. W., & Ozgul, A. (2020). African Wild Dog Dispersal and Implications for Management. *The Journal of Wildlife Management*, *84*(4), 614–621. https://doi.org/10.1002/jwmg.21841

Cozzi, G., Börger, L., Hutter, P., Abegg, D., Beran, C., McNutt, J. W., & Ozgul, A. (2015). Effects of Trophy Hunting Leftovers on the Ranging Behaviour of Large Carnivores: A Case Study on Spotted Hyenas. *PLOS ONE*, *10*(3), e0121471. https://doi.org/10.1371/journal.pone.0121471

Cozzi, G., Broekhuis, F., McNutt, J. W., Turnbull, L. A., Macdonald, D. W., & Schmid, B. (2012). Fear of the Dark or Dinner by Moonlight? Reduced Temporal Partitioning

Among Africa's Large Carnivores. *Ecology*, *93*(12), 2590–2599. https://doi.org/10.1890/12-0017.1

Cozzi, G., Maag, N., Börger, L., Clutton-Brock, T. H., & Ozgul, A. (2018). Socially Informed Dispersal in a Territorial Cooperative Breeder. *Journal of Animal Ecology*, *87*(3), 838–849. https://doi.org/10.1111/1365-2656.12795

DeCesare, N. J., Hebblewhite, M., Schmiegelow, F., Hervieux, D., McDermid, G. J., Neufeld, L., Bradley, M., Whittington, J., Smith, K. G., Morgantini, L. E., Wheatley, M., & Musiani, M. (2012). Transcending Scale Dependence in Identifying Habitat with Resource Selection Functions. *Ecological Applications*, *22*(4), 1068–1083. https://doi.org/10.1890/11-1610.1

DeCesare, N. J., Squires, J. R., & Kolbe, J. A. (2005). Effect of Forest Canopy on GPS-Based Movement Data. *Wildlife Society Bulletin*, *33*(3), 935–941. https://doi.org/10.2193/0091-7648(2005)33[935:EOFCOG]2.0.CO;2

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, *64*(4), 1–34. https://doi.org/10.18637/jss.v064.i04

D'eon, R. G., & 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. https://doi.org/10.1111/j.1365-2664.2005.01010.x

Dickie, M., McNay, S. R., Sutherland, G. D., Cody, M., & Avgar, T. (2020). Corridors or Risk? Movement Along, and Use of, Linear Features Varies Predictably Among Large Mammal Predator and Prey Species. *Journal of Animal Ecology*, *89*(2), 623–634. https://doi.org/10.1111/1365-2656.13130

Duchesne, T., Fortin, D., & 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. https://doi.org/10.1371/journal.pone.0122947

Eisaguirre, J. M., Williams, P. J., & Hooten, M. B. (2024). Rayleigh Step-Selection Functions and Connections to Continuous-Time Mechanistic Movement Models. *Movement Ecology*, *12*(1), 14. https://doi.org/10.1186/s40462-023-00442-w

Elliot, N. B., Cushman, S. A., Macdonald, D. W., & 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. https://doi.org/10.1111/1365-2664.12282

Enns, G. E., Jex, B., & Boyce, M. S. (2023). Diverse Migration Patterns and Seasonal Habitat Use of Stone's Sheep (Ovis dalli stonei). *PeerJ*, *11*, e15215. https://doi.org/10.7717/peerj.15215

Fattebert, J., Robinson, H. S., Balme, G., Slotow, R., & Hunter, L. (2015). Structural Habitat Predicts Functional Dispersal Habitat of a Large Carnivore: How Leopards Change Spots. *Ecological Applications*, *25*(7), 1911–1921. https://doi.org/10.1890/14-1631.1

Fieberg, J., Signer, J., Smith, B., & Avgar, T. (2021). A 'How to' Guide for Interpreting Parameters in Habitat-Selection Analyses. *Journal of Animal Ecology*, *90*(5), 1027–1043. https://doi.org/10.1111/1365-2656.13441

Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T., & Mao, J. S. (2005). Wolves Influence Elk Movements: Behavior Shapes a Trophic Cascade in Yellowstone National Park. *Ecology*, *86*(5), 1320–1330. https://doi.org/10.1890/04-0953

Frair, J. L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N. J., & Pedrotti, L. (2010). Resolving Issues of Imprecise and Habitat-Biased Locations in Ecological Analyses Using GPS Telemetry Data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1550), 2187–2200. https://doi.org/10.1098/rstb.2010.0084

Graves, T. A., & Waller, J. S. (2006). Understanding the Causes of Missed Global Positioning System Telemetry Fixes. *Journal of Wildlife Management*, *70*(3), 844–851. https://doi.org/10.2193/0022-541X(2006)70[844:UTCOMG]2.0.CO;2

Hansen, M. C., & Riggs, R. A. (2008). Accuracy, Precision, and Observation Rates of Global Positioning System Telemetry Collars. *The Journal of Wildlife Management*, *72*(2), 518–526. https://doi.org/10.2193/2006-493

Hijmans, R. J., Bivand, R., Pebesma, E., & Sumner, M. D. (2023, September 6). *Terra: Spatial Data Analysis* (Version 1.7-38). https://cran.r-project.org/web/packages/terra/index.html

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., ... Balkenhol, N. (2019). Right on Track? Performance of Satellite Telemetry in Terrestrial Wildlife Research. *PLoS ONE*, *14*(5), e0216223. https://doi.org/10.1371/journal.pone.0216223

Hofmann, D. D., Behr, D. M., McNutt, J. W., Ozgul, A., & Cozzi, G. (2021). Bound Within Boundaries: Do Protected Areas Cover Movement Corridors of Their Most Mobile, Protected Species? *Journal of Applied Ecology*, *58*(6), 1133–1144. https://doi.org/10.1111/1365-2664.13868

Hofmann, D. D., Cozzi, G., & Fieberg, J. R. (2023). *R Code Associated with Methods for Implementing Integrated Step-Selection Functions with Incomplete Data*. Data Repository for the University of Minnesota (DRUM). https://doi.org/10.13020/6WCD-6S43

Hofmann, D. D., Cozzi, G., McNutt, J. W., Ozgul, A., & Behr, D. M. (2023). A Three-Step Approach for Assessing Landscape Connectivity Via Simulated Dispersal: African Wild Dog Case Study. *Landscape Ecology*, *38*(4), 981–998. https://doi.org/10.1007/s10980-023-01602-4

Johnson, D., London, J. M., & McClintock, B. (2022, October 6). *NMML/crawl: Last CRAN release* (Version v2.3.0-1). https://doi.org/10.5281/ZENODO.596464

Johnson, D. S., London, J. M., Lea, M.-A., & Durban, J. W. (2008). Continuous-Time Correlated Random Walk Model for Animal Telemetry Data. *Ecology*, *89*(5), 1208–1215. https://doi.org/10.1890/07-1032.1

Johnson, D. H. (1980). The Comparison of Usage and Availability Measurements for Evaluating Resource Preference. *Ecology*, *61*(1), 65–71. https://doi.org/10.2307/1937156

Kassambara, A. (2023, February 10). *Ggpubr: 'ggplot2' Based Publication Ready Plots* (Version 0.6.0). https://cran.r-project.org/web/packages/ggpubr/index.html

Kays, R., Crofoot, M. C., Jetz, W., & Wikelski, M. (2015). Terrestrial Animal Tracking as an Eye on Life and Planet. *Science*, *348*(6240), aaa2478–aaa2478. https://doi.org/10.1126/science.aaa2478

Kéry, M., & Royle, J. A. (2016). Introduction to Data Simulation. In *Applied Hierarchical Modeling in Ecology* (pp. 123–143). Elsevier. https://doi.org/10.1016/B978-0-12-801378-6.00004-7

Kim, D., Thompson, P. R., Wolfson, D., Merkle, J., Oliveira-Santos, L. G. R., Forester, J. D., Avgar, T., Lewis, M. A., & Fieberg, J. (2023, August 17). *Identifying Signals of Memory from Observations of Animal Movements in Plato's Cave* (preprint). Ecology. https://doi.org/10.1101/2023.08.15.553411

Klappstein, N. J., Thomas, L., & Michelot, T. (2023). Flexible Hidden Markov Models for Behaviour-Dependent Habitat Selection. *Movement Ecology*, *11*(1), 30. https://doi.org/10.1186/s40462-023-00392-3

Klappstein, N., Michelot, T., Fieberg, J., Pedersen, E., Field, C., & Flemming, J. M. (2024, January 10). *Step Selection Analysis with Non-Linear and Random Effects in mgcv*. https://doi.org/10.1101/2024.01.05.574363

Kuang, K. (2022, April 28). *Pbmcapply: Tracking the Progress of Mc*pply with Progress Bar* (Version 1.5.1). https://cran.r-project.org/web/packages/pbmcapply/index.html

Levin, S. A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology*, *73*(6), 1943–1967. https://doi.org/10.2307/1941447

Lewis, J. S., Rachlow, J. L., Garton, E. O., & 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. https://doi.org/10.1111/j.1365-2664.2007.01286.x

Mattisson, J., Andrén, H., Persson, J., & Segerström, P. (2010). Effects of Species Behavior on Global Positioning System Collar Fix Rates. *Journal of Wildlife Management*, *74*(3), 557–563. https://doi.org/10.2193/2009-157

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. https://doi.org/10.1007/s13253-017-0285-6

McGarigal, K., Wan, H. Y., Zeller, K. A., Timm, B. C., & Cushman, S. A. (2016). Multi-Scale Habitat Selection Modeling: A Review and Outlook. *Landscape Ecology*, *31*(6), 1161–1175. https://doi.org/10.1007/s10980-016-0374-x

Michelot, T., Klappstein, N. J., Potts, J. R., & Fieberg, J. (2024). Understanding Step Selection Analysis Through Numerical Integration. *Methods in Ecology and Evolution*, *15*(1), 24–35. https://doi.org/10.1111/2041-210X.14248

Michelot, T., Langrock, R., & Patterson, T. A. (2016). moveHMM: An R Package for the Statistical Modelling of Animal Movement Data Using Hidden Markov Models. *Methods in Ecology and Evolution*, *7*(11), 1308–1315. https://doi.org/10.1111/2041-210X.12578

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*(1), 80–92. https://doi.org/10.1111/1365-2656.13087

Munden, R., Börger, L., Wilson, R. P., Redcliffe, J., Brown, R., Garel, M., & 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. https://doi.org/10.1111/2041-210X.13574

Nathan, R. (2008). An Emerging Movement Ecology Paradigm. *Proceedings of the National Academy of Sciences*, *105*(49), 19050–19051. https://doi.org/10.1073/pnas.0808918105

Nathan, R., Monk, C. T., Arlinghaus, R., Adam, T., Alós, J., Assaf, M., Baktoft, H., Beardsworth, C. E., Bertram, M. G., Bijleveld, A. I., Brodin, T., Brooks, J. L., Campos-Candela, A., Cooke, S. J., Gjelland, K. Ø., Gupte, P. R., Harel, R., Hellström, G., Jeltsch, F., ... Jarić, I. (2022). Big-Data Approaches Lead to an Increased Understanding of the Ecology of Animal Movement. *Science*, *375*(6582), eabg1780. https://doi.org/10.1126/science.abg1780

Phillips, K. A., Elvey, C. R., & Abercrombie, C. L. (1998). Applying GPS to the Study of Primate Ecology: A Useful Tool? *American Journal of Primatology*, *46*(2), 167–172. https://doi.org/10.1002/(SICI)1098-2345(1998)46:2⟨167::AID-AJP6⟩3.0.CO;2-U

Pitman, R. T., Fattebert, J., Williams, S. T., Williams, K. S., Hill, R. A., Hunter, L. T. B., Robinson, H., Power, J., Swanepoel, L., Slotow, R., & Balme, G. A. (2017). Cats, Connectivity and Conservation: Incorporating Data Sets and Integrating Scales for Wildlife Management. *Journal of Applied Ecology*, *54*(6), 1687–1698. https://doi.org/10.1111/1365-2664.12851

Pohle, J., Signer, J., Eccard, J. A., Dammhahn, M., & Schlägel, U. E. (2023). How to Account for Behavioural States in Step-Selection Analysis: A Model Comparison. https://doi.org/10.48550/ARXIV.2304.12964

Potts, J. R., Börger, L., Scantlebury, D. M., Bennett, N. C., Alagaili, A., & Wilson, R. P. (2018). Finding Turning-Points in Ultra-High-Resolution Animal Movement Data. *Methods in Ecology and Evolution*, *9*(10), 2091–2101. https://doi.org/10.1111/2041-210X.13056

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. manual. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

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., & Trombulak, S. C. (2012). The Role of Landscape Connectivity in Planning and Implementing Conservation and Restoration Priorities. *Issues in*

Ecology, (16), 1–23. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84868160117&partnerID=40&md5=1aeb7d32cfc24db10a506ee2caa82267

Rumiano, F., Wielgus, E., Miguel, E., Chamaillé-Jammes, S., Valls-Fox, H., Cornélis, D., Garine-Wichatitsky, M. D., Fritz, H., Caron, A., & 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. https://doi.org/10.3390/rs12193218

Signer, J., Fieberg, J., Reineking, B., Schlägel, U., Smith, B., Balkenhol, N., & Avgar, T. (2023). Simulating Animal Space Use from Fitted Integrated Step-Selection Functions (iSSF). *bioRxiv: the Preprint Server for Biology.* https://doi.org/10.1101/2023.08.10.552754

Signer, J., Fieberg, J., & Avgar, T. (2017). Estimating Utilization Distributions from Fitted Step-Selection Functions. *Ecosphere*, *8*(4), e01771. https://doi.org/10.1002/ecs2.1771

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*(2), 880–890. https://doi.org/10.1002/ece3.4823

Therneau, T. M. (2022, October 3). *Coxme: Mixed Effects Cox Models* (Version 2.2-18.1). https://cran.r-project.org/web/packages/coxme/index.html

Therneau, T. M., Lumley, T., Elizabeth, A., & Cynthia, C. (2023, August 14). *Survival: Survival Analysis* (Version 3.5-7). https://cran.r-project.org/web/packages/survival/index.html

Thurfjell, H., Ciuti, S., & Boyce, M. S. (2014). Applications of Step-Selection Functions in Ecology and Conservation. *Movement Ecology*, *2*(1), 4. https://doi.org/10.1186/2051-3933-2-4

Tomkiewicz, S. M., Fuller, M. R., Kie, J. G., & 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. https://doi.org/10.1098/rstb.2010.0090

Toth, C., & Jóźków, G. (2016). Remote Sensing Platforms and Sensors: A Survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*, 22–36. https://doi.org/10.1016/j.isprsjprs.2015.10.004

Vales, D. J., Nielson, R. M., & Middleton, M. P. (2022). Black-Tailed Deer Seasonal Habitat Selection: Accounting for Missing Global Positioning System Fixes. *The Journal of Wildlife Management*, *86*(8). https://doi.org/10.1002/jwmg.22305

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., Posit, & PBC. (2023, August 14). *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* (Version 3.4.3). https://cran.r-project.org/web/packages/ggplot2/index.html

Wiens, J. A. (1989). Spatial Scaling in Ecology. *Functional Ecology*, *3*(4), 385. https://doi.org/10.2307/2389612

Williams, H. J., Taylor, L. A., Benhamou, S., Bijleveld, A. I., Clay, T. A., Grissac, S., Demšar, U., English, H. M., Franconi, N., Gómez-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., & Börger, L. (2019). Optimizing the Use of Biologgers for Movement Ecology Research. *Journal of Animal Ecology.* https://doi.org/10.1111/1365-2656.13094

Wood, S. N. (2011). Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (Second Edition). CRC Press/Taylor & Francis Group.

Zeller, K. A., Vickers, T. W., Ernest, H. B., & Boyce, W. M. (2017). Multi-Level, Multi-Scale Resource Selection Functions and Resistance Surfaces for Conservation Planning: Pumas as a Case Study. *PLoS ONE*, *12*(6), e0179570.

879 Zeller, K. A., Wattles, D. W., Bauder, J. M., & DeStefano, S. (2020). Forecasting Seasonal
880   Habitat Connectivity in a Developing Landscape. *Land*, *9*(7), 233. https://doi.org/
881   10.3390/land9070233

Figure 1: (**a**) Demonstration of how missingness affects the number of valid steps that can be used for step-selection analyses under different levels of forgiveness. The upper panel depicts a trajectory with zero missing locations. That is, all aspired locations were successfully collected on a regular interval (yielding a regular step duration of $\Delta t = 1$). This trajectory produces four valid steps that can be included in the iSSF model and one invalid step that has to be omitted because it has no turning angle associated with it. In the central panel, animal location 4 was not obtained, introducing a missingness of 16.7%. If the modeler has a forgiveness of one, 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 exhibited a forgiveness of two, such as in the lower panel, a total of three steps could be obtained for further analysis. (**b**) Conceptual illustration of how increasing the forgiveness allows one to retain additional steps that can be used for step-selection analysis. The sequence of dots resembles the sequence of locations that were scheduled to be collected (e.g. using a GPS device), with the lines representing hypothetical steps. Because not all locations were successfully obtained (gray dots), there is missingness. Depending on the forgiveness level, already a single missing location enforces the introduction of a new burst, which leads to the loss of several steps. In addition, some of the remaining steps are invalid (dotted) because they are lacking a turning angle. By increasing the forgiveness, a modeler is willing to retain steps that exceed the regular step duration by a certain threshold, which enables them to obtain longer bursts and increase the number of steps that can be used for further analysis. In the figure, forgiveness increases from left to right.
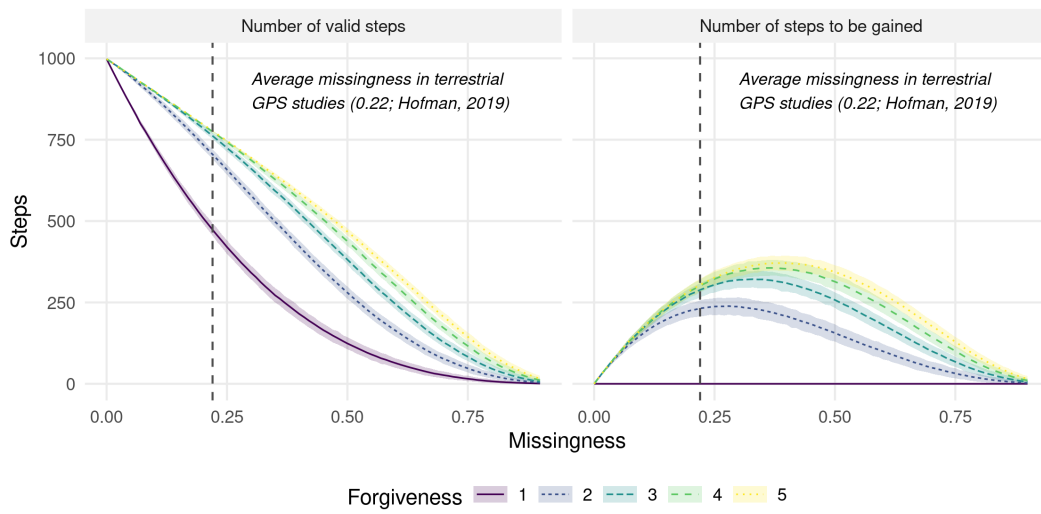
Figure 2: Illustration of how missingness in animal location data reduces the number of valid steps that can be used in step-selection analyses (left panel) and how increasing forgiveness helps to retain additional steps that are otherwise omitted (right panel). At a missingness of 0, 998 valid steps can be obtained from the total of 1,000 animal locations. At higher missingness, step durations become irregular, which means that the number of valid steps decreases substantially. However, if the modeler is willing to increase their forgiveness, additional steps can be gained. The right panel shows the number of valid steps that is gained when increasing the forgiveness from 1 to 2, 3, 4, and 5, respectively. Ribbons indicate the 95%-percentile intervals derived from 1,000 replicates.
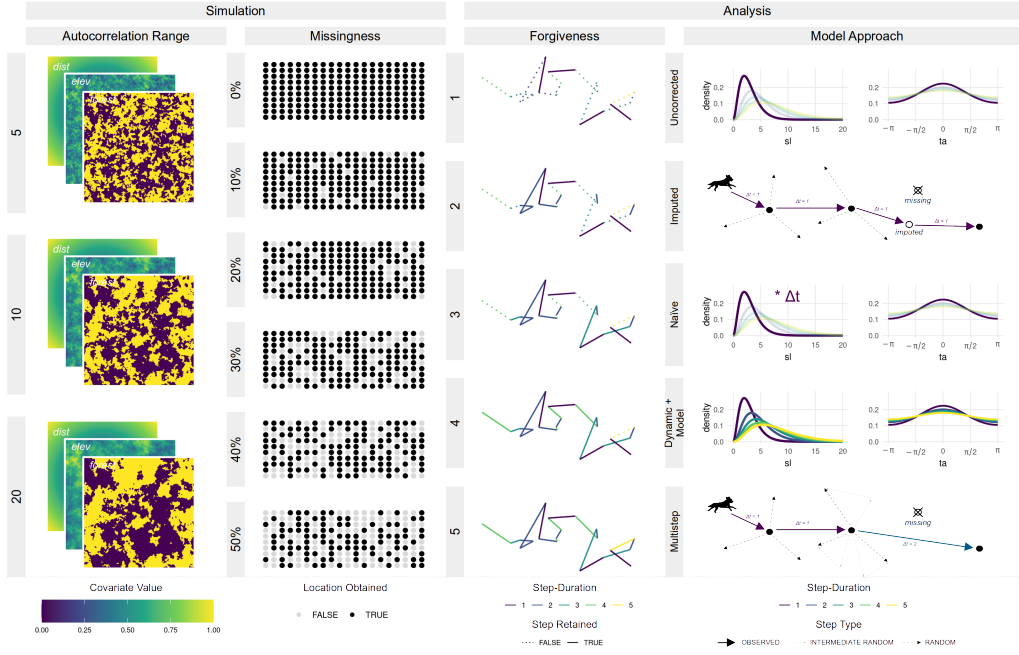
Figure 3: Illustration of the study design. We varied the autocorrelation range when simulating spatial covariates from 5 to 20 and tested for different missingness scenarios (ranging from 0% to 50% missing locations). To investigate how increasing forgiveness (i.e., the willingness to include steps with duration above the regular step duration) influenced model results, we varied its value from 1 (regular step selection) to 5 (considering steps that are five times the regular step duration). Finally, we tested five different methods to account for potential biases introduced by including irregular steps. This gave 3 x 6 x 5 x 5 = 450 combinations, each of which we replicated 100 times. We assumed step lengths (sl) to follow a gamma distribution, whereas turning angles (ta) followed a von Mises distribution.
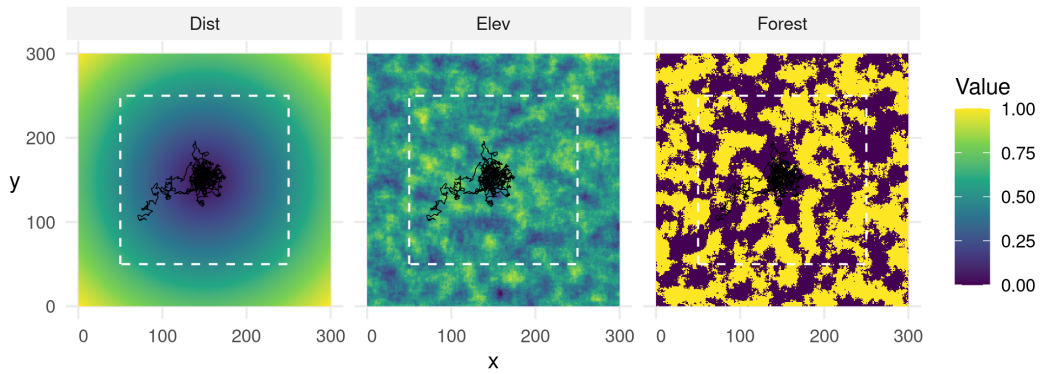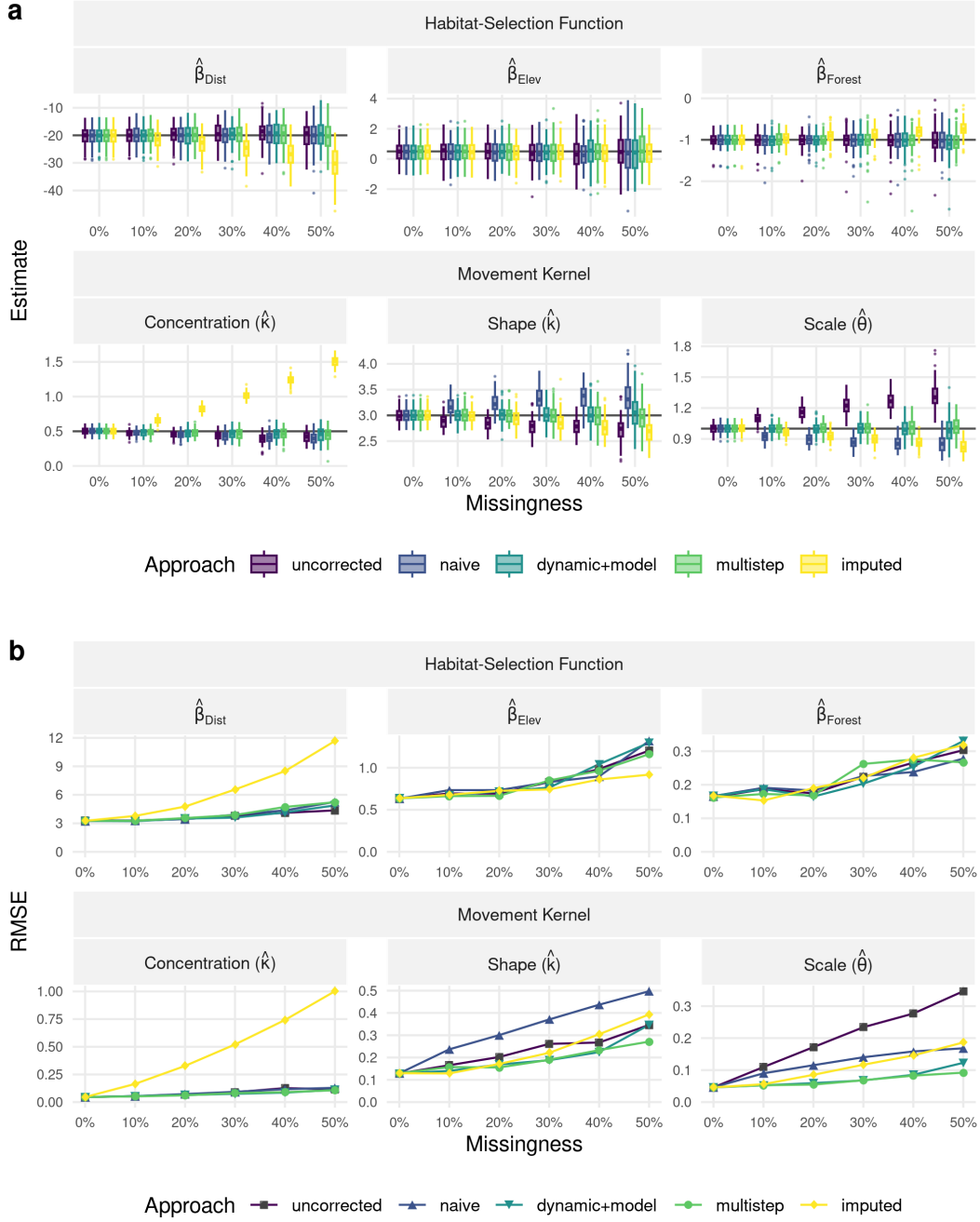
Figure 4: Example of a landscape configuration across which we simulated movement trajectories. All simulated layers had a resolution of 300 x 300 pixels. The distance layer indicated the distance to the center of the landscape and served to simulate attraction. The elevation and forest layers were simulated by sampling pixel-values from a normal distribution and applying a moving window to achieve spatial autocorrelation. Simulated individuals were initiated within the white dashed rectangle, which ensured that they would not be released directly at a map border. Simulated individuals were attracted to the landscape's center, preferred elevated areas, and avoided areas covered by forest. The black line shows the simulated trajectory associated with the visualized landscape configuration (cfr. Section 2.2).
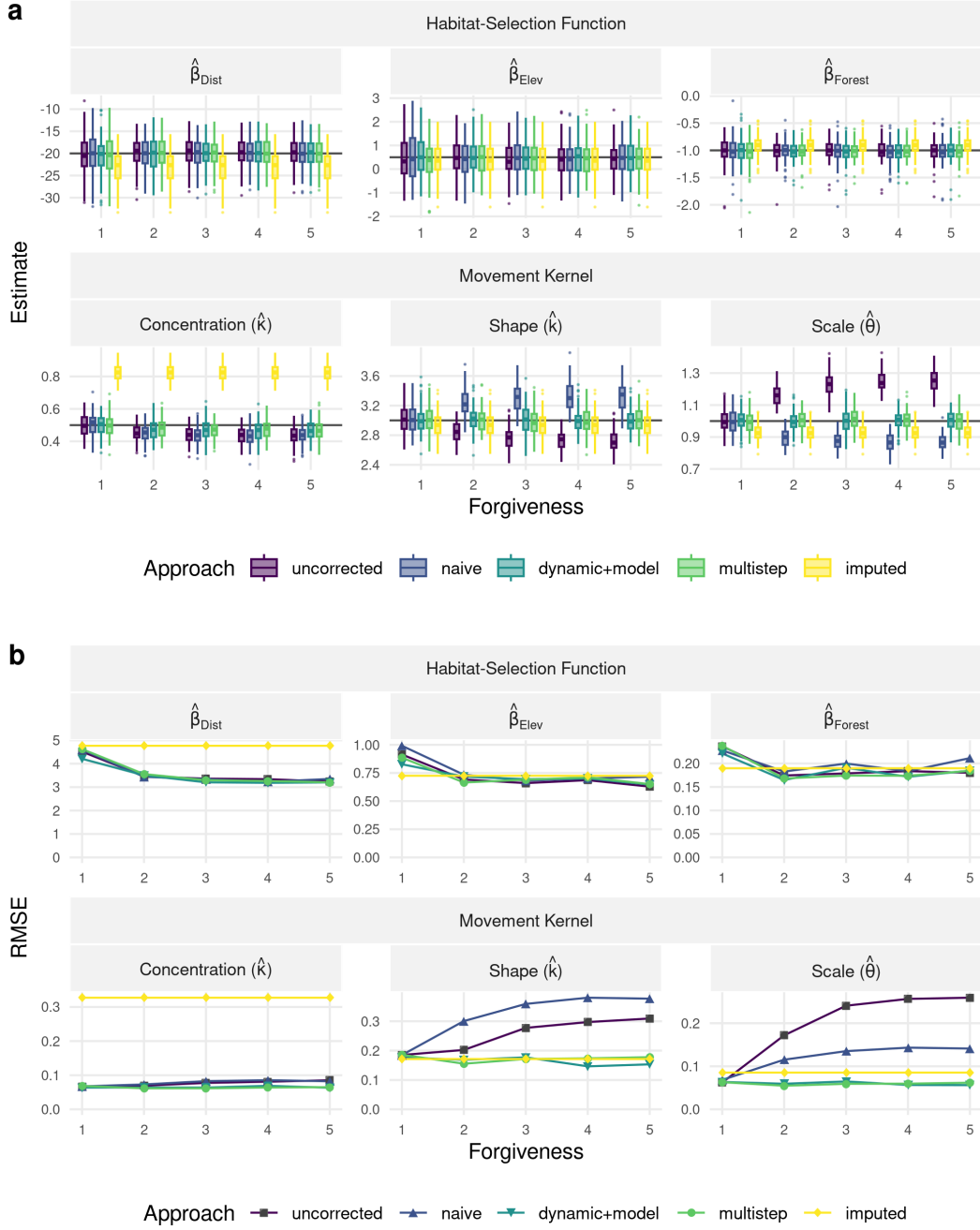
Figure 5: (**a**) Parameter estimates and (**b**) Root mean-square error (RMSE) with regard to the movement kernel and habitat-selection function as a function of forgiveness. Results are shown for the scenario with landscape autocorrelation of 20 and missingness of 20%. The movement kernel comprised of a gamma distribution with shape parameter k and scale parameter θ governing the step-length distribution and a von Mises distribution with concentration parameter κ governing the turning-angle distribution. Habitat-selection was based on three covariates, namely a Distance, Elevation, and a Forest layer. Estimates are shown for the five different approaches we tested for. The uncorrected approach ignored the fact that higher forgiveness implied temporal irregularity in the data, while all other approaches attempted to correct for the potential biases introduced by temporal irregularity. Note, the imputation approach is not affected by the forgiveness level, since it always starts with a full trajectory.

Figure 6: (**a**) Parameter estimates and (**b**) Root mean-square error (RMSE) with regard to the movement kernel and habitat-selection function as a function of missingness. Results are shown for the scenario with landscape autocorrelation of 20 and forgiveness of 2. The movement kernel comprised of a gamma distribution with shape parameter k and scale parameter θ governing the step-length distribution and a von Mises distribution with concentration parameter κ governing the turning-angle distribution. Habitat-selection was based on three covariates, namely a Distance, Elevation, and a Forest layer. Estimates are shown for the five different approaches we tested for. The uncorrected approach ignored the fact that higher forgiveness implied temporal irregularity in the data, while all other approaches attempted to correct for the potential biases introduced by temporal irregularity. Note, the imputation approach is not affected by the forgiveness level, since it always starts with a full trajectory.
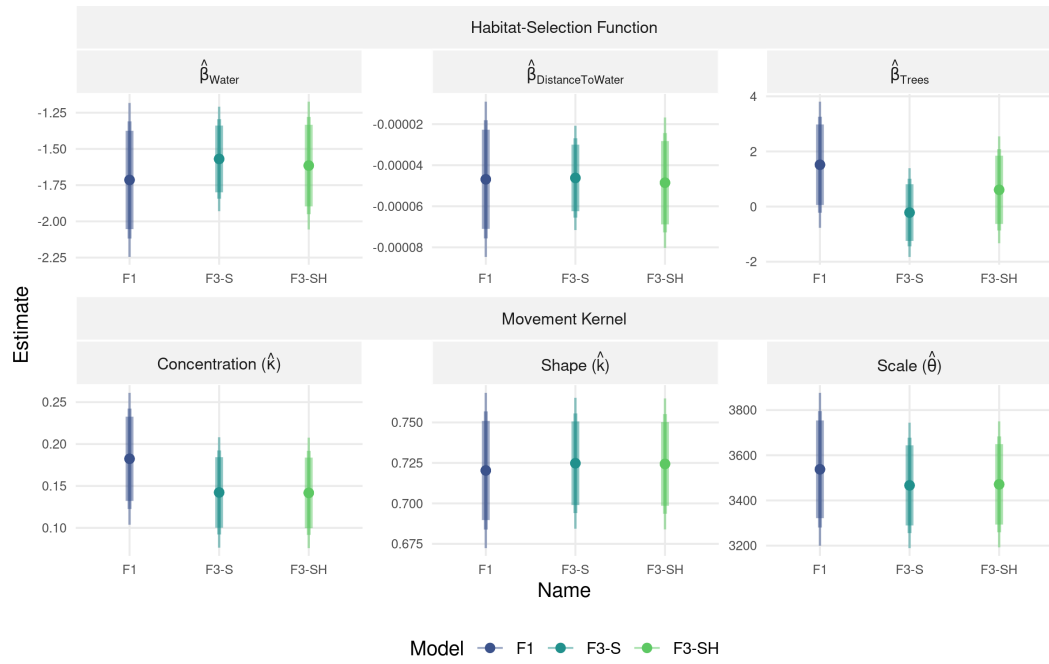
Figure 7: Model results from the case study using GPS data collected on Apollo. In F1, forgiveness was set to one (only 2-hour steps were considered), whereas in F3-S and F3-SH a forgiveness of three was employed (allowing for step durations of up to 6 hours). In model F3-S, the step duration was interacted with step descriptors. In model F3-SH, step duration was interacted with step descriptors and habitat covariates. The bars indicate the 90%, 95%, and 99% confidence intervals. Note that for simplicity, we omitted interactions with the step duration from this figure.

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 (SSFs).

Term	Definition
Animal locations	A series of telemetry data points that include date, time, longitude, and latitude information, describing when and where an animal was observed or recorded.
Step	A straight line connecting two consecutive locations.
Observed step	A step that connects two observed animal locations.
Random step	A step that connects an observed animal location with a random location. Random locations are typically generated by combining an observed animal location with random step lengths and turning angles.
Step length	The Euclidean distance of a step.
Turning angle	A measure of the change in direction between two consecutive steps.
Habitat-selection function	A probabilistic description of an animal's habitat preferences. It describes how an animal selects habitat when not constrained by its movement capacity. Also referred to as movement-free habitat-selection function.
Movement kernel	A probabilistic description of an animal's movement capacity. It describes how an animal would move when not constrained by habitat selection. Also referred to as selection-free movement kernel.
Trajectory	A sequence of animal locations collected on the same individual.
Step duration	The time interval associated with a particular step, i.e., the time elapsed between two consecutive animal locations.
Regular animal locations	A series of animal locations that have been obtained at regularly spaced time intervals, such as every hour.
Irregular animal locations	A series of animal locations collected at irregular time intervals.
Regular step durations	Step durations that occur when animal locations are successfully collected at regular time intervals.
Irregular step durations	Step durations that occur when animal locations are not successfully collected at regular time intervals.
Missingness	The fraction of animal locations that should have been collected but, for some reason, were not. For example, if only eight out of ten expected animal locations were successfully collected, the missingness would be 0.2 (i.e., 20%).
Forgiveness	The maximum step duration, measured in multiples of the regular step duration, a modeler is willing to include in the step-selection analysis. A modeler with a forgiveness of one, for instance, only considers regular steps, while a modeler with a forgiveness of two would consider irregular steps up to twice the regular step duration.
Burst	A sequence of consecutive animal locations equally spaced in time and with steps were the step duration does not exceed the forgiveness.
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.