ORIGINAL RESEARCH

WILEY Ecology and Evolution · Open Access

Animal movement tools (`amt`): R package for managing tracking data and conducting habitat selection analyses

Johannes Signer[1] (iD) | **John Fieberg**[2] (iD) | **Tal Avgar**[3] (iD)

[1]Wildlife Sciences, University of Goettingen, Göttingen, Germany

[2]Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota, St. Paul, Minnesota

[3]Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada

Correspondence
Johannes Signer, Wildlife Sciences, University of Goettingen, Göttingen, Germany.
Email: jsigner@gwdg.de

Abstract

Advances in tracking technology have led to an exponential increase in animal location data, greatly enhancing our ability to address interesting questions in movement ecology, but also presenting new challenges related to data management and analysis. Step-selection functions (SSFs) are commonly used to link environmental covariates to animal location data collected at fine temporal resolution. SSFs are estimated by comparing observed steps connecting successive animal locations to random steps, using a likelihood equivalent of a Cox proportional hazards model. By using common statistical distributions to model step length and turn angle distributions, and including habitat- and movement-related covariates (functions of distances between points, angular deviations), it is possible to make inference regarding habitat selection and movement processes or to control one process while investigating the other. The fitted model can also be used to estimate utilization distributions and mechanistic home ranges. Here, we present the R package `amt` (animal movement tools) that allows users to fit SSFs to data and to simulate space use of animals from fitted models. The `amt` package also provides tools for managing telemetry data. Using fisher (*Pekania pennanti*) data as a case study, we illustrate a four-step approach to the analysis of animal movement data, consisting of data management, exploratory data analysis, fitting of models, and simulating from fitted models.

KEYWORDS
habitat selection, home range, movement ecology, resource-selection function, step-selection function, telemetry

1 | INTRODUCTION

Advances in technology have led to large collections of fine-scale animal biotelemetry data (Cagnacci, Boitani, Powell, & Boyce, 2010; Kays, Crofoot, Jetz, & Wikelski, 2015), fueling the development of new quantitative methods for studying animal movement (Hooten, Johnson, McClintock, & Morales, 2017). Nathan et al. (2008)

introduced the movement ecology paradigm that conceptually connects different factors shaping the realized movement path of animals (e.g., the internal state of an animal, interaction with intra- and conspecifics, and varying environmental conditions). The movement ecology paradigm can serve as a framework for generating new hypotheses about animal movements. To test these hypotheses, efficient and straightforward tools for the management and analyses

of movement data are required. Although a large number of R packages have been developed for analyzing animal movement data (e.g., Calabrese, Fleming, & Gurarie, 2016; Gurarie, Andrews, & Laidre, 2009; Michelot, Langrock, & Patterson, 2016), these packages often utilize domain-specific data formats and focus on a narrow subset of analytical methods (e.g., methods for fitting discrete or continuous time movement models or trajectory segmentation). We had two primary objectives in developing the `amt` R package, namely to provide: (a) a set of functions for exploratory analyses of movement data in R, and (b) functions for inferring habitat selection from movement data in R. For the second objective, we focus on the analysis of fine-scale animal location data using step-selection functions (SSFs), but `amt` also provides methods for other commonly used analytical approaches (e.g., home-range analysis, resource-selection functions). SSFs are powerful tools for modeling animal movement and habitat selection, but are not currently available in open-source software packages, despite their popularity.

Methods that quantify habitat selection by linking environmental covariates to location data of animals have been around for a long time. Traditionally, resource-selection functions (RSF; Boyce & McDonald, 1999; Manly, McDonald, Thomas, McDonald, & Erickson, 2007) were used to study habitat selection of animals. RSFs compare covariates associated with locations where the animal was observed with covariates associated with random locations within the availability domain, a spatial domain within which any location is assumed available for the animal to use at any given time. Despite the sensitivity of the resulting inference to habitat availability (Beyer et al., 2010), no consensus exists as to the most suitable approach to delineate the spatial domain of availability (Northrup, Hooten, Anderson, & Wittemyer, 2013; Paton & Matthiopoulos, 2016; Prokopenko, Boyce, & Avgar, 2017b). Moreover, the assumption that the availability domain can be considered temporally static might have been justifiable for very coarse sampling rates (e.g., daily or weekly positions of the animal), but is challenging for modern GPS data with sampling rates <1 hr. SSFs (Fortin et al., 2005; Thurfjell, Ciuti, & Boyce, 2014) resolve these issues by pairing each observed location with a set of random locations deemed accessible from the previously observed location. SSFs estimate conditional selection coefficients using a likelihood equivalent of a Cox proportional hazards model (Gail, Lubin, & Rubinstein, 1981).

Until recently, SSFs were fitted by sampling random points based on the empirical (observed) distribution of steps (straight lines connecting consecutive locations). This approach has come under some scrutiny as it implicitly assumes habitat selection is conditional on animal movement but not vice versa, potentially leading to biased inference (Forester, Im, & Rathouz, 2009). A recent extension, termed integrated SSF (iSSF), alleviates this concern and allows for simultaneous inference of habitat selection and movement processes (Avgar, Potts, Lewis, & Boyce, 2016). This is accomplished by requiring that random steps are sampled under one of several analytical distributions, and also by including, in addition to habitat-related covariates, movement-related covariates (functions of distances between points, angular

deviations) resulting in likelihood-based estimates of the shape and scale of the underlying analytical distributions (Avgar et al., 2016; Duchesne, Fortin, & Rivest, 2015; Forester et al., 2009). Unlike SSFs (that do not include an explicit movement component), a fitted iSSF is a fully fledged biased random walk model that can be used to simulate animal space-use (Avgar et al., 2016; Duchesne et al., 2015; Signer, Fieberg, & Avgar, 2017). Integrated SSFs are attractive because one can make inferences regarding both habitat selection and movement processes, or to control for one process while investigating the other. Hence, iSSFs offer a flexible framework for testing complex hypotheses about animal space-use behavior (e.g., Prokopenko, Boyce, & Avgar, 2017a; Scrafford, Avgar, Heeres, & Boyce, 2018; Viana et al., 2018), while also providing a mechanistic predictive tool for projecting anticipated space-use patterns across novel or modified landscapes. Integrated SSFs can be seen as discrete approximations to spatio-temporal point process models (Johnson, Thomas, Ver Hoef, & Christ 2008; Johnson, Hooten, & Kuhn 2013; Brost, Hooten, Hanks, & Small 2015). Similar methods exist for continuous time movement models (Hanks et al., 2015; Hooten, Johnson, Hanks, & Lowry, 2010), but these methods typically require a discretization of space (rather than of time).

Step-selection functions (SSFs and iSSFs) are usually straightforward to fit (using any conditional logistic regression routine) once data are appropriately structured, but data preparation itself tends to be more complex and confusing and may thus become a limiting step in the application of this approach. Here, we describe the `amt` package for R, which provides a flexible and coherent workflow for efficient analysis of animal tracking data. We make heavy use of piped workflows and list columns as introduced to R through the `tidyverse` package-family (Wickham, 2017). We illustrate a typical workflow for fitting a (i)SSF using fisher (*Pekania pennanti*) data from LaPoint, Gallery, Wikelski, and Kays (2013a). Detailed vignettes, help files, sample data and analyses are available in the `amt` package available on CRAN (https://cran.r-project.org/package=amt).

2 | FUNCTIONALITY

A typical workflow to analyze animal tracking data can be divided into four main steps (described in detail below):

1. Data preparation, inspection, and management: Load and inspect gaps in the data, resample tracks if needed, and adjust coordinate reference systems.
2. Exploratory data analysis and descriptive analyses: Explore patterns in the data graphically, consider multiple movement characteristics (e.g., step-length distribution, net square displacement, or home-range size) across several animals and/or time periods.
3. Modeling: Fit models to answer questions or test hypothesis related to movement and space use of animals.
4. Simulation: Use fitted models to simulate derived quantities (e.g., space use) and assess model fit.

2.1 | Data preparation, inspection, and management

After loading data into R, users should perform a variety of data quality checks and possibly remove fixes with missing coordinates (although this information could potentially be used to test if fixes are missing at random). We provide functions to quantify variability in sampling rates over time and among individuals, inspect the data visually for obvious outliers (e.g., determined by screening for unreasonable speeds), remove periods at the beginning and the end of the track to exclude possible capture effects, and resample the data to form regular bursts (i.e., partition the track into groups of observations with regular sampling rates, within some specified level of tolerance). Regular sampling rates are required for SSFs, because selection is not scale invariant (Barnett & Moorcroft, 2008; Signer et al., 2017), and thus, sampling rates should be similar for different animals in a given study. One option to overcome the problem of missing fixes would be to use data imputation before fitting (i)SSFs (McClintock, 2017), as has been suggested and implemented for hidden Markov models (McClintock & Michelot, 2018).

2.2 | Exploratory data analysis and descriptive analyses

Once data have been cleaned, the next logical step is to explore the data by looking at different movement-related statistics (e.g., distributions of turning angles or step lengths) and trajectory and space-use summaries (e.g., net squared displacement, path sinuosity, home-range area). These summaries may be calculated for the whole trajectory or on a subset of points (a track might be split by time of the day, season, year, or any other biologically meaningful factor).

2.3 | Modeling

In the next step, we fit models to test hypotheses about animal movement and habitat selection. Importantly, `amt` provides functionality necessary for data development steps prior to fitting RSFs and (i)SSFs (e.g., methods for generating random points or steps, and extract environmental covariates for the observed and random steps). For many other analyses (e.g., behavioral change point analyses, fitting continuous time movement models or identification of hidden behavioral states with hidden Markov models), `amt` provides coercion functions to translate location data into objects of classes required by the respective packages.

2.4 | Simulation

As a final step, new data can be simulated from fitted models. Simulations can be used to obtain estimates of space use (i.e., the utilization distribution), identify corridors of high use, asses the power of the model (testing how well parameters can be recovered as a function of sample size), or perform model validation (Fieberg et al., 2018). Many packages that fit models also provide methods to simulate from fitted models (e.g., `ctmm` or `moveHMM`). `amt` provides means to simulate space use from fitted iSSFs.

3 | CASE STUDY

We illustrate a subset of the above steps using data from radio collared fishers available through movebank (LaPoint, Gallery, Wikelski, & Kays, 2013b; LaPoint et al., 2013a). For details about the data and the capture of the animals, we refer to Brown et al. (2012) and LaPoint et al. (2013a). We begin by analyzing the space use of Ricky T (id 1016), and then illustrate how similar analyses can be extended to several animals for population-level inference (Fieberg, Matthiopoulos, Hebblewhite, Boyce, & Frair, 2010; Hooten, Buderman, Brost, Hanks, & Ivan, 2016).

3.1 | From data cleaning to simulated space use for one animal

We begin with loading the data from all fishers, remove observations with missing spatial coordinates (longitude, latitude), and subset relocations for Ricky T (id: 1016).

```
library(raster)
library(lubridate)
library(amt)
dat <- read_csv("data/Martes pennanti LaPoint New York.csv") %>%
  filter(!is.na(`location-lat`)) %>%
  select(x = `location-long`, y = `location-lat`,
         t = `timestamp`, id = `tag-local-identifier`) %>%
  filter(id %in% c(1465, 1466, 1072, 1078, 1016, 1469))
dat_1 <- dat %>% filter(id == 1016)
```

The function `amt::make_track` creates a track (the basic building block of the `amt` package), given the names of the columns containing _x_ and _y_ coordinates, time (_t_), and we can set a coordinate reference system (CRS). The original data were provided in geographical coordinates (EPSG code: 4326). Here, we shall transform this original CRS (using function `amt::transform_coords`) to the projected North American Datum (NAD83, EPSG code: 5070).

```
dat_1 <- amt::make_track(dat_1, .x = x, .y = y, .t = t,
crs = sp::CRS("+init=epsg:4326")) %>%
  amt::transform_coords(sp::CRS("+init=epsg:5070"))
```

We then summarize the distribution of time intervals between successive locations to get a general impression for the sampling rate.

```
amt::summarize_sampling_rate(dat_1)

## # A tibble: 1 x 9
##   min           q1                median mean    q3      max     sd      n     unit
##   <S3: table>   <S3: table>       <S3: > <S3: >  <S3:>   <S3:>   <dbl>   <int> <chr>
## 1 0.1           1.93333333333333  2.033... 8.041... 2.56... 1208... 44.0  8957  min
```

We see that we have 8,957 total locations, the shortest interval between locations is 0.1 min and the largest time interval between locations is 1,208 min, with median interval length equal to roughly 2 min. Despite the 2 min temporal resolution, we choose to resample the track to 10 min with a tolerance of 1 min (`amt::track_resample`), in order to conduct the analyses on the same temporal scale as the next example (some individuals had a median sampling rate of 10 min).

The function `minutes` from the package `lubridate` (Grolemund & Wickham, 2011), is used here to create an object of class `Period` that is then passed to `amt::track_resample`. `Period`s can be specified using all common time units; thus, it is straightforward to specify a sampling rate and an acceptable tolerance. We will also choose to keep only those bursts (subsets of the track with constant sampling rate, within the specified tolerance) with at least three relocations, the minimum required to calculate a turn angle (`amt::filter_min_n_burst`). The following code implements those choices and translates from a point representation to a step (step length, turn angle) representation of the data. In the final line of the code snippet, we use the function `amt::time_of_day` (a wrapper around `maptools::sunriset` and `maptools::crepuscule`; Bivand & Lewin-Koh, 2017) to calculate if a location was taken during the day or night. If the argument `include.crepuscule` is set to `TRUE`, the function not only considers day and night, but also dawn and dusk.

```
stps <- amt::track_resample(dat_1, rate = minutes(10),
tolerance = seconds(60)) %>%
  amt::filter_min_n_burst(min_n = 3) %>%
  amt::steps_by_burst() %>%
  amt::time_of_day(include.crepuscule = FALSE)
```

We then use the `str` function to inspect the structure of `stps`.

```
str(stps)
Classes 'steps', 'tbl_df', 'tbl' and 'data.frame': 1501 obs. of  11 variables:
 $ burst_  : num  16 16 16 16 18 18 21 21 23 23 ...
 $ x1_     : num  1780410 1780469 1780468 1780480 1780471 ...
 $ x2_     : num  1780469 1780468 1780480 1780469 1780469 ...
 $ y1_     : num  2412415 2412224 2412230 2412213 2412223 ...
 $ y2_     : num  2412224 2412230 2412213 2412217 2412217 ...
 $ sl_     : num  199.47 6.01 20.8 11.15 6.85 ...
 $ ta_     : num  NA 2.96 -2.7 -2.56 NA ...
 $ t1_     : POSIXct, format: "2010-02-10 03:00:36" "2010-02-10 03:10:40" ...
 $ t2_     : POSIXct, format: "2010-02-10 03:10:40" "2010-02-10 03:20:10" ...
 $ dt_     :Class 'difftime'  atomic [1:1501] 10.1 9.5 10.2 10.4 9.5 ...
 $ tod_end_: chr  "night" "night" "night" "night" ...
```

`stps` is a regular `data_frame` with 11 attributes of steps (e.g., start, end, and step length; columns) and 1501 steps (rows). For each step, the start ($x1\_$, $y1\_$) and end ($x2\_$, $y2\_$) coordinates, as well as the start and end time ($t1\_$, $t2\_$) are given. In addition, the following derived quantities are calculated: step length ($sl\_$; in CRS units), turning angles ($ta\_$; in degrees; notice that it cannot be calculated for steps that are not preceded by a valid step), the time difference ($dt\_$), and the burst ($burst\_$) to which the step belongs. We proceed by preparing the environmental data. We hypothesized that Ricky T prefers forested wetlands over other landuse classes. We used the National Landcover Database (which is freely available at https://www.mrlc.gov/nlcd11_data.php). We first load the landuse raster and create a layer called wet that is 1 for forested wetlands (category 90) and 0 otherwise (using the `raster` package; Hijmans, 2017).

```
land_use <- raster("data/landuse_study_area.tif")
wet <- land_use == 90
```

For convenience and readability, we give the layer a meaningful name.

```
names(wet) <- "wet"
```

Before proceeding to modeling space use and habitat selection of Ricky T, we perform some exploratory data analysis based on step length and turning angles in different habitat types (forested wetlands versus other areas) and time of the day (day and night). We will have to extract the covariate value at the start point of each step (using the function `amt::extract_covariates`) and plot the density of step lengths per habitat class and time of day (Figure 1; for the full code to replicate Figure 1 see Supporting information Data S1). Note that the function `amt::extract_covariates` takes an argument `where` that indicates whether covariate values should be extracted at the beginning or the end of a step ("both" can be used to extract the covariate at the start and the end of a step). Depending on the target process under investigation (habitat selection or movement), covariates might be extracted at the end of the step (habitat selection process) or at the start of the step (movement process). If covariates are extracted at the end of the step, they are typically included in the model as main effects, to answer questions of the type: How do covariates influence where the animal moves? In contrary, if covariates are extracted at the beginning of the step, they are typically included in the model as an interaction with movement characteristics (step length, log of the step length, or the cosine of the turn angle), to test hypotheses of the type: Do animals move faster/more directed, if they start in a given habitat? Finally, covariate values at the start and the end of a step can also be included in the model as an interaction with each other, to test hypotheses of the type: Are animals more likely to stay in a given habitat, if they are already in that habitat?

To fit SSFs, the observed covariates associated with observed steps are compared to covariates associated with random (or control) steps. Random steps can be generated by either (a) sampling from the observed turn step-length and turn angle distribution (resulting in a traditional SSF), or (b) by fitting a parametric distribution to the observed step lengths (either a negative-exponential, a half-normal, a log-normal, or a gamma; see Avgar et al., 2016, Appendix 2) and turn angles (a von Mises; Duchesne et al., 2015). As mentioned above, an iSSF is arguably less biased and also provides the user with a mechanistic movement model that can be used to simulate space use, and hence utilization distributions (Avgar et al., 2016; Signer et al., 2017). Currently, `amt` only implements the iSSFs with gamma and von Mises distributions.

Thus, we proceed by fitting a gamma distribution to the step lengths and a von Mises distribution to the turn angles using maximum likelihood (Agostinelli & Lund, 2017; Delignette-Muller & Dutang, 2015), and use these distributions to generate and pair nine random steps with each observed step. The number of random steps effects the estimation error; the more the steps, the lower the error, but the higher the computational burden (Avgar et al., 2016). We then extract the covariates at the end point of each step (observed and random) using the function `amt::extract_covariates`, and fit a conditional logistic regression model to the resulting data including movement-related covariates with the function `amt::fit_issf` (a wrapper to `survival::clogit`; Therneau & Grambsch, 2000; see Supporting information Data S2 for details of the model).

We included two main effects in the model, the environmental covariate wet, and the log of the step length ($log\_sl\_$) as a modifier
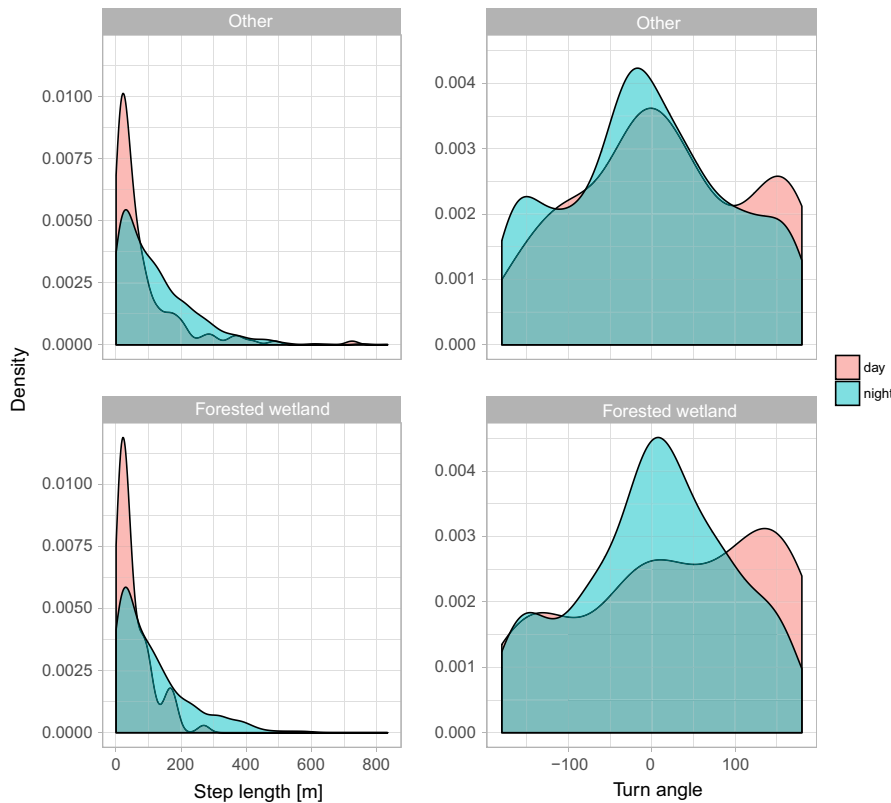
FIGURE 1 Exploratory data analysis of one individual fisher, Ricky T (id: 1016): empirical distributions of step lengths (first column) and turning angles (second column) are shown for forested wetland (second row) and other habitats (first row) and for day and night (colors)

of the shape parameter of the underlying gamma distribution. The estimated coefficient of `log_sl_` can be used to adjust the tentative shape estimate (i.e., the estimate of the shape parameter using the observed step lengths) of the underlaying gamma distribution for the step lengths. We also include interactions between `wet` and `tod_`, a factor with two levels—day (the reference category) and night, and between `tod_` and `log_sl_`. These interactions are included to the test the hypotheses that habitat selection and displacement rate, respectively, differ between day and night. The target variable `case_` is one for observed steps and zero for random (or control) steps. Each step is paired with several (here 9) control steps that form together a stratum (indicated by `strat(step_id_)` in the model formula). The function `amt::random_steps` automatically creates a new column, `step_id_`, that identifies different strata.

```
m1 <-stps %>% amt::random_steps(n = 9) %>%
  amt::extract_covariates(wet) %>%
  amt::time_of_day(include.crepuscule = FALSE) %>%
  mutate(log_sl_ = log(sl_)) %>%
  amt::fit_issf(case_ ~ wet + log_sl_ + wet:tod_end_
    + log_sl_:tod_end_ + strata(step_id_))
```

We could have also included cosines of the turning angles and their interaction with day. This choice would modify the concentration parameter of the underlying von Mises distribution for the turning angles and allow the degree of directional persistence to depend on time of day; the data summarized in Figure 1 suggest that this could be a sensible choice. For the sake of simplicity, however, we have assumed we have correctly modeled the degree of directional persistence and that it does not differ between day and night. It would be straightforward

to perform model selection (e.g., using Akaike Information Criterion (AIC); Akaike, 1974) to identify the best model structure.

Inspecting the fitted model (Table 1), we make the following observations. (a) There is evidence to suggest that the animal prefers forested wetlands over other landuse classes, (b) there is no difference in Ricky's preference for wetlands between day and night, (c) there is evidence to modify the shape of the gamma distribution fit to the observed step lengths (through the log of the step length), and (d) the modification of the shape parameter should be done separately for day and night, indicating that expected movement speeds differ between day and night.

Besides inspecting the coefficients and their standard errors, we can calculate derived quantities, such as the expected speed. Because we included an interaction between parameters of the step-length distribution and time of the day, we have to account for this interaction when calculating the expected speed for day and night. We begin by retrieving the tentative parameter estimates (i.e., the estimated parameters before correcting for habitat selection; see Avgar et al. (2016) for more details) for the gamma distribution of the step-length distribution:

```
shape <- amt::sl_shape(m1)
scale <- amt::sl_scale(m1)
```

We then adjust the shape parameter separately for day and night with the estimates of the corresponding coefficients from the fitted model (Avgar et al., 2016).

```
shape_adj_day <- amt::adjust_shape(shape, coef(m1)["log_sl_"])
shape_adj_night <- amt::adjust_shape(shape, coef(m1)["log_sl_"]) +
    coef(m1)["log_sl_:tod_end_night"]
```

The underlying gamma distributions for the step lengths vary by time of day (Table 1). The expected speed for day and night is thus given by the product of the tentative scale parameter (no adjustment is needed here, because we did not include step length in the model and hence assumed that the scale of the gamma distribution describing the step lengths remains unchanged) and the adjusted shape parameter. To obtain 95% confidence intervals for the mean speed, we bootstrapped the model `m1` 1,000 times by resampling (with replacement) the strata (for full code see Supporting information Data S1). Results suggest that Ricky T moves significantly faster during nights (11.0 m/min, 95% CI = 10.7, 11.4 m/min) than during days (8.57 m/min, 95% CI = 7.8, 9.32 m/min).

In a final step, we simulated space-use from the fitted model `m1` to obtain a model-based estimate of the animal's utilization distribution (UD; Avgar et al., 2016; Signer et al., 2017). Generally, two types of UDs can be simulated: the transient UD and the steady-state UD. The transient UD describes the expected space-use distribution of the animal within a short time period and is hence conditional on the starting position. The steady-state UD describes the expected space-use distribution of the animal in the long-term. In order to simulate UDs, one has to ensure that the animals stay within the study domain. We see three possible methods for achieving this goal–all implemented in `amt`: (a) use a covariate that attracts the animal toward one or more centers of activity (e.g., the squared distance to the mean of all coordinates), (b) use a very large landscape, or (c) use a wrapped landscape (torus). Here, we illustrate the simulation of steady-state and transient UDs. For the steady-state UD, we simulate from the first observed location 10^7 time steps on a toroid landscape, once for day and once for night. For the transient UD, we are interested in the UD up to 10 hr after last observation, we therefore simulated 72 steps (at a 10 min sampling rate) 5×10^3 times.

We describe the simulation for the steady-state and transient UD for daytime. First, we create a movement kernel (Figure 2a) that is used to determine the animal's movement ability at each time step. Note, we use the tentative scale estimate and the shape estimate adjusted for day.

```
mk <- amt::movement_kernel(scale = scale, shape = shape_adj_day,
template = wet)
```

Next, we create a habitat kernel (that is for each pixel we calculate the estimated selection coefficients times the resources and exponentiate the product; Figure 2b).

```
hk <- amt::habitat_kernel(coef = list(forest = coef(m1)["wet"]),
resources = wet, exp = TRUE)
```

We then estimate the steady-state UD (Figure 2c and e) with the function `simulate_ud`:

```
ssud_day <- amt::simulate_ud(movement_kernel = mk,
    habitat_kernel = hk,
    start = as.numeric(stps[1, c("x1_", "y1_")]),
    n = 1e7)
```

In order to simulate the transient UD (Figure 2c and e), we have to repeatedly simulate short tracks starting at the same point, and then sum individual UDs and normalize, which we do with the function `amt::simulate_tud`.

```
tud_day <- amt::simulate_tud(mk, hk, as.numeric
(stps[1501, c("x1_", "y1_")]), n = 72, n_rep = 5e3)
```

All simulations took <1 min on a standard laptop.

3.2 | Many animals: quantifying population-level effects

We start again with the same data set (`dat`), containing data from six individual fishers. This time we are interested in quantifying among-animal variability in the selection coefficients. We proceed using nearly all the same steps as in the first example, but with a different data structure: `data_frames` with list columns (Müller & Wickham, 2018). List columns are best thought of as regular columns of a `data_frame` that are R lists and can contain any objects (in our case tracks and fitted models). The `purrr::nest` command can be used to nest data into a list column (Henry & Wickham, 2017).

```
dat_all <- dat %>% nest(-id)
```

`dat_all` is now a `data_frame` with 6 rows (one for each individual) and two columns. In the first column the animal id is given, and in the second column (by default named data) the relocations of the corresponding animal are saved. We start by assigning the sex of each animal.

```
dat_all$sex <- c("f", "f", "f", "m", "m", "m")
```

We can now apply the steps as before for all animals. We first create a track for each animal and transform the coordinate reference system using the function `amt::transform_coords`.

```
dat_all <- dat_all %>%
  mutate(trk = lapply(data, function(d) {
  amt::make_track(d, x, y, t,
  crs = sp::CRS("+init=epsg:4326")) %>%
  amt::transform_coords(sp::CRS("+init=epsg:5070"))
}))
```

TABLE 1 Coefficients of fitted integrated step-selection function

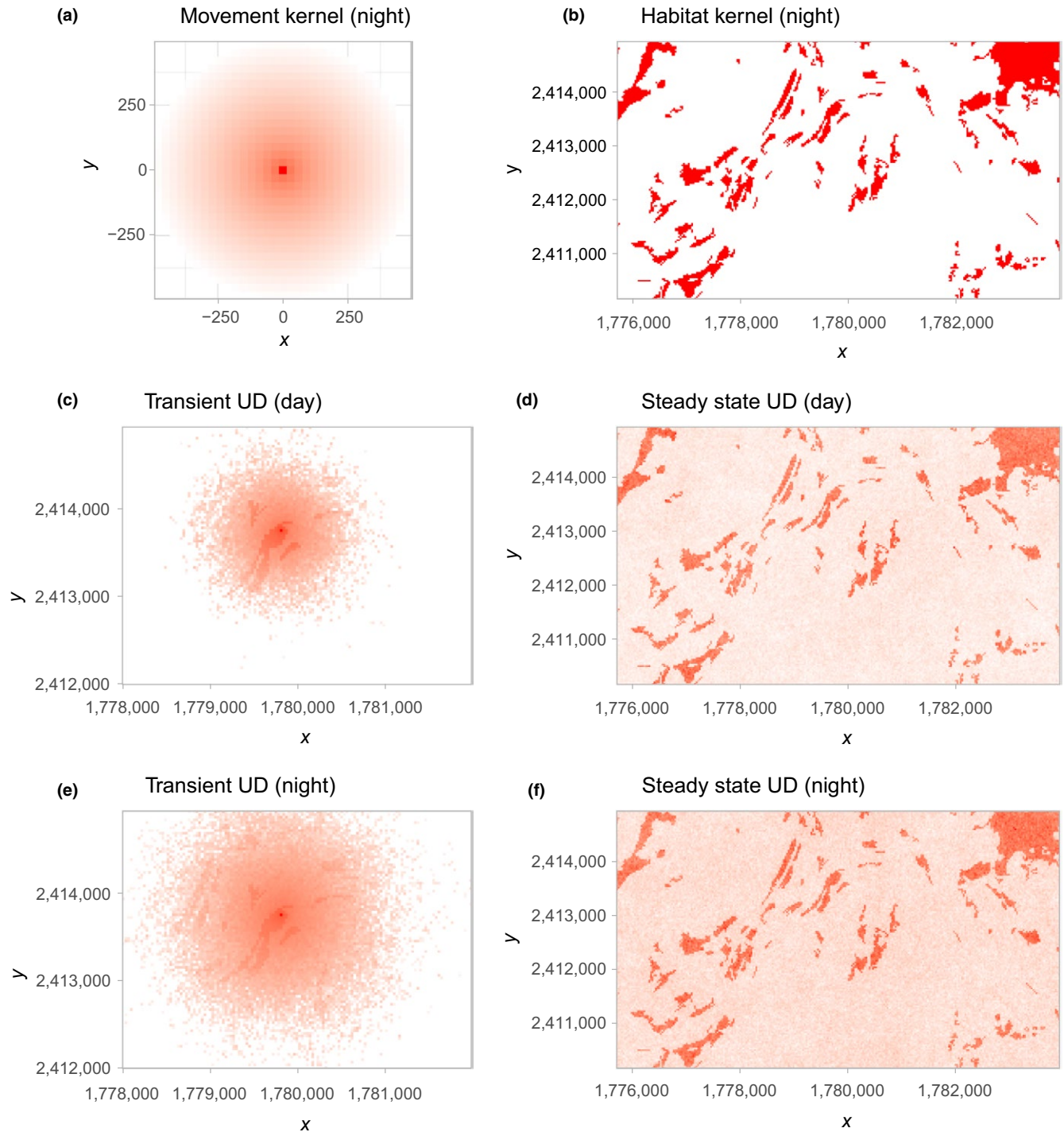|	coef	exp(coef)	SE(coef)	z	Pr(>|z|)
wet	0.9765	2.6552	0.2672	3.6551	0.0003
log_sl_	−0.2775	0.7577	0.0600	−4.6259	0.0000
wet:tod_end_night	−0.3656	0.6938	0.2831	−1.2914	0.1966
log_sl_:tod_end_night	0.3529	1.4231	0.0655	5.3839	0.0000

FIGURE 2 Simulated utilization distributions. To obtain simulated Utilization Distributions (UD), a movement kernel (panel a) and a habitat kernel (panel b) are needed. The movement kernel is always placed at the current position of the animal. The next step of the animal is then sampled with probability proportional to the product of two kernels. Expected differences in movement speeds between night and day are reflected in the transient UD (panels c and e) and to a lesser extend in steady-state UD (panels d and f). Note, for better visualization, fills were log10 transformed for panels a, c, and e
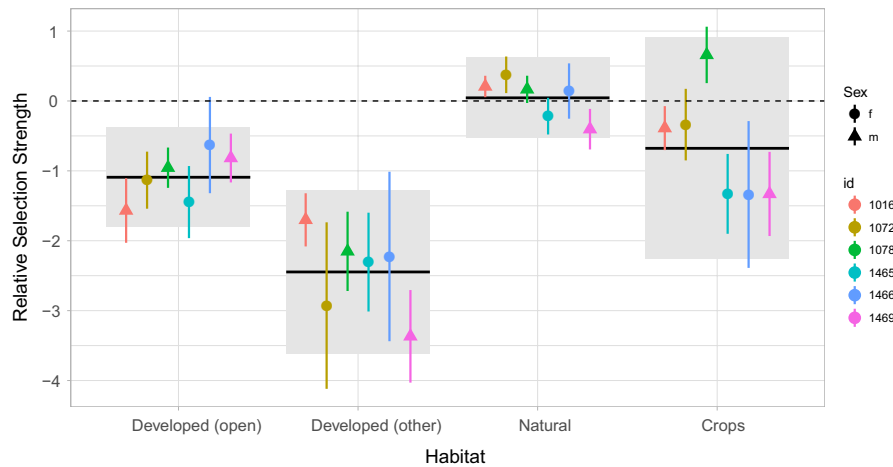
FIGURE 3 Point estimates with 95% confidence intervals for the relative selection strength (Avgar, Lele, Keim, & Boyce, 2017) for different landuse classes (we used wetland forests and wet areas as the reference class). Different colors indicate the id of the animals and symbols the sex (circles for female and triangles for males). Population-level estimates are given by solid horizontal lines and 95% confidence intervals at population level are given by the light gray boxes. The dashed horizontal line indicates no preference relative to wetland forest (the reference category)

Next, we prepare again the landuse data. This time we reclassify the landuse raster (using `raster::reclassify`) into five categories: water and wetland forests, developed open spaces, other developed areas, forests and shrubs, and crops.

```
land_use <- raster("data/landuse_study_area.tif")
rcl <- cbind(c(11, 12, 21:24, 31, 41:43, 51:52, 71:74, 81:82, 90, 95),
             c(1, 1, 2, 3, 3, 3, 2, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 1, 1))
lu <- reclassify(land_use, rcl)
names(lu) <- "landuse"
```

We again first inspect the sampling rate of the six individuals:

```
dat_all %>% mutate(sr = lapply(trk, summarize_sampling_rate)) %>%
  select(id, sr) %>% unnest
```

```
# A tibble: 6 x 10
     id   min    q1 median  mean    q3   max    sd     n unit
  <int> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <int> <chr>
1  1072  3.92  9.80   10.0  20.7  10.3  1650  95.2  1348 min
2  1465 0.400  1.97   2.03  8.93  3.98  1080  48.1  3003 min
3  1466 0.417  1.97   2.07  15.7  4.08  2167 104    1500 min
4  1078  1.35  9.78   10.0  21.7  10.3  2111  87.2  1637 min
5  1469 0.417  1.97   2.17  13.3  5.42  2889  90.2  2435 min
6  1016 0.100  1.93   2.03  8.04  2.57  1209  44.0  8957 min
```

This time we see that some individuals have a 2 min sample rate and others a 10 min one. Thus, we decided to resample the tracks to the same sampling rate of 10 min (noting that (i)SSF inference is scale dependent; Signer et al., 2017) using `amt::track _ resample`. We then filter again bursts, keeping only those with at least three points (`filter _ min _ n _ burst`), convert from a point to a step representation of the tracks (`amt::steps _ by _ burst`) and generate nine random steps for each observed step (`amt::random _ steps`), extract the environmental covariates (`amt::extract _ covariates`), convert landuse to a factor (`mutate`), and fit a SSF (`amt::fit _ issf`). The main difference to the previous example here, is that the all the steps from above are wrapped into one `mutate` call. This call creates a new column to `dat _ all` called `ssf`. This is a list column and each entry in this column contains a fitted SSF.

```
m1 <- dat_all %>% mutate(ssf = lapply(trk, function(x) {
  x %>% amt::track_resample(rate = minutes(10),
  tolerance = minutes(2)) %>%
    amt::filter_min_n_burst(min_n = 3) %>%
    amt::steps_by_burst() %>% random_steps() %>%
    amt::extract_covariates(lu) %>%
    mutate(landuse_end = factor(landuse)) %>%
    amt::fit_issf(case_ ~ landuse_end + strata(step_id_))
}))
```

`m1` is still a `data _ frame` with one new column: `ssf` that is again a list column with a fitted SSF. From here, it is easy to investigate coefficients for several animals and look at population-level effects. The results suggest that there are some general population-level trends (Figure 3). All fishers seem to prefer wetland forests and natural areas relative to developed areas (of either type), whereas considerable among-animal variability in the coefficients for crops makes it difficult to draw firm conclusions about this landuse type. Lastly, there seems to be little differentiation based on sex (Figure 3, code provided in Supporting information Data S1).

4 | DISCUSSION AND OUTLOOK

We have illustrated how `amt` can be used to fit SSFs and explore temporal movement and habitat selection patterns at the individual and population levels. We demonstrated how an iSSF, fit to a single fisher, can be used to simulate utilization distributions (UDs; Signer et al., 2017). The UD map (Figure 2) can be thought of as a stochastic approximation of a mechanistic home-range model (Moorcroft & Lewis, 2013). Whereas traditional home-range estimators offer static summaries of space-use patterns, mechanistic home-range estimators can provide insights into the movement and habitat selection processes that give rise to these patterns. In our model, we included an interaction between parameters of the movement model and time of the

day (day/night), allowing us to explore time-dependent space-use patterns (Figure 2b,d). We then showed how `amt` can be used to conduct similar analyses with more than one animal, allowing us to investigate population-level effects by looking at the distribution of animal-specific coefficients (Figure 3 Fieberg et al., 2010; Hooten et al., 2016). In our second example (modeling six fisher), we restricted the analysis to habitat selection; incorporation a movement model would be straight forward here as well (Prokopenko et al., 2017a; Scrafford et al., 2018).

We expect `amt` will contribute to movement ecology in two ways. First, `amt` is likely to help researchers manage their data and analyses using a more reproducible workflow, a much discussed issue (e.g., Cooper & Hsing, 2017; Lewis, Vander Wal, & Fifield, 2018). Second, `amt` will facilitate the use of iSSFs by a wider community of ecologists and also allow them to more fully realize the power of these methods (e.g., by modeling how landscape features influence both movement and habitat selection processes). Prior to `amt`, software for implementing iSSFs was not available. Therefore, use of iSSFs required custom-written code. `amt` provides functions that make it easy to fit iSSFs and to explore predicted space-use patterns from fitted models.

Besides the introduced functions to fit SSFs, `amt` provides additional functions for calculating home ranges, estimating RSFs and other utility functions to work with telemetry data and interface with other packages. Future development of `amt` will focus on increased functionality by adding more functions for data quality assurance. We also hope to implement UD estimation for more sophisticated models. Currently, `amt` does not allow simulating UDs from models that include interactions between movement (speed and/or turning angles) and other covariates.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

None declared.

AUTHORS CONTRIBUTION

JS, JF, and TA conceived the ideas and designed the package; JS programmed the package; JS, JF, and TA analyzed the case study; JS, JF, and TA wrote the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

DATA ACCESSIBILITY

The fisher data can be downloaded from movebank LaPoint et al. (2013b) and the landcover data are freely available at https://www.mrlc.gov/nlcd11_data.php.

ORCID

Johannes Signer (ID) https://orcid.org/0000-0002-1771-7775

John Fieberg (ID) https://orcid.org/0000-0002-3180-7021

Tal Avgar (ID) https://orcid.org/0000-0002-8764-6976

REFERENCES

Agostinelli, C., & Lund, U. (2017). R package circular: Circular Statistics (version 0.4-93). CA: Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA.

Akaike, H. (1974). A new look at the statistical model identification. _IEEE transactions on automatic control_, 19, 716–723. https://doi.org/10.1109/TAC.1974.1100705

Avgar, T., Lele, S. R., Keim, J. L., & Boyce, M. S. (2017). Relative selection strength: Quantifying effect size in habitat-and step-selection inference. _Ecology and Evolution_, 7, 5322–5330. https://doi.org/10.1002/ece3.3122

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, 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, 139–159. https://doi.org/10.1007/s00285-007-0149-8

Beyer, H. L., Haydon, D. T., Morales, J. M., Frair, J. L., Hebblewhite, M., Mitchell, M., & Matthiopoulos, J. (2010). The interpretation of habitat preference metrics under use-availability designs. _Philosophical Transactions of the Royal Society B: Biological Sciences_, 365, 2245–2254. https://doi.org/10.1098/rstb.2010.0083

Bivand, R., & Lewin-Koh, N. (2017). maptools: Tools for Reading and Handling Spatial Objects. R package version 0.9-2.

Boyce, M. S., & McDonald, L. L. (1999). Relating populations to habitats using resource selection functions. _Trends in Ecology & Evolution_, 14, 268–272. https://doi.org/10.1016/S0169-5347(99)01593-1

Brost, B. M., Hooten, M. B., Hanks, E. M., & Small, R. J. (2015). Animal movement constraints improve resource selection inference in the presence of telemetry error. _Ecology_, 96, 2590–2597. https://doi.org/10.1890/15-0472.1

Brown, D. D., LaPoint, S., Kays, R., Heidrich, W., Kümmeth, F., & Wikelski, M. (2012). Accelerometer-informed gps telemetry: Reducing the trade-off between resolution and longevity. _Wildlife Society Bulletin_, 36, 139–146. https://doi.org/10.1002/wsb.111

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 of London. _Series B_, _Biological sciences_, 365, 2157–2162. https://doi.org/10.1098/rstb.2010.0107

Calabrese, J. M., Fleming, C. H., & Gurarie, E. (2016). ctmm: An R package for analyzing animal relocation data as a continuous-time stochastic process. _Methods in Ecology and Evolution_, 7, 1124–1132. https://doi.org/10.1111/2041-210X.12559

Cooper, N., & Hsing, P. Y. (2017). A guide to reproducible code in ecology and evolution. Technical report, British Ecological Society. Available at https://www.britishecologicalsociety.org/wp-content/uploads/2017/12/guide-to-reproducible-code.pdf.

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. _Journal of Statistical Software_, 64, 1–34.

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*, e0122947. https://doi.org/10.1371/journal.pone.0122947

Fieberg, J. R., Forester, J. D., Street, G. M., Johnson, D. H., ArchMiller, A. A., & Matthiopoulos, J. (2018). Used-habitat calibration plots: A new procedure for validating species distribution, resource selection, and step-selection models. *Ecography*, *41*, 737–752. https://doi.org/10.1111/ecog.03123

Fieberg, J., Matthiopoulos, J., Hebblewhite, M., Boyce, M. S., & Frair, J. L. (2010). Correlation and studies of habitat selection: Problem, red herring or opportunity? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *365*, 2233–2244. https://doi.org/10.1098/rstb.2010.0079

Forester, J. D., Im, H. K., & Rathouz, P. J. (2009). Accounting for animal movement in estimation of resource selection functions: Sampling and data analysis. *Ecology*, *90*, 3554–3565. https://doi.org/10.1890/08-0874.1

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*, 1320–1330. https://doi.org/10.1890/04-0953

Gail, M. H., Lubin, J. H., & Rubinstein, L. V. (1981). Likelihood calculations for matched case-control studies and survival studies with tied death times. *Biometrika*, *68*, 703–707. https://doi.org/10.1093/biomet/68.3.703

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*, 1–25.

Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, *12*, 395–408. https://doi.org/10.1111/j.1461-0248.2009.01293.x

Hanks, E. M., Hooten, M. B., Alldredge, M. W. (2015). Continuous-time discrete space models for animal movement. *The Annals of Applied Statistics*, *9*, 145–165. https://doi.org/10.1214/14-AOAS803

Henry, L., & Wickham, H. (2017). purrr: Functional Programming Tools. R package version 0.2.2.2.

Hijmans, R. J. (2017). raster: Geographic Data Analysis and Modeling. R package version 2.6-7.

Hooten, M. B., Buderman, F. E., Brost, B. M., Hanks, E. M., & Ivan, J. S. (2016). Hierarchical animal movement models for population-level inference. *Environmetrics*, *27*, 322–333. https://doi.org/10.1002/env.2402

Hooten, M. B., Johnson, D. S., Hanks, E. M., & Lowry, J. H. (2010). Agent-based inference for animal movement and selection. *Journal of Agricultural, Biological and Environmental Statistics*, *15*, 523–538. https://doi.org/10.1007/s13253-010-0038-2

Hooten, M. B., Johnson, D. S., McClintock, B. T., & Morales, J. M. (2017). *Animal movement: Statistical models for telemetry data*. Boca Raton, FL: CRC Press.

Johnson, D. S., Hooten, M. B., & Kuhn, C. E. (2013). Estimating animal resource selection from telemetry data using point process models. *Journal of Animal Ecology*, *82*, 1155–1164. https://doi.org/10.1111/1365-2656.12087

Johnson, D. S., Thomas, D. L., Ver Hoef, J. M., & Christ, A. (2008). A general framework for the analysis of animal resource selection from telemetry data. *Biometrics*, *64*, 968–976. https://doi.org/10.1111/j.1541-0420.2007.00943.x

Kays, R., Crofoot, M. C., Jetz, W., & Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. *Science*, *348*, aaa2478. https://doi.org/10.1126/science.aaa2478

LaPoint, S., Gallery, P., Wikelski, M., & Kays, R. (2013a). Animal behavior, cost-based corridor models, and real corridors. *Landscape Ecology*, *28*, 1615–1630. https://doi.org/10.1007/s10980-013-9910-0

LaPoint, S., Gallery, P., Wikelski, M., & Kays, R. (2013b). Data from: Animal behavior, cost-based corridor models, and real corridors. *Landscape Ecology*, *28*, 1615–1630. https://doi.org/10.1007/s10980-013-9910-0

Lewis, K. P., Vander Wal, E., & Fifield, D. A. (2018). Wildlife biology, big data, and reproducible research. *Wildlife Society Bulletin.*, *42*, 172–179. https://doi.org/10.1002/wsb.847

Manly, B., McDonald, L., Thomas, D. L., McDonald, T. L., & Erickson, W. P. (2007). *Resource selection by animals: Statistical design and analysis for field studies*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

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*, 249–269. https://doi.org/10.1007/s13253-017-0285-6

McClintock, B. T., & Michelot, T. (2018). momentuhmm: R package for generalized hidden markov models of animal movement. *Methods in Ecology and Evolution*, *9*, 1518–1530. https://doi.org/10.1111/2041-210X.12995

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*, 1308–1315. https://doi.org/10.1111/2041-210X.12578

Moorcroft, P. R., & Lewis, M. A. (2013). *Mechanistic home range analysis.(MPB-43)*. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9781400849734

Müller, K., & Wickham, H. (2018). tibble: Simple Data Frames. R package version 1.4.2.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, *105*, 19052–19059. https://doi.org/10.1073/pnas.0800375105

Northrup, J. M., Hooten, M. B., Anderson, C. R., & Wittemyer, G. (2013). Practical guidance on characterizing availability in resource selection functions under a use–availability design. *Ecology*, *94*, 1456–1463. https://doi.org/10.1890/12-1688.1

Paton, R. S., & Matthiopoulos, J. (2016). Defining the scale of habitat availability for models of habitat selection. *Ecology*, *97*, 1113–1122.

Prokopenko, C. M., Boyce, M. S., & Avgar, T. (2017a). Characterizing wildlife behavioural responses to roads using integrated step selection analysis. *Journal of Applied Ecology*, *54*, 470–479. https://doi.org/10.1111/1365-2664.12768

Prokopenko, C. M., Boyce, M. S., & Avgar, T. (2017b). Extent-dependent habitat selection in a migratory large herbivore: Road avoidance across scales. *Landscape ecology*, *32*, 313–325. https://doi.org/10.1007/s10980-016-0451-1

Scrafford, M., Avgar, T., Heeres, R., & Boyce, M. (2018). Roads elicit negative movement and habitat-selection responses by wolverines (gulo gulo luscus). *Behavioral Ecology*, *29*, 534–542.

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

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer. https://doi.org/10.1007/978-1-4757-3294-8

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

Viana, D. S., Granados, J. E., Fandos, P., Pérez, J. M., Cano-Manuel, F. J., Burón, D., … Soriguer, R. C. (2018). Linking seasonal home range size with habitat selection and movement in a mountain ungulate. *Movement Ecology*, *6*, 1. https://doi.org/10.1186/s40462-017-0119-8

Wickham, H. (2017). tidyverse: Easily Install and Load 'Tidyverse' Packages. R package version 1.1.1.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.