Step Selection Functions
A brief Introduction

AfriMove Meeting

2025-05-22

Johannes Signer (jsigner@uni-goettingen.de)

UNIVERSITÄT
GÖTTINGEN

Agenda

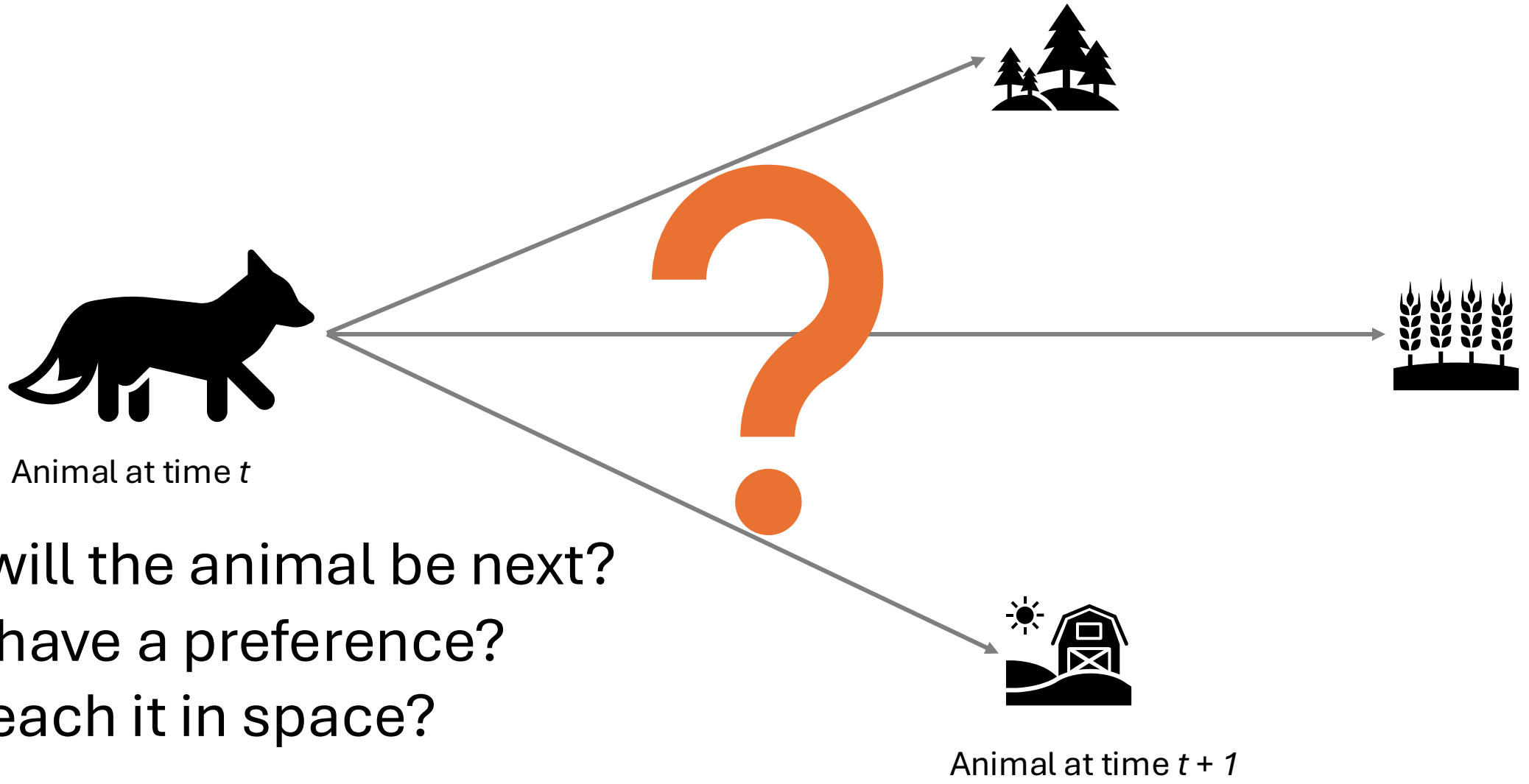
- Some background about SSF
- A practical example of red deer from northern Germany

- Slides, data and coded examples are available here: https://github.com/jmsigner/2025_afrimove_talk

Who am I?

- Research in Wildlife Science (University of Göttingen, German; Group of Niko Balkenhol).
- Interest in statistical ecology, particularly in the analysis of telemetry data.
- Involved in SASSCAL Antelope project.

What is it all about?

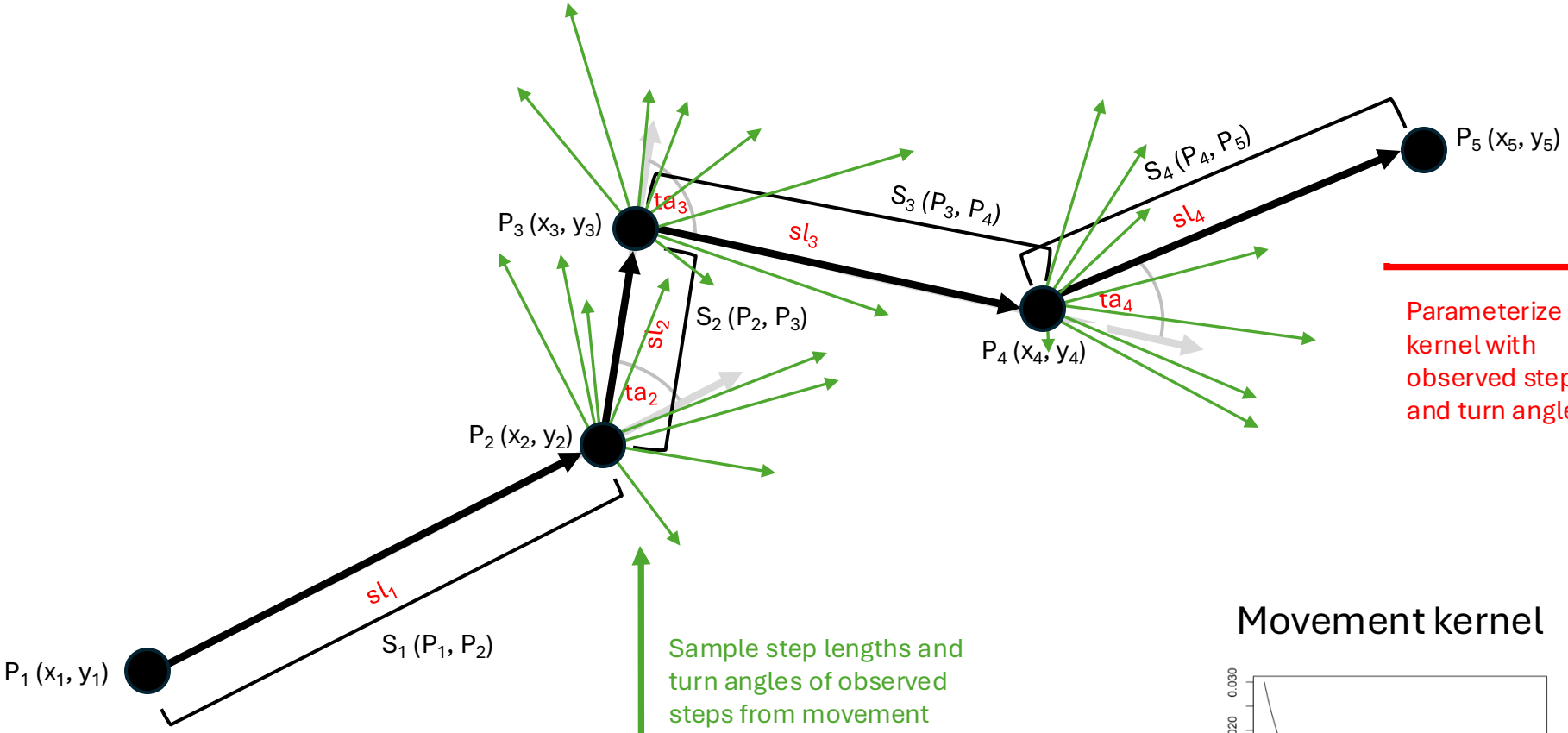


Where will the animal be next?
Does it have a preference?
Can it reach it in space?

What is the difference:

- Resource Selection Function (RSF; also sometimes referred to as Habitat Selection Function; HSF): **Global availability**, i.e., the animal assumed to reach each available point at each time step.
- Step Selection Functions (SSF): **Conditional availability**, available steps are conditioned on the current position of the animal.
- Integrated Step Selection Function (iSSF): **Conditional availability** and also estimating a movement kernel, i.e., where to expect the animal in the next time step, if there were no selection.

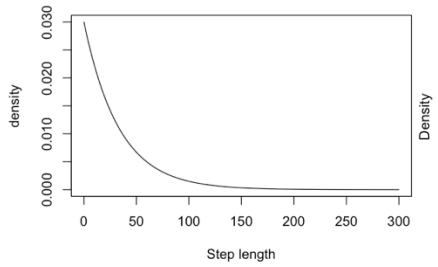
SSF the idea



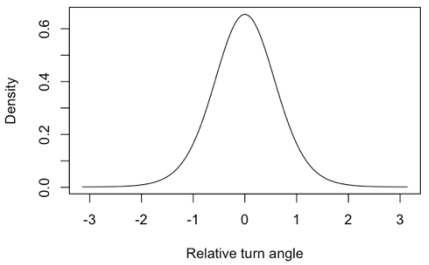
Parameterize movement kernel with observed step lengths (sl) and turn angles (ta)

Sample step lengths and turn angles of observed steps from movement kernel.

Movement kernel



Step length distribution: usually exponential or gamma.



Turn angle distribution: usually von Mises.

GPS relocations

Steps connecting two consecutive points

Next steps

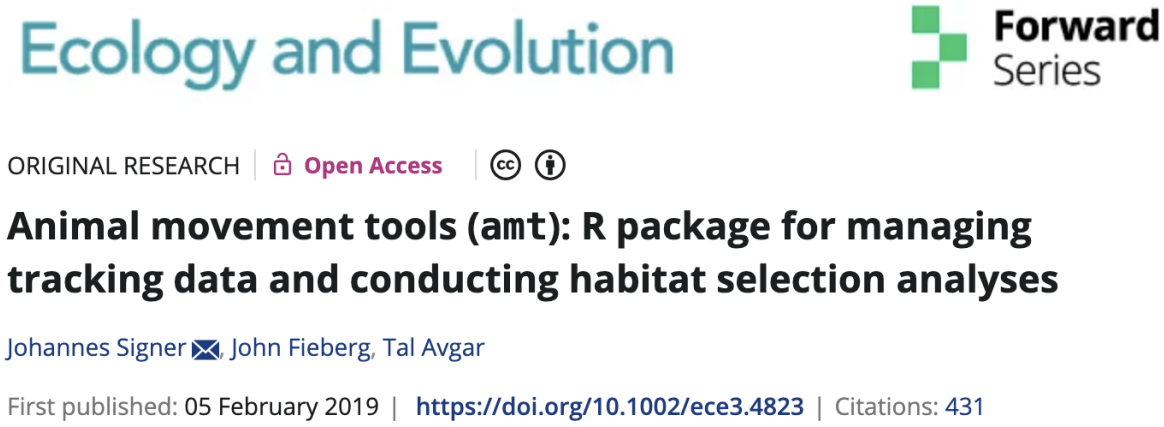
- Extract covariate values at the end (habitat selection) or start (movement kernel) of random and observed steps
- Estimate selection coefficients with using a conditional logistic regression (other frameworks can be used).
- Work with the results

Adding movement to Habitat Selection

- Avgar 2015 showed, that we can (and in fact) should disentangle habitat selection and movement.

Implementation of an iSSF

- iSSFs are not hard, but some steps (especially data preparation can be tedious).
- **amt** was designed to make this step easier.
- Strong dependency on other R packages to perform geospatial analysis (e.g., **sf**, **terra**).



Typical workflow

- Load data (here we use data from the amt package)
- Make a track (amt data format)
- *Optionally, regularize the track (for regular sampling intervals).*
- Create steps
- Create random steps
- Extract covariates
- Annotate time of day

```
data(deer)

dat1 <-
  deer |>
  steps() |>
  random_steps() |>
  extract_covariates(dist_forest) |>
  time_of_day()
```

This yields the following data

```
> head(dat1)
# A tibble: 6 × 13
      x1_      x2_      y1_      y2_     sl_    ta_ t1_                 t2_                 dt_       case_ step_id_ dist_forest tod_end_
*   <dbl>    <dbl>    <dbl>    <dbl>   <dbl>  <dbl> <dttm>              <dttm>              <drtn>    <lgl>    <dbl>       <dbl> <fct>
1 4314053. 4314105. 3445768. 3445859. 104.     3.00 2008-03-30 06:00:54 2008-03-30 12:01:47 6.014722 … TRUE         3      -0.398 day
2 4314053. 4314053. 3445768. 3445768.   0.149 -0.218 2008-03-30 06:00:54 2008-03-30 12:01:47 6.014722 … FALSE        3      -0.787 day
3 4314053. 4314299. 3445768. 3445243. 580.     0.819 2008-03-30 06:00:54 2008-03-30 12:01:47 6.014722 … FALSE        3       0.119 day
4 4314053. 4314157. 3445768. 3446139. 385.    -3.03  2008-03-30 06:00:54 2008-03-30 12:01:47 6.014722 … FALSE        3       0.545 day
5 4314053. 4314175. 3445768. 3445890. 172.     2.74  2008-03-30 06:00:54 2008-03-30 12:01:47 6.014722 … FALSE        3      -0.202 day
6 4314053. 4313506. 3445768. 3446225. 712.    -1.89  2008-03-30 06:00:54 2008-03-30 12:01:47 6.014722 … FALSE        3      -0.398 day
```

Fitting a fist model (m0)

```r
m0 <- fit_clogit(
  case_ ~ # This is the response
    dist_forest + # distance to forest
    strata(step_id_), # each stratum is an observed step with its random steps
  data = dat1, # the data
  model = TRUE # to save the input data
)
```

```
              coef exp(coef) se(coef)      z Pr(>|z|)
dist_forest -1.4222    0.2412   0.1118 -12.72   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The closer to the forest (negative values) the better.

Relative Selection Strength (RSS)

- The RSS indicates how much more likely it is to select a location x_1 relative to a reference location x_2.

- $RSS(x_2, x_1) = w(x_2) / w(x_1)$

```
> x1 <- data.frame(dist_forest = 1)
> x2 <- data.frame(dist_forest = 0)
> exp(log_rss(m0, x1, x2)$df$log_rss)
[1] 0.2411831
> exp(coef(m0))
dist_forest
  0.2411831
```

Ecology and Evolution

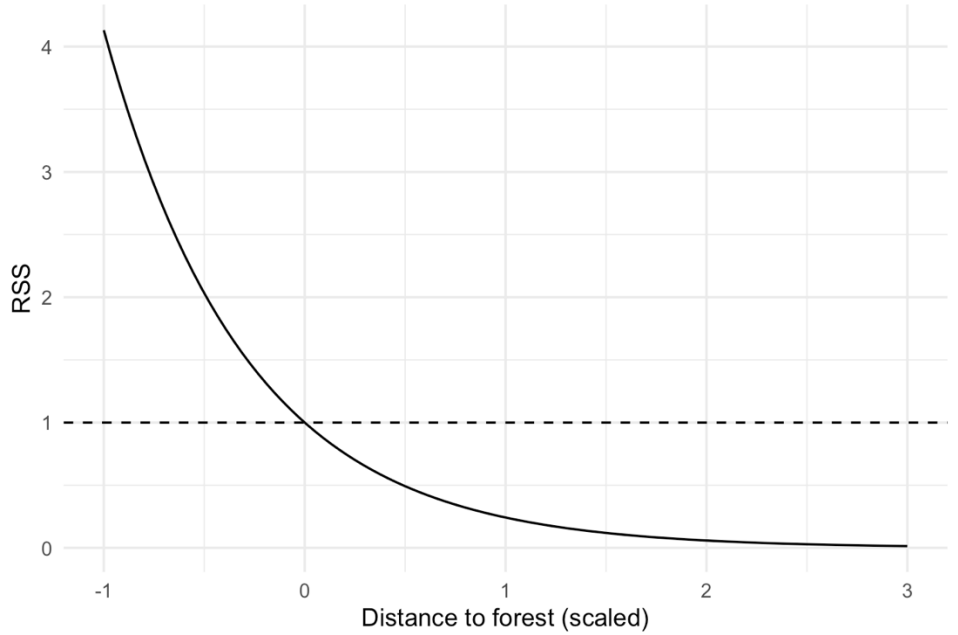
Forward Series

ORIGINAL RESEARCH | Open Access

Relative Selection Strength: Quantifying effect size in habitat- and step-selection inference

Tal Avgar, Subhash R. Lele, Jonah L. Keim, Mark S. Boyce

First published: 14 June 2017 | https://doi.org/10.1002/ece3.3122 | Citations: 151



Let’s add an interaction with time of day

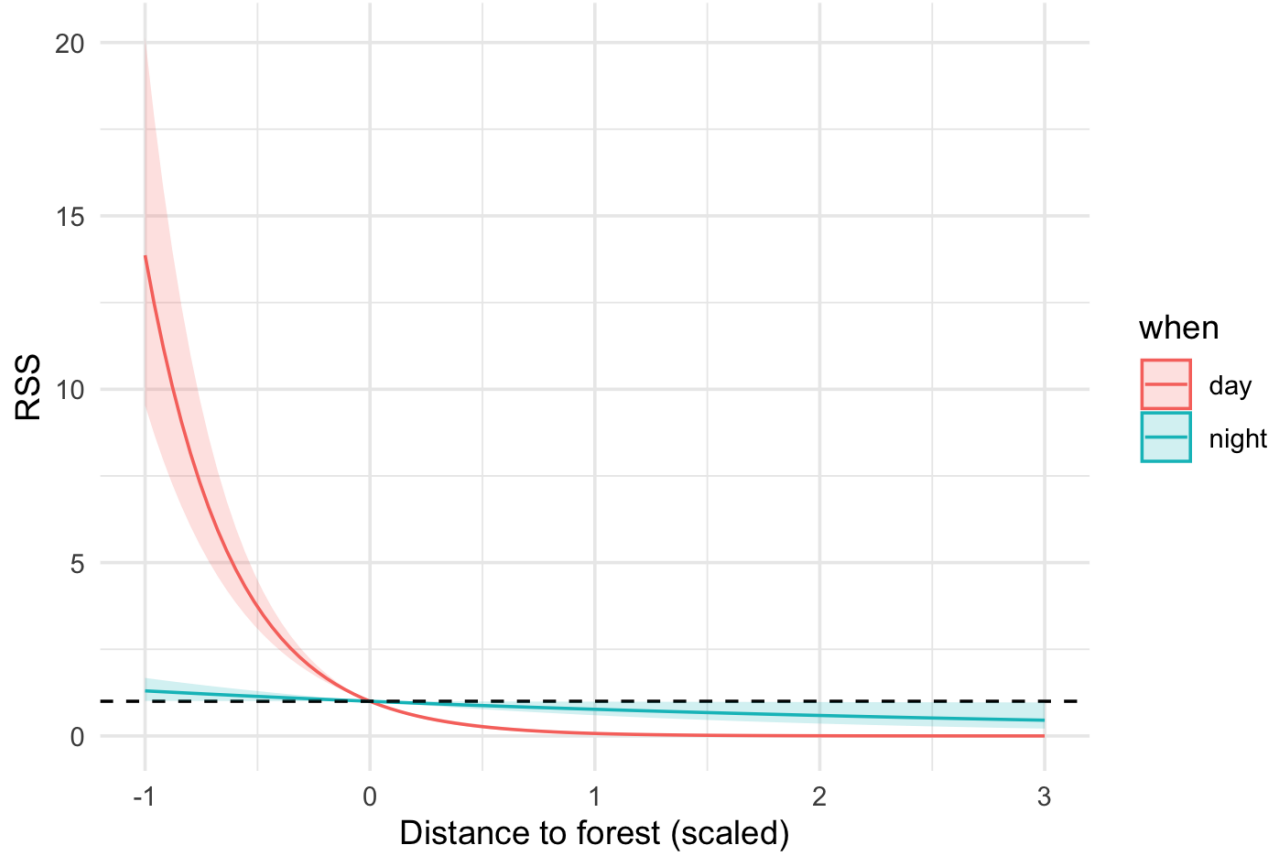
- Maybe, there is different selection behavior for distance to forest during day than during the night?

```
m1 <- fit_clogit(
  case_ ~ # This is the response
    dist_forest + # distance to forest
    dist_forest:tod_end_ + # as interaction with time of day
    strata(step_id_), # each stratum is an observed step with its random
steps
  data = dat1, # the data
  model = TRUE # to save the input data
)
```

difference between day and night (2)

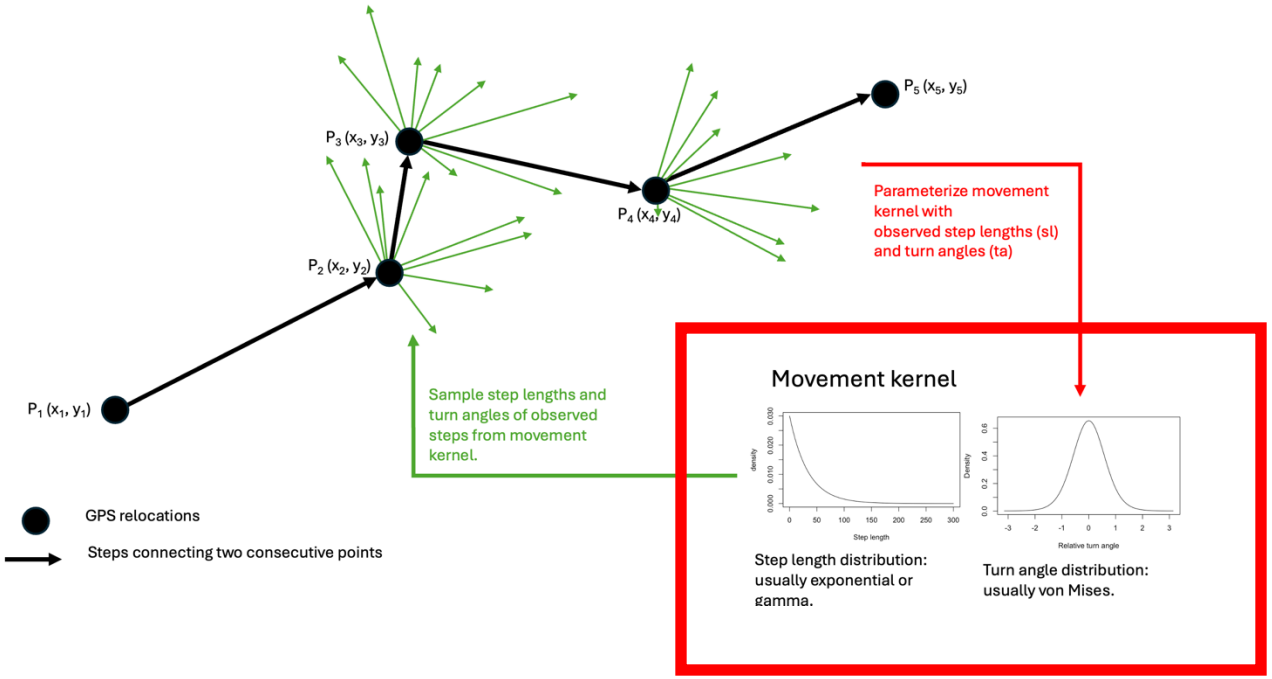
```
                              coef exp(coef) se(coef)     z Pr(>|z|)
dist_forest                -2.62923   0.07213  0.19056 -13.8   <2e-16 ***
dist_forest:tod_end_night   2.36636  10.65857  0.22968  10.3   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’
```

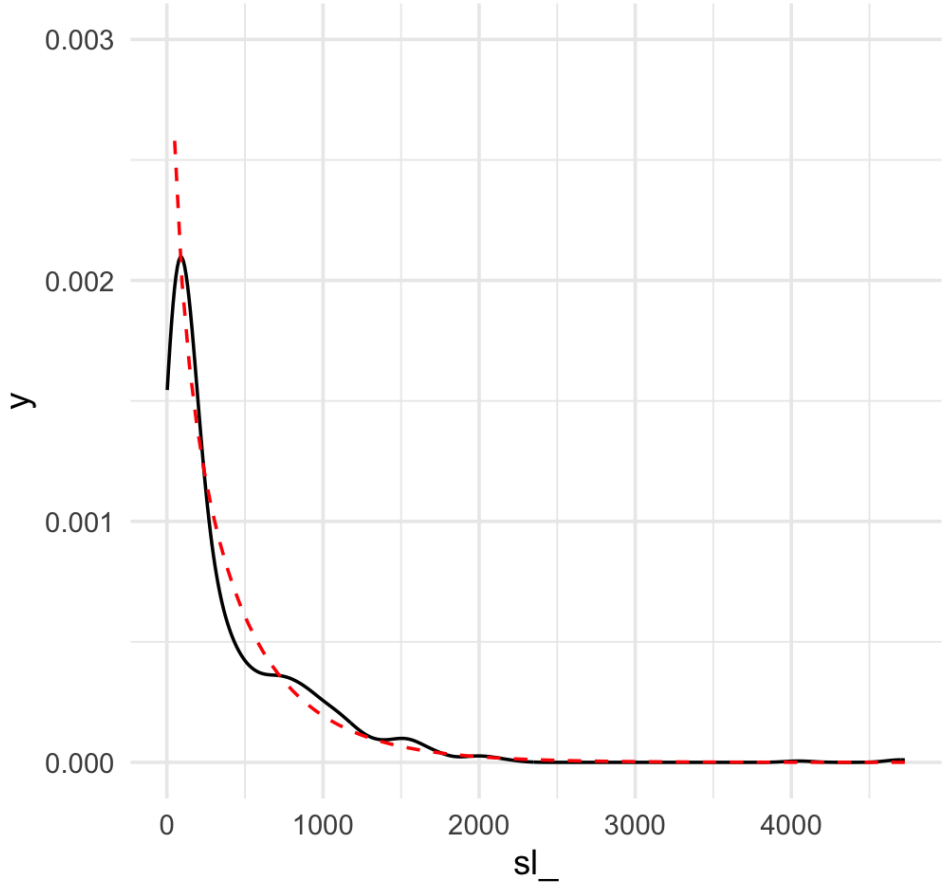
During day: Even stronger selection for forest.
During night?



Include movement: The tentative movement kernel

Step length distribution (shown here) and turn angle distribution estimated from the data





Extending SSF to iSSF

```r
m2 <- fit_clogit(
  case_ ~ # This is the response
    dist_forest + # distance to forest
    dist_forest:tod_end_ + # as interaction with time of day

    # movement model
    sl_ +
    log(sl_) +

    strata(step_id_), # each stratum is an observed step with its random steps
  data = dat1, # the data
  model = TRUE # to save the input data
)
```

We use a Gamma distribution to model step lengths: sl_ links to the shape and log(sl_) to the scale parameter.

```
                                coef  exp(coef)   se(coef)       z Pr(>|z|)
dist_forest                -2.6255942  0.0723967  0.1879899 -13.967   <2e-16 ***
sl_                         0.0001785  1.0001785  0.0001249   1.429    0.153
log(sl_)                    0.0497895  1.0510498  0.0360540   1.381    0.167
dist_forest:tod_end_night   2.2453567  9.4437832  0.2289693   9.806   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Linking estimated coefficients to tentative distributions

BRITISH ECOLOGICAL SOCIETY | Journal of Animal Ecology

HOW TO… | Open Access

A ‘How to’ guide for interpreting parameters in habitat-selection analyses

John Fieberg, Johannes Signer, Brian Smith, Tal Avgar

First published: 14 February 2021 | https://doi.org/10.1111/1365-2656.13441 | Citations: 177

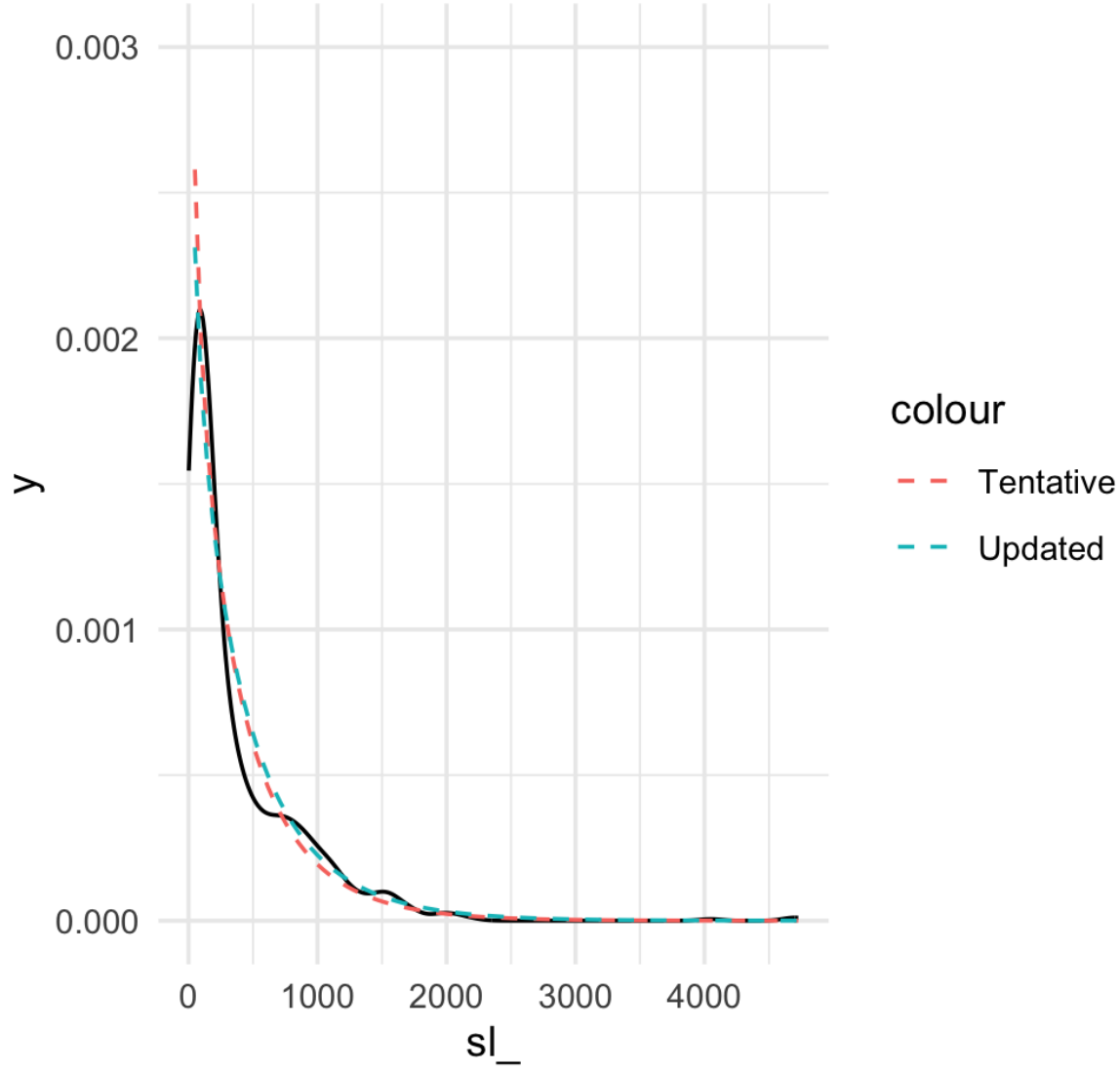
Gamma

If available step lengths were sampled from a gamma distribution with tentative shape k_0 and scale q_0, and the step length (l) and its log-transform ($ln[l]$) were included as covariates in the analysis, with resulting coefficient estimates β_l and $\beta_{ln[l]}$ (respectively), the adjusted (selection-free) step length Gamma shape ($\hat{k}$) and scale ($\hat{q}$) parameters are given by:

$$\begin{cases} \hat{k} = k_0 + \beta_{ln[l]} \\ \hat{q} = \frac{1}{\left(\frac{1}{q_0} - \beta_l\right)} \end{cases}$$

Updating the tentative distribution

- The function *update_sl_distr()* updates distributions, if there are no interactions (green dashed line).



Interaction with movement:

- We can now start to test, if the step length distribution and thus the displacement depends on covariates.
- We will start by looking, if time of day has an effect.

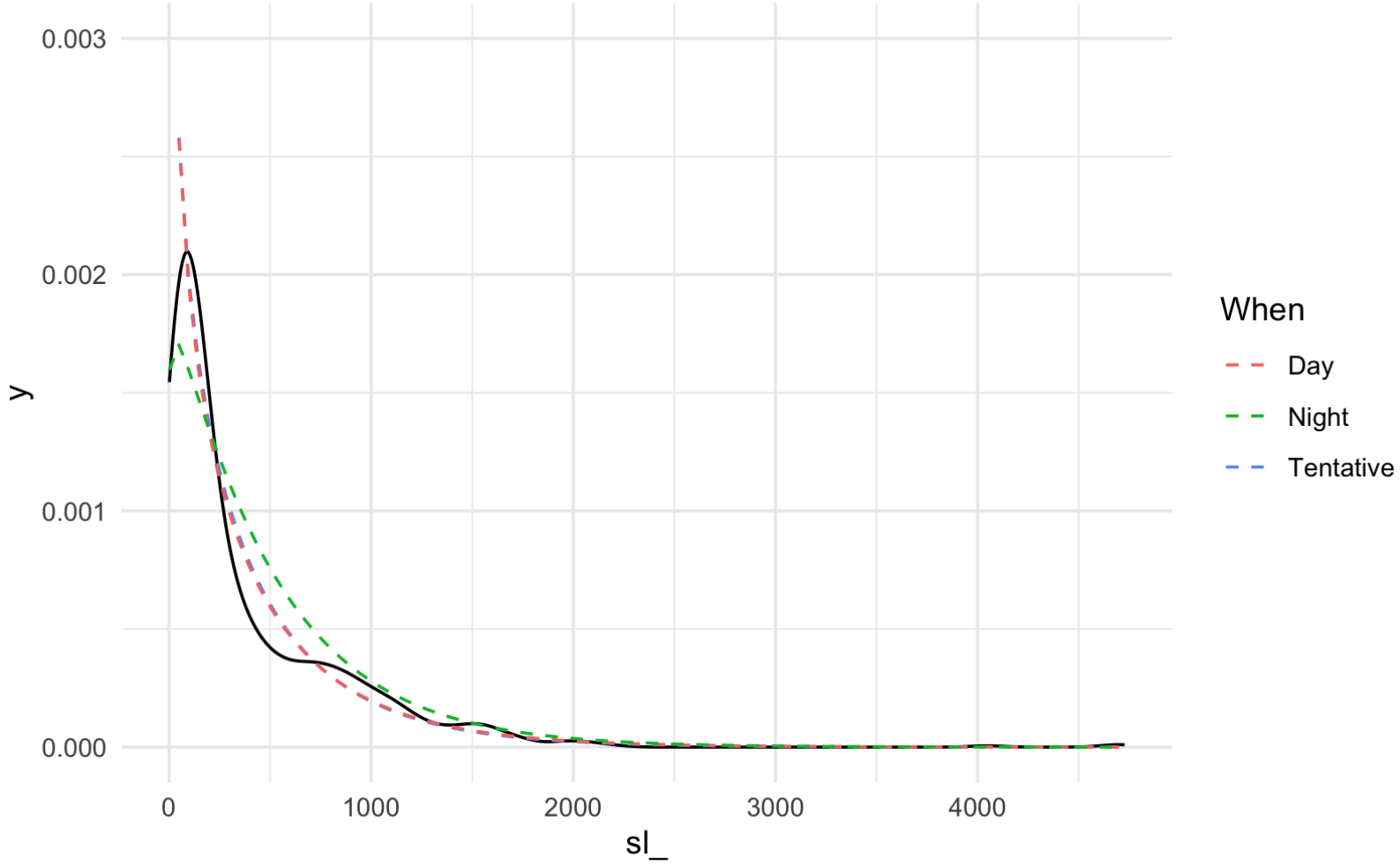
```
m3 <- fit_clogit(
  case_ ~ # This is the response
    dist_forest + # distance to forest
    dist_forest:tod_end_ + # as interaction with time of day

    # Movement kernel
    sl_ +
    log(sl_) +

    # Is movement different for different times?
    sl_:tod_start_ +
    log(sl_):tod_start_ +
    strata(step_id_), # each stratum is an observed step with its r
  data = dat1, # the data
  model = TRUE # to save the input data
)
```

Results of model 3

```
                              coef  exp(coef)   se(coef)       z Pr(>|z|)    
dist_forest                -2.5742862  0.0762082  0.1836126 -14.020  < 2e-16 ***
sl_                         0.0001149  1.0001149  0.0001546   0.743 0.457413    
log(sl_)                   -0.0305555  0.9699066  0.0392698  -0.778 0.436515    
dist_forest:tod_end_night   2.2760978  9.7386039  0.2231104  10.202  < 2e-16 ***
sl_:tod_start_night        -0.0002365  0.9997635  0.0002714  -0.871 0.383546    
log(sl_):tod_start_night    0.3304006  1.3915255  0.0901510   3.665 0.000247 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Updating the step-length distribution

- Updating becomes a bit more tedious.



Lets add some more complexity by adding twilight

- In a final model, I added a third time of day for steps (6 h sampling rate) where the start and the end time of day where different.

```
dat2 <- dat1 |> mutate(
  tod2 = case_when(
    tod_start_ == "night" & tod_end_ == "night" ~ "night",
    tod_start_ == "day" & tod_end_ == "day" ~ "day",
    .default = "twilight"
  )
)
```

Twilight model

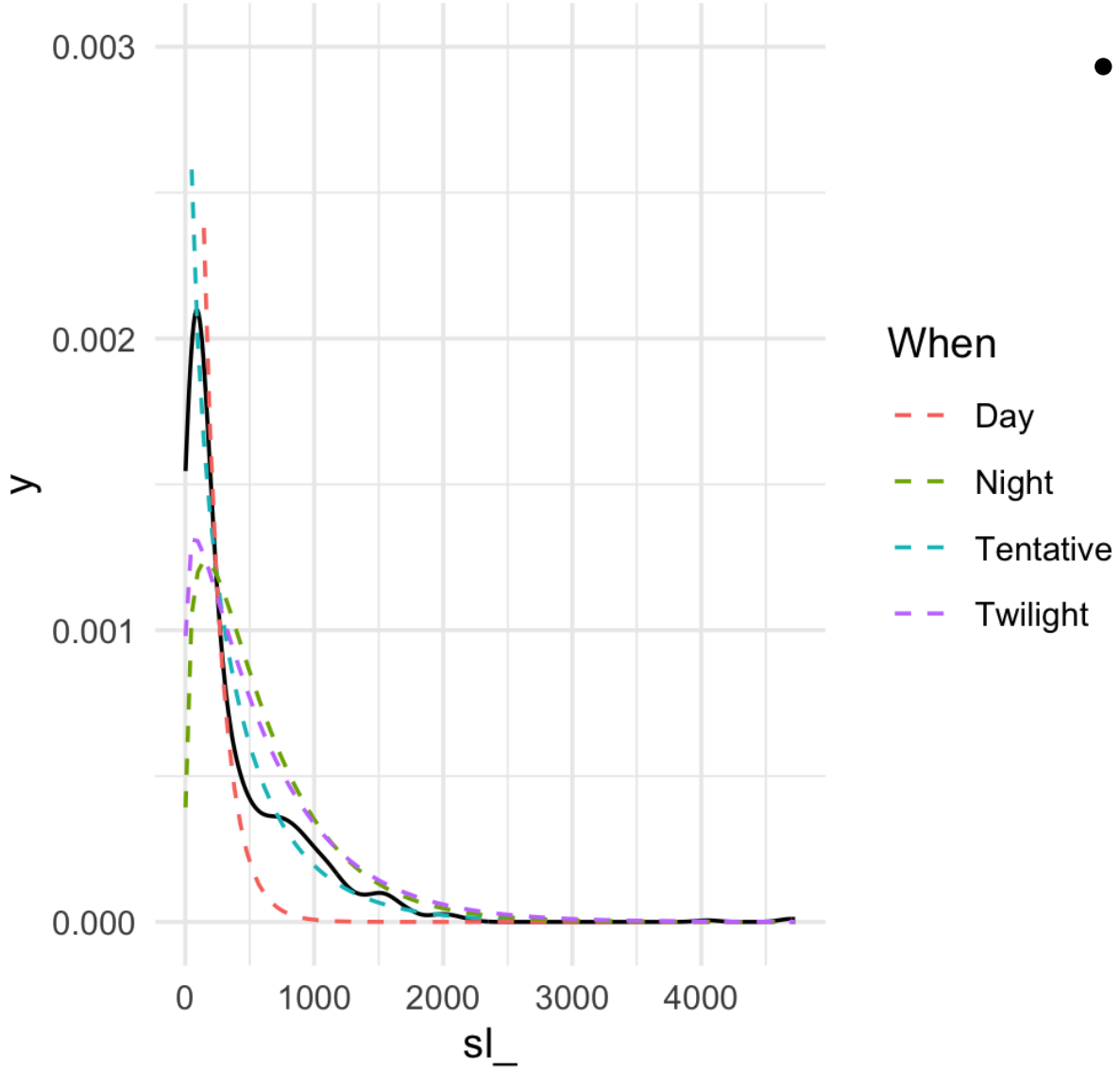
- The model as such is the same, just that tod2 gains one additional level.

```
m4 <- fit_clogit(
  case_ ~ # This is the response
    dist_forest + # distance to forest
    dist_forest:tod2 + # as interaction with time of day

    # Movement model
    sl_ +
    log(sl_) +

    sl_:tod2 +
    log(sl_):tod2 +
    strata(step_id_),
  data = dat2, # the data
  model = TRUE # to save the input data
)
```

Results of model 4

```
                              coef  exp(coef)   se(coef)      z Pr(>|z|)    
dist_forest                -1.9625610  0.1404981  0.3072839 -6.387 1.69e-10 ***
sl_                        -0.0045351  0.9954752  0.0007024 -6.456 1.07e-10 ***
log(sl_)                    0.1339335  1.1433168  0.0688472  1.945   0.0517 .  
dist_forest:tod2night       0.9187482  2.5061511  0.3989237  2.303   0.0213 *  
dist_forest:tod2twilight    0.5594194  1.7496563  0.3356013  1.667   0.0955 .  
sl_:tod2night               0.0042071  1.0042160  0.0008051  5.225 1.74e-07 ***
sl_:tod2twilight            0.0046655  1.0046764  0.0007210  6.471 9.73e-11 ***
log(sl_):tod2night          0.4670711  1.5953148  0.1843746  2.533   0.0113 *  
log(sl_):tod2twilight       0.2353464  1.2653470  0.0988580  2.381   0.0173 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Adjusting the step length distribution



- Clear differences between day and night and twilight.

What about the turn angle distribution?

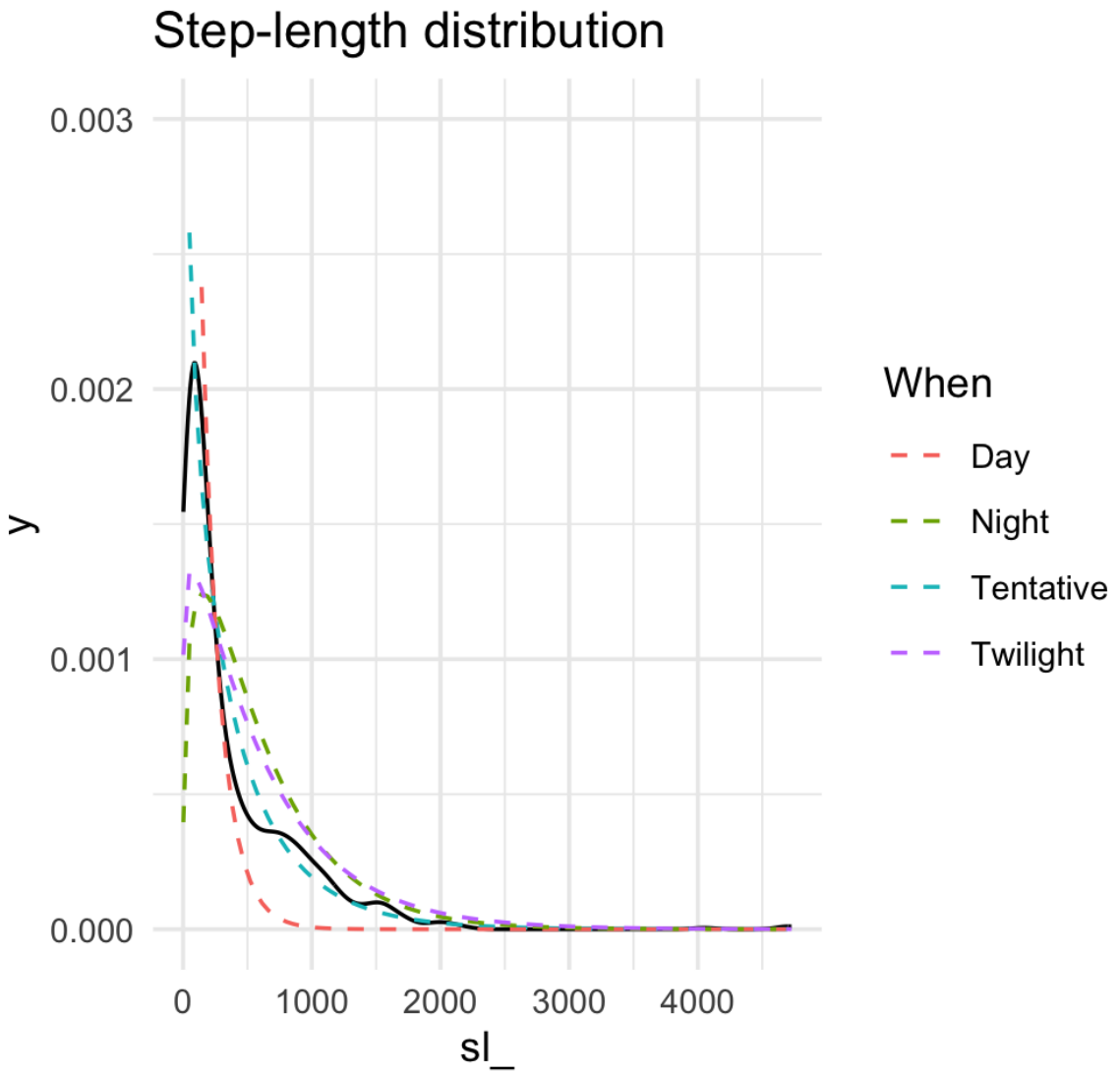
- We can add the turn-angle distribution too to the model.
- Model the concentration parameter (kappa), that is linked to the cosine of turn angle (ta_).

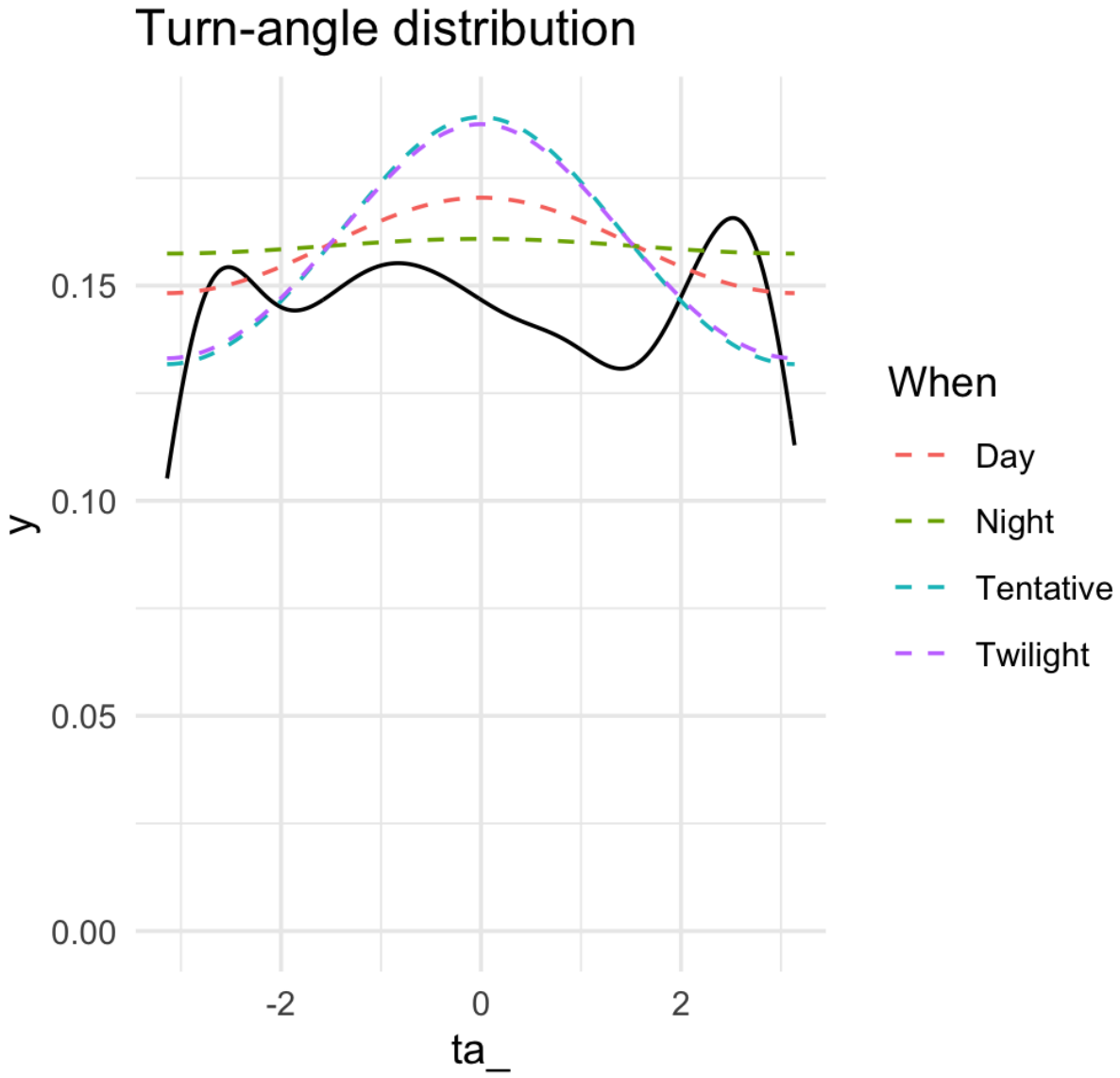
```
m5 <- fit_clogit(
  case_ ~ # This is the response
    dist_forest + # distance to forest
    dist_forest:tod2 + # as interaction with time of day

    # Movement model
    sl_ +
    log(sl_) +
    cos(ta_) +

    sl_:tod2 +
    log(sl_):tod2 +
    cos(ta_):tod2 +
    strata(step_id_),
  data = dat2, # the data
  model = TRUE # to save the input data
)
```

Results of model 5

```
                              coef  exp(coef)   se(coef)      z Pr(>|z|)
dist_forest             -1.9347384  0.1444621  0.3077478 -6.287 3.24e-10 ***
sl_                     -0.0045143  0.9954959  0.0006990 -6.459 1.06e-10 ***
log(sl_)                 0.1329933  1.1422423  0.0684373  1.943  0.05198 .
cos(ta_)                -0.2506553  0.7782906  0.0890444 -2.815  0.00488 **
dist_forest:tod2night    0.9197492  2.5086613  0.3997191  2.301  0.02139 *
dist_forest:tod2twilight 0.5916826  1.8070263  0.3361230  1.760  0.07835 .
sl_:tod2night            0.0041849  1.0041936  0.0008014  5.222 1.77e-07 ***
sl_:tod2twilight         0.0046628  1.0046736  0.0007175  6.499 8.09e-11 ***
log(sl_):tod2night       0.4661871  1.5939051  0.1838312  2.536  0.01121 *
log(sl_):tod2twilight    0.2270334  1.2548718  0.0983935  2.307  0.02103 *
cos(ta_):tod2night       0.0804597  1.0837852  0.1733877  0.464  0.64262
cos(ta_):tod2twilight   -0.1016746  0.9033235  0.1169136 -0.870  0.38449
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Movement kernel model 5





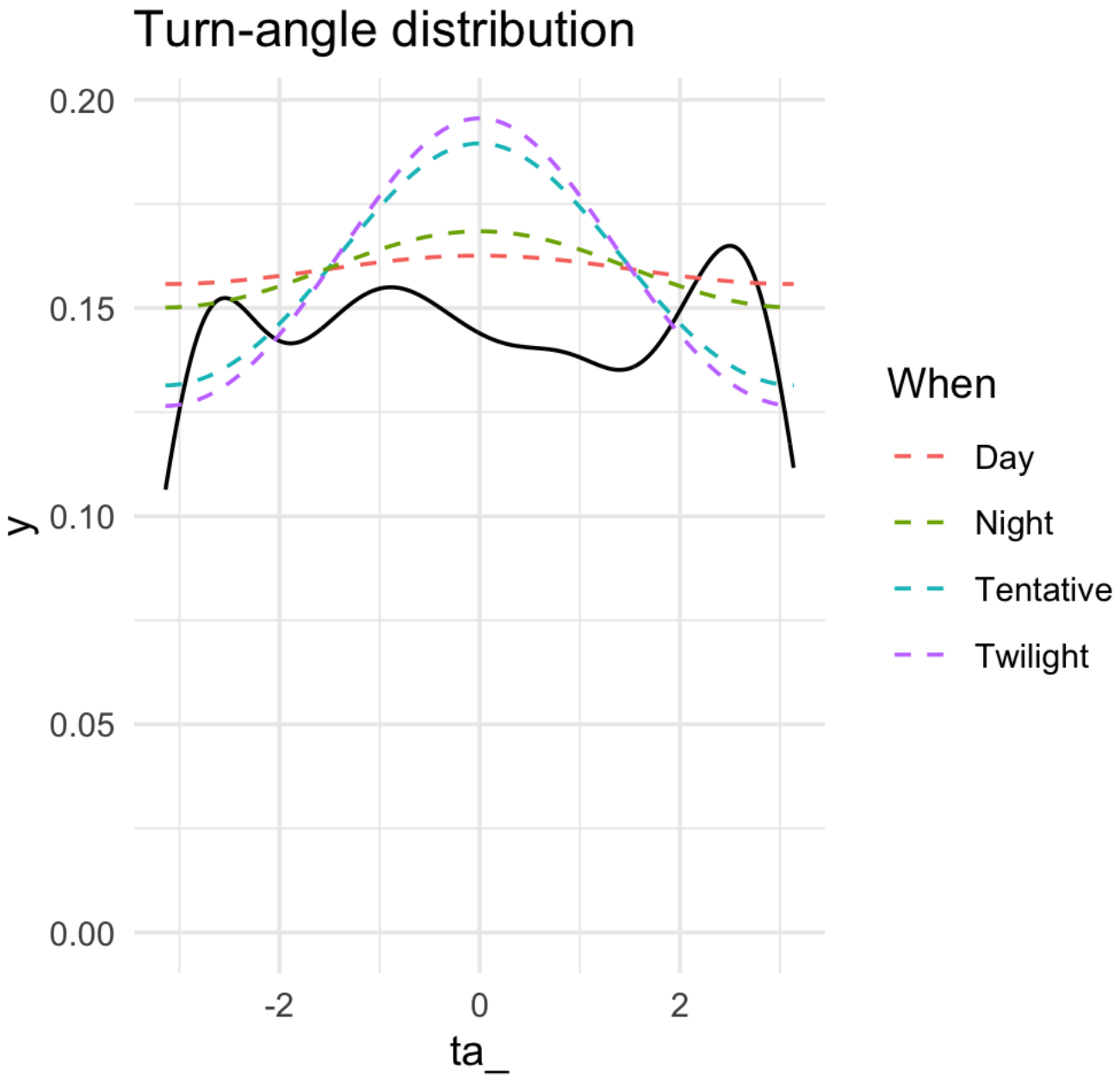
Which model to choose?

- Model 5 seems to be plausible, and is also supported by the AIC.

	df	AIC	dAIC
m0$model	1	3426.735	274.35552
m1$model	2	3322.727	170.34781
m2$model	4	3319.949	167.56913
m3$model	6	3300.236	147.85643
m4$model	9	3177.107	24.72733
m5$model	12	3152.380	0.00000

Conclusion

- Clear difference of habitat selection between different time of days.
- Larger displacement during twilight and night, than during the day.
- Most directed movement during twilight.



Beyond the basics

- Non-linear effects
- Accounting for individual variation
- Account for different behavioral states
- Simulate from a fitted model
- Validate models

ZOOLOGICAL SCIENCE

How to account for behavioral states in step-selection

RESEARCH ARTICLE | Open Access

Using lineups to evaluate goodness of fit of animal movement models

John Fieberg, Smith Freeman, Johannes Signer

Thank you for your attention

- Happy to take questions now or also later (jsigner@uni-goettingen.de)