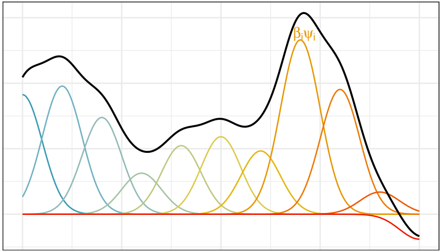
Step selection functions in mgcv

OTN ECR workshop

April 29, 2025



Outline

1. Introduction to SSFs
2. SSF implementation
3. SSFs with non-linear and random effects
4. Tutorial in R

Introduction to SSFs

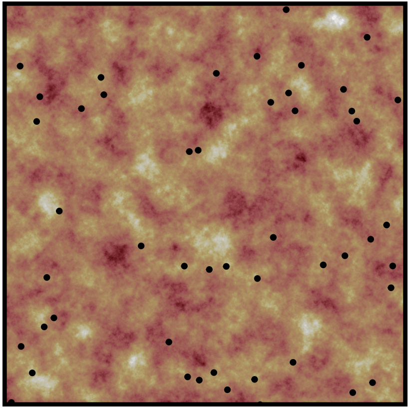
Habitat selection

Process by which animals show preference or avoidance of spatial features

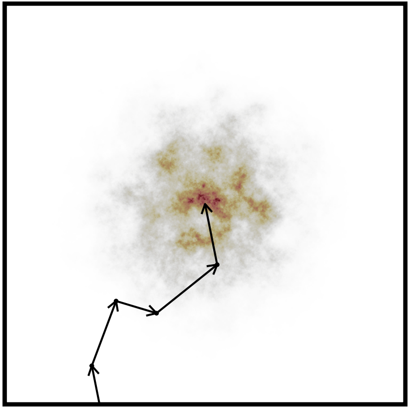
- Resources (e.g., food)
- Risks (e.g., roads, costly environments)


Habitat selection

Global scale (e.g., RSF)



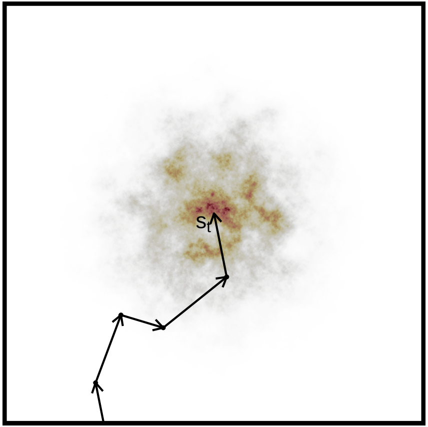
Local scale (e.g., **SSF**)



Step selection functions

Model to describe movement decisions at the scale of a step

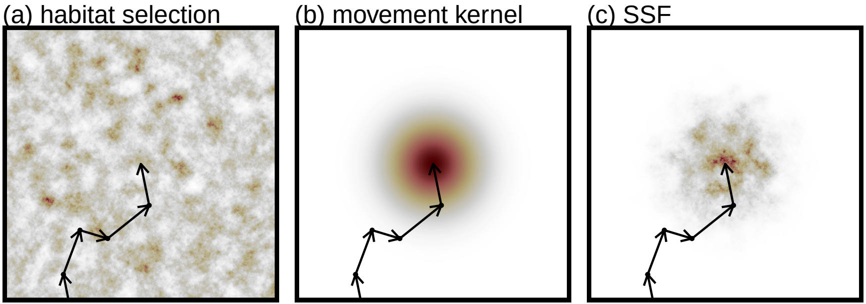
- i.e., likelihood of step from $\boldsymbol{s}_t$ to $\boldsymbol{s}_{t+1}$



SSF formulation

Defines the step likelihood based on movement constraints and habitat preferences

$$w(\mathbf{s}_t, \mathbf{s}_{t+1})\phi(\mathbf{s}_{t+1} \mid \mathbf{s}_{1:t}) \propto p(\mathbf{s}_{t+1}|\mathbf{s}_{1:t})$$



Habitat selection function w

w describes preference/avoidance of environmental features

- Usually main inference of interest

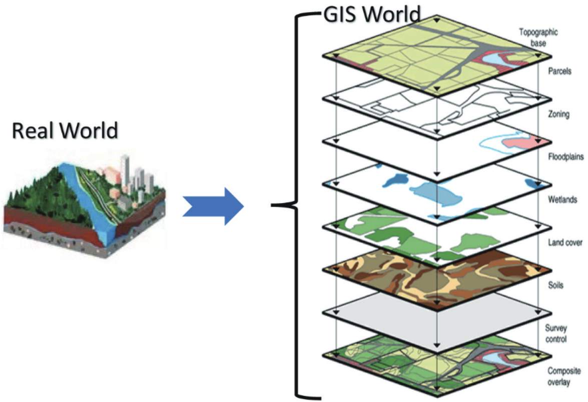


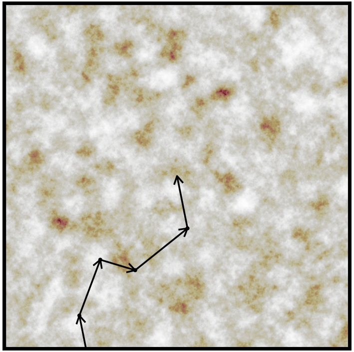
Figure from Acharya & Lee 2019

Habitat selection function w

Common to specify as a log-linear model,

$$w(\boldsymbol{s}_t, \boldsymbol{s}_{t+1}) = \exp\{\beta_h^\top \boldsymbol{c}_h(\boldsymbol{s}_t, \boldsymbol{s}_{t+1})\}$$

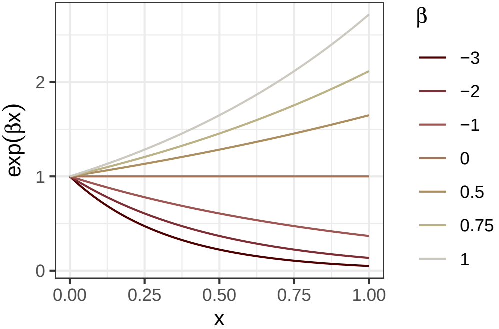
- $\boldsymbol{c}_h(\boldsymbol{s}_t, \boldsymbol{s}_{t+1})$ are the covariates at the step
- β_h are linear effects of covariates
  - positive = preference
  - negative = avoidance



Habitat selection: relative selection strength

$\exp(\beta)$ = RSS

- How many times more likely a step is (*when x increases by 1*)
- Assuming equal availability
- All other covariates held constant
- **e.g.,** $\beta = 1; \quad \exp(\beta) = 2.72$
  $\beta = -1; \ \exp(\beta) = 0.37$

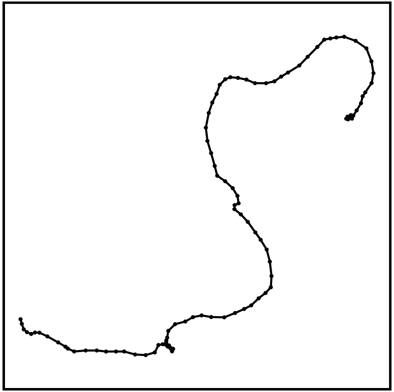


Fieberg, J, J Signer, B Smith, & T Avgar. 2021. A 'how to' guide for interpreting parameters in habitat-selection analyses. *Journal of Animal Ecology*.

Choice of ϕ

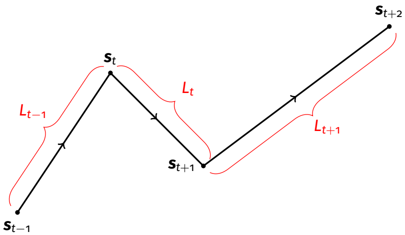
Movement kernel ϕ is a model of (movement-based) availability

Correlated random walk



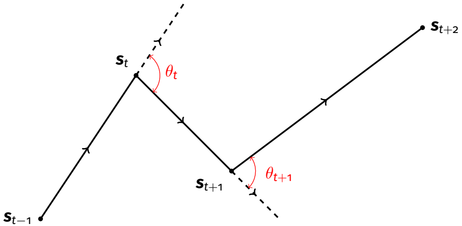
Movement: correlated random walks

Step lengths



Turning angles

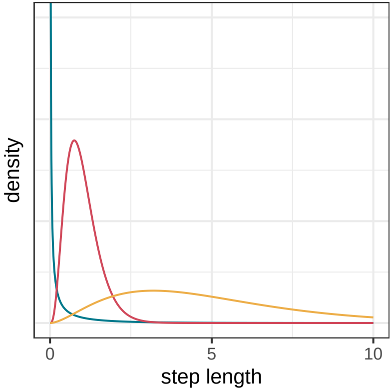
($\theta = 0 =$ directional persistence)



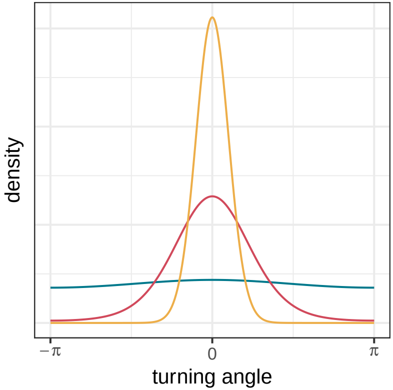
Movement: correlated random walks

Probability distributions for L and θ to capture movement patterns

e.g., gamma (with μ, σ)



e.g., von Mises (with concentration κ)

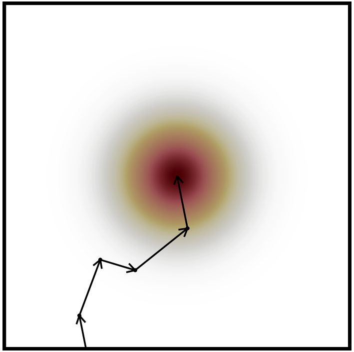


Movement kernel ϕ

Can write a CRW as a log-linear model

$$\phi(\boldsymbol{s}_{t+1} \mid \boldsymbol{s}_{1:t}) = \exp\{\boldsymbol{\beta}_m^\top \boldsymbol{c}_m(\boldsymbol{s}_t, \boldsymbol{s}_{t+1})\}$$

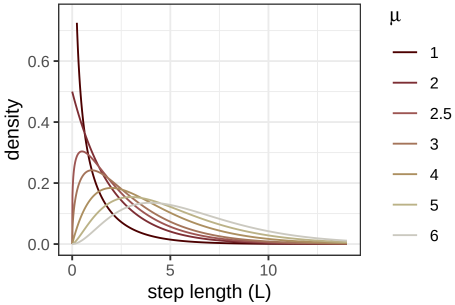
- $\boldsymbol{c}_m(\boldsymbol{s}_t, \boldsymbol{s}_{t+1})$ are movement covariates
- $\boldsymbol{\beta}_m$ parameterise CRW (distributions of step lengths and turning angles)

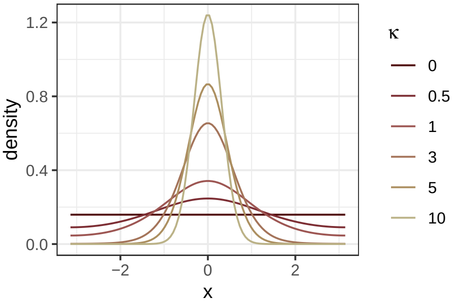


Movement kernel

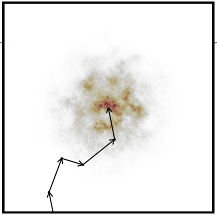
E.g., a CRW with a gamma distribution of L and von Mises of θ:

$$\beta_1 L_t + \beta_2 \log(L_t) + \beta_3 \cos(\theta)$$





SSF formulation

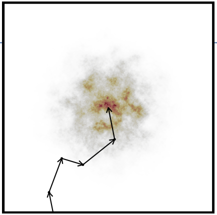


Then we can write the SSF as,

$$p(\boldsymbol{s}_{t+1} \mid \boldsymbol{s}_{1:t}) \propto \exp\{\boldsymbol{\beta}^\top \boldsymbol{c}(\boldsymbol{s}_t, \boldsymbol{s}_{t+1})\}$$

- $\boldsymbol{\beta} = \{\boldsymbol{\beta}_h, \boldsymbol{\beta}_m\}$ (i.e., habitat/movement coefficients)
- $\boldsymbol{c}(\boldsymbol{s}_t, \boldsymbol{s}_{t+1}) = \{\boldsymbol{c}_h(\boldsymbol{s}_t, \boldsymbol{s}_{t+1}), \boldsymbol{c}_m(\boldsymbol{s}_t, \boldsymbol{s}_{t+1})\}$ (i.e., habitat/movement covariates)

SSF formulation



Then we can write the SSF as,

$$p(\boldsymbol{s}_{t+1} \mid \boldsymbol{s}_{1:t}) \propto \exp\{\boldsymbol{\beta}^\top \boldsymbol{c}(\boldsymbol{s}_t, \boldsymbol{s}_{t+1})\}$$

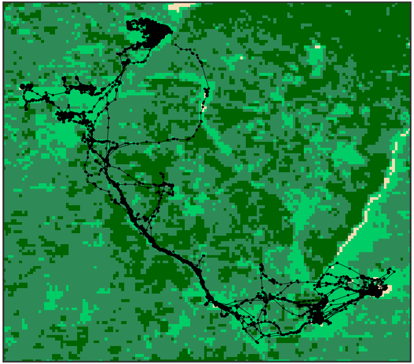
- $\boldsymbol{\beta} = \{\boldsymbol{\beta}_h, \boldsymbol{\beta}_m\}$ (i.e., habitat/movement coefficients)
- $\boldsymbol{c}(\boldsymbol{s}_t, \boldsymbol{s}_{t+1}) = \{\boldsymbol{c}_h(\boldsymbol{s}_t, \boldsymbol{s}_{t+1}), \boldsymbol{c}_m(\boldsymbol{s}_t, \boldsymbol{s}_{t+1})\}$ (i.e., habitat/movement covariates)

example linear predictor with habitat covariate x:

$$\beta_1 L_t + \beta_2 \log(L_t) + \beta_3 \cos(\theta_t) + \beta_4 x_t$$

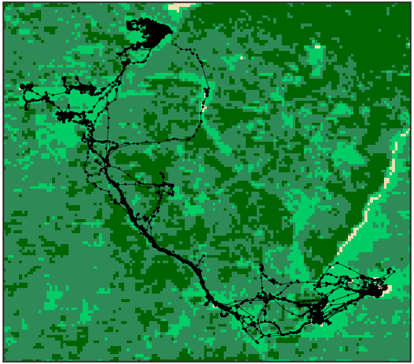
SSF formulation

CRW biased by habitat features



SSF formulation

CRW biased by habitat features



Main assumptions:

1. Discrete-time model (scale-dependence)
2. Locations are without error
3. Movement: distributions from the exponential family*
4. Habitat selection: exponential function*

SSF implementation

Likelihood

Want to find $\hat{\beta}$

- The full track likelihood is the product of all step likelihoods,

$$\mathcal{L}(\beta) = \prod_{t=1}^{T} p(\mathbf{s}_{t+1} | \mathbf{s}_{1:t})$$

Likelihood

Want to find $\hat{\beta}$

- The full track likelihood is the product of all step likelihoods,

$$\mathcal{L}(\beta) = \prod_{t=1}^{T} p(\boldsymbol{s}_{t+1} | \boldsymbol{s}_{1:t})$$

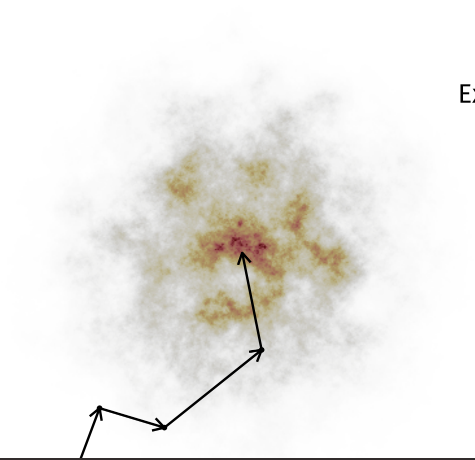
- Each step likelihood is a probability density function with respect to $\boldsymbol{s}_{t+1}$,

$$p(\boldsymbol{s}_{t+1} | \boldsymbol{s}_{1:t}) = \frac{\exp\{\beta^\top \boldsymbol{c}(\boldsymbol{s}_t, \boldsymbol{s}_{t+1})\}}{\int_{\boldsymbol{r} \in \Omega} \exp\{\beta^\top \boldsymbol{c}(\boldsymbol{s}_{1:t}, \boldsymbol{r})\} d\boldsymbol{r}}$$

- **Intuition:** comparison between the observed step and every other possibility

Sampling random points

Problem: to calculate the exact likelihood, we need to integrate this complex function

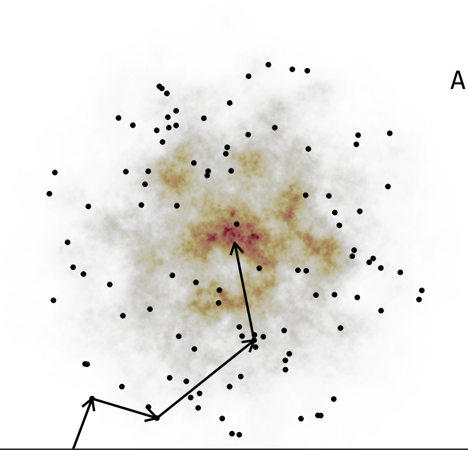


Exact likelihood:

$$p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_{1:t}) = \frac{\exp\{\beta^\top \boldsymbol{c}(\boldsymbol{s}_t, \boldsymbol{s}_{t+1})\}}{\int_{\boldsymbol{r}\in\Omega} \exp\{\beta^\top \boldsymbol{c}(\boldsymbol{s}_{1:t}, \boldsymbol{r})\} d\boldsymbol{r}}$$

Sampling random points

Solution: sample random points as an approximation

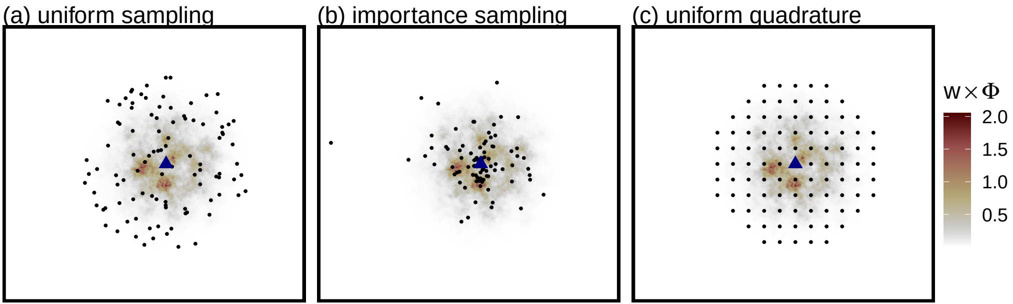


Approximate likelihood:

$$\tilde{p}(\mathbf{s}_{t+1}|\mathbf{s}_{1:t}) = \frac{\exp\{\beta^\top \mathbf{c}(\mathbf{s}_t, \mathbf{s}_{t+1})\}}{\sum_{i=0}^{N} \exp\{\beta^\top \mathbf{c}(\mathbf{s}_{1:t}, \mathbf{r}_{it})\}}$$

Sampling random points

A note on "availability": ϕ defines availability, *not* the random points



Michelot T, NJ Klappstein, JR Potts, & J Fieberg. 2024. Understanding SSA through numerical integration. *Methods in Ecology and Evolution*.

Data format

```
head(data, 10)

##           ID stratum obs       step       angle    habitat hab_type
## 1  animal1       1   1 10.8822255  2.90224116 -4.0714827   forest
## 2  animal1       1   0  5.4848448 -2.69155307  2.8945138   forest
## 3  animal1       1   0  4.3370893  2.93253046 -2.7174871    grass
## 4  animal1       1   0  1.9944188 -1.82062400  1.1764761   forest
## 5  animal1       1   0  2.9415715 -2.41304345  0.2840002    grass
## 6  animal1       2   1  1.8732339  0.11101795  1.0777558   forest
## 7  animal1       2   0  2.7423604  2.89445643  2.5447914    grass
## 8  animal1       2   0  0.8913317  0.07167756 -0.7991826   forest
## 9  animal1       2   0  2.9881229 -0.04271794 -0.6219923     bush
## 10 animal1       2   0  8.5755296 -2.97343825 -4.1804313     bush
```

Implementation as a Cox PH model

At each time t, each location is matched with N random points

$$\Pr(y_{it} = 1 \mid \mathbf{X}_t) = \frac{\exp(\beta^\top \mathbf{x}_{it})}{\sum_{n=0}^{N} \exp(\beta^\top \mathbf{x}_{nt})}$$

where

- $y_{it} = \begin{cases} 1 & \text{if the } i\text{-th point in stratum } t \text{ is the observed location,} \\ 0 & \text{otherwise} \end{cases}$
- $\boldsymbol{X}_t$ is the design matrix for the fixed effects β

```r
fit <- gam(cbind(dummy, stratum) ~ step + angle + habitat,
           data = data,
           family = cox.ph,
           weights = obs)
```

SSFs with smooth effects

SSFs with smooth effects

An SSF with smooth/random effects takes the form,

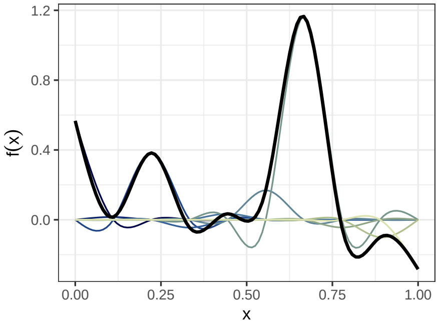
$$\Pr(y_{it} = 1 \mid \mathbf{X}_t, \mathbf{Z}_t) = \frac{\exp(\boldsymbol{\beta}^\top \mathbf{x}_{it} + \boldsymbol{\gamma}^\top \mathbf{z}_{it})}{\sum_{n=0}^{N} \exp(\boldsymbol{\beta}^\top \mathbf{x}_{nt} + \boldsymbol{\gamma}^\top \mathbf{z}_{nt})}, \quad \boldsymbol{\gamma} \sim MVN\left(\mathbf{0}, \frac{1}{\lambda}\mathbf{S}^-\right)$$

where

- $\mathbf{X}_t$ is the design matrix for the fixed effects $\boldsymbol{\beta}$
- $\mathbf{Z}_t$ is the design matrix for the (normally distributed) **random effects** $\boldsymbol{\gamma}$
- $\mathbf{S}$ is a penalty matrix and λ is a smoothing parameter

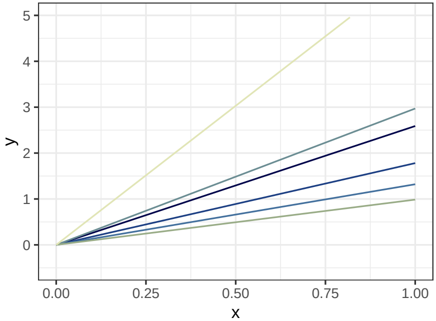
Choice of penalty and λ interpretation

Penalised splines:



- **S** is specific to basis functions
- λ = wiggliness of the function

Random slopes:



- **S** is identity matrix
- λ = inverse of variance

SSFs with non-linear and random effects

The SSF is now a GAM-like model and can be implemented in `mgcv`

- software for implementing GAMs
- very flexible, wide range of options

```
fit <- gam(cbind(dummy, stratum) ~ s(step) + s(angle) + s(habitat),
           data = data,
           family = cox.ph,
           weights = obs)
```

Pedersen et al. (2019). Hierarchical generalized additive models in ecology: an introduction with mgcv. PeerJ.

SSFs with GAMs

1. Smooth covariate effects
   - Non-linear patterns of habitat selection
   - Non-parametric movement distributions
2. Inter-individual variability
   - Random slopes
   - Hierarchical smooths
3. Accounting for unexplained spatial pattern
   - Spatial smoothing
4. Smoothly varying coefficient models

Methods in Ecology and Evolution

RESEARCH ARTICLE

Step selection functions with non-linear and random effects

Natasha J. Klappstein[1] | Théo Michelot[1] | John Fieberg[2] | Eric J. Pedersen[3] | Joanna Mills Flemming[1]

[1]Department of Mathematics & Statistics, Dalhousie University, Halifax, Nova Scotia, Canada
[2]Department of Fisheries, Wildlife, and Conservation Biology, University of Minnesota, Minneapolis, Minnesota, USA
[3]Department of Biology, Concordia University, Montreal, Quebec, Canada

Correspondence
Natasha J. Klappstein
Email: nklappstein@dal.ca

Funding information
National Sciences and Engineering Research Council of Canada; National Aeronautics and Space Administration, Grant/Award Number: 80NSSC21K1182; Minnesota Agricultural Experimental Station

Handling Editor: Theoni Photopoulou

Abstract

1. Step selection functions (SSFs) are used to jointly describe animal movement patterns and habitat preferences. Recent work has extended this framework to model inter-individual differences, account for unexplained structure in animals' space use and capture temporally varying patterns of movement and habitat selection.
2. In this paper, we formulate SSFs with penalised smooths (similar to generalised additive models) to unify new and existing extensions, and conveniently implement the models in the popular, open-source `mgcv` R package.
3. We explore non-linear patterns of movement and habitat selection, and use the equivalence between penalised smoothing splines and random effects to implement individual-level and spatial random effects. This framework can also be used to fit varying-coefficient models to account for temporally or spatially heterogeneous patterns of selection (e.g. resulting from behavioural variation), or any other non-linear interactions between drivers of the animal's movement decisions.
4. We provide the necessary technical details to understand several key special cases of smooths and their implementation in `mgcv`, showcase the ecological relevance using two illustrative examples and provide R code to facilitate the adoption of these methods. This paper offers a broad overview of how smooth effects can be applied to increase the flexibility and biological realism of SSFs.

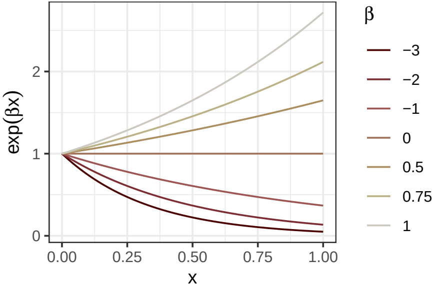
KEYWORDS
animal movement, generalised additive models, habitat selection, integrated step selection analysis, `mgcv`, penalised splines, random effects, step selection functions

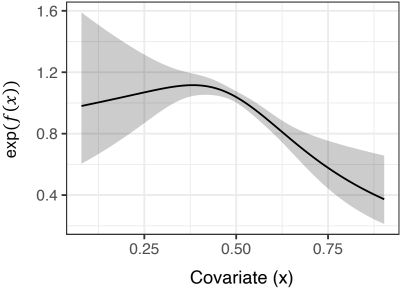
Smooth covariate effects: **habitat**

Allow non-linear patterns of habitat selection

- **e.g.,** include smooth effect of x:

$$\beta_1 L_t + \beta_2 \log(L_t) + \beta_3 \cos(\theta_t) + \beta_4 x \quad \rightarrow \quad \beta_1 L_t + \beta_2 \log(L_t) + \beta_3 \cos(\theta_t) + f(x)$$

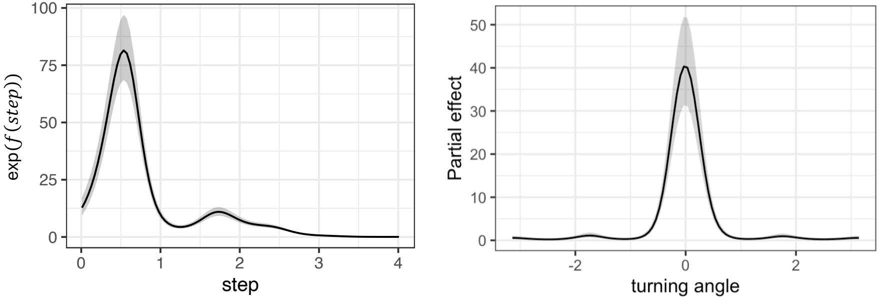




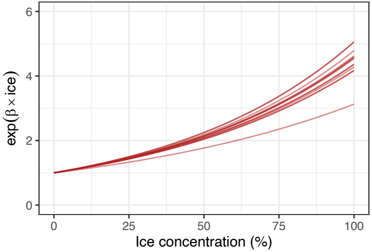
Smooth covariate effects: **movement**

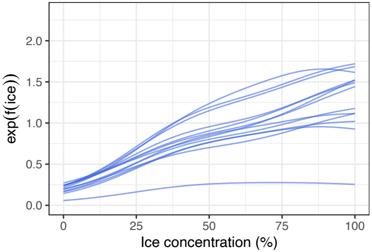
Can also be for more flexible movement kernels with non-parametric distributions:

$$\beta_1 L_t + \beta_2 \log(L_t) + \beta_3 \cos(\theta_t) + \beta_4 x \quad \rightarrow \quad f(L_t) + f(\theta) + \beta_4 x$$



Accounting for inter-individual variability



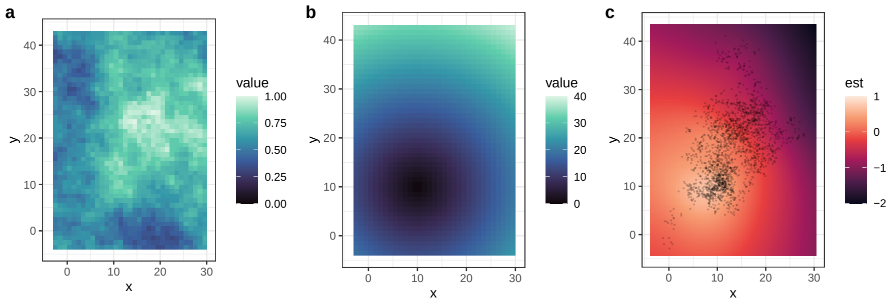


Accounting for spatial heterogeneity

Spatial random effects can be used to account for unexplained spatial pattern

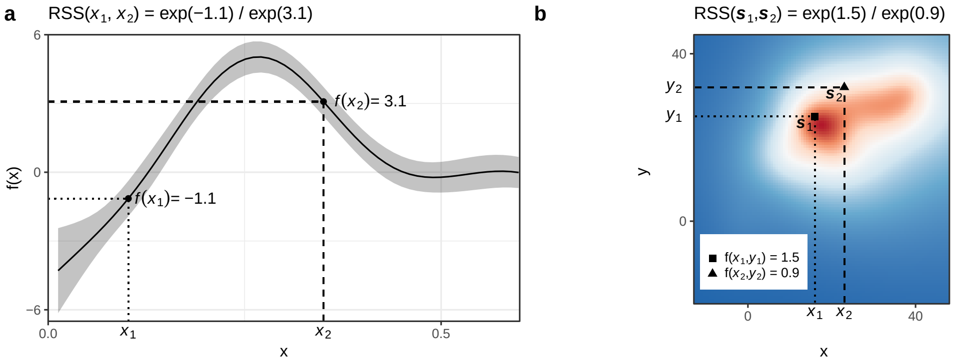
- **e.g.,** accounting for an unknown centre of attraction

$$\beta_1 L_t + \beta_2 \log(L_t) + \beta_3 \cos(\theta_t) + \beta_4 \mathrm{cov}_t + f(x_t, y_t)$$



Smooths interpretation

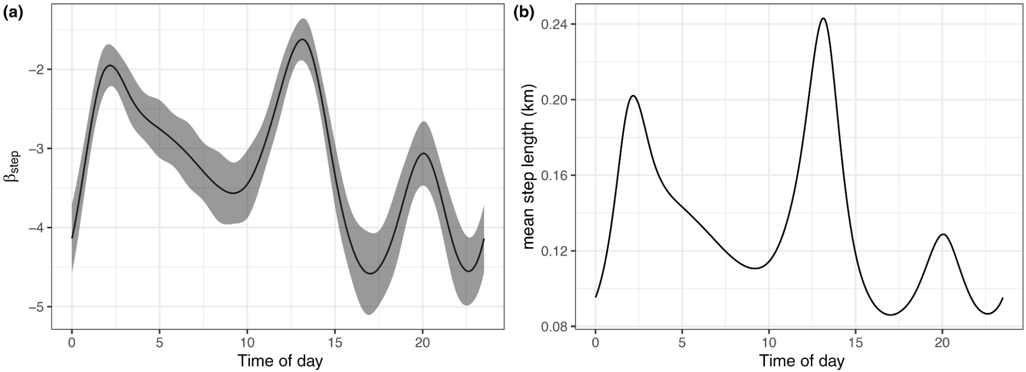
Can be interpreted as relative selection strength



Interactions: **varying-coefficient models**

Allow linear effect of covariate to smoothly vary with another covariate

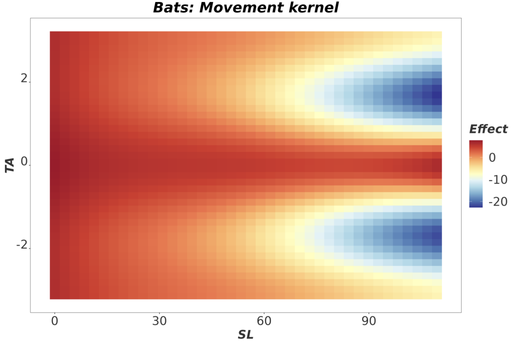
- **e.g.,** to account for spatial or temporal variability in selection/movement



Interactions: **tensor products**

Smooth interaction between two covariates (with different units)

- **e.g.,** to account for dependence in movement variables



Arce Guillen R., et al. 2024. Flexible movement kernel estimation in habitat selection analyses with generalized additive models. *bioRxiv*.

Workflow

1. Prepare movement data (using `amt`)
   - Add random points and calculate movement variables
   - Interpolate spatial covariates
2. Fit model in `mgcv`
   - Implementation as a Cox PH model
3. Interpret model outputs using `gratia` and `mgcv`
   - Smooths visualisation
   - Derive relative selection strength and movement distributions
   - Note: lack of model checking for SSFs

Further documentation and resources

- Github repository
  - `github.com/NJKlappstein/smoothSSF`

- Klappstein, NJ., et al. 2024. Step selection functions with non-linear and random effects. *Methods in Ecology and Evolution*.

SSF TUTORIAL WITH MGCV