Using telemetry data for estimating animal resource selection: a comparison of different statistical approaches

Ana Couto[1], Thomas Cornulier[1], David Miller[1,2], Katherine Whyte[1], Adam Butler[1], Esther Jones[1]

ana.couto@bioss.ac.uk
offshore.renewables@bioss.ac.uk

Telemetry data are often highly autocorrelated in space and time, thus challenging the assumption of independence between observations and subsequent inference using resource selection frameworks. Failure to correctly account for autocorrelation in tracking data can lead to overly narrow confidence intervals, bias in parameter estimates and model selection.
Here we compare different common and newer approaches for mitigating the statistical problems posed by autocorrelation, using simulated datasets.

Data simulations

We simulated a binary habitat taking values 0 or 1 in a checkerboard pattern, to ensure homogeneous habitat availability, with no spatial bias relative to home range centre.
We then simulated 25 trips with 4 states), including:

- State 1 – outbound from home range centre,
- State 2 – searching behaviour (faster and more directed), associated with habitat = 0
- State 3 – foraging behaviour (slower, more sinuous), associated with habitat = 1
- State 4 – inbound trip, with probability increasing with time since departure.

Stage 2 and 3 were modelled using correlated random walks (CRWs) while stage 1 and 4, biased random walks (BRWs) were used (Michelot et al., 2017). At each time step, the state process is simulated based on a transition probability matrix. Step length and bearing are simulated using a gamma and von Mises distributions, respectively.

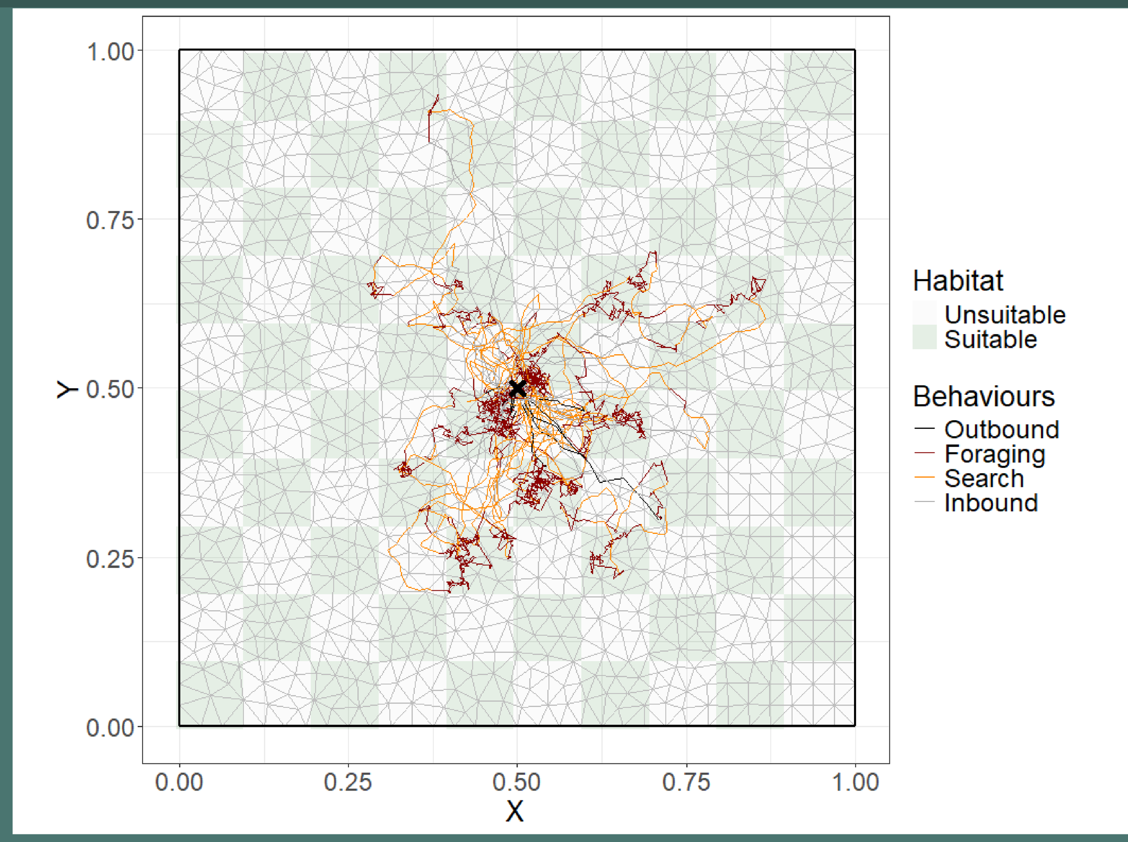

Figure 1 – Simulated habitat suitability and tracking data, with a triangle mesh. The nodes of the triangles correspond to the quadrature points (available points).

Models compared*

Inhomogeneous Poisson Process Regression Models

(Dovers et al., 2023, Matthiopoulos et al., 2023)

$$ln\,\lambda(s_i) = \beta_0 + \beta_1 Habitat(s_i)$$

$$\ell_{\text{IPP}} \approx \sum_{i=1}^{n+Q} y_i \ln \lambda(s_i) - w_i \lambda(s_i)$$

$$y_i = \begin{cases} 1 \text{ if } s_i \text{ is an observation} \\ 0 \text{ if } s_i \text{ is a quadrature point} \end{cases}$$

$$w_i = \begin{cases} 1^{-6} \text{ for } i = 1, \dots, n \\ w_q \text{ for } i = n+1, \dots, n+Q \text{ and } q = 1, \dots, Q \end{cases}$$

w_q = area of each triangle in the mesh
Q = number of quadrature points
n= number of points
s_i= point events found in some spatial domain

IPP assumes independence between points. Data thinning often used to meet independence assumption, resulting in loss of useful information and may lead to bias in parameter estimates.

Log-Gaussian Cox Process Regression Models

(Dovers et al., 2023)

$$ln\,\lambda(s_i) = \beta_0 + \beta_1 Habitat(s_i) + \xi(s)$$

$$\xi(s) \approx \sum_{j=1}^{k} B_j(s) b(j)$$

$\xi(s)$= zero-mean latent Gaussian field

An IPP that includes a Gaussian field ξ(s) to account for unexplained spatial autocorrelation. Not usually designed to account for serial correlation in telemetry data. Data thinning often used to meet independence assumption.

Resource selection functions with autocorrelation-adjusted weights

(Alston et al., 2022)

$$ln\,\lambda(s_i) = \beta_0 + \beta_1 Habitat(s_i)$$

$$\ell_{\text{RSF-W}} \approx \sum_{i=1}^{n} w_i \ln \lambda(s_i) - \int_D \lambda(t)dt$$

$$w_i = w_{AKDE_i} N$$

w_{AKDE_i} = weights optimized for non-parametric autocorrelated kernel density estimation at each sampled location
N = the effective sample size of the autocorrelated Gaussian area estimate

An IPP model that uses weights from non-parametric autocorrelated kernel density estimate (w_{AKDE}) in the log-likelihood, to mitigate pseudo-replication in the dataset without resorting to data thinning.

Estimates comparison

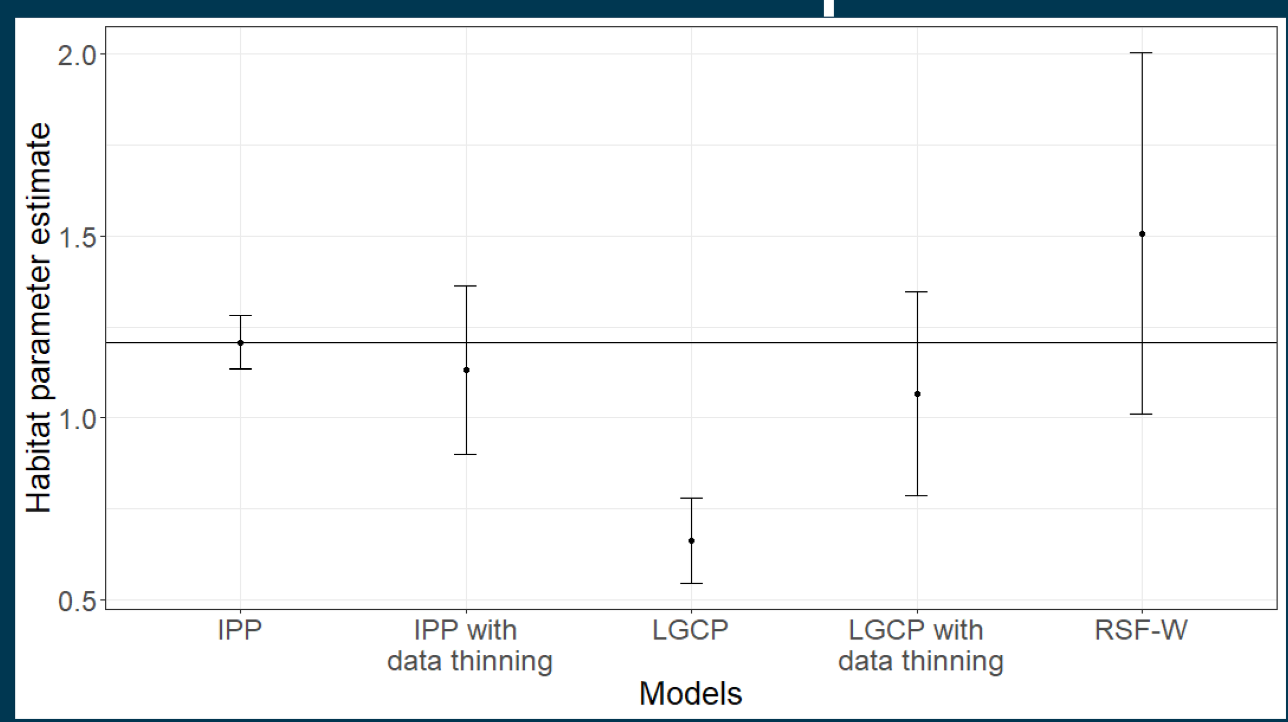

Figure 2 – Estimates for habitat parameter with 95% confidence intervals for the different models. The horizontal line corresponds to the estimate for the IPP, which is assumed to be unbiased

Both IPP and LGCP produce optimistically narrow confidence intervals when using the full dataset, widening as expected with data thinning.
LGCP habitat selection estimates are biased relative to full IPP, likely due to spatial confounding.
RSF-W produces the widest confidence interval, but a much higher estimate than the IPP, indicating bias. A likely cause is the failure of the model to account for varying levels of movement autocorrelation among habitats.

LGCP with Neighbourhood Cross-Validation (NCV)

Model fitting by cross-validation, based on splitting the data by arbitrary "neighbourhoods". May be used to account for serial autocorrelation, by defining neighbourhoods to account for the spatio-temporal structure in the data. Implemented in mgcv with the new method="NCV" option (Wood, 2024) .

Can we use this approach to estimating animal resource selection from telemetry data?
How sensitive to neighbourhood designs?

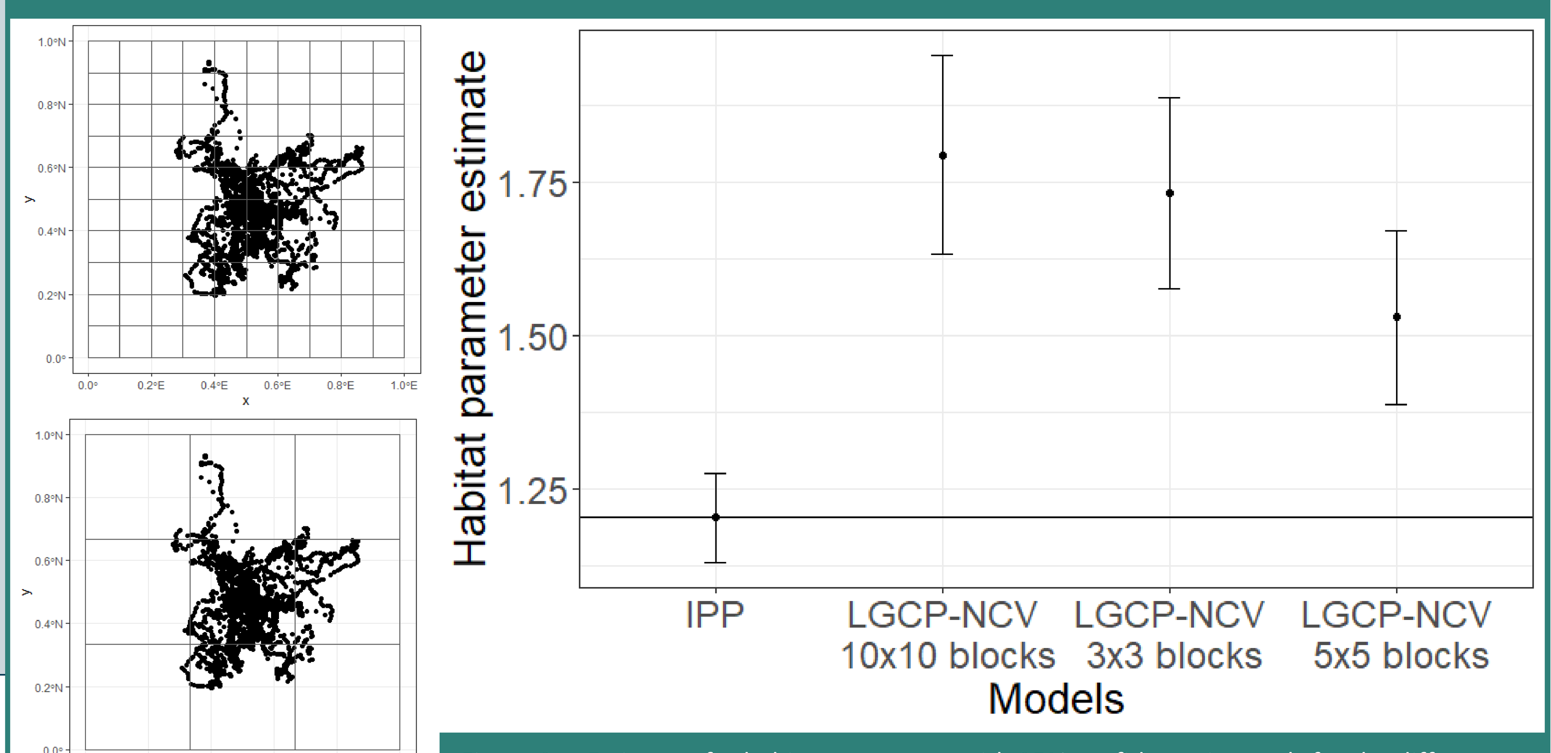

Figure 3 – Estimates for habitat parameter with 95% confidence intervals for the different models. The horizontal line corresponds to the estimate for the IPP, which is assumed to be unbiased.

1 BioSS
2 UK Centre for Ecology & Hydrology

*not an exhaustive list of methods

Alston, Jesse M., et al. "Mitigating pseudoreplication and bias in resource selection functions with autocorrelation-informed weighting." Methods in Ecology and Evolution 14.2 (2023): 643-654.
Dovers, Elliot, Jakub Stoklosa, and David I. Warton. "Fitting log-Gaussian Cox processes using generalized additive model software." The American Statistician (2024): 1-16.
Matthiopoulos, Jason; Fieberg, John; Aarts, Geert. (2023). Species-Habitat Associations: Spatial data, predictive models, and ecological insights, 2nd edition. University of Minnesota Libraries Publishing. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/217469.
Michelot, Théo, et al. "Estimation and simulation of foraging trips in land-based marine predators." *Ecology* 98.7 (2017): 1932-1944.
Wood (2024) On Neighbourhood Cross Validation arXiv:2404.16490v1 [stat.ME] 25 Apr 2024