HZDR HELMHOLTZ-ZENTRUM DRESDEN-ROSSENDORF
CASUS CENTER FOR ADVANCED SYSTEMS UNDERSTANDING

MARE-Madeira 2025

Continuous-time movement models

Using the 'ctmm' R package

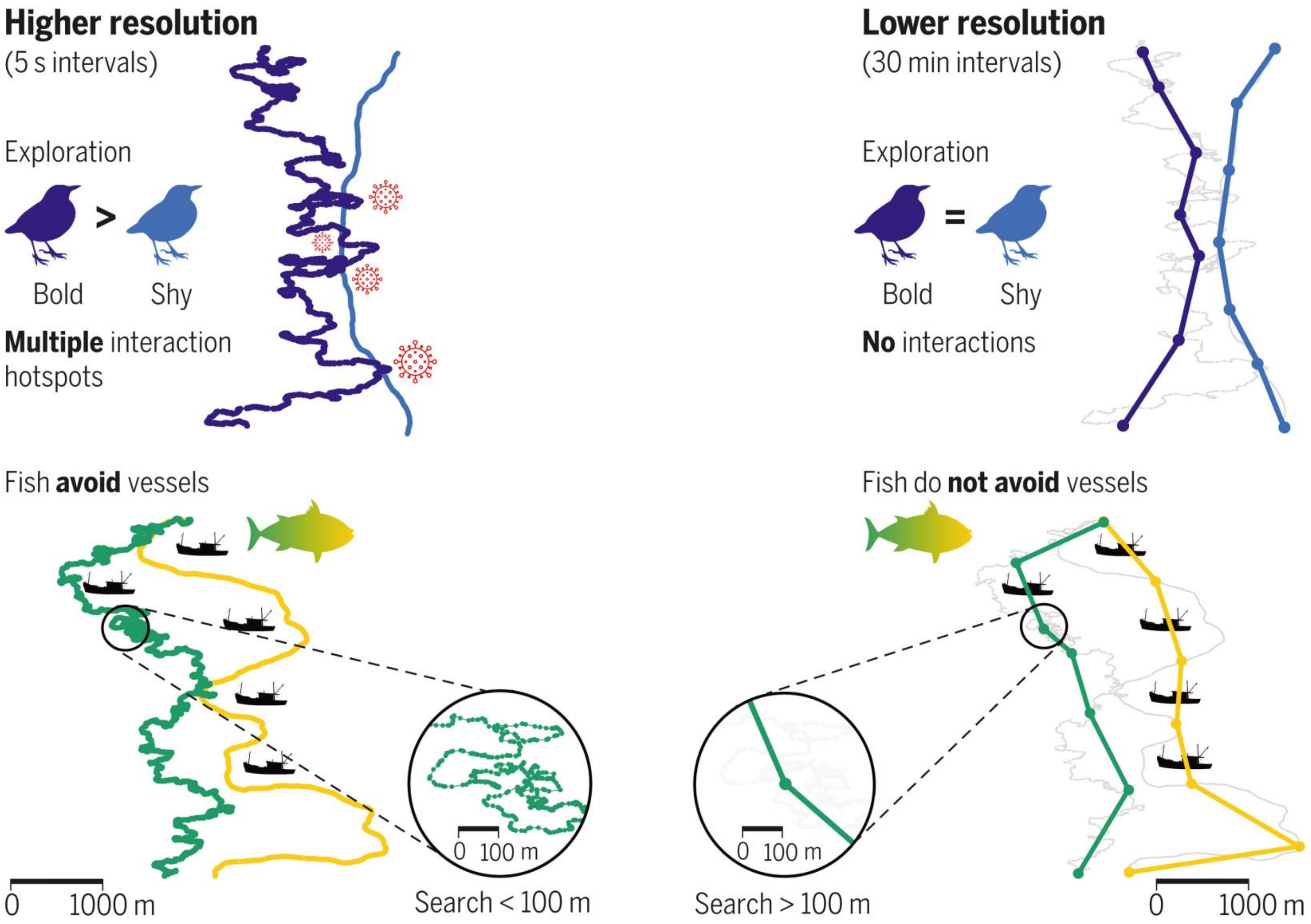
Inês Silva
i.simoes-silva@hzdr.de

TECHNISCHE UNIVERSITÄT DRESDEN
CBG Max Planck Institute of Molecular Cell Biology and Genetics
UFZ HELMHOLTZ Centre for Environmental Research
Uniwersytet Wrocławski
SPONSORED BY THE Federal Ministry of Education and Research
STAATSMINISTERIUM FÜR WISSENSCHAFT KULTUR UND TOURISMUS
Freistaat SACHSEN

Account for autocorrelation	(stochastic process models)
Account for sampling irregularity and mismatch	(continuous-time models)
Account for location error if necessary	(calibrated error models)
Don’t assume a model	(model selection, test assumptions)
Don’t be satisfied with point estimates	(estimate uncertainties)
Reducing bias and error as much as possible	(debiased estimators)
Propagate individual uncertainties into population estimates	(hierarchical models)

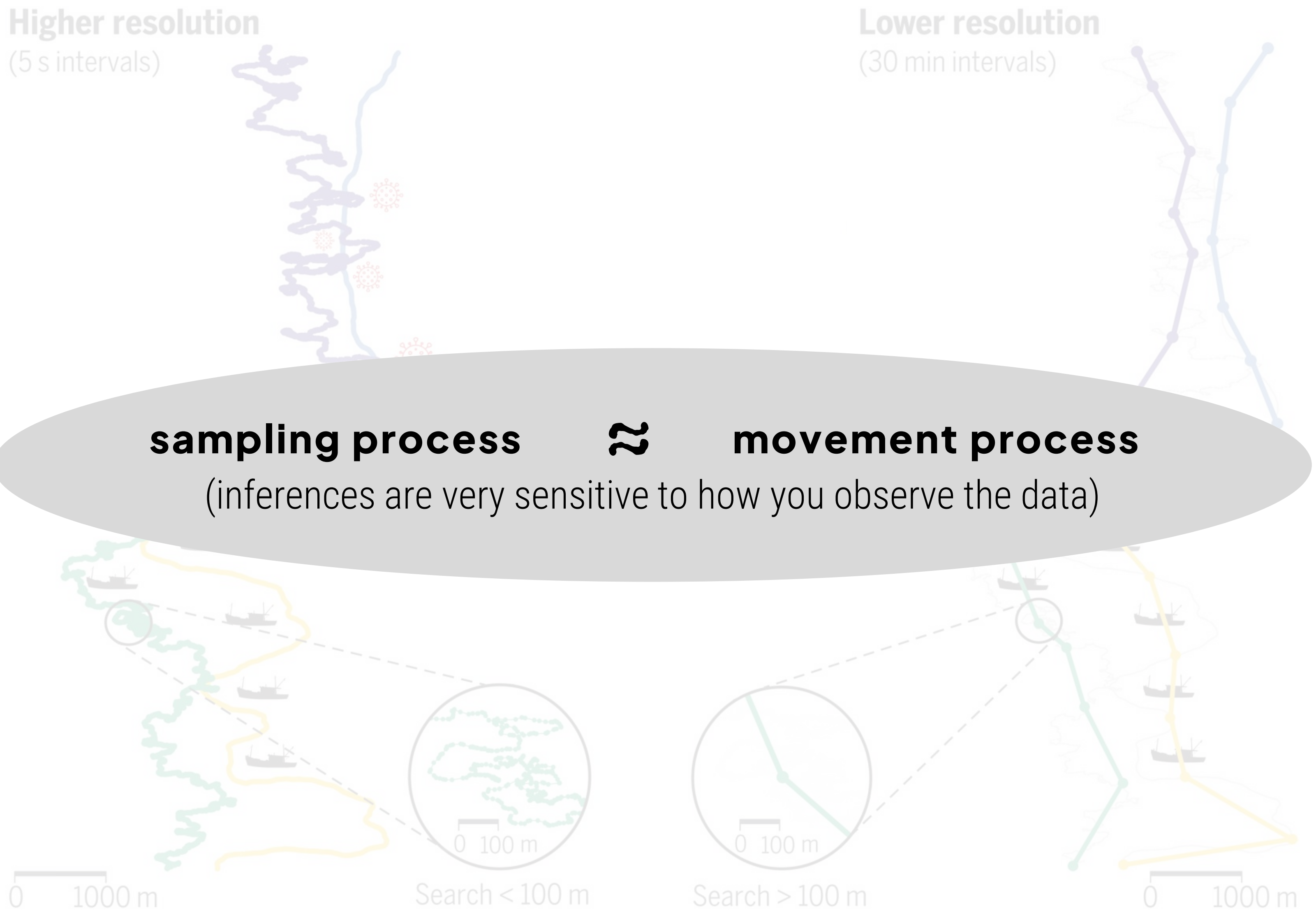
i.simoes-silva@hzdr.de

Nathan *et al.* (2022)



i.simoes-silva@hzdr.de

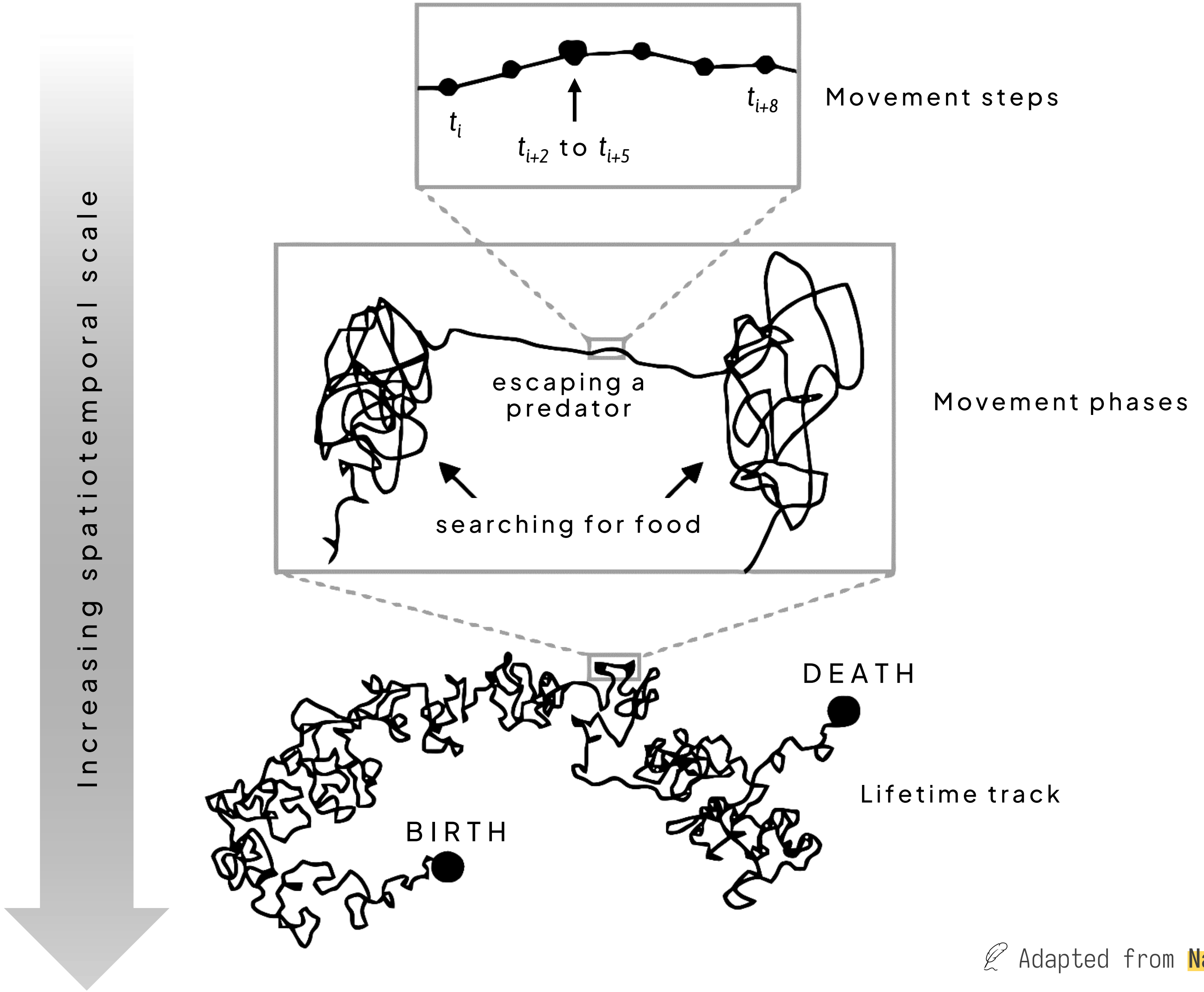
Nathan *et al.* (2022)



sampling process ≈ movement process

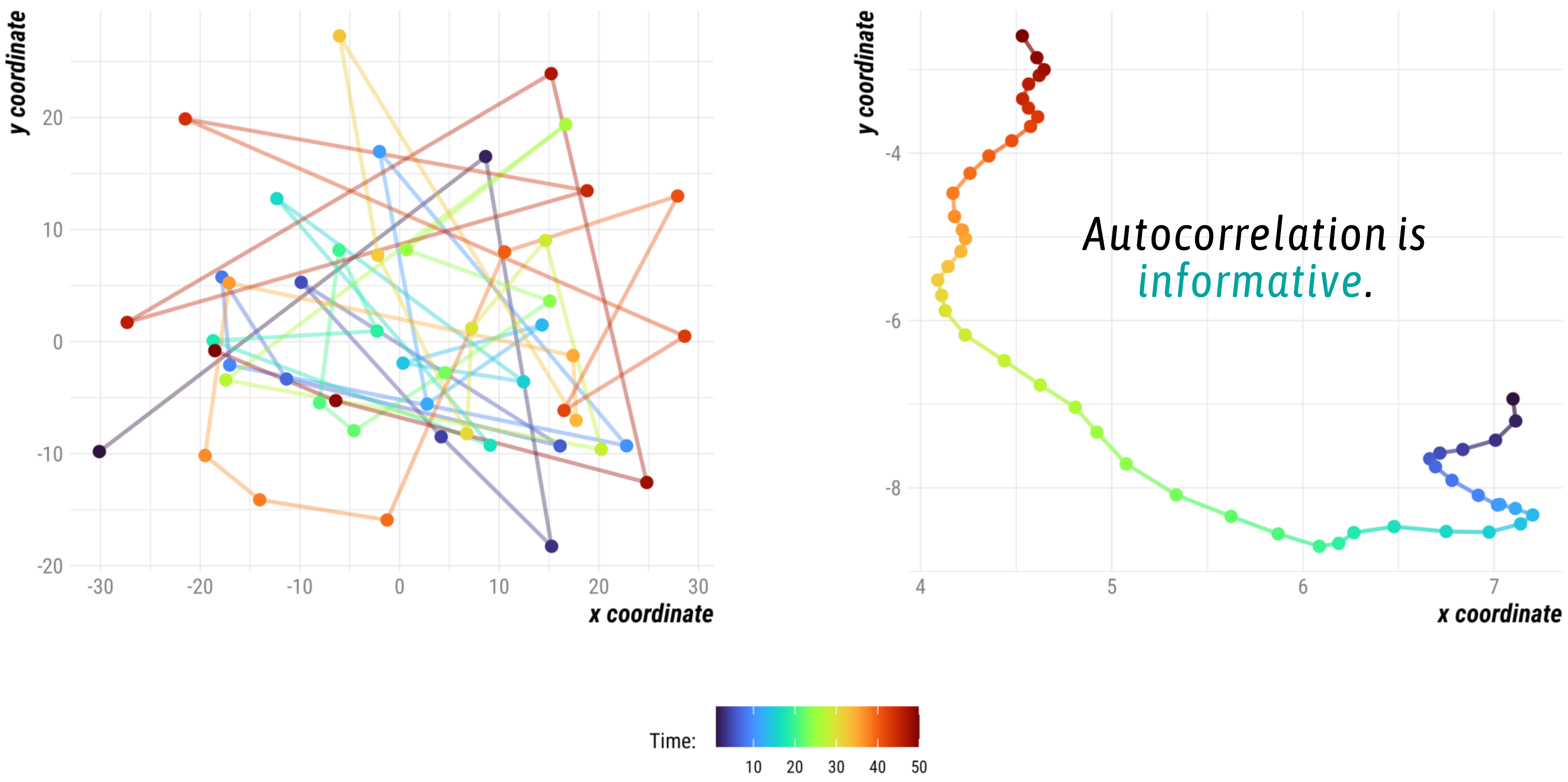
(inferences are very sensitive to how you observe the data)

i.simoes-silva@hzdr.de



Adapted from Nathan *et al.* (2008)

Strong (short- and long-term) ***autocorrelation*** in path



i.simoes-silva@hzdr.de

How can we accurately characterize animal movement?
(*i.e.*, location, velocity and direction over a period of time)

sampling process

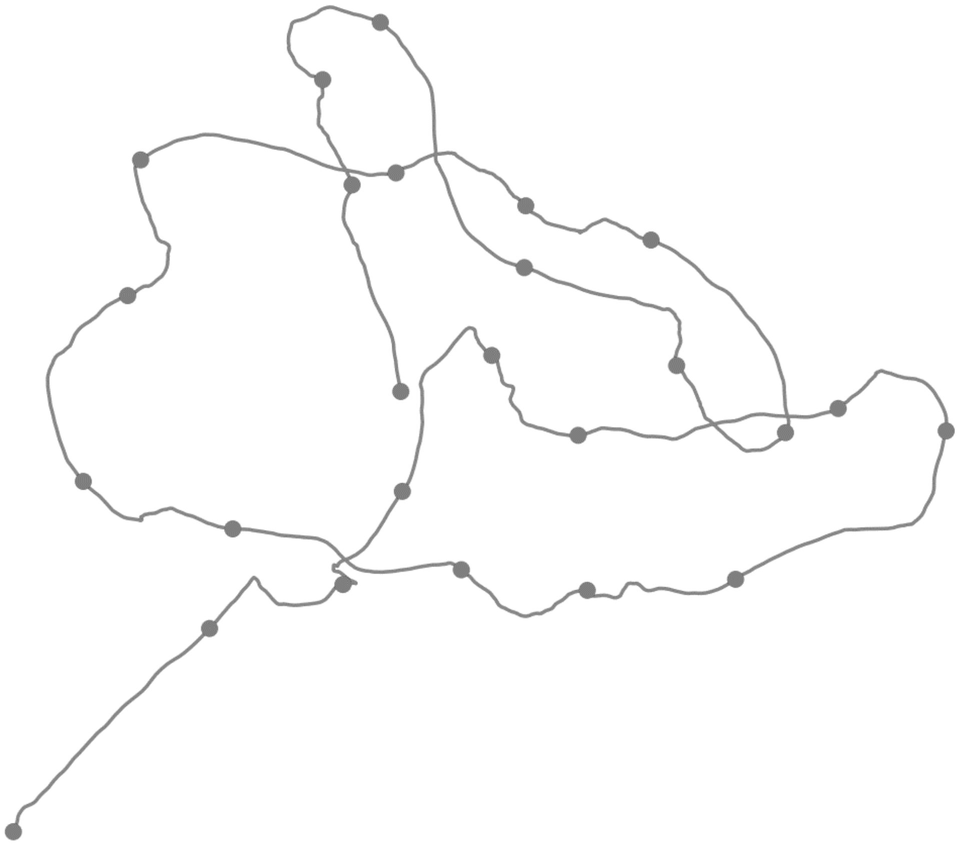
movement process

Only way around all these problems is a method that separates sampling and movement processes + uses all time lags in the data.

i.simoes-silva@hzdr.de

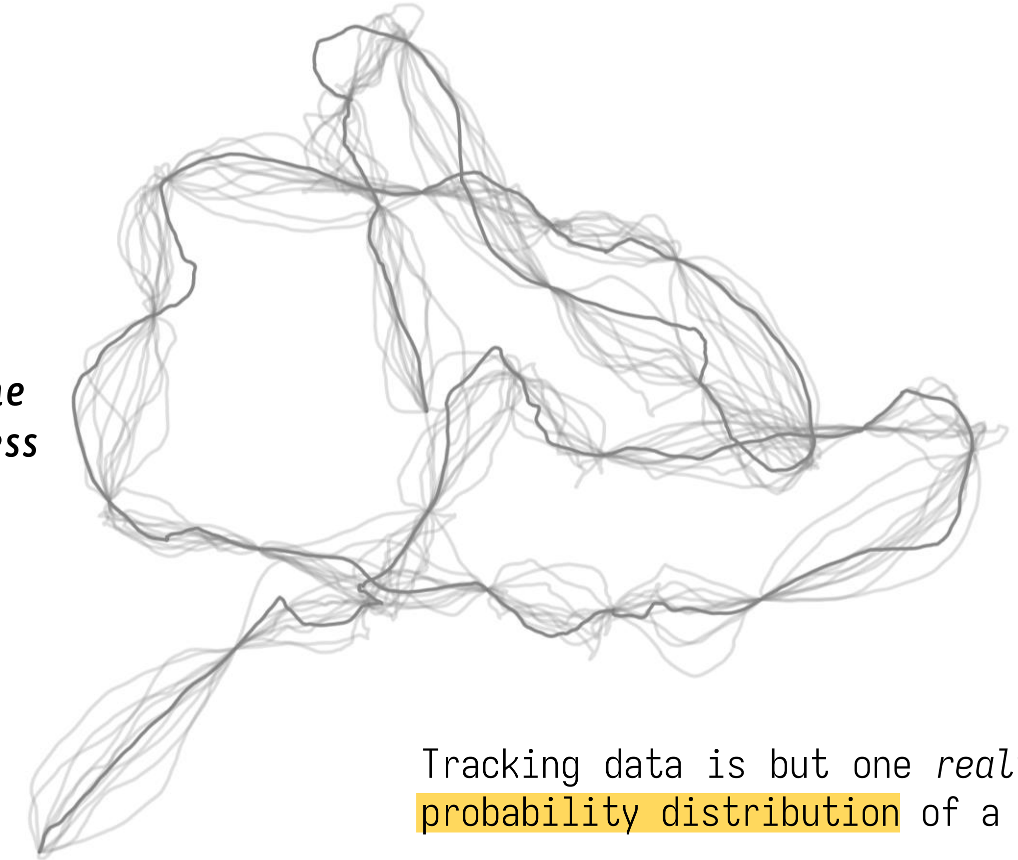
How can we accurately characterize animal movement?

(*i.e.*, location, velocity and direction over a period of time)



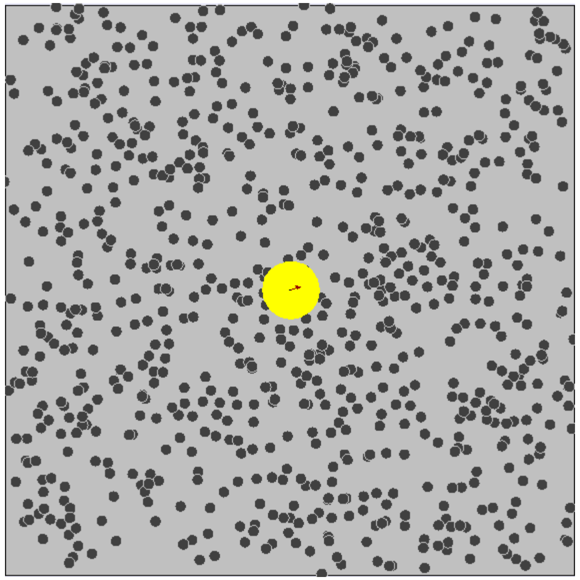
How can we accurately characterize animal movement?
(*i.e.*, location, velocity and direction over a period of time)

Continuous-time stochastic process



Tracking data is but one *realization* of the probability distribution of a movement path.

i.simoes-silva@hzdr.de



Brownian motion

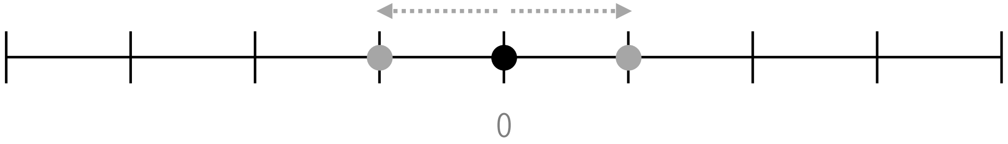
Particles move **randomly** in all directions, *i.e.*, change in position is random and continuous over time.

A ***stochastic process*** describes the evolution of some phenomenon over time. "***Stochastic***" is generally interchangeable with "***random***", indicating an element of chance or unpredictability.

Does not imply that we have no information about the dynamics; a stochastic process provides structure with some ***randomness***.

Markov process

Type of stochastic process where the future state of the system depends ***only on its current state***, not on the sequence of events that preceded it.



Focus on fundamental statistics (moments or cumulants) of the process:

mean $\mu(t)$:

$$\boldsymbol{\mu}(t) = \langle \boldsymbol{r}(t) \rangle \qquad \boldsymbol{r}(t) \equiv ((\boldsymbol{x}(t), \boldsymbol{y}(t))$$

and the autocorrelation function $\sigma(t,t')$:

$$\boldsymbol{\sigma}(t,t') = \langle [\boldsymbol{r}(t) - \boldsymbol{\mu}(t)] \, [\boldsymbol{r}(t') - \boldsymbol{\mu}(t')]^T \rangle$$

where:
t denotes the time index,
r = (x, y) denotes the location, and $\langle \cdots \rangle$ denotes the average over realizations of the process.

Different movement models predict different functional forms for the first two cumulants.

i.simoes-silva@hzdr.de

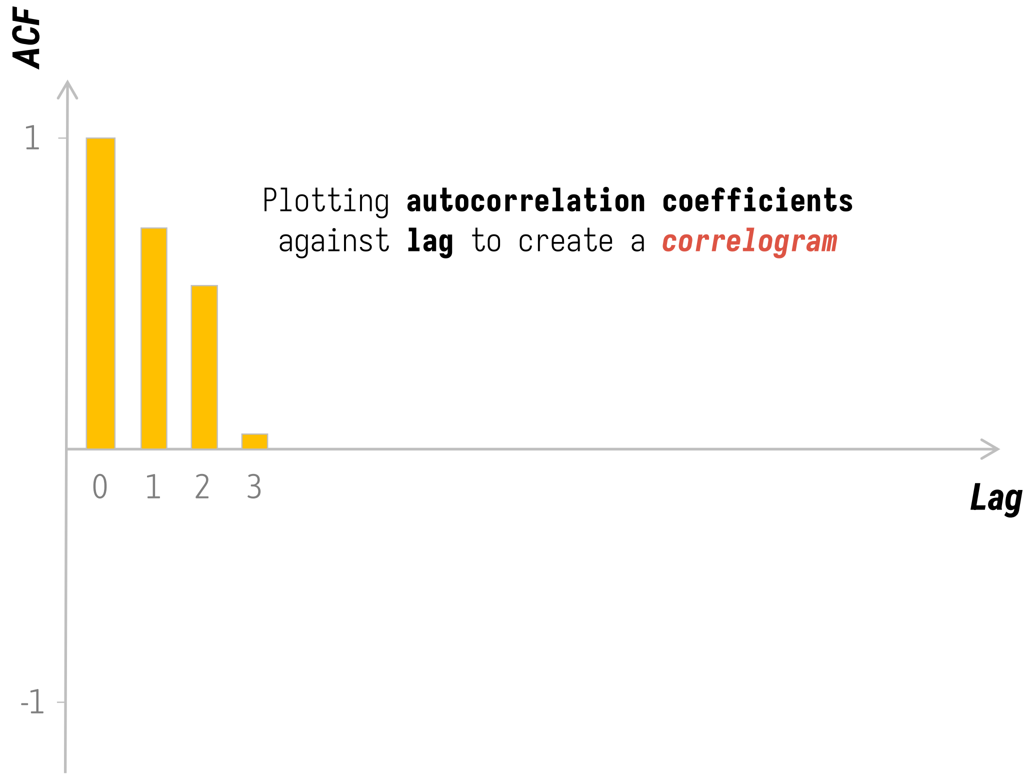
How can we measure and visualize autocorrelation?

The ***autocorrelation function*** **(ACF)** is a valuable tool for investigating properties of an empirical time series. We typically visualize it through the ***correlogram***.

$ACF_3 = var_{t+3} / var_t$

Lag = 3

Represents a shift in time.



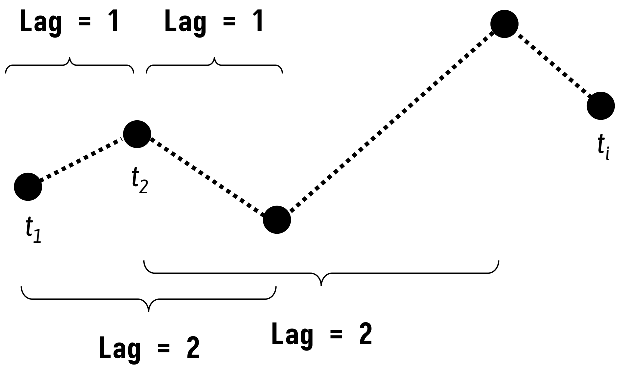
i.simoes-silva@hzdr.de

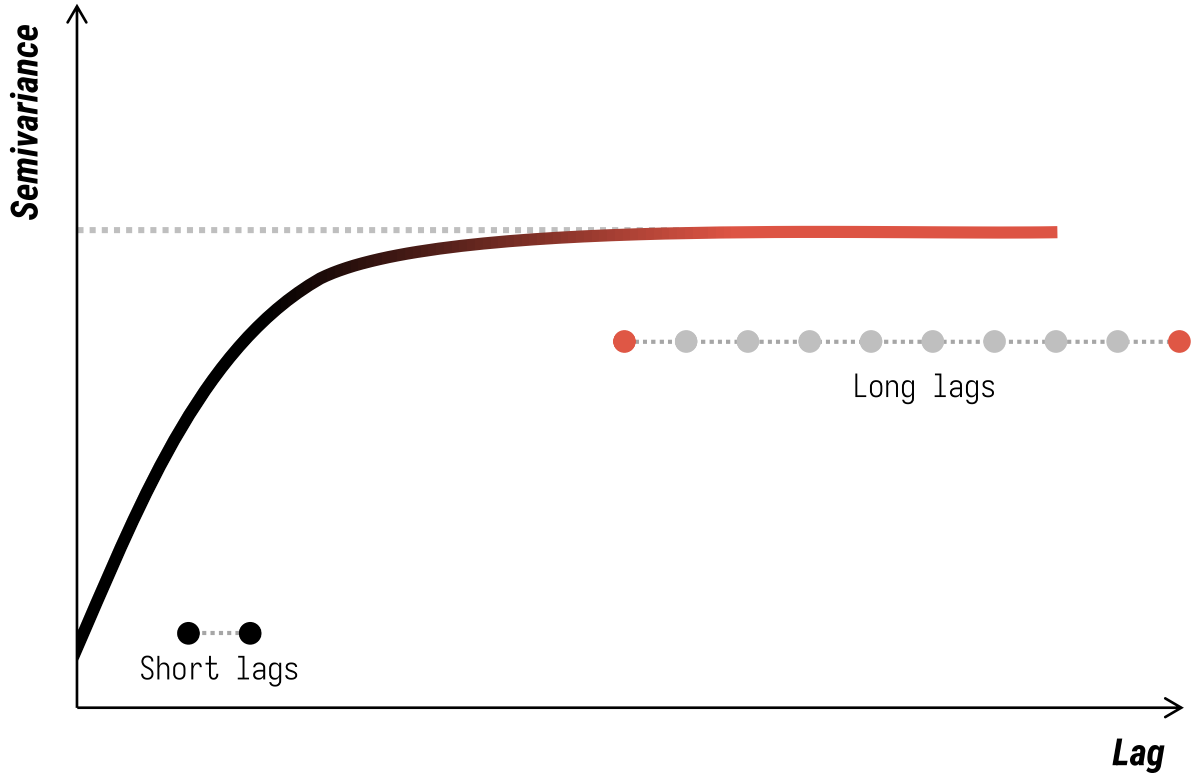
The **SVF** contains similar information to the **ACF**, but works better for ***autocorrelated movement data*** due to its superior statistical properties:

- Can be readily estimated from movement paths via ***variogram***;
- Contains information on the ***mixture of movement modes*** present;
- Uses ***all*** possible timescales (lags) in the data, not just the sampling interval (and is therefore inherently less sensitive to the particular choice of sampling rate).
- Can also be ***averaged over multiple animals*** and ***over time*** when movement behavior is highly nonstationary, further increasing its statistical power.

How can we measure and visualize autocorrelation in tracking data?

Semivariograms (or **variograms**) are used to detect autocorrelation. We plot **lags** on the **x-axis** for all pairs of observations against their **semivariance** (average square distance between any two locations with a given lag) on the **y-axis**.



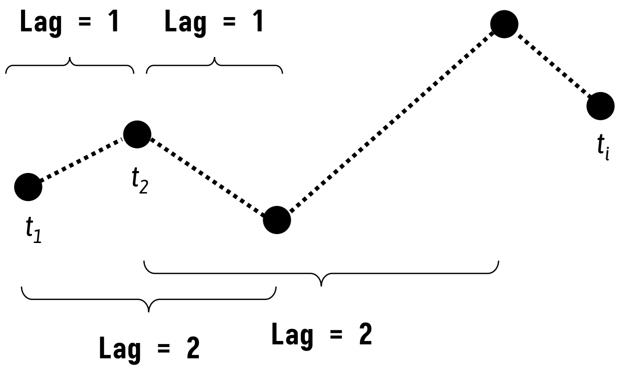


The more **similar** the pairs of locations are per lag, the **lower the semivariance** for that lag.

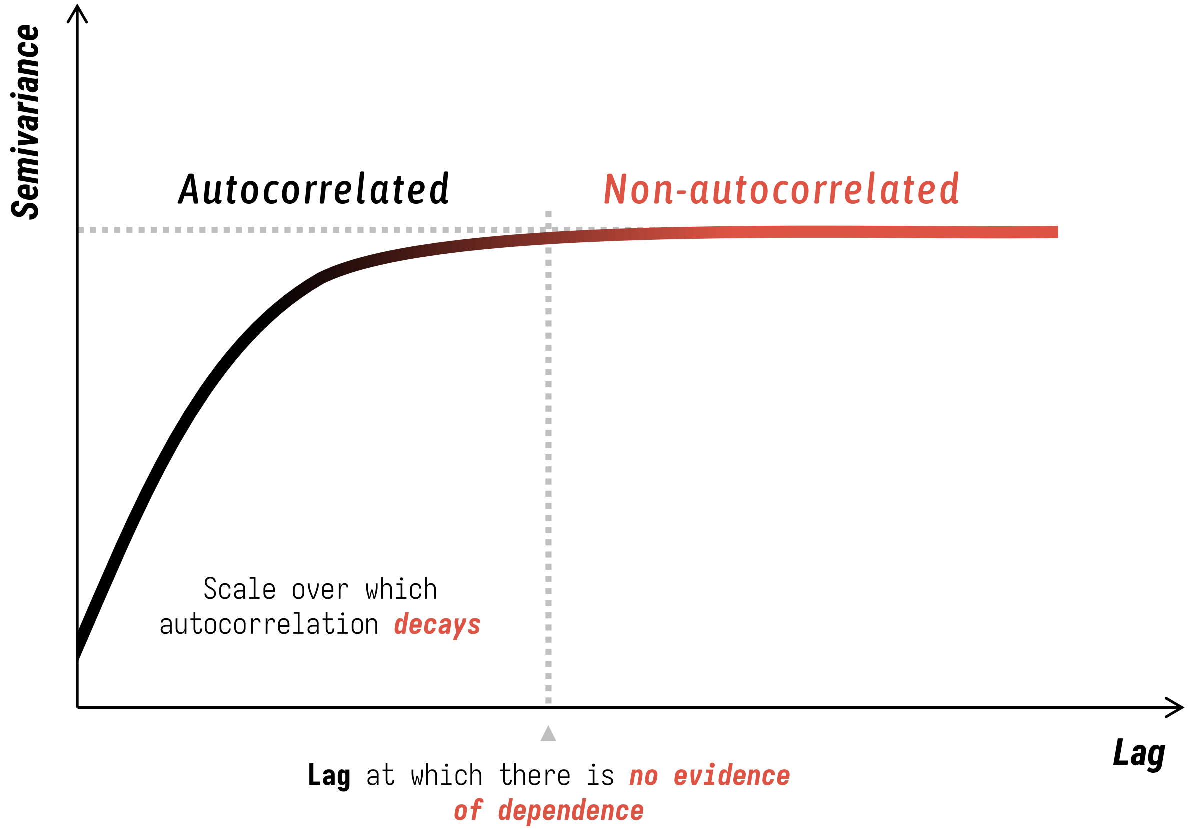
i.simoes-silva@hzdr.de

How can we measure and visualize autocorrelation in tracking data?

Semivariograms (or ***variograms***) are used to detect autocorrelation. We plot **lags** on the **x-axis** for all pairs of observations against their ***semivariance*** (average square distance between any two locations with a given lag) on the **y-axis**.



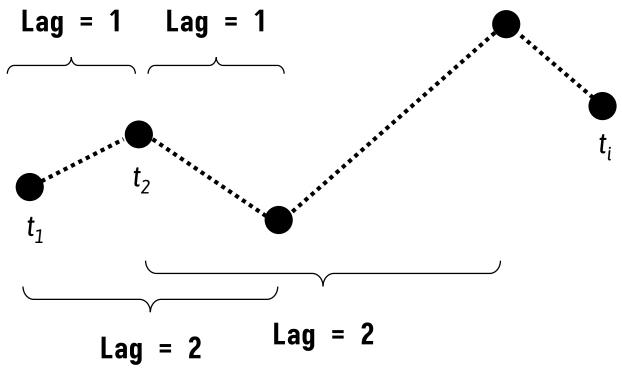
The more **similar** the pairs of locations are per lag, the **lower the semivariance** for that lag.



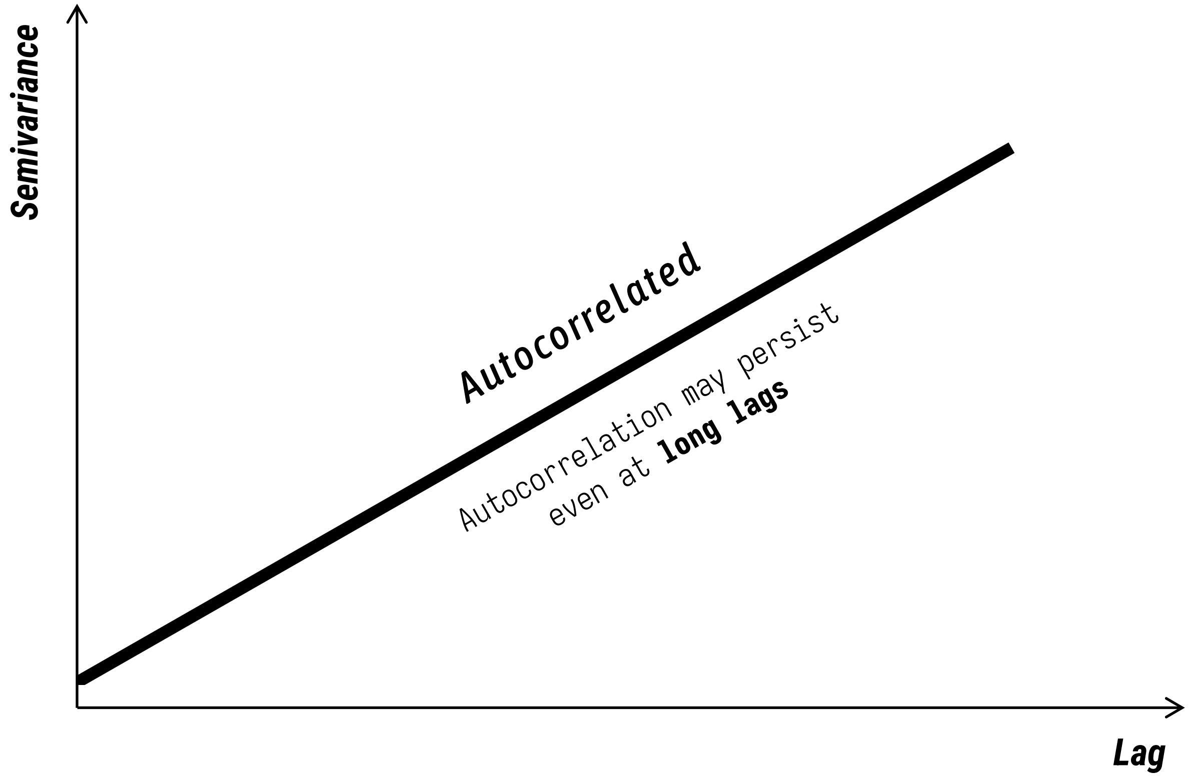
i.simoes-silva@hzdr.de

How can we measure and visualize autocorrelation in tracking data?

Semivariograms (or ***variograms***) are used to detect autocorrelation. We plot **lags** on the **x-axis** for all pairs of observations against their ***semivariance*** (average square distance between any two locations with a given lag) on the **y-axis**.



The more **similar** the pairs of locations are per lag, the **lower the semivariance** for that lag.



Different movement models can be expressed via SVFs.

Why continuous time?

Sampling may be **irregular** in time,
Sampling **schedules may differ** across individuals
— model parameters are biological, not sampling-dependent
Discrete-time models can be intrinsically **scale-dependent** (e.g., SSFs)
Continuous-time models are more **realistic** and have a **wider scope of inference**
Can easily model temporal and spatial scales that span orders of magnitude
Can accommodate **speed**, **distance**, **acceleration**, *etc.*
Location **error** is easier to model.

Home ranging behavior is a
prevalent pattern in ***space-use***.

Konrad Wothe

“

(...) it may be here remarked that most animals and plants keep to their proper homes, and do not needlessly wander about; we see this even with migratory birds, which almost always return to the same spot.

Darwin (1861)

The size and configuration of an animal's home range are key to understanding its ***space-use requirements***.

First defined as:

“

(...) the area traversed by the individual in its normal activities of food gathering, mating, and caring for young. Occasional sallies outside the area, perhaps exploratory in nature, should not be considered as in part of the home range.

Burt (1943)

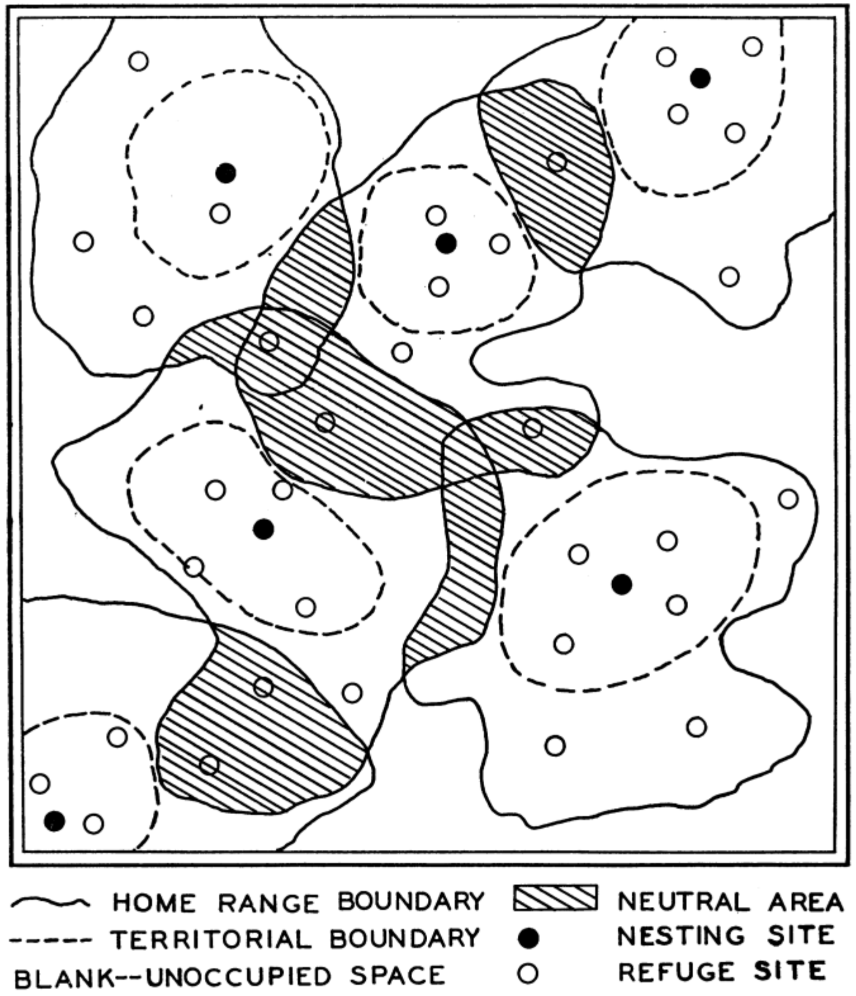


FIG. 1. Theoretical quadrat with six occupants of the same species and sex, showing territory and home range concepts as presented in text.

First defined as:

“

(...) the area traversed by the individual in its normal activities of food gathering, mating, and caring for young. Occasional sallies outside the area, perhaps exploratory in nature, should not be considered as in part of the home range.

Burt (1943)

Home range not actively defended ≠ **Territory** actively defended

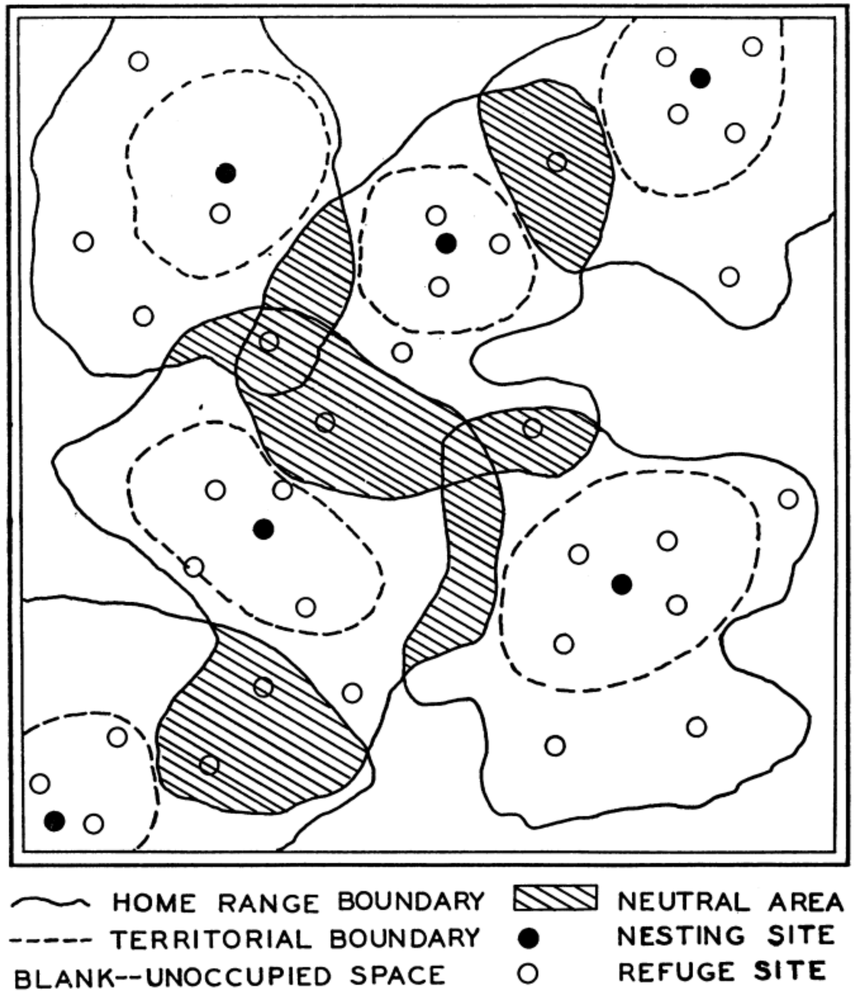


FIG. 1. Theoretical quadrat with six occupants of the same species and sex, showing territory and home range concepts as presented in text.

First defined as:

> “ (...) the area traversed by the individual in its normal activities of food gathering, mating, and caring for young. Occasional sallies outside the area, perhaps exploratory in nature, should not be considered as in part of the home range.
>
> Burt (1943)

How to quantify home range area?
What constitutes an exploratory movement?
How to identify these exploratory movements?

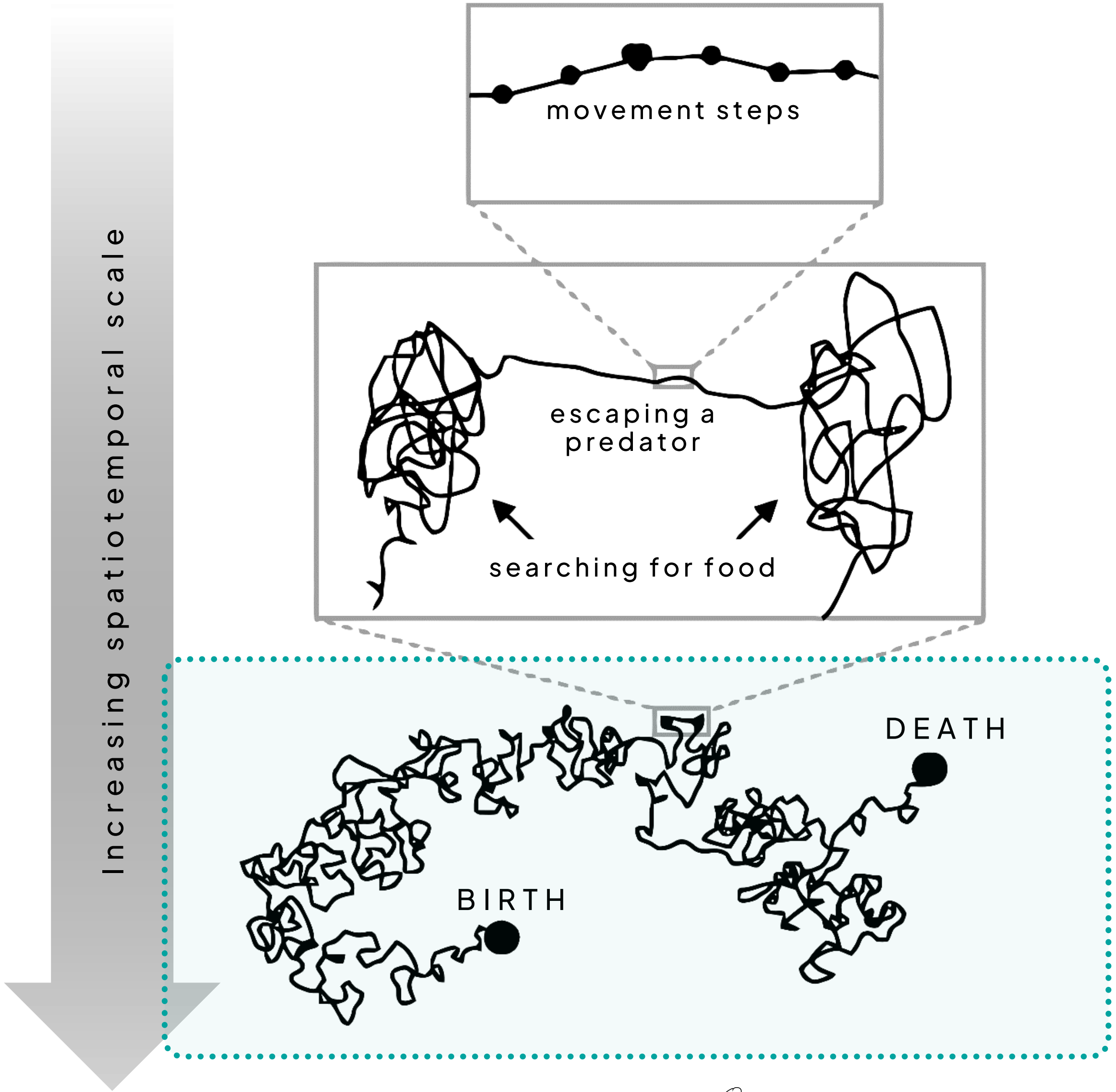
In practice, it is hard to define when a move is purely ***exploratory***.

Exclusion of infrequent *outlying relocation points*.

i.simoes-silva@hzdr.de

Here, we follow the definition of home range as the area repeatedly used throughout an animal's lifetime for all its normal behaviors and activities, excluding occasional exploratory excursions.

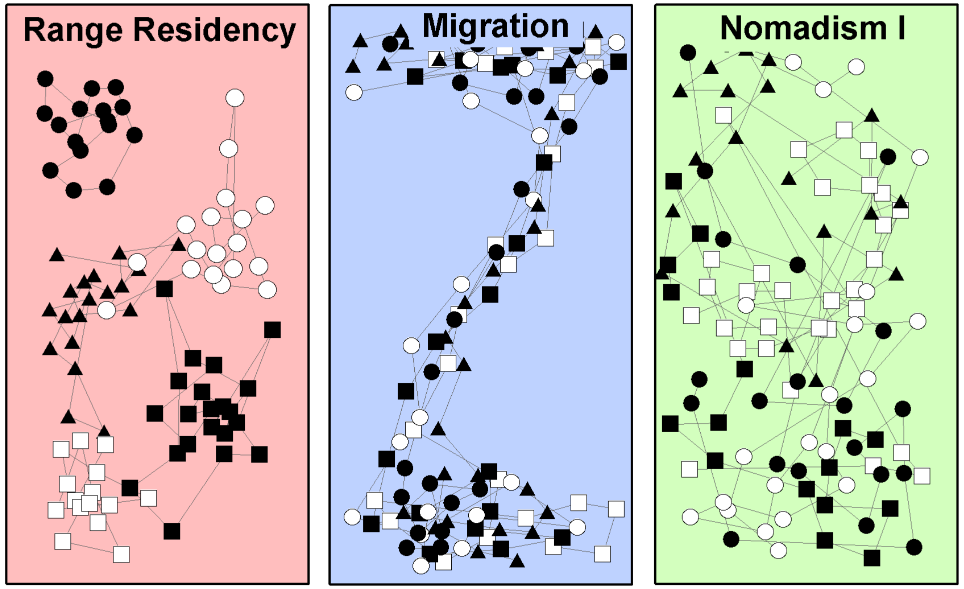
Home range area can be expected to include future locations.



Adapted from Nathan *et al.* (2008)

However, not all animals have ***home ranges***…
There are three main behaviors that animals may exhibit:

- individual occupies the ***same area*** throughout its ***lifetime*** (*i.e.*, range resident);
- regular movement to and from ***spatially disjoint ranges*** (migratory);
- does ***not*** follow regular temporal and spatial patterns (nomadic),
- ***permanently*** moves from one to another (dispersive).



Mueller *et al.* (2008)

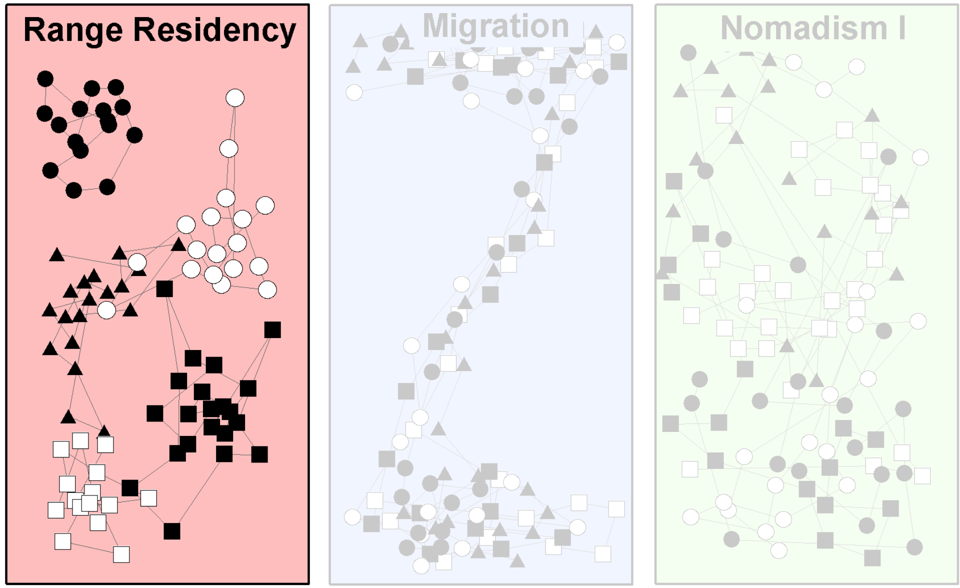
However, not all animals have ***home ranges***…
There are three main behaviors that animals may exhibit:

- individual occupies the ***same area*** throughout its ***lifetime*** (*i.e.*, range resident);
- regular movement to and from ***spatially disjoint ranges*** (migratory);
- does ***not*** follow regular temporal and spatial patterns (nomadic),
- ***permanently*** moves from one to another (dispersive).

intra-individual variation
inter-individual variation
inter-population variation

or

sampling schedule



Mueller *et al.* (2008)

Home-range area estimates may inform:

- Protected area delineation,
- Land-use decisions,
- Conservation policy and initiatives,
  (*e.g.*, related to human-wildlife conflict).

It is vital to accurately capture the area repeatedly used throughout an ***animal's lifetime***.

Dennis Hamilton

Avijan Saha

Aleksei Volkov

i.simoes-silva@hzdr.de

Benjamin Marshall

Northern king cobra
(Ophiophagus hannah)

Northern king cobra
(*Ophiophagus hannah*)

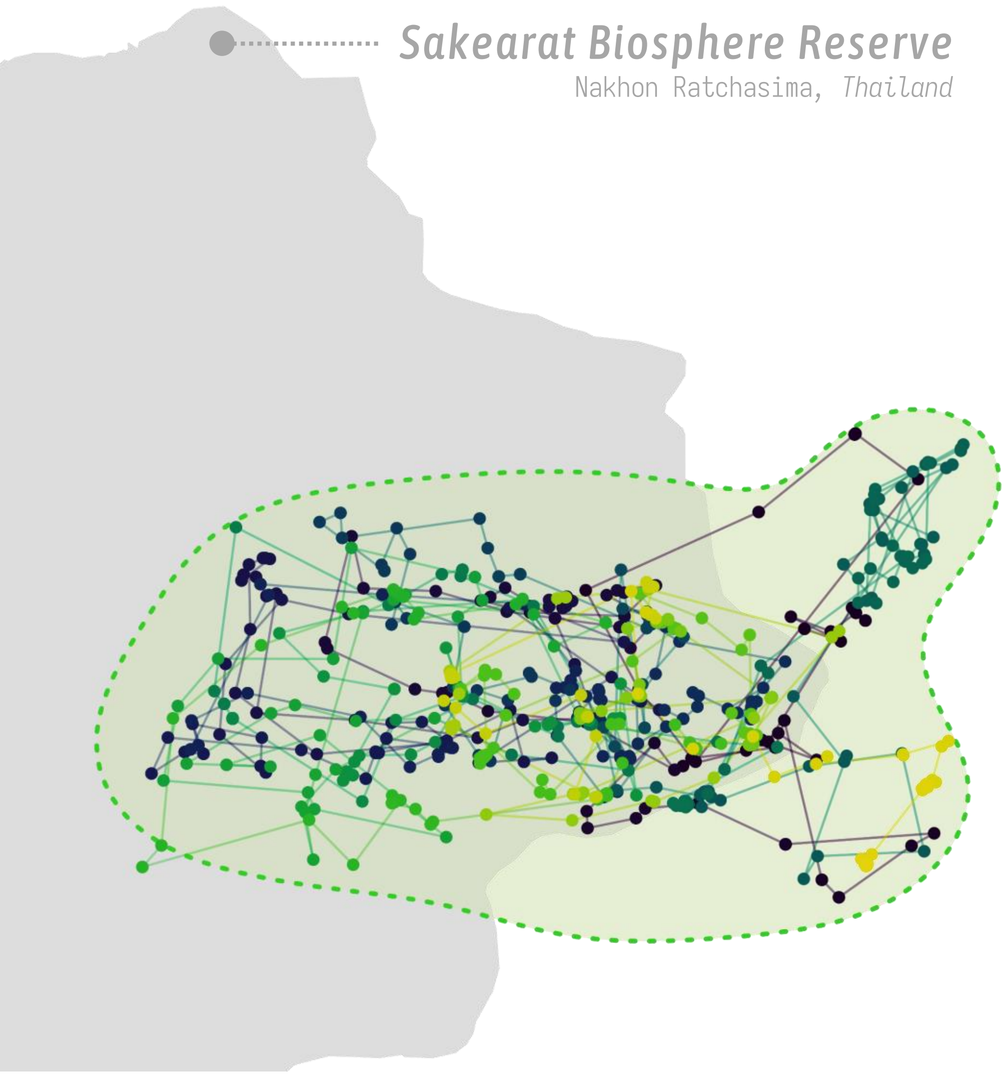
BIRTH DEATH

Sampling *duration* ≈ 1 year

start end

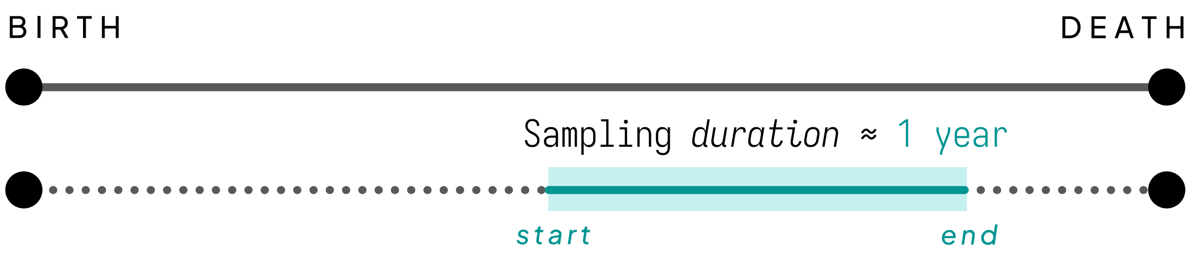
How representative is this period?
Is the movement behavior stationary?

Why estimate home range?



Benjamin Marshall

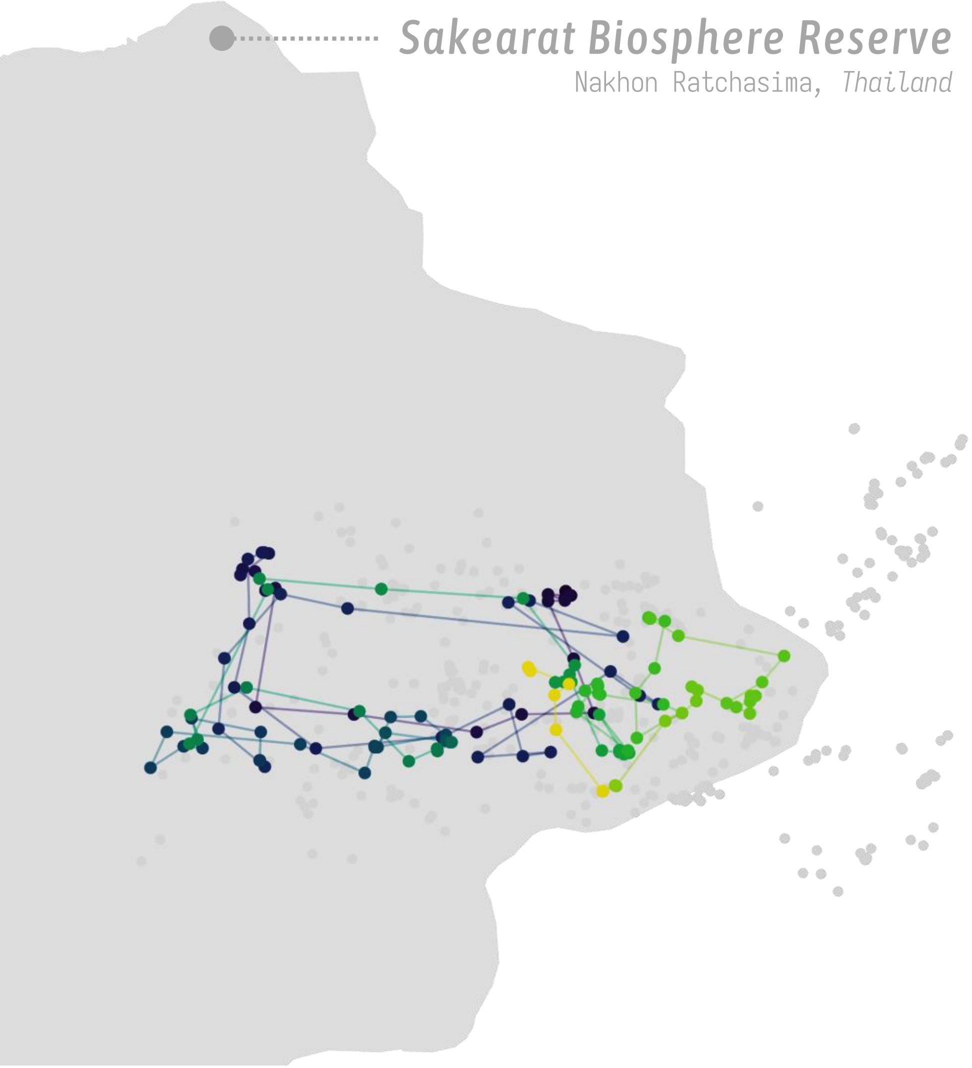
Northern king cobra
(Ophiophagus hannah)



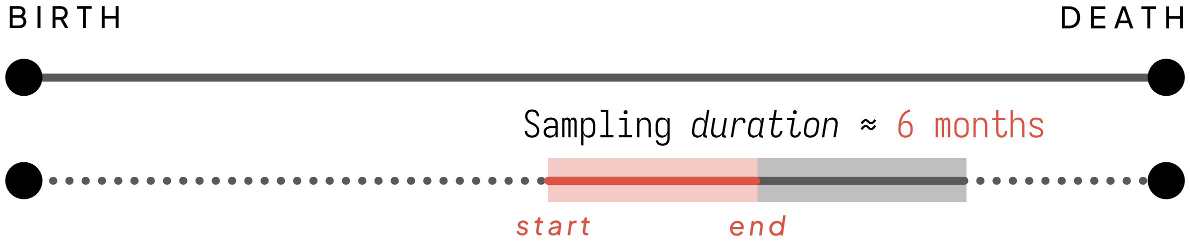
How representative is this period?
Is the movement behavior stationary?

i.simoes-silva@hzdr.de

Why estimate home range?

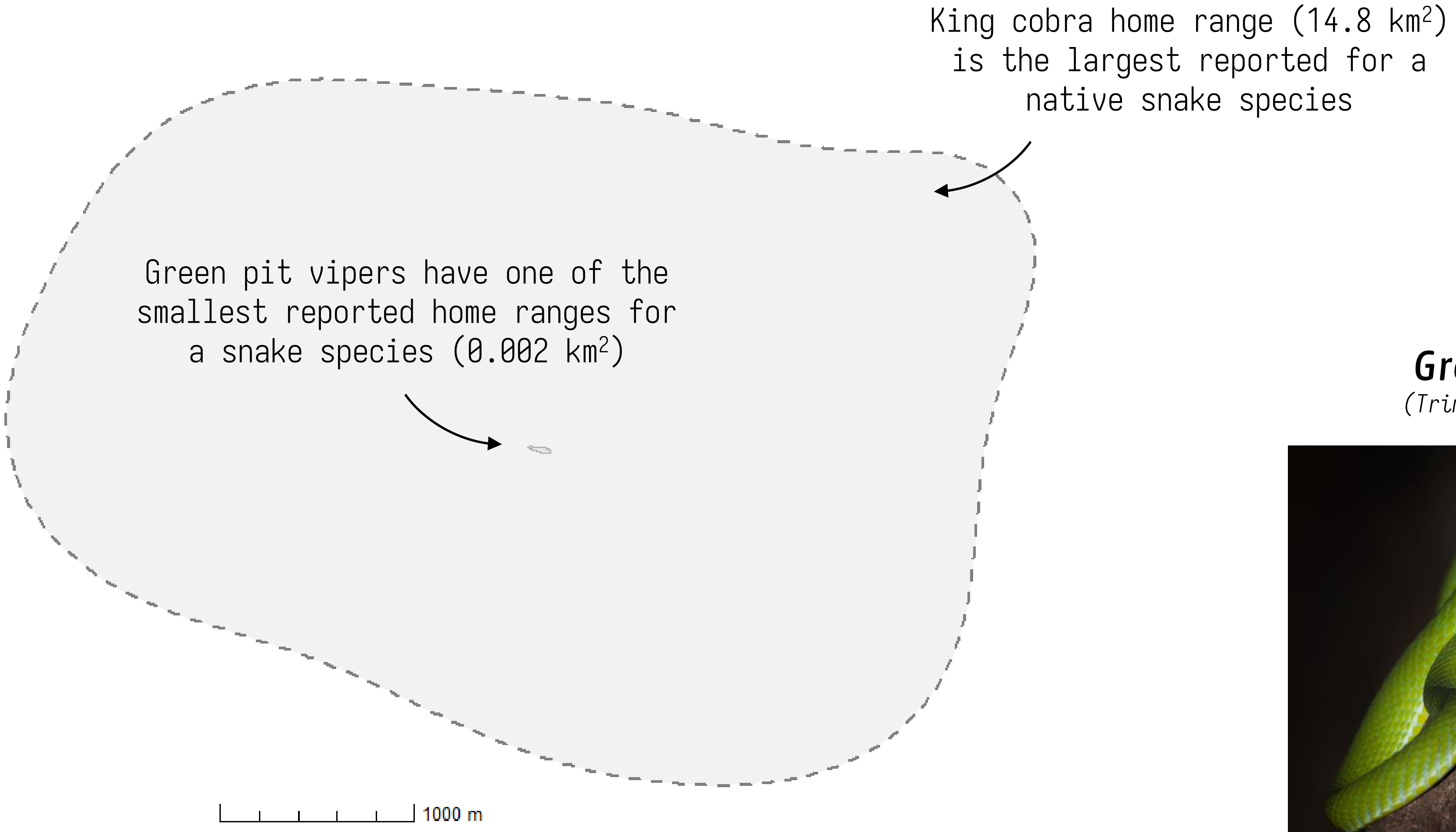


Northern king cobra
(*Ophiophagus hannah*)



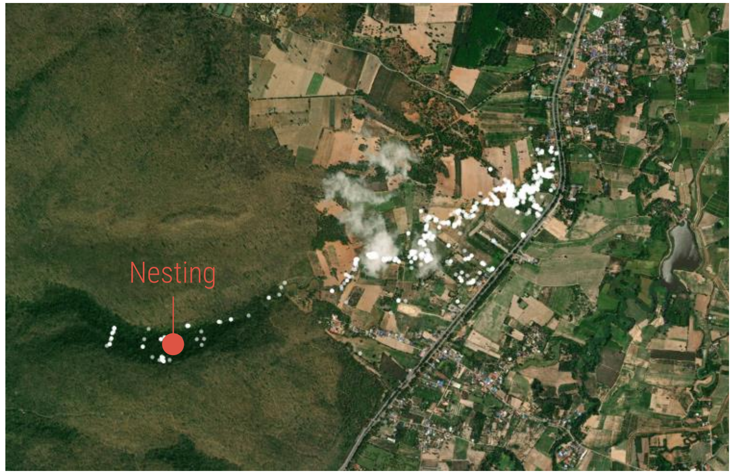
How *representative* is this period?
Is the movement behavior *stationary*?

Why estimate home range?

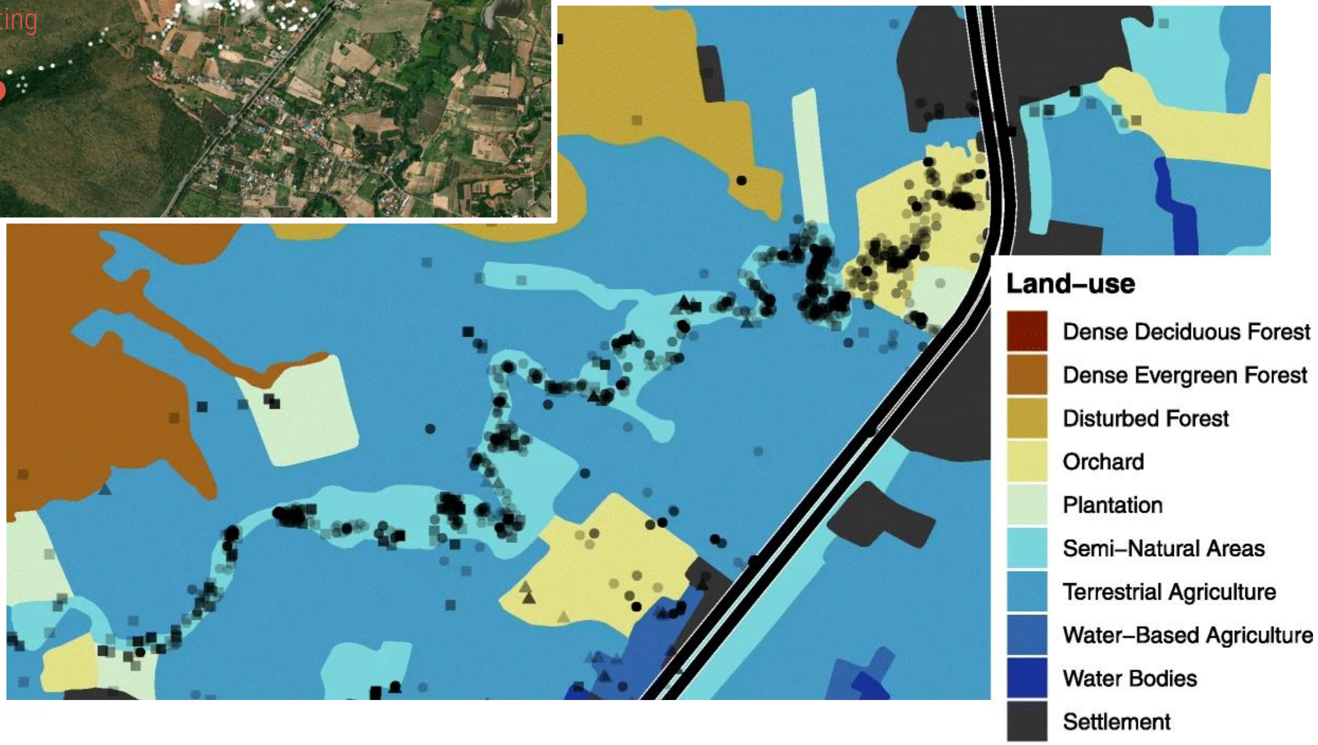


Green pit viper
(Trimeresurus macrops)

i.simoes-silva@hzdr.de



Marshall et al. (2020)

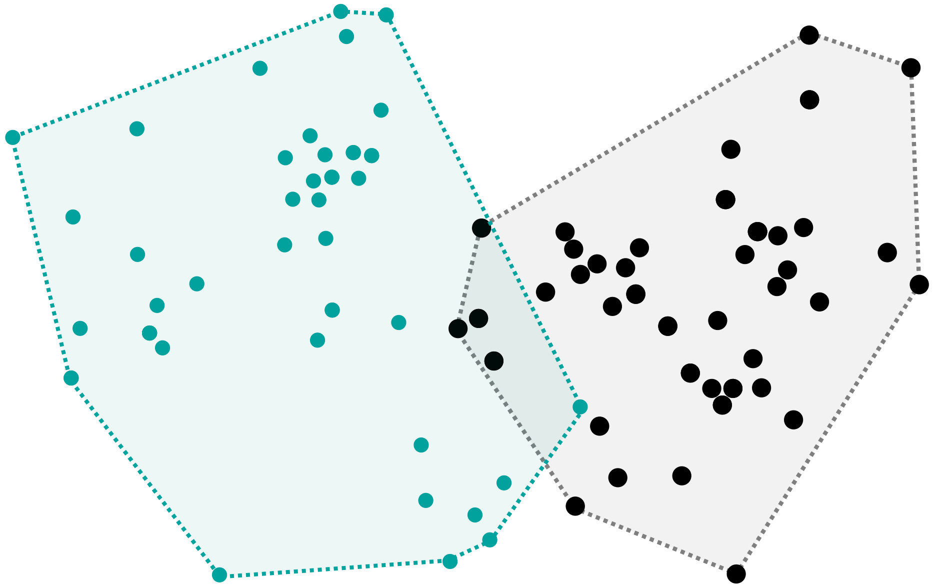


Movements are largely occurring within semi-natural areas.

Females almost exclusively utilize agricultural canals as ***movement corridors***.

— Minimum Convex Polygon (MCP):

Mohr (1947)



The smallest polygon drawn around tracking locations with all interior angles less than 180 degrees.

Burt (1943)

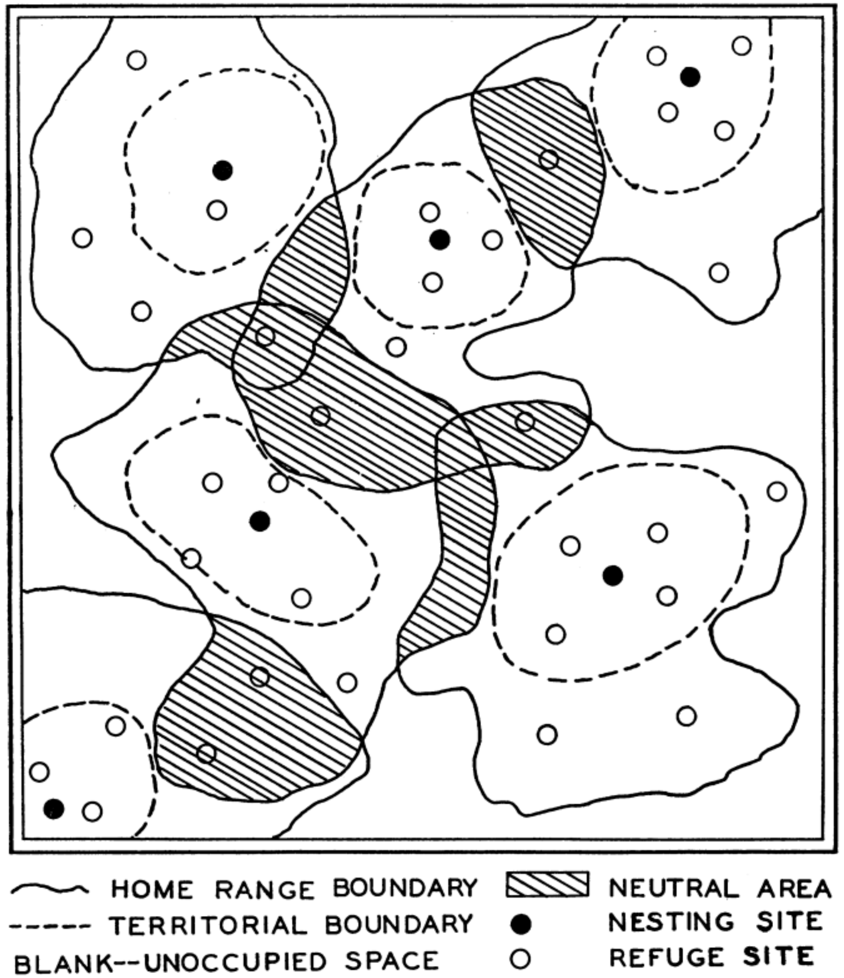
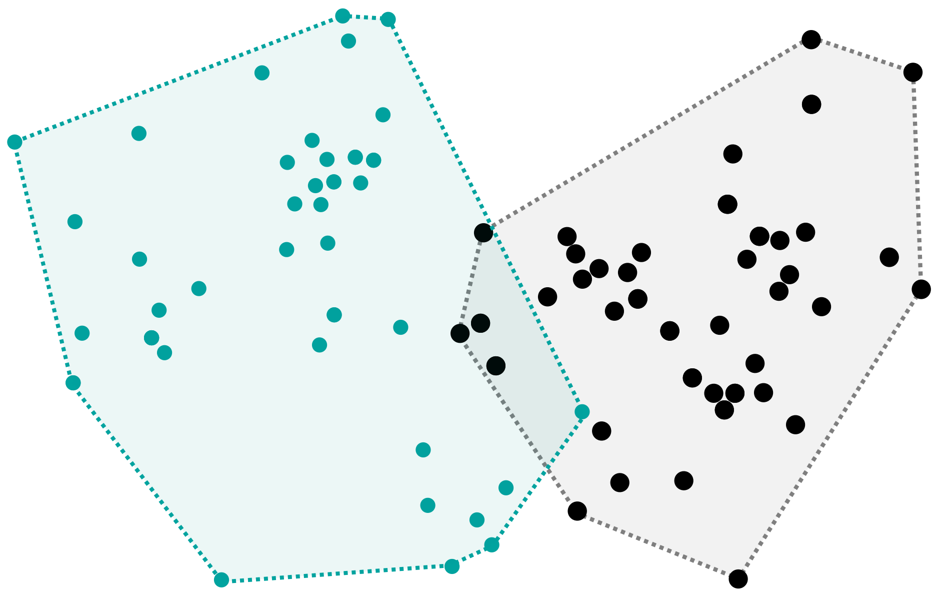


FIG. 1. Theoretical quadrat with six occupants of the same species and sex, showing territory and home range concepts as presented in text.

i.simoes-silva@hzdr.de

— Minimum Convex Polygon **(MCP)**:

Mohr (1947)



The smallest polygon drawn around tracking locations with all interior angles less than 180 degrees.

Assumes ***uniform use***,
Assumes locations are ***independent***;
Sensitive to ***outliers*** and point geometry.

i.simoes-silva@hzdr.de

“

It seems that an understanding of the biological significance of an animal's home range must include some knowledge of the intensity of use, by the animal, of various parts of the area.

Hayne (1949)

The use of frequently visited areas within an animal's home range, not resulting from random or continuous movement, should be statistically clustered.

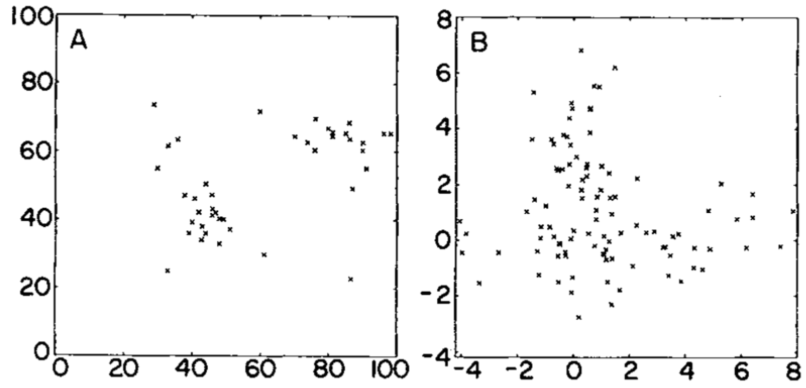


FIG. 1. Plots of (A) the DC data set and (B) the SIM data set.

Worton (1989)

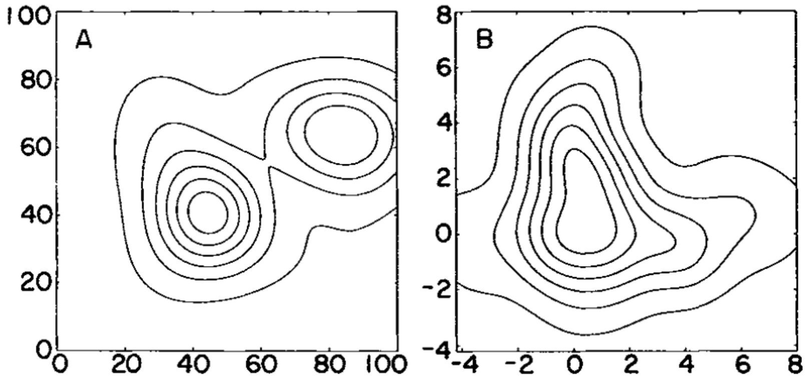
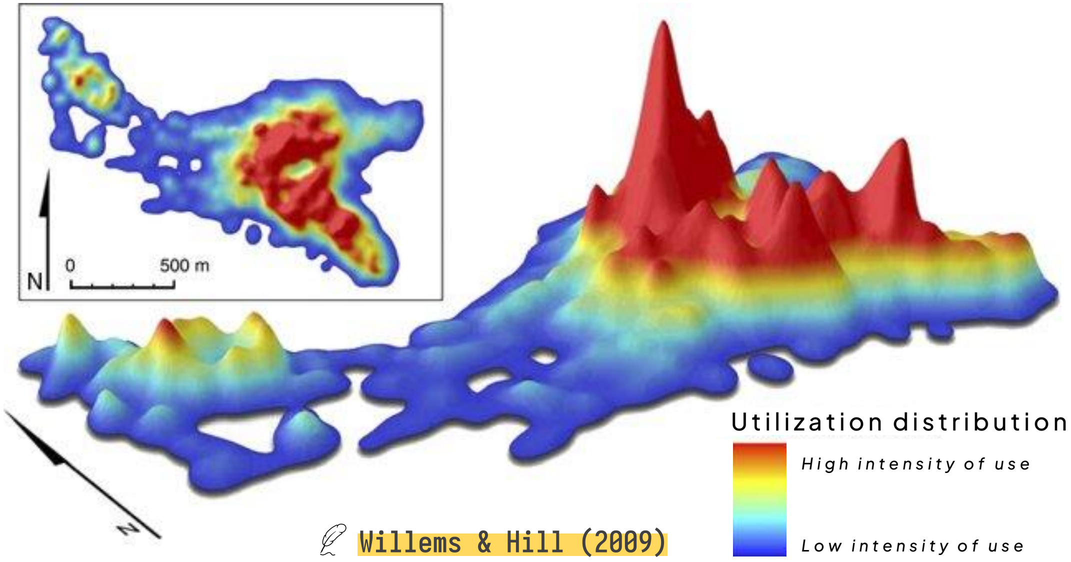


FIG. 2. Fixed kernel density estimates of the UD densities with the ad hoc choice of smoothing parameters for (A) the DC data set (h = 10.) and (B) the SIM data set (h = 1.).

“

It seems that an understanding of the biological significance of an animal's home range must include some knowledge of the intensity of use, by the animal, of various parts of the area.

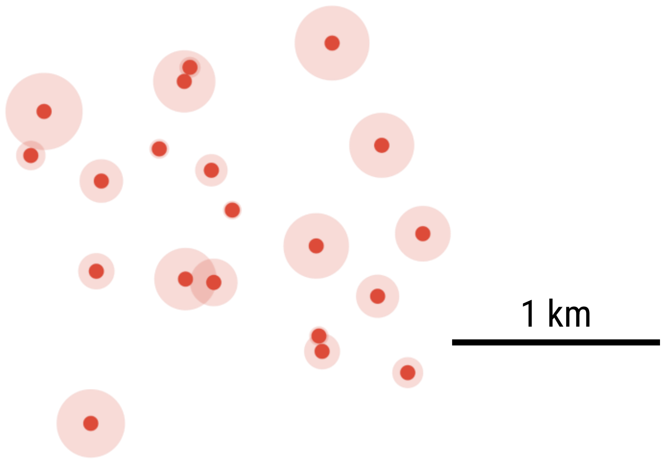
Hayne (1949)

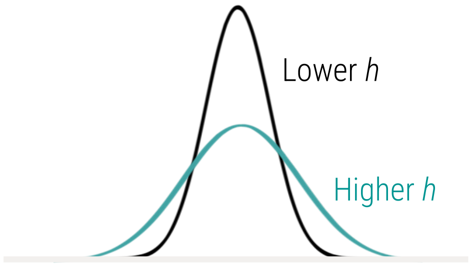


Willems & Hill (2009)

Utilization distributions (UDs):
Characterizing space-use as a **density function** (accounts for intensity of use).

KDEs describe not just the borders of the home range, but the ***intensity of use***.

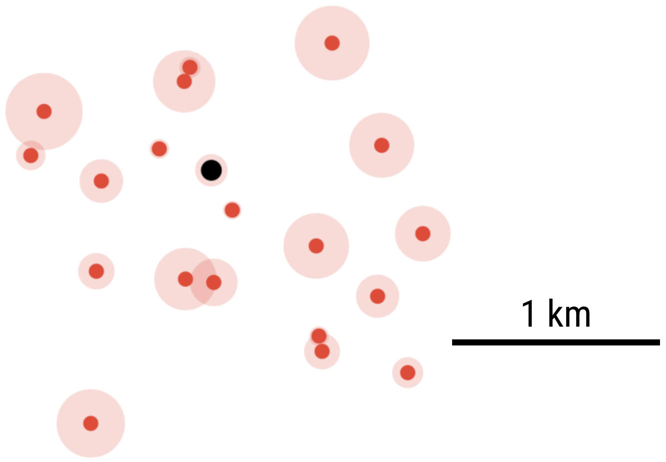


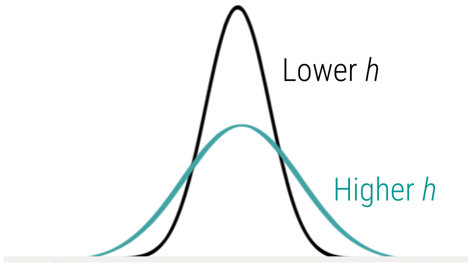


On an x—y plane, each location has a three-dimensional "hill", the ***kernel***.
The shape and width of the kernel, called the ***bandwidth*** (*h*), can be selected using algorithms.

i.simoes-silva@hzdr.de

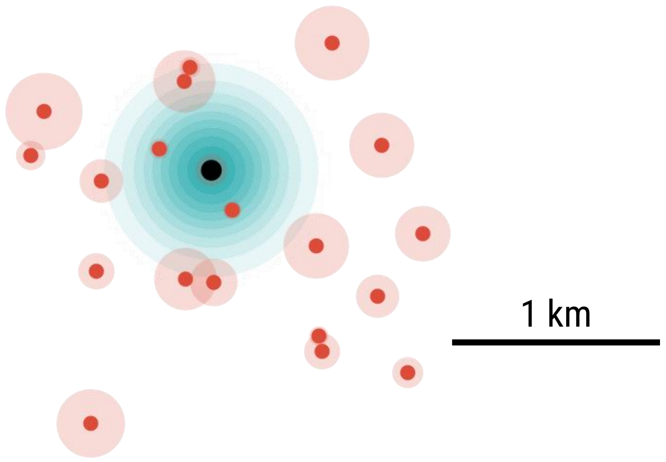
KDEs describe not just the borders of the home range, but the ***intensity of use***.

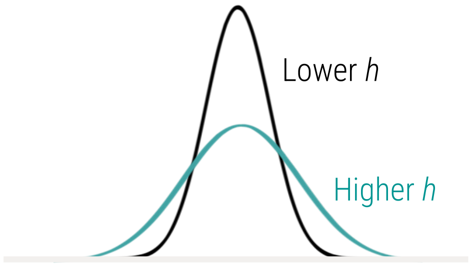




On an x—y plane, each location has a three-dimensional "hill", the ***kernel***.
The shape and width of the kernel, called the ***bandwidth*** (*h*), can be selected using algorithms.

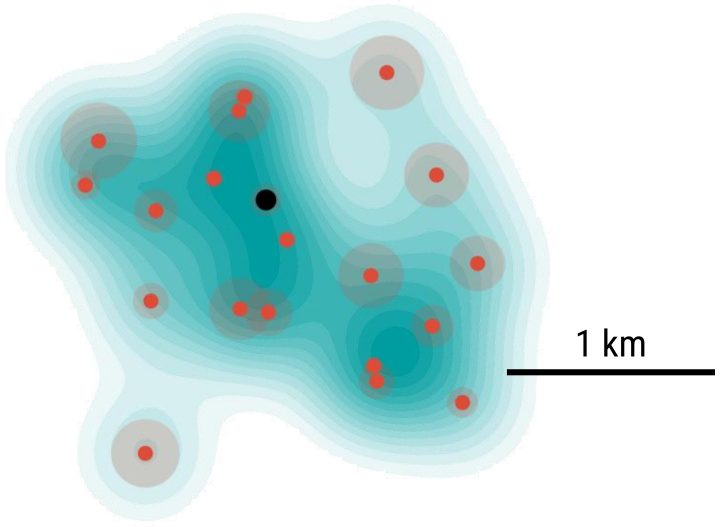
KDEs describe not just the borders of the home range, but the ***intensity of use***.

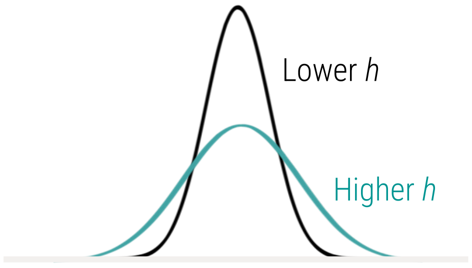




On an x—y plane, each location has a three-dimensional "hill", the ***kernel***.
The shape and width of the kernel, called the ***bandwidth*** (*h*), can be selected using algorithms.

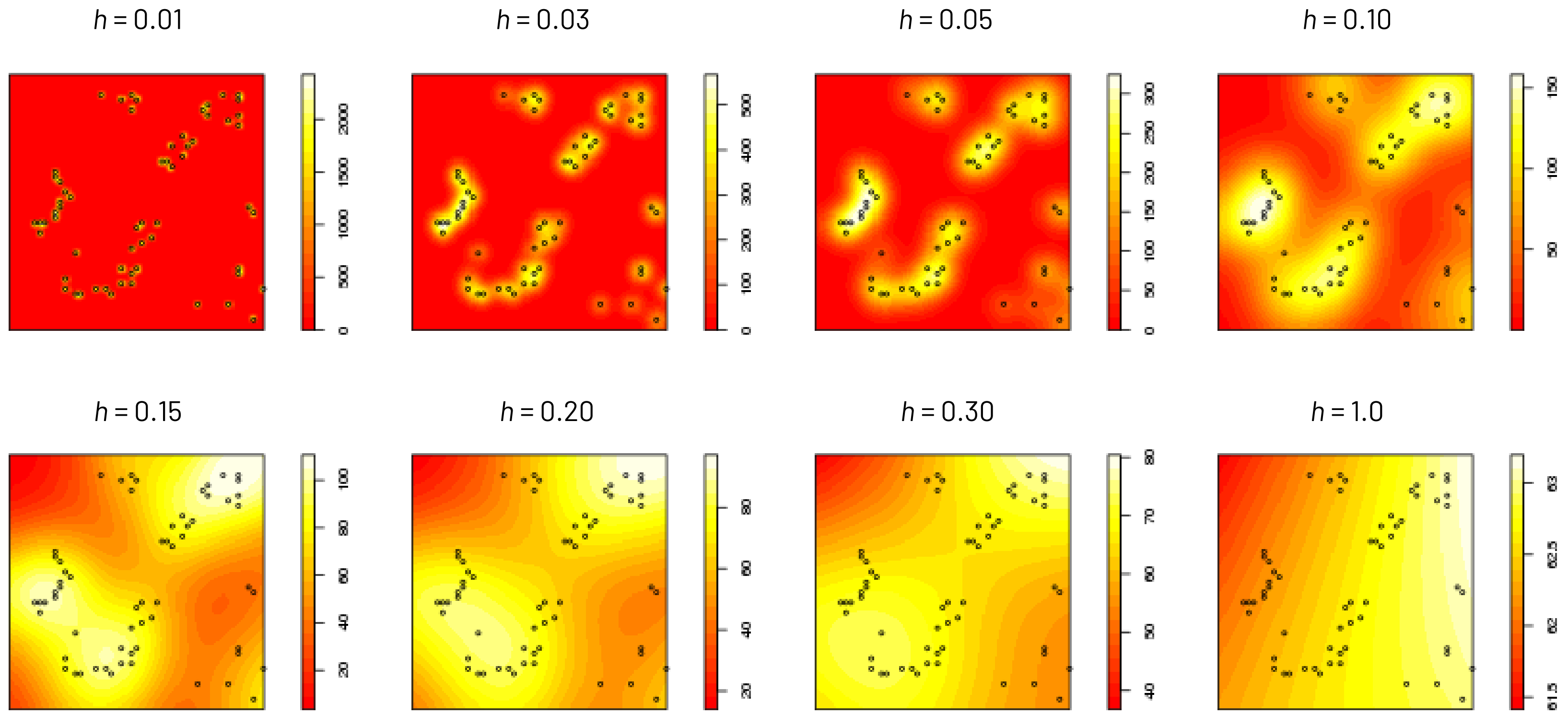
KDEs describe not just the borders of the home range, but the ***intensity of use***.





On an x—y plane, each location has a three-dimensional "hill", the ***kernel***.
The shape and width of the kernel, called the ***bandwidth*** (*h*), can be selected using algorithms.

i.simoes-silva@hzdr.de



Increasing the ***bandwidth (h)***, or ***smoothing parameter***

i.simoes-silva@hzdr.de

KDE is the most statistically efficient ***non-parametric distribution estimator***.

not explicitly modeling the causes of space use

The objective is typically to minimize the ***'mean integrated square error' (MISE)***:

$$\mathrm{MISE}(H) = \mathrm{E}\left[\iint \left(\hat{p}(x, y|H) - p(x, y)\right)^2 dx\, dy\right]$$

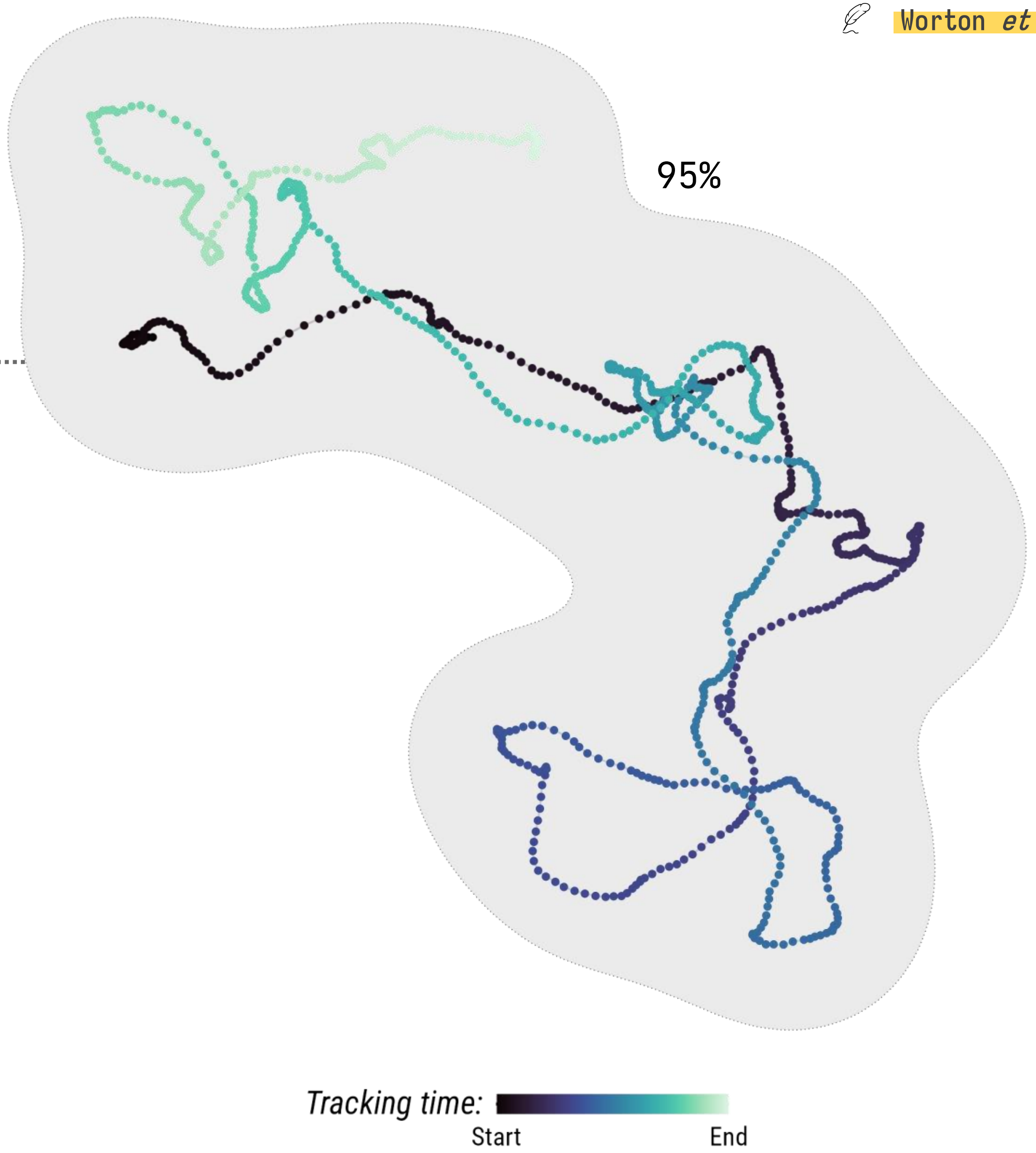
with respect to the ***bandwidth*** or ***smoothing***, ***h***

h is not a model parameter— there is no true value of ***h*** that best characterizes an animal! You do not choose your bandwidth. ***You choose your bandwidth optimizer.***

Worton *et al.* (1989)

Sensitive to ***bandwidth selection***.

Gaussian reference function
(**GFR** or h_{ref})



It assumes:
– locations are independent,
– data follows a normal (Gaussian) distribution.

$$h_{ref} = 1.06 \cdot \sigma_i \cdot n^{-1/(4+d)}$$

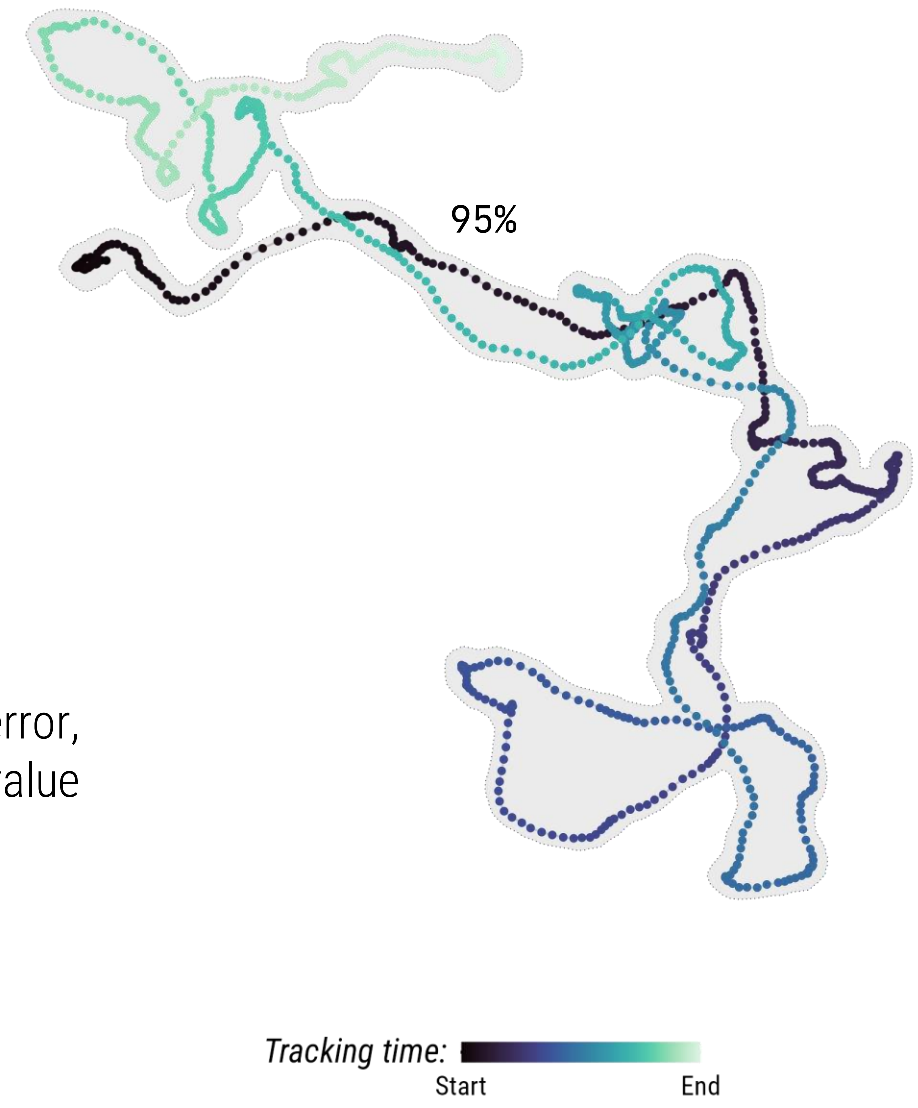
where:
σ is the standard deviation of your data
n is the number of observations
n is the number of dimensions of the data

i.simoes-silva@hzdr.de

Seaman & Powell (1996)

Sensitive to ***bandwidth selection***.

Least-squares cross-validation (LSCV)
Attempts to prevent oversmoothing...



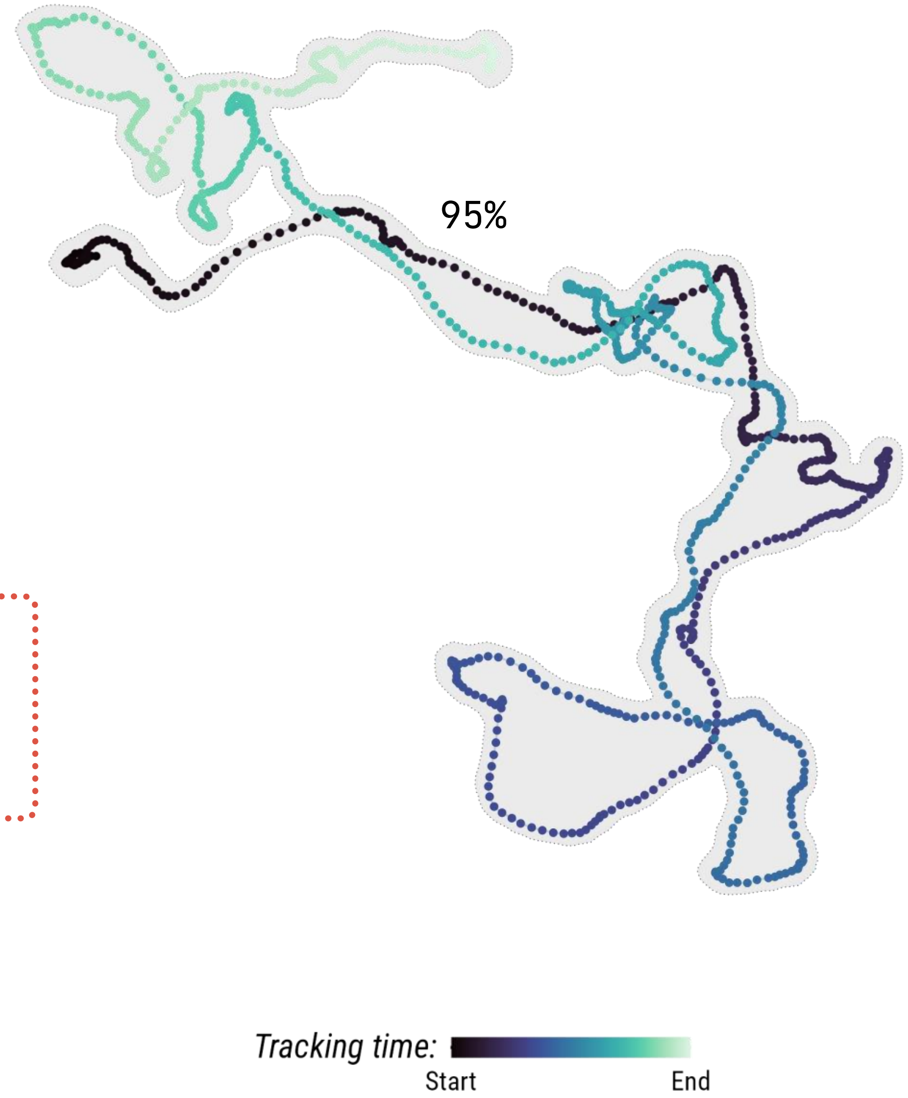
It assumes:
— locations are independent.

LSCV finds a bandwidth that minimizes estimation error, by leaving out each point and trying to predict its value from the rest.

Seaman & Powell (1996)

Sensitive to ***bandwidth selection***.

Least-squares cross-validation (**LSCV**)



This algorithm performs poorly with ***large sample sizes***, and still does not account for the locations' temporal structure.

What about *data thinning*?

Thinning the data to achieve independence:

Fig. Tracking data representing **hourly** locations over **one month**.

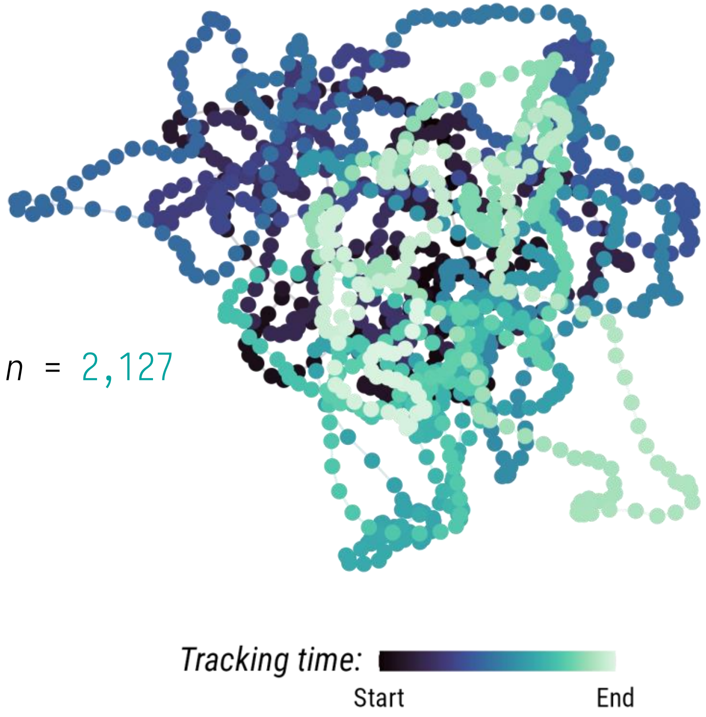
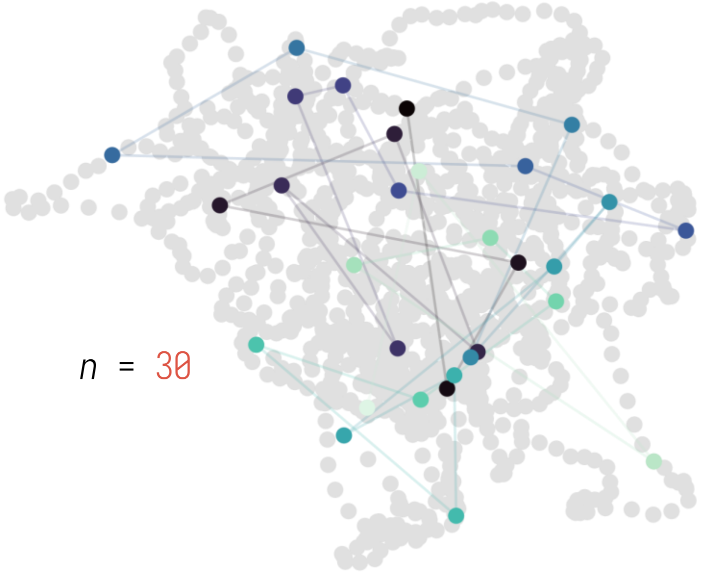


Fig. Tracking data subsampled so there is only **one location per day**.



Data loss of ≈ 70.4%

i.simoes-silva@hzdr.de

Thinning the data to achieve independence:

Fig. Tracking data representing **hourly** locations over **one month**.

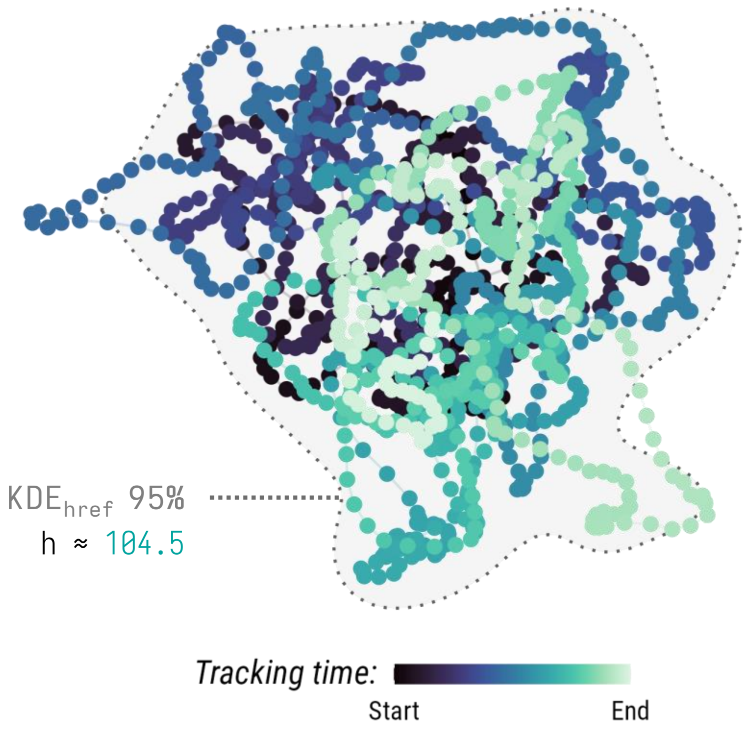
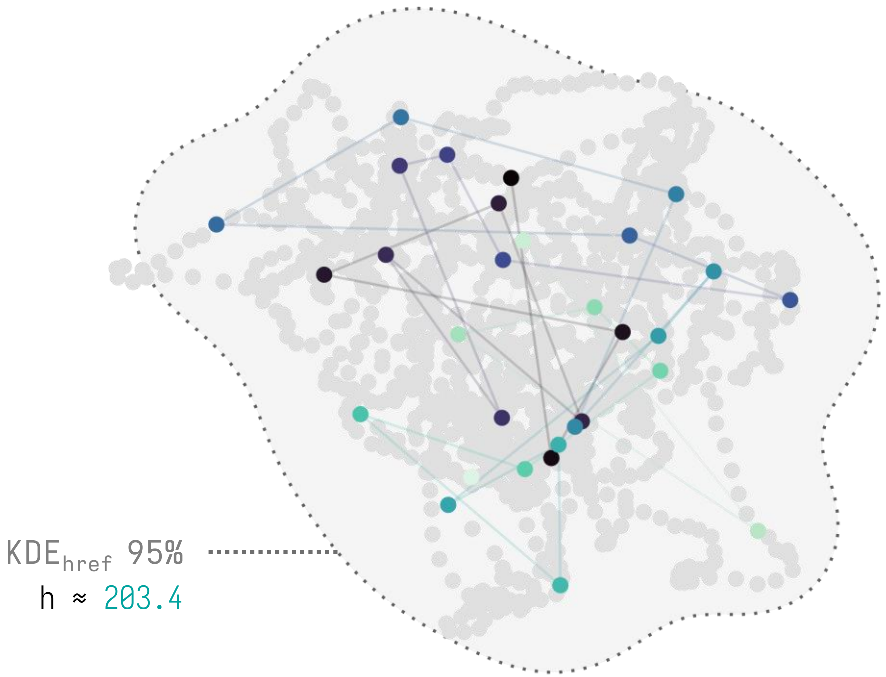


Fig. Tracking data subsampled so there is only **one location per day**.



Data loss of ≈ 70.4%

i.simoes-silva@hzdr.de

Thinning the data to achieve independence:

Fig. Tracking data representing **hourly** locations over **one month**.

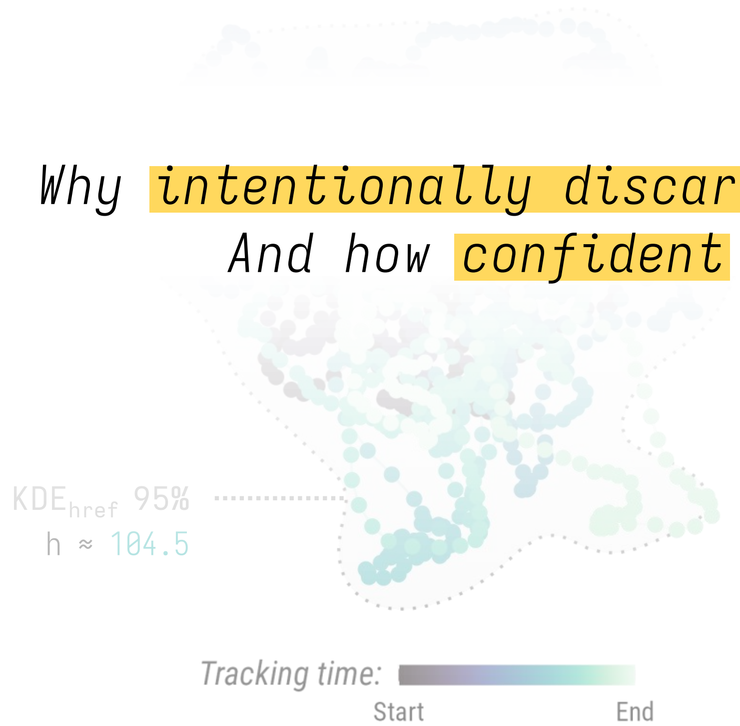
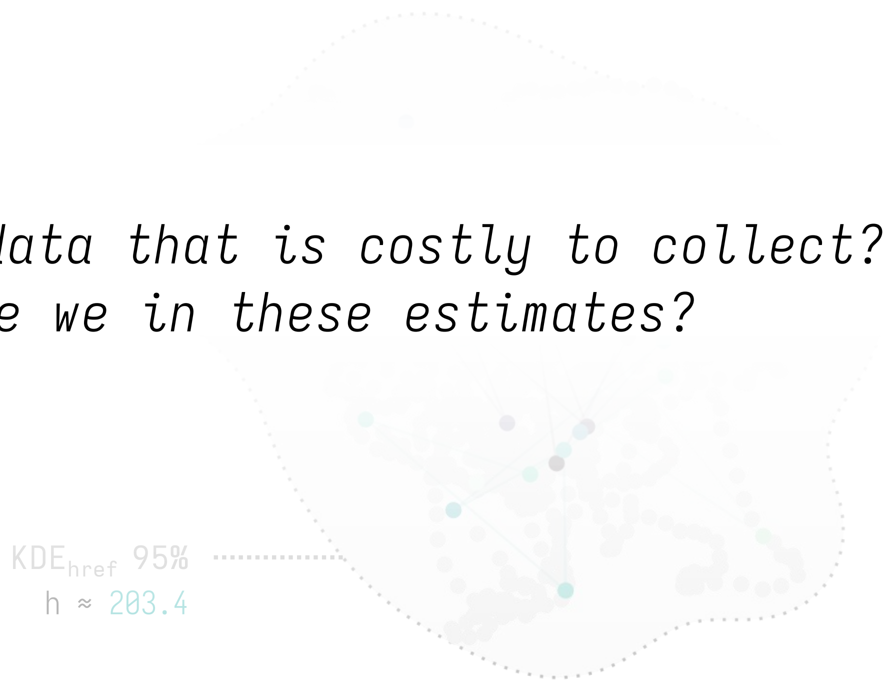


Fig. Tracking data subsampled so there is only **one location per day**.



Data loss of ≈ 70.4%

Why intentionally discard data that is costly to collect?
And how confident are we in these estimates?

i.simoes-silva@hzdr.de

Utilization distribution

represents an animal's distribution and the probability of use throughout an area

Range distribution

extrapolate space use into the future

≠

Occurrence distribution

interpolate between data points in the past

"How much space does an animal need over the long term?"

"Where did an animal go during a period of observation?"

What is an animal's home range area?
What is their the population range area?
Are protected areas sufficiently large?

Where did an animal cross a linear feature?
How likely is it to visit a location of interest?
How much time did it spend in a specific habitat?

Alston *et al.* (2022)

i.simoes-silva@hzdr.de

Utilization distribution

represents an animal's distribution and the probability of use throughout an area

Range distribution

extrapolate space use into the future

≠

Occurrence distribution

interpolate between data points in the past

Estimates ***home ranges***

Considers all realizations of the movement process that could occur.

Sampling-independent!

Estimates ***occurrence regions***

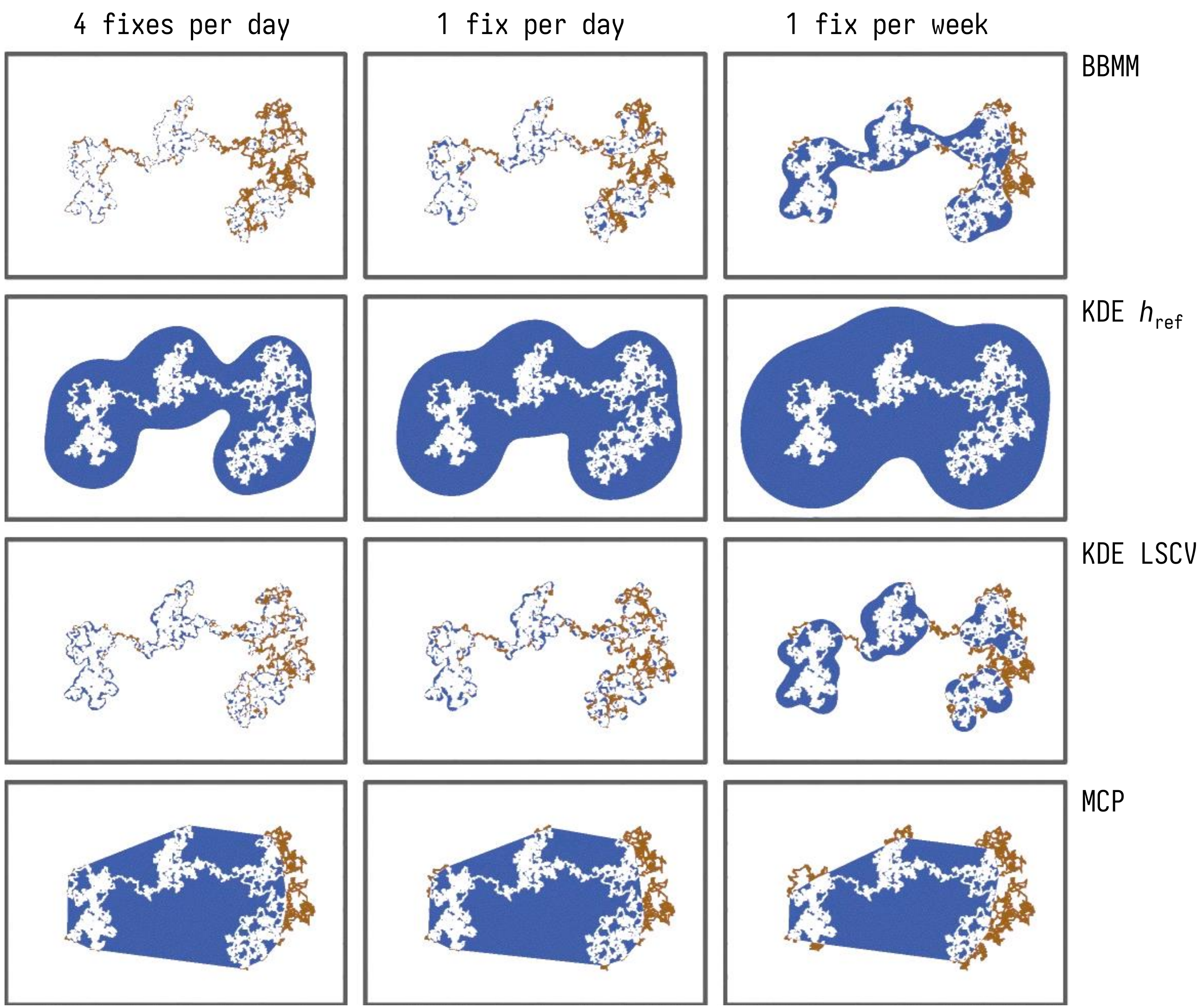
Pinpoints the one realization of the movement process that did occur.

Sampling-dependent!

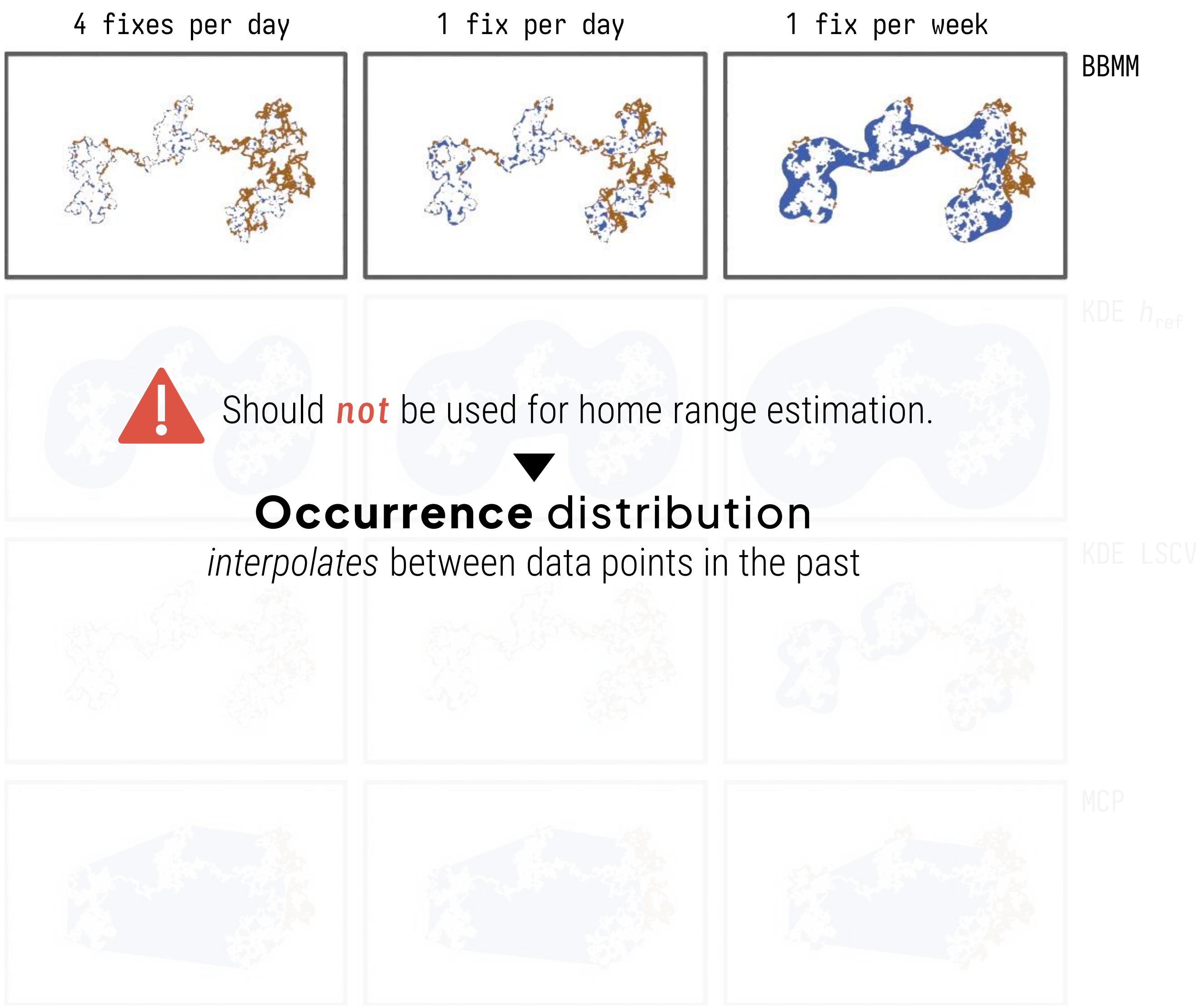
Alston *et al.* (2022)

i.simoes-silva@hzdr.de

*What is each method **actually** estimating?*



Silva *et al.* (2021)



Should **not** be used for home range estimation.

Occurrence distribution

interpolates between data points in the past

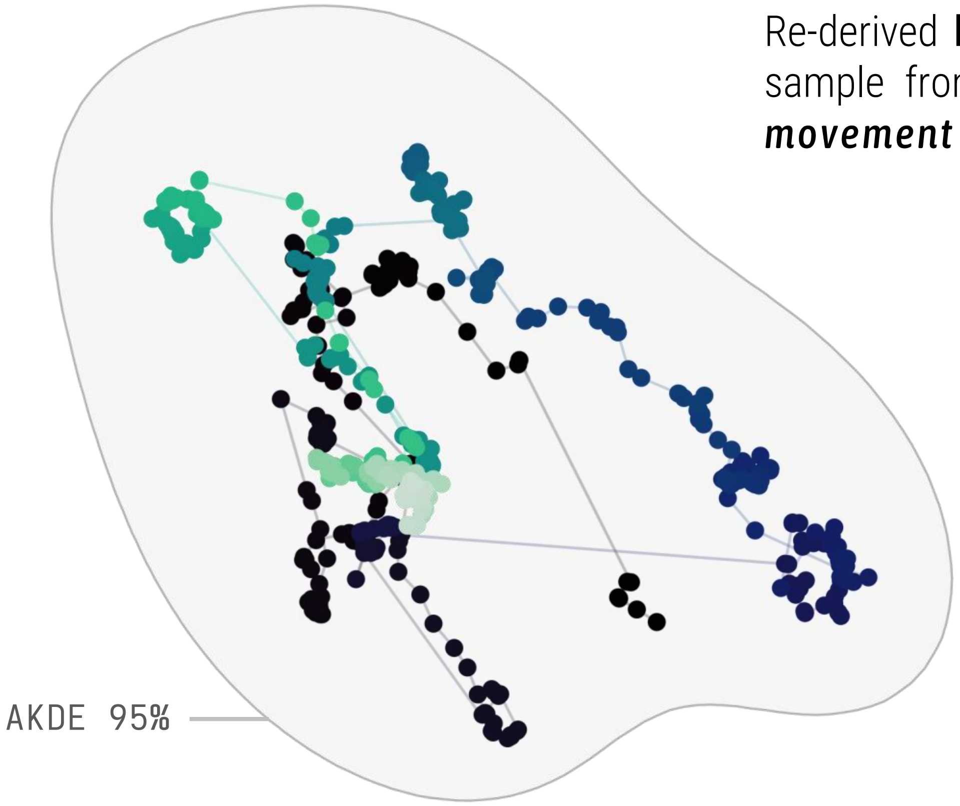
What is each method actually estimating?

Not all methods are answering the same questions…

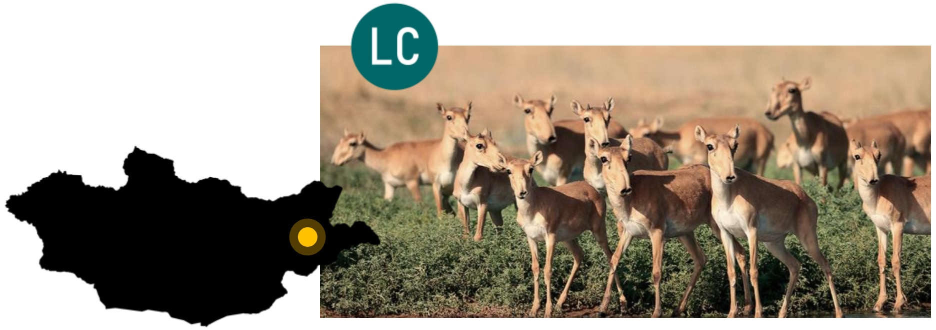
Silva *et al.* (2021)

Autocorrelated Kernel Density Estimator **(AKDE)**:

Fleming *et al.* (2015)



Re-derived **KDE** that explicitly assumes the data represents a sample from a ***nonstationary***, ***autocorrelated***, ***continuous movement process***.



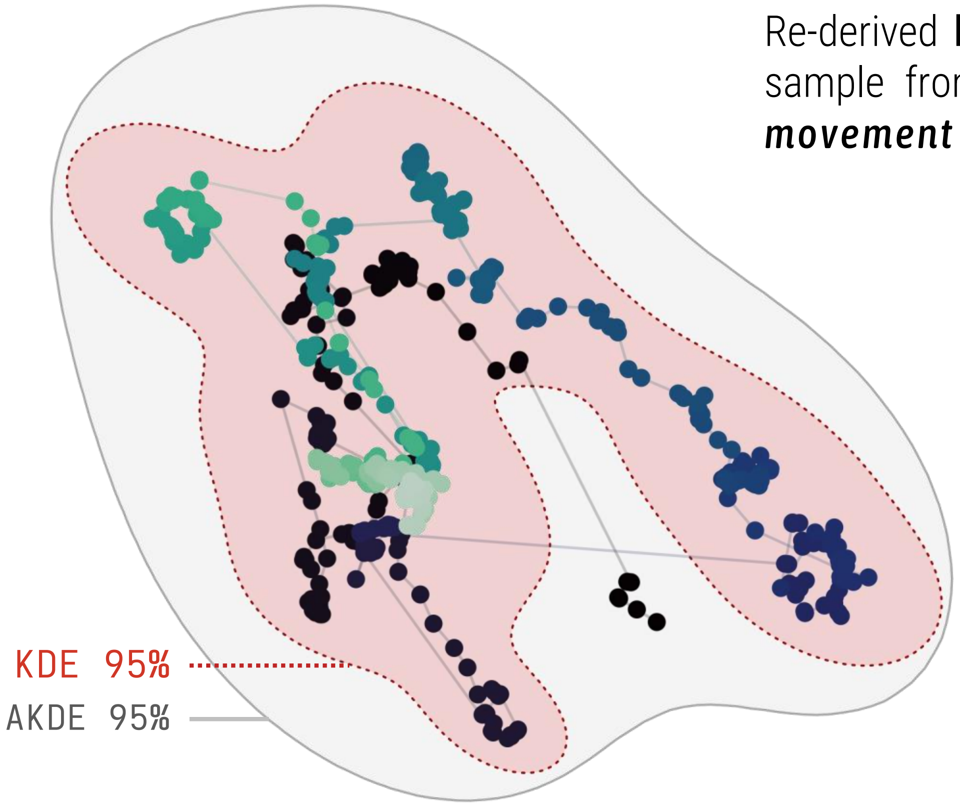
Mongolian gazelle
(Procapra gutturosa)

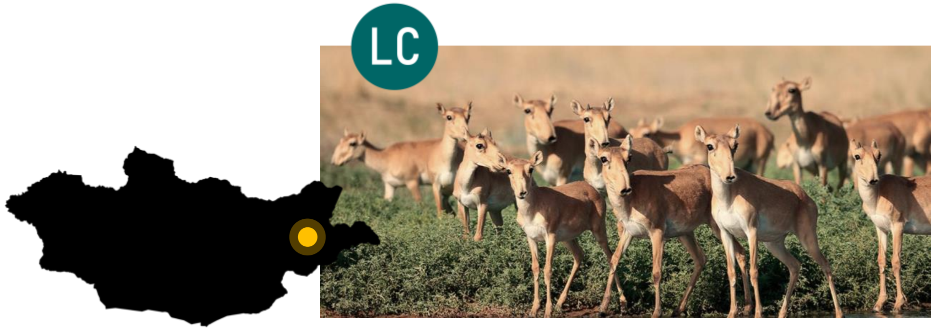
Duration ≈ 1 year
τ_p ≈ 2.4 months

Autocorrelated Kernel Density Estimator **(AKDE)**:

Fleming *et al.* (2015)

Re-derived **KDE** that explicitly assumes the data represents a sample from a ***nonstationary***, ***autocorrelated***, ***continuous movement process***.



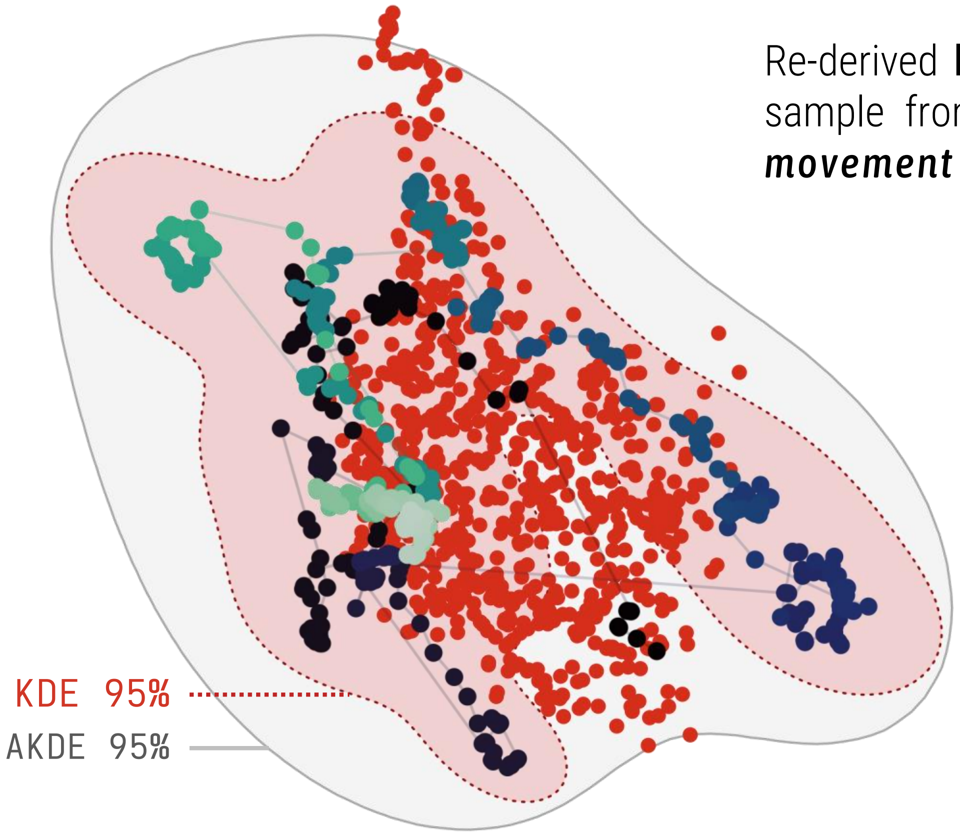


Mongolian gazelle
(Procapra gutturosa)

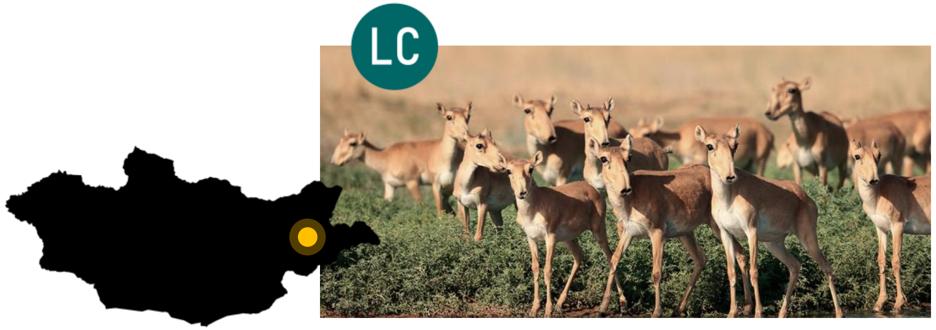
Duration ≈ 1 year
τ_p ≈ 2.4 months

Autocorrelated Kernel Density Estimator **(AKDE)**:

Fleming *et al.* (2015)



Re-derived **KDE** that explicitly assumes the data represents a sample from a ***nonstationary***, ***autocorrelated***, ***continuous movement process***.

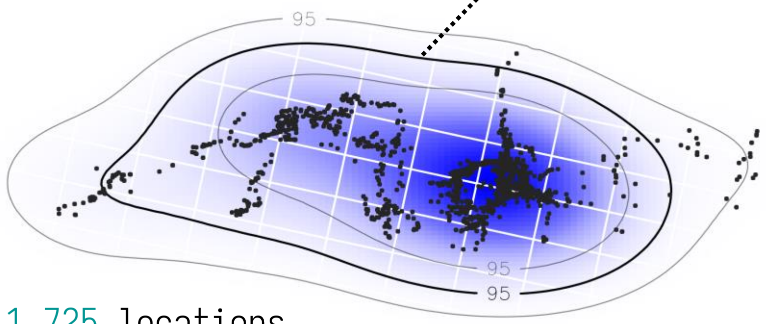


Mongolian gazelle
(*Procapra gutturosa*)

Duration ≈ 2 years → showing an ***extra year of data***!
τ_p ≈ 2.4 months

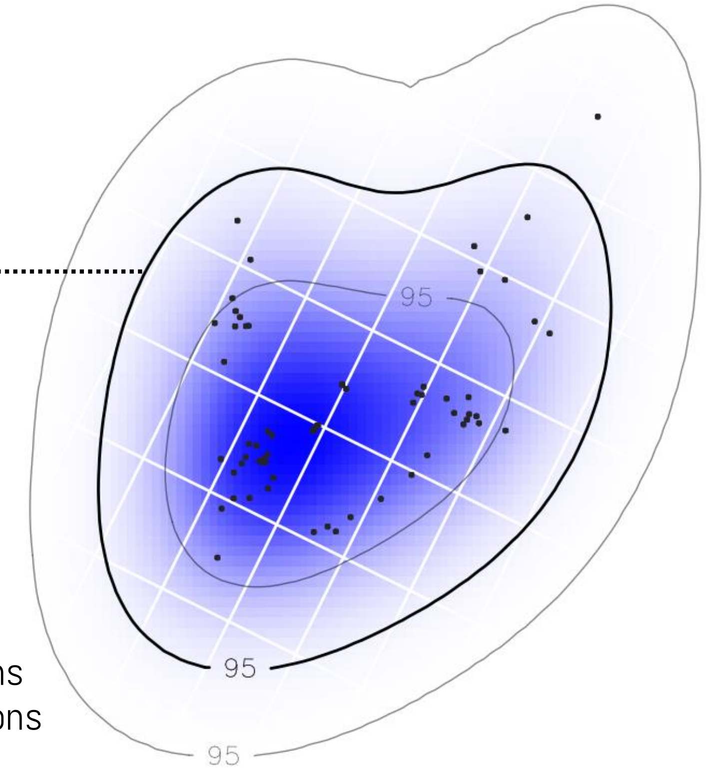
AKDEs also provide accurate ***confidence intervals*** that can diagnose situations where the data are insufficient to provide a reasonable home range estimate.

The 95% contour represents the maximum likelihood area where the animal spends 95% of its time.



n = 1,725 locations
N ≈ 15.7 locations

Home range area estimate
757.5 km² (430.0 – 1,176.2)



n = 66 locations
N ≈ 6.9 locations

Home range area estimate
4,232.3 km² (1,681.7 – 7,939.5)

Sample size, ***n*** is the denominator when calculating both SEs and CIs

Standard error (SE):

$$SE = \frac{\sigma}{\sqrt{\boldsymbol{n}}}$$

Confidence intervals (CI):

$$95\%\ CI = \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{\boldsymbol{n}}}$$

All else equal: as ***sample size increases***, both ***SE*** and ***CI width decrease***.

But with autocorrelated data each new datapoint is related to a previously collected datapoint and does not bring a full independent datapoint worth of information (*e.g.*, 90% autocorr. ≈ 10% new info).

i.simoes-silva@hzdr.de

Therefore, when data are autocorrelated, SEs and CIs are underestimated. Effect is usually strongest on SEs and CIs, but autocorrelation can also impact the mean.

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$$

Why does this matter?

Ignoring autocorrelation can lead to overly optimistic results, either by ***underestimating uncertainty*** (researchers might conclude that their estimates are more precise than they truly are) or by ***misleading statistical inferences*** (incorrect conclusions about significance).

It is important to distinguish two ***sample size*** concepts:

Absolute sample size

Total number of locations

n

Effective sample size

$\neq$

T + Δt

Sampling ***duration***

How long is an animal tracked for?

Sampling ***frequency***

How frequently are locations collected?

It is important to distinguish two ***sample size*** concepts:

Absolute **sample size**
Total number of locations

n

T + Δt

Sampling ***duration***

How long is an animal tracked for?

Sampling ***frequency***

How frequently are locations collected?

$\neq$

Effective **sample size**

N_{area}

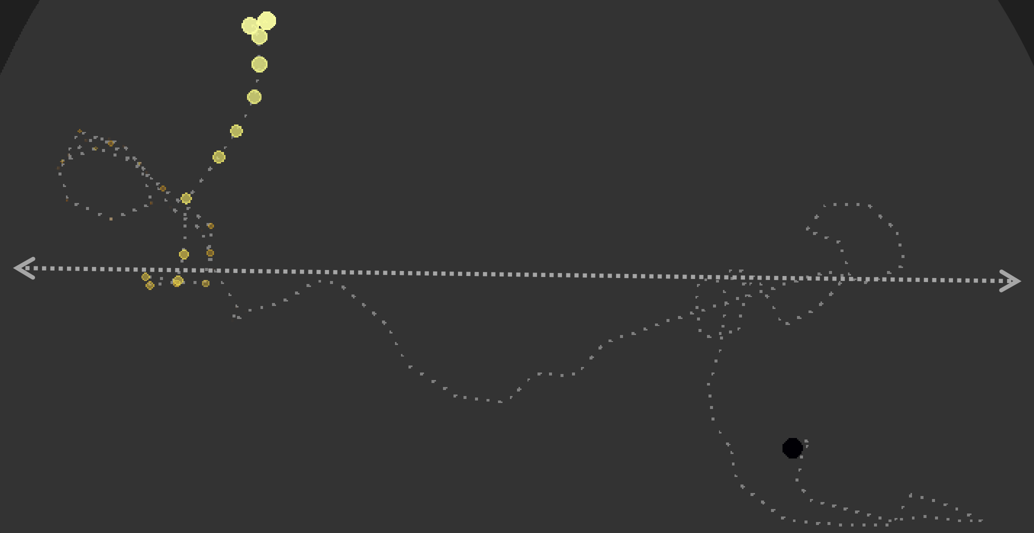
N_{speed}

roughly estimated as T/τ_p

T is the sampling duration
τ_p is the average ***home range crossing time***

τ_p

Position autocorrelation timescale



How long does it take for an animal to traverse the linear extent of its home range?

Duration = 1 day

τ_p = 1 day

τ_p

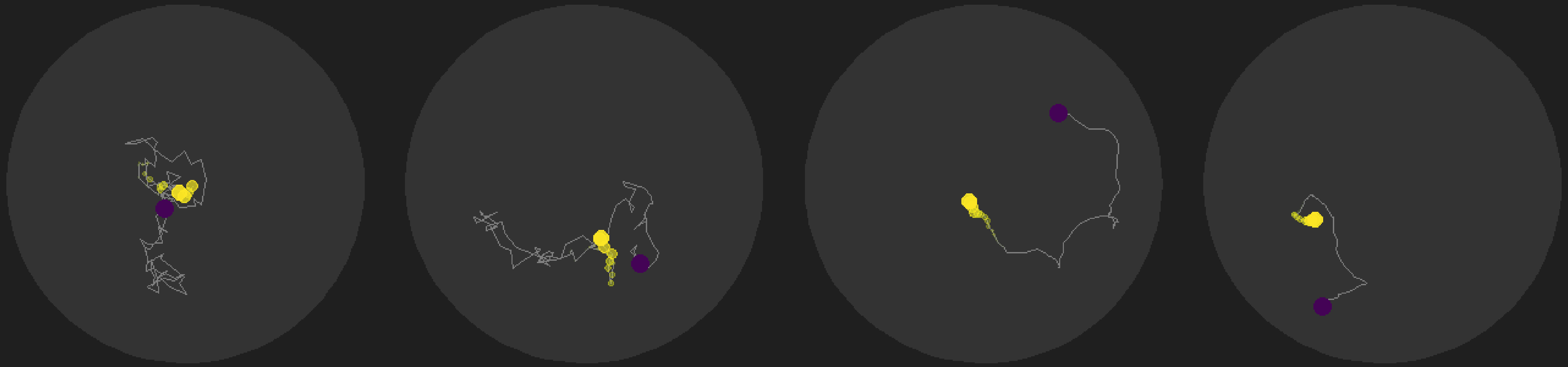
Position autocorrelation timescale

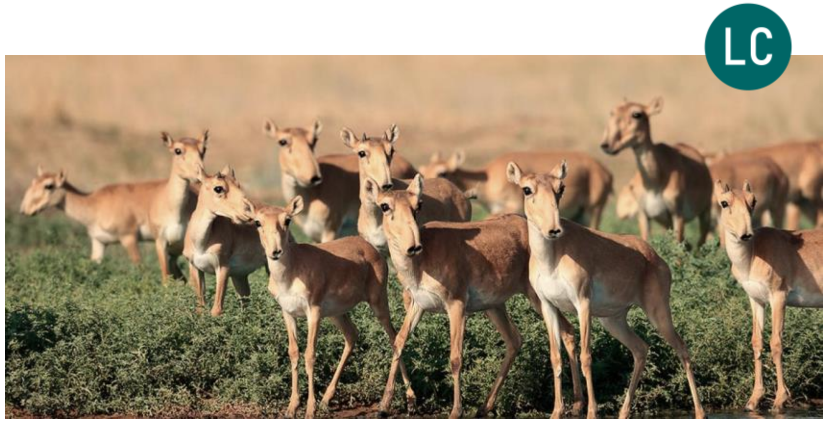
Effective sample size (N) decreases as the *home range crossing time parameter* (τ_p) increases.

τ_v

Velocity autocorrelation timescale

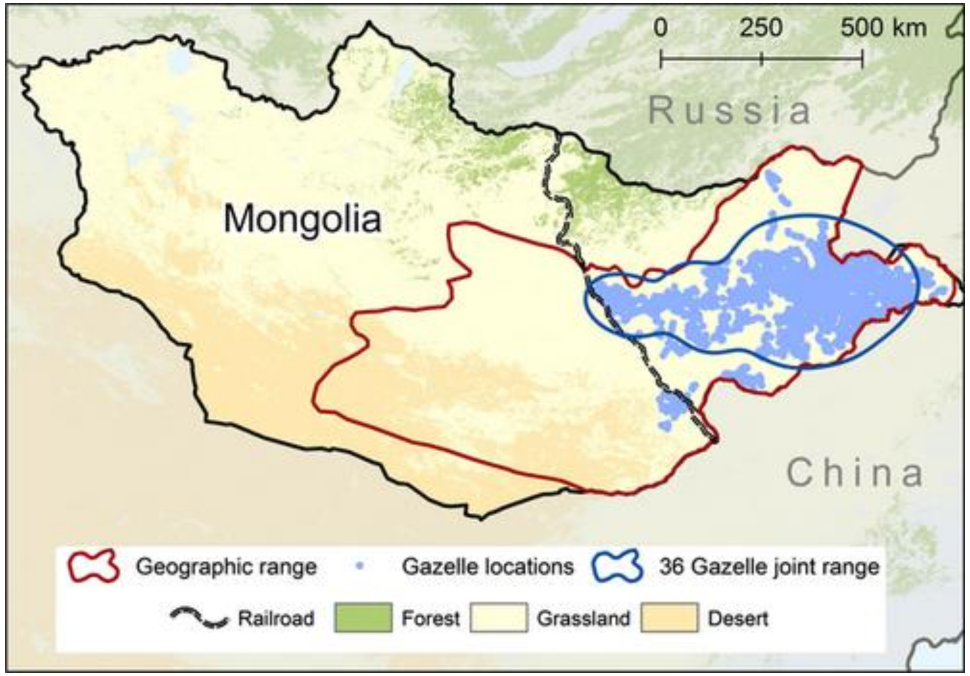
τ_v = 1 minute τ_v = 1 hour τ_v = 12 hours τ_v = 1 day





Sampling *duration* ≈ tracked for 1 year (360 days)

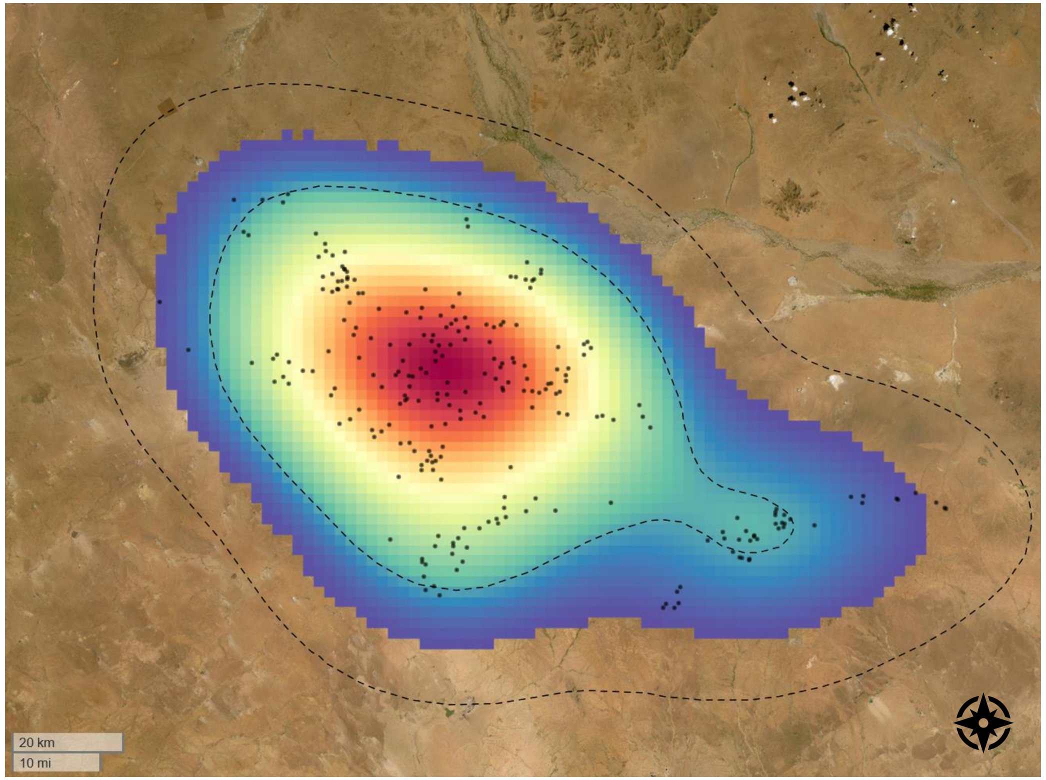
Sampling *frequency* ≈ 1 fix every hour



LC

Sampling duration ≈ tracked for 1 year (360 days)

Sampling *frequency* ≈ 1 fix every hour

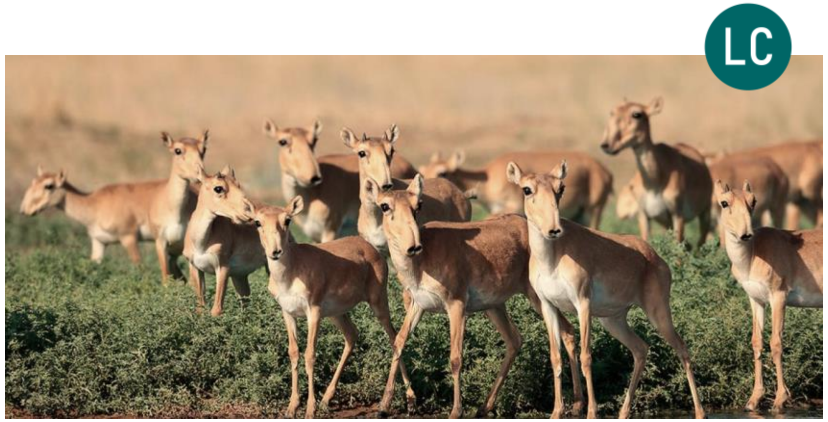


Absolute sample size, n = 866 locations

Effective sample size, $N \approx 2.4$ locations

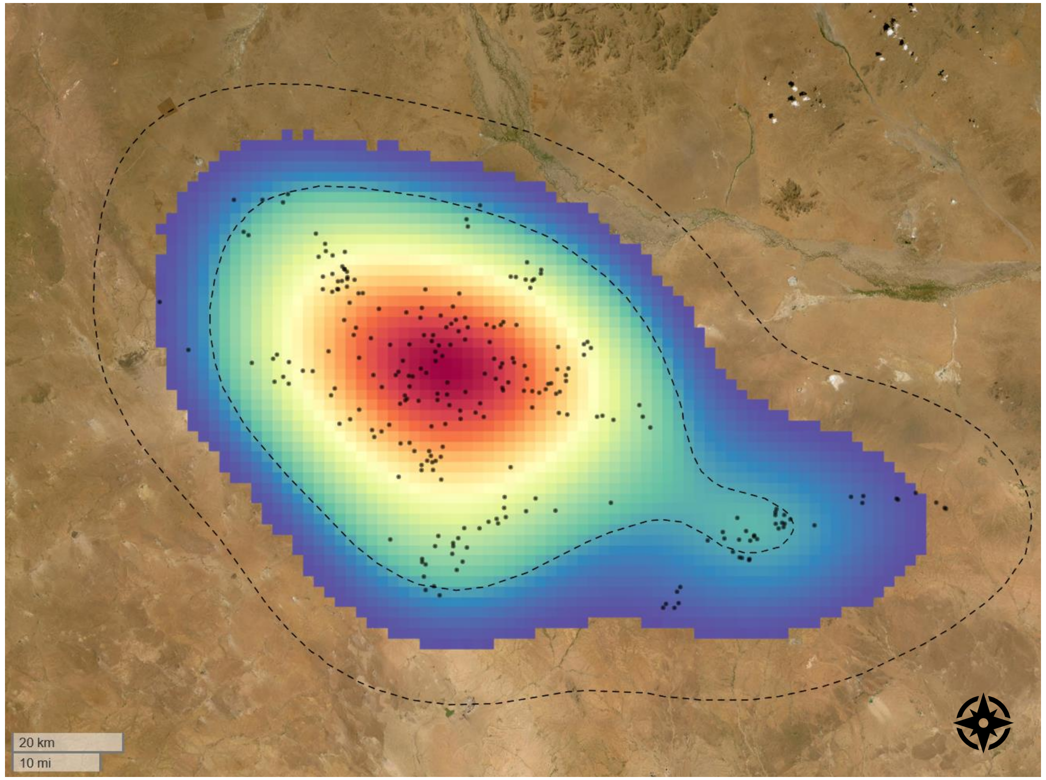
Home range area estimate

61,692.5 km^2



Sampling duration ≈ tracked for 1 year (360 days)

Sampling frequency ≈ 1 fix every hour



Absolute sample size, n = 866 locations

Effective sample size, N ≈ 2.4 locations

τ_p ≈ 141.2 days

Home range area estimate

61,692.5 km^2

For ***independent*** data,

$$n = N$$

For ***autocorrelated*** data,

$$n \gg N$$

n = *absolute* sample size
N = *effective* sample size

Many biases, including most that affect home range estimation, are exacerbated by ***small sample sizes***.

i.simoes-silva@hzdr.de

- 1. *Check range residency **assumption**;*
  Verify if the data is from a ***range-resident*** animal
- 2. *Select **movement model**;*
  Selecting the best-fit model through ***model selection***
- 3. *Estimate **home range** area;*
  Reconstructing ***range distribution*** from sampled locations

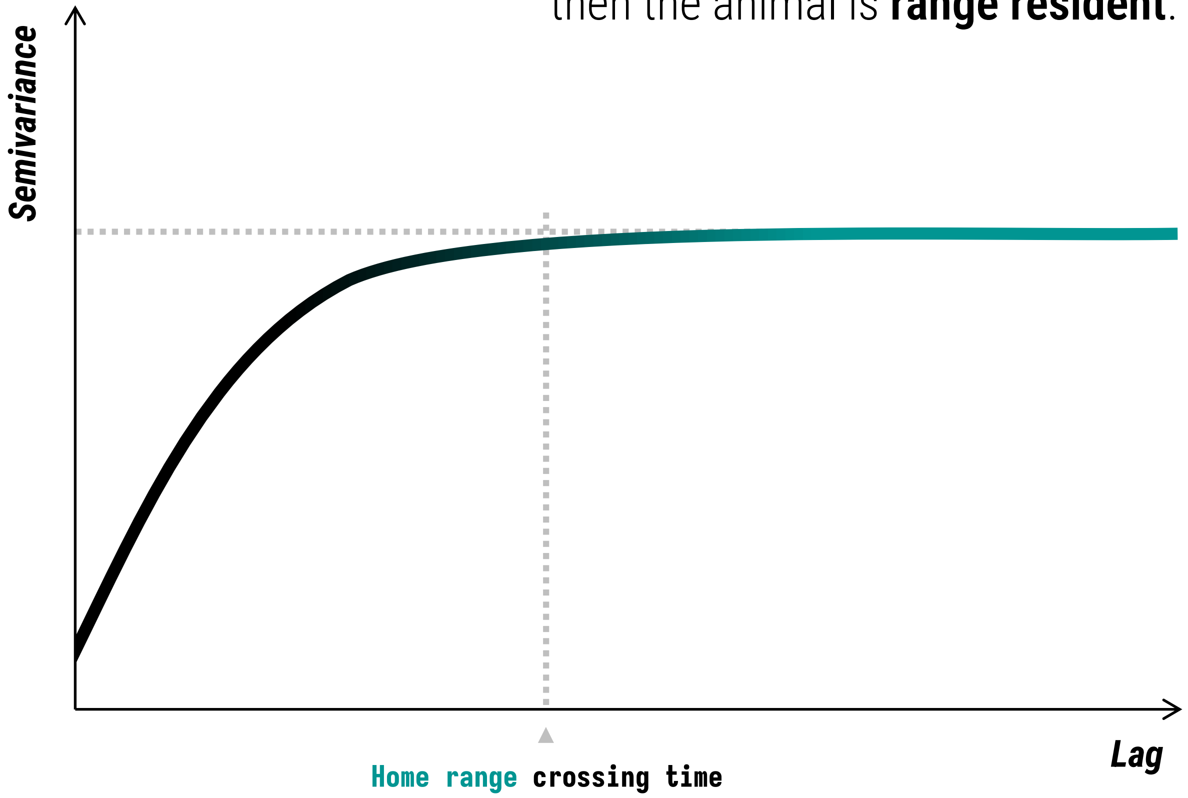
i.simoes-silva@hzdr.de

1. *Check range residency **assumption**;*
   Verify if the data is from a ***range-resident*** animal

2. *Select **movement model**;*
   Selecting the best-fit model through ***model selection***

3. *Estimate **home range** area;*
   Reconstructing ***range distribution*** from sampled locations

If semivariance reaches an asymptote,
then the animal is **range resident**.



1. *Check range residency assumption;*
   Verify if the data is from a ***range-resident*** animal

2. *Select movement model;*
   Selecting the best-fit model through ***model selection***

3. *Estimate home range area;*
   Reconstructing ***range distribution*** from sampled locations

i.simoes-silva@hzdr.de

Conventional methods:

Tracking data → Home range estimate

Assuming independent process.

Continuous-time methods:

Tracking data → **Movement model** → Home range estimate

What process best explains a given animal movement dataset?

- Independent and Identically Distributed (**IID**)
- Brownian Motion (**BM**)
- Ornstein-Uhlenbeck (**OU**)
- Integrated Ornstein-Uhlenbeck (**IOU**)
- Ornstein-Uhlenbeck with Foraging (**OUF**)

i.simoes-silva@hzdr.de

SPATIAL DEPENDENCY

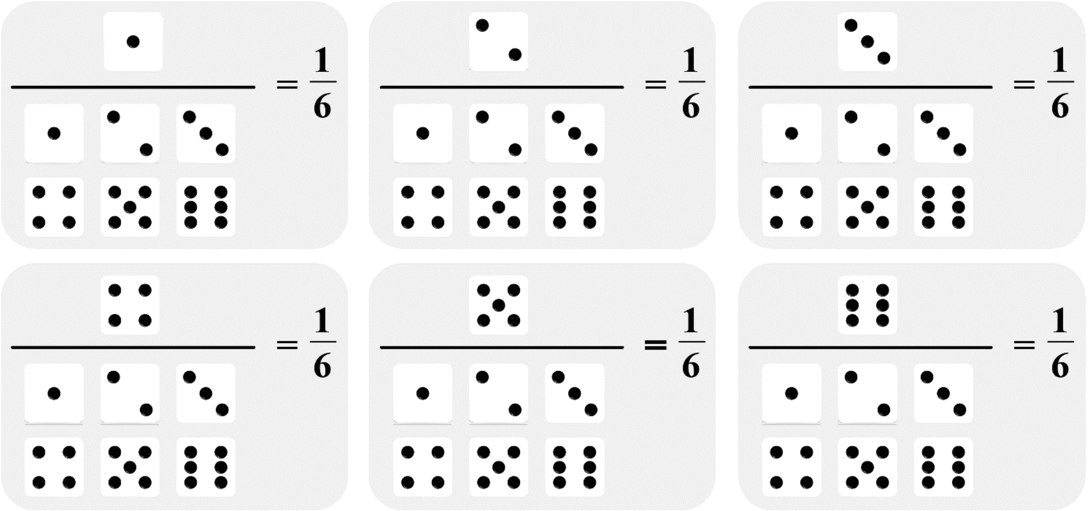
TEMPORAL DEPENDENCY

RESTRICTED

— *Independent and Identically Distributed* (IID)

Stochastic process where each location has the same probability distribution as all others, and all are ***mutually independent***.

For example,



Dice rolls are independent and identically distributed (IID).

SPATIAL DEPENDENCY

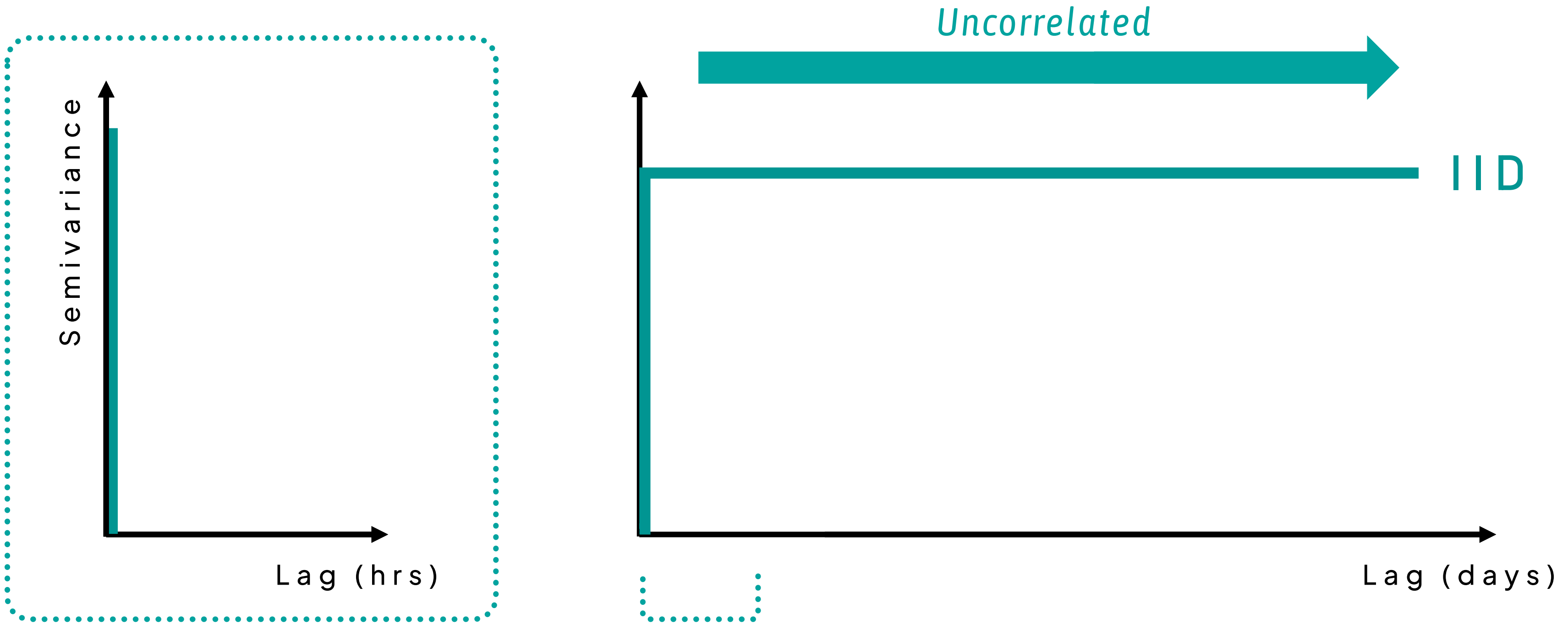
TEMPORAL DEPENDENCY

RESTRICTED

— *Independent and Identically Distributed* **(IID)**

Stochastic process where each location has the same probability distribution as all others, and all are ***mutually independent***.

How would the variogram of a IID process look like?



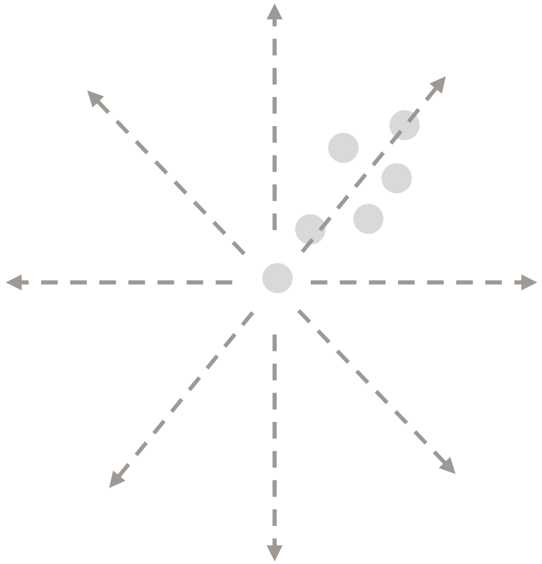
SPATIAL DEPENDENCY

TEMPORAL DEPENDENCY

RESTRICTED

— *Brownian Motion* **(BM)**

Stochastic process with stationary and independent increments, *i.e.*, no “memory” — the future behavior of a Brownian motion process does not depend on its past. Diffusion is ***constant***.



As time goes on, the animal is more likely to be further away from its starting location.

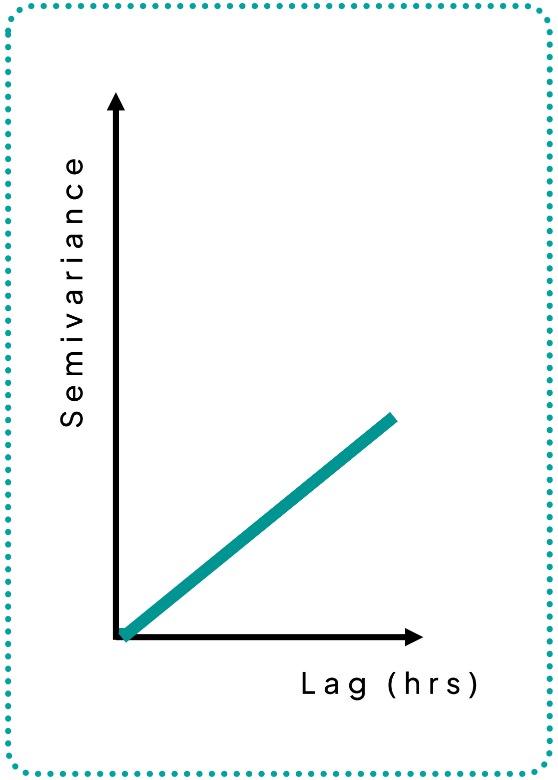
SPATIAL DEPENDENCY

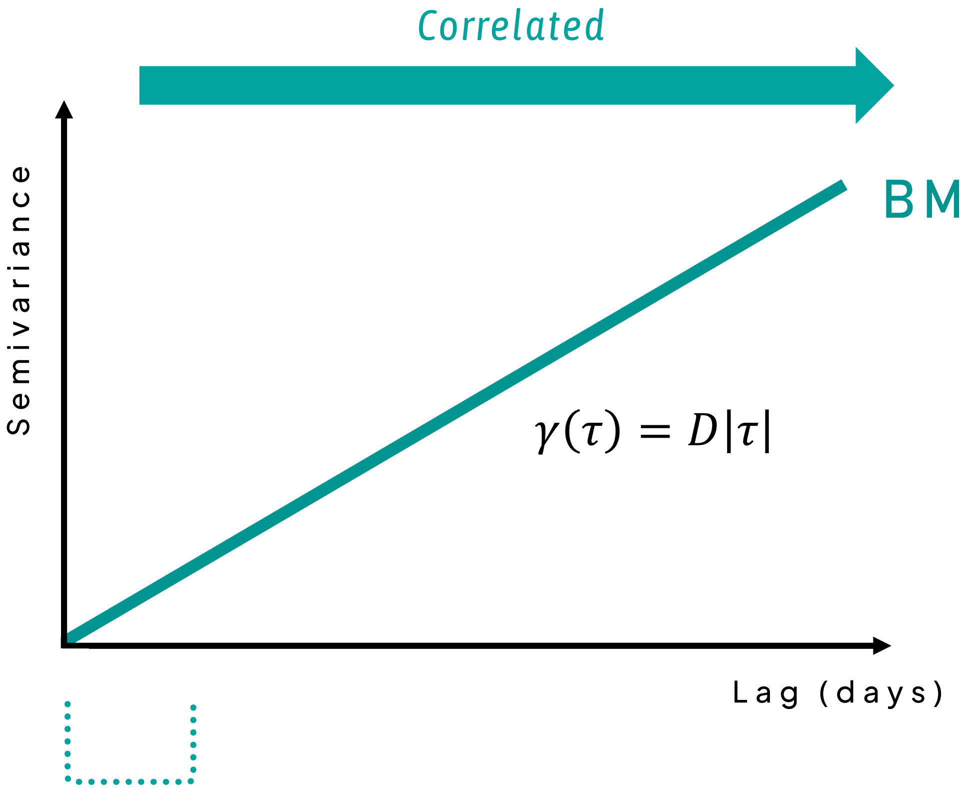
TEMPORAL DEPENDENCY

RESTRICTED

— *Brownian Motion* **(BM)**

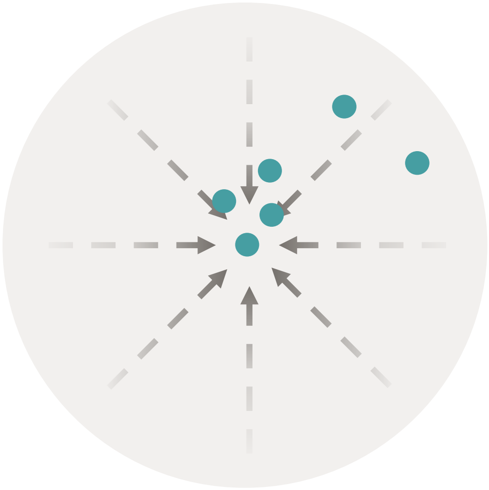
How would the variogram of a BM process look like?





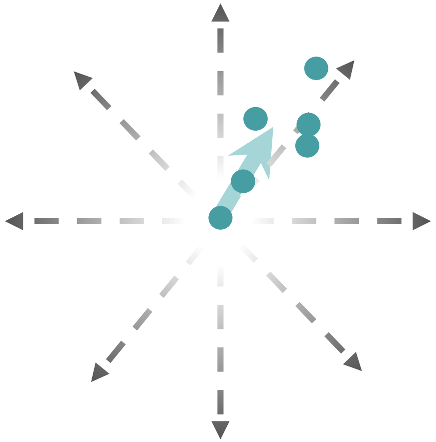
These processes are all modifications of **BM** process:

Ornstein-Uhlenbeck **(OU)**



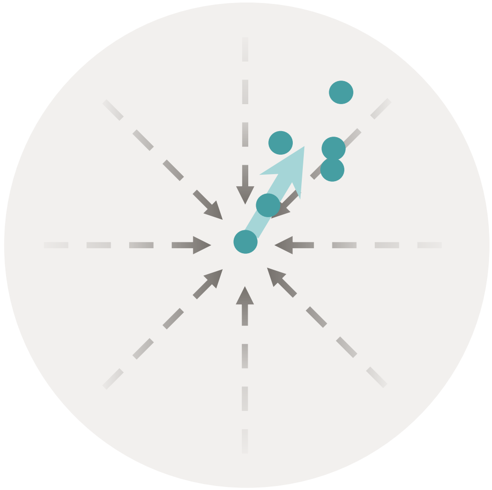
Unlike **BM**, the **OU** process tends towards a central location (***bounded diffusion***).

Integrated OU **(IOU)**



Like **BM**, the **IOU** process exhibits ***unbounded diffusion*** — but unlike **BM**, it has ***persistence of motion***.

OU with Foraging **(OUF)**



Unlike **BM**, the **OUF** process exhibits ***bounded diffusion*** and ***persistence of motion***.

i.simoes-silva@hzdr.de

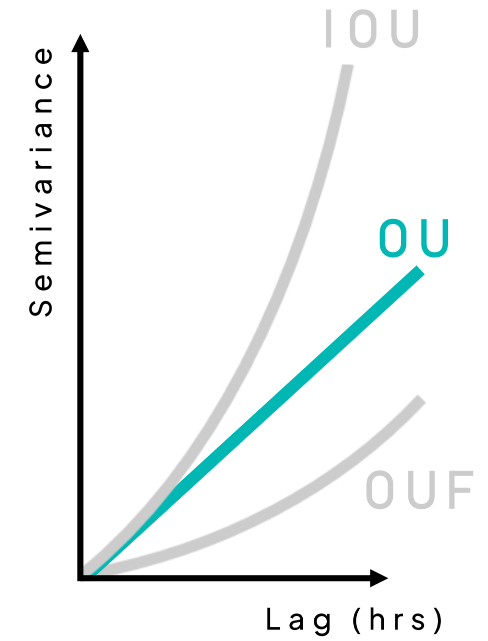
These processes are all modifications of **BM** process:

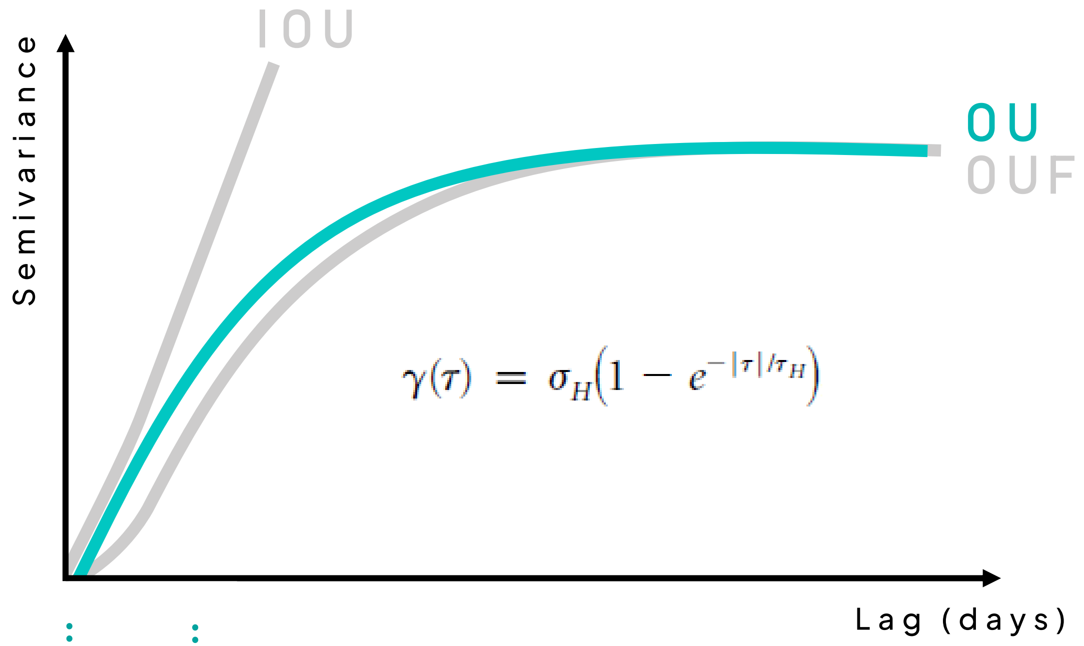
Ornstein-Uhlenbeck **(OU)**

Integrated OU **(IOU)**

OU with Foraging **(OUF)**

How would the variogram of these OU processes look like?





SPATIAL DEPENDENCY

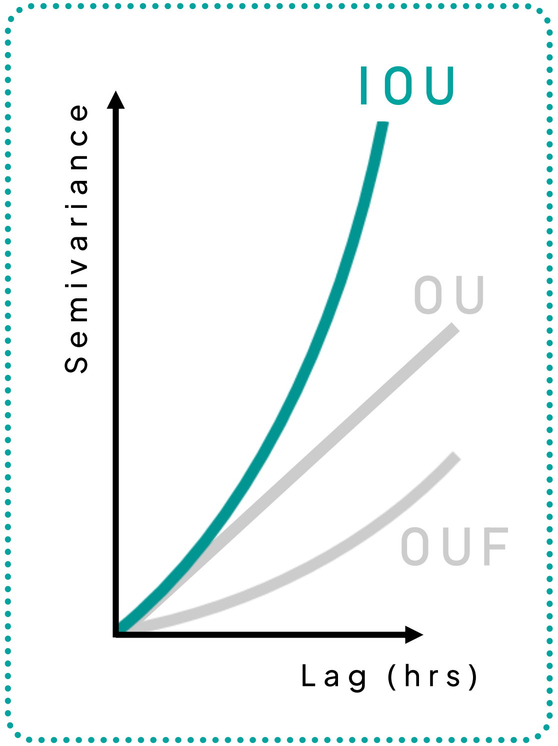
TEMPORAL DEPENDENCY

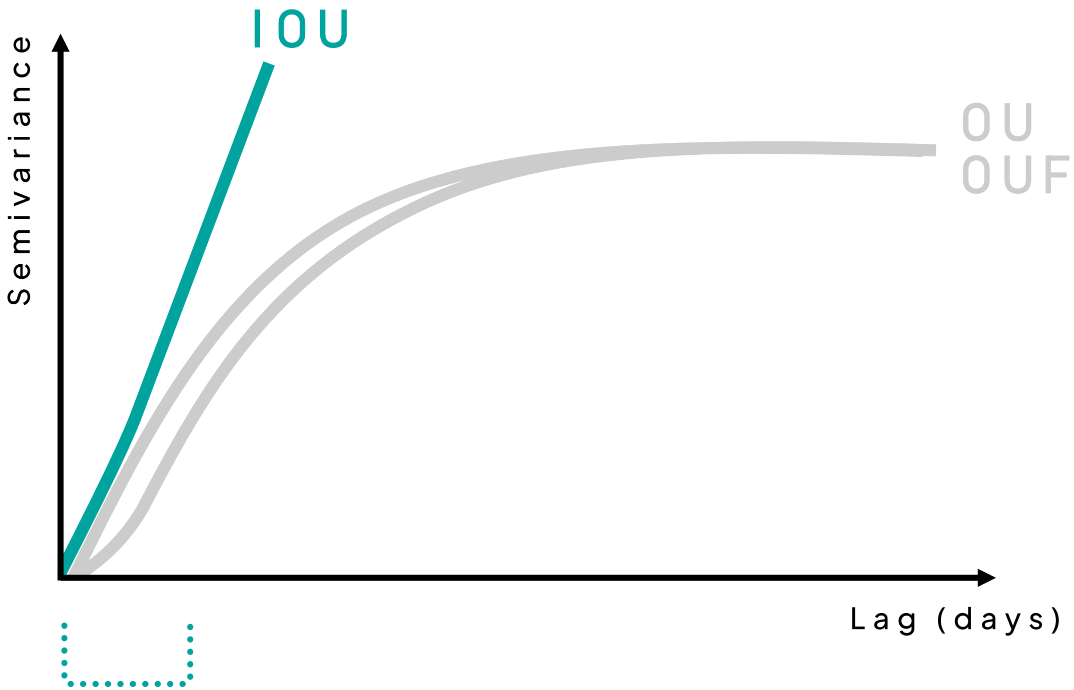
RESTRICTED

These processes are all modifications of **BM** process:

Ornstein-Uhlenbeck **(OU)** *Integrated OU* **(IOU)** *OU with Foraging* **(OUF)**

How would the variogram of these OU processes look like?





SPATIAL DEPENDENCY

TEMPORAL DEPENDENCY

RESTRICTED

i.simoes-silva@hzdr.de

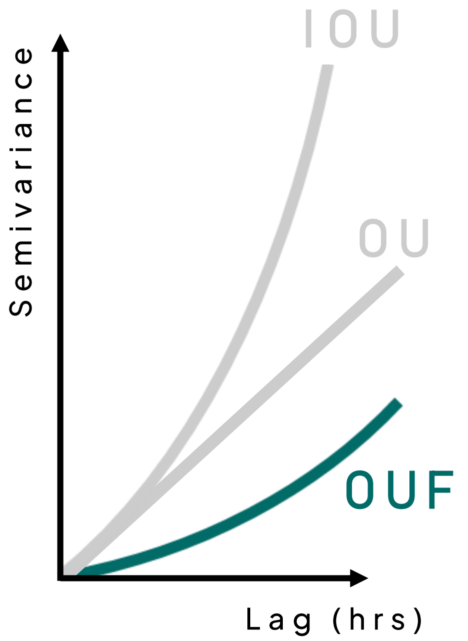
These processes are all modifications of **BM** process:

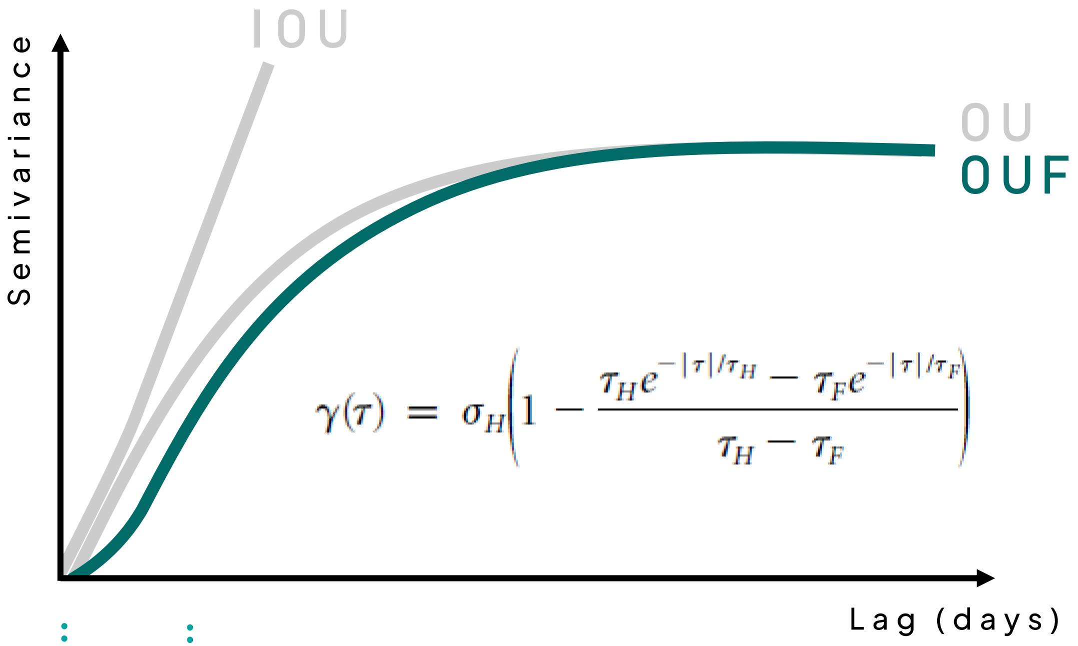
Ornstein-Uhlenbeck **(OU)**

Integrated OU **(IOU)**

OU with Foraging **(OUF)**

How would the variogram of these OU processes look like?





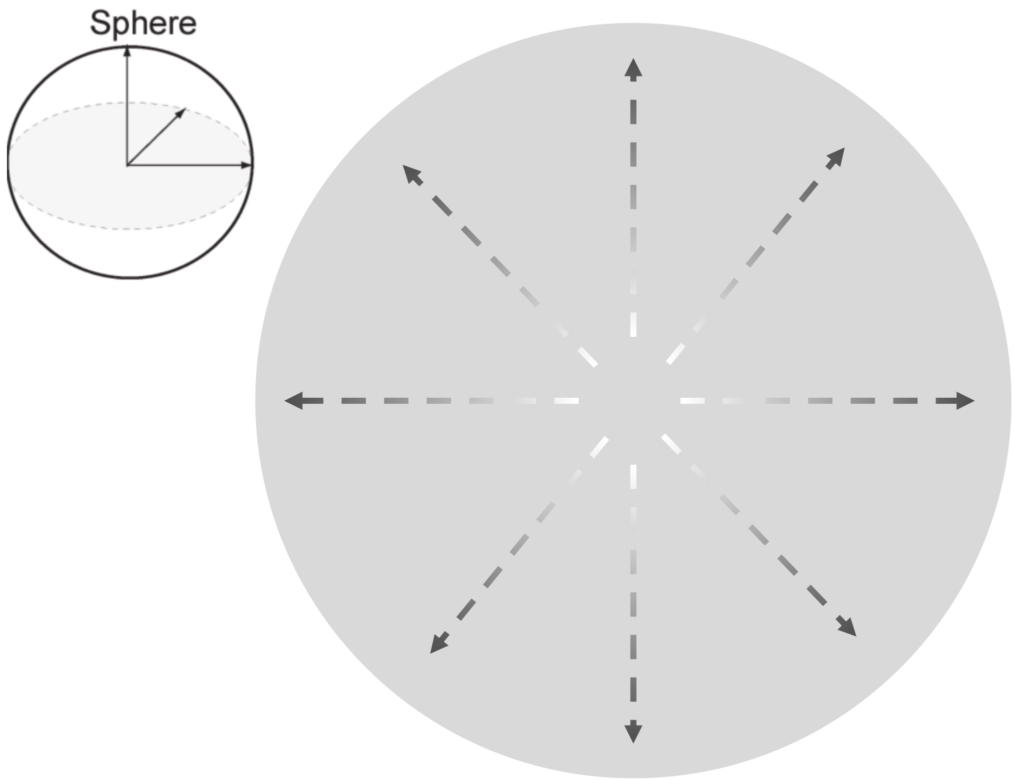
$$\gamma(\tau) = \sigma_H \left(1 - \frac{\tau_H e^{-|\tau|/\tau_H} - \tau_F e^{-|\tau|/\tau_F}}{\tau_H - \tau_F}\right)$$

SPATIAL DEPENDENCY

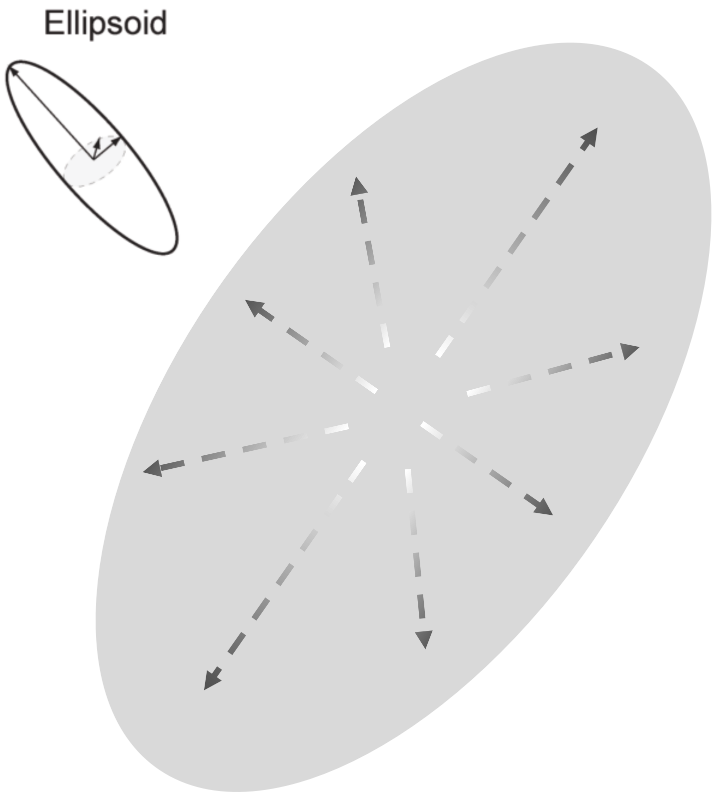
TEMPORAL DEPENDENCY

RESTRICTED

Isotropic refers to the properties of a material which is independent of the direction; whereas ***anisotropic*** is direction-dependent.



Isotropic diffusion



Anisotropic diffusion

i.simoes-silva@hzdr.de

1. *Check range residency **assumption**;*
   Verify if the data is from a ***range-resident*** animal

2. *Select **movement model**;*
   Selecting the best-fit model through ***model selection***

3. *Estimate **home range** area;*
   Reconstructing ***range distribution*** from sampled locations

i.simoes-silva@hzdr.de

Tracking data → Movement model → **Home range estimate**

Autocorrelation

Model	Position	Velocity	Restricted	Parameters
IID	No	No	Yes	τ = NULL
BM	Yes	No	No	$\tau = \infty$
OU	Yes	No	Yes	$\tau = \tau_p$
IOU	Yes	Yes	No	$\tau = \{\infty, \tau_v\}$
OUF	Yes	Yes	Yes	$\tau = \{\tau_p, \tau_v\}$

AKDE explicitly requires a movement model that accounts for ***autocorrelation***.

AKDE reduces to the **(conventional) KDE** in the limit where autocorrelation vanishes, and locations are truly ***independent***.

“ All models are wrong, but some are useful.

Box *et al.* (1987)

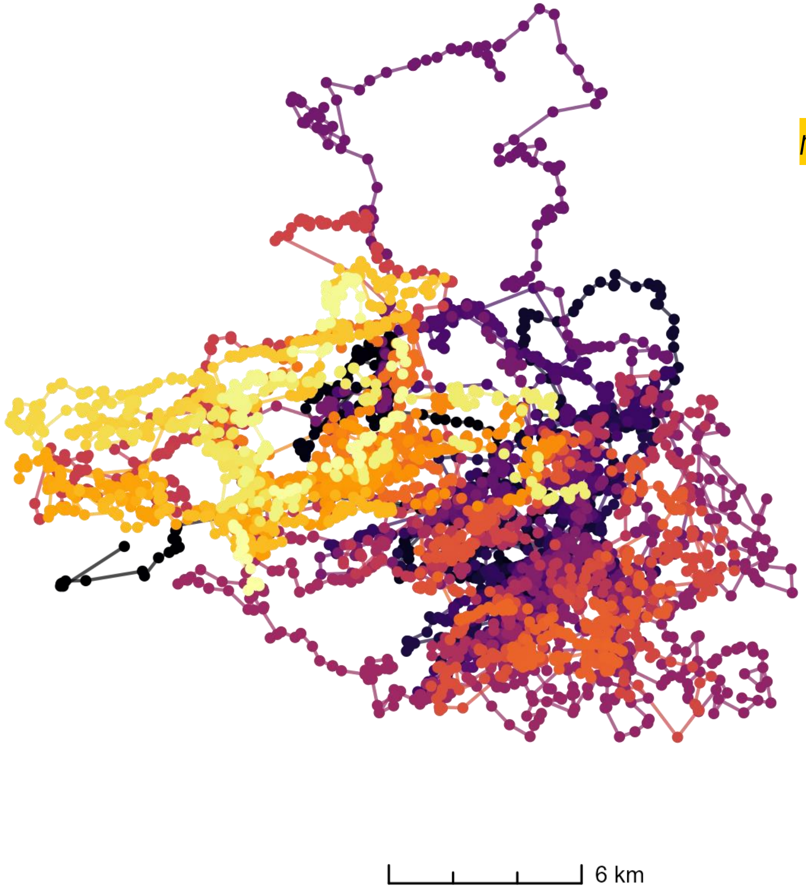
Estimate home range

Tracking time:

Start End

African Buffalo

(Syncerus caffer)



OUF process:

n = 5,677 locations

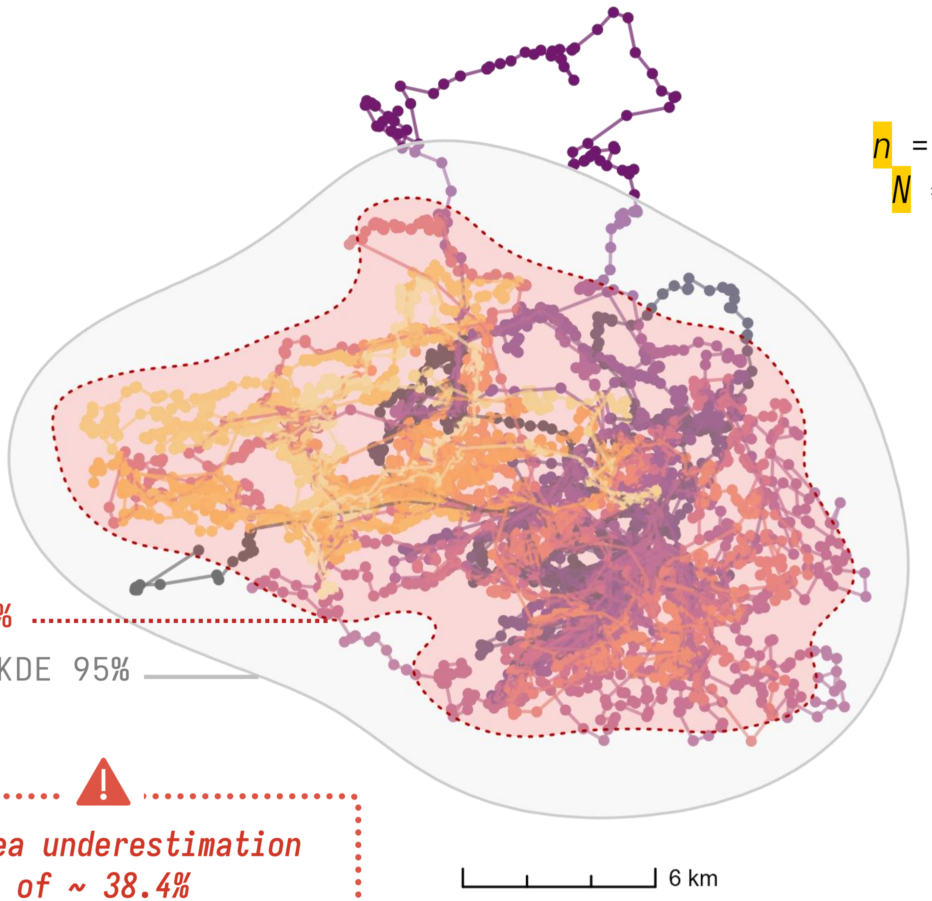
N ≈ 22.3 locations

i.simoes-silva@hzdr.de

Estimate home range

African Buffalo
(Syncerus caffer)

Tracking time:
Start End



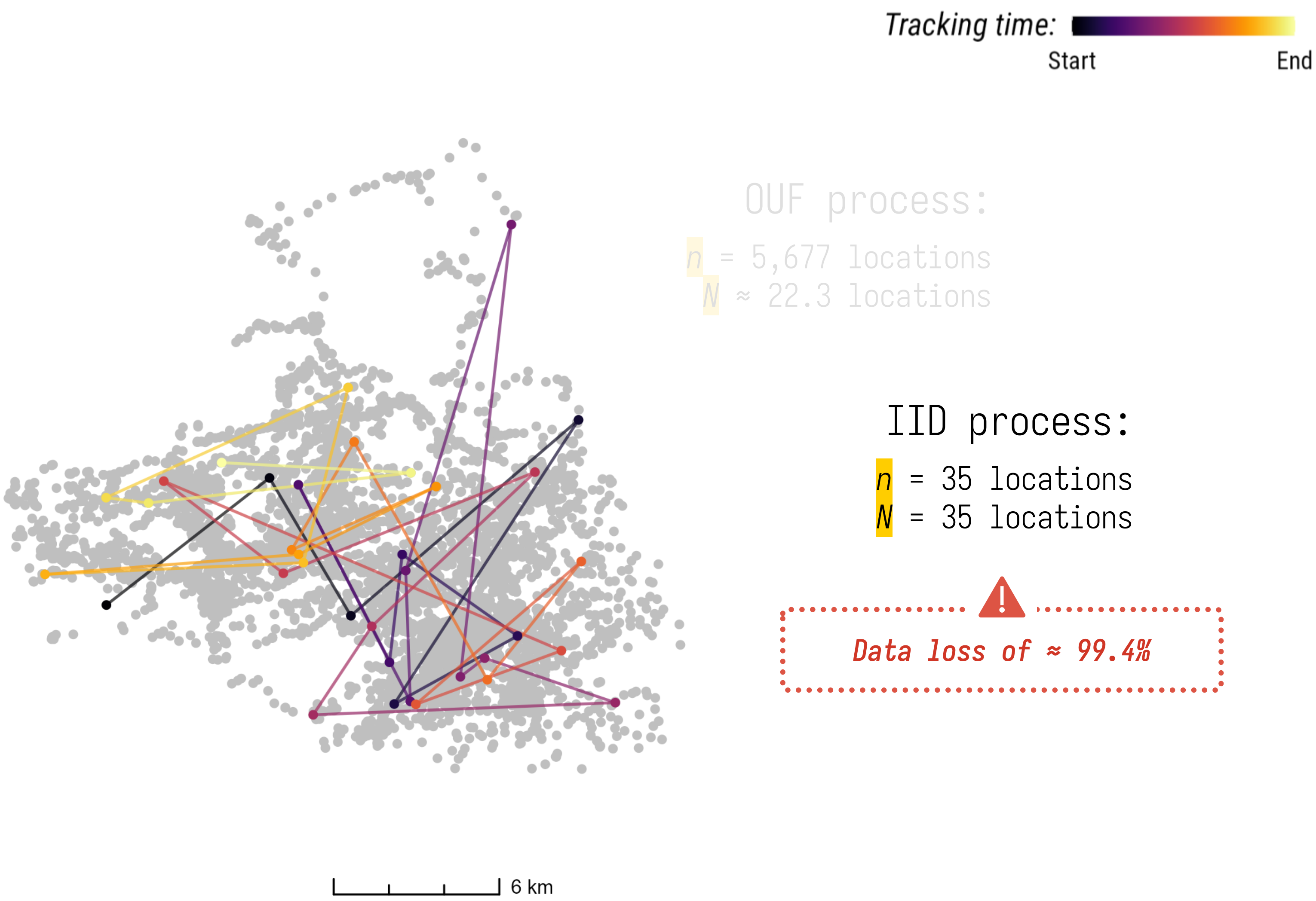
KDE 95%
AKDE 95%

OUF process:
n = 5,677 locations
N ≈ 22.3 locations

HR area underestimation of ~ 38.4%

i.simoes-silva@hzdr.de

African Buffalo
(Syncerus caffer)



OUF process:

n = 5,677 locations

N ≈ 22.3 locations

IID process:

n = 35 locations

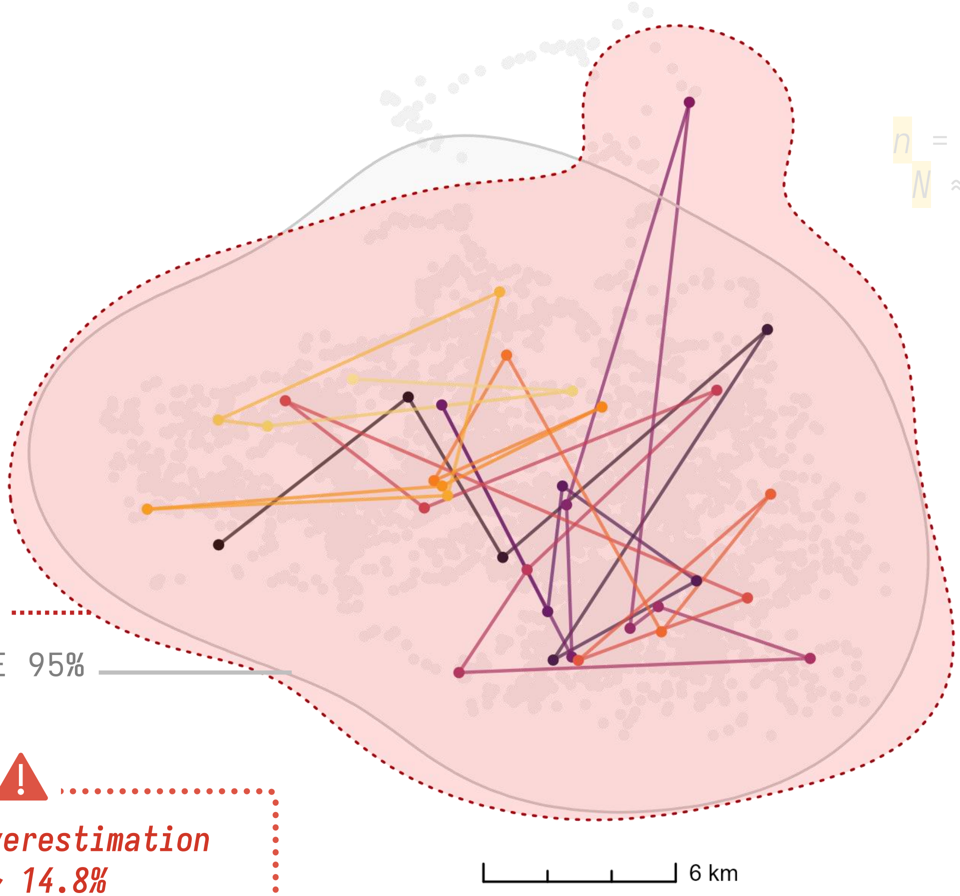
N = 35 locations

Data loss of ≈ 99.4%

Estimate home range

Tracking time: Start — End

African Buffalo
(*Syncerus caffer*)



OUF process:
n = 5,677 locations
N ≈ 22.3 locations

IID process:
n = 35 locations
N = 35 locations

Data loss of ≈ 99.4%

HR area overestimation of ~ 14.8%

1. ***Check range residency assumption;***
   Verify if the data is from a ***range-resident*** animal
2. ***Select movement model;***
   Selecting the best-fit model through ***model selection***
3. ***Estimate home range area;***
   Reconstructing ***range distribution*** from sampled locations

i.simoes-silva@hzdr.de

1. *Check range residency **assumption**;*

   Verify if the data is from a ***range-resident*** animal

2. *Select **movement model**;*

   Selecting the best-fit model through ***model selection***

3. *Estimate **home range** area;*

   Reconstructing ***range distribution*** from sampled locations

4. *Apply **mitigation measures**;*

   Accounting for ***common biases*** in movement data

i.simoes-silva@hzdr.de

Many biases, including most that affect home range estimates, are exacerbated by ***small sample sizes***. Conversely, ***large sample sizes*** in modern tracking datasets exacerbate autocorrelation.

Bias sources (in order of their general importance):		***Mitigation measures:***
Unmodelled autocorrelation	▶	AKDE
Oversmoothing	▶	AKDE_c (default)
Autocorrelation estimation bias	▶	pHREML (default)
		Parametric bootstrapping
Unrepresentative sampling in time	▶	Weighted AKDE, or wAKDE

Many biases, including most that affect home range estimates, are exacerbated by ***small sample sizes***. Conversely, ***large sample sizes*** in modern tracking datasets exacerbate autocorrelation.

Bias sources (in order of their general importance):		***Mitigation measures:***
Unmodelled autocorrelation	▶	AKDE
Oversmoothing	▶	AKDE_C (default)
Autocorrelation estimation bias	▶	pHREML (default)
		Parametric bootstrapping
Unrepresentative sampling in time	▶	Weighted AKDE, or wAKDE

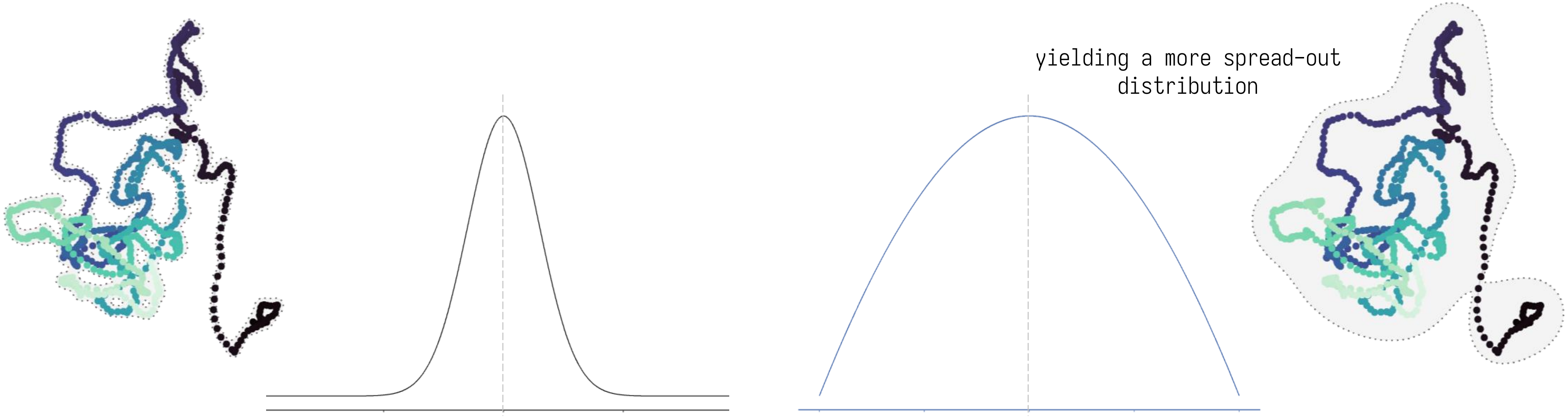
i.simoes-silva@hzdr.de

Tracking data → Movement model → **Home range estimate**

Area-corrected AKDE or **AKDE$_C$**

Deals with: ***oversmoothing***

Even when we account for autocorrelation, **GRF-KDEs** remain biased due to the natural tendency of the **GRF** approximation to ***oversmooth***.



i.simoes-silva@hzdr.de

Tracking data ····> Movement model ····> **Home range estimate**

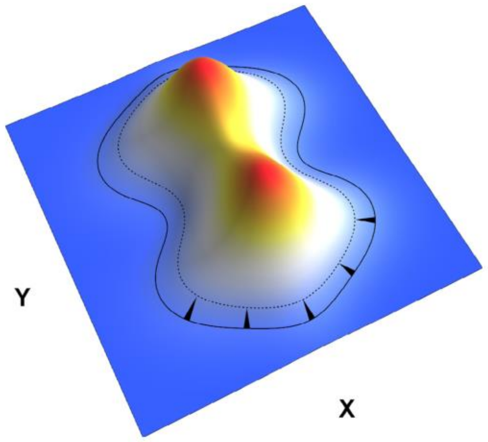
Area-corrected AKDE or **AKDE$_C$**

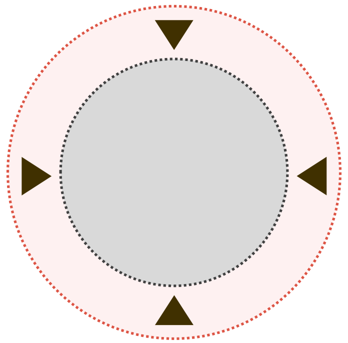
Deals with: ***oversmoothing***

Even when we account for autocorrelation, **GRF-KDEs** remain biased due to the natural tendency of the **GRF** approximation to **oversmooth**.

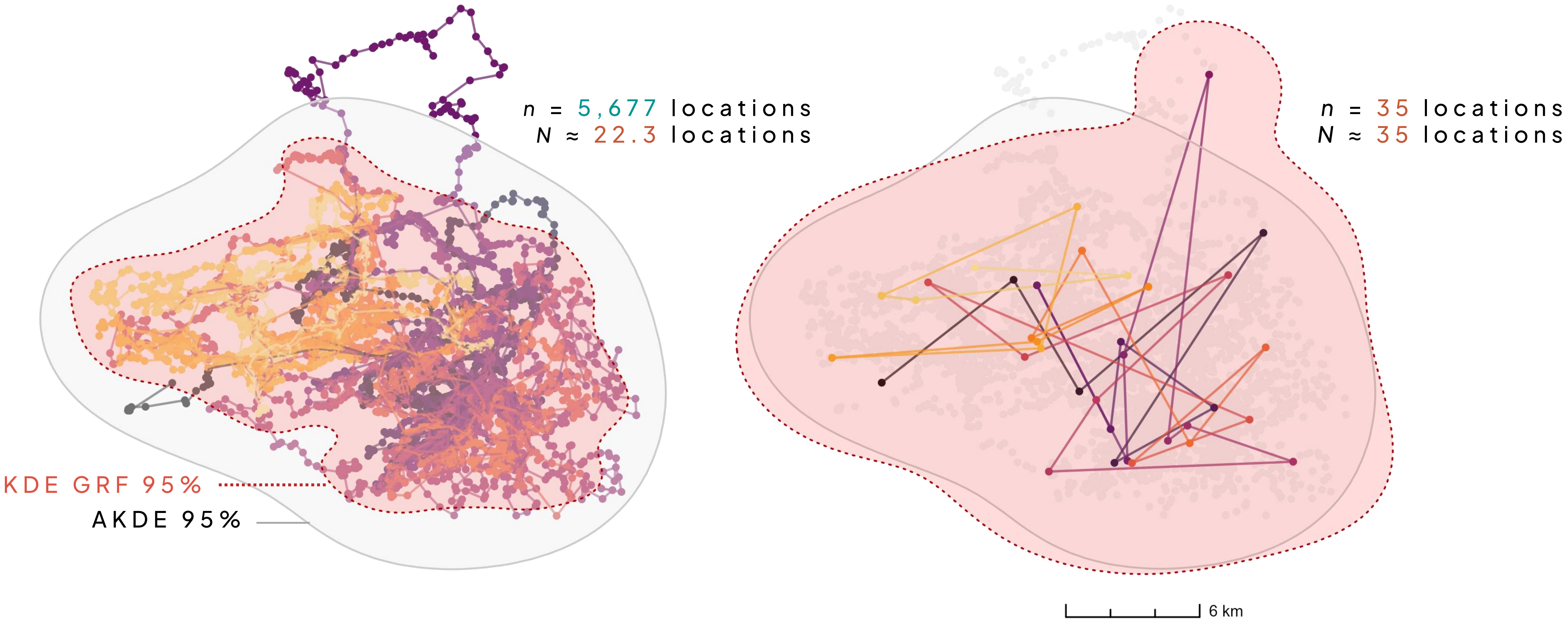
Derived an ***improved (A)KDE*** that pulls the contours of the location distribution estimate inward towards the data without distorting its shape.

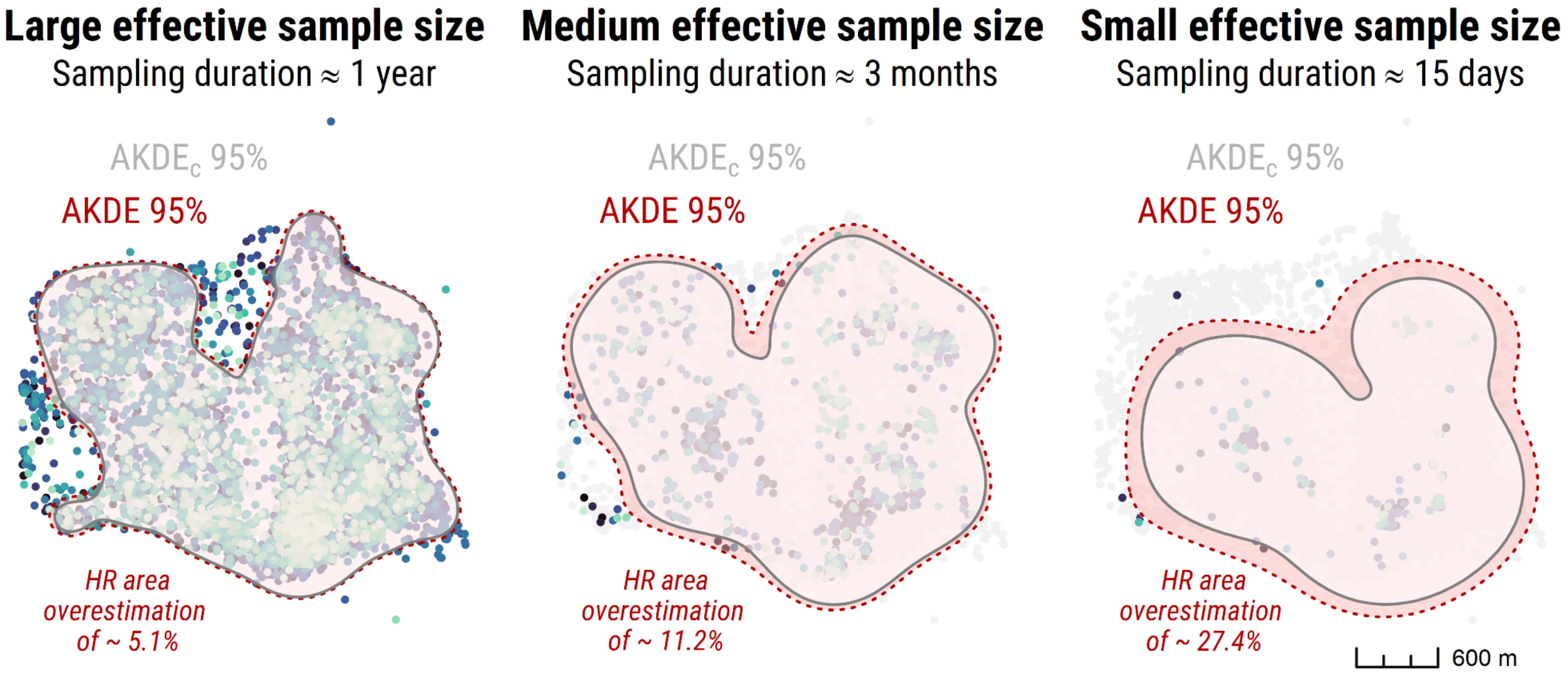
Fleming & Calabrese (2017)





The oversmoothing (positive) bias can be ***masked*** by the often-stronger negative bias caused by unmodeled autocorrelation.



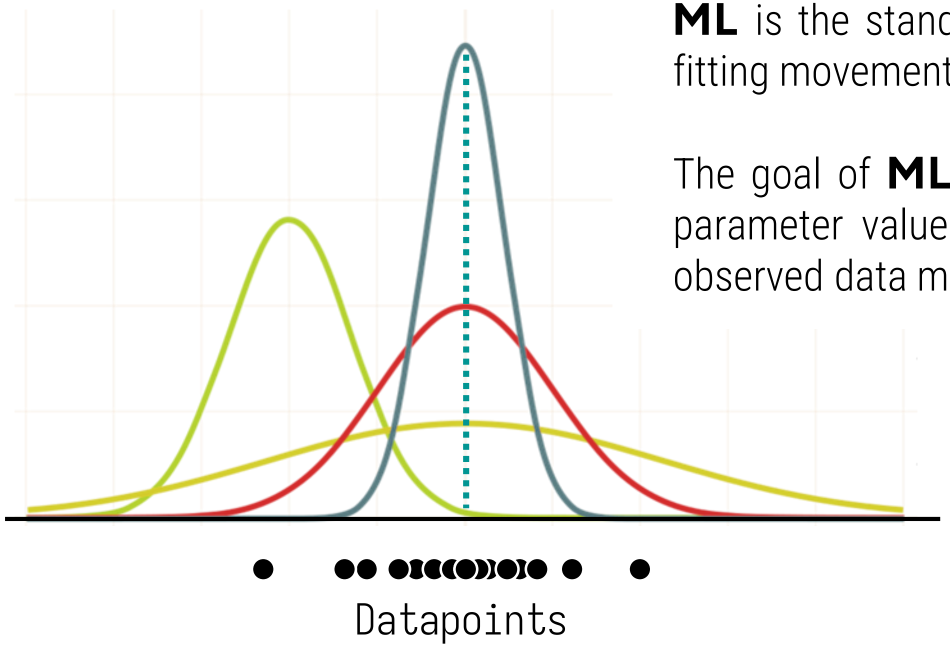


i.simoes-silva@hzdr.de

Tracking data → **Movement model** → Home range estimate

What ***movement model parameters*** are most likely to characterize a given tracking dataset?

— ***Maximum Likelihood* (ML)**:



ML is the standard approach to fitting movement models.

The goal of **ML** is to select the parameter values that make the observed data most probable.

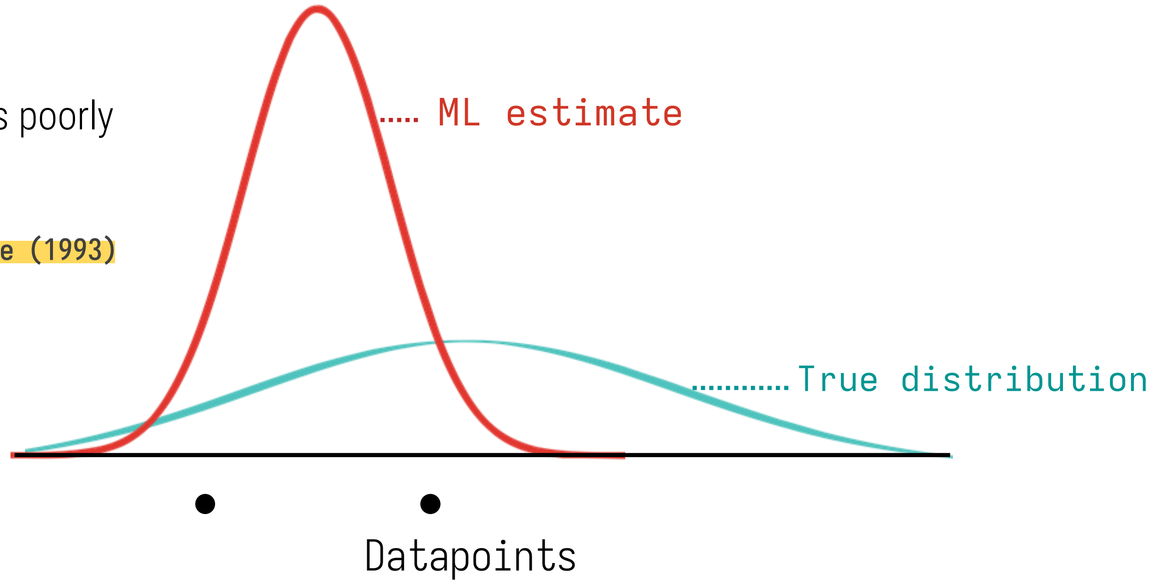
Tracking data → **Movement model** → Home range estimate

What ***movement model parameters*** are most likely to characterize a given tracking dataset?

— ***Maximum Likelihood* (ML)**:

Unfortunately, **ML** performs poorly at ***small sample sizes***.

Cressie (1993)



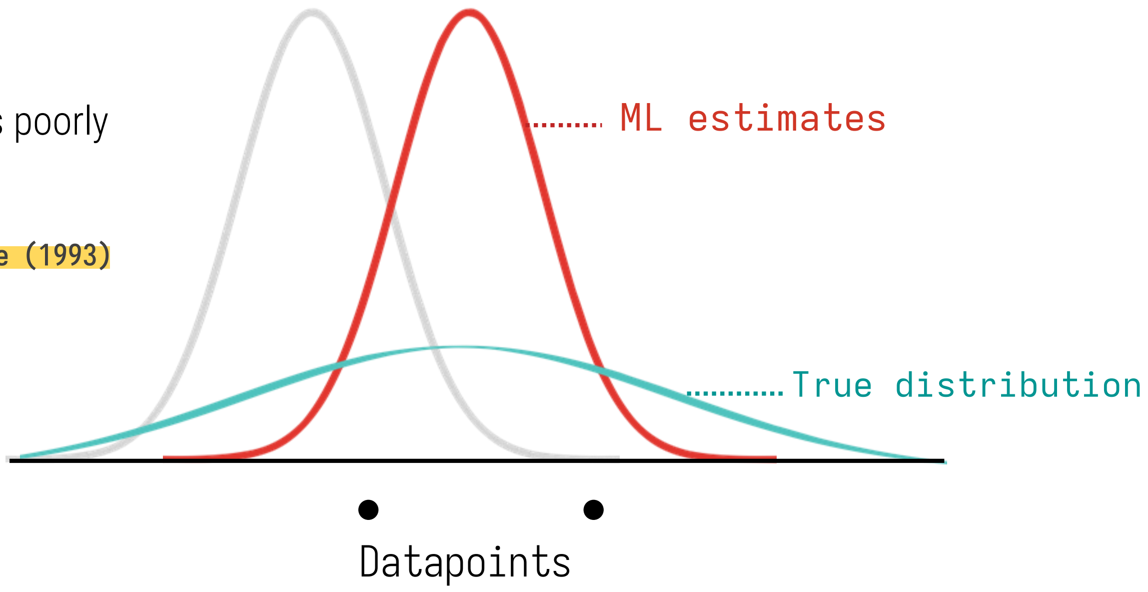
Tracking data ⇢ **Movement model** ⇢ Home range estimate

What ***movement model parameters*** are most likely to characterize a given tracking dataset?

— *Maximum Likelihood* **(ML)**:

Unfortunately, **ML** performs poorly at ***small sample sizes***.

Cressie (1993)



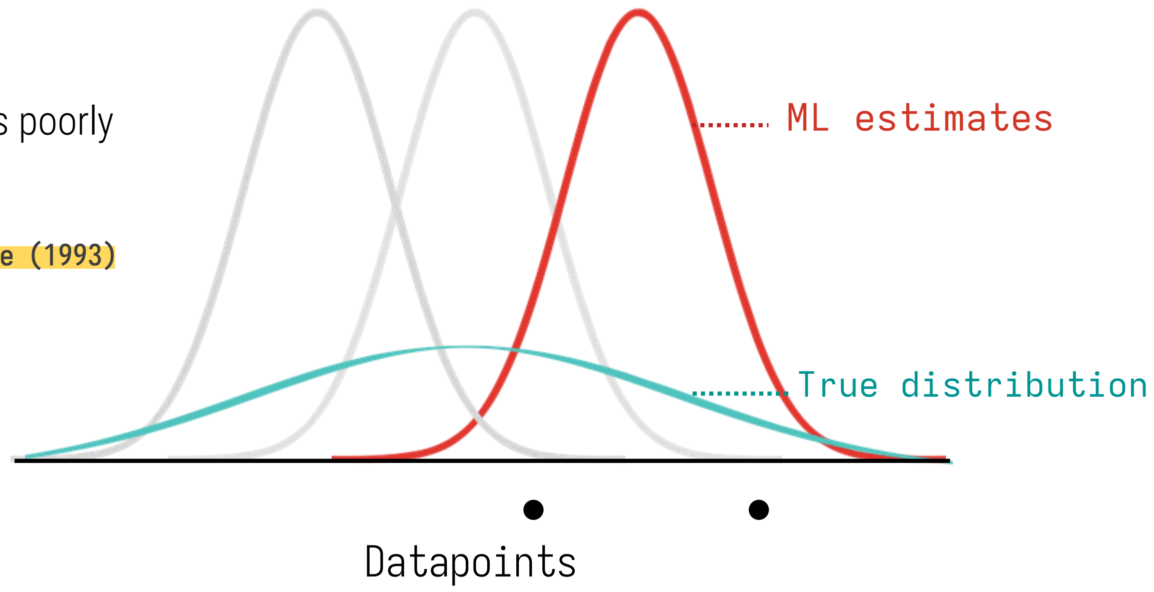
Tracking data → **Movement model** → Home range estimate

What ***movement model parameters*** are most likely to characterize a given tracking dataset?

— ***Maximum Likelihood* (ML)**:

Unfortunately, **ML** performs poorly at ***small sample sizes***.

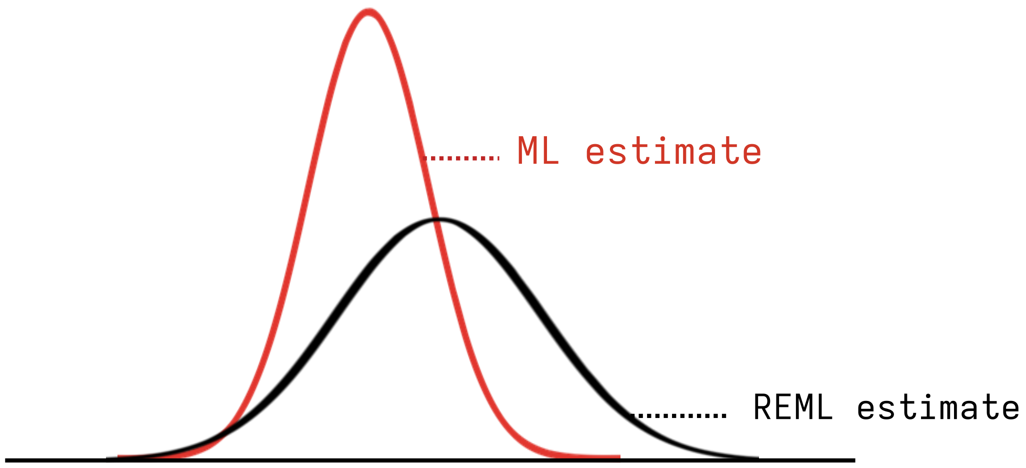
Cressie (1993)



We can tackle the **ML** bias in several ways.
For example:

Residual **ML** (or **REML**)

Widely used method for reducing bias in **ML** variance estimation, by maximizing the likelihood of residuals rather than the data.
Essentially, it prioritizes ***reducing bias*** at the ***cost of increasing variability*** in parameter estimates.

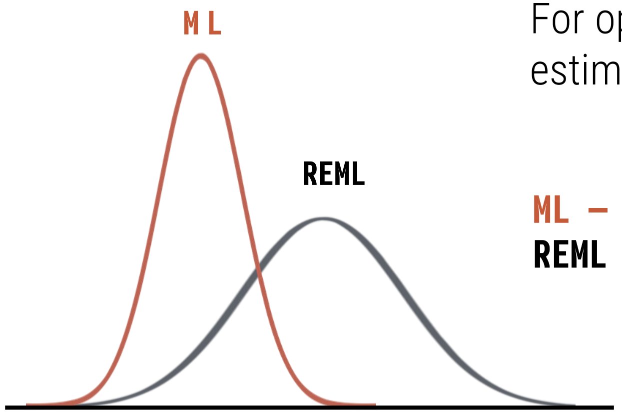


Bartlett (1937)

Tracking data → Movement model → **Home range estimate**

pHREML AKDE

Deals with: ***autocorrelation estimation bias***



For optimal performance, we need to estimate autocorrelation ***correctly***.

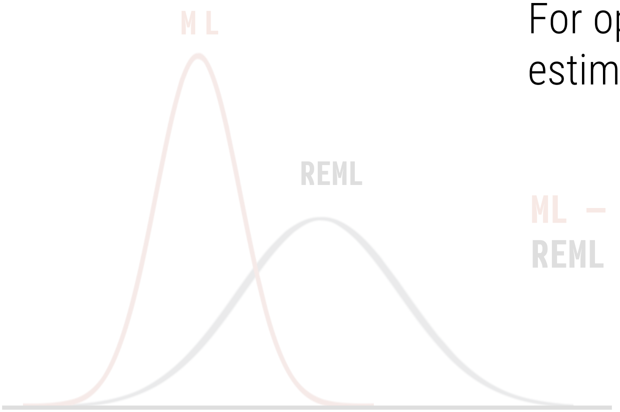
ML – performs poorly at ***small*** **sample sizes**.
REML – performs poorly at ***small effective*** **sample sizes**.

Tracking data → Movement model → **Home range estimate**

pHREML AKDE

Deals with: ***autocorrelation estimation bias***

For optimal performance, we need to estimate autocorrelation ***correctly***.



ML – performs poorly at *small* **sample sizes**.
REML – performs poorly at *small* ***effective*** **sample sizes**.

Focus on:

small ***effective*** sample sizes — perturbative REML (**pREML**)
small ***absolute*** sample sizes — Hybrid REML (**HREML**)
small ***absolute*** and ***effective*** sample sizes — perturbative Hybrid REML (**pHREML**)

As such, we consider other parameter estimation methods:

- perturbative REML (**pREML**)
- Hybrid REML (**HREML**)
- perturbative Hybrid REML (**pHREML**)

Fleming et al. (2019)

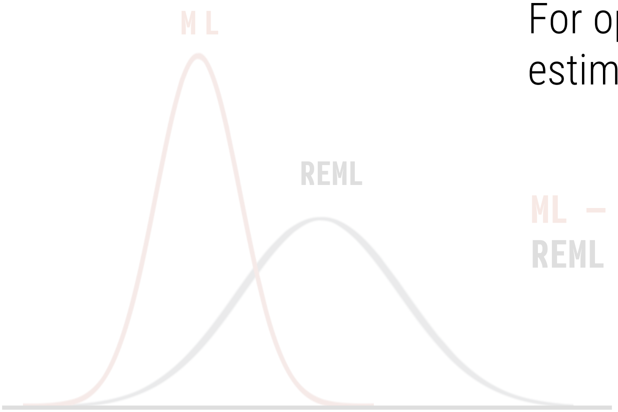
i.simoes-silva@hzdr.de

Tracking data → Movement model → ***Home range estimate***

pHREML AKDE

Deals with: ***autocorrelation estimation bias***

For optimal performance, we need to estimate autocorrelation ***correctly***.



ML – performs poorly at ***small*** **sample sizes**.
REML – performs poorly at ***small effective*** **sample sizes**.

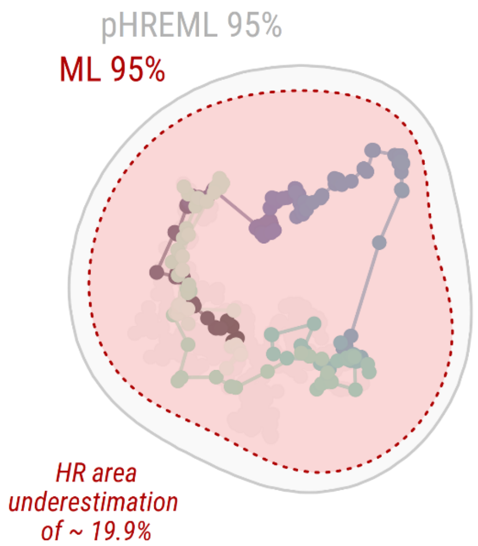
As such, we consider other parameter estimation methods:

Focus on:

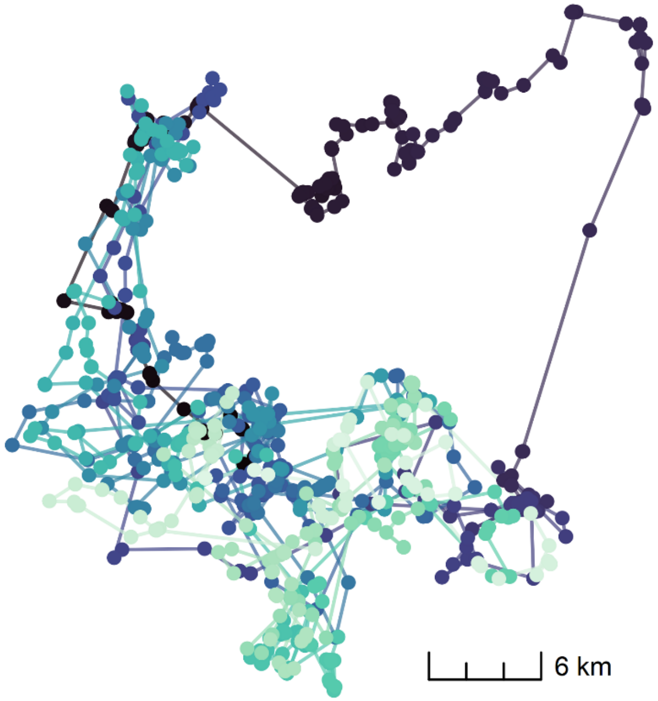
small ***effective*** sample sizes — perturbative REML (**pREML**)
small ***absolute*** sample sizes — Hybrid REML (**HREML**)
small ***absolute*** and ***effective*** sample sizes — perturbative Hybrid REML (**pHREML**)

Fleming et al. (2019)

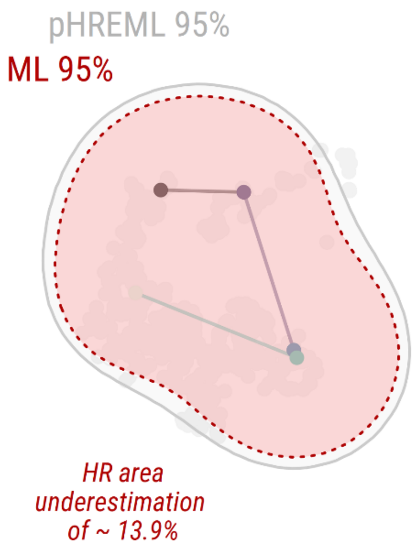
Large absolute &
Small effective sample size
$n = 363 \quad N \approx 3.1$



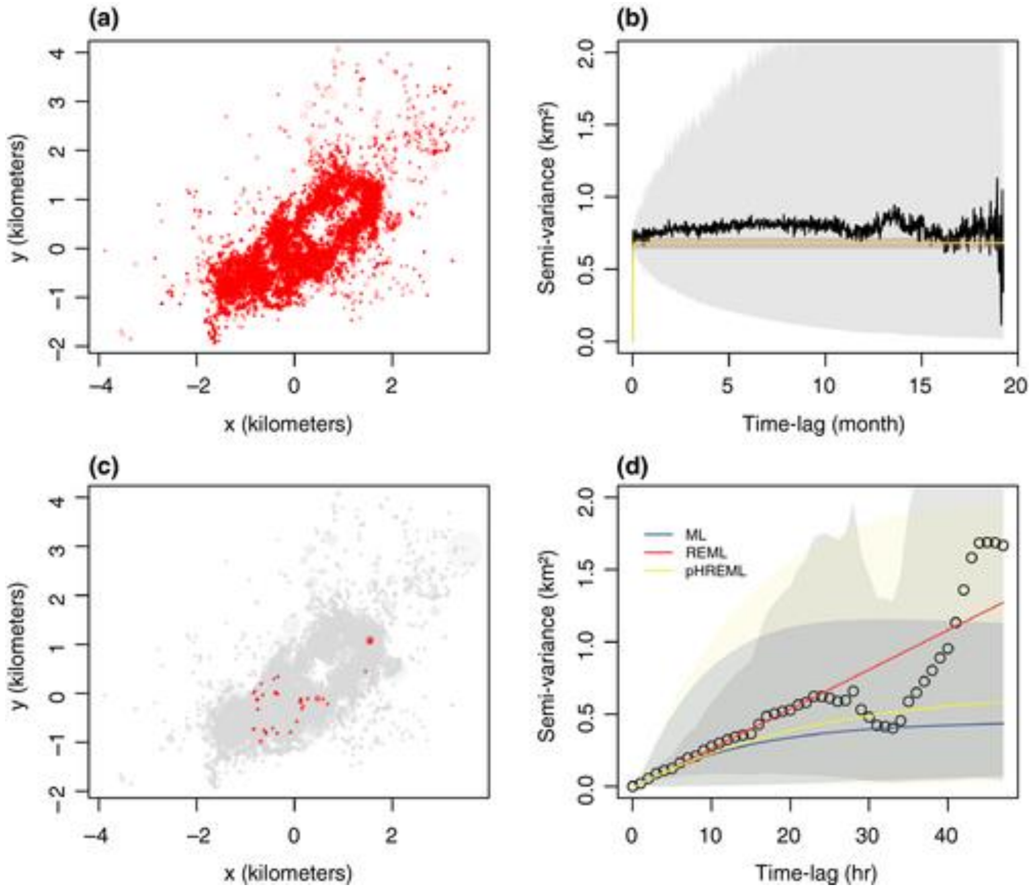
$n = 1010 \quad N \approx 14.5$



Small absolute &
Small effective sample size
$n = 5 \quad N \approx 4$

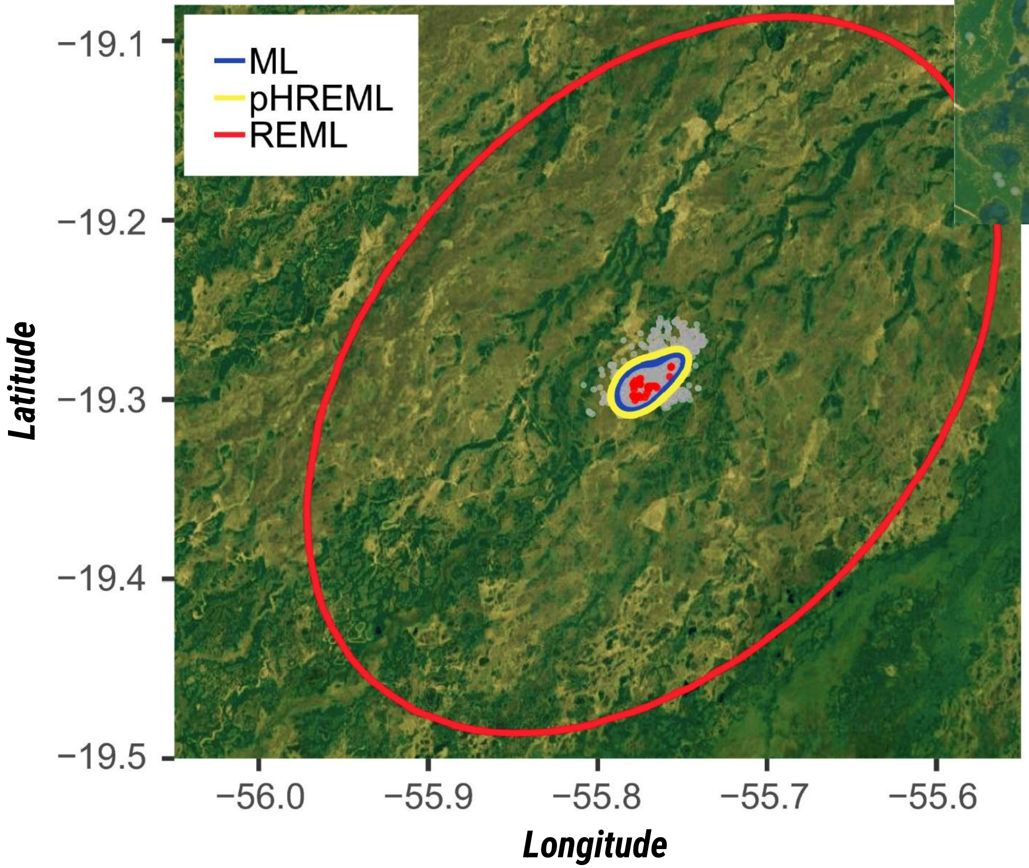


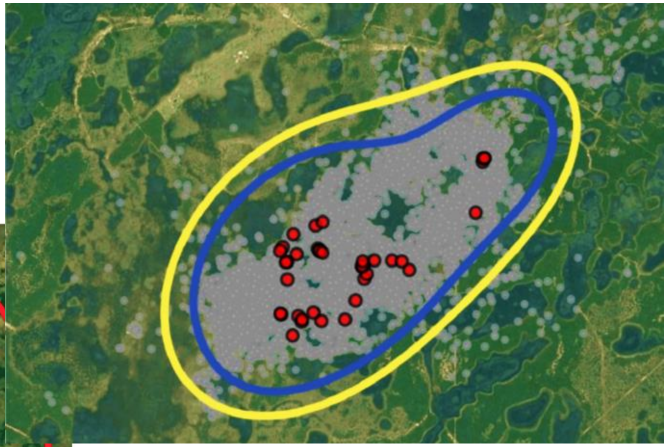
Tracking data (1-hr intervals for *19 months*), reduced to *2 days*.



Lowland tapir
(Tapirus terrestris)

Tracking data (1-hr intervals for *19 months*), reduced to *2 days*.





ML
pHREML
REML

Lowland tapir
(Tapirus terrestris)

Tracking data → Movement model → *Home range estimate*

Weighted AKDE or **wAKDE**

Deals with: ***unrepresentative sampling in time***

For optimal performance, we need to estimate autocorrelation ***correctly***.

- Many real-world issues can lead to irregular sampling:

  duty-cycling tags to avoid wasting battery,
  acceleration-informed sampling,
  device malfunction,
  habitat-related signal loss,
  and other causes.

- Shifting sampling schedules (based on behavioral or seasonal patterns) is also a common strategy.

wAKDE optimally ***upweights*** observations that occur during ***under-sampled times***, while optimally ***downweighting*** observations occurring during ***over-sampled times***.

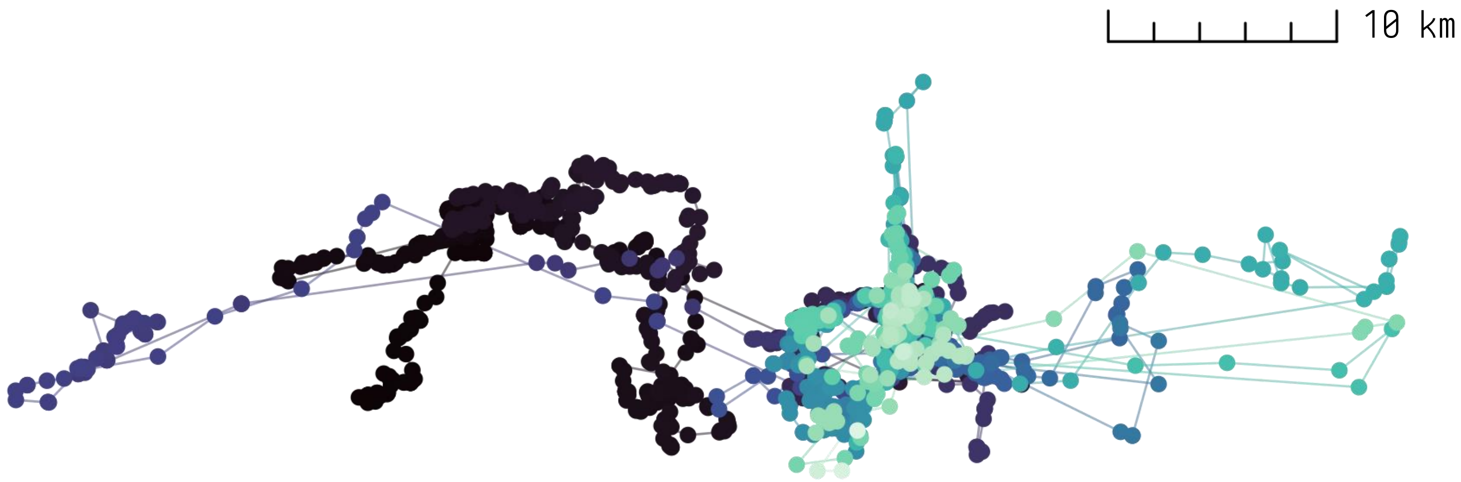


Fig. African buffalo dataset (nicknamed “Pepper”) with an ***irregular*** sampling schedule.

wAKDE optimally ***upweights*** observations that occur during ***under-sampled times***, while optimally ***downweighting*** observations occurring during ***over-sampled times***.

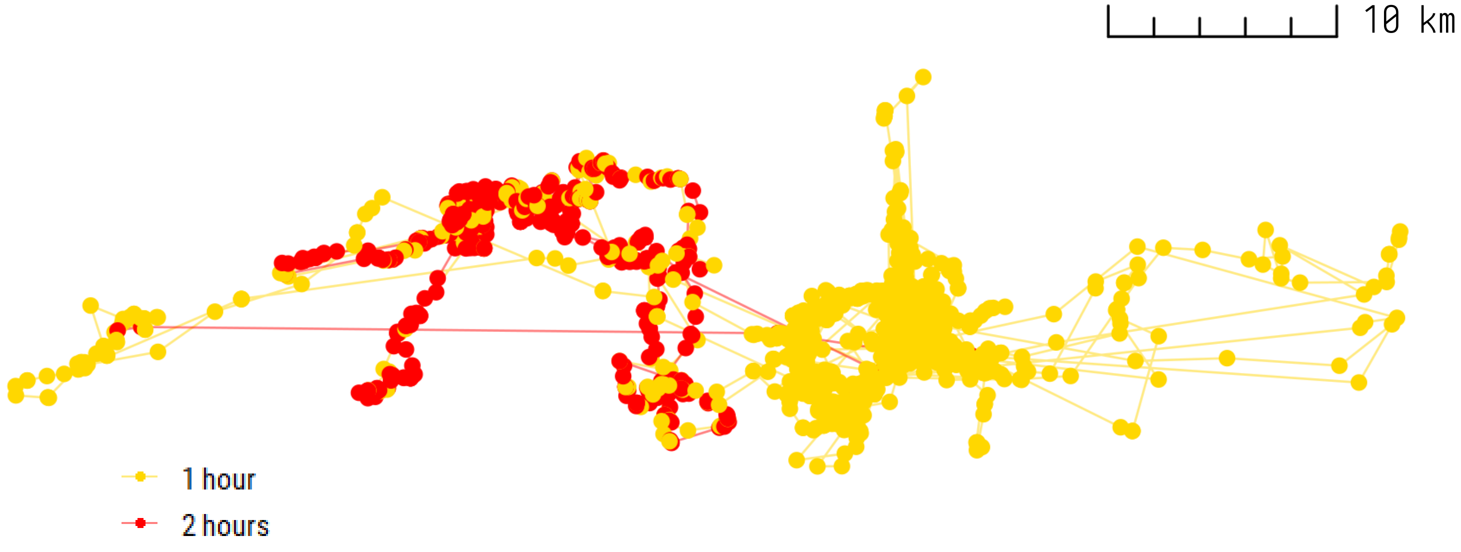


Fig. African buffalo dataset (nicknamed "Pepper") with an ***irregular*** sampling schedule. Sampling rate shifted from ***1 fix every hour*** to ***1 fix every 2 hours***.

wAKDE optimally ***upweights*** observations that occur during ***under-sampled times***, while optimally ***downweighting*** observations occurring during ***over-sampled times***.

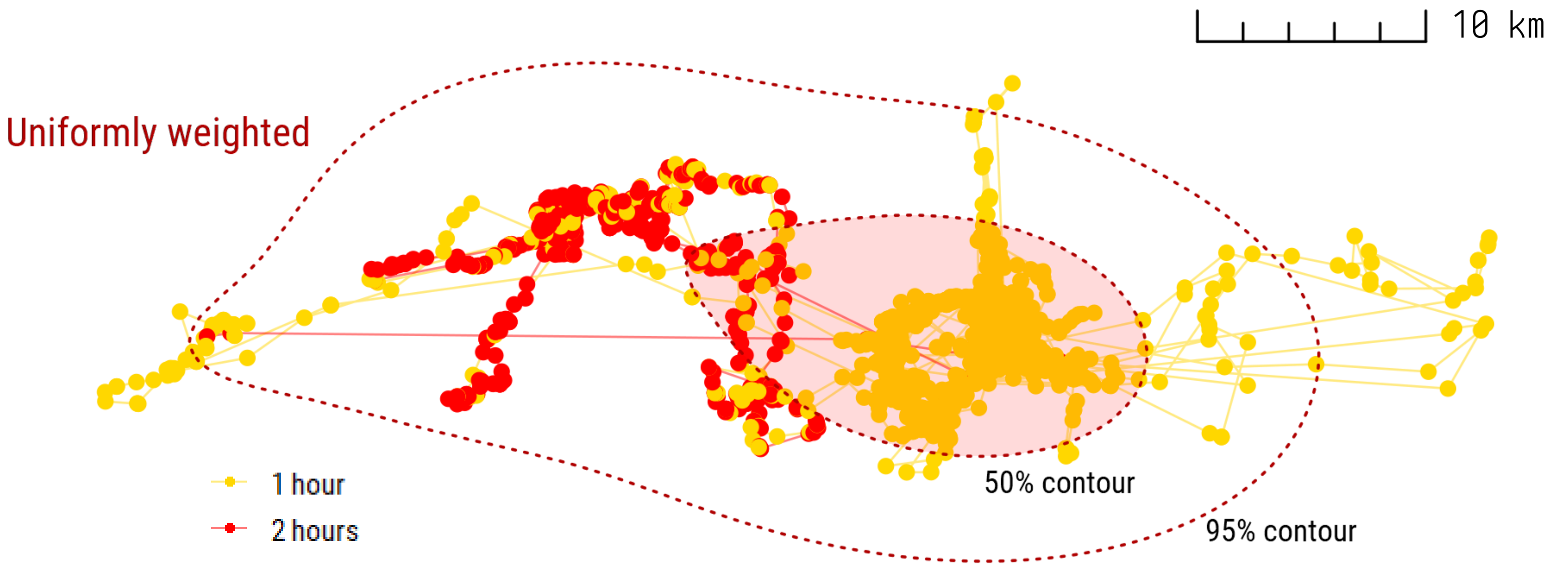


Fig. African buffalo dataset (nicknamed "Pepper") with an ***irregular*** sampling schedule. Sampling rate shifted from ***1 fix every hour*** to ***1 fix every 2 hours***.

wAKDE optimally ***upweights*** observations that occur during ***under-sampled times***, while optimally ***downweighting*** observations occurring during ***over-sampled times***.

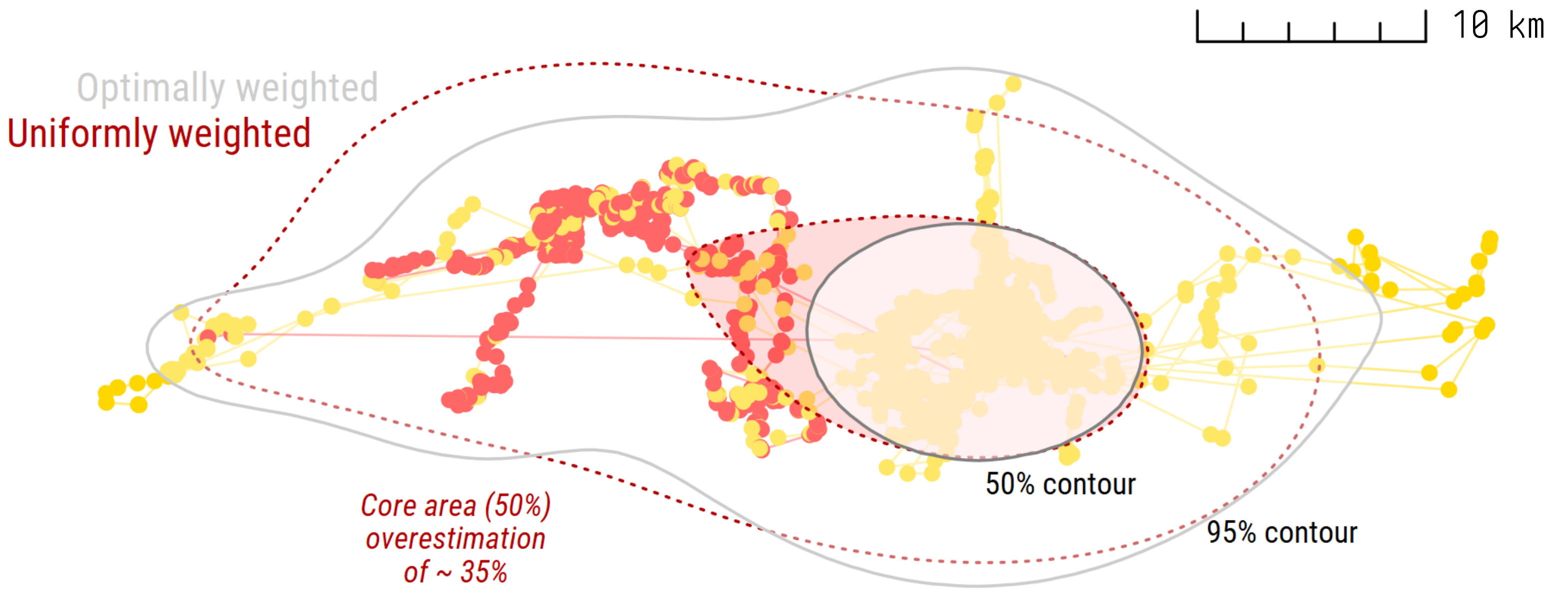


Fig. African buffalo dataset (nicknamed "Pepper") with an ***irregular*** sampling schedule. Sampling rate shifted from ***1 fix every hour*** to ***1 fix every 2 hours***.

i.simoes-silva@hzdr.de

Tracking data → Movement model → ***Home range estimate***

Parametric bootstrapping AKDE

Deals with: ***very low effective sample size***

Bartlett (1937)

Residual **ML (or REML)**

Fleming *et al.* (2019)

perturbative Hybrid REML **(pHREML)**

\+

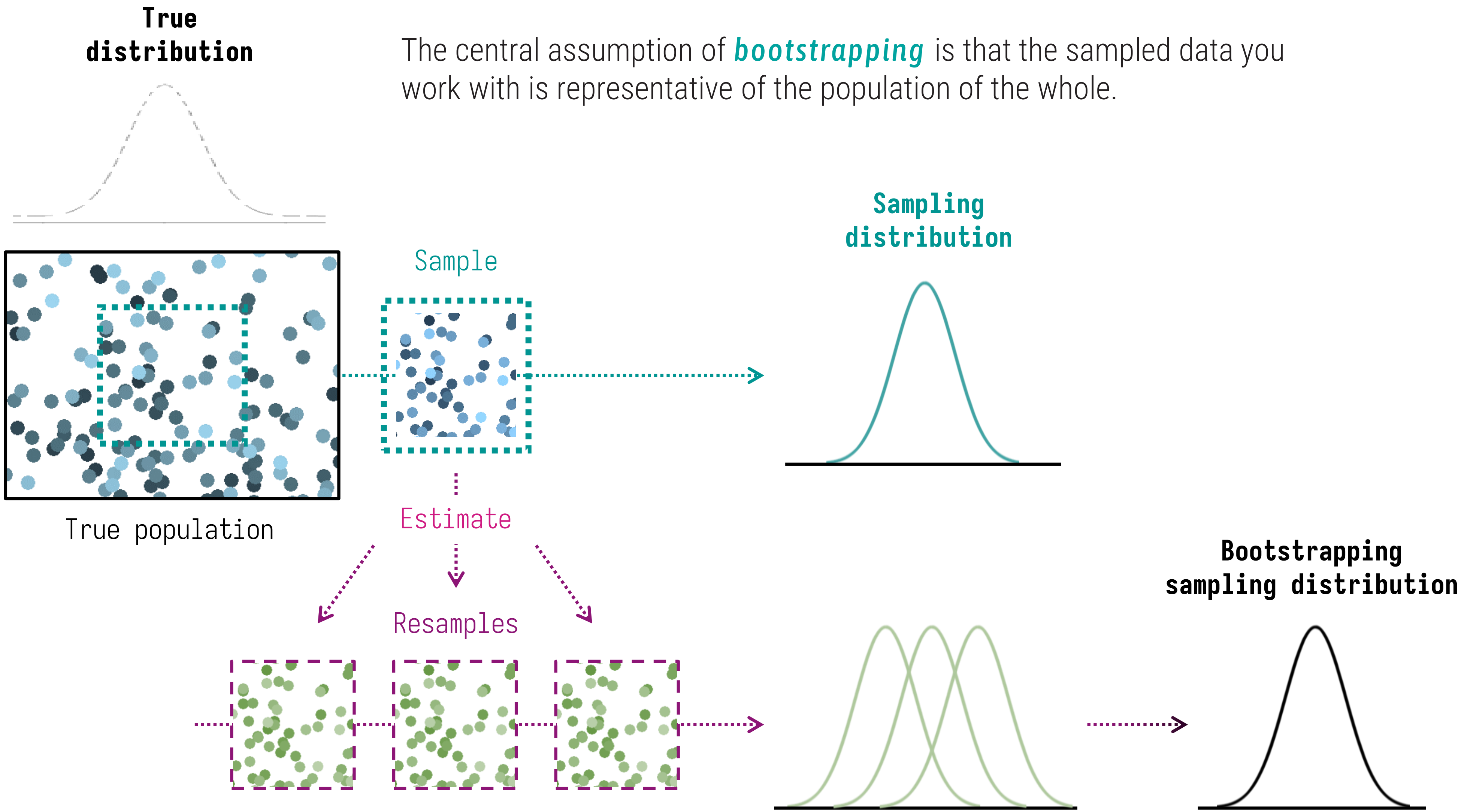
Efron & Efron (1982)

Parametric bootstrapping

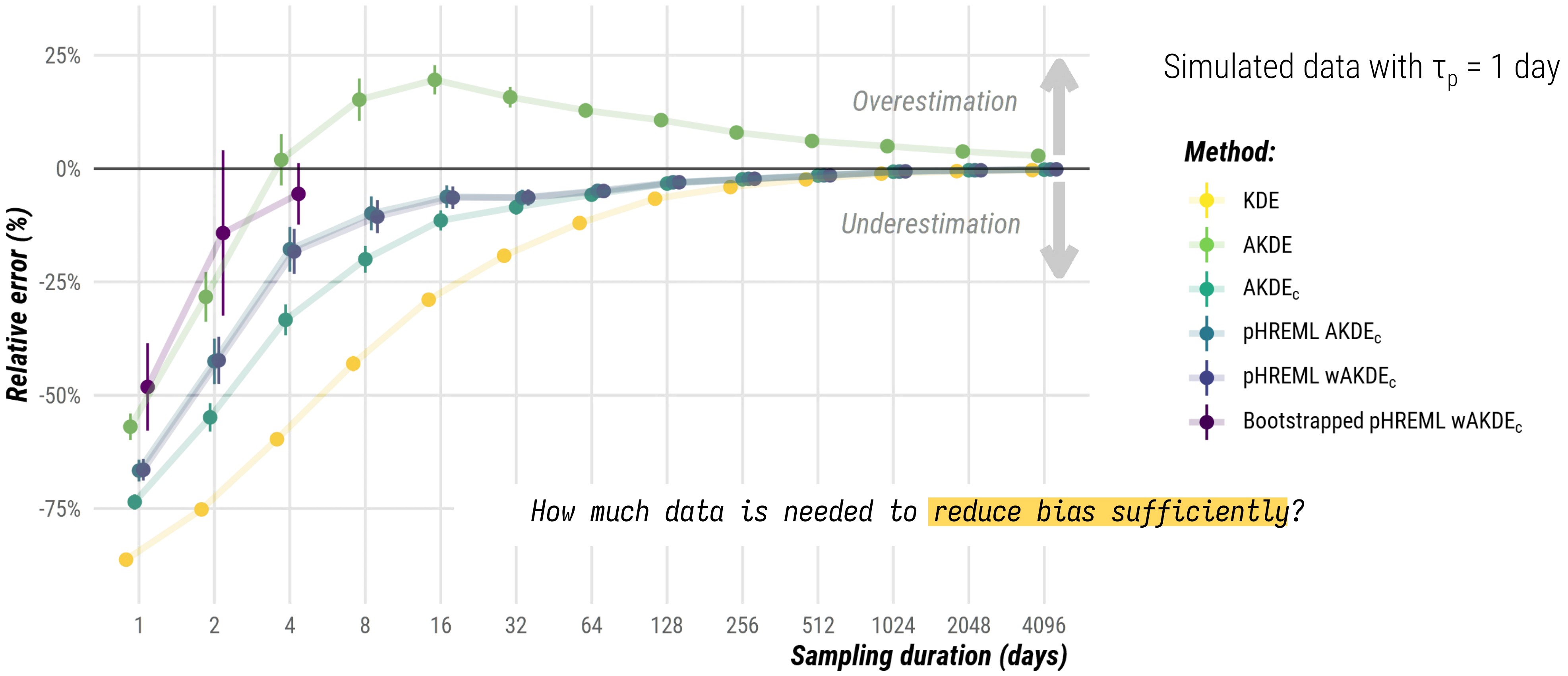
The parametric bootstrap estimates the bias and variance of an estimator by *approximating* the sampling distribution of the true movement model with that of the best-fit model.

i.simoes-silva@hzdr.de

The central assumption of ***bootstrapping*** is that the sampled data you work with is representative of the population of the whole.

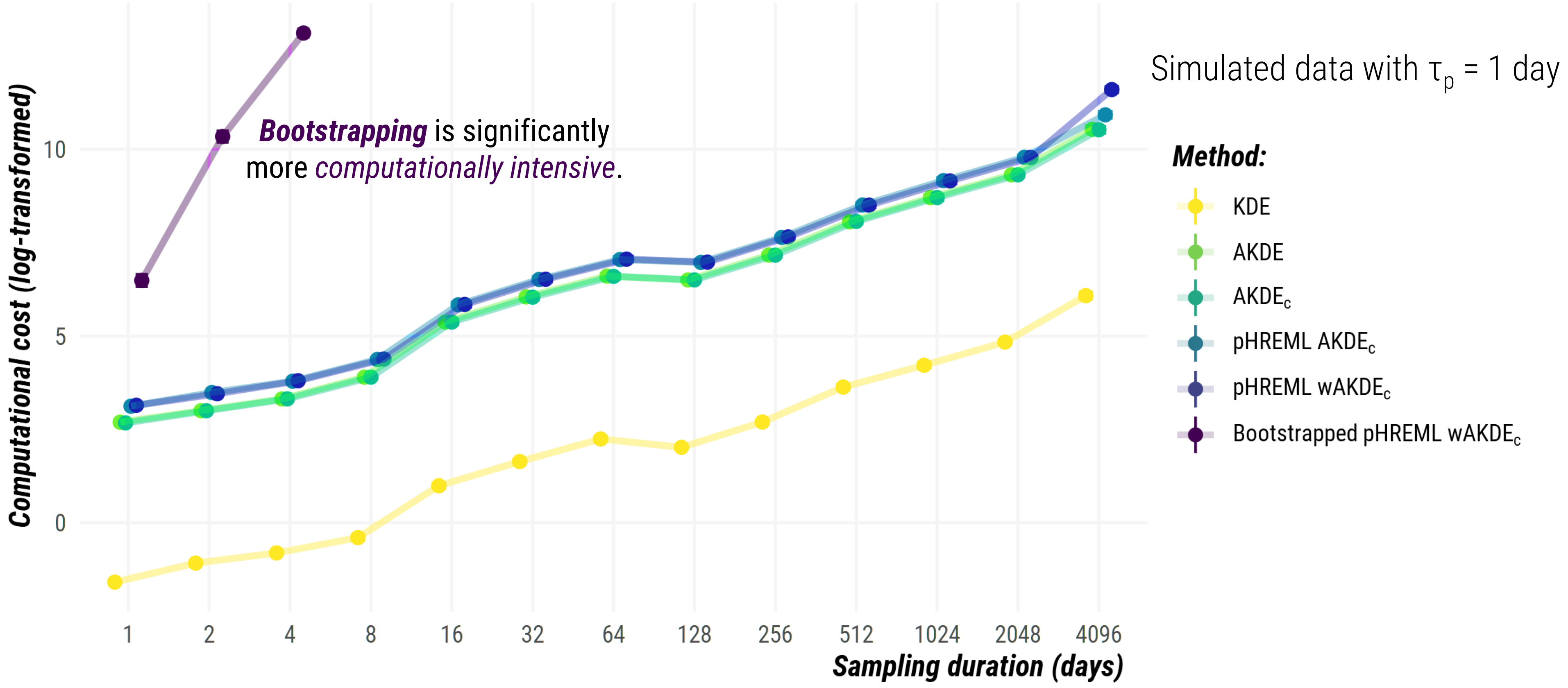


Relative error vs. sampling duration



How much data is needed to reduce bias sufficiently?

Computational cost vs. sampling duration



i.simoes-silva@hzdr.de