

MARE-Madeira 2025



# *Home range estimation*

Using the '`ctmm`' R package

*Inês Silva*

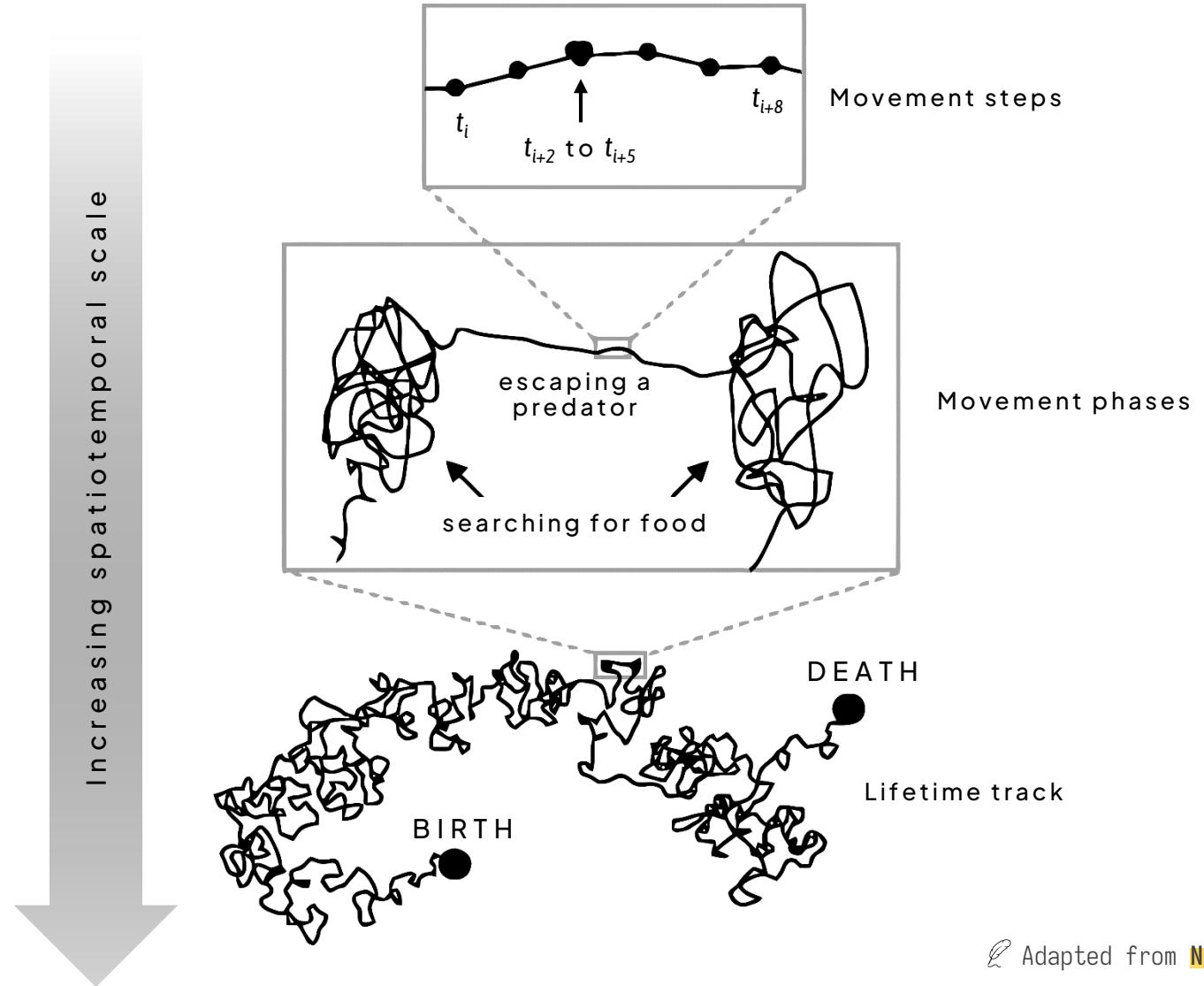
✉ [i.simoes-silva@hzdr.de](mailto:i.simoes-silva@hzdr.de)



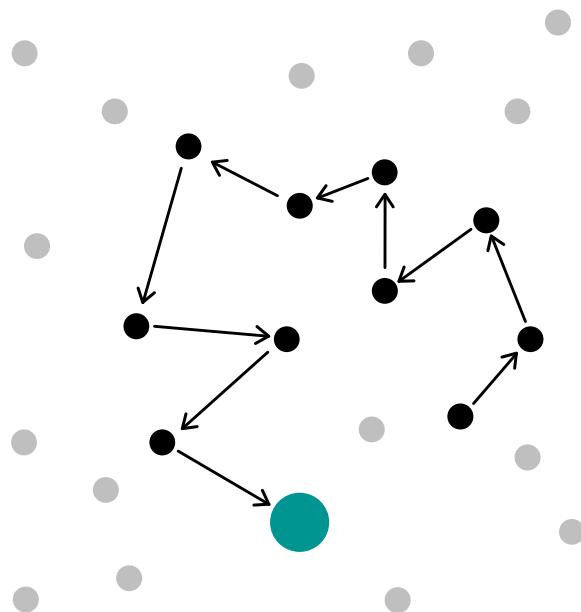
- |  |                                     |
|--|-------------------------------------|
| Account for autocorrelation                                  | (stochastic process models)         |
| Account for sampling irregularity and mismatch               | (continuous-time models)            |
| Account 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)            |
| Propagate individual uncertainties into population estimates | (hierarchical models)               |
| Reducing bias and error as much as possible                  | (debiased estimators)               |



## Introduction



↬ Adapted from [Nathan et al. \(2008\)](#)



## Brownian motion

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

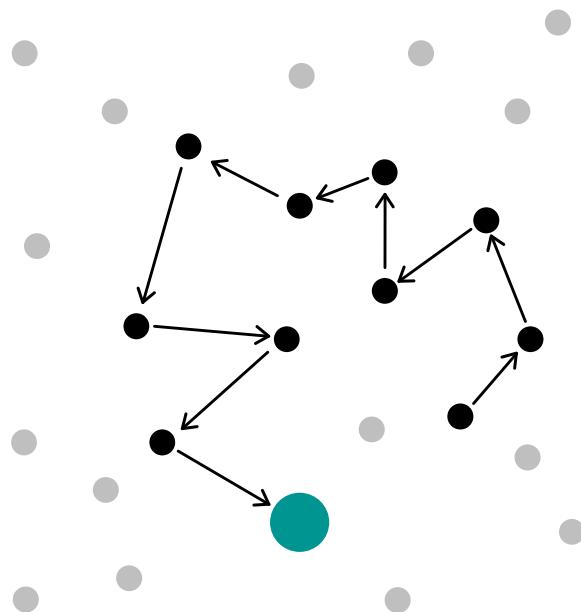
A **stochastic process** describes the evolution of some phenomenon over time. “**Stochastic**” is generally interchangeable with “**random**”, indicating an element of **chance or unpredictability**.



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.



## Brownian motion

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

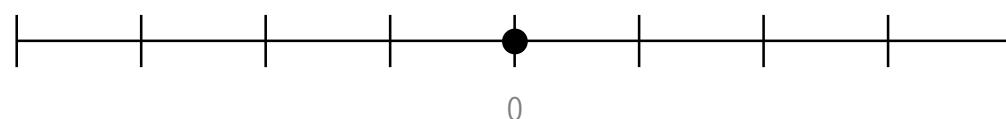
A **stochastic process** describes the evolution of some phenomenon over time. “**Stochastic**” is generally interchangeable with “**random**”, indicating an element of **chance or unpredictability**.

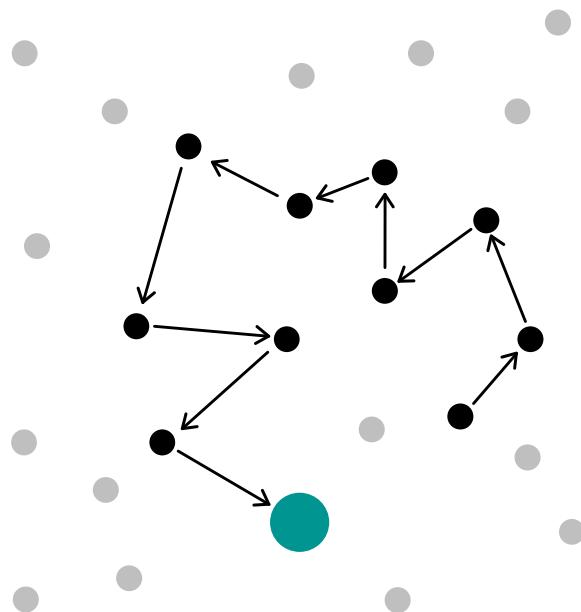


Does not imply that we have no information about the dynamics; a stochastic process provides structure with some **randomness**.

## 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.





## Brownian motion

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

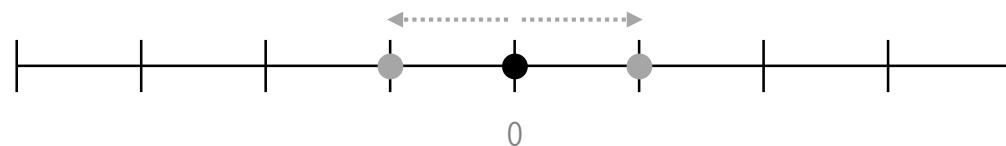
A **stochastic process** describes the evolution of some phenomenon over time. “**Stochastic**” is generally interchangeable with “**random**”, indicating an element of **chance or unpredictability**.

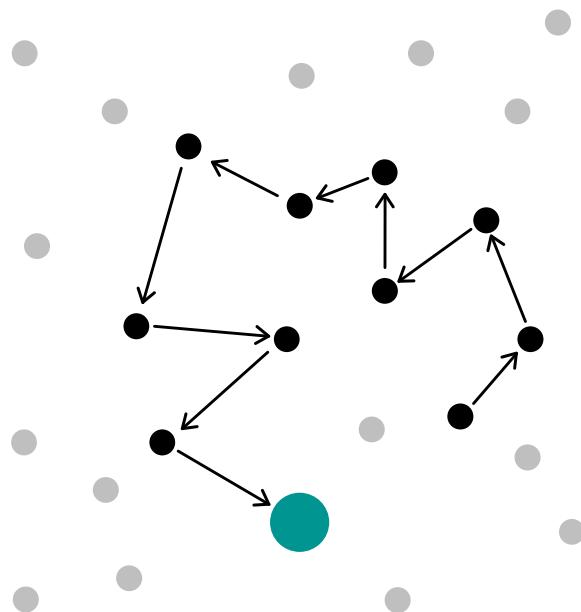


Does not imply that we have no information about the dynamics; a stochastic process provides structure with some **randomness**.

## 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.





## Brownian motion

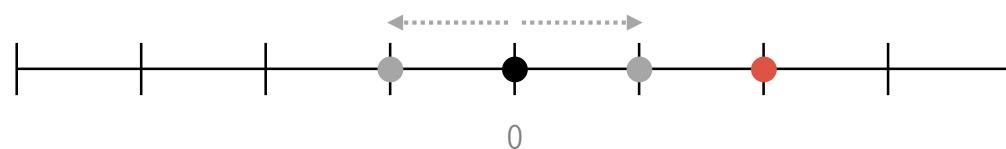
Particles move **randomly** in all directions, *i.e.*, change in position is random and continuous over time.

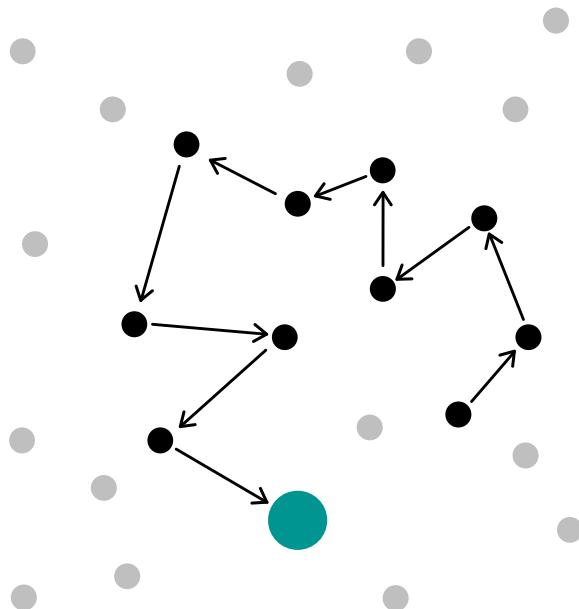
A **stochastic process** describes the evolution of some phenomenon over time. “**Stochastic**” is generally interchangeable with “**random**”, indicating an element of **chance or unpredictability**.

 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.





## Brownian motion

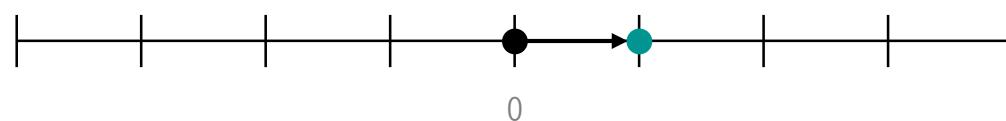
Particles move **randomly** in all directions, i.e., change in position is random and continuous over time.

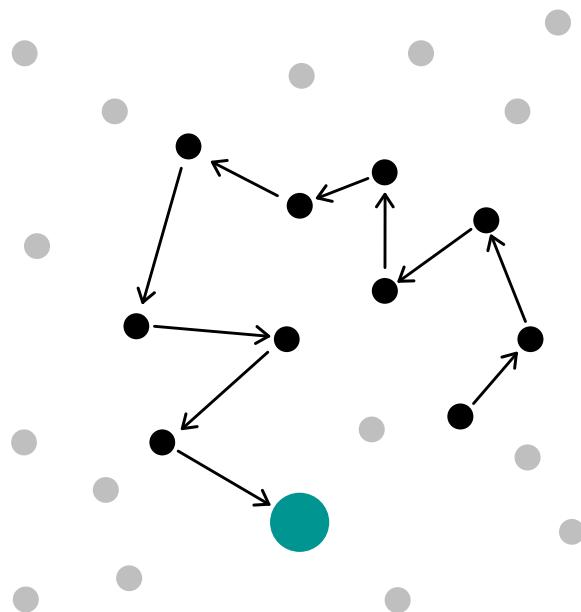
A **stochastic process** describes the evolution of some phenomenon over time. “**Stochastic**” is generally interchangeable with “**random**”, indicating an element of **chance or unpredictability**.

 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.





## Brownian motion

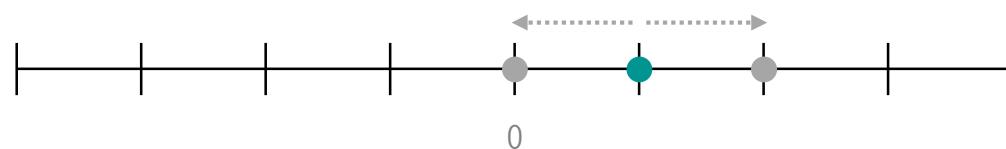
Particles move **randomly** in all directions, *i.e.*, change in position is random and continuous over time.

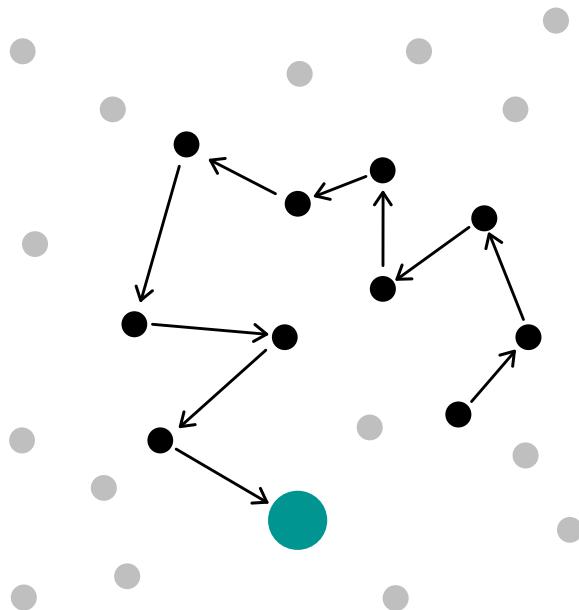
A **stochastic process** describes the evolution of some phenomenon over time. “**Stochastic**” is generally interchangeable with “**random**”, indicating an element of **chance or unpredictability**.

 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.





## Brownian motion

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

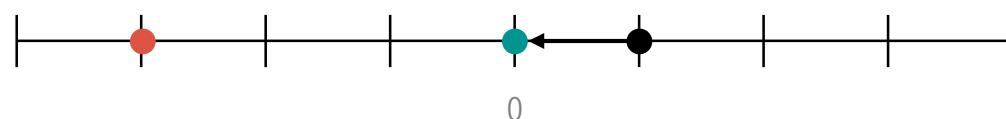
A **stochastic process** describes the evolution of some phenomenon over time. “**Stochastic**” is generally interchangeable with “**random**”, indicating an element of **chance or unpredictability**.



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

## Markov process

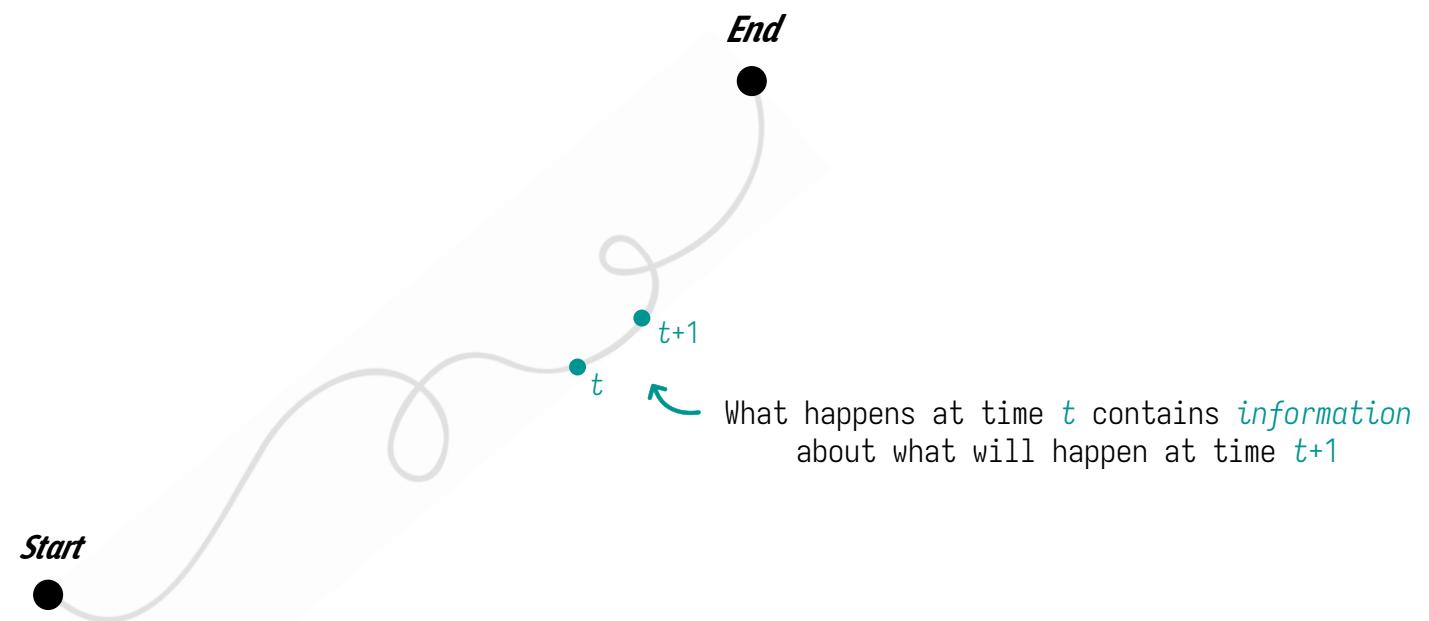
Type of stochastic process where the future state of the system depends **only on its current state**, not on the sequence of events that preceded it.





### Formal description of movement:

Movement is defined as an individual's spatial locations over time,  
forming a **continuous trajectory**.



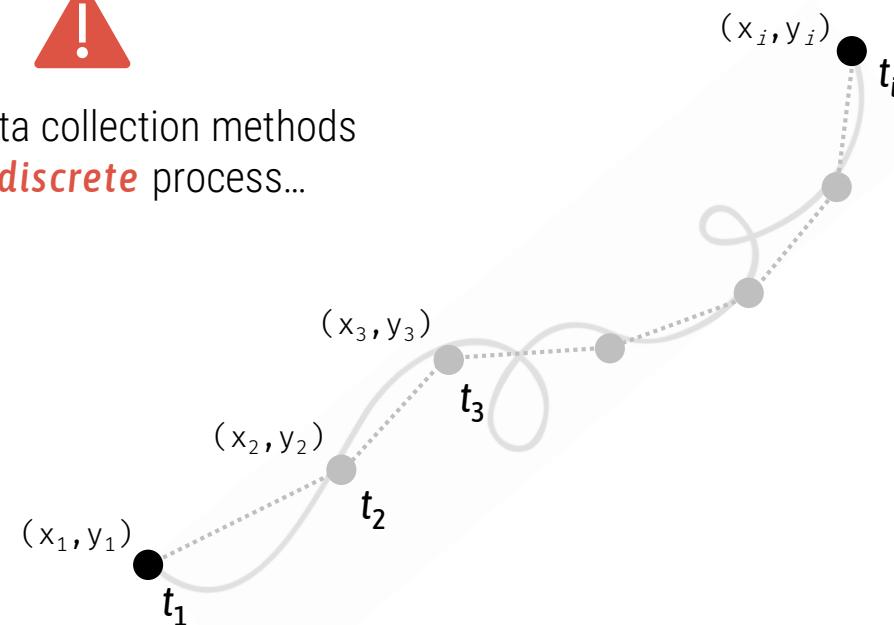


### Formal description of movement:

Movement is defined as an individual's spatial locations over time,  
forming a **continuous trajectory**.



Current data collection methods  
result in a **discrete** process...



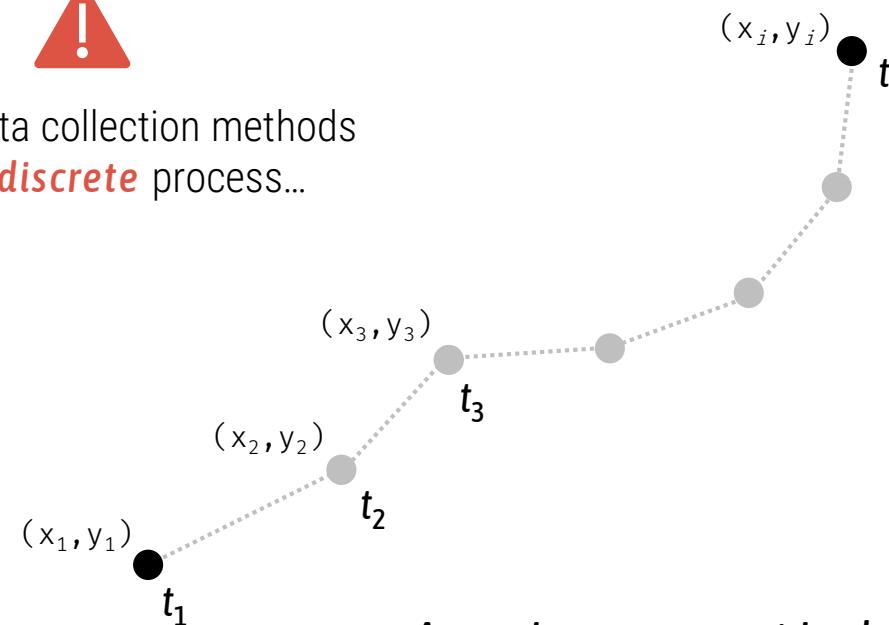


### Formal description of movement:

Movement is defined as an individual's spatial locations over time,  
forming a **continuous trajectory**.



Current data collection methods  
result in a **discrete** process...

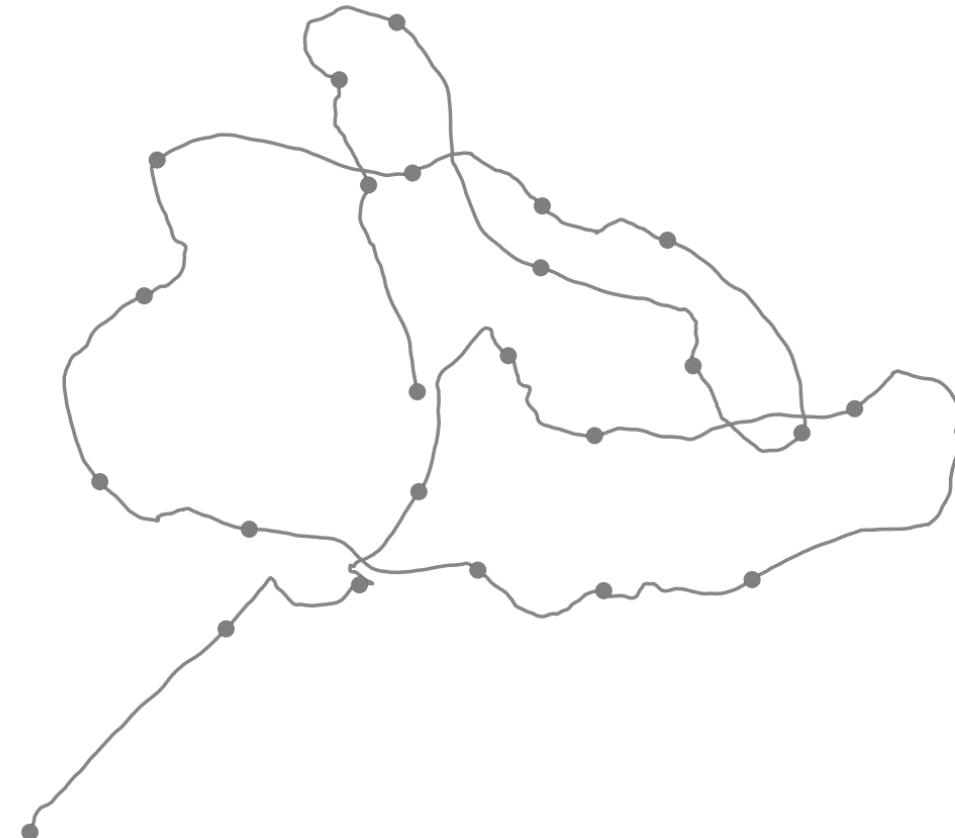


**Assuming movement is also discrete-time may cause problems when:**

- sampling is irregular in time,
- and/or sampling schedules differ across individuals.



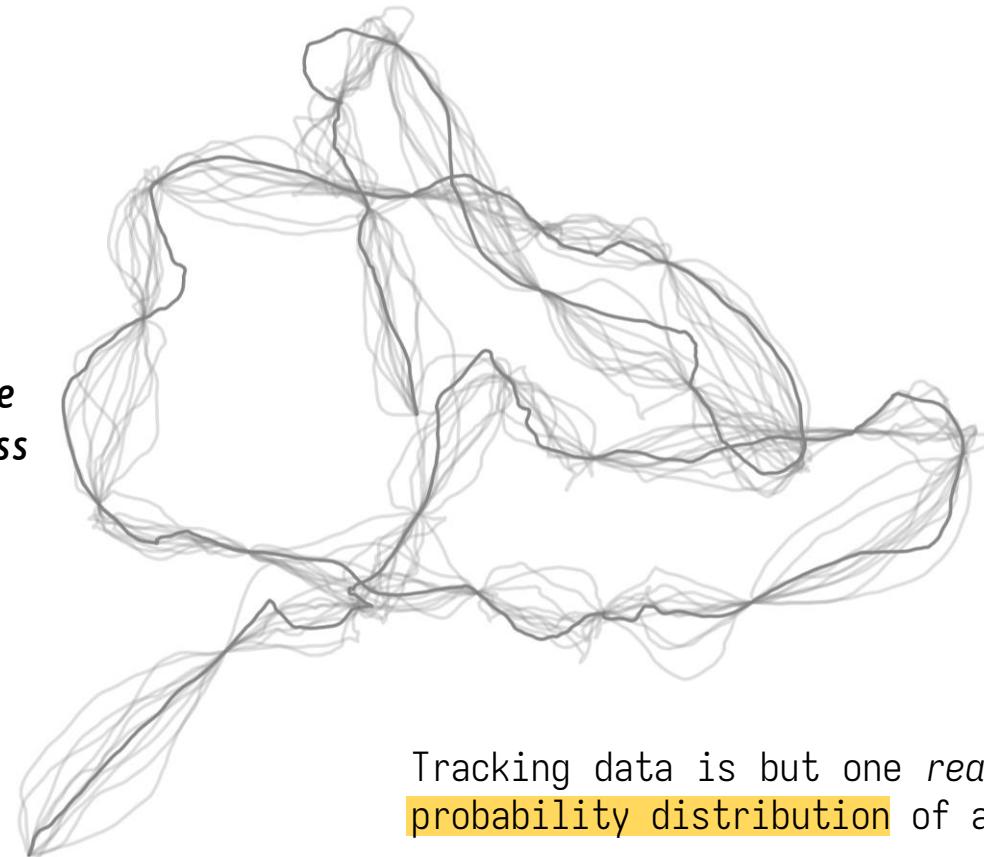
**How can we accurately characterize animal movement?**  
(i.e., location, velocity and direction over a period of time)





**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.



## 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

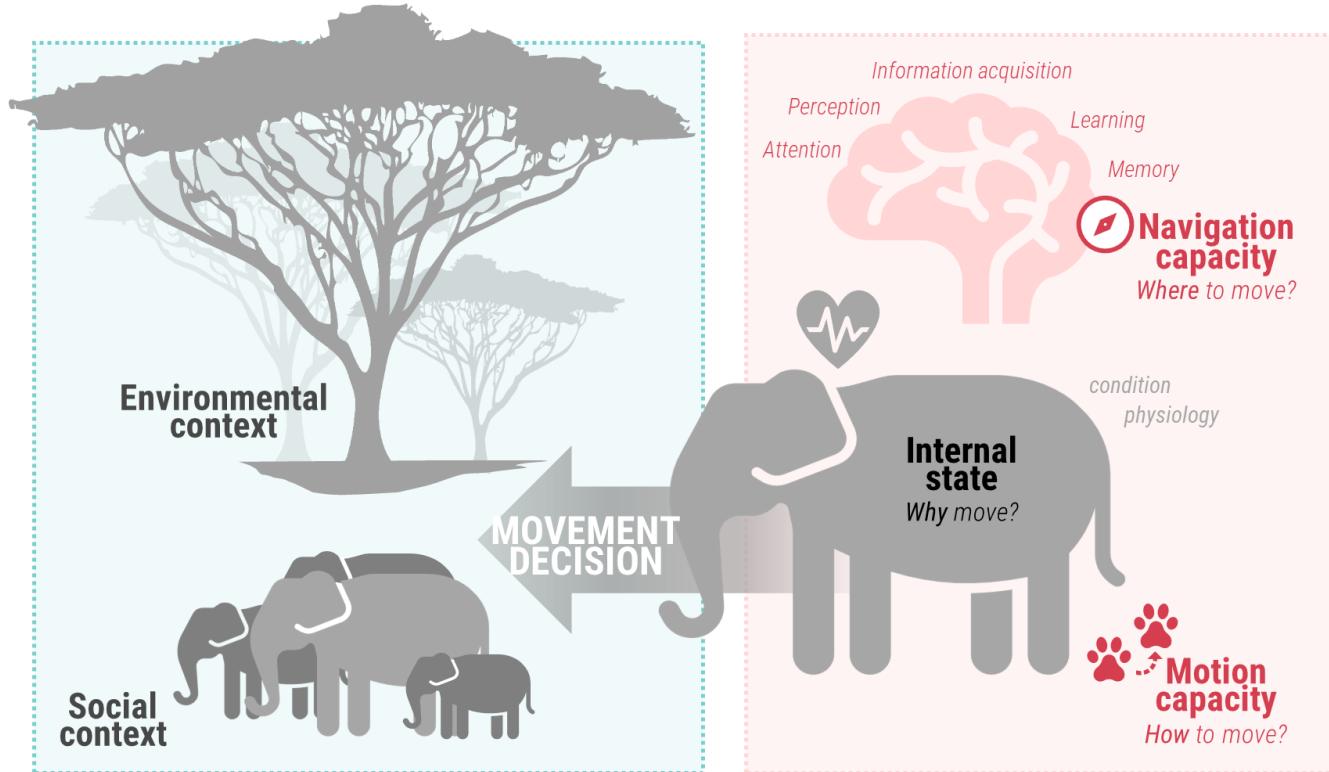
Can accommodate **speed, distance, acceleration, etc.**

Location **error** is easier to model.



What is a **home range**?

Characterizing spatial-temporal patterns of **space-use** is fundamental.



Adapted from **Lewis et al. (2021)**  
DOI: 10.3389/fevo.2021.681704



What is a *home range*?

**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**.



## What is a *home range*?

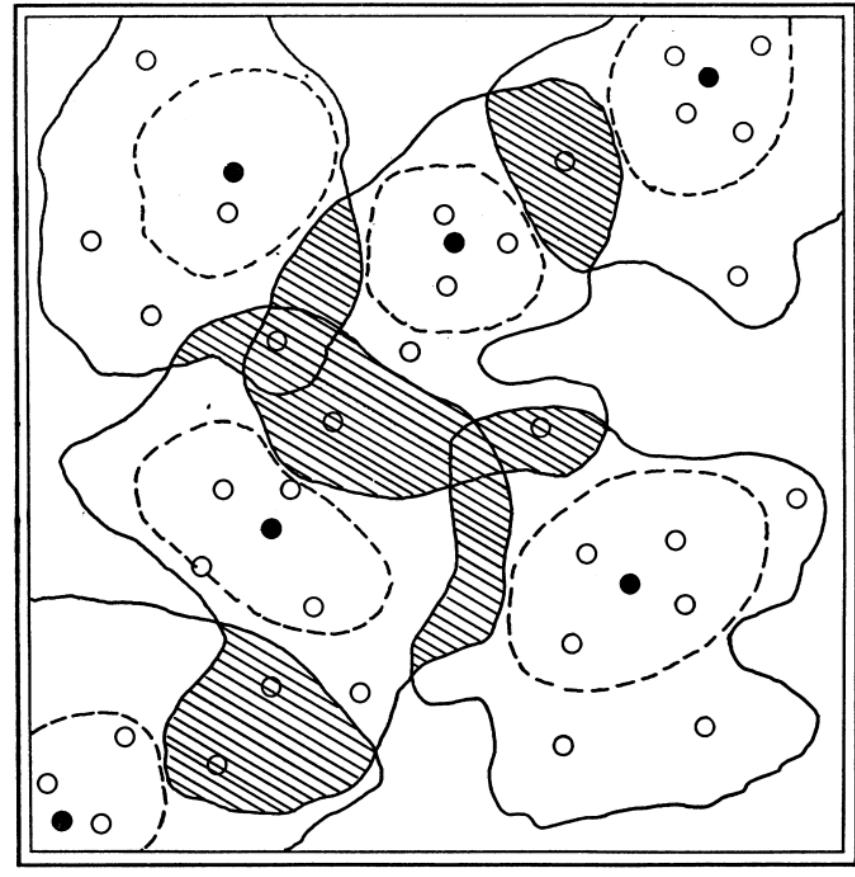
First defined as:

“

(...) 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 BOUNDARY    ┌─────────┐ NEUTRAL AREA  
- - - TERRITORIAL BOUNDARY    ● NESTING SITE  
BLANK--UNOCCUPIED SPACE    ○ REFUGE SITE

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



## What is a *home range*?

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.

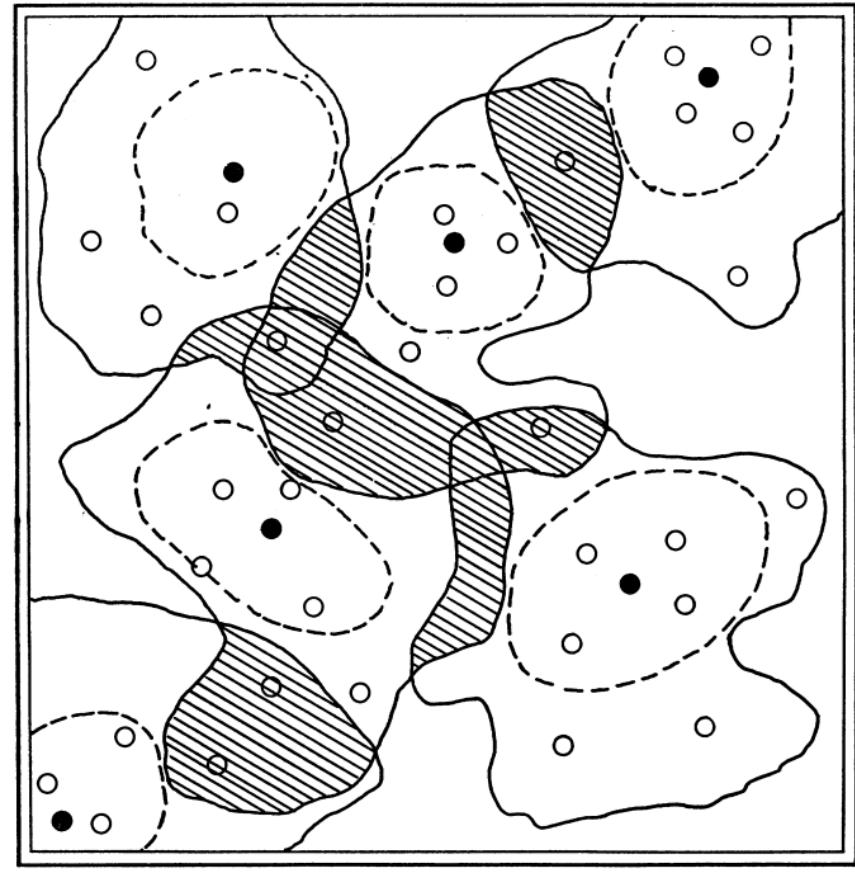


Burt (1943)

Home range  
not actively defended



Territory  
actively defended



— HOME RANGE BOUNDARY    ■ NEUTRAL AREA  
--- TERRITORIAL BOUNDARY    ● NESTING SITE  
BLANK--UNOCCUPIED SPACE    ○ REFUGE SITE

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



## What is a *home range*?

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**.

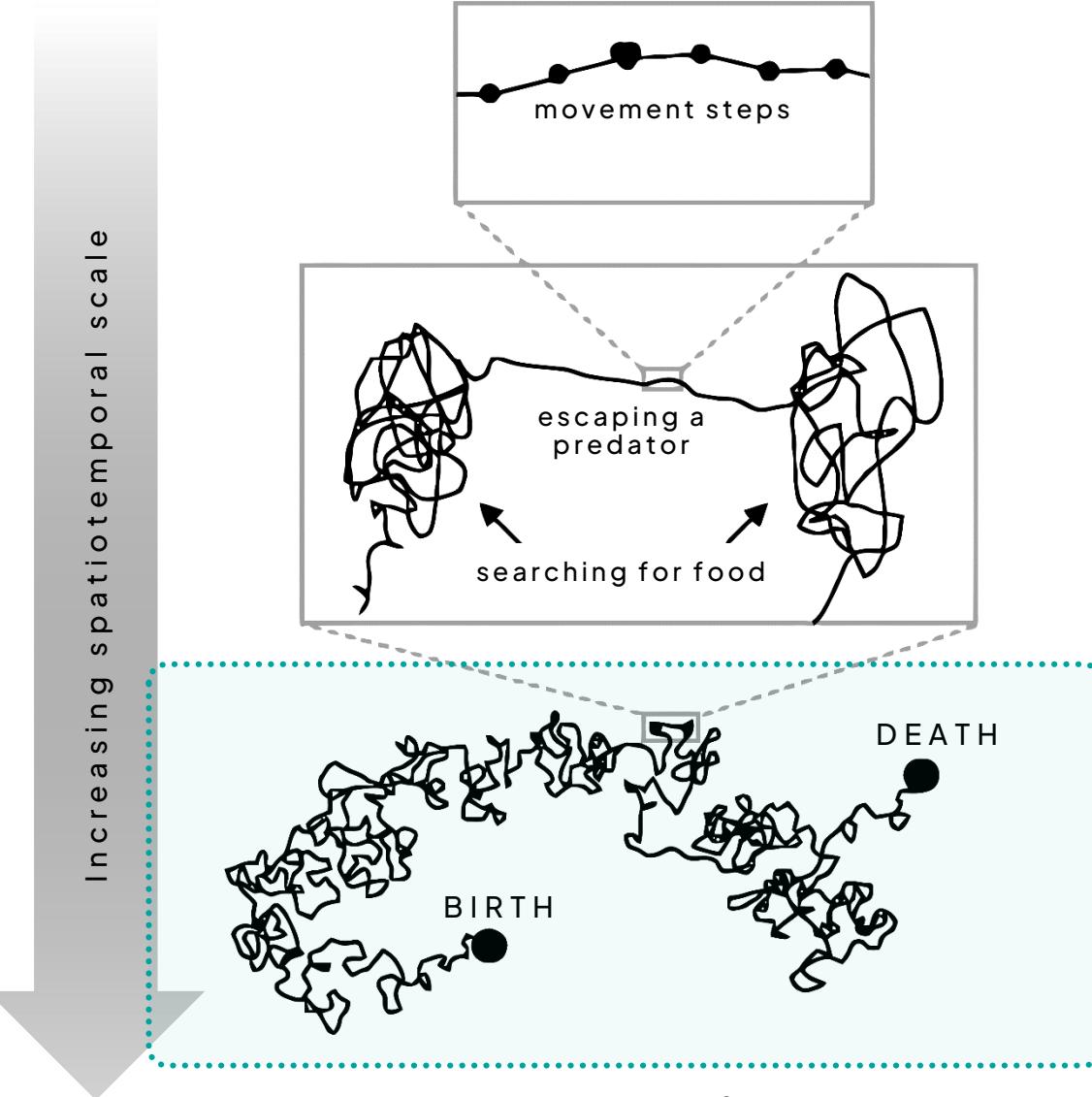


Exclusion of infrequent  
*outlying relocation points*.



## What is a *home range*?

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**.



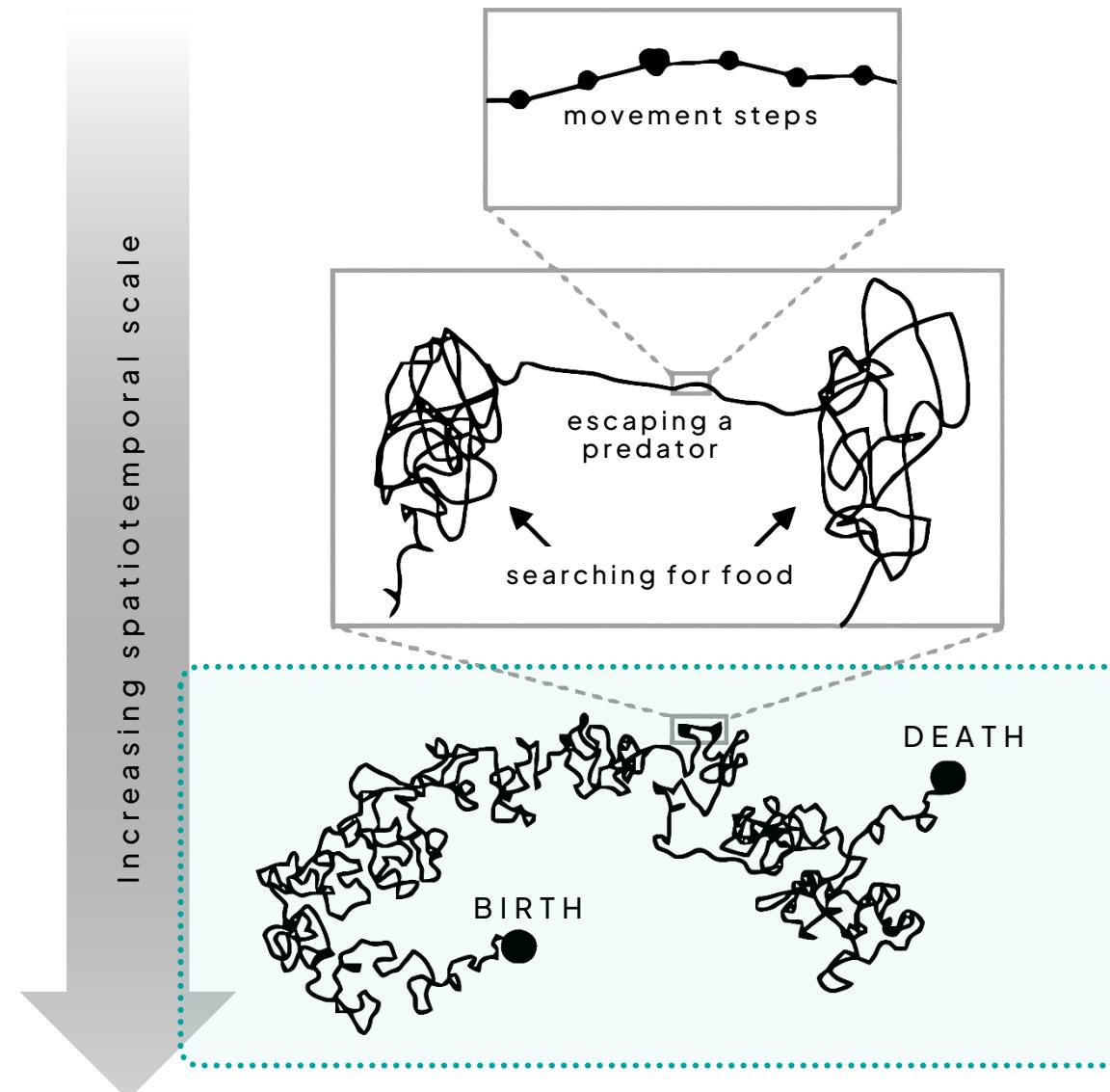
Adapted from [Nathan et al. \(2008\)](#)



## What is a *home range*?

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. \(2008\)](#)

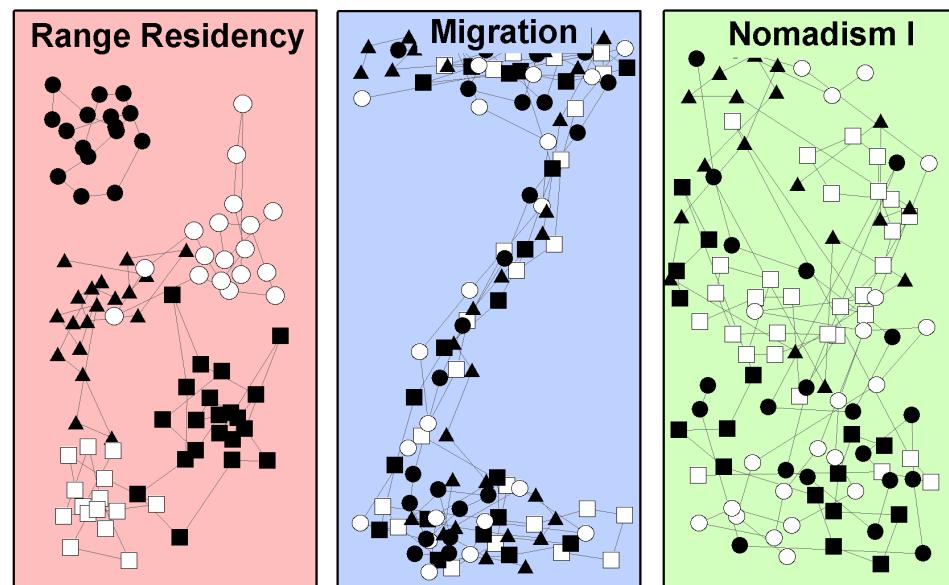


What is a **home range**?

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)



What is a **home range**?

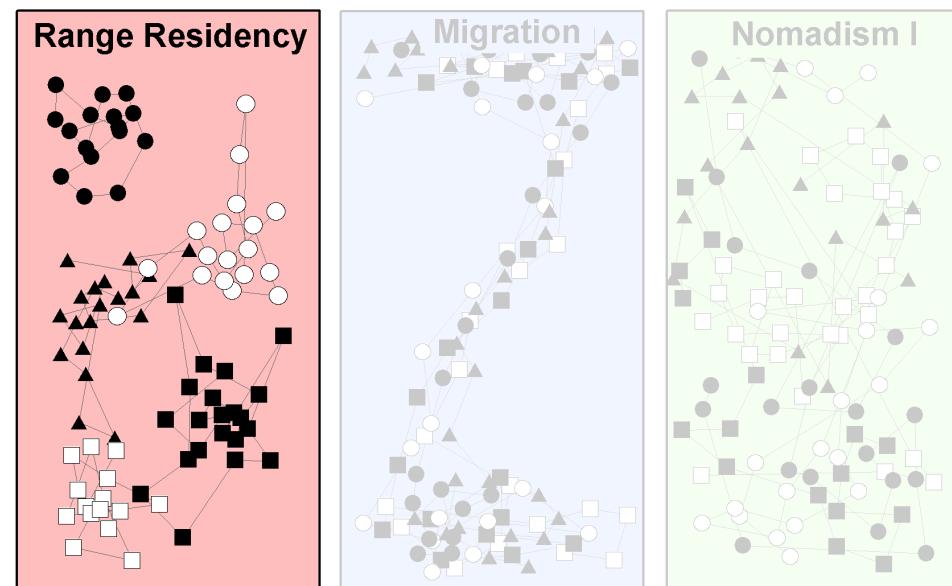
However, not all animals have **home ranges**...

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)



## Why estimate home range?

### 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*.



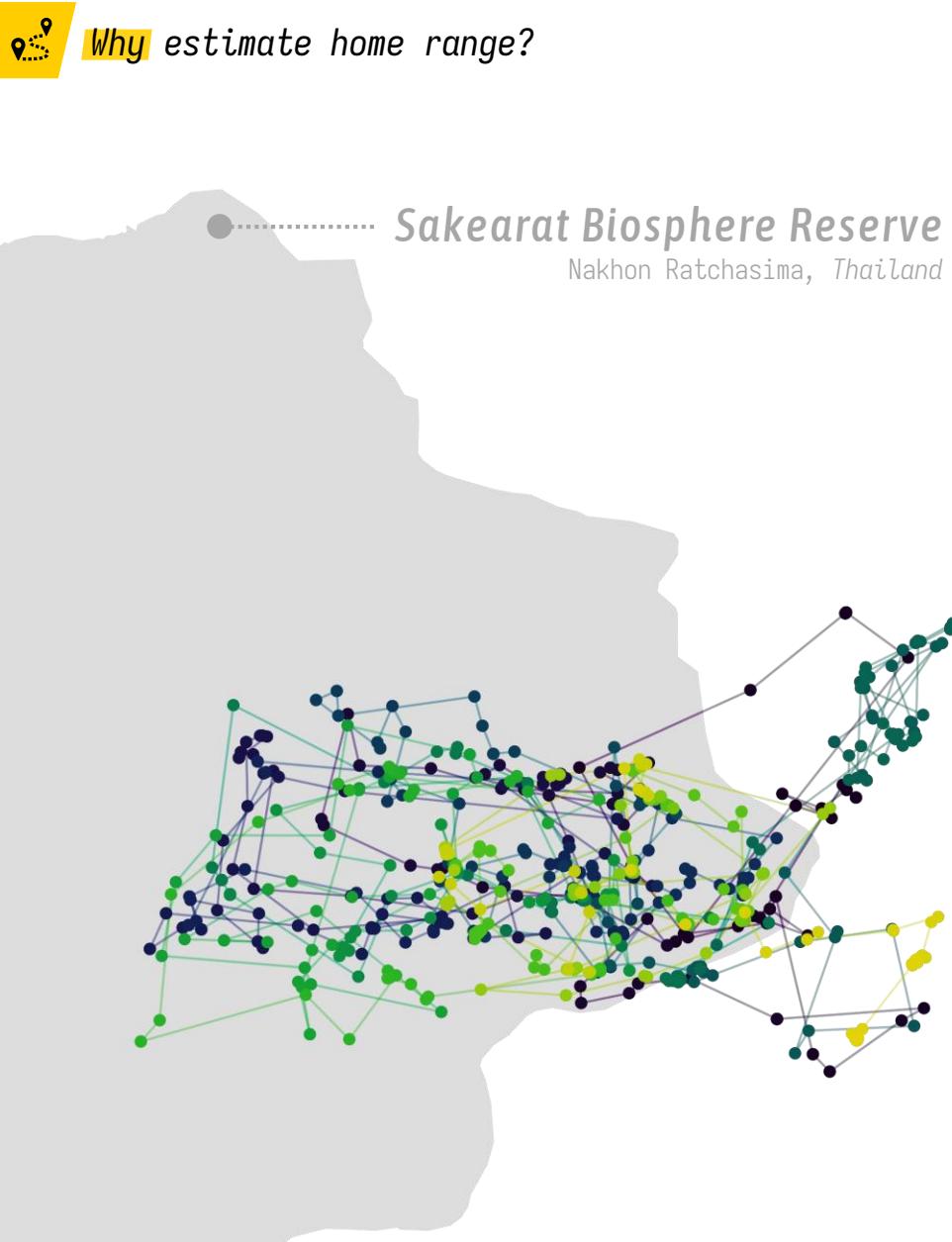
Dennis Hamilton



Avijan Saha



Aleksei Volkov

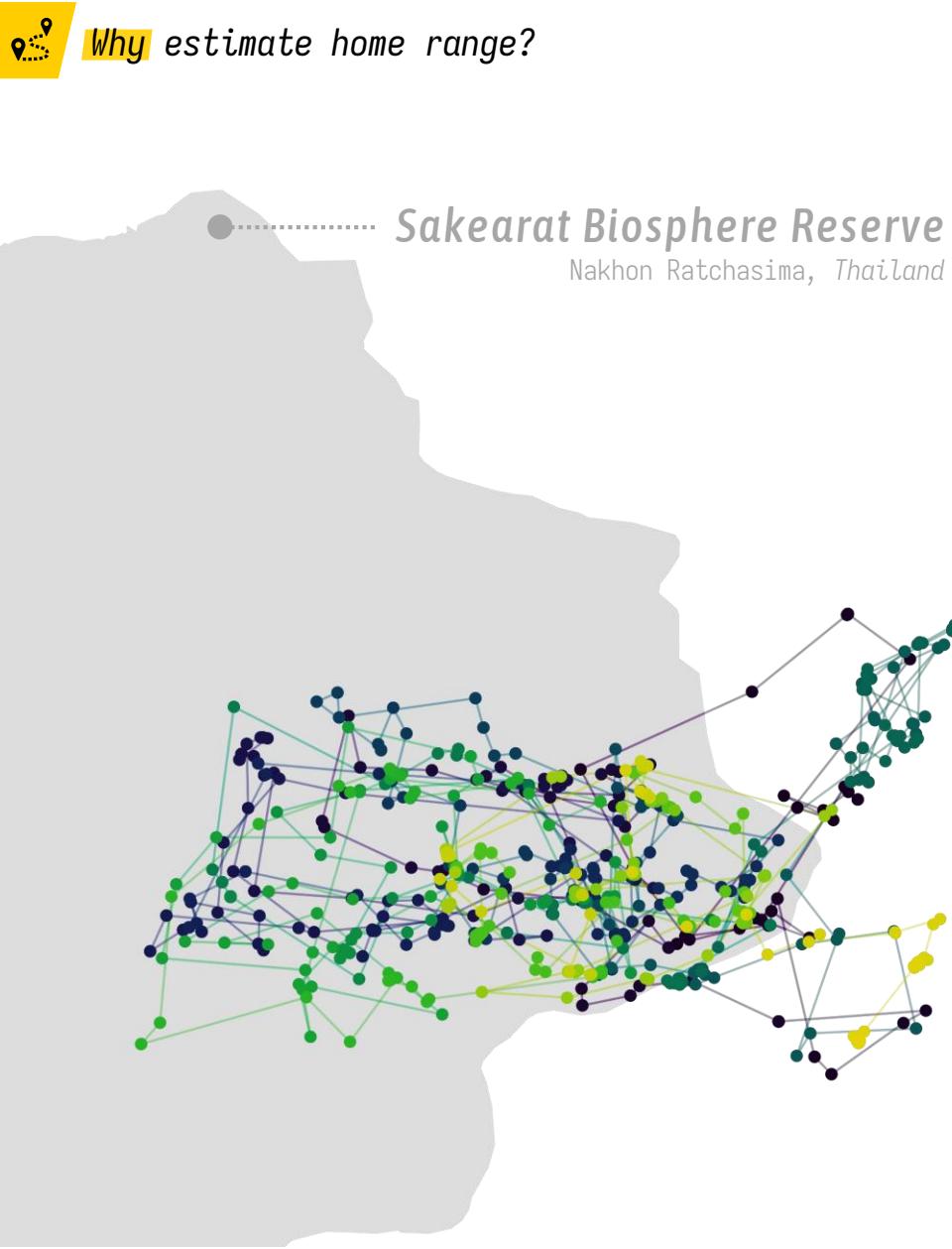


📍 Why estimate home range?

📷 Benjamin Marshall



**Northern king cobra**  
(*Ophiophagus hannah*)

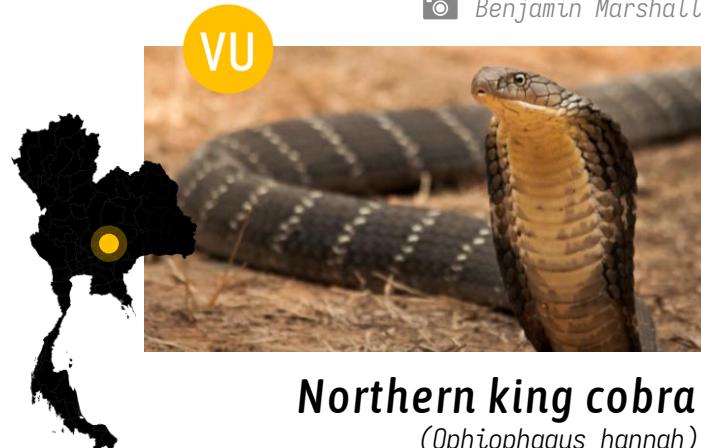


## Why estimate home range?

# Sakearat Biosphere Reserve

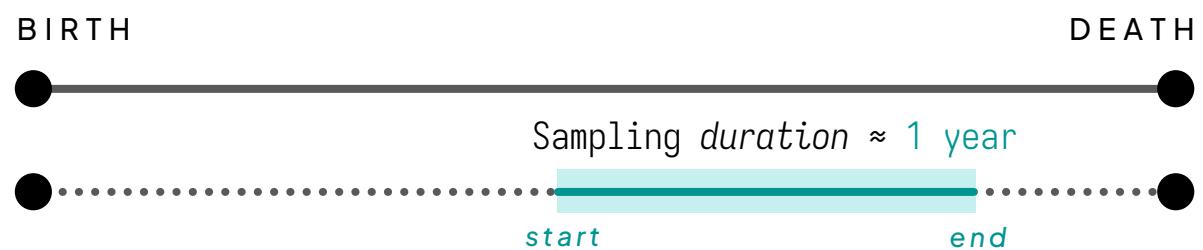
Nakhon Ratchasima, Thailand

 Benjamin Marshall



# **Northern king cobra**

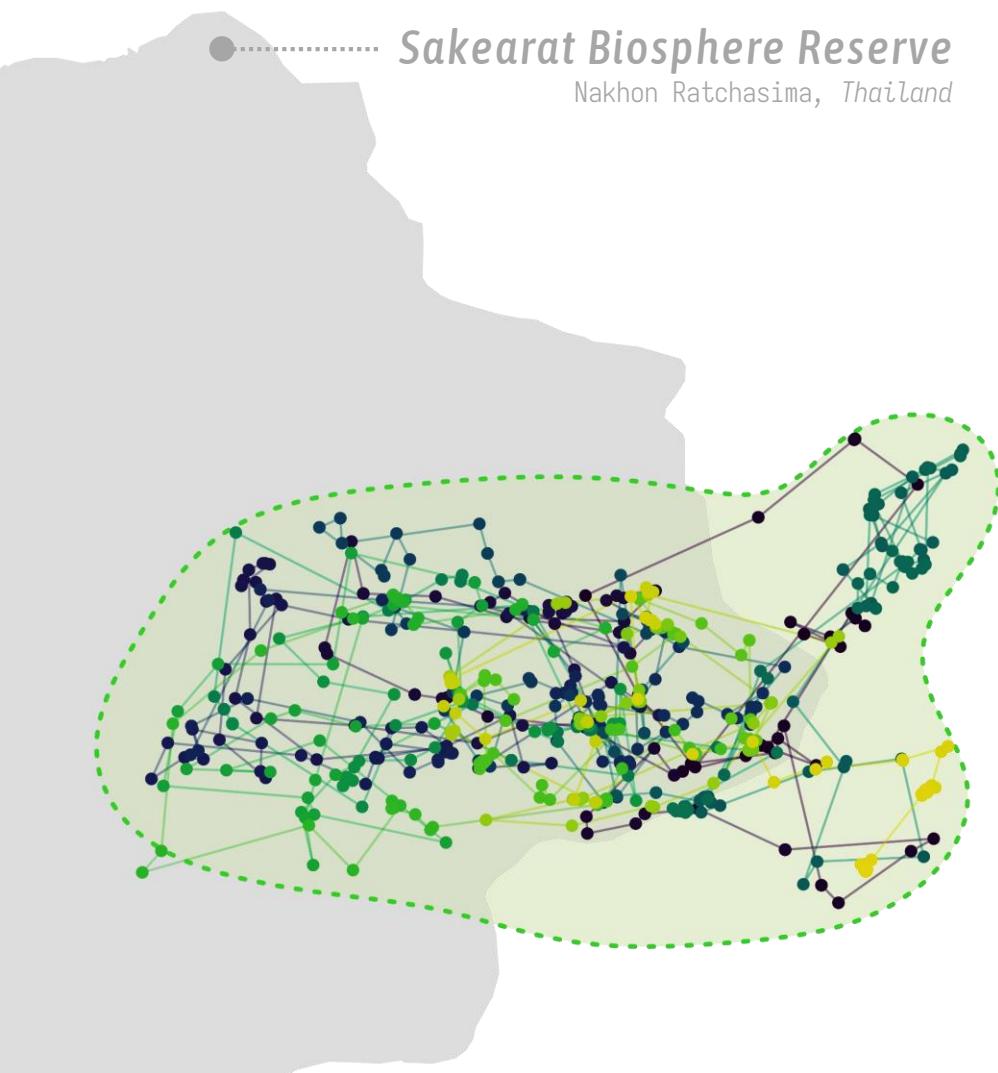
(*Ophiophagus hannah*)



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



Why estimate home range?



BIRTH



DEATH



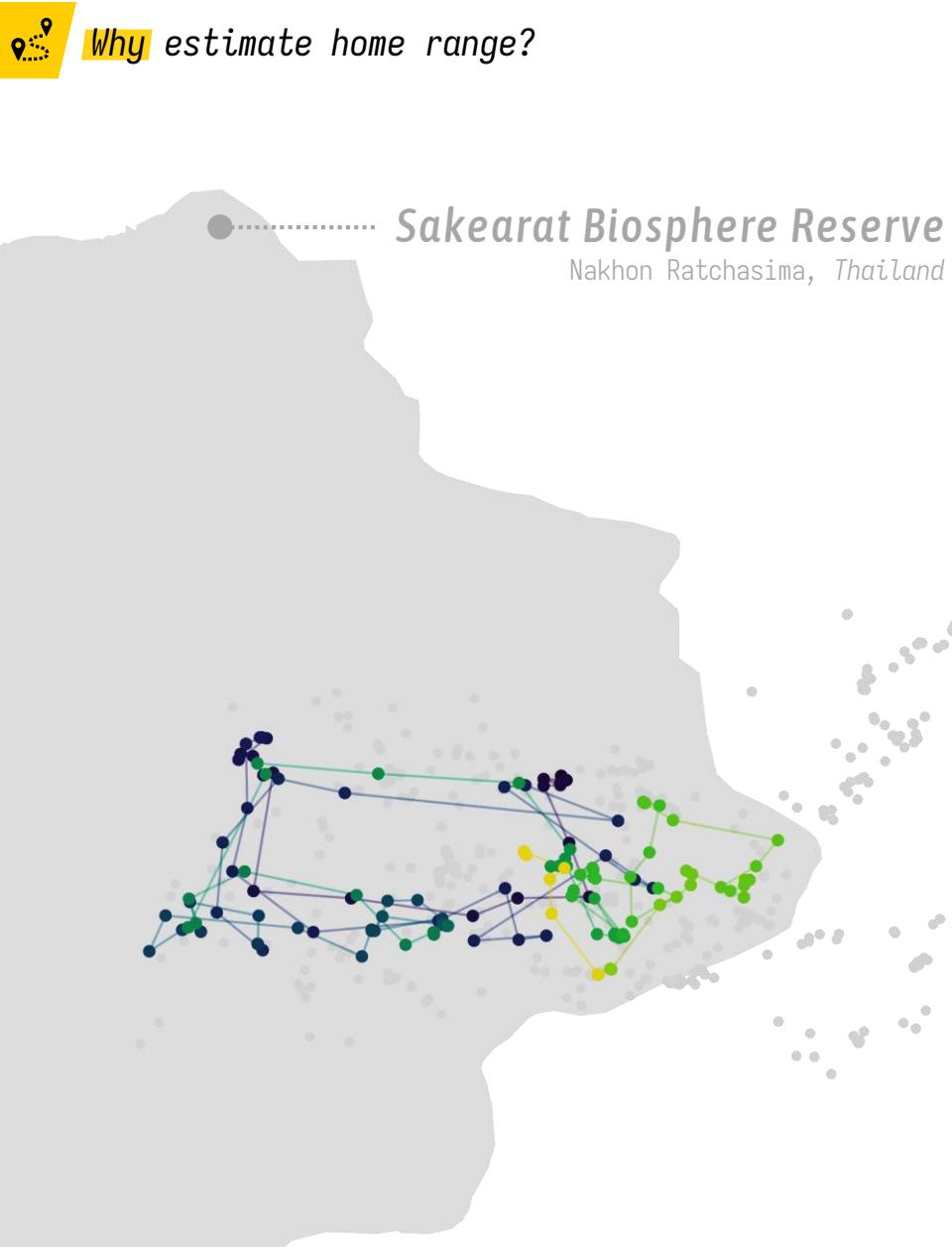
Sampling duration  $\approx$  1 year  
start end

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

Benjamin Marshall



**Northern king cobra**  
(*Ophiophagus hannah*)



## Why estimate home range?

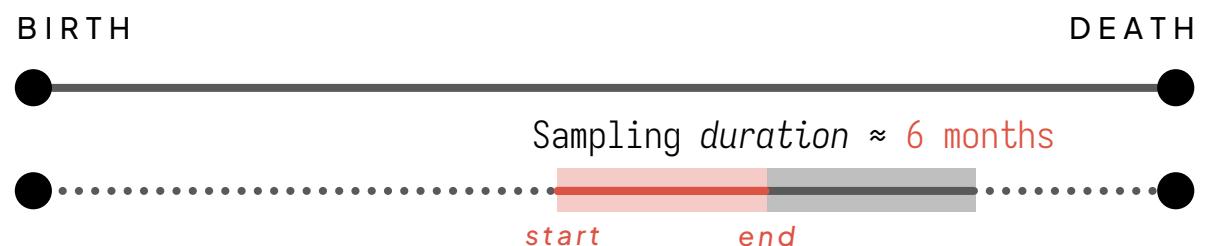
# Sakearat Biosphere Reserve

Nakhon Ratchasima, Thailand



# **Northern king cobra**

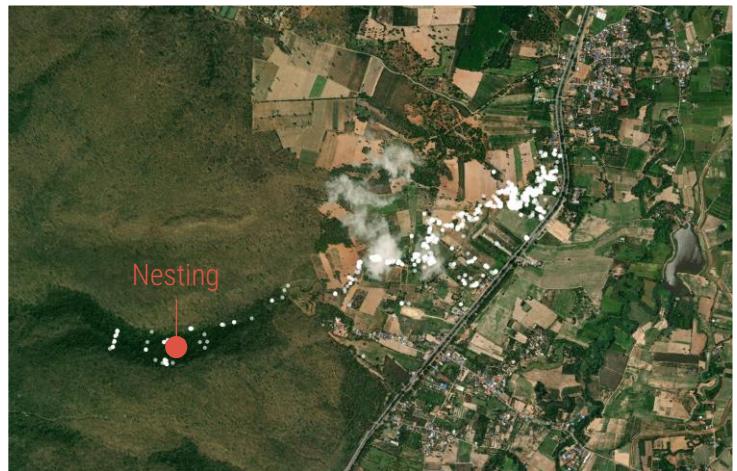
(*Ophiophagus hannah*)



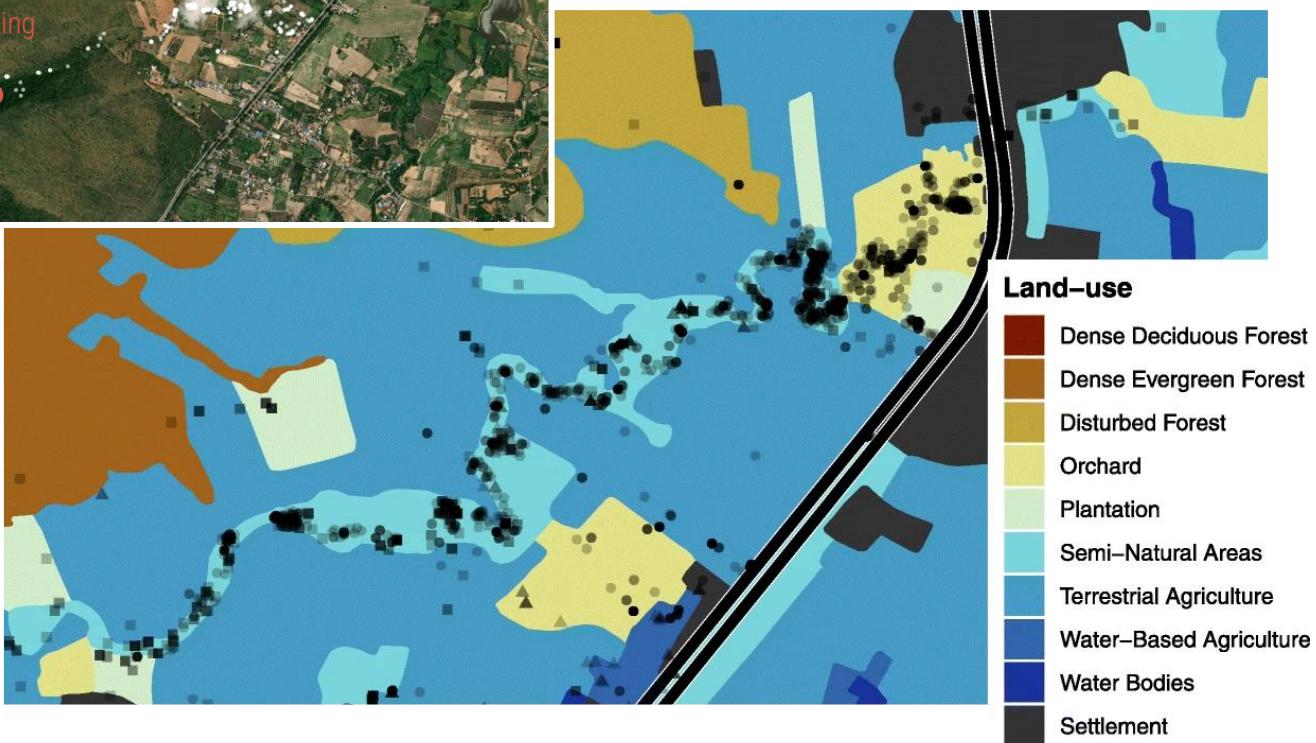
How representative is this period?  
Is the movement behavior stationary?



## Why estimate home range?



Marshall et al. (2020)



Movements are largely occurring within semi-natural areas.



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



Why estimate home range?



**Giant anteater**  
*(Myrmecophaga tridactyla)*



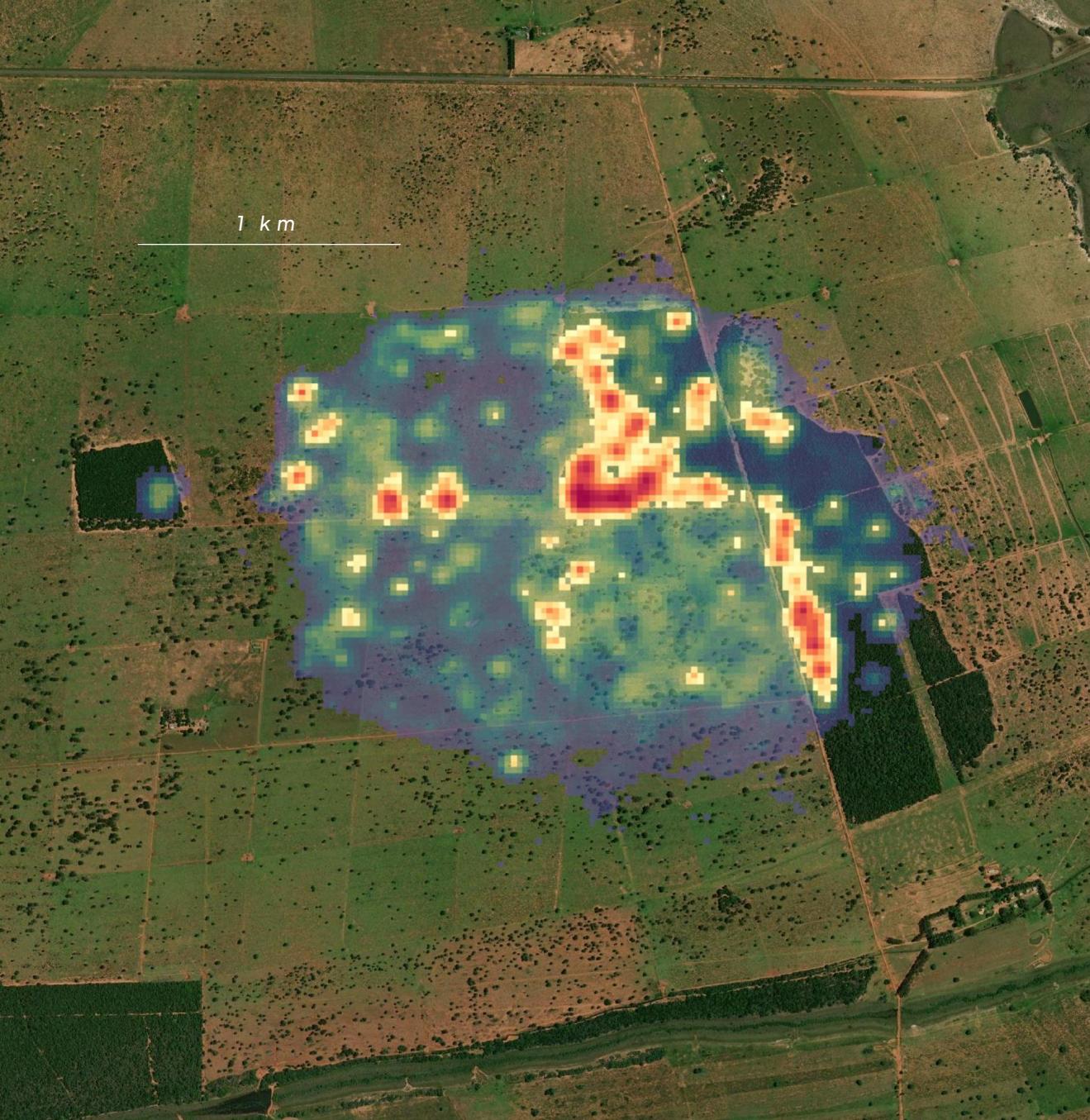


## Why estimate home range?

**Fragmentation** creates patches of suitable habitat surrounded by unsuitable habitat.



**Giant anteater**  
*(Myrmecophaga tridactyla)*



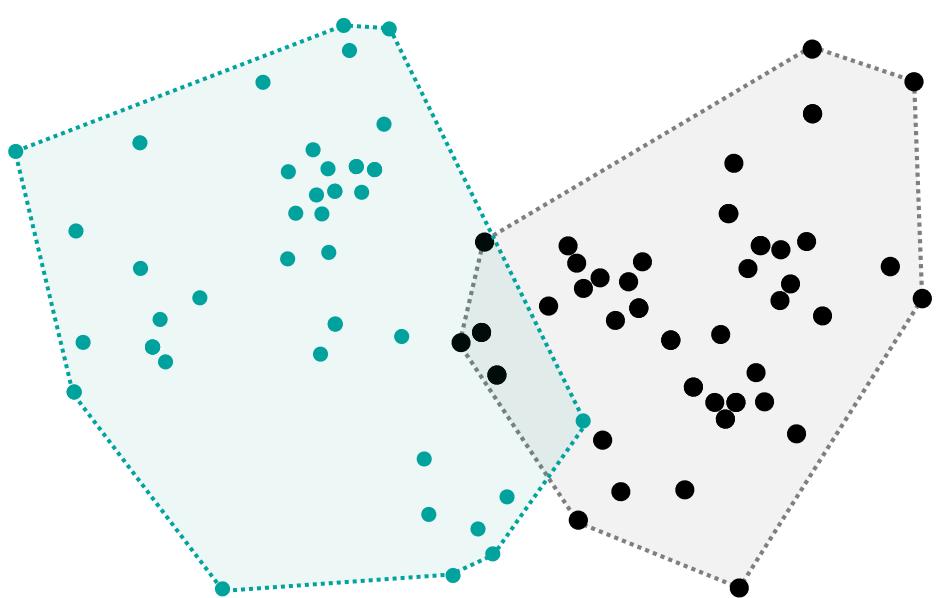


## How to estimate home range?

### – Minimum Convex Polygon (MCP):



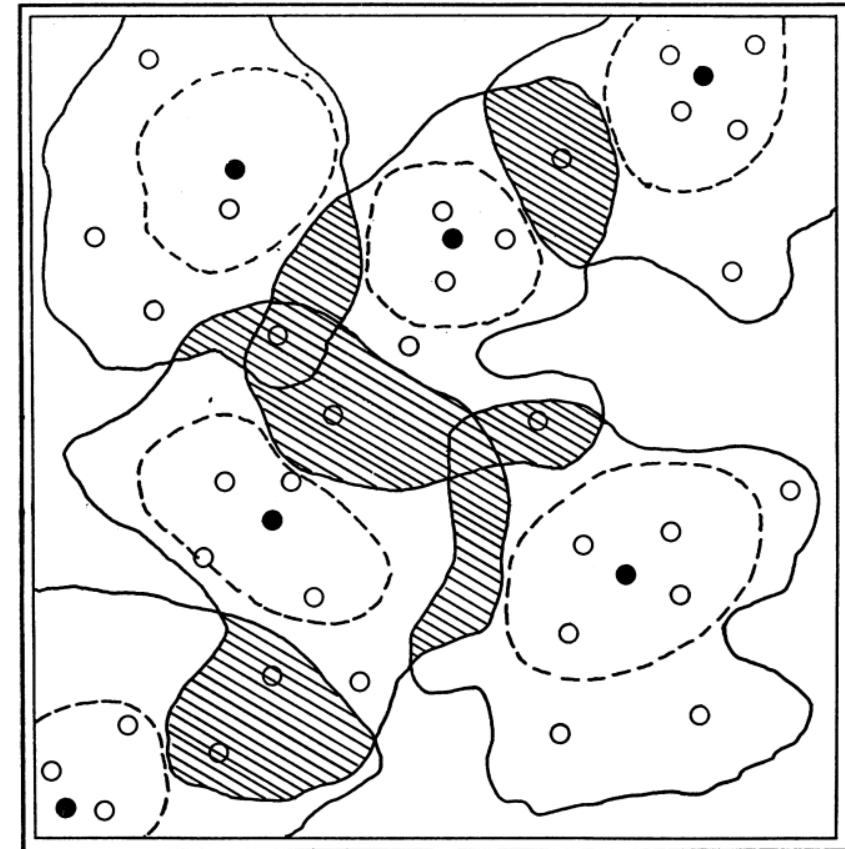
Mohr (1947)



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



Burt (1943)



— HOME RANGE BOUNDARY    ■ NEUTRAL AREA  
--- TERRITORIAL BOUNDARY    ● NESTING SITE  
BLANK--UNOCCUPIED SPACE    ○ REFUGE SITE

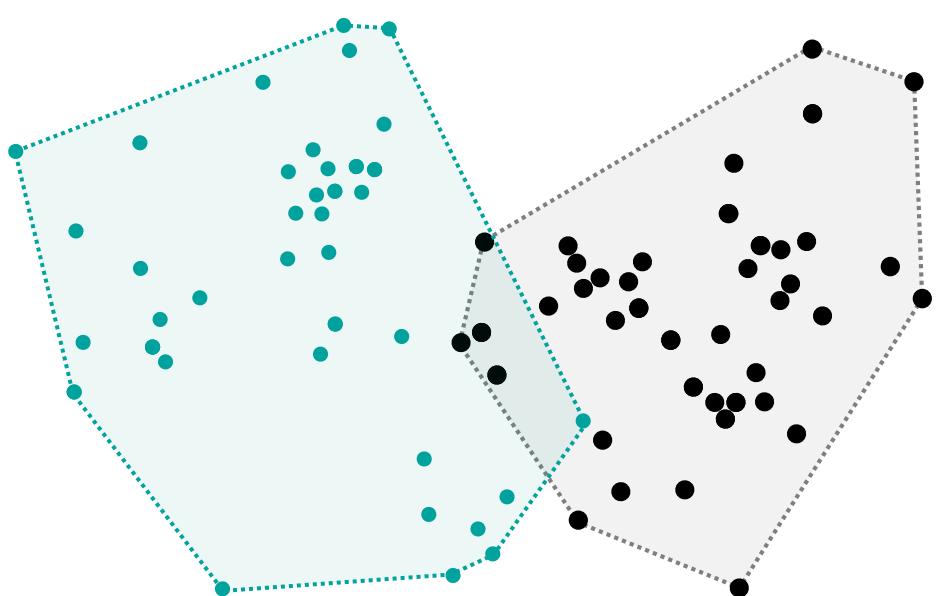
FIG. 1. Theoretical quadrat with six occupants of the same species and sex, showing territory and home range concepts as presented in text.



## How to estimate home range?

### – 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.



## How to estimate home range?

“

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.

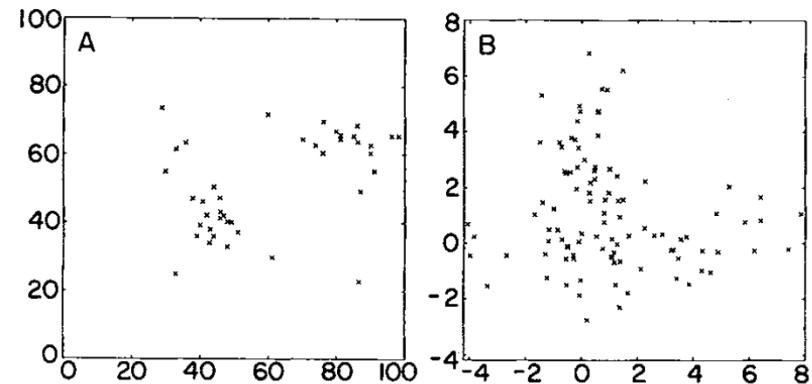


FIG. 1. Plots of (A) the DC data set and (B) the SIM data set.

Worton (1989)

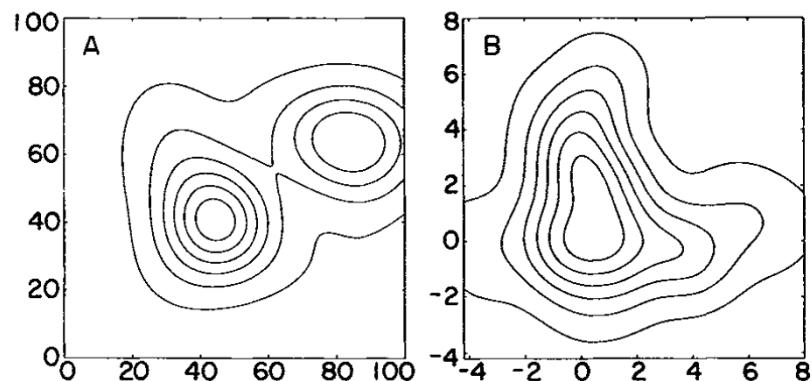


FIG. 2. **Fixed kernel density estimates** of the UD densities with the ad hoc choice of smoothing parameters for (A) the DC data set ( $h = 10.0$ ) and (B) the SIM data set ( $h = 1.0$ ).

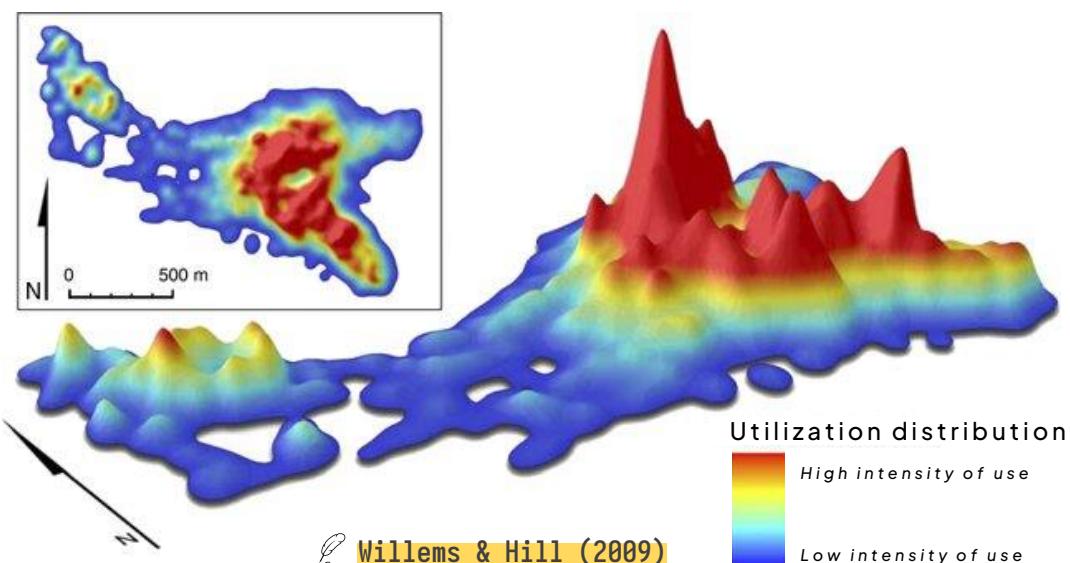


## How to estimate home range?

“

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)



### Utilization distributions (UDs):

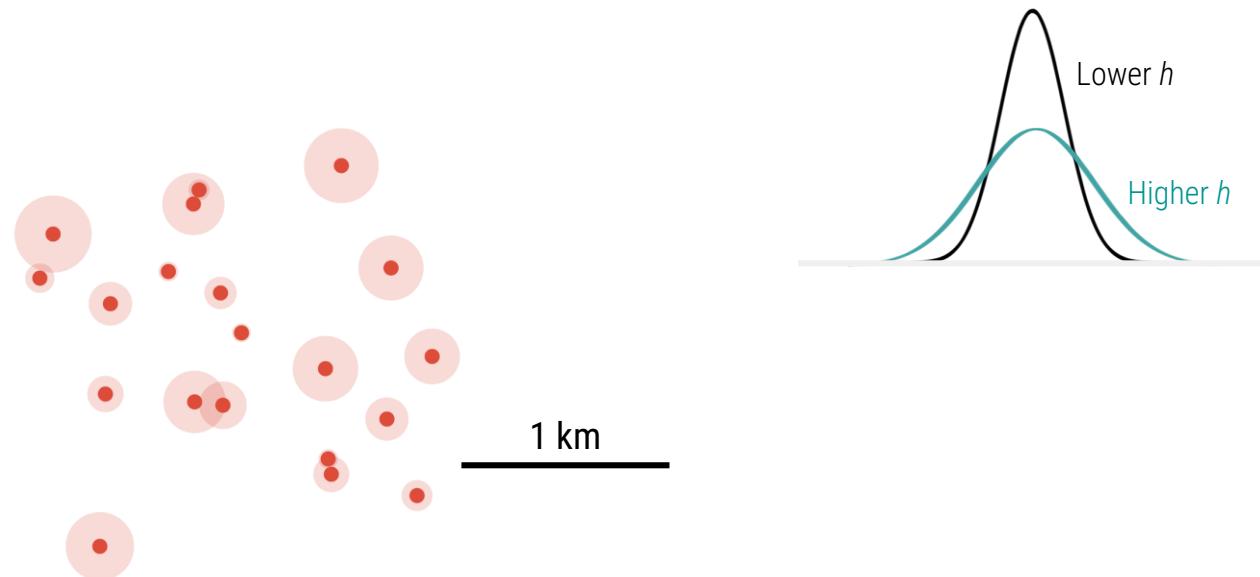
Characterizing space-use as a density function (accounts for intensity of use).





## Kernel Density Estimator (KDE)

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

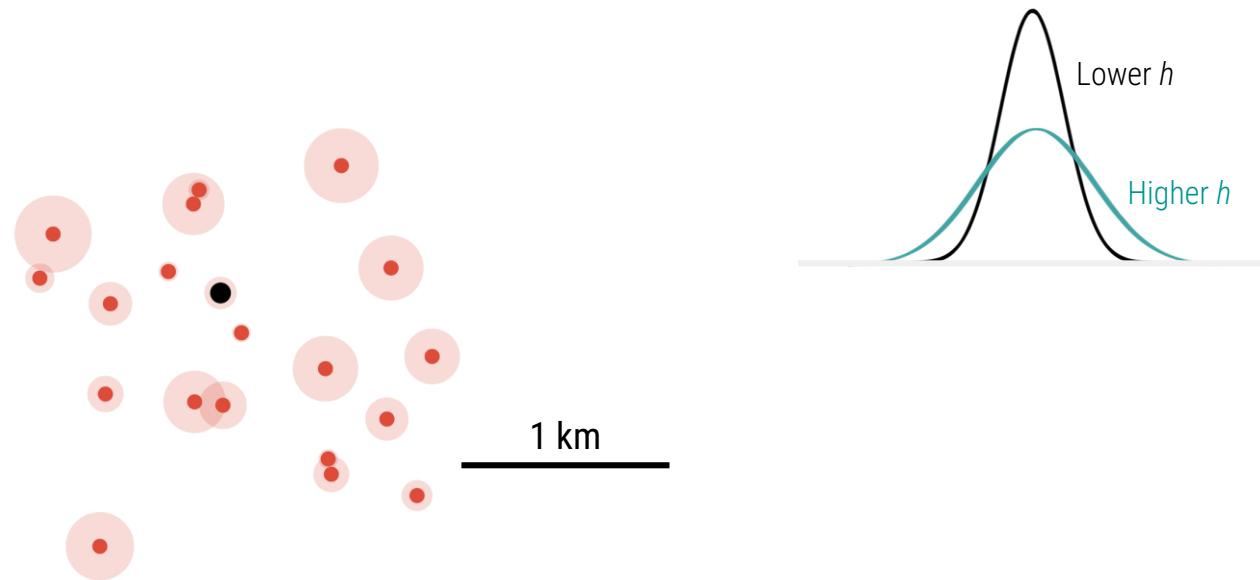


On an x–y plane, each location has a three-dimensional "hill", the **kernel**.  
The shape and width of the kernel, called the **bandwidth** ( $h$ ), can be selected using algorithms.



## Kernel Density Estimator (KDE)

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

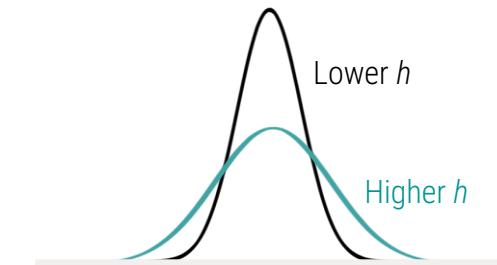
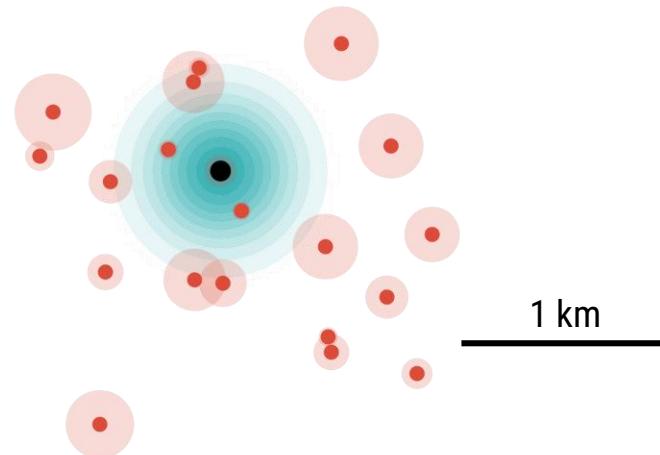


On an x–y plane, each location has a three-dimensional "hill", the **kernel**.  
The shape and width of the kernel, called the **bandwidth** ( $h$ ), can be selected using algorithms.



## Kernel Density Estimator (KDE)

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

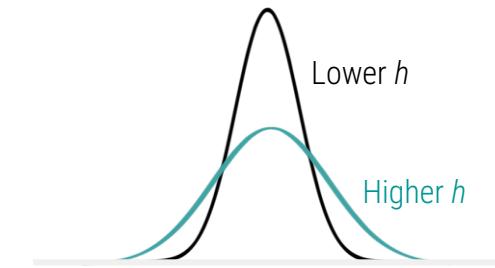
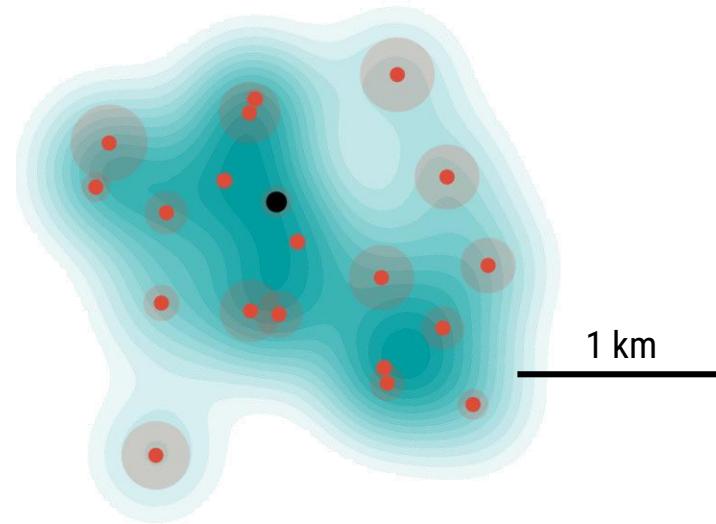


On an x–y plane, each location has a three-dimensional "hill", the **kernel**.  
The shape and width of the kernel, called the **bandwidth** ( $h$ ), can be selected using algorithms.



## Kernel Density Estimator (KDE)

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

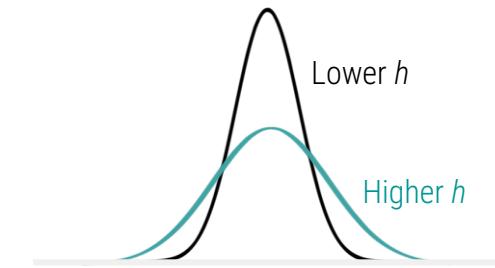
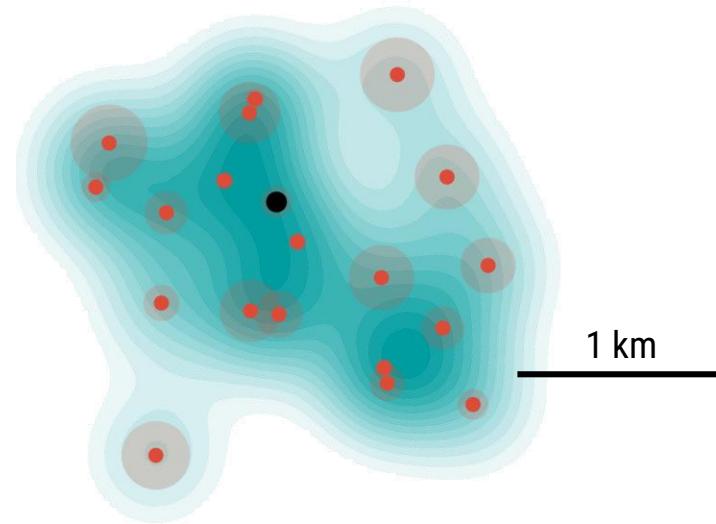


On an x–y plane, each location has a three-dimensional "hill", the **kernel**.  
The shape and width of the kernel, called the **bandwidth** ( $h$ ), can be selected using algorithms.



## Kernel Density Estimator (KDE)

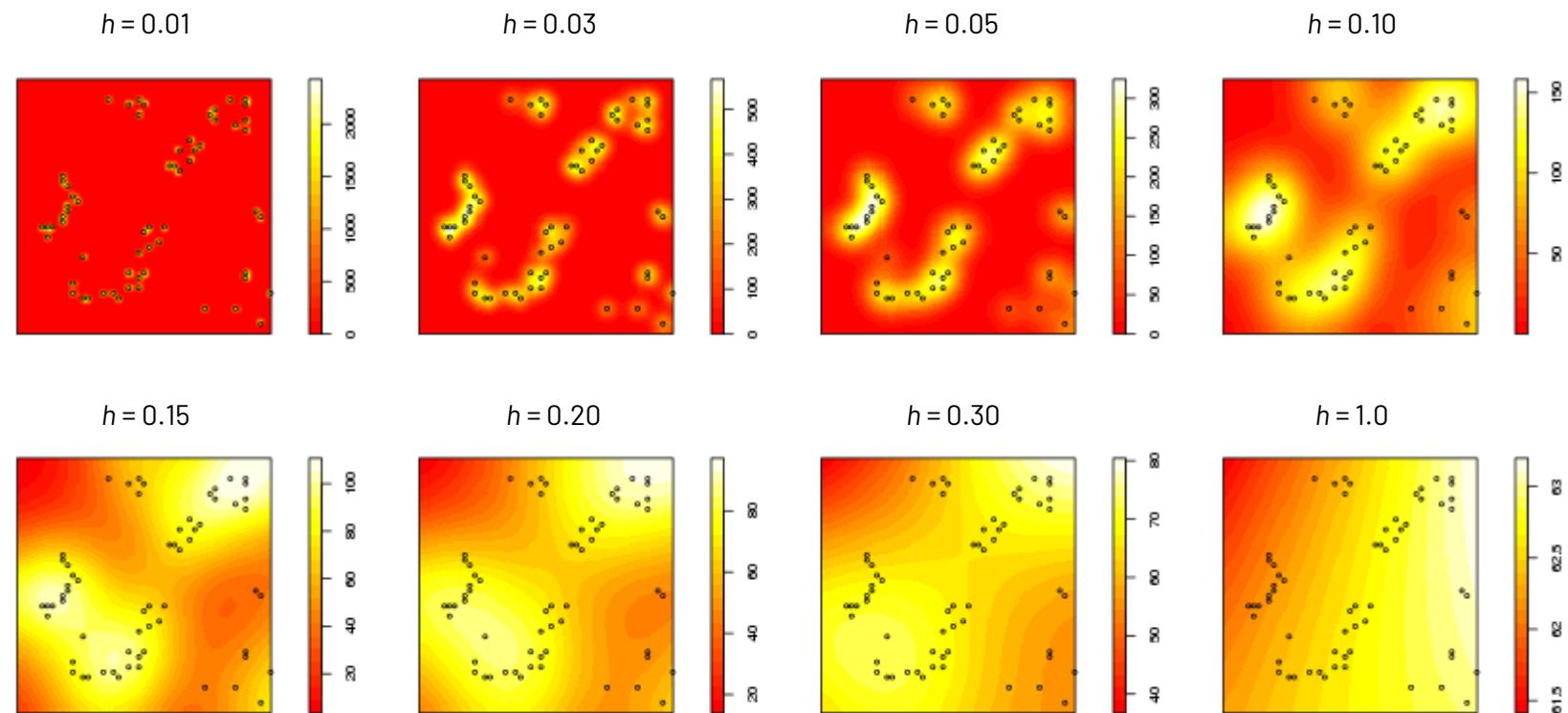
**KDEs** describe not just the borders of the home range, but the *intensity of use*.



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.



## Kernel Density Estimator (KDE)



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



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

↪ not explicitly modeling the causes of space use



**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**):

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

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



**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**):

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

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.**



Sensitive to **bandwidth selection**.

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

It assumes:

- locations are **independent**,
- data follows a normal (Gaussian) distribution.

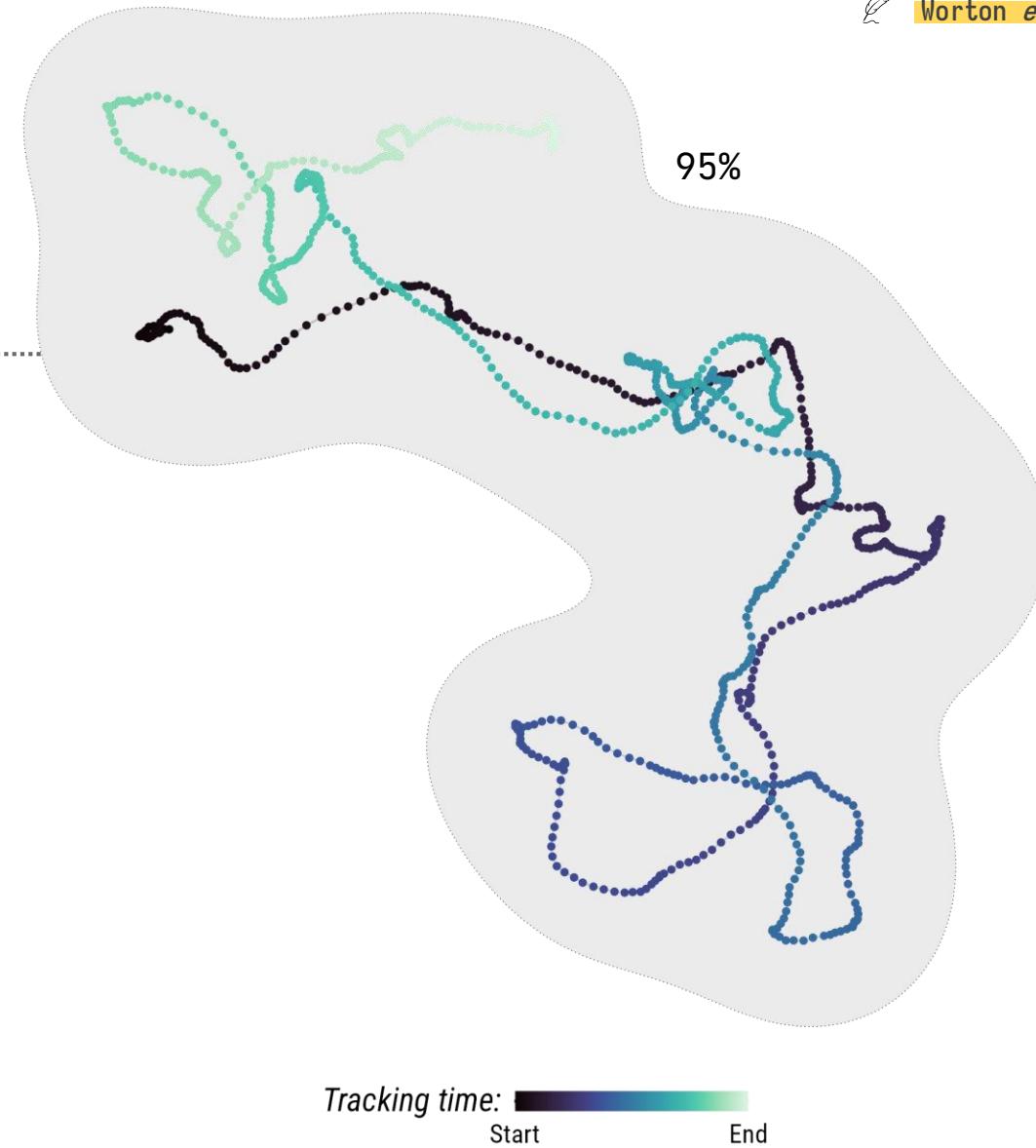
$$h_{ref} = 1.06 \cdot \sigma_i \cdot n^{-1/(4+d)}$$

where:

$\sigma$  is the standard deviation of your data

$n$  is the number of observations

$d$  is the number of dimensions of the data





Sensitive to **bandwidth selection**.

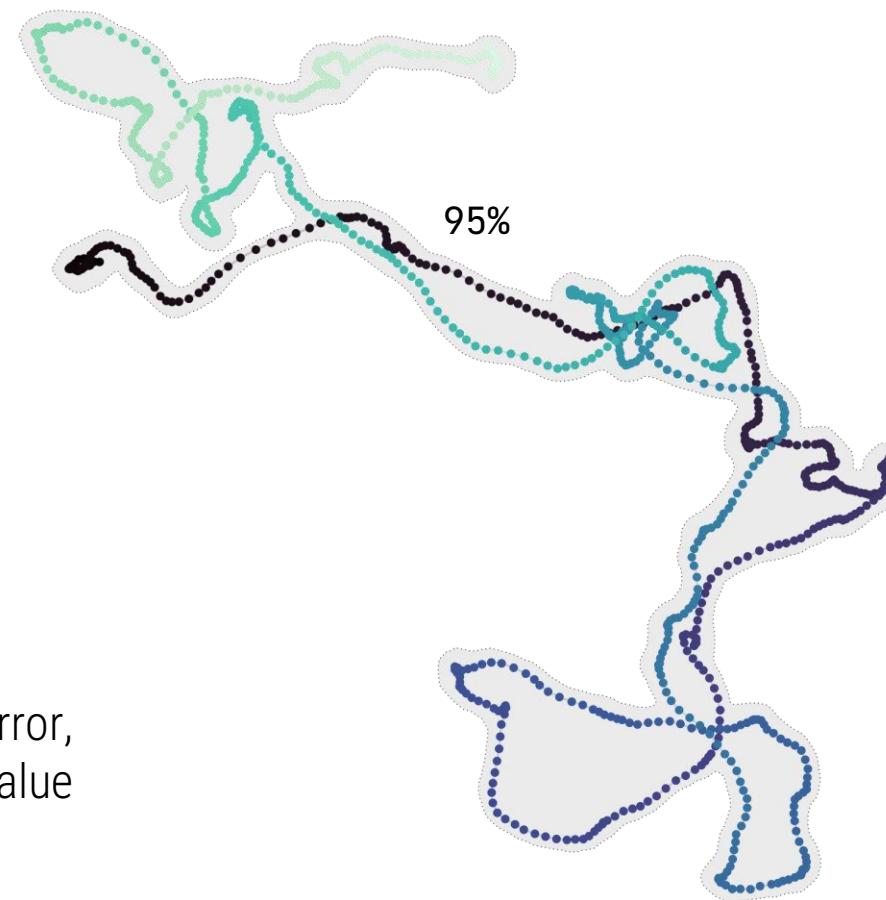
Least-squares cross-validation (**LSCV**) .....

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.



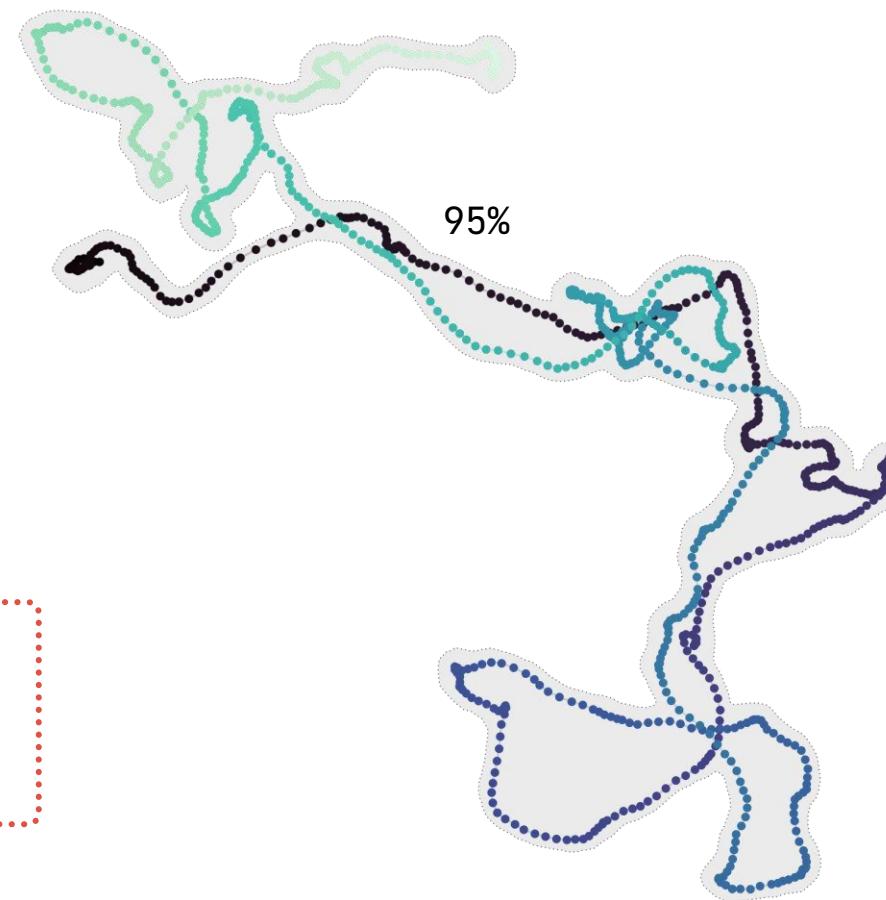
Tracking time: Start End



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.



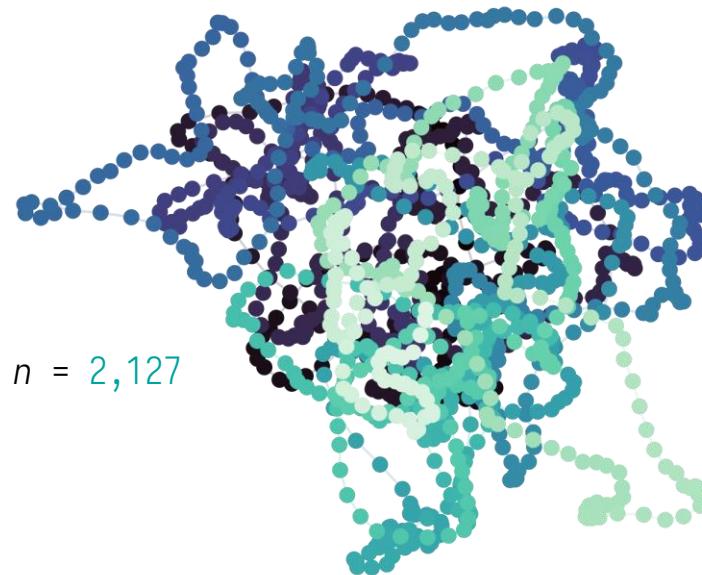
Tracking time: Start End



What about **data thinning**?

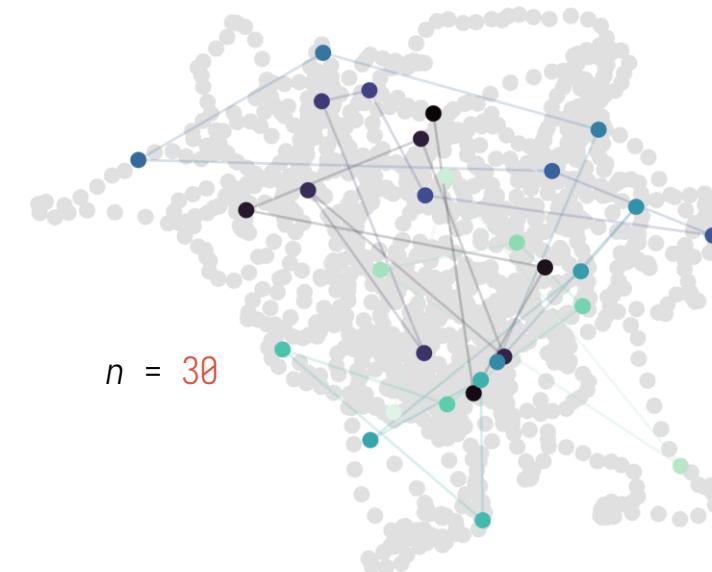
**Thinning** the data to achieve independence:

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



Tracking time: Start End

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



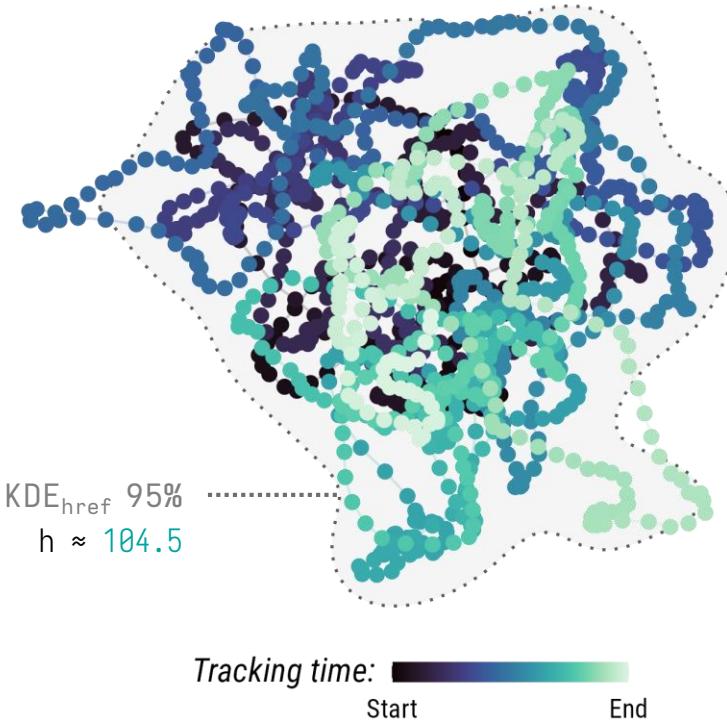
⚠  
Data loss of  $\approx 70.4\%$



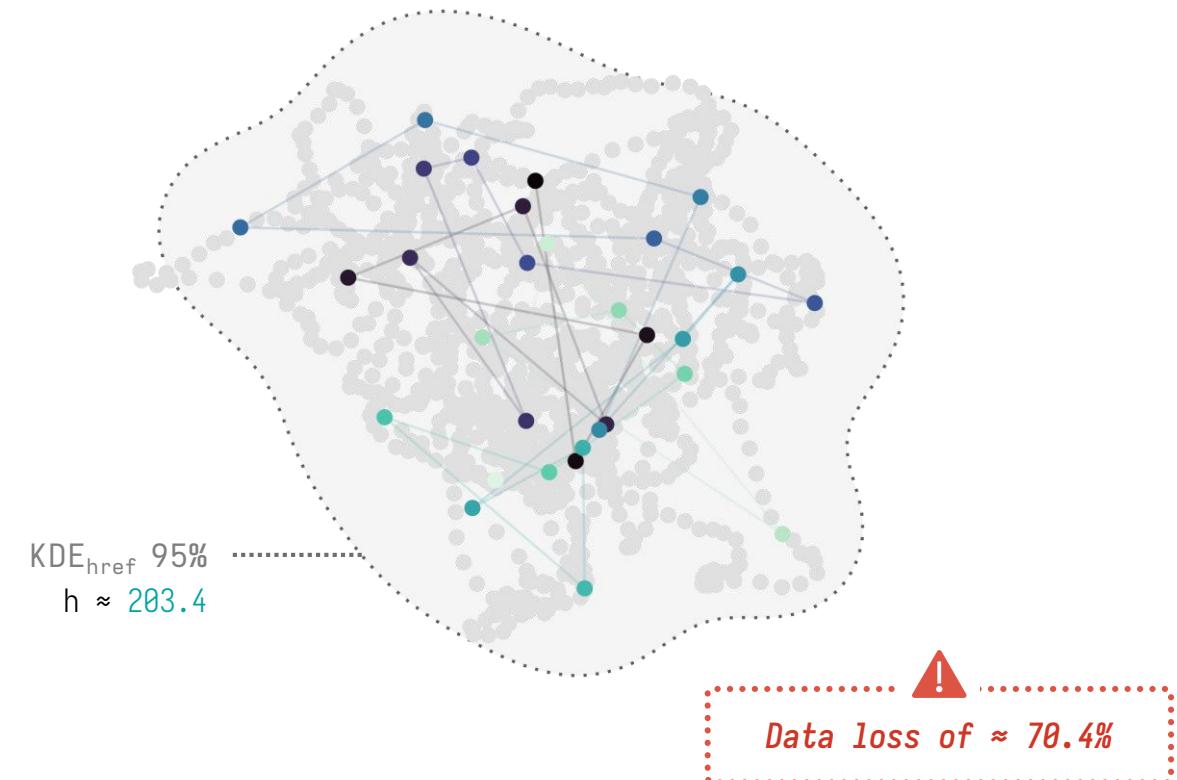
What about **data thinning**?

**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**.





What about **data thinning**?

**Thinning** the data to achieve independence:

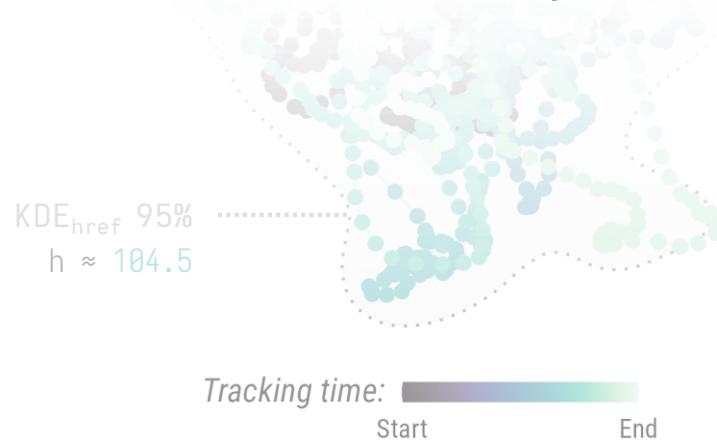
**Fig.** Tracking data representing hourly locations over **one month**.



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



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





## Utilization distribution

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

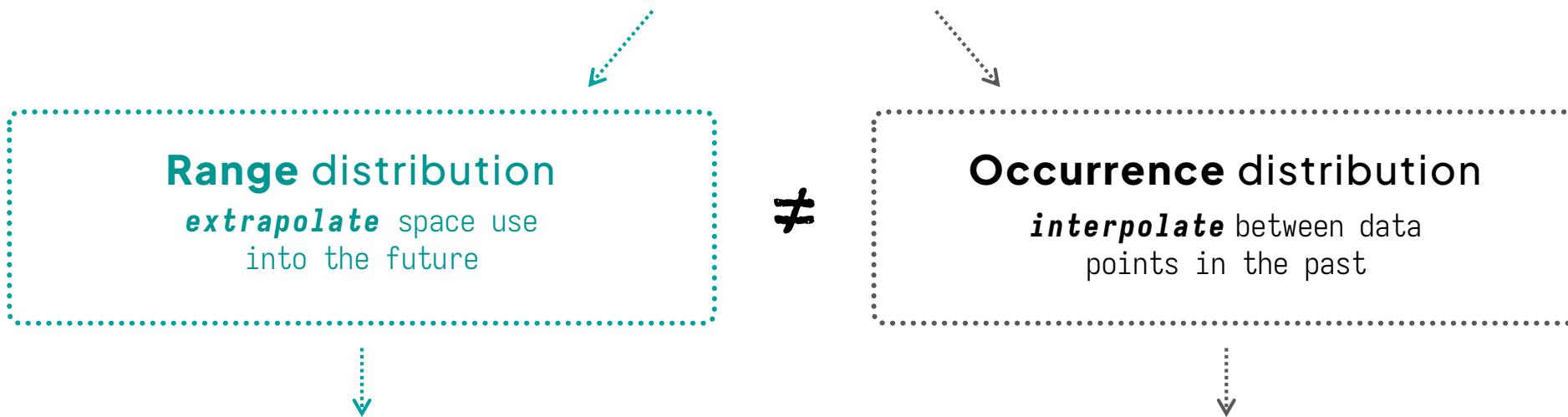


Alston et al. (2022)



## Utilization distribution

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



### 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.

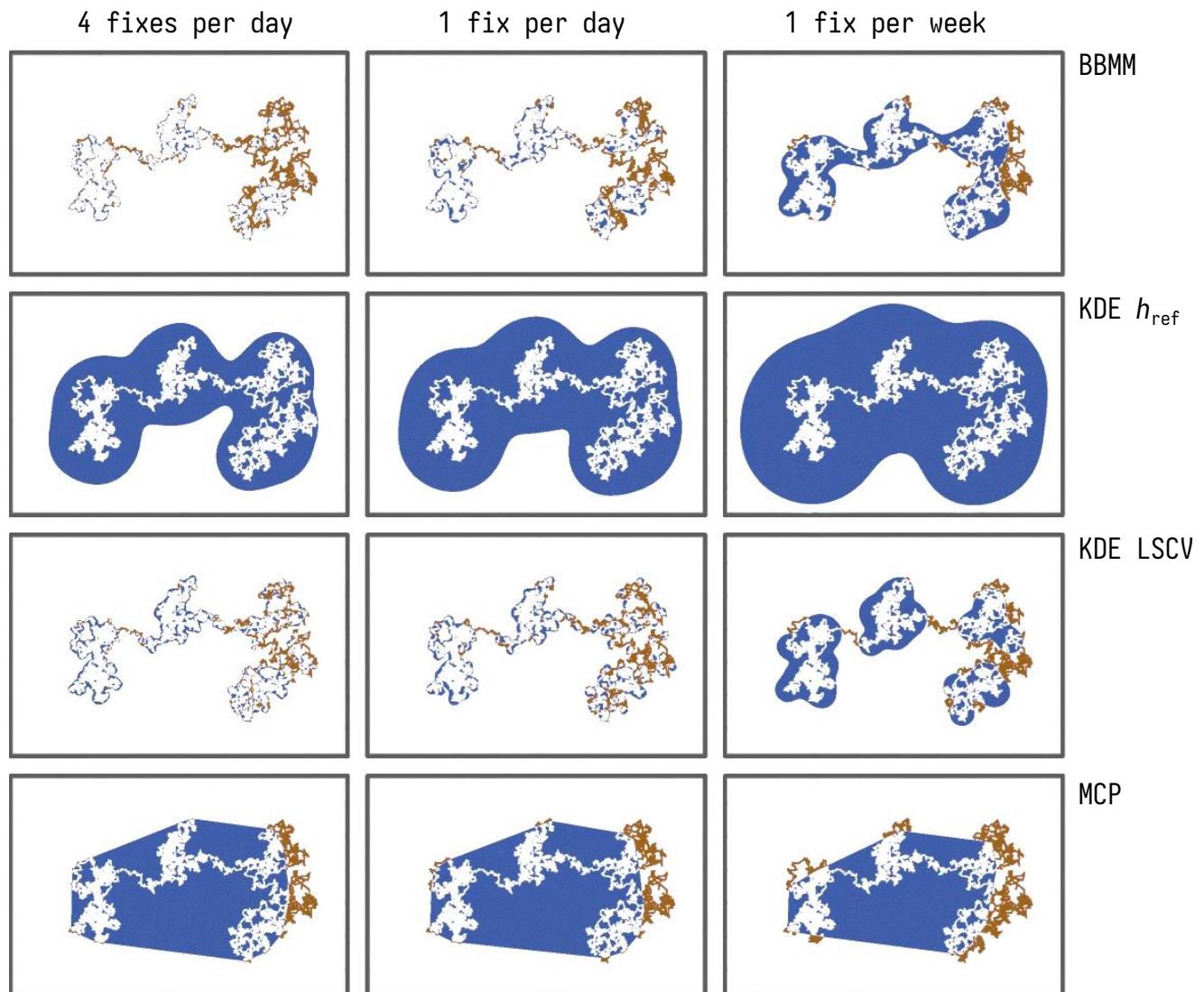
Sampling-dependent!

Alston et al. (2022)



## Other estimators

What is each method  
**actually** estimating?

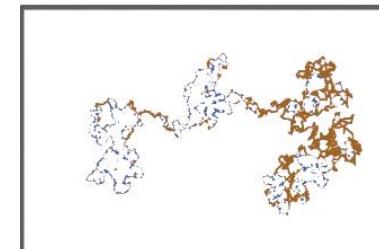


Silva et al. (2021)

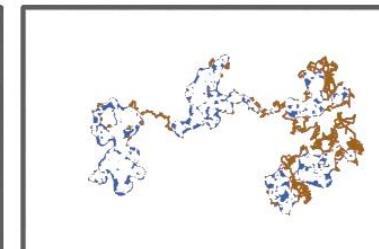


## Other estimators

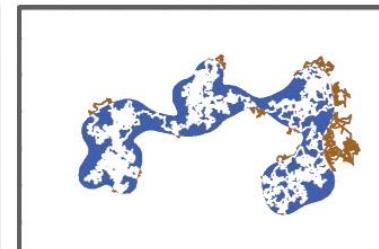
4 fixes per day



1 fix per day



1 fix per week



BBMM

KDE  $h_{ref}$



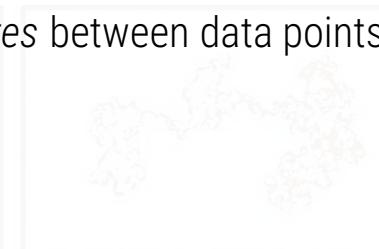
Should **not** be used for home range estimation.



## Occurrence distribution

interpolates between data points in the past

KDE LSCV



MCP



## What is each method actually estimating?

Not all methods are answering the same questions...



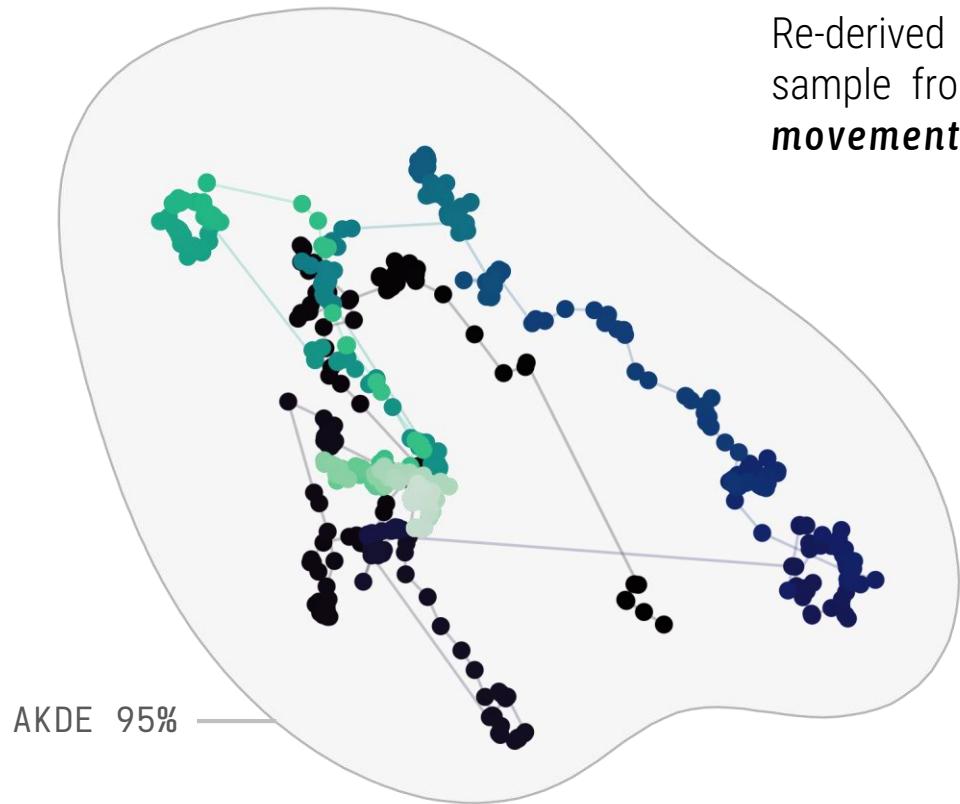
Silva et al. (2021)



What about **autocorrelated** data?

## 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*)

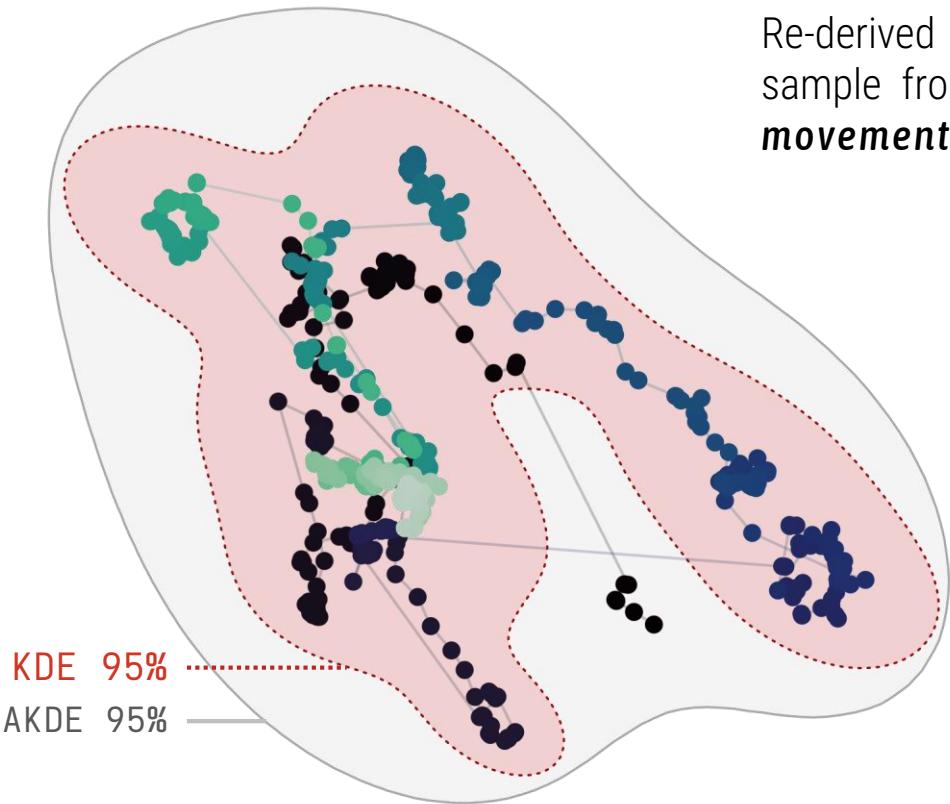
Duration  $\approx 1$  year  
 $\tau_p \approx 2.4$  months



What about **autocorrelated** data?

## 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*)

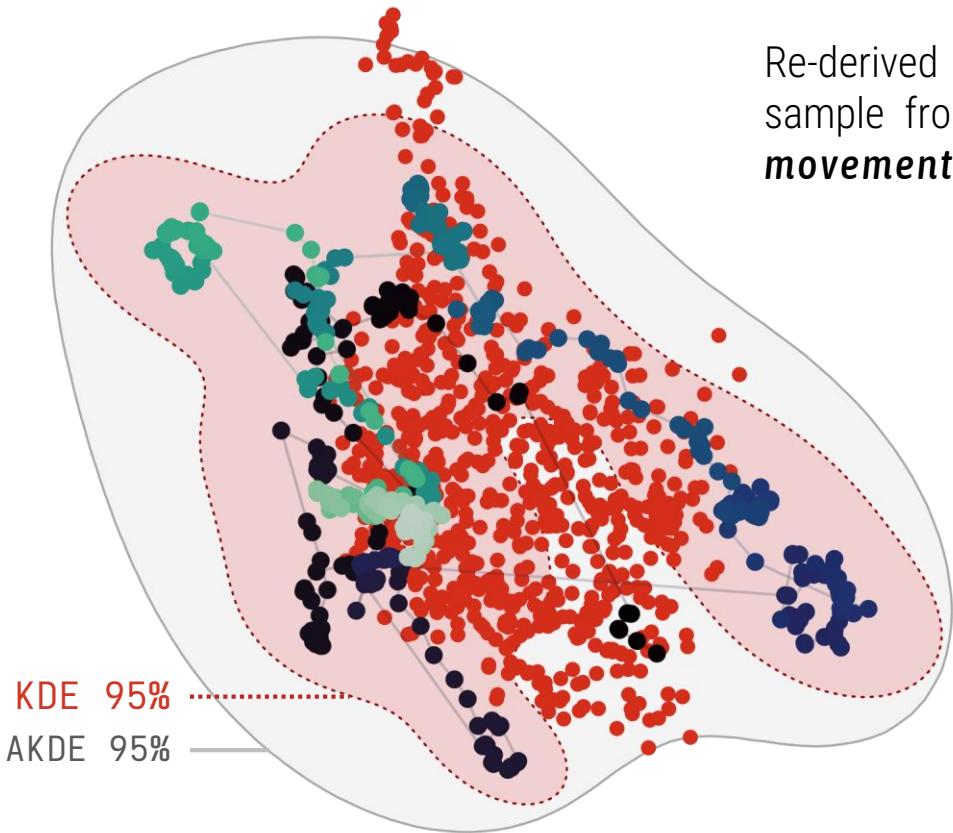
Duration  $\approx 1$  year  
 $\tau_p \approx 2.4$  months



What about **autocorrelated** data?

## 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*.



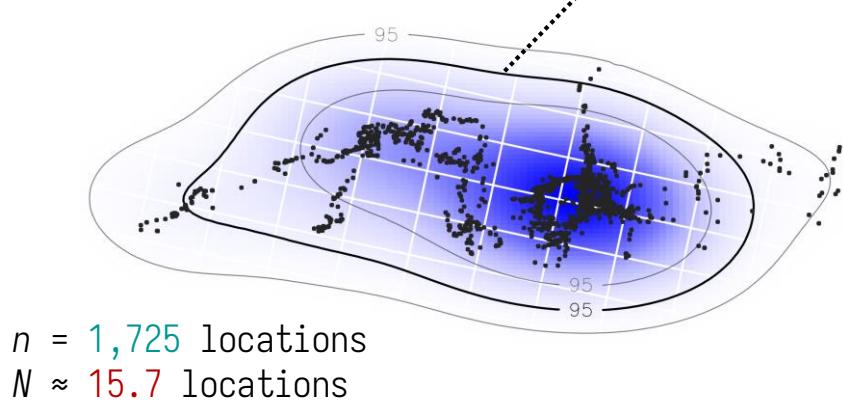
**Mongolian gazelle**  
(*Procapra gutturosa*)

Duration  $\approx$  2 years ..... showing an *extra year of data!*  
 $\tau_p \approx 2.4$  months



What about **autocorrelated** data?

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

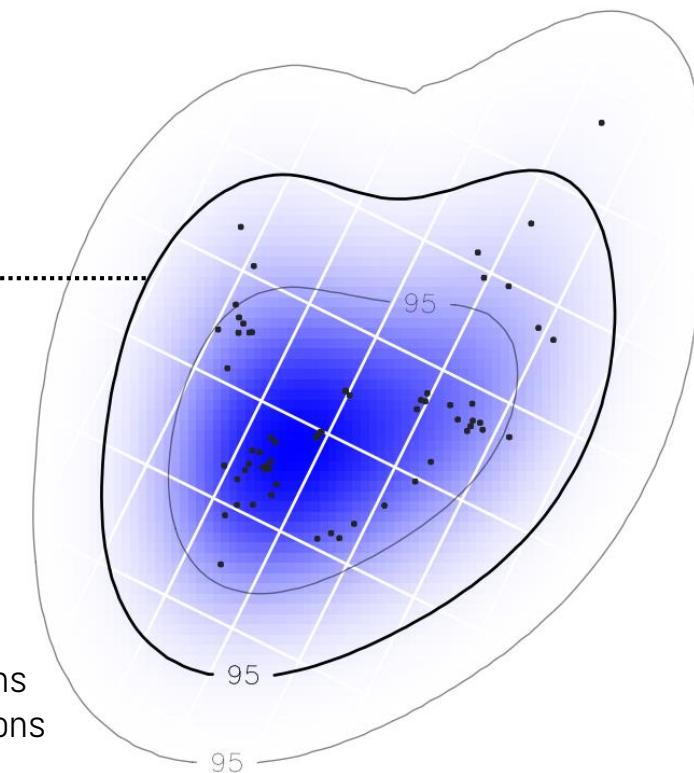


Home range area estimate  
757.5 km<sup>2</sup> (430.0 – 1,176.2)

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

$n = 66$  locations  
 $N \approx 6.9$  locations

Home range area estimate  
4,232.3 km<sup>2</sup> (1,681.7 – 7,939.5)





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

**Absolute sample size**

Total number of locations

$\neq$

**Effective sample size**

$n$

$T$

$+$

$\Delta t$

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

$$n = T + \Delta t$$

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  $T/\tau_p$

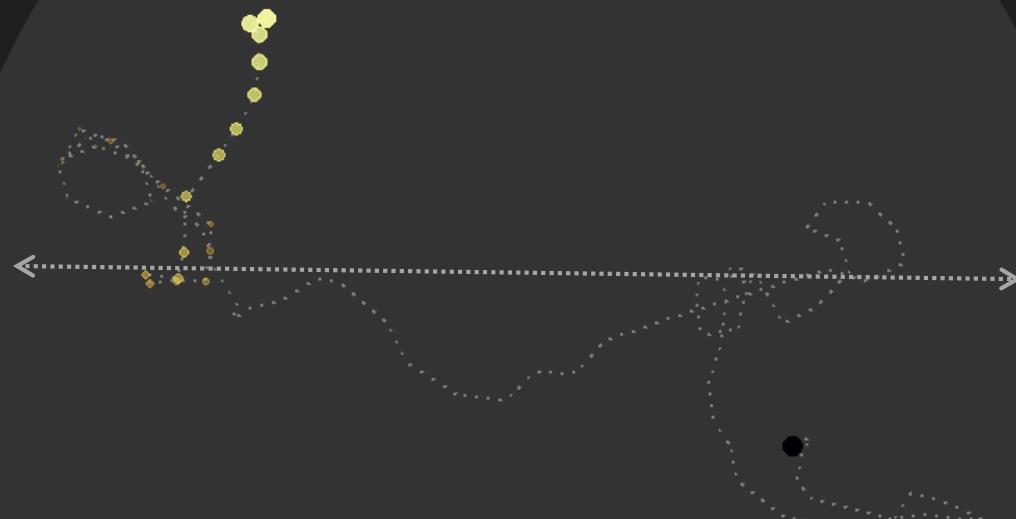
$T$  is the sampling duration

$\tau_p$  is the average **home range crossing time**



$\tau_p$

*Position autocorrelation  
timescale*



How long does it take for an animal to  
traverse the [linear extent of](#)  
[its home range?](#)

Duration = 1 day  
 $\tau_p$  = 1 day



$\tau_p$

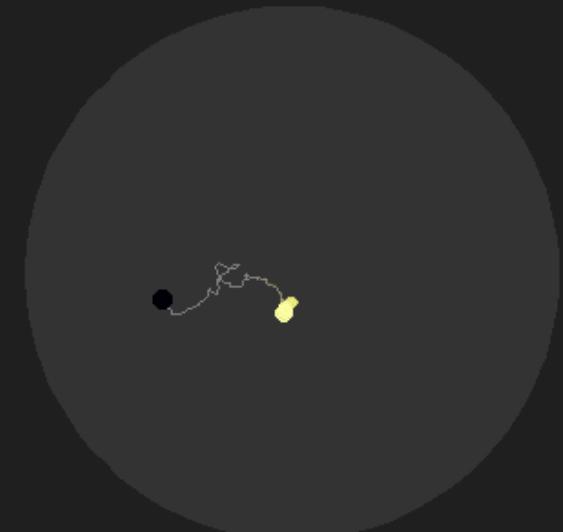
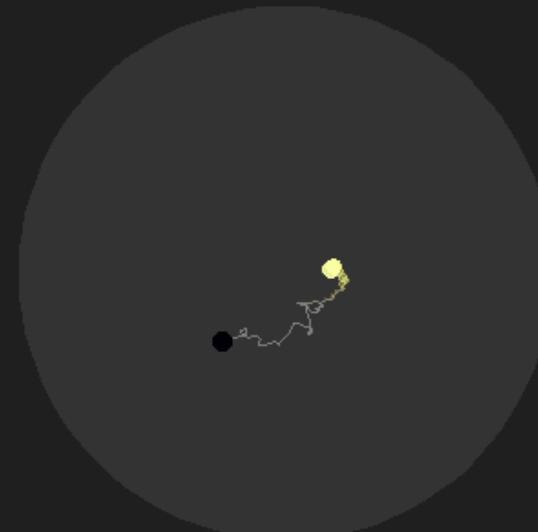
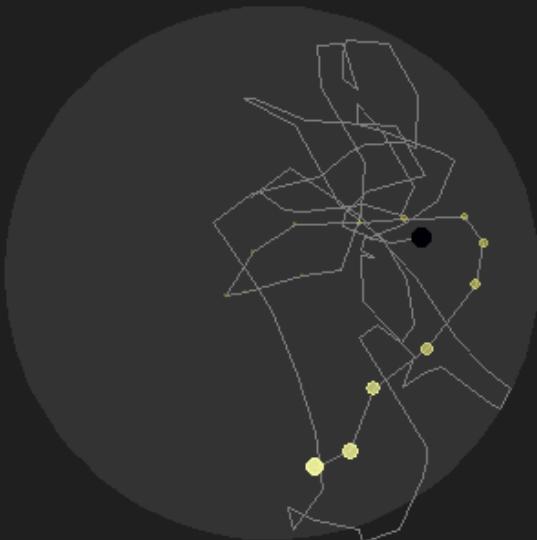
*Position autocorrelation  
timescale*

$\tau_p = 1$  hour

$\tau_p = 1$  day

$\tau_p = 5$  days

$\tau_p = 10$  days



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



$\tau_v$

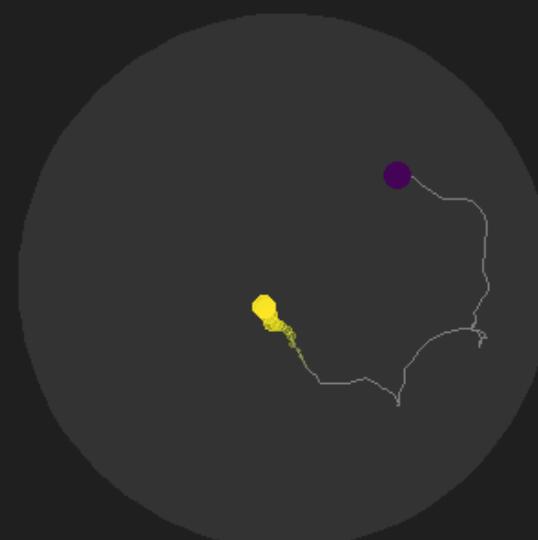
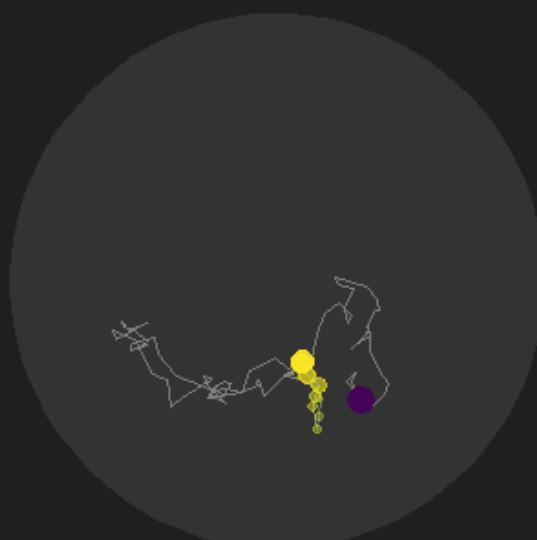
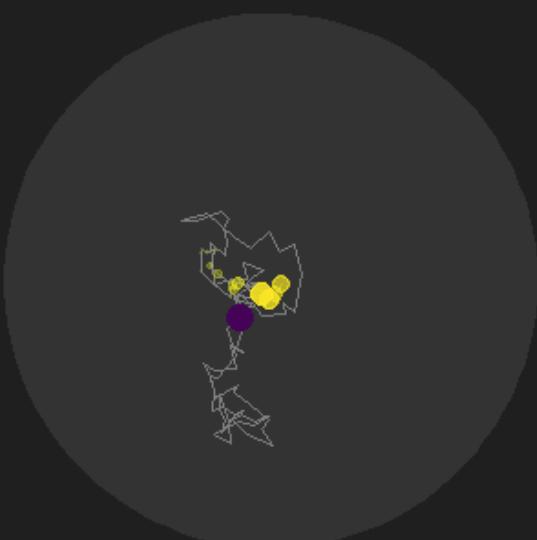
*Velocity autocorrelation  
timescale*

$\tau_v = 1$  minute

$\tau_v = 1$  hour

$\tau_v = 12$  hours

$\tau_v = 1$  day





## Understanding sample sizes



Sampling duration ≈ tracked for 1 year (360 days)

Sampling frequency ≈ 1 fix every hour



## Understanding sample sizes



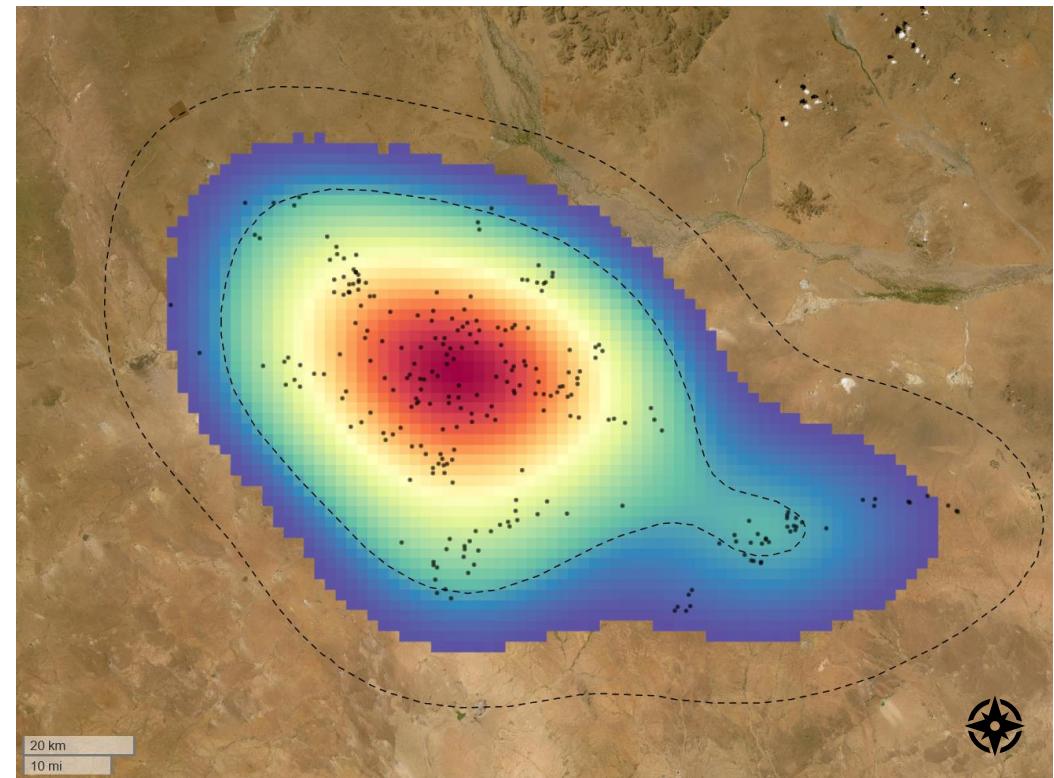
LC

Sampling duration  $\approx$  tracked for 1 year (360 days)

Sampling frequency  $\approx$  1 fix every hour

Absolute sample size,  $n = 866$  locations  
Effective sample size,  $N \approx 2.4$  locations

Home range area estimate  
 $61,692.5 \text{ km}^2$





## Understanding sample sizes



LC

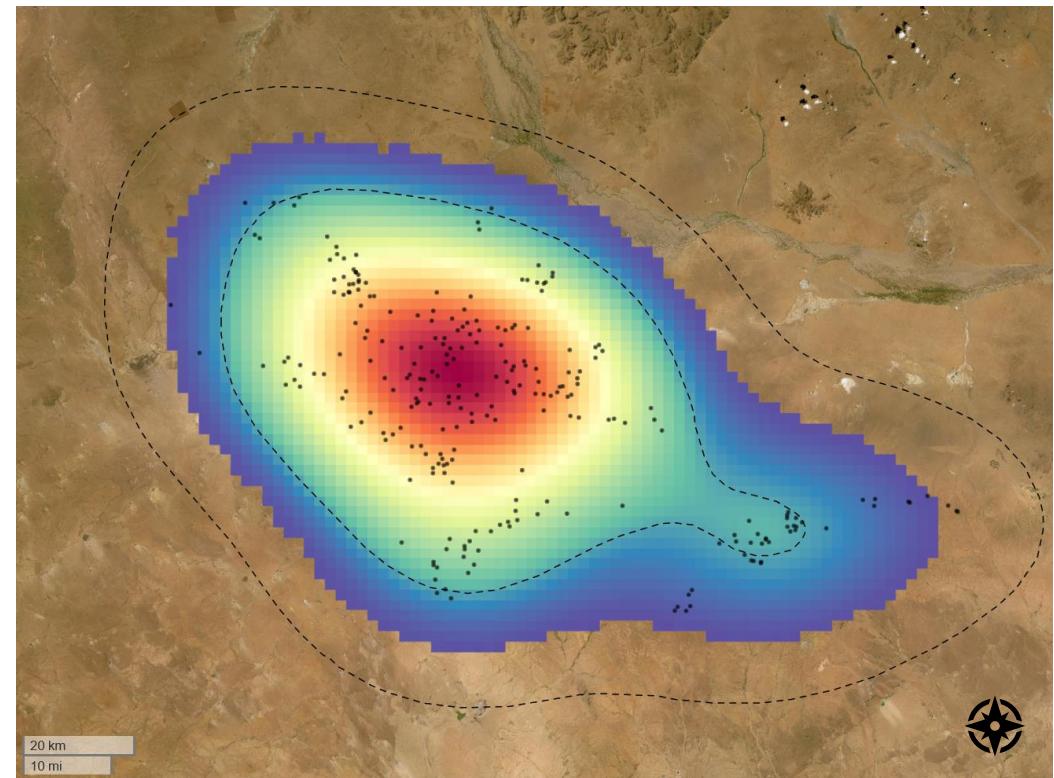
Sampling duration  $\approx$  tracked for 1 year (360 days)

Sampling frequency  $\approx$  1 fix every hour

Absolute sample size,  $n = 866$  locations  
Effective sample size,  $N \approx 2.4$  locations

$$\tau_p \approx 141.2 \text{ days}$$

Home range area estimate  
 $61,692.5 \text{ km}^2$





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**.



How can we measure and visualize autocorrelation?

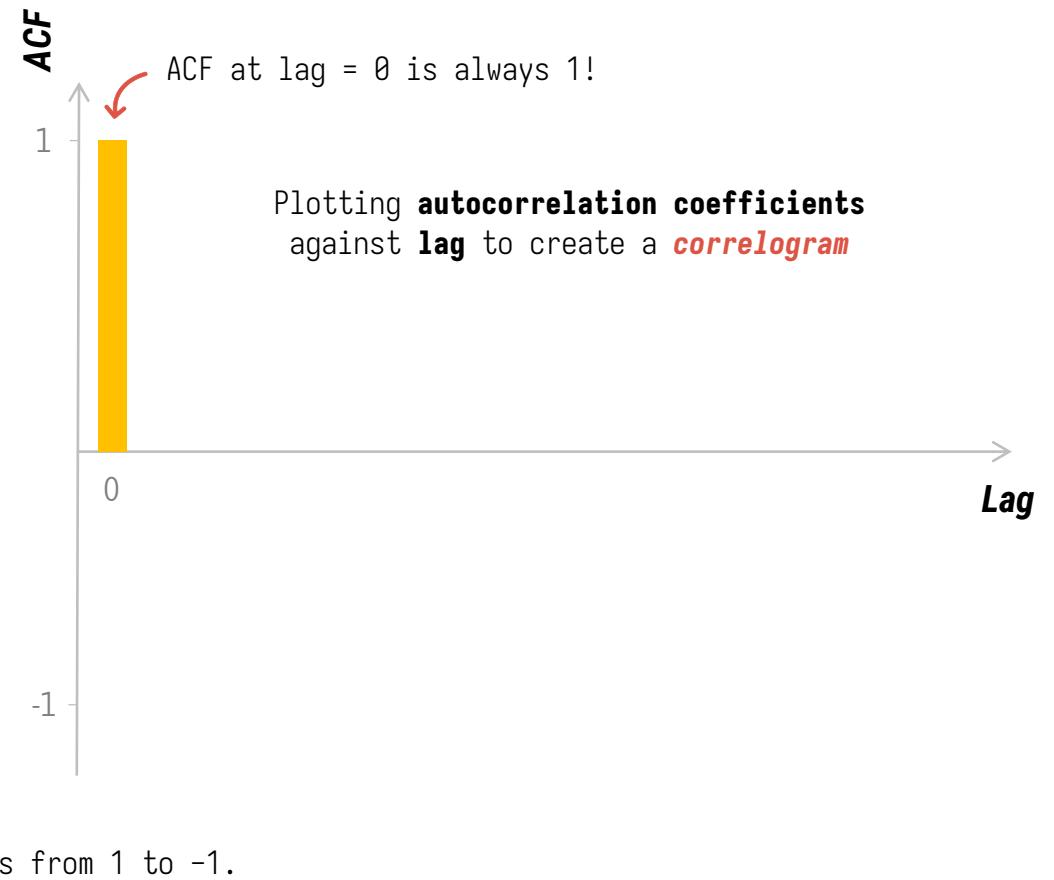
The **autocorrelation function (ACF)** is a valuable tool for investigating properties of an empirical time series. We typically visualize it through the **correlogram**.

$$ACF_0 = \text{var}_{t+0} / \text{var}_t$$

**Lag = 0**



$t_1$





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**.

$$ACF_1 = \text{var}_{t+1} / \text{var}_t$$

Lag = 1



Represents a shift in time.





## Autocorrelation

How can we measure and visualize autocorrelation?

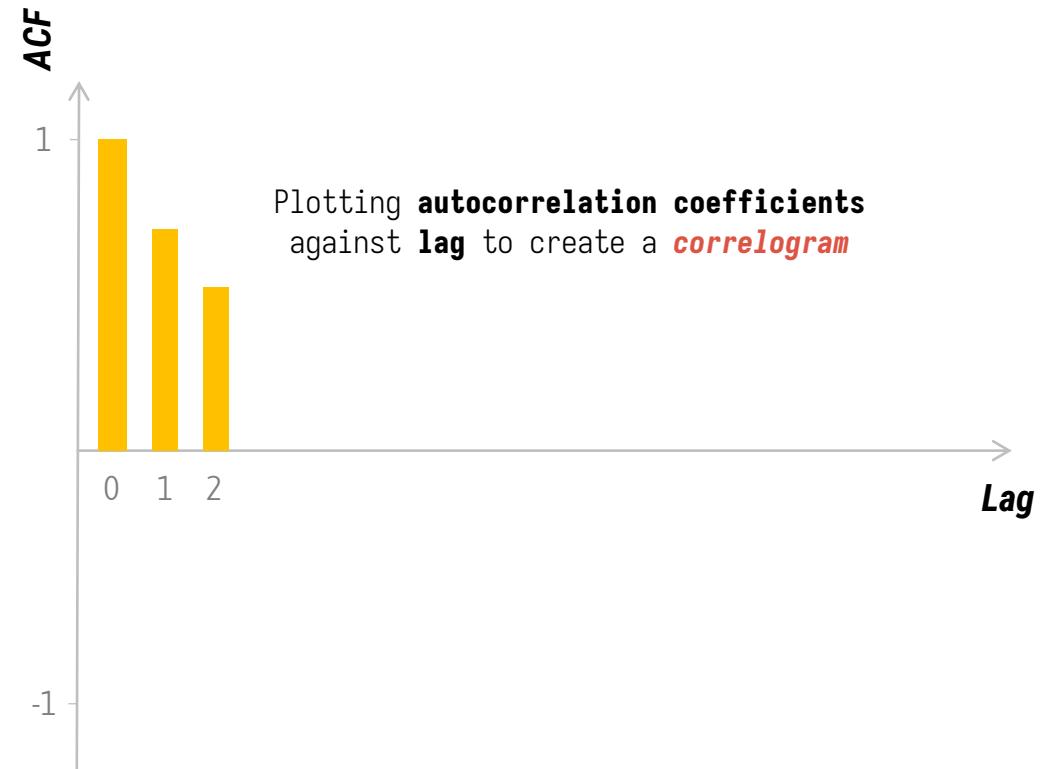
The **autocorrelation function (ACF)** is a valuable tool for investigating properties of an empirical time series. We typically visualize it through the **correlogram**.

$$ACF_2 = \text{var}_{t+2} / \text{var}_t$$

Lag = 2



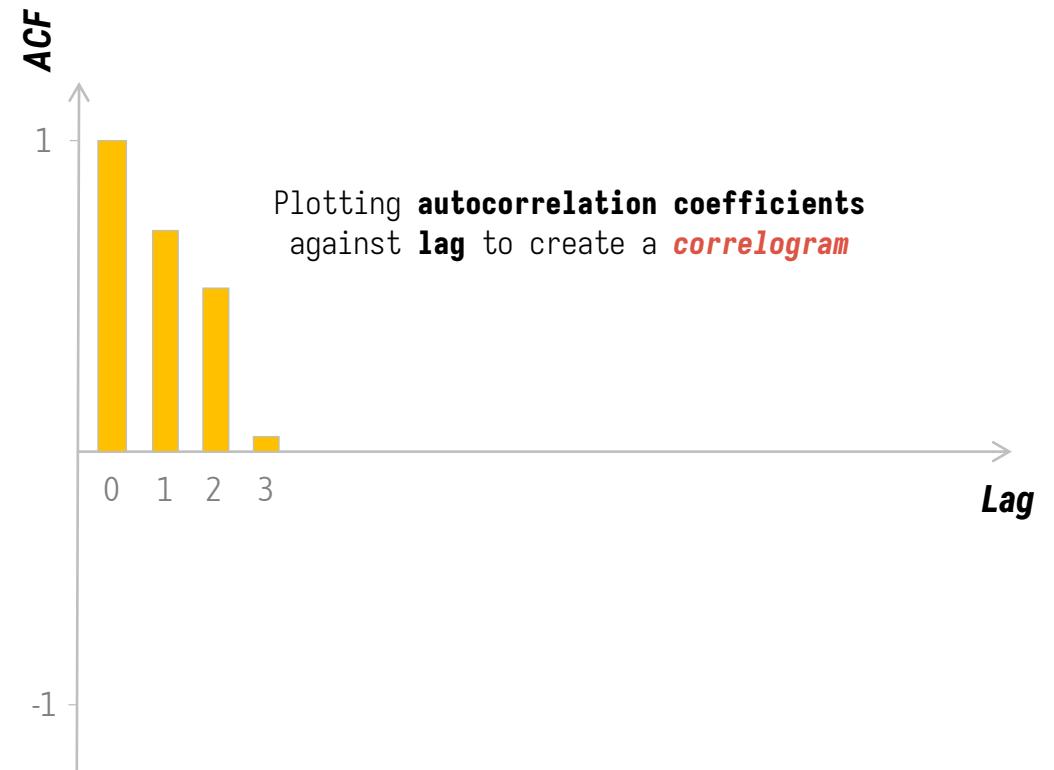
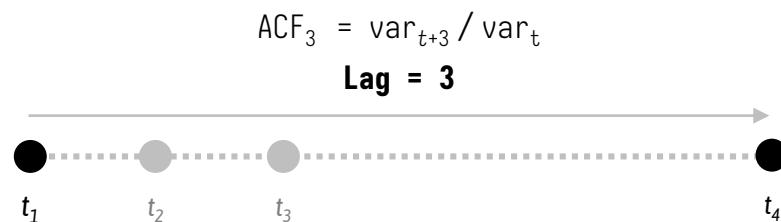
Represents a shift in time.





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**.

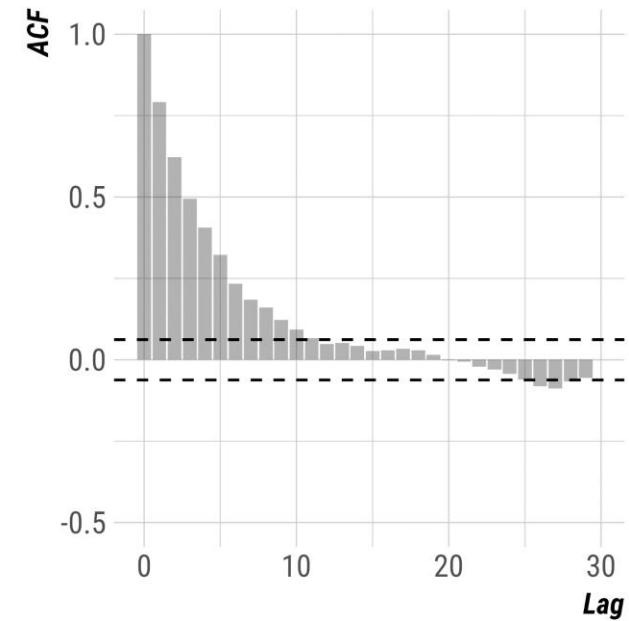
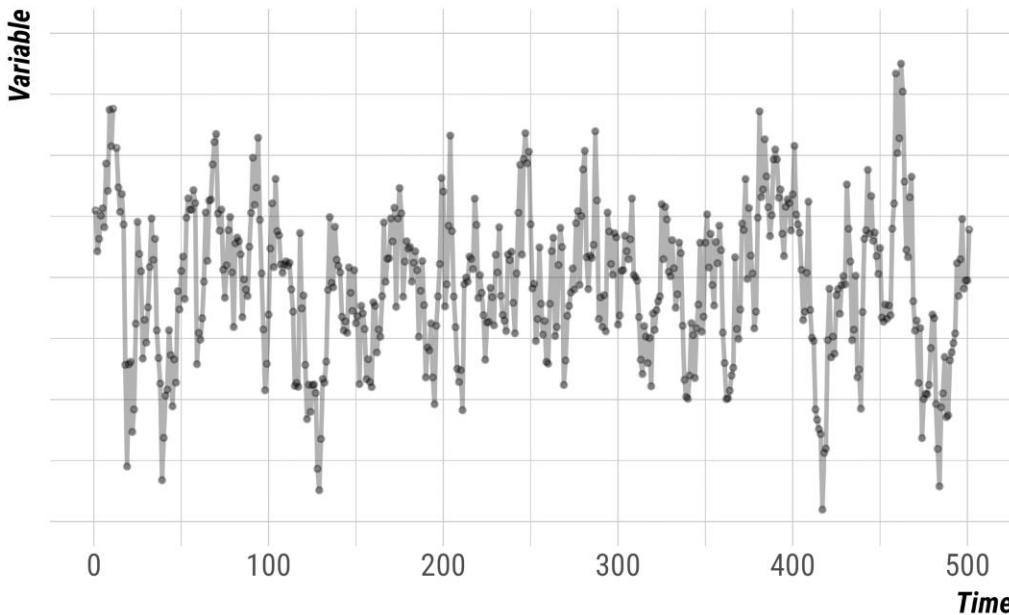




## Autocorrelation

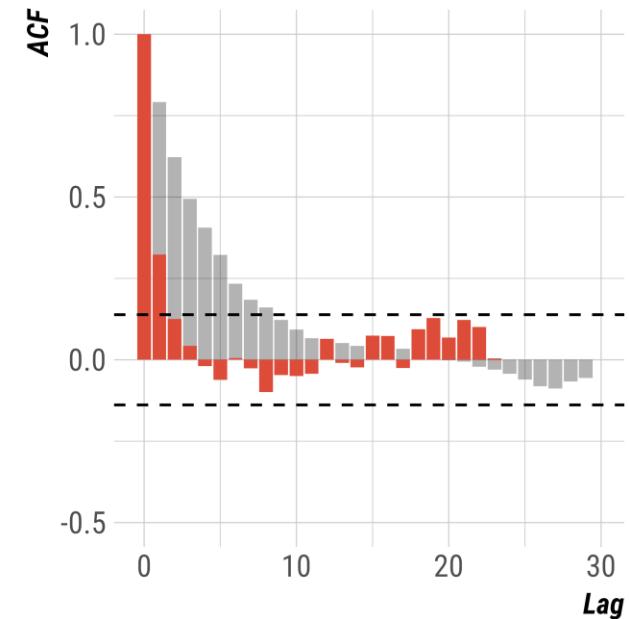
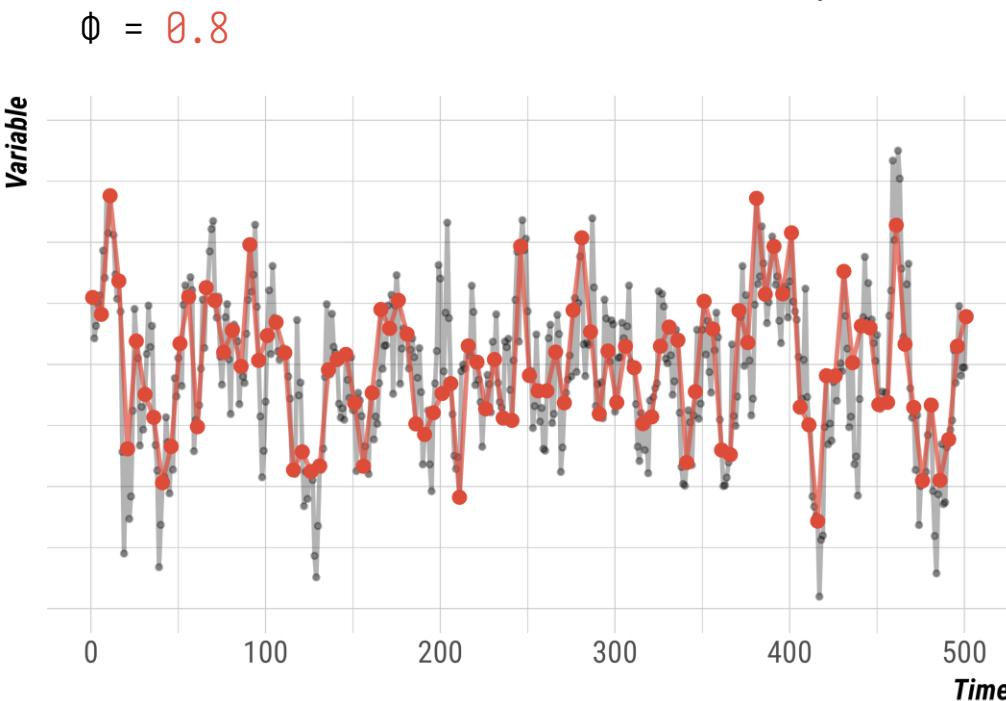
The **sampling frequency** (or **duration**) of data collection can influence our ability to detect autocorrelation.

$$\phi = 0.8$$



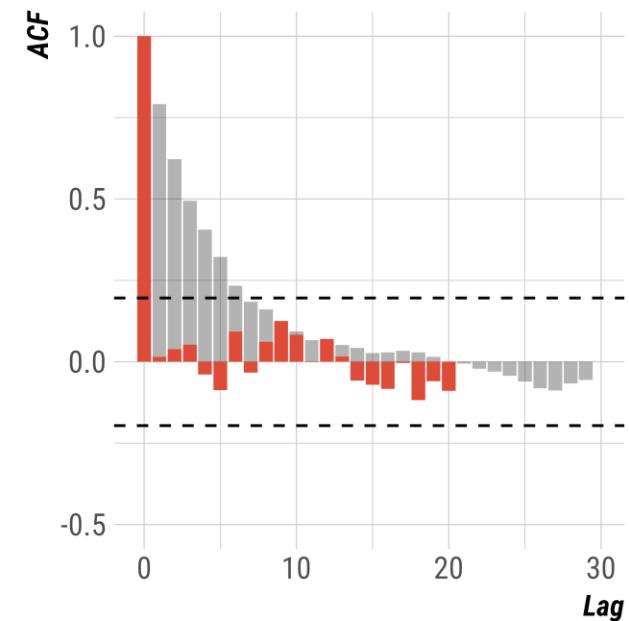
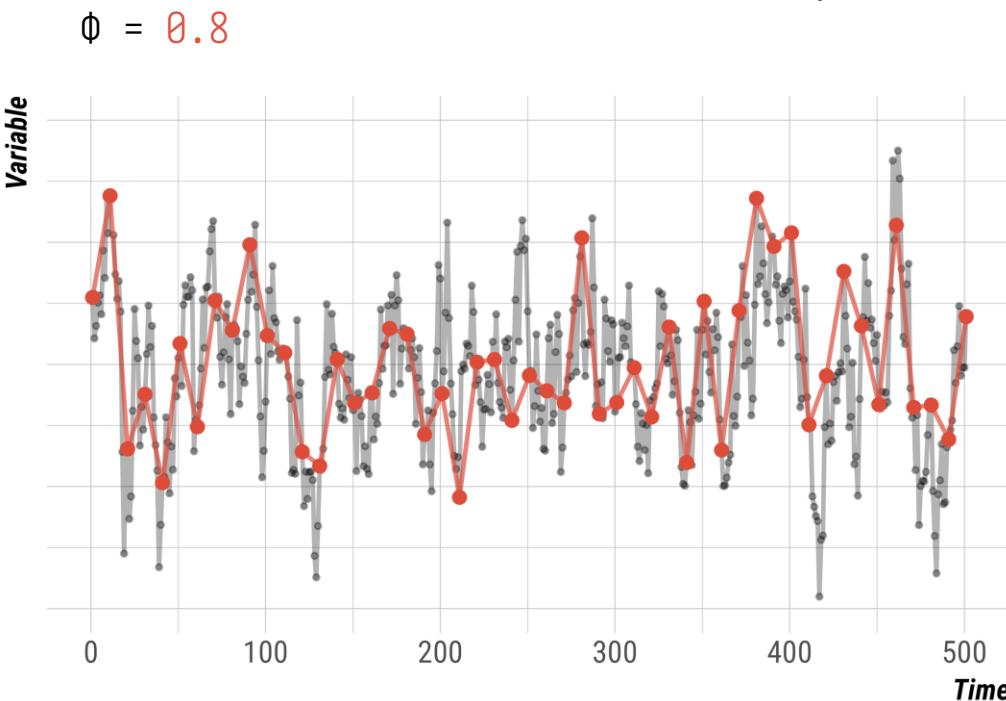


The **sampling frequency** (or **duration**) of data collection can influence our ability to detect autocorrelation.



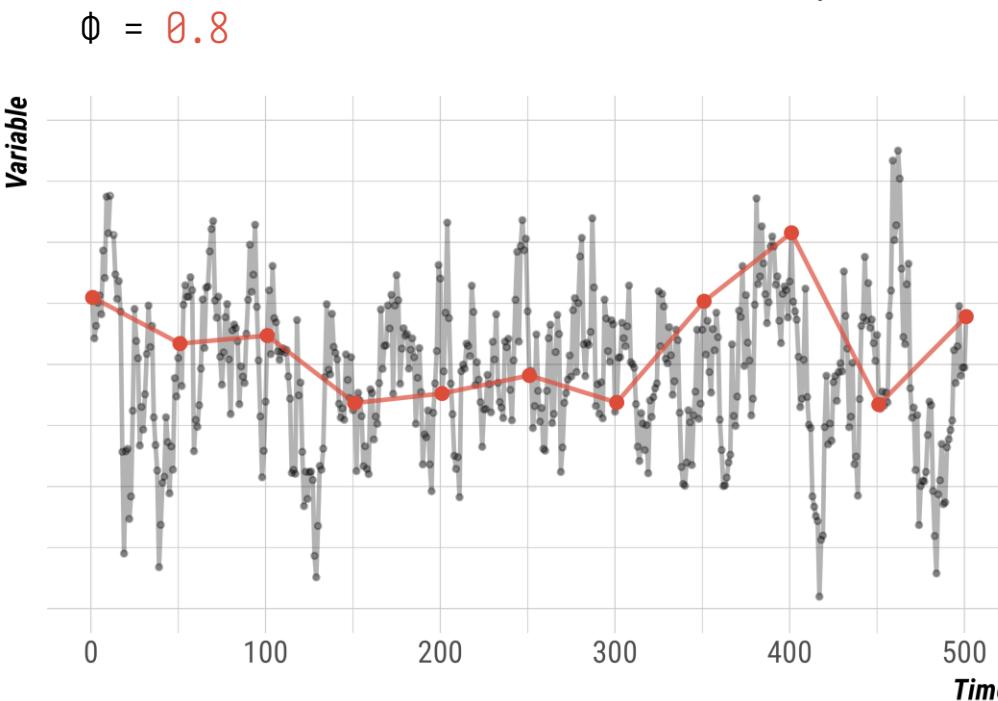


The **sampling frequency** (or **duration**) of data collection can influence our ability to detect autocorrelation.

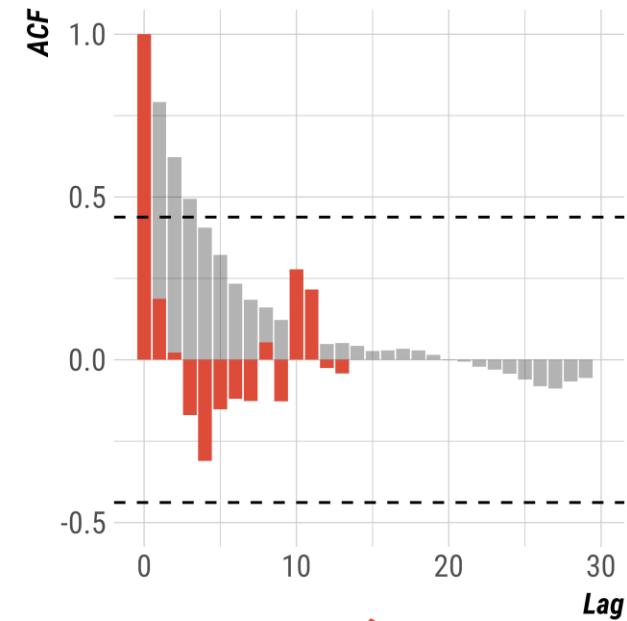




The **sampling frequency** (or **duration**) of data collection can influence our ability to detect autocorrelation.



**Sampling frequency** should encapsulate the **process timescale** for accurate representation.

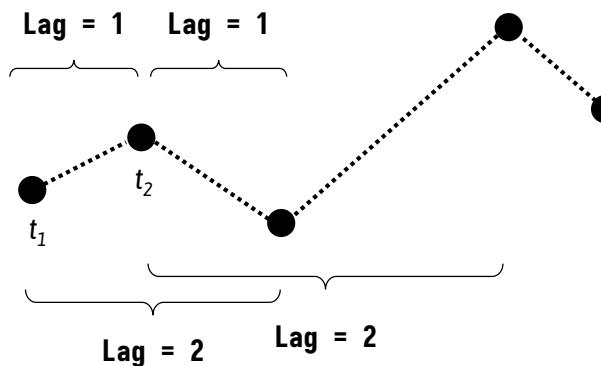


Confidence intervals increase because sample size decreases (*i.e.*, less data points).



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**.



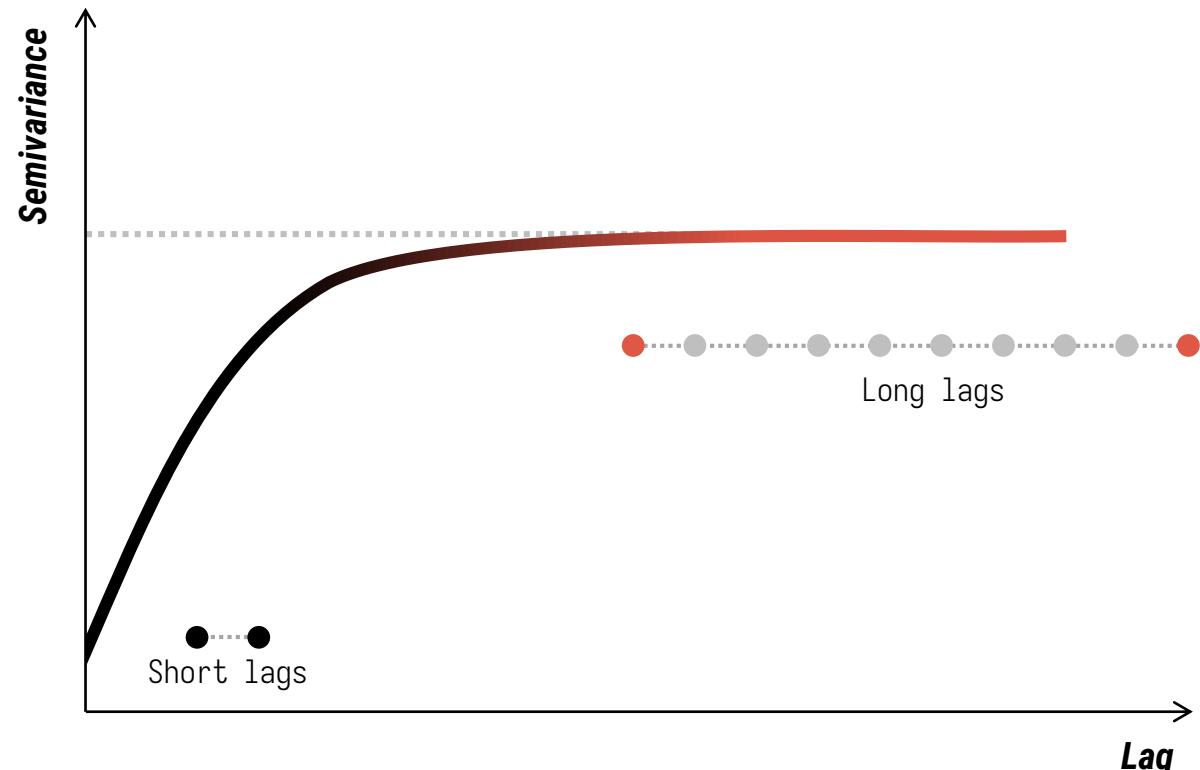
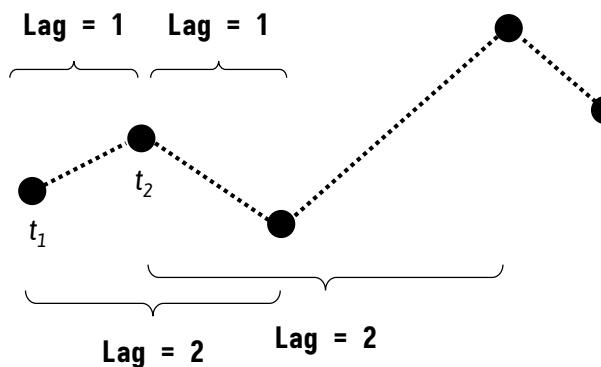
The more **similar** the pairs of locations are per lag, the **lower the semivariance** for that lag.



## Autocorrelation

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**.



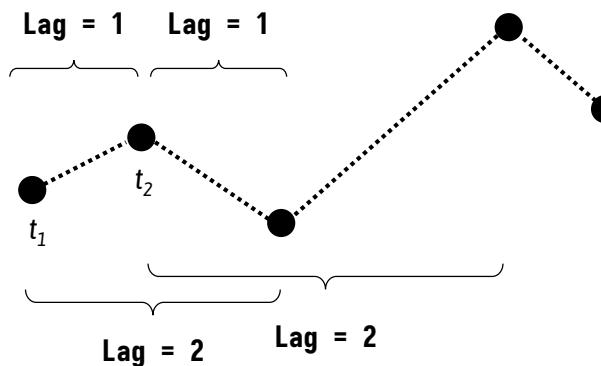
The more **similar** the pairs of locations are per lag, the **lower the semivariance** for that lag.



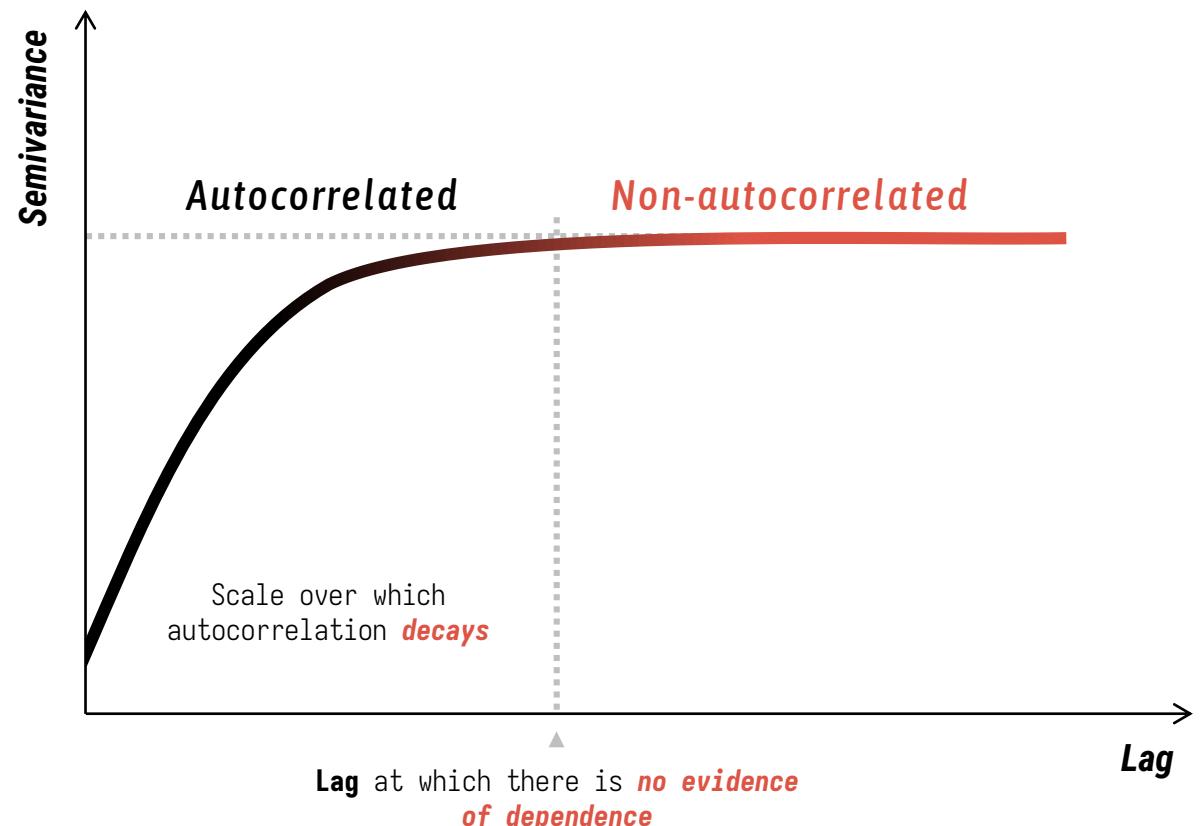
## Autocorrelation

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.

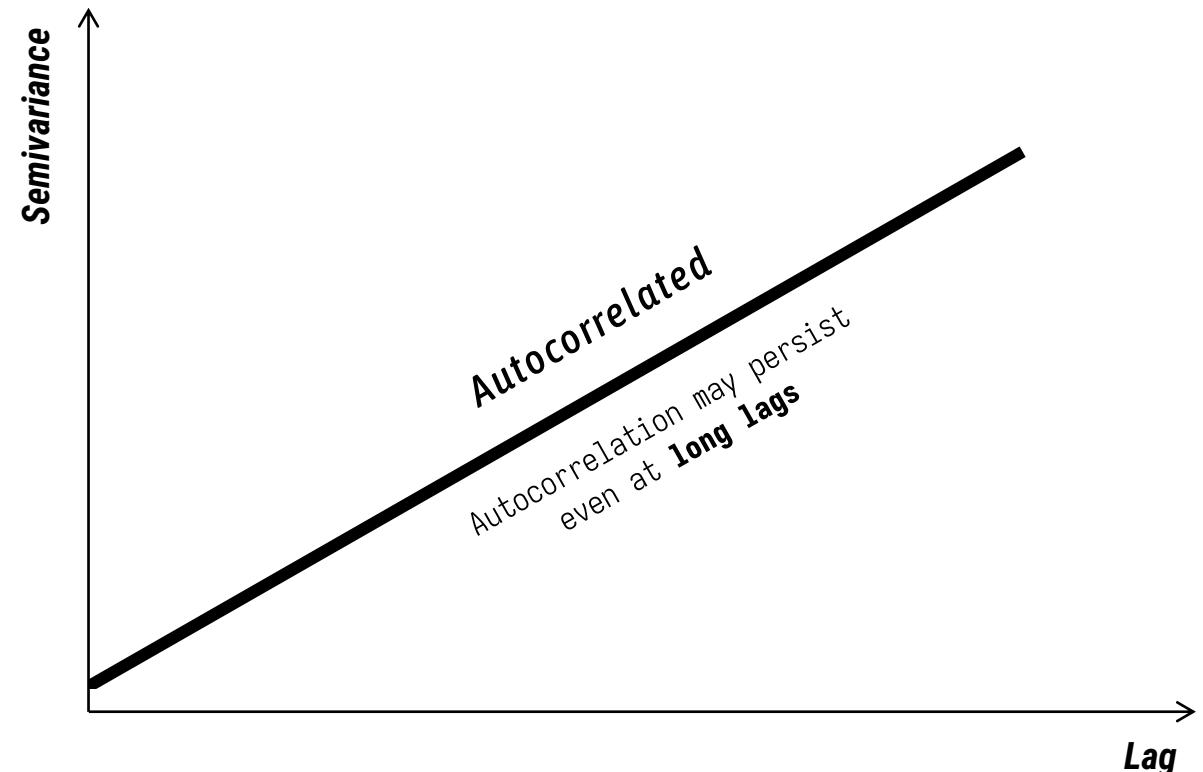
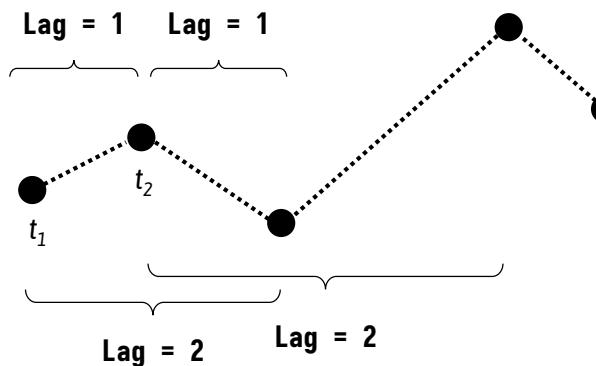




## Autocorrelation

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.



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

Standard error (SE):

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

Confidence intervals (CI):

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

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.  $\approx$  10% new info).



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.

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

*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).



- 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

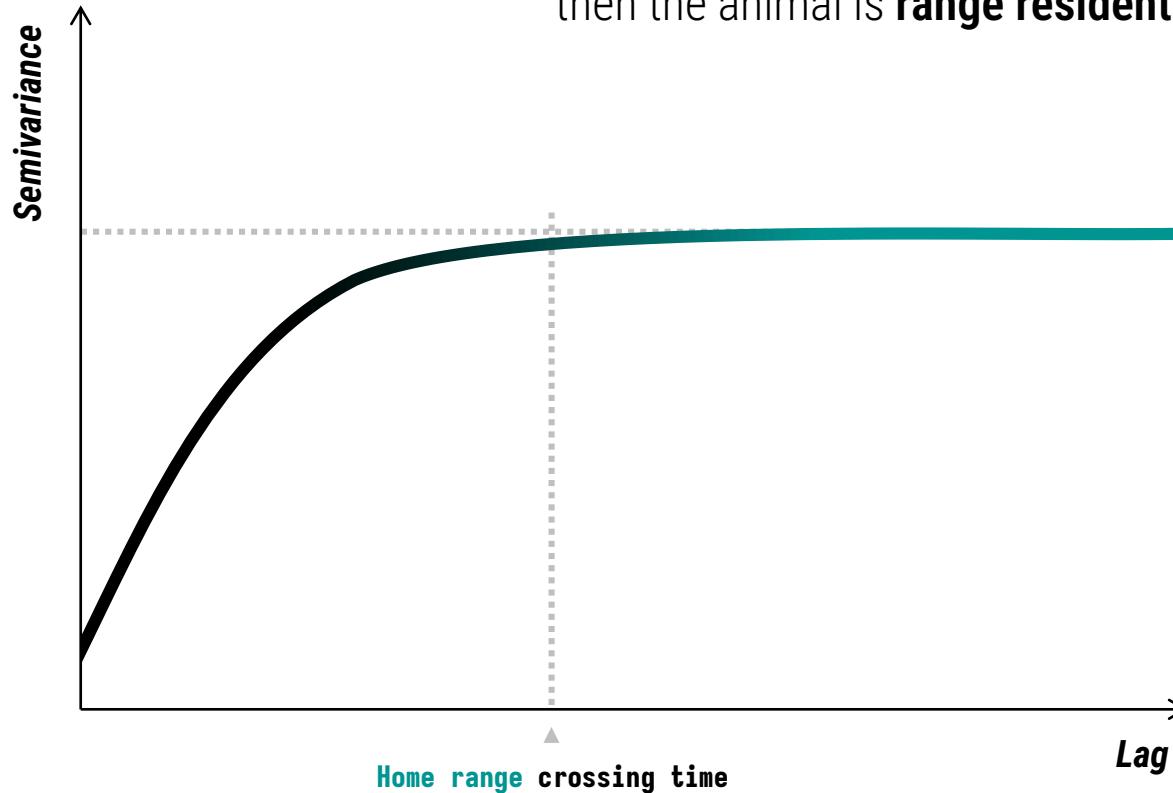


- 1. *Check range residency assumption;*  
Verify if the data is from a **range-resident** animal
  
- 2. *Select movement model;*  
Selecting the best-fit model through **model selection**
  
- 3. *Estimate home range area;*  
Reconstructing **range distribution** from sampled locations



## Range residency assumption

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



*Conventional methods:*



*Continuous-time methods:*



*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)



## – Independent and Identically Distributed (IID)

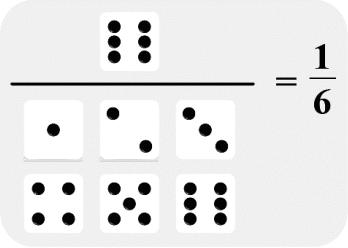
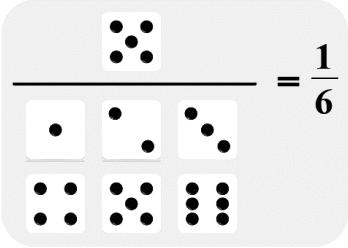
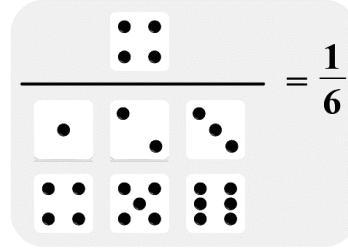
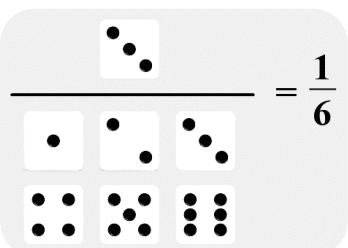
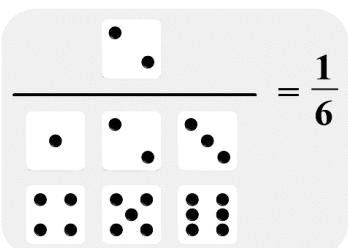
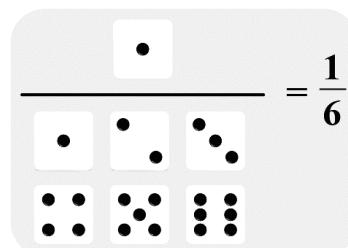
Stochastic process where each location has the same probability distribution as all others, and all are **mutually independent**.

SPATIAL DEPENDENCY

TEMPORAL DEPENDENCY

RESTRICTED

For example,



Dice rolls are independent and identically distributed (IID).

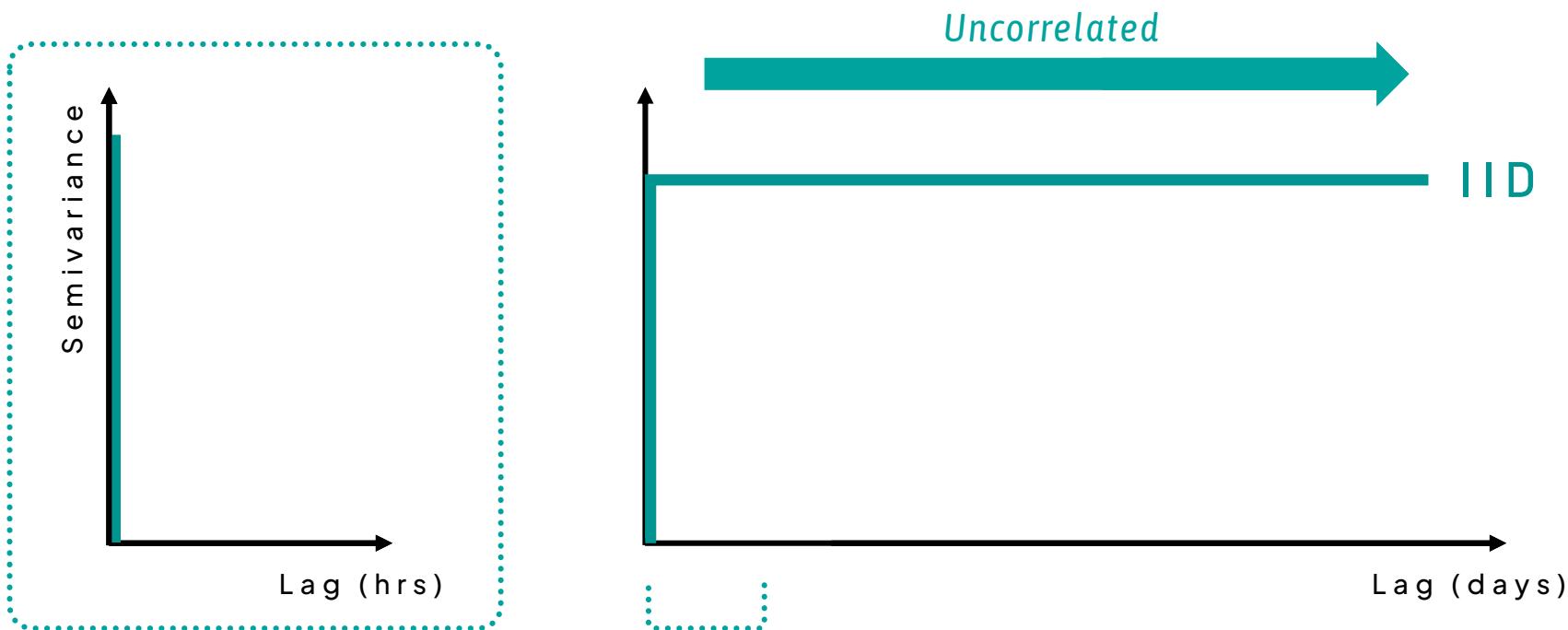


### – Independent and Identically Distributed (IID)

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

SPATIAL DEPENDENCY  
TEMPORAL DEPENDENCY  
RESTRICTED

How would the **variogram** of a IID process look like?

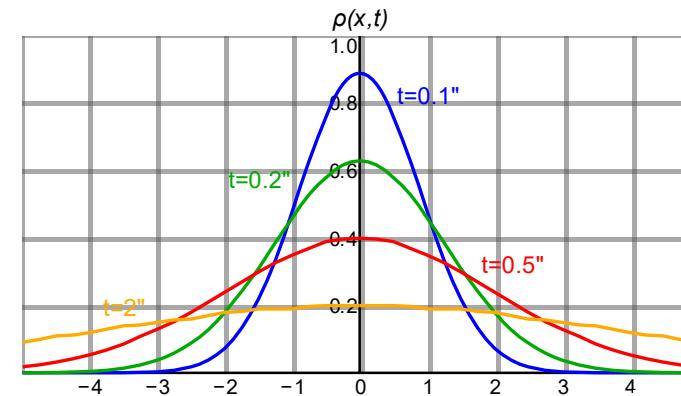
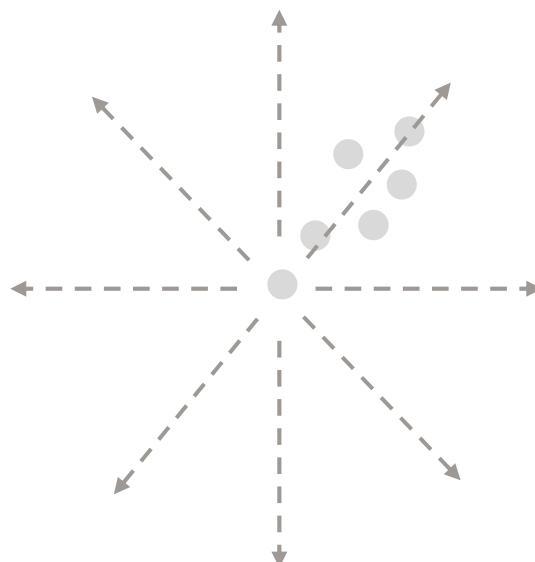




## — Brownian Motion (BM)

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**.

SPATIAL DEPENDENCY  
TEMPORAL DEPENDENCY  
RESTRICTED



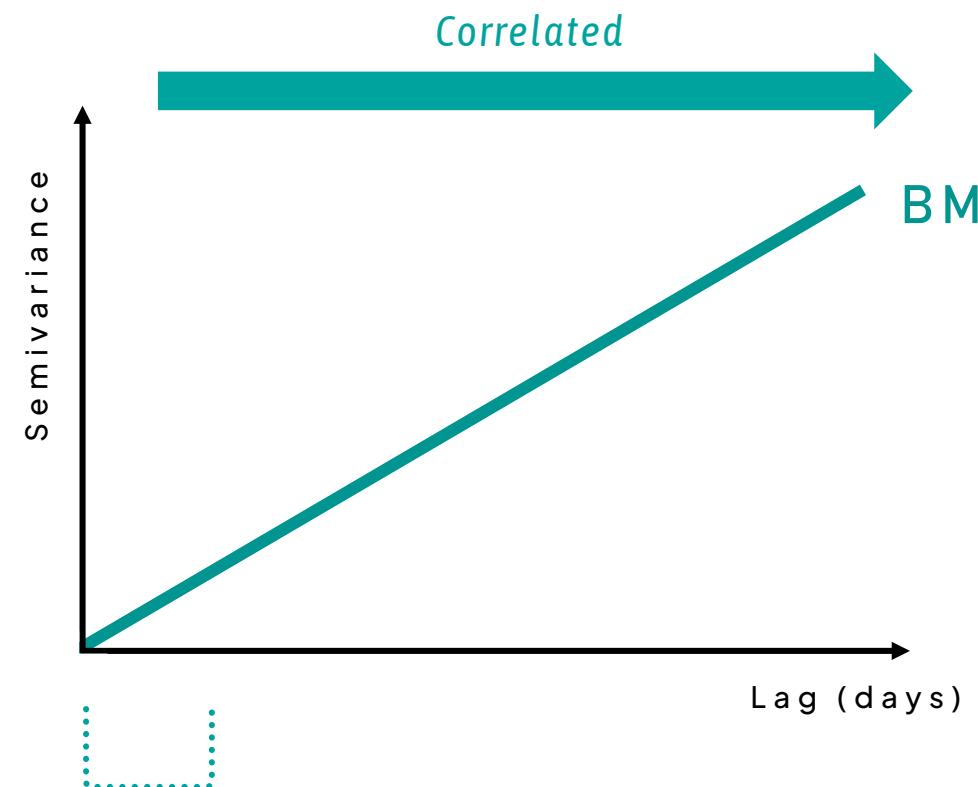
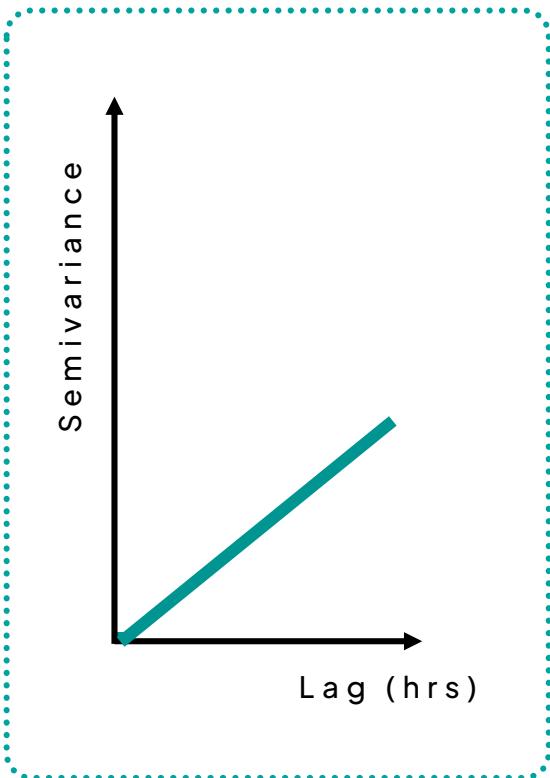
As time goes on, the animal is more likely to be further away from its starting location.



## – Brownian Motion (BM)

How would the **variogram** of a BM process look like?

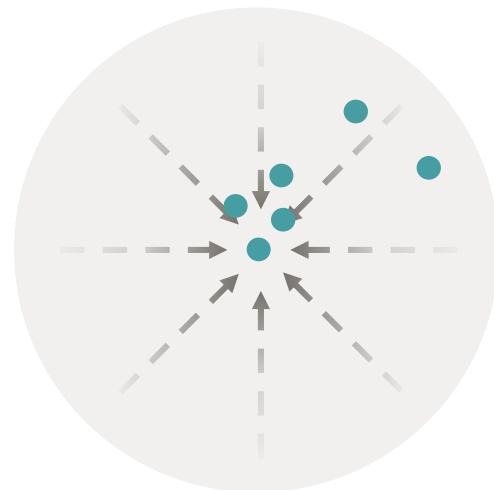
|                     |
|---------------------|
| SPATIAL DEPENDENCY  |
| TEMPORAL DEPENDENCY |
| RESTRICTED          |





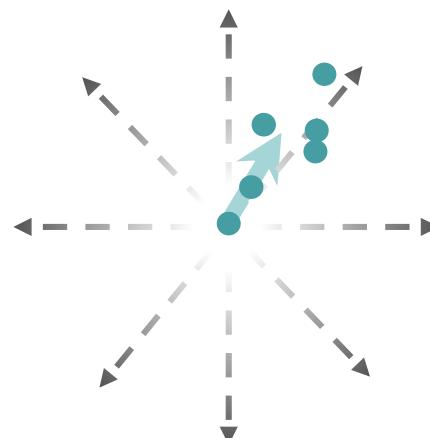
These processes are all modifications of **BM** process:

**Ornstein-Uhlenbeck (OU)**



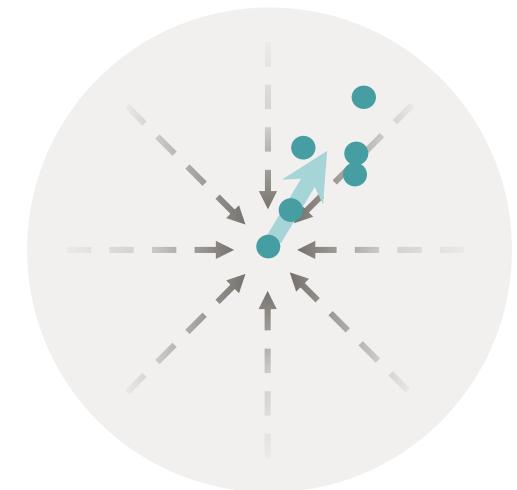
Unlike **BM**, the **OU** process tends towards a central location (**bounded diffusion**).

**Integrated OU (IOU)**



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

**OU with Foraging (OUF)**

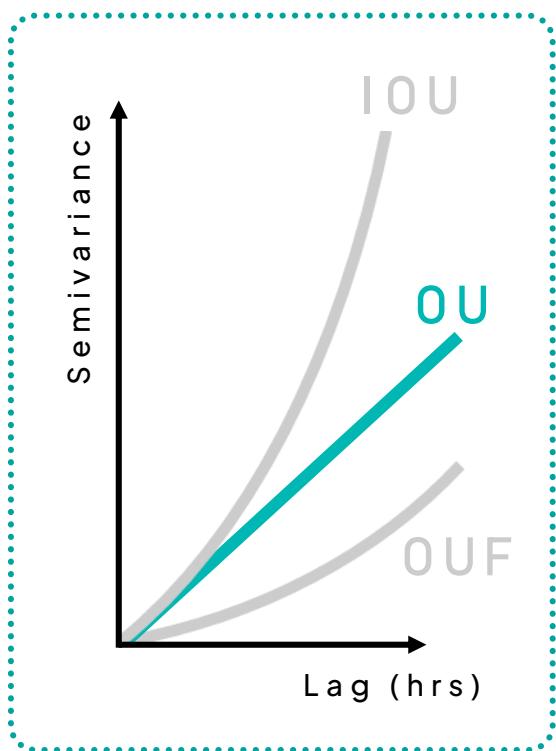


Unlike **BM**, the **OUF** process exhibits **bounded diffusion** and **persistence of motion**.

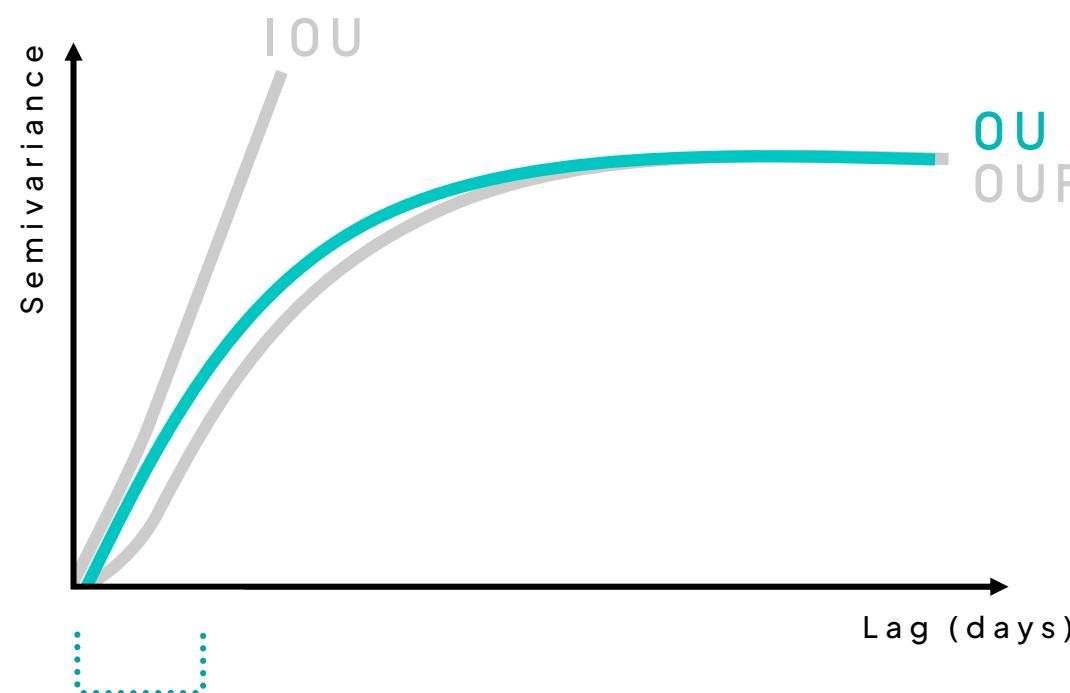


These processes are all modifications of **BM** process:

Ornstein-Uhlenbeck (OU)



Integrated OU (IOU)



OU with Foraging (OUF)

- SPATIAL DEPENDENCY
- TEMPORAL DEPENDENCY
- RESTRICTED



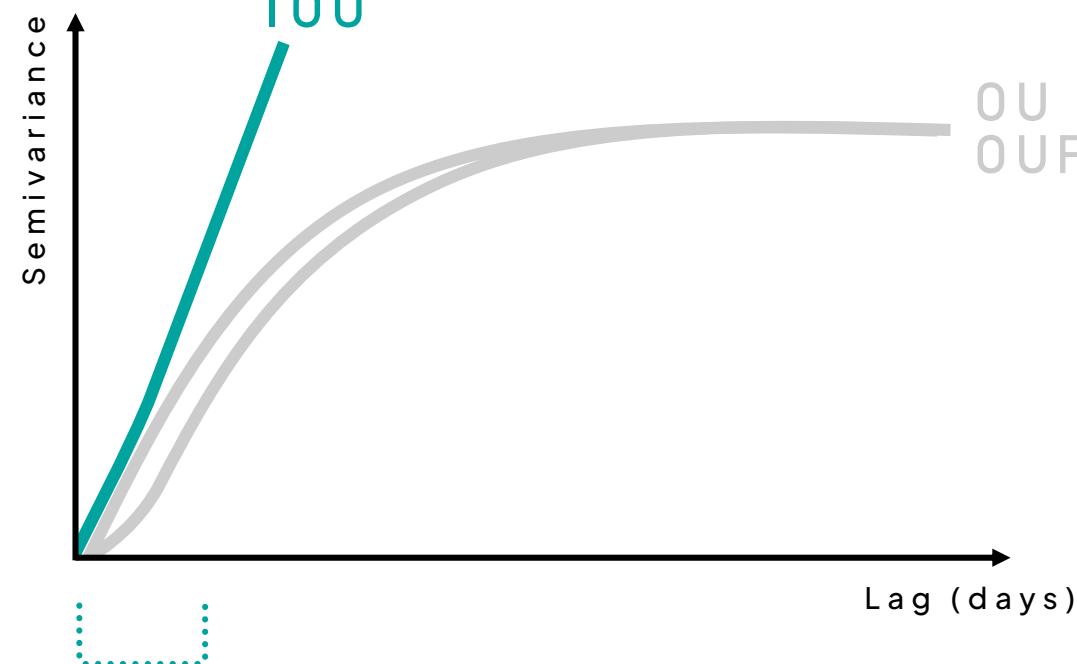
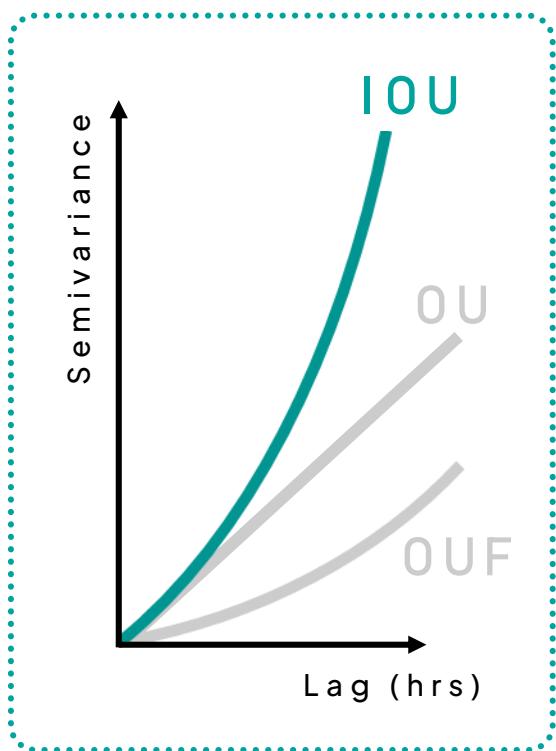
These processes are all modifications of **BM** process:

### Ornstein-Uhlenbeck (OU)

### Integrated OU (IOU)

### OU with Foraging (OUF)

How would the **variogram** of these OU processes look like?

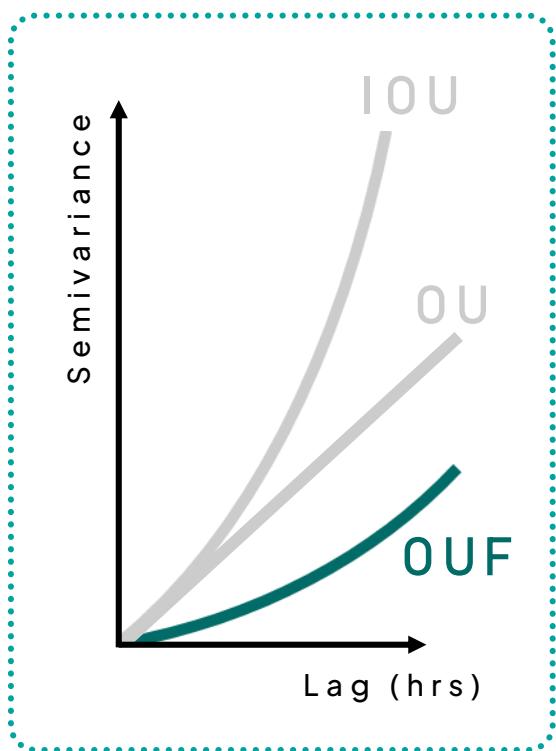


- SPATIAL DEPENDENCY
- TEMPORAL DEPENDENCY
- RESTRICTED

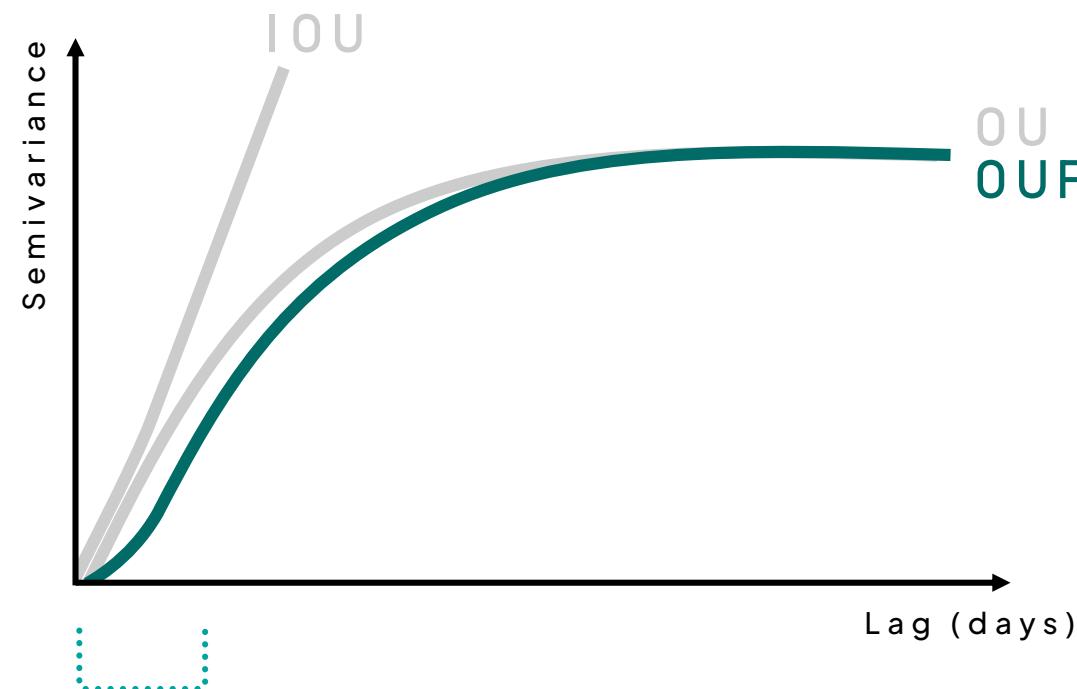


These processes are all modifications of **BM** process:

Ornstein-Uhlenbeck (OU)



Integrated OU (IOU)



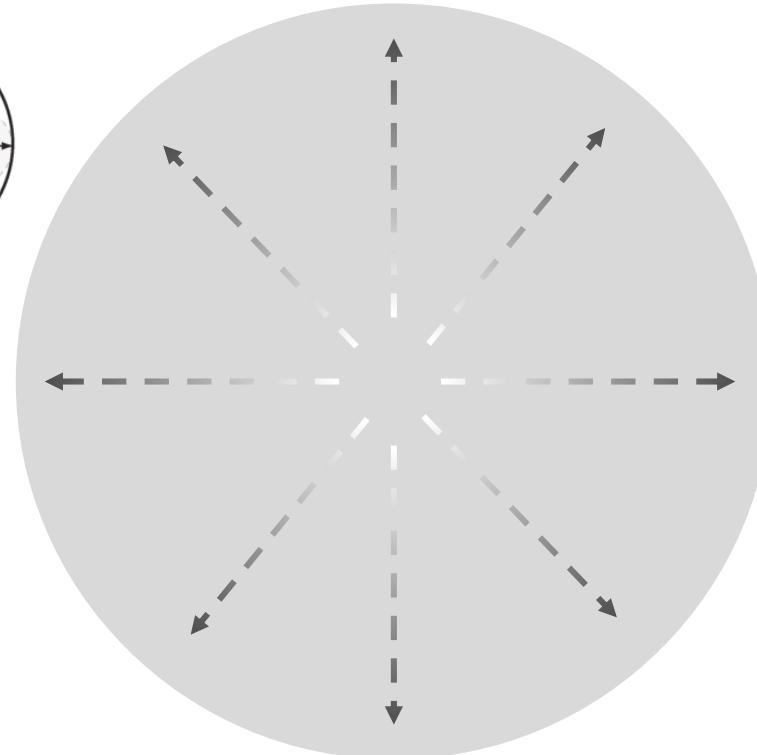
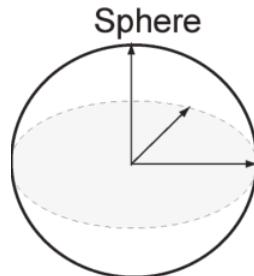
OU with Foraging (OUF)

How would the **variogram** of these OU processes look like?

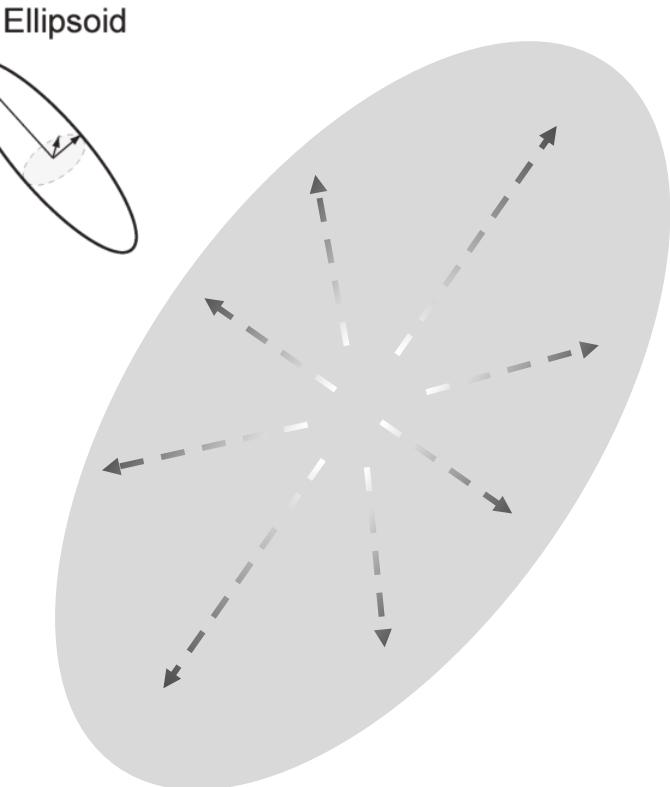
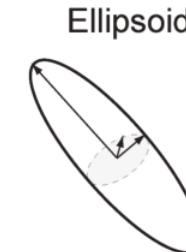
- SPATIAL DEPENDENCY
- TEMPORAL DEPENDENCY
- RESTRICTED



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



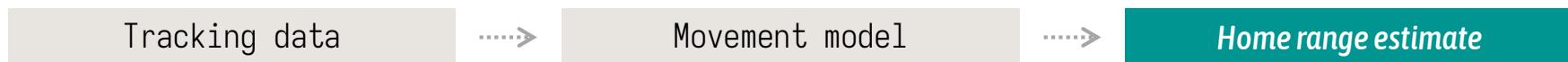
**Isotropic diffusion**



**Anisotropic diffusion**



- 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



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

| Model | Position | Velocity | Restricted | Parameters                  |
|-------|----------|----------|------------|-----------------------------|
| IID   | No       | No       | Yes        | $\tau = \text{NULL}$        |
| 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 reduces to the **(conventional) KDE** in the limit where autocorrelation vanishes, and locations are truly *independent*.

“

All models are wrong, but **some are useful.**

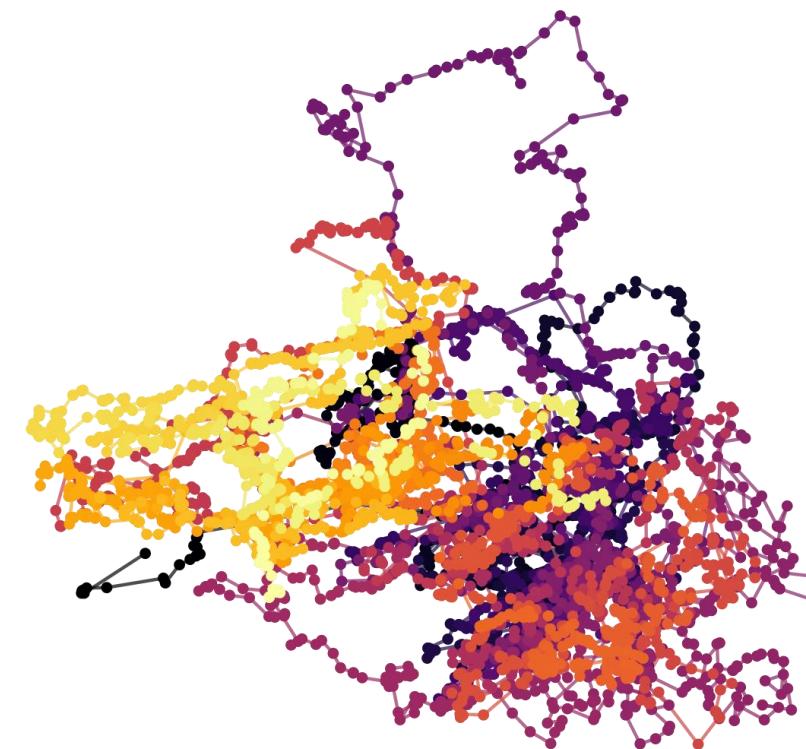
Box et al. (1987)



Estimate home range



African Buffalo  
(*Syncerus caffer*)



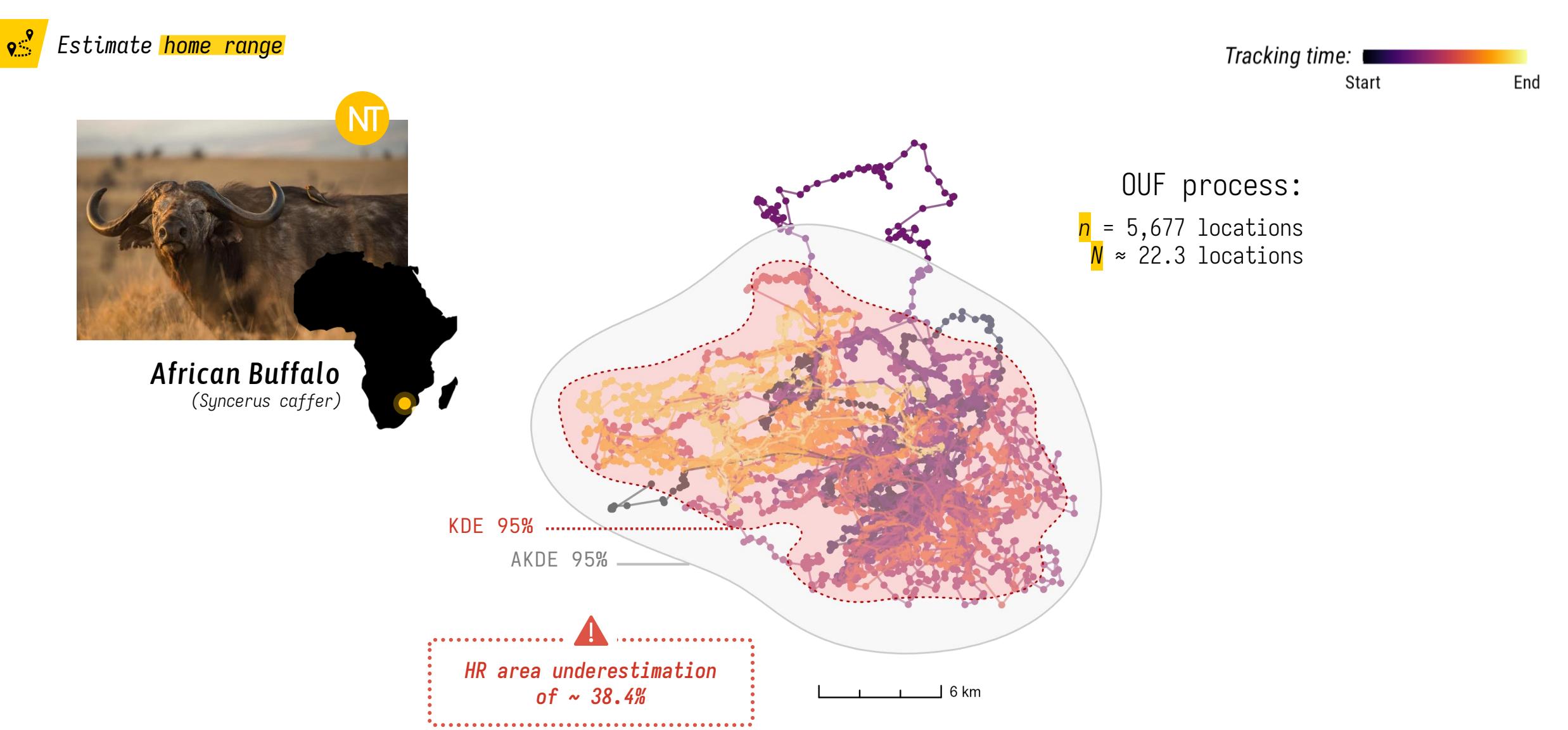
6 km

Tracking time: Start End

OUF process:

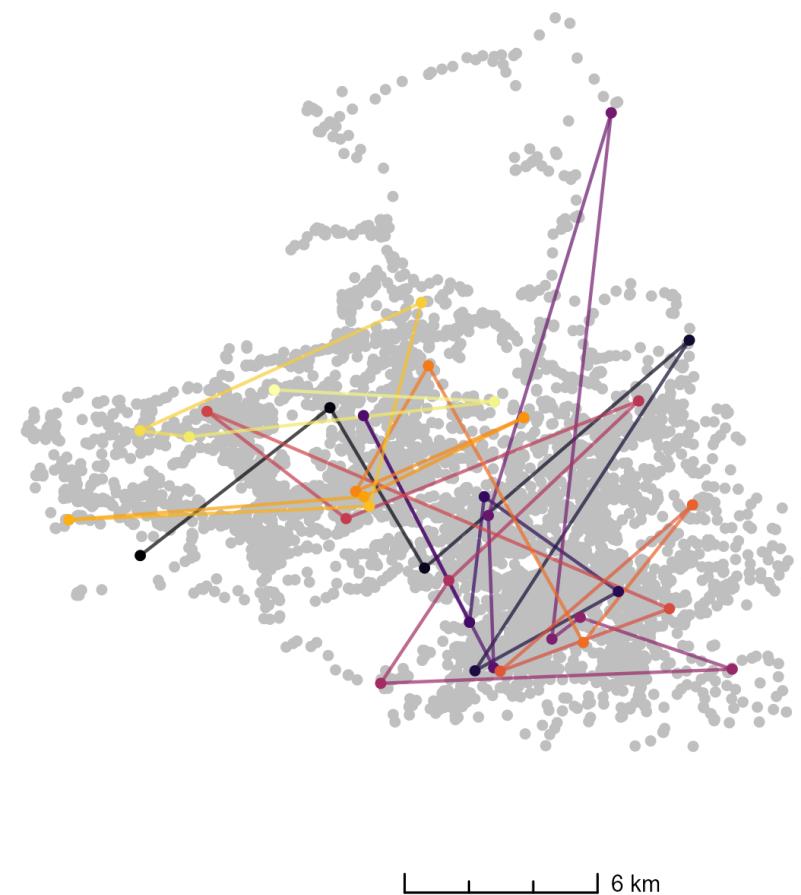
$n = 5,677$  locations

$N \approx 22.3$  locations





Estimate home range



Tracking time: Start End

OUF process:

$n$  = 5,677 locations  
 $N \approx 22.3$  locations

IID process:

$n$  = 35 locations  
 $N$  = 35 locations

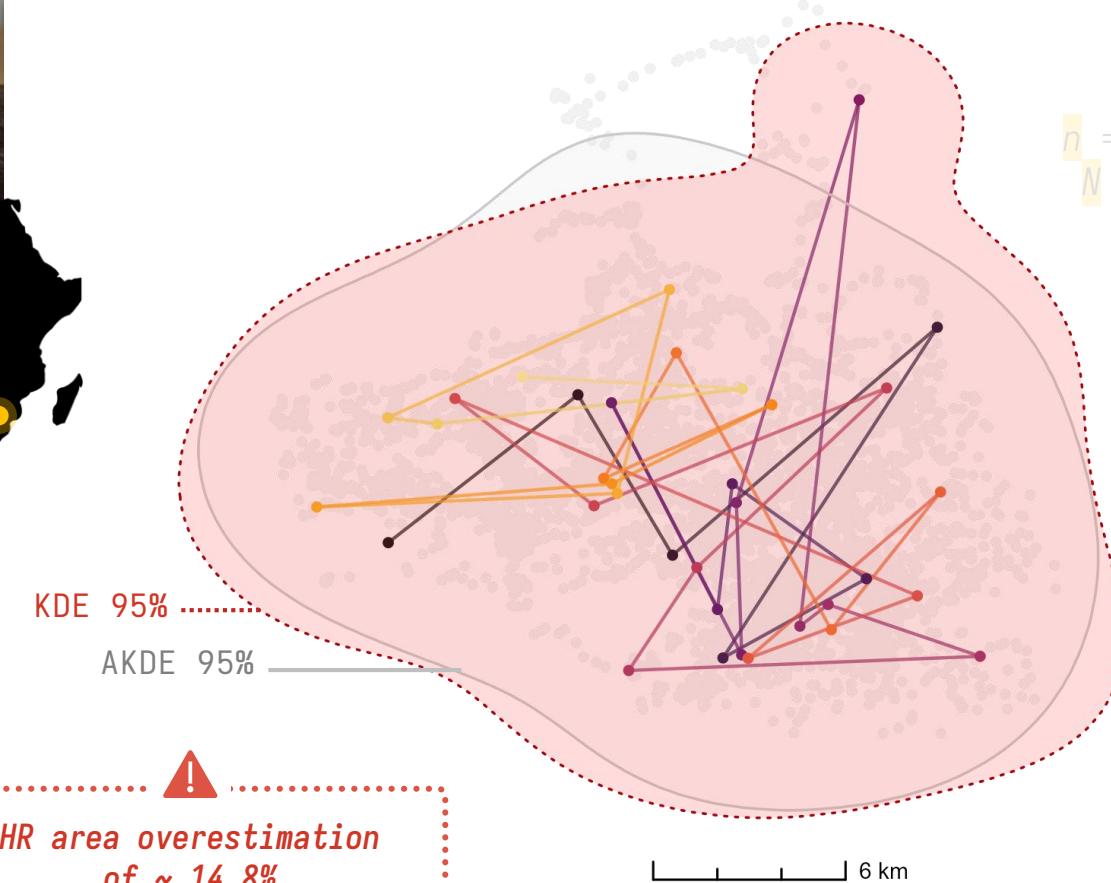
Data loss of  $\approx 99.4\%$



Estimate home range



**African Buffalo**  
(*Syncerus caffer*)



KDE 95%  
AKDE 95%

HR area overestimation  
of ~ 14.8%

6 km

Tracking time: Start End

OUF process:

$n = 5,677$  locations  
 $N \approx 22.3$  locations

IID process:

$n = 35$  locations  
 $N = 35$  locations

Data loss of ≈ 99.4%



- 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



- 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



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 <sub>c</sub> (default)                  |
| Autocorrelation estimation bias   | ► pHREML (default)<br>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 <sub>c</sub> (default)                  |
| Autocorrelation estimation bias   | ▶ pHREML (default)<br>Parametric bootstrapping |
| Unrepresentative sampling in time | ▶ Weighted AKDE, or wAKDE                      |



Tracking data



Movement model



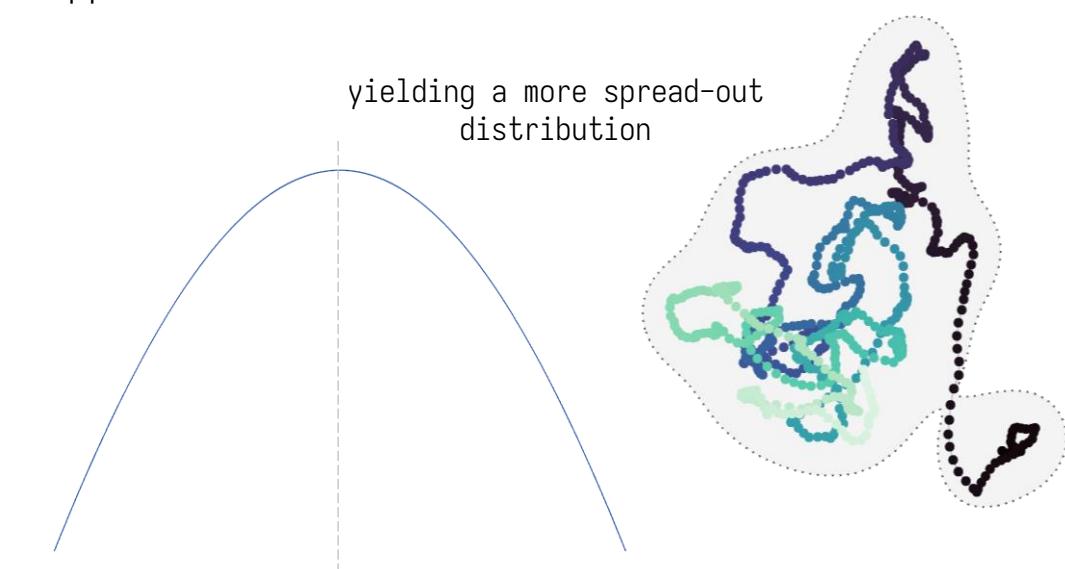
**Home range estimate**

## Area-corrected AKDE or **AKDE<sub>C</sub>**

Deals with: **oversmoothing**



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





Tracking data



Movement model



**Home range estimate**

## Area-corrected AKDE or **AKDE<sub>C</sub>**

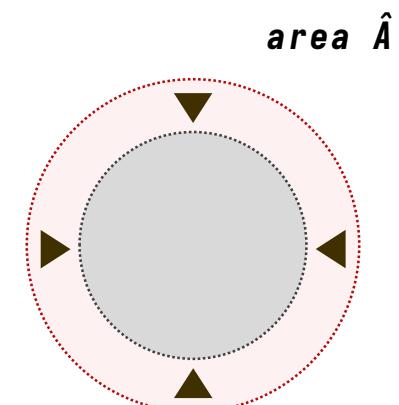
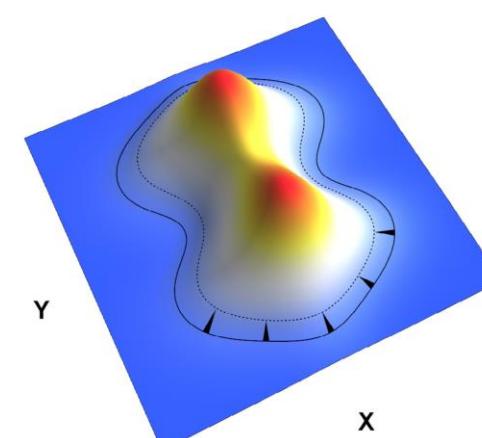
Deals with: **oversmoothing**



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

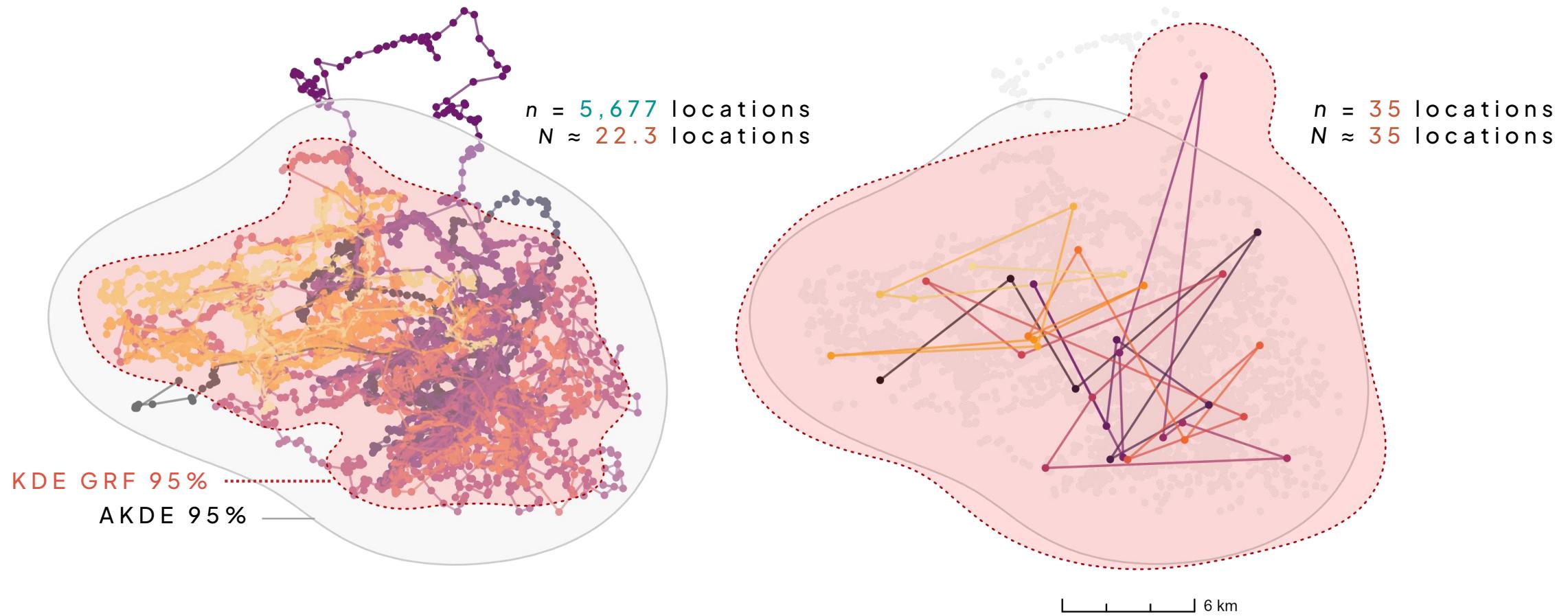
Derived an **improved (A)KDE** that pulls the contours of the location distribution estimate inward towards the data without distorting its shape.

Fleming & Calabrese (2017)





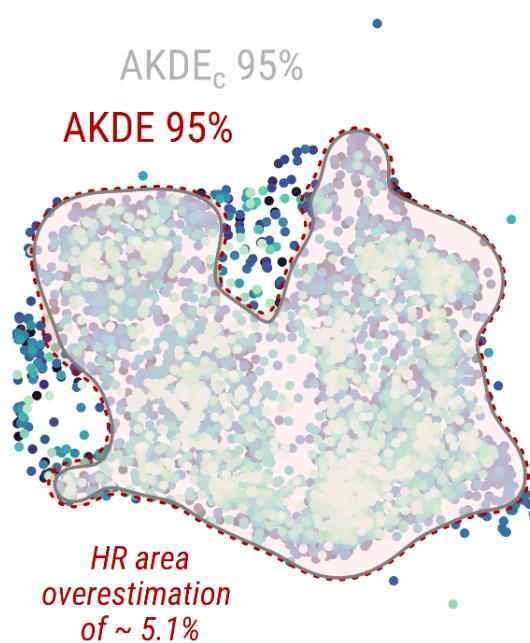
The oversmoothing (positive) bias can be **masked** by the often-stronger negative bias caused by unmodeled autocorrelation.





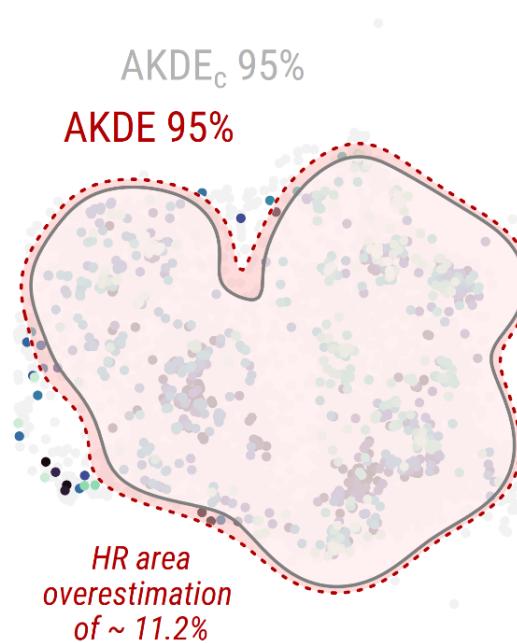
### Large effective sample size

Sampling duration  $\approx$  1 year



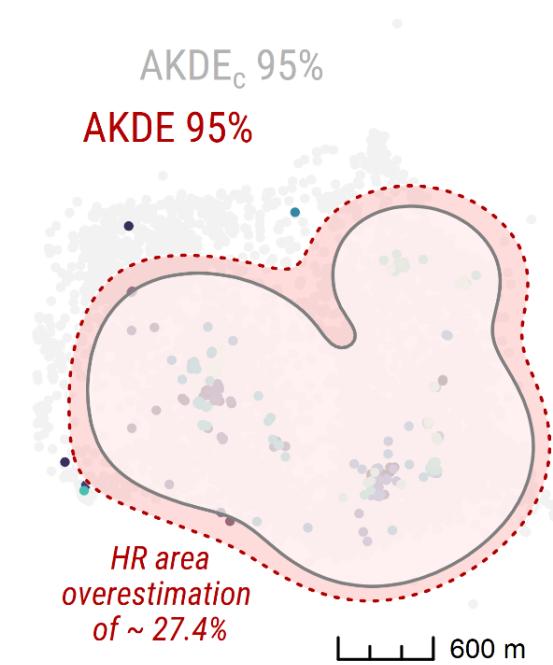
### Medium effective sample size

Sampling duration  $\approx$  3 months



### Small effective sample size

Sampling duration  $\approx$  15 days





Tracking data



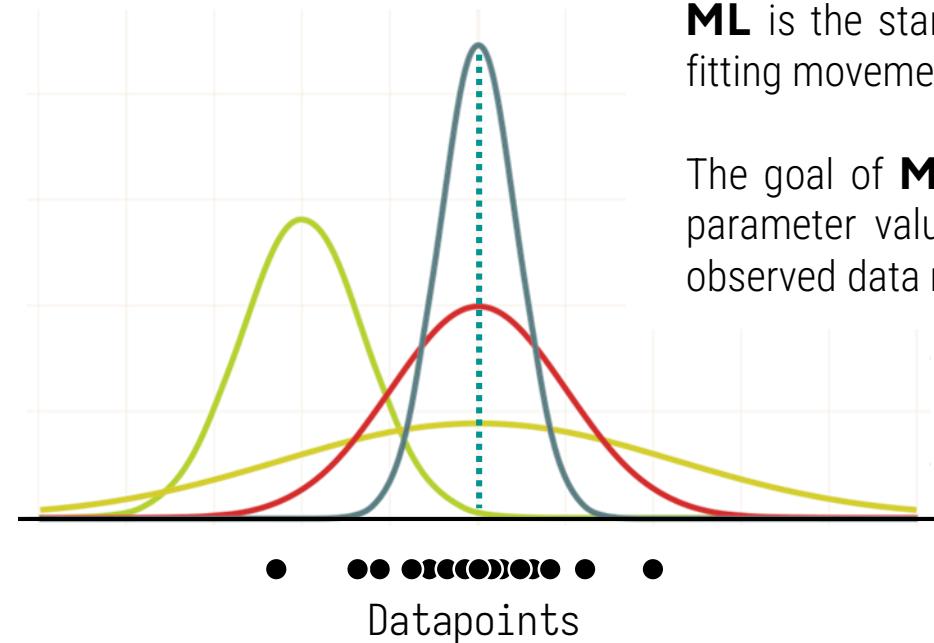
**Movement model**



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

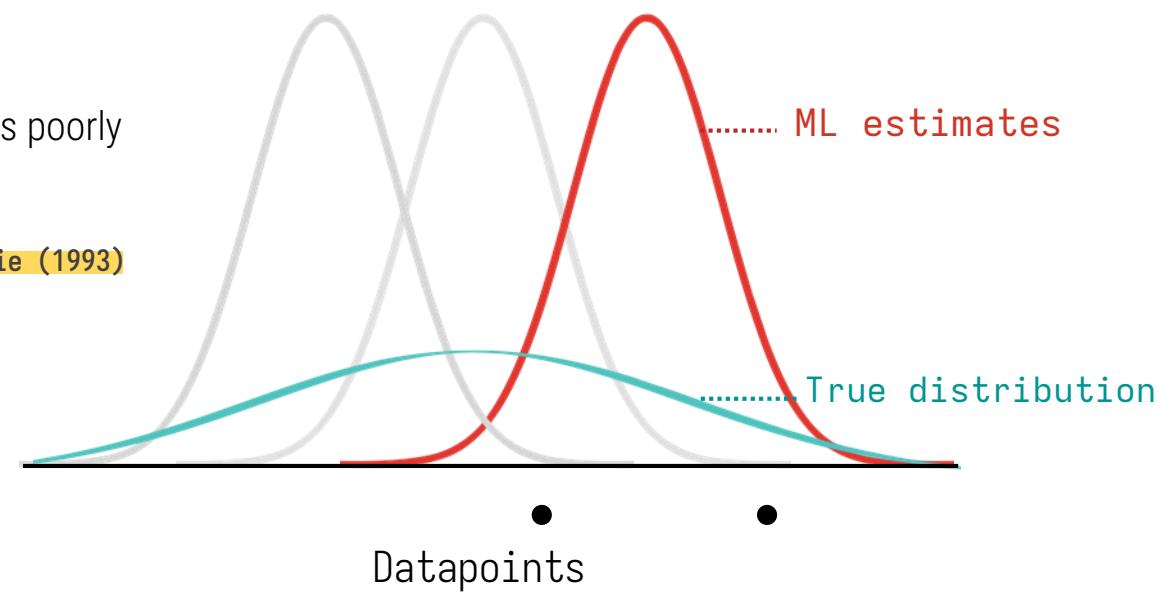
What **movement model parameters** are most likely to characterize a given tracking dataset?

— Maximum Likelihood (**ML**):

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



Cressie (1993)



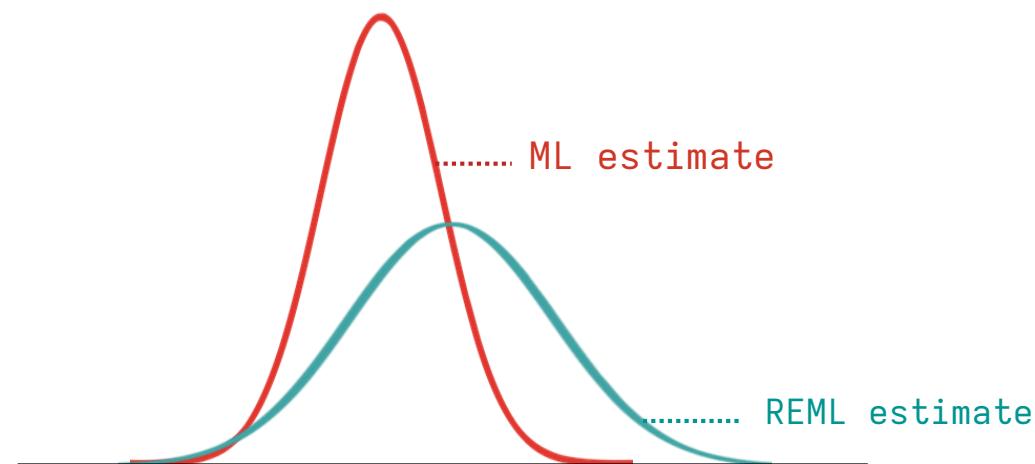


We can tackle the **ML** bias in several ways.  
For example:

### 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



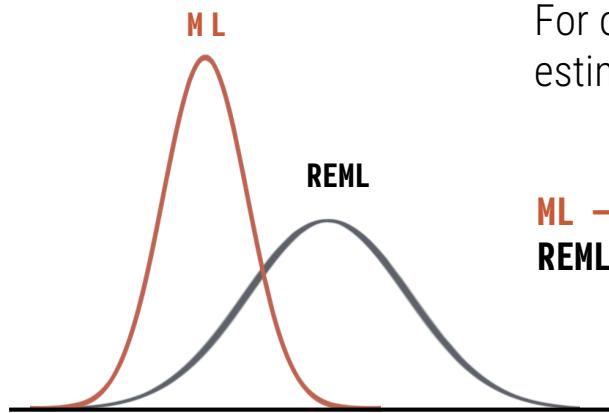
**Home range estimate**

## 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

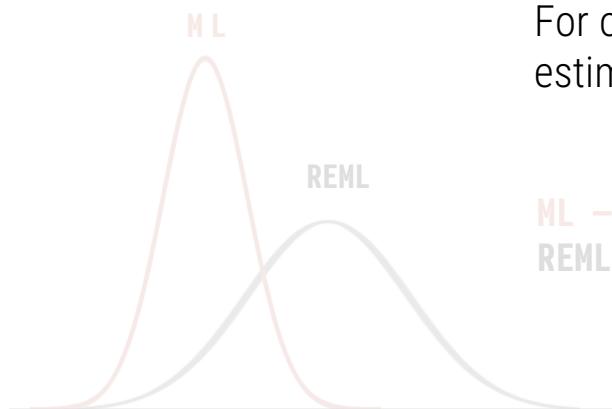
Home range estimate

## 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  
small **absolute** sample sizes  
small **absolute** and **effective** sample sizes

As such, we consider other parameter estimation methods:

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

Fleming et al. (2019)



Tracking data



Movement model



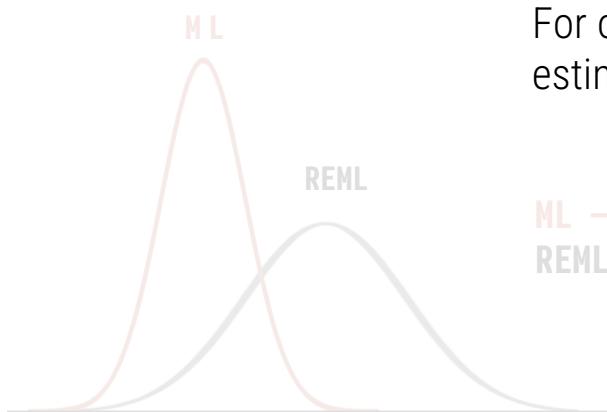
**Home range estimate**

## pHREML AKDE

Deals with: **autocorrelation estimation bias**



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



**Focus on:**  
small **effective** sample sizes  
small **absolute** sample sizes  
small **absolute** and **effective** sample sizes

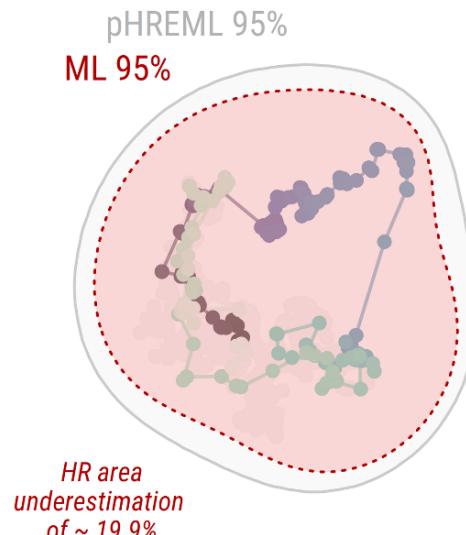
As such, we consider other parameter estimation methods:

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

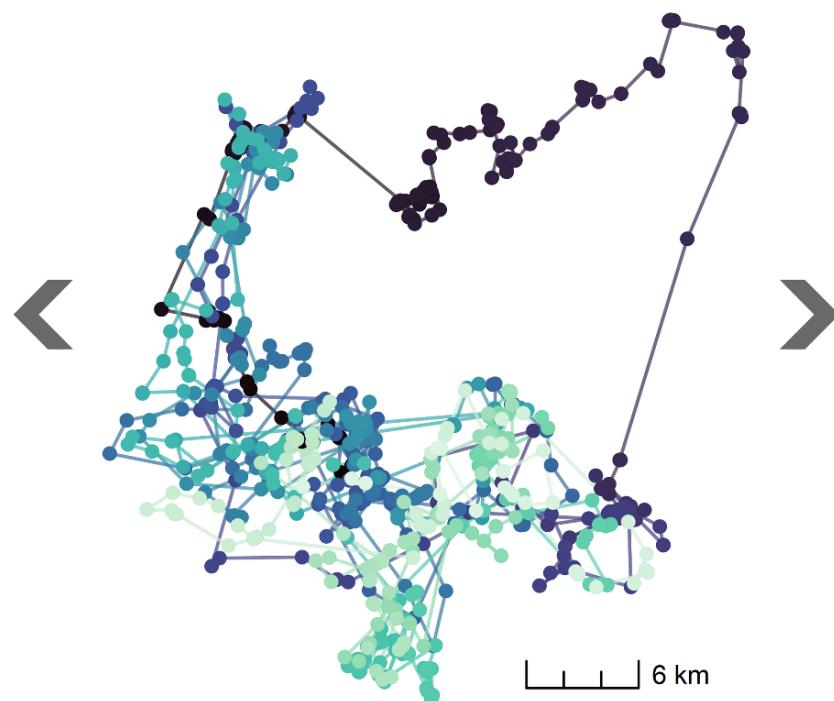
↳ Fleming et al. (2019)



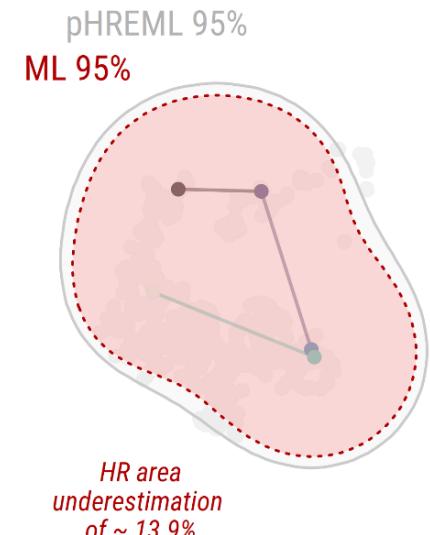
**Large absolute &  
Small effective sample size**  
 $n = 363$     $N \approx 3.1$



$n = 1010$     $N \approx 14.5$

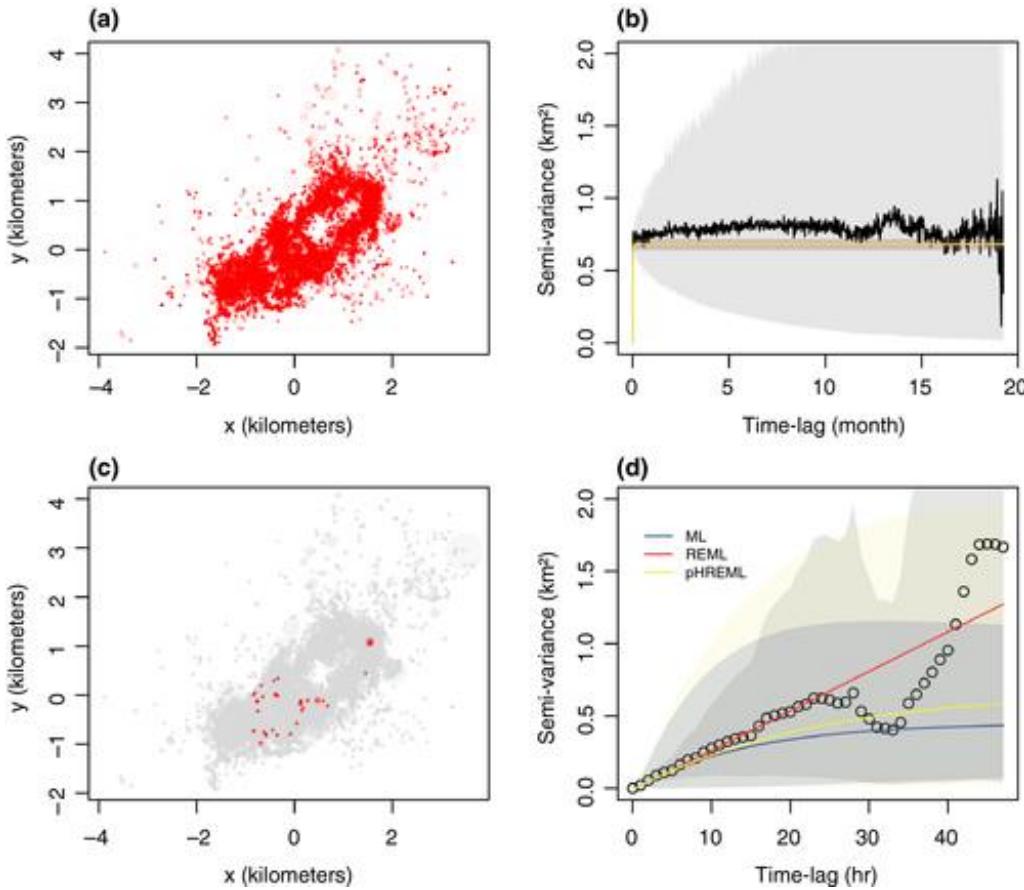


**Small absolute &  
Small effective sample size**  
 $n = 5$     $N \approx 4$





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



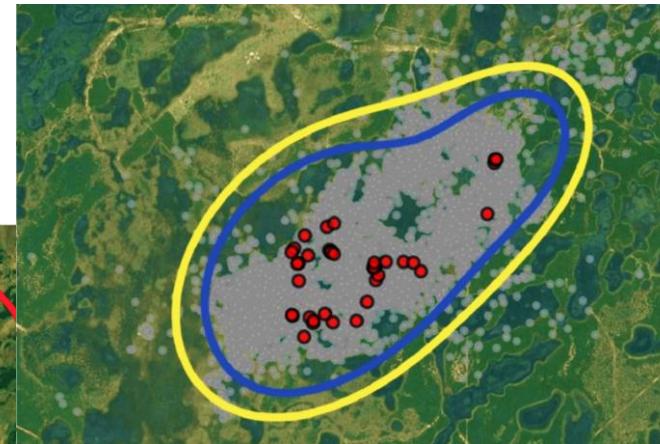
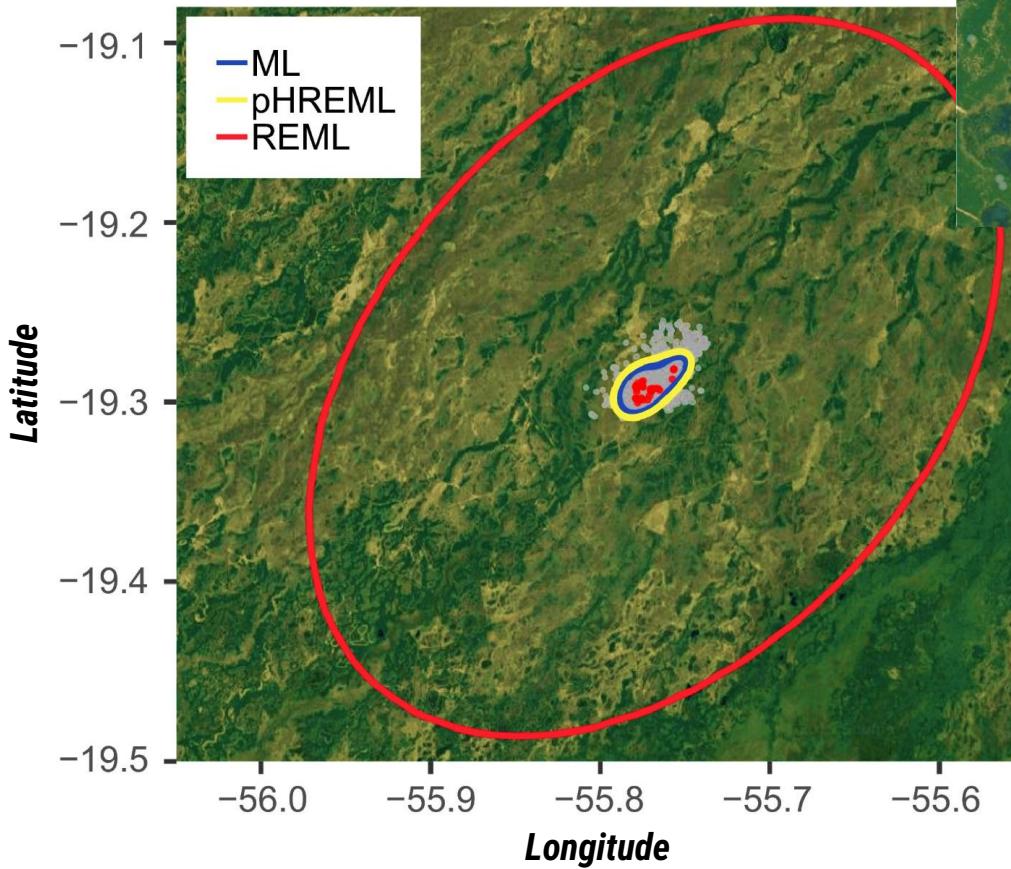
**Lowland tapir**  
(*Tapirus terrestris*)





## Dealing with *biases*

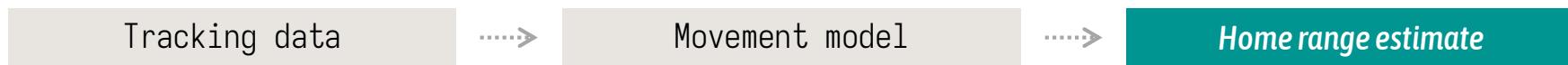
Tracking data (1-hr intervals for **19 months**), reduced to **2 days**.



ML  
pHREML  
REML

**Lowland tapir**  
(*Tapirus terrestris*)





## Weighted AKDE or wAKDE

Deals with: *unrepresentative sampling in time*

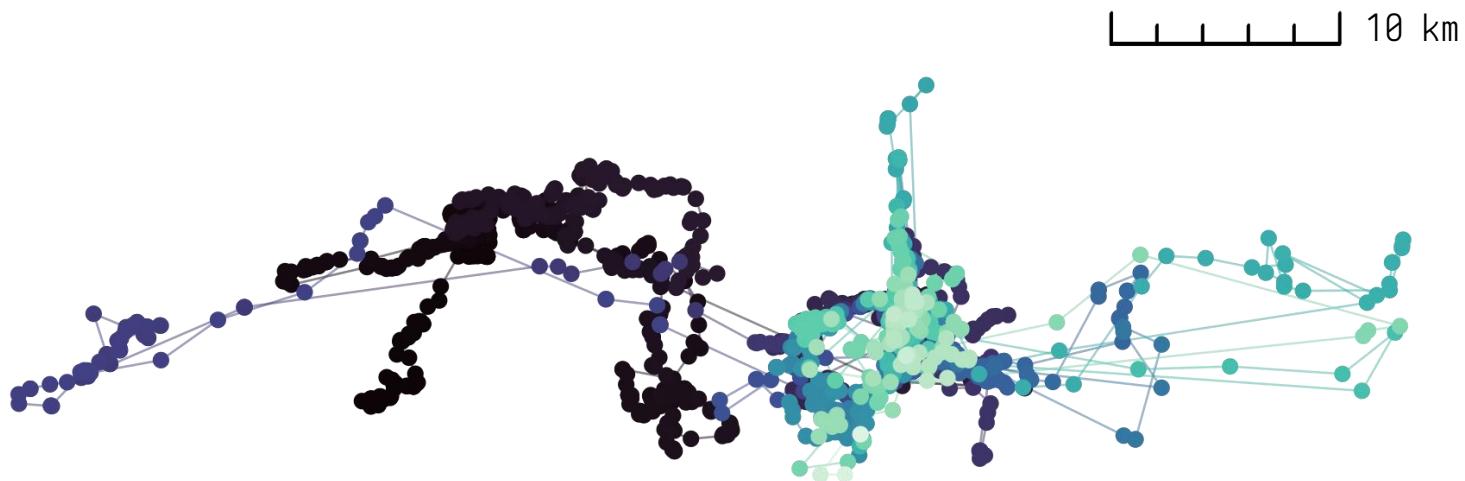


For optimal performance, we need to estimate autocorrelation **correctly**.

- ▶ 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.



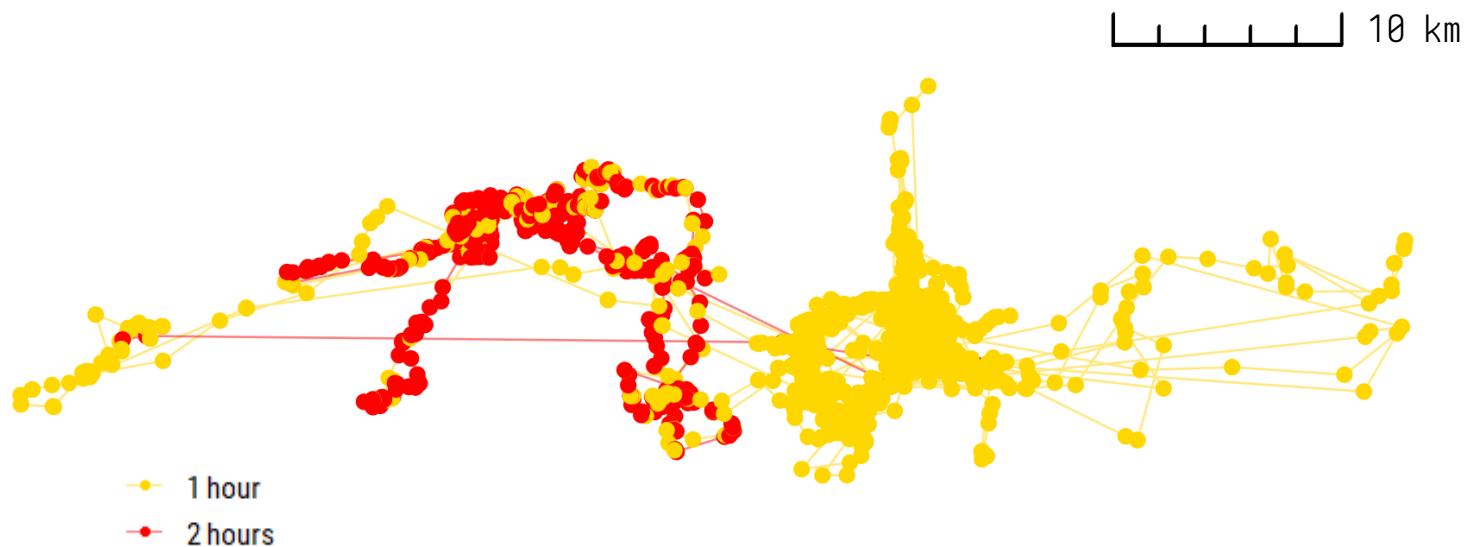
**wAKDE** optimally *upweights* observations that occur during *under-sampled times*, while optimally *downweighting* observations occurring during *over-sampled times*.



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



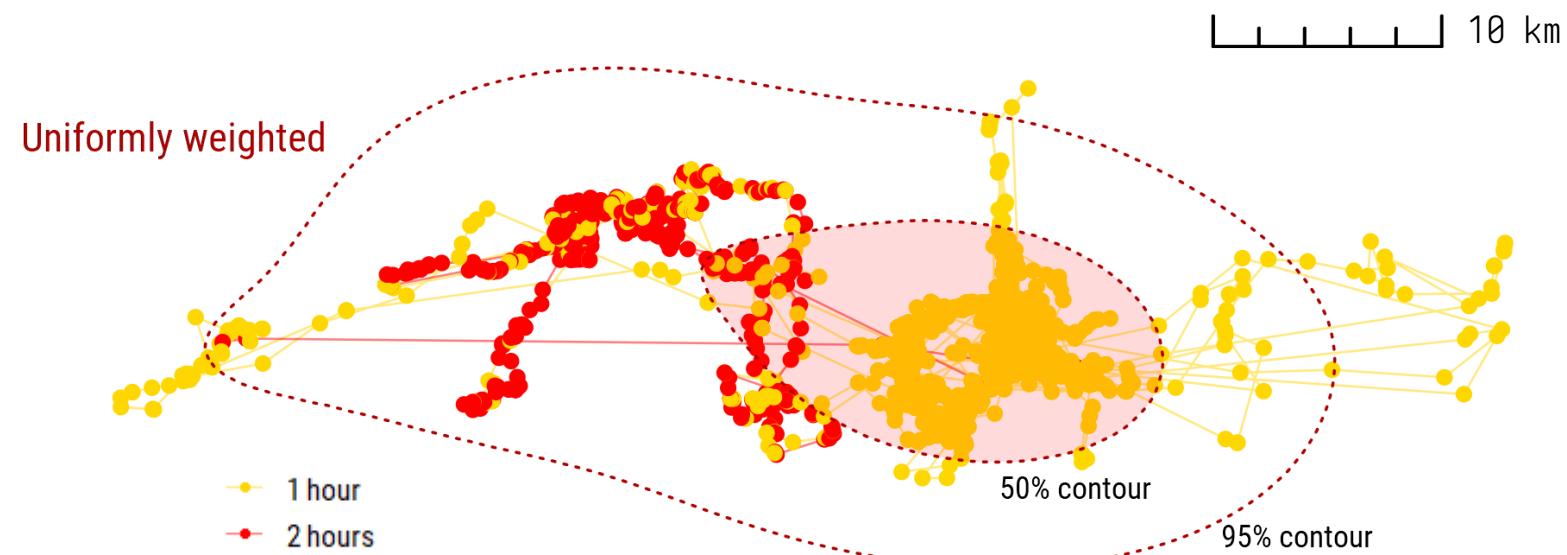
**wAKDE** optimally *upweights* observations that occur during *under-sampled times*, while optimally *downweighting* observations occurring during *over-sampled times*.



**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**.



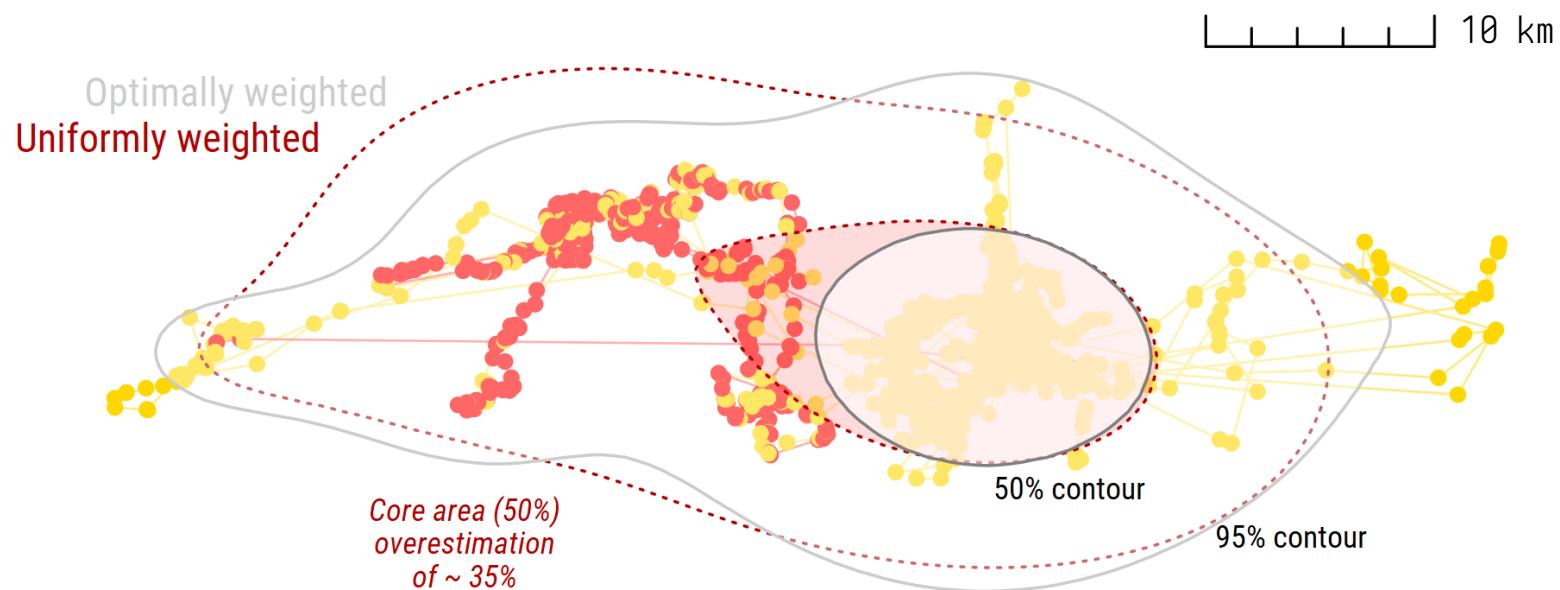
wAKDE optimally *upweights* observations that occur during *under-sampled times*, while optimally *downweighting* observations occurring during *over-sampled times*.



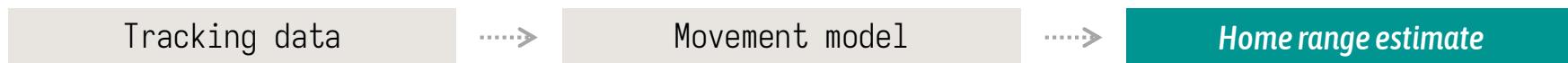
**Fig.** African buffalo dataset (nicknamed “Pepper”) with an *irregular* sampling schedule.  
Sampling rate shifted from **1 fix every hour** to **1 fix every 2 hours**.



**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**.



## Parametric bootstrapping AKDE

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

↳ **Bartlett (1937)**

**Residual ML (or REML)**

↳ **Fleming et al. (2019)**

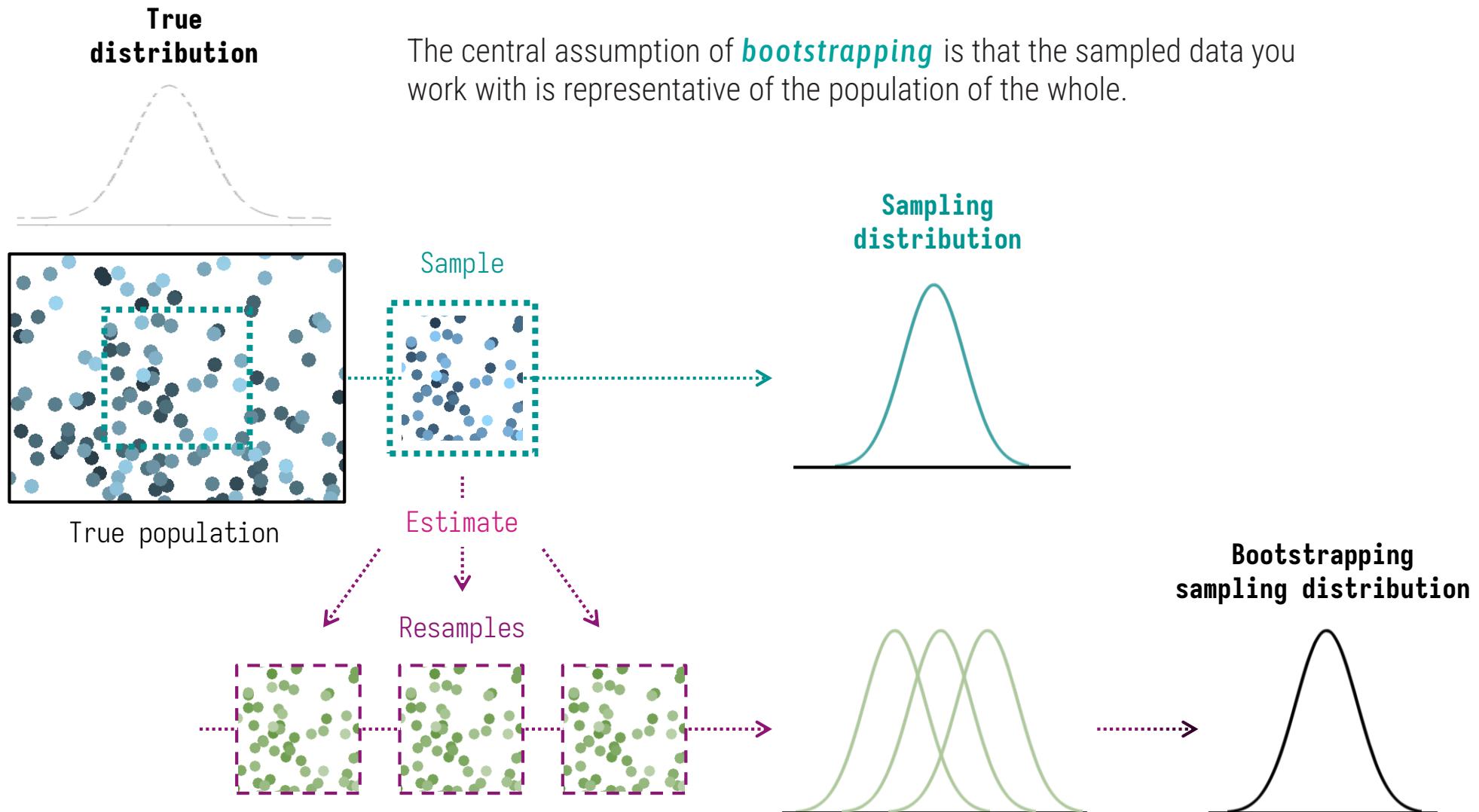
**perturbative Hybrid REML (pHREML)**



↳ **Efron & Efron (1982)**

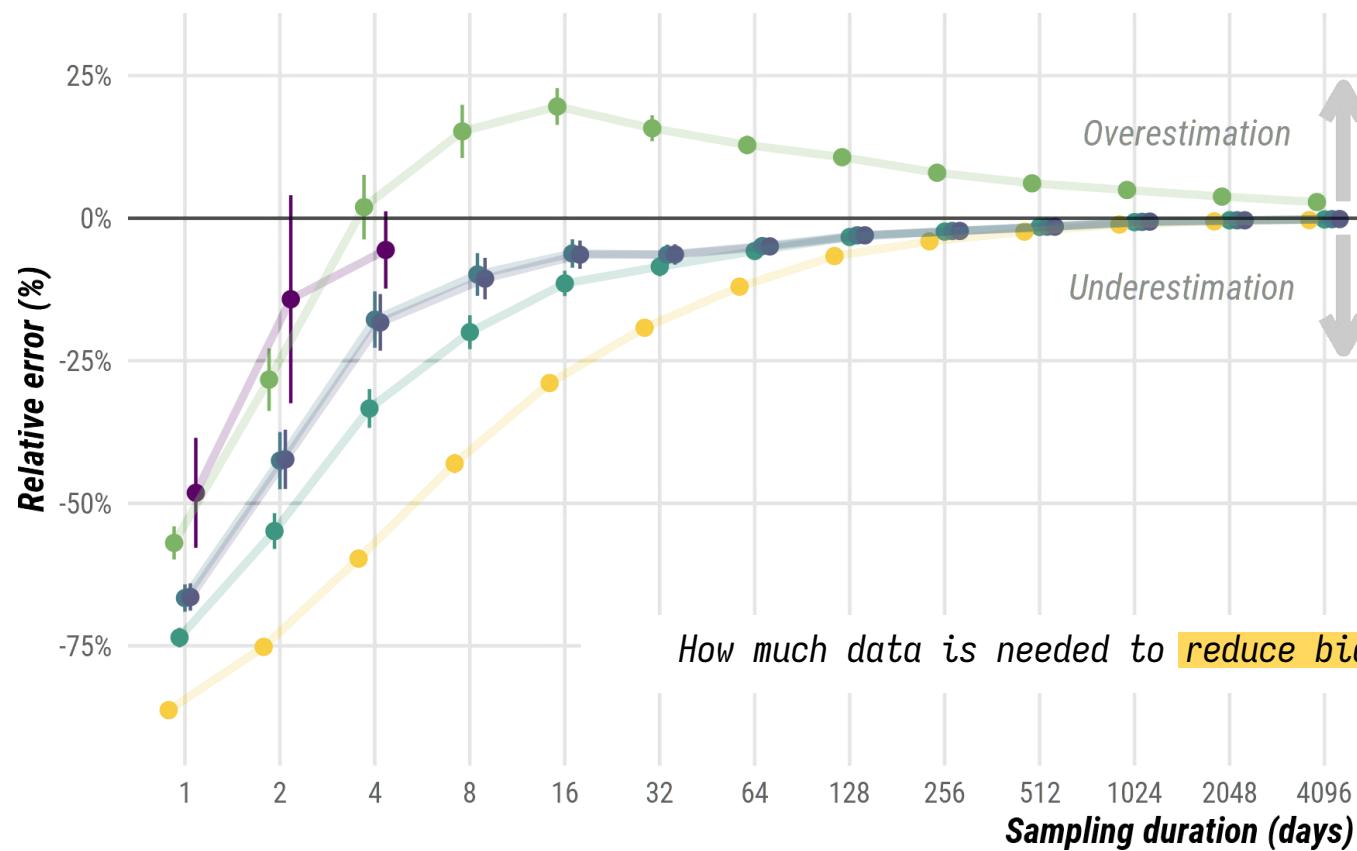
## Parametric bootstrapping

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.





## Relative error vs. sampling duration

Simulated data with  $\tau_p = 1$  day**Method:**

- KDE
- AKDE
- AKDE<sub>c</sub>
- pHREML AKDE<sub>c</sub>
- pHREML wAKDE<sub>c</sub>
- Bootstrapped pHREML wAKDE<sub>c</sub>

How much data is needed to reduce bias sufficiently?

Computational *cost* vs. sampling duration