Center for Advanced Systems Understanding (CASUS)
Helmholtz-Zentrum Dresden-Rossendorf (*HZDR*)

# CTMM STUDY DESIGN

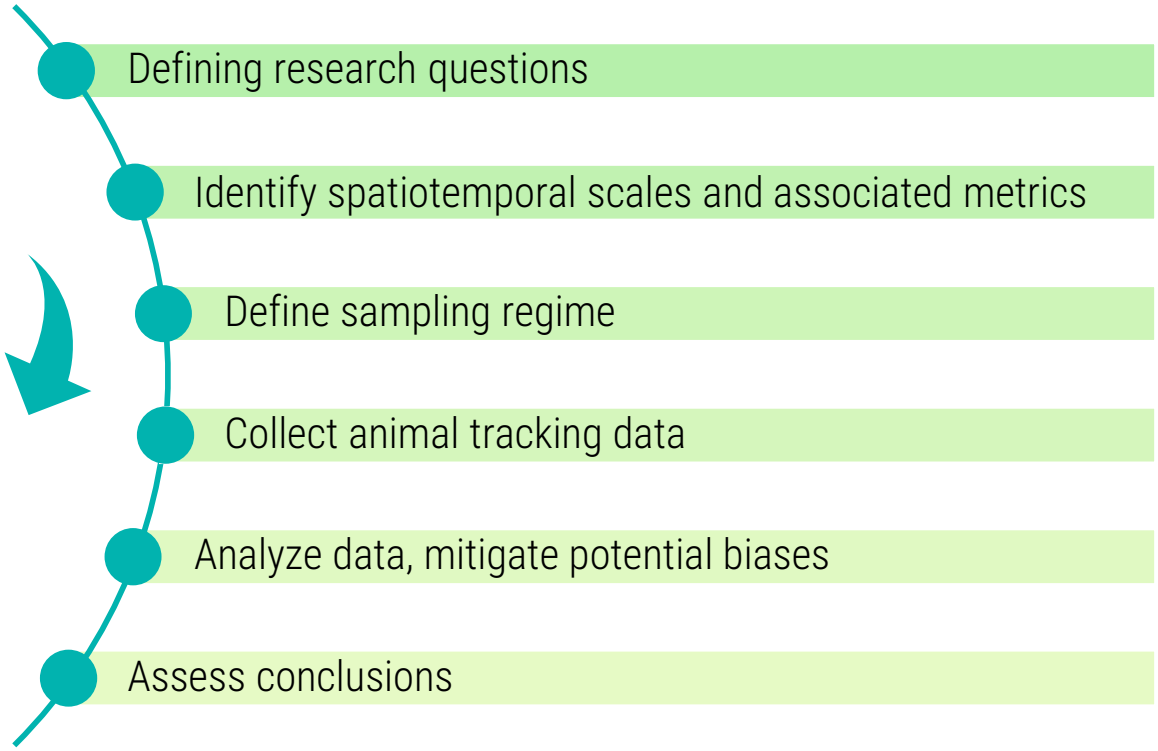
## USING THE MOVEDESIGN APP

Inês Silva, Chris Fleming

i.simoes-silva@hzdr.de

# WORKFLOW






**(1)** Cost of GPS devices (and data transfer),
**(2)** challenges of deployment, and
**(3)** technological limitations,
can all constrain study design.

# WORKFLOW



- Defining research questions
- Identify spatiotemporal scales and associated metrics
- Define sampling regime
- Collect animal tracking data
- Analyze data, mitigate potential biases
- Assess conclusions

*What do I do with the data I've already collected?*

"To consult the statistician after an experiment is finished is often merely to ask him to conduct a postmortem examination.

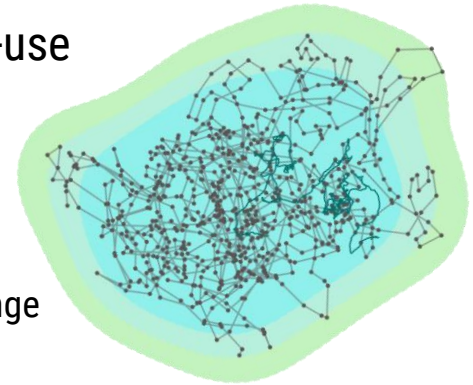
He can perhaps say what the experiment died of."

# RESEARCH QUESTION



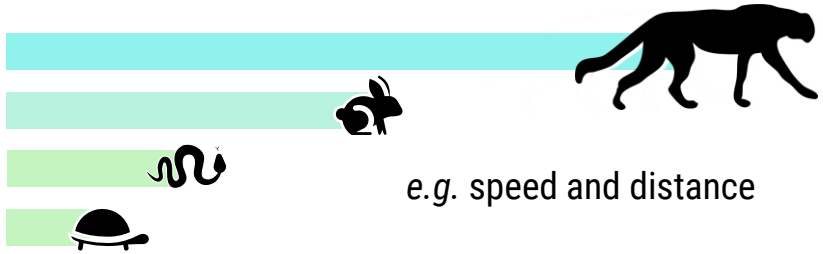
## WHAT DO YOU WANT TO KNOW?

### Space-use



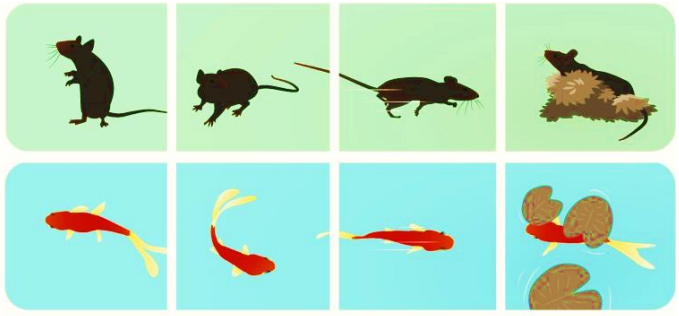
*e.g.* home range

### Movement behavior

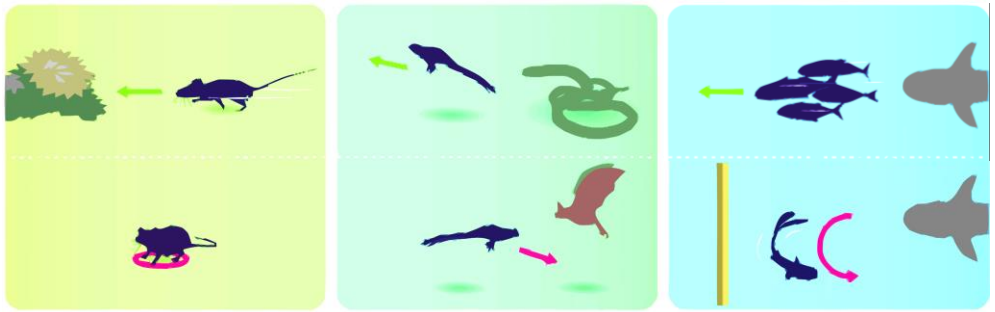


*e.g.* speed and distance

### Behavioral states



### Collective behavior, interactions



Modified from Evan *et al.* (2019)
DOI: 10.1016/j.tics.2019.01.012

# RESEARCH QUESTION



## UTILIZATION DISTRIBUTION

### RANGE DISTRIBUTION

*extrapolate* space use into the future

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

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

≠

### OCCURRENCE DISTRIBUTION

*interpolate* between data points in the past

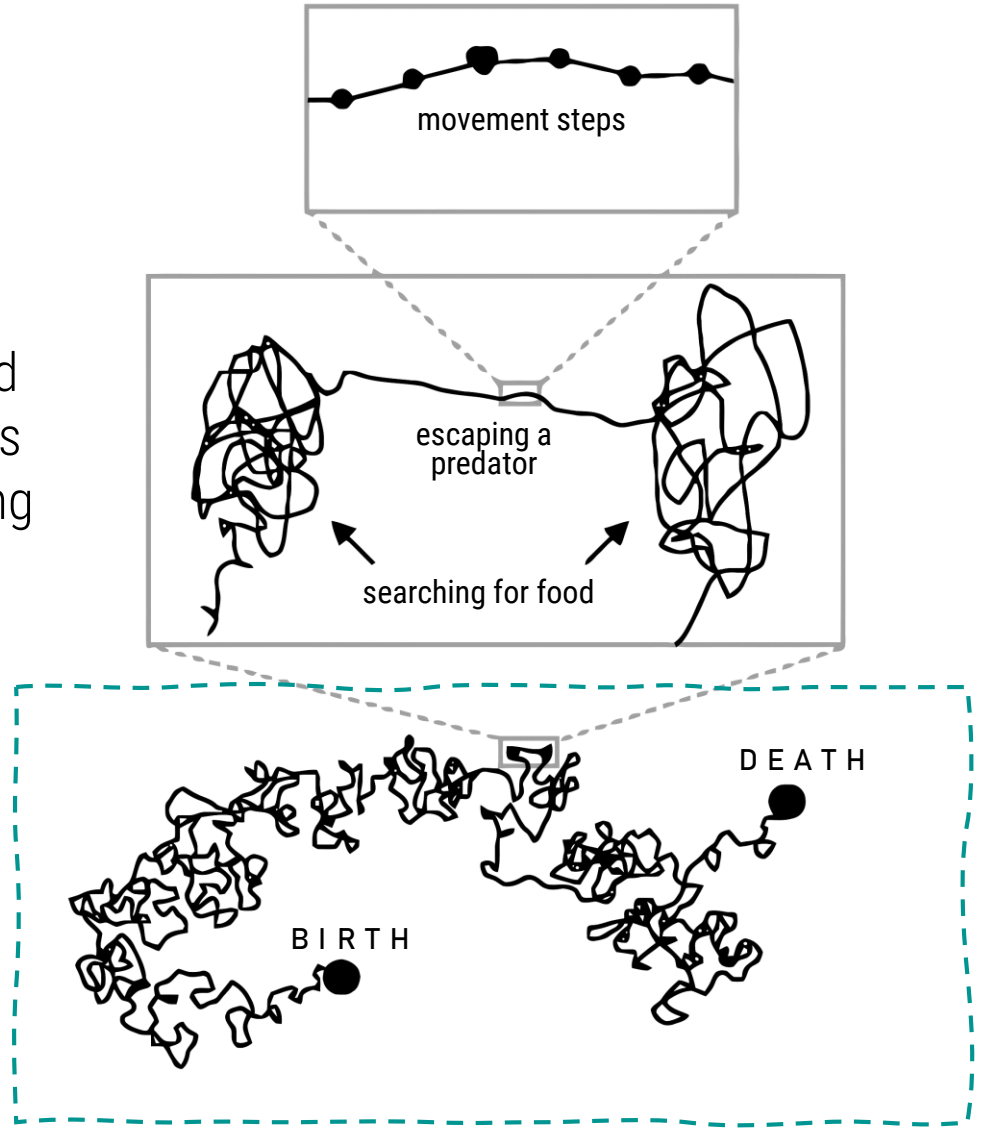
*"Where did an animal go during a period of observation?"*

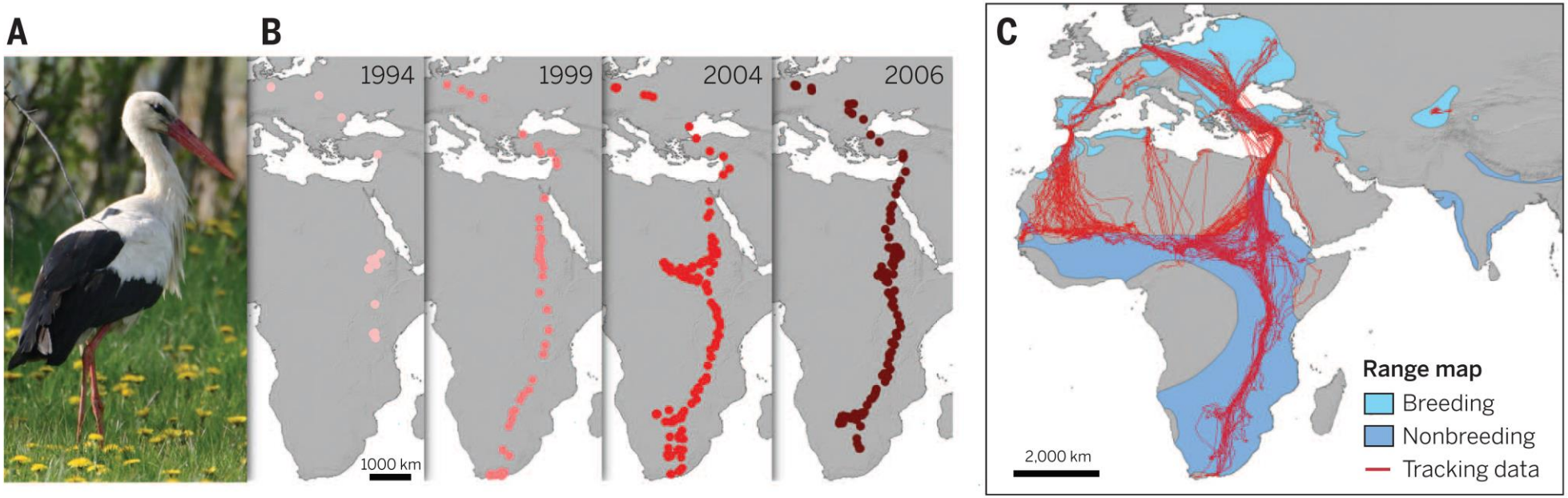
Where did an animal cross a linear feature?
How likely is it that it visited a location of interest?
How much time did it spend in a specific habitat?

# SCALES & METRICS



Home range as the area repeatedly used throughout an animal's **lifetime** for all its normal behaviors and activities, excluding occasional exploratory excursions.





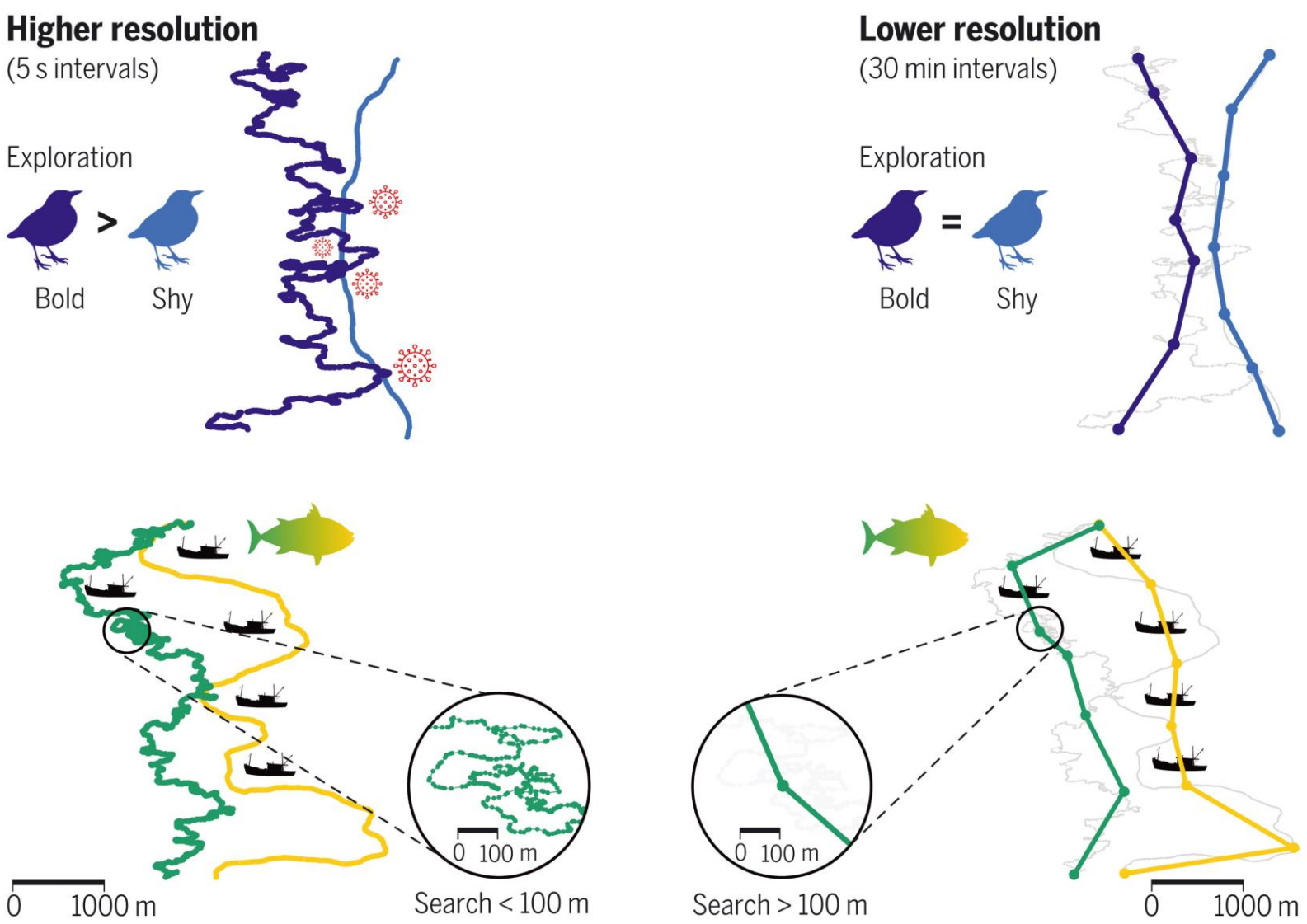

Kays *et al.* (2015)

DOI: 10.1126/science.abg1780

# SCALES & METRICS



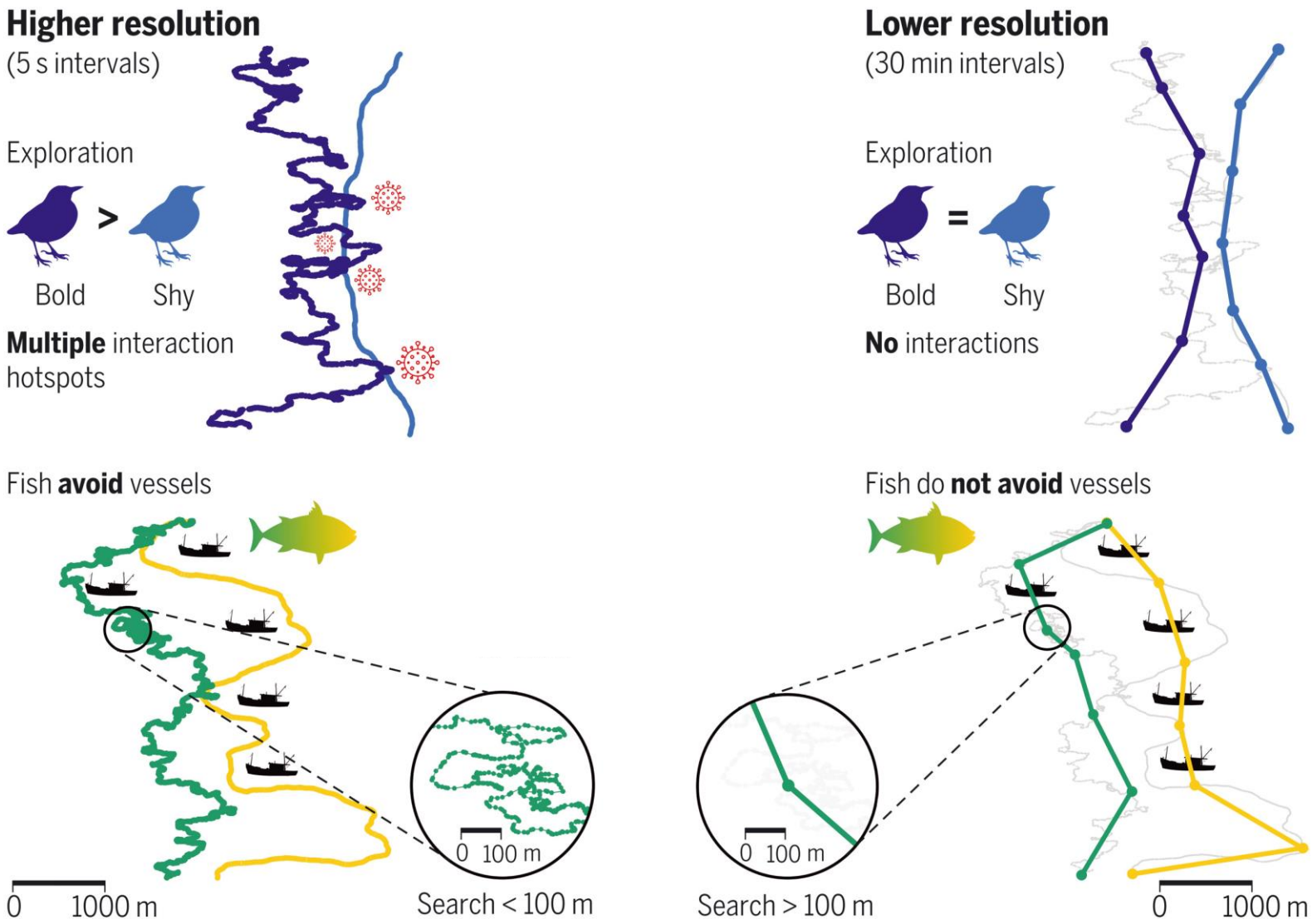
Nathan *et al.* (2022)
DOI: 10.1126/science.abg1780



# SCALES & METRICS

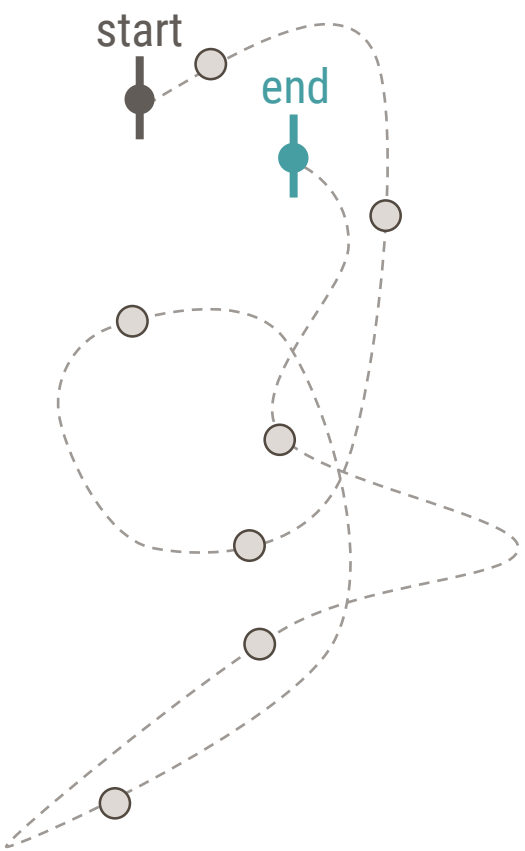


Nathan *et al.* (2022)
DOI: 10.1126/science.abg1780



# SAMPLING REGIME






▸ WHAT TRACKING REGIME DO WE SET?

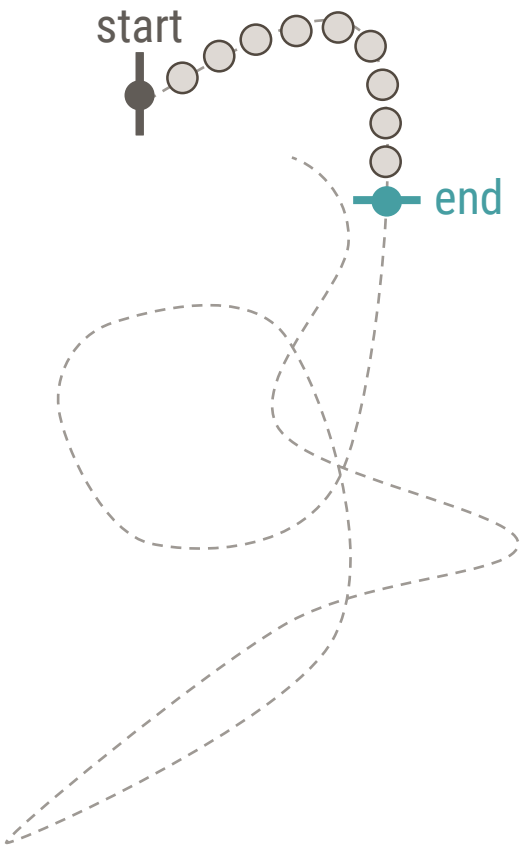
**SAMPLING DURATION**
*How long is an animal tracked for?*

**SAMPLING FREQUENCY**
*How frequently are locations collected?*

Long duration but coarse/low-resolution.

# SAMPLING REGIME






## WHAT TRACKING REGIME DO WE SET?

**SAMPLING DURATION**
*How long is an animal tracked for?*

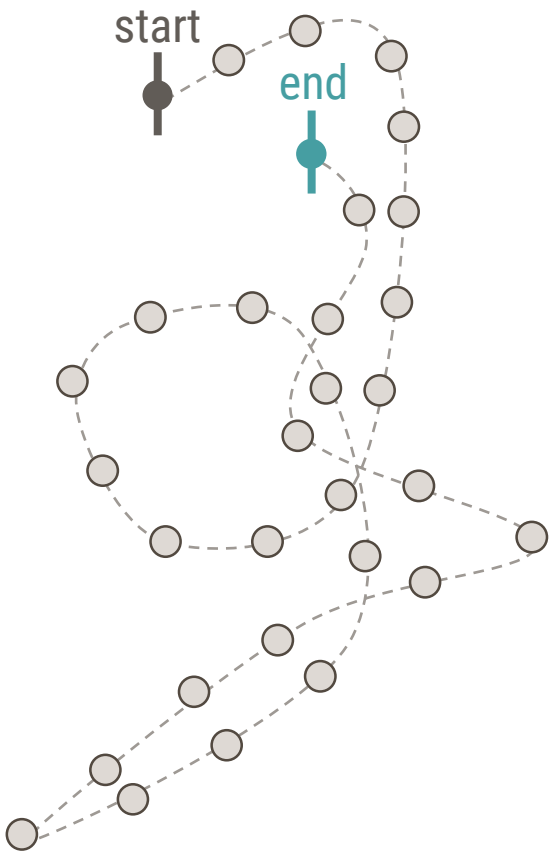
**SAMPLING FREQUENCY**
*How frequently are locations collected?*

Long duration but coarse/low-resolution.

Short-duration but fine/high-resolution.

# SAMPLING REGIME






## WHAT TRACKING REGIME DO WE SET?

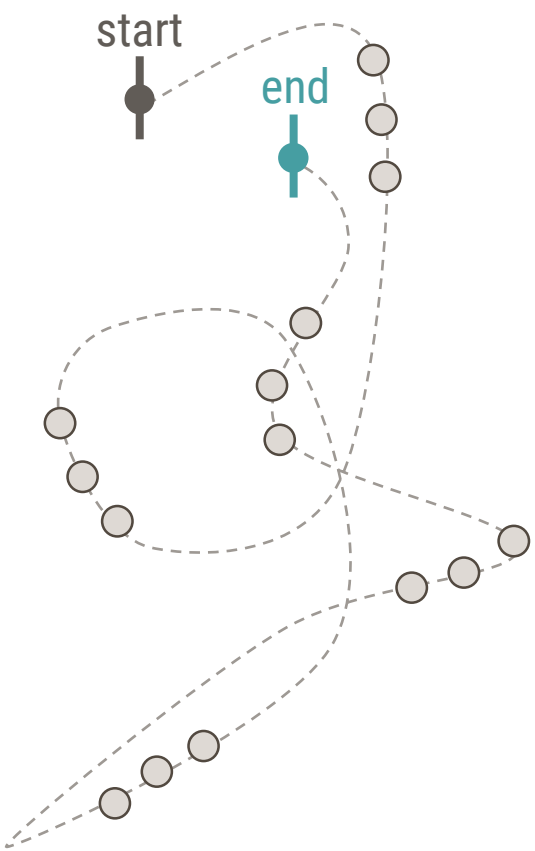
**SAMPLING DURATION**
*How long is an animal tracked for?*

**SAMPLING FREQUENCY**
*How frequently are locations collected?*

Long duration but coarse/low-resolution.
Short-duration but fine/high-resolution.
Long duration and fine/high-resolution.

# SAMPLING REGIME






## WHAT TRACKING REGIME DO WE SET?

**SAMPLING DURATION**
*How long is an animal tracked for?*

**SAMPLING FREQUENCY**
*How frequently are locations collected?*

Long duration but coarse/low-resolution.
Short-duration but fine/high-resolution.
Long duration and fine/high-resolution.
Duty-cycled, acceleration-informed, etc.

# SAMPLING REGIME



*n*

## ABSOLUTE SAMPLE SIZE

### SAMPLING DURATION

*How long is an animal tracked for?*

`1 month`

### SAMPLING FREQUENCY

*How frequently are locations collected?*

`1 fix every hour`

`709 locations`

Total number of locations

How much *new* information are we collecting?
Is this data *sufficient* to answer our research question?

# SAMPLING REGIME



*n*

## ABSOLUTE SAMPLE SIZE

### SAMPLING DURATION

*How long is an animal tracked for?*

6 months

### SAMPLING FREQUENCY

*How frequently are locations collected?*

1 fix every 20 minutes

17,520 locations

Total number of locations

How much *new* information are we collecting?
Is this data *sufficient* to answer our research question?

# SAMPLING REGIME



*n*

## ABSOLUTE SAMPLE SIZE

### SAMPLING DURATION

*How long is an animal tracked for?*

1 year

### SAMPLING FREQUENCY

*How frequently are locations collected?*

1 fix every minute

525,949 locations

Total number of locations

How much *new* information are we collecting?
Is this data *sufficient* to answer our research question?

# SAMPLING REGIME



$n$

## ABSOLUTE SAMPLE SIZE

### $N_{area}$ EFFECTIVE SAMPLE SIZE (AREA)

Corresponds to the number of *home range crossing events*

### $N_{speed}$ EFFECTIVE SAMPLE SIZE (SPEED)

Corresponds to the number of *statistically independent speed observations*

For autocorrelated data, $N < n$, and often $N \ll n$

# SAMPLING REGIME



$n$

## ABSOLUTE SAMPLE SIZE

$N_{area}$

## EFFECTIVE SAMPLE SIZE (AREA)

Corresponds to the number of *home range crossing events*

Out of 72% of 369 empirical data sets with >1,000 total observations:

Only 4% had an $N_{area}$ >1,000,
30% had an $N_{area}$ < 30.

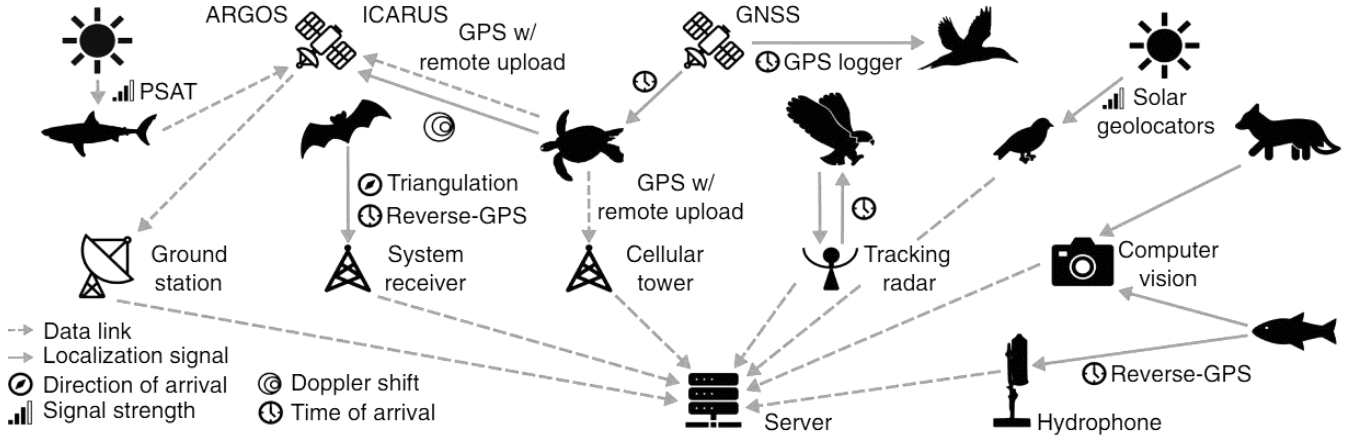
Noonan *et al*. (2019)

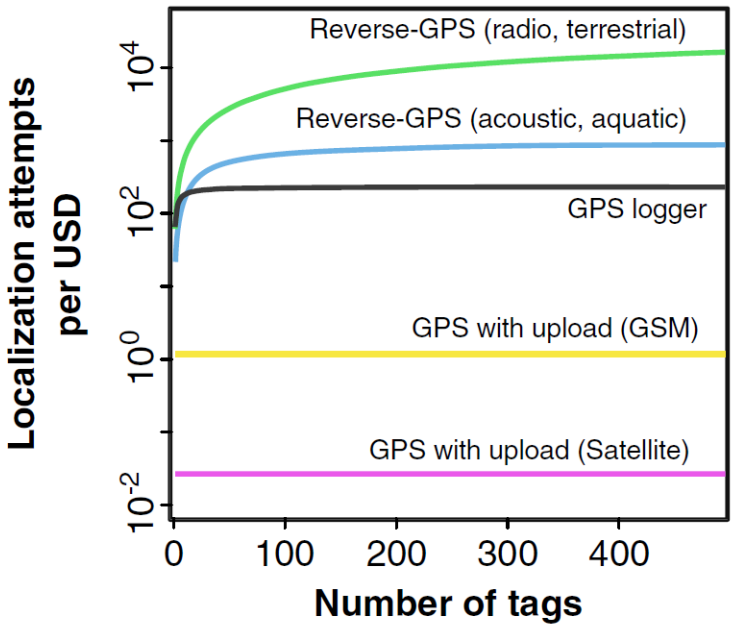
For autocorrelated data, $N < n$, and often $N \ll n$

# SAMPLING REGIME



## WHAT DEVICE DO WE USE?



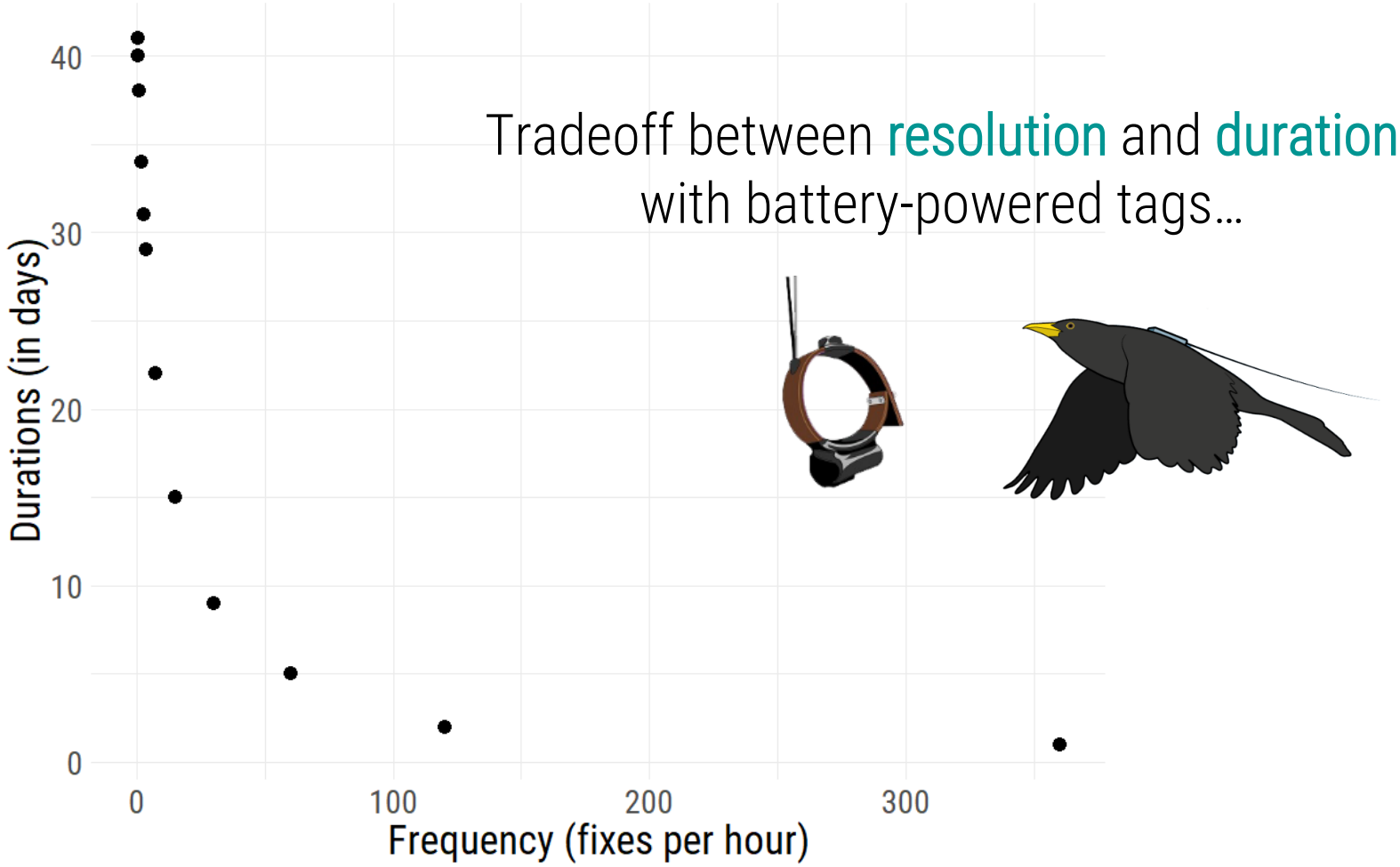




What is cost-effective?

Nathan *et al.* (2022)
DOI: 10.1126/science.abg1780

# SAMPLING REGIME



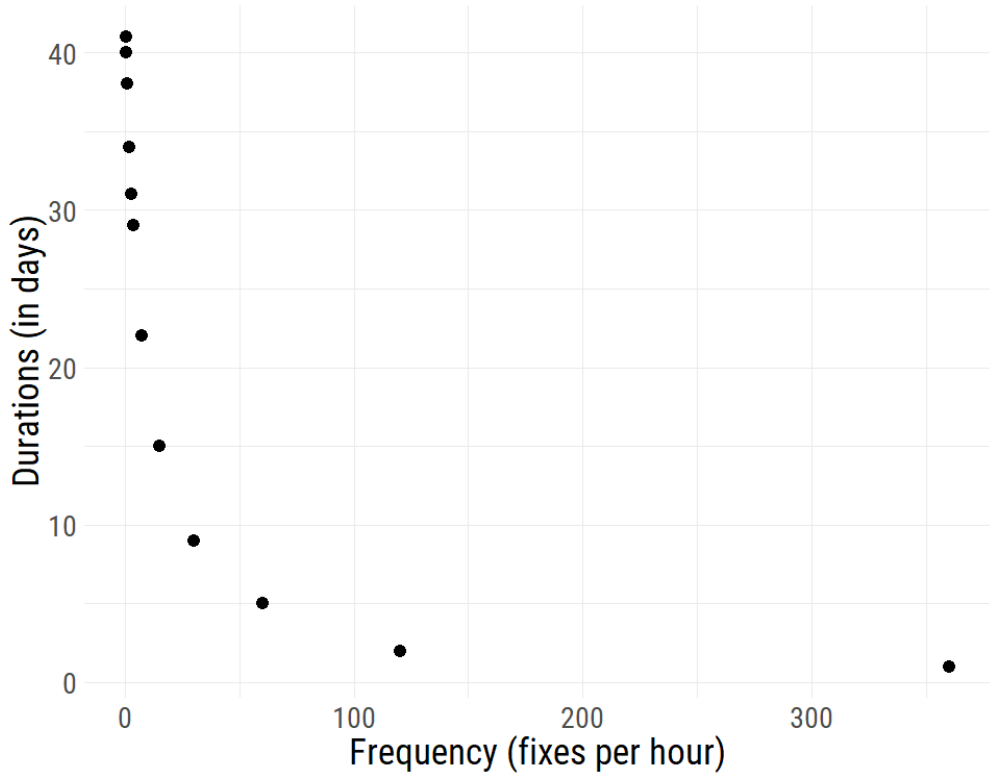


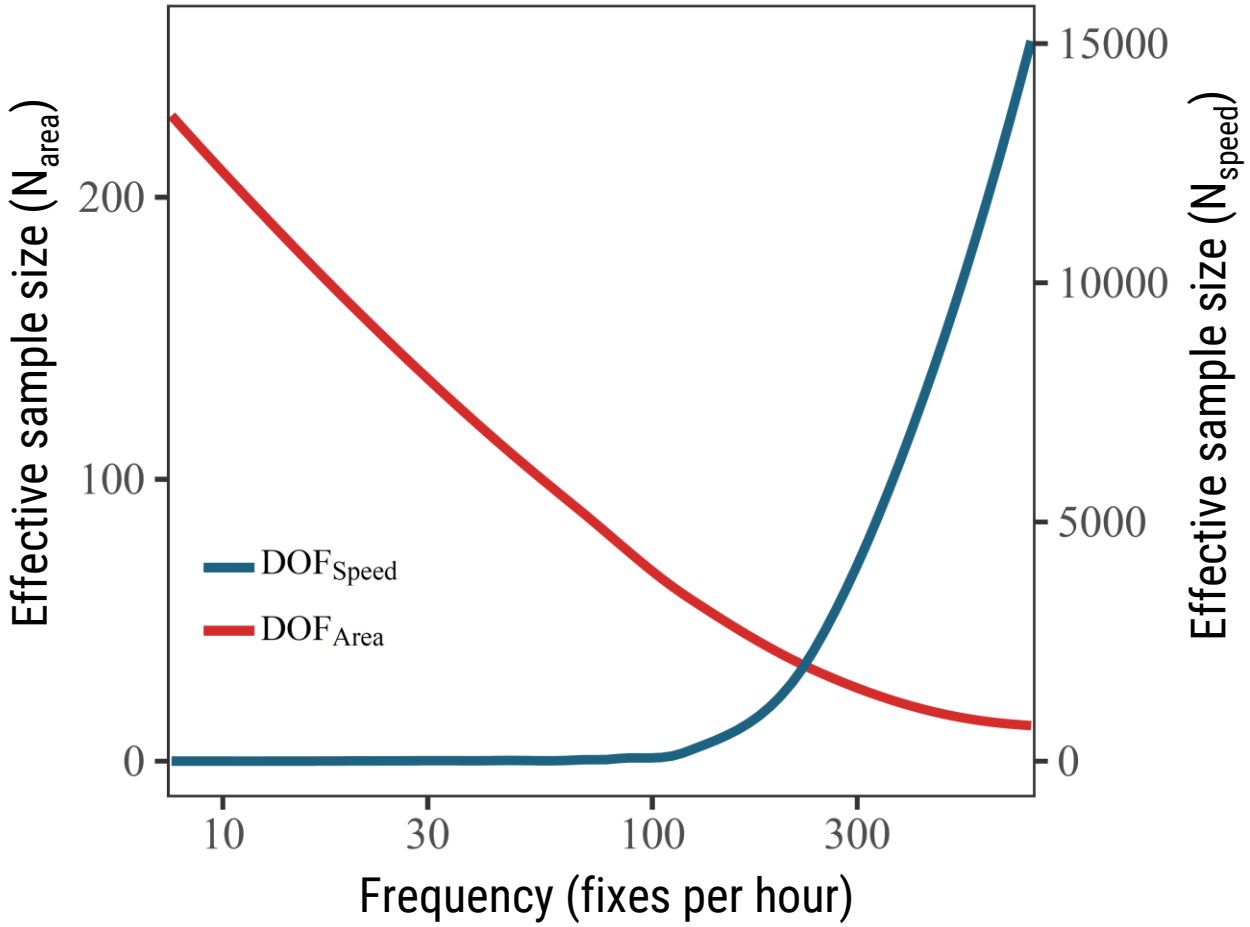

Tradeoff between resolution and duration
with battery-powered tags…

# SAMPLING REGIME



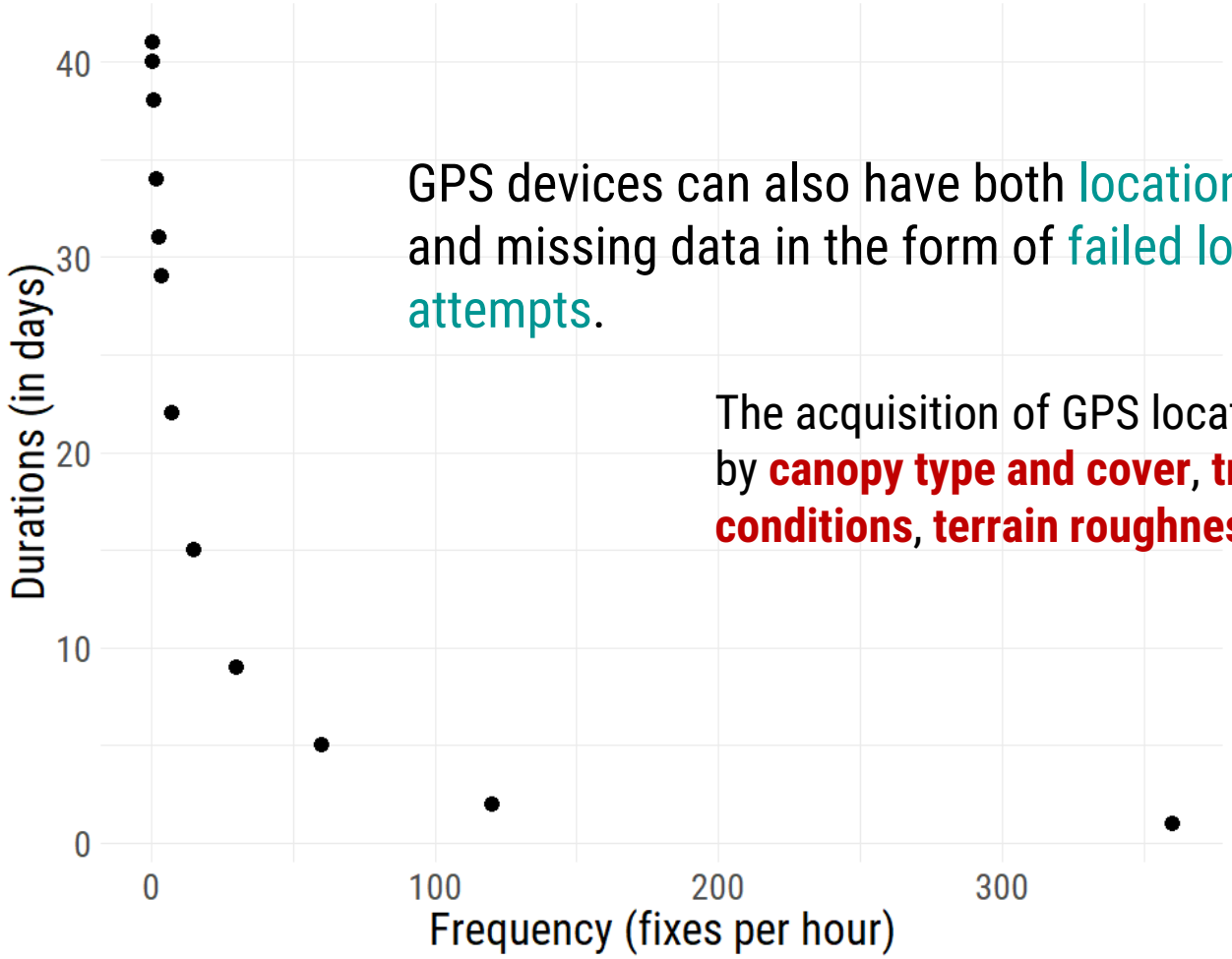
## EFFECTIVE SAMPLE SIZE TRADE-OFFS



# SAMPLING REGIME



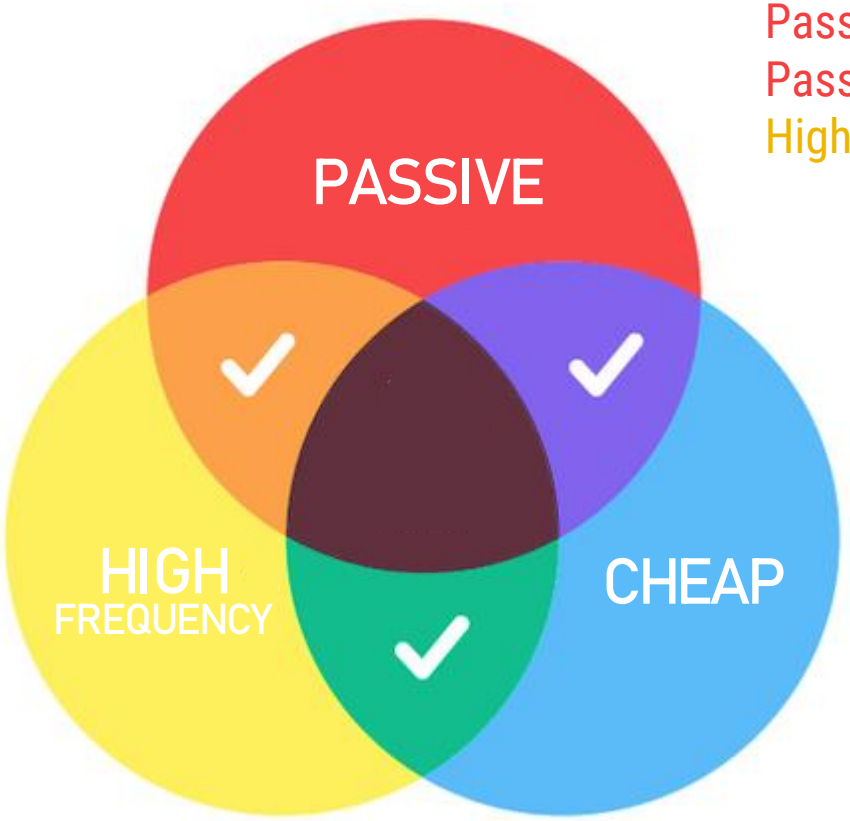



GPS devices can also have both location error and missing data in the form of failed location attempts.

The acquisition of GPS locations can be affected by **canopy type and cover**, **tree density**, **weather conditions**, **terrain roughness**, etc.

# SAMPLING REGIME






Modified from InReach Solutions

There's always a trade-off:

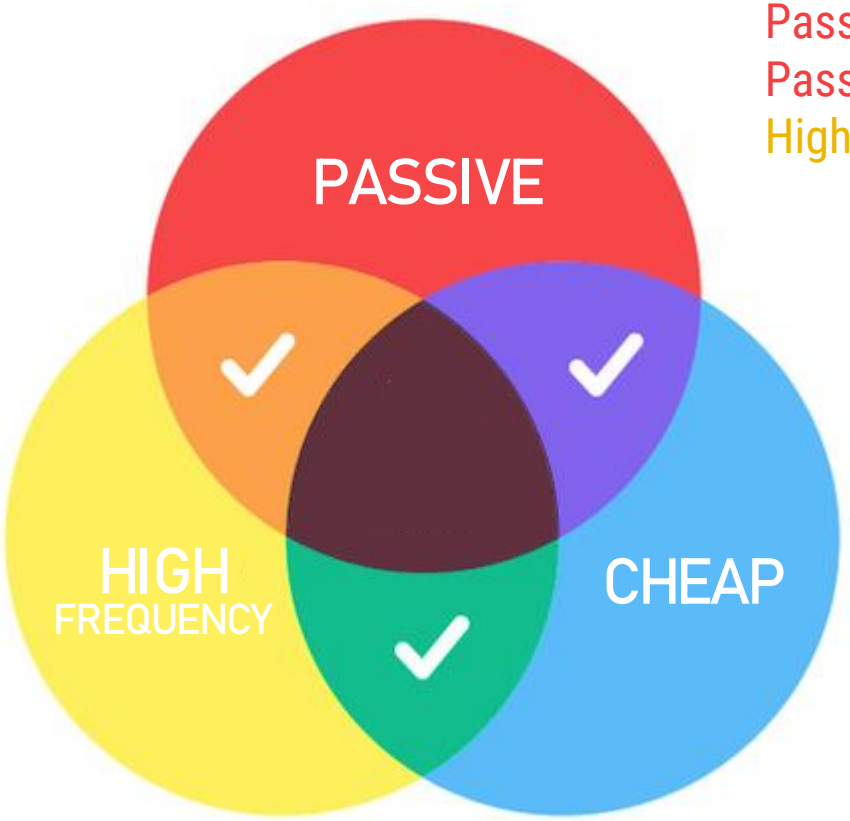
Passive + High-frequency data = expensive tracking devices,
Passive + Cheap = low frequency of data collection,
High-frequency data + Cheap = more time in the field.

▸ WHAT ARE YOUR PRIORITIES?

*"I want max info across as many scales as possible."*

# SAMPLING REGIME






Modified from InReach Solutions

There's always a trade-off:

Passive + High-frequency data = expensive tracking devices,
Passive + Cheap = low frequency of data collection,
High-frequency data + Cheap = more time in the field.

## ▸ WHAT ARE YOUR PRIORITIES?

*"I want max info across as many scales as possible."*

**Even sampling** is best and gives you the most analytical options. Accept tradeoffs and realize you can't have it all.

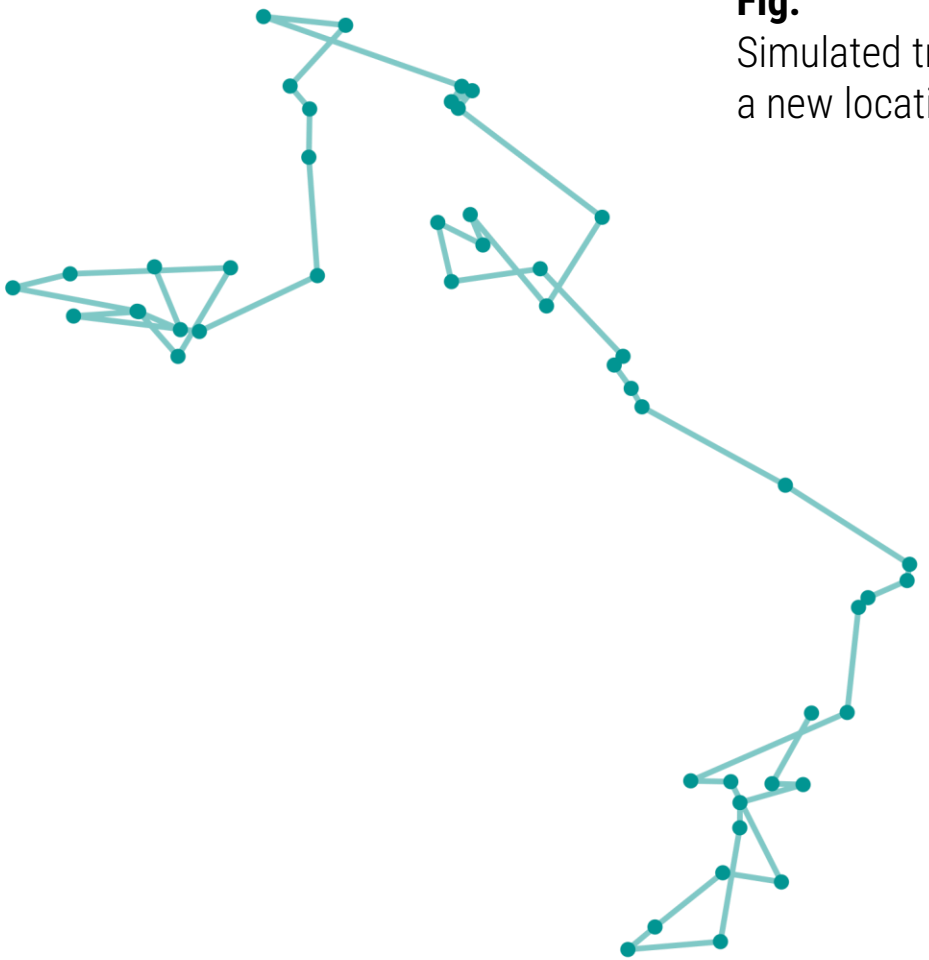


**Fig.**
Simulated tracking dataset with a new location once per day.

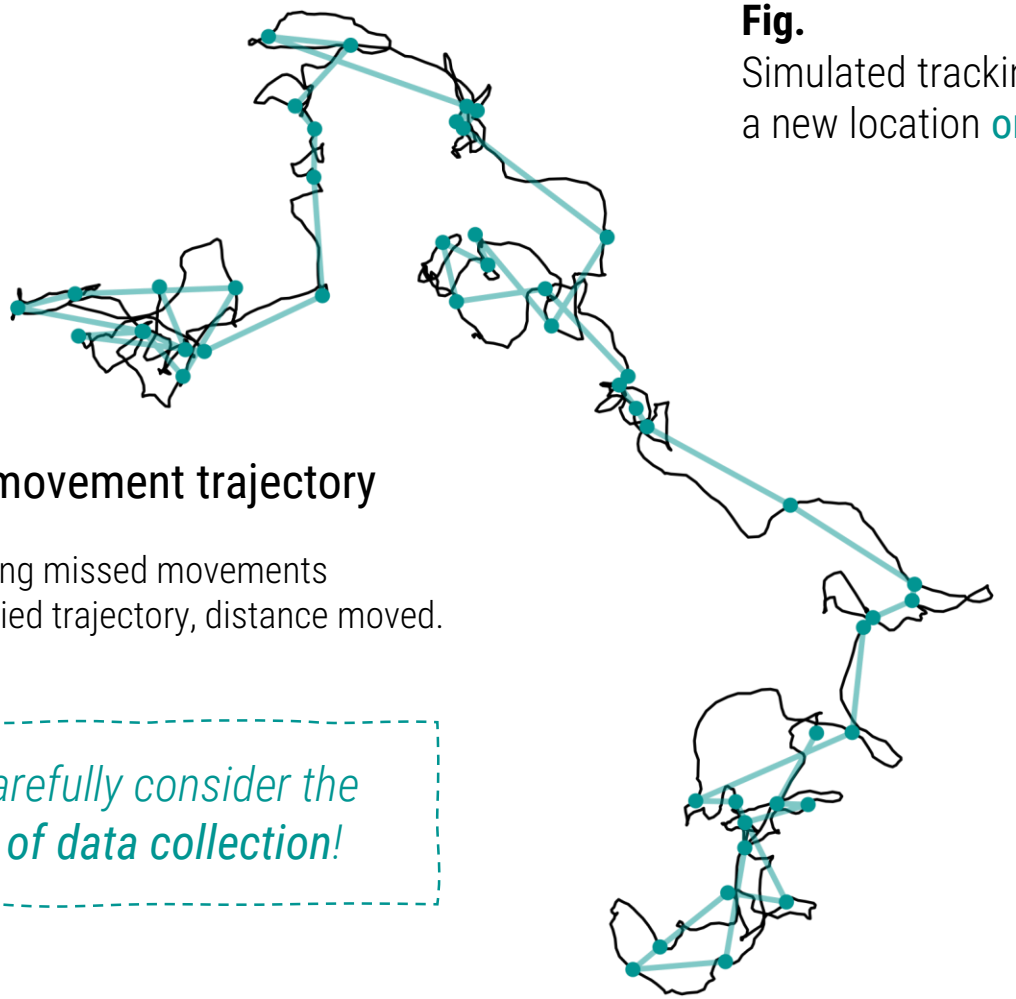


**Fig.**
Simulated tracking dataset with a new location once per day.

Underlying movement trajectory

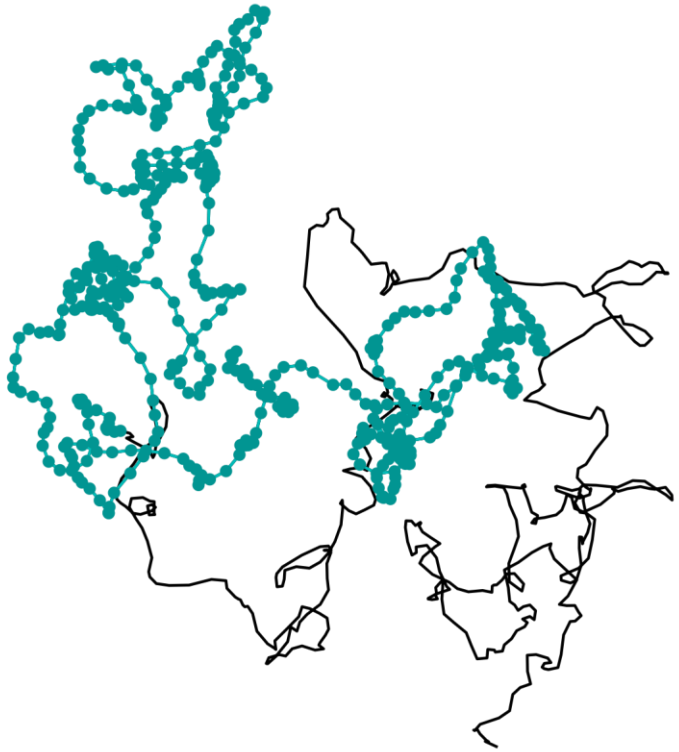
- Sampling missed movements
- Simplified trajectory, distance moved.

*We must carefully consider the* ***frequency of data collection****!*

$\Delta t$ = 1 day
$\tau_v$ = 10 minutes

**Fig.**
Simulated tracking dataset with a sampling duration of 6 months.

**Fig.**
Simulated tracking dataset with a sampling duration of 6 months.

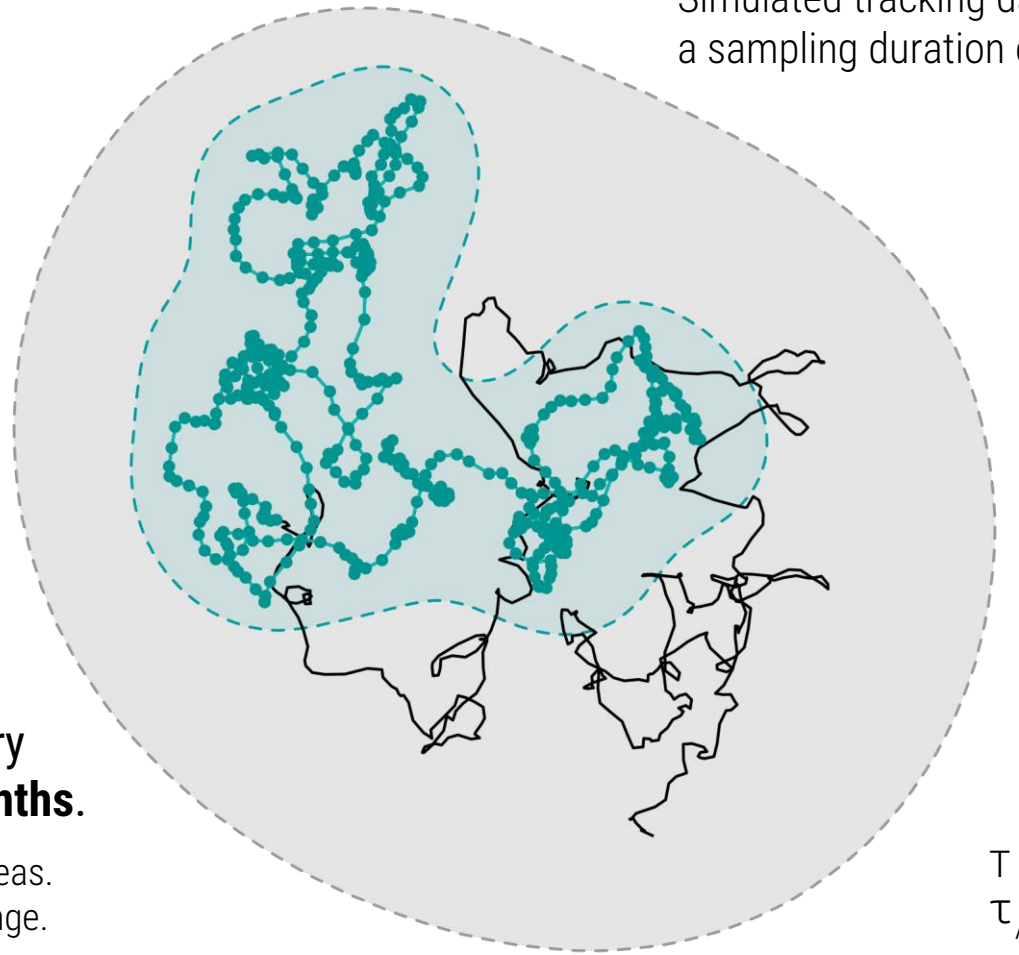


Movement trajectory for the following **6 months**.

- Sampling missed used areas.

*We must carefully consider* ***study duration****!*

**Fig.**
Simulated tracking dataset with a sampling duration of 6 months.



Movement trajectory for the following **6 months**.

- Sampling missed used areas.
- Underestimated home range.

$T$ = 1 year
$\tau_p$ = 8 months

## Autocorrelation is the *rule* in modern movement data

ctmm addresses these problems by elevating autocorrelation from the role of a nuisance factor to that of a **central and informative statistical characteristic** of the movement process.

Multiple quantities can be *autocorrelated*:
Positions, velocities, accelerations, *etc*…

It is not physically possible for these quantities to be uncorrelated.
Only question is: *can you see these correlations in your data?*

Tau (or *τ)* refers to the *characteristic timescales* of autocorrelation.
***i.e.*, time required for the given autocorrelation to decay significantly:**

Here, we refer to two different timescales:

$\tau_p$

POSITION AUTOCORRELATION TIMESCALE

Long-term behavior of a movement process

$\tau_v$

VELOCITY AUTOCORRELATION TIMESCALE

Short-term behavior of a movement process

Notably, autocorrelation in position tends to decay more slowly than autocorrelation in velocity.

$$\tau_p \gg \tau_v$$

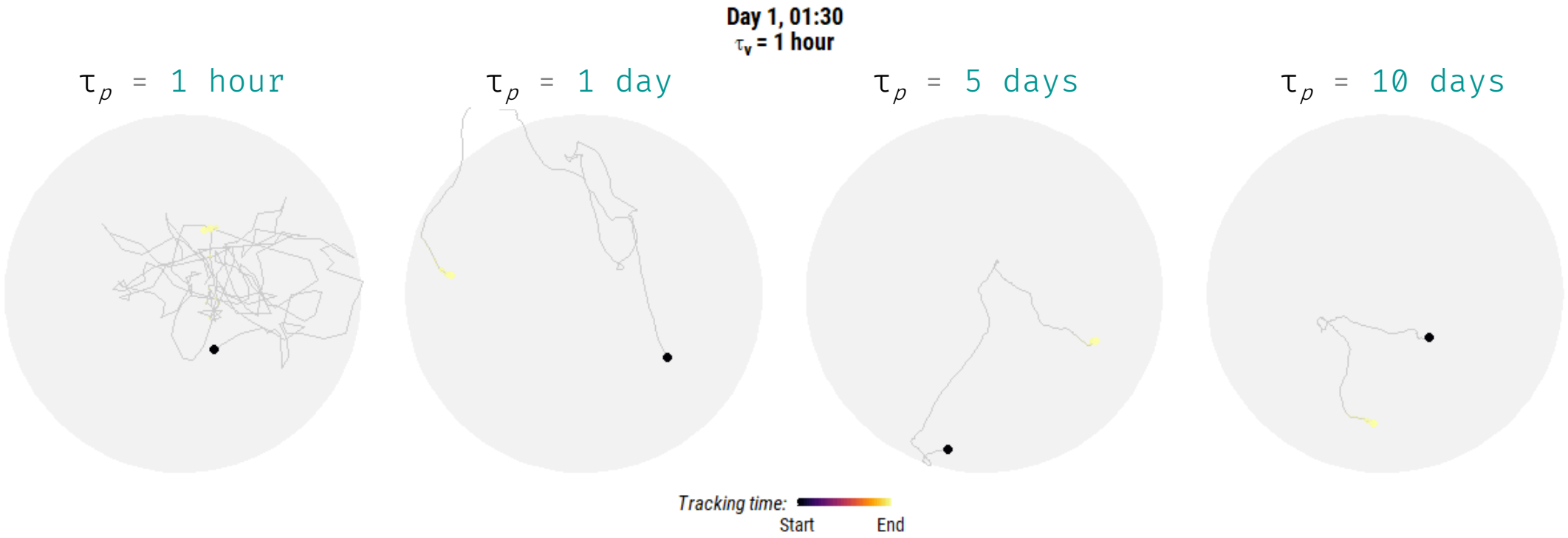
**How long does it take for an animal to traverse the linear extent of its home range?**

$\tau_p$
HOME RANGE CROSSING
PARAMETER

Duration = 1 day
$\tau_p$ = 1 day

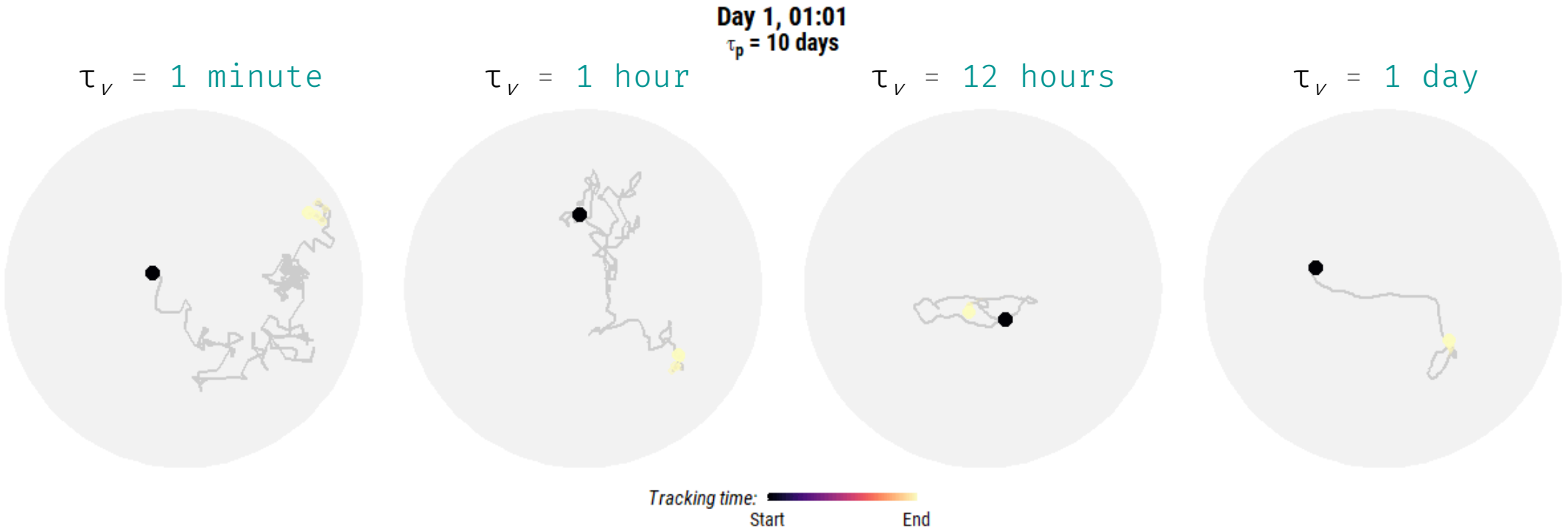
*Tracking time:*
Start End

# $\tau_p$ HOME RANGE CROSSING PARAMETER




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

# $\tau_V$
# DIRECTIONAL PERSISTENCE PARAMETER




**How long does an animal persists in its motion?**

DESIGN

# PARAMETERS

NT

AFRICAN BUFFALO

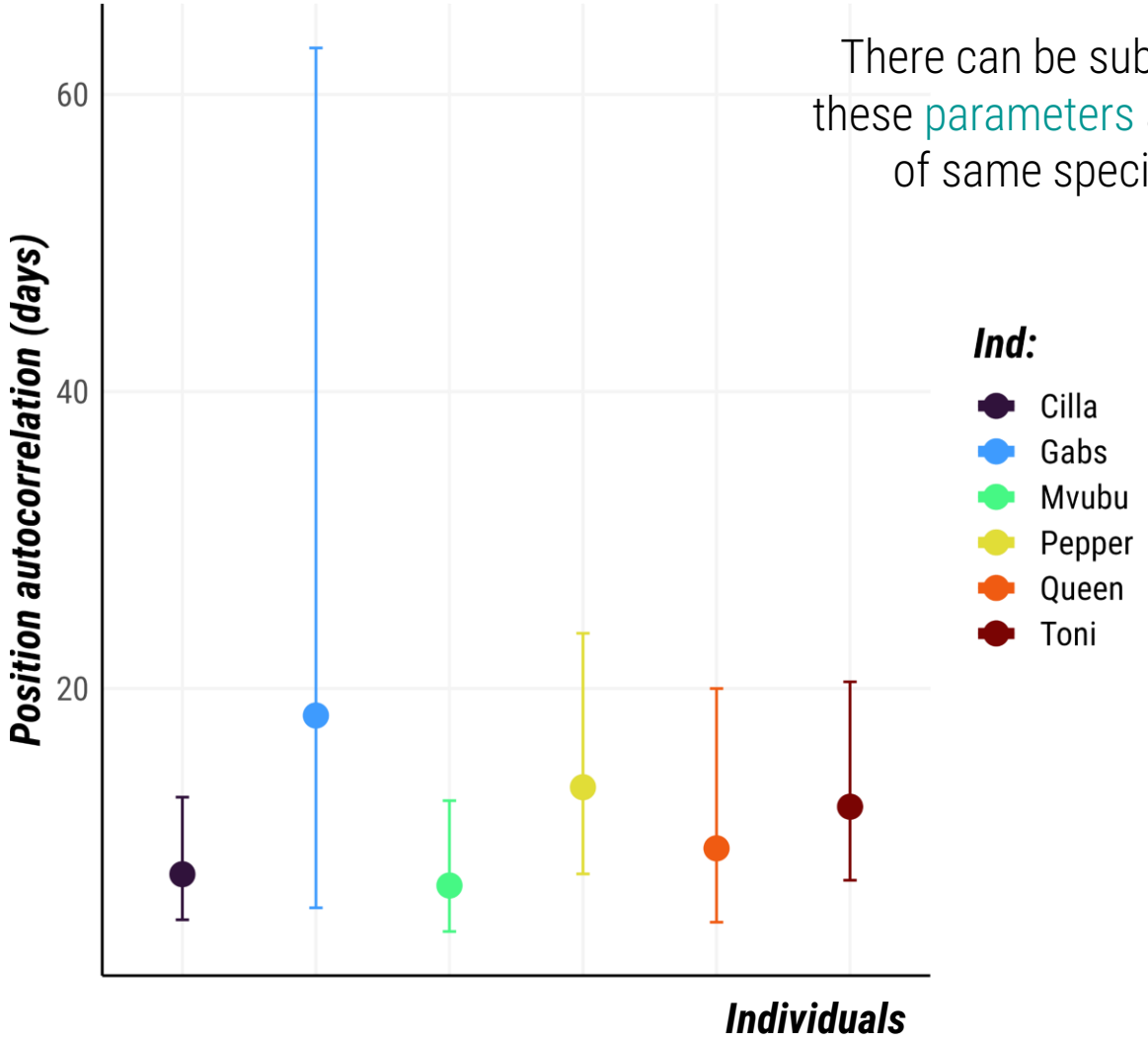
SYNCERUS CAFFER



LC

MONGOLIAN GAZELLE

PROCOPRA GUTTUROSA




There can be substantial variability in these parameters among the individuals of same species, or population.

# PARAMETERS



NT

AFRICAN BUFFALO

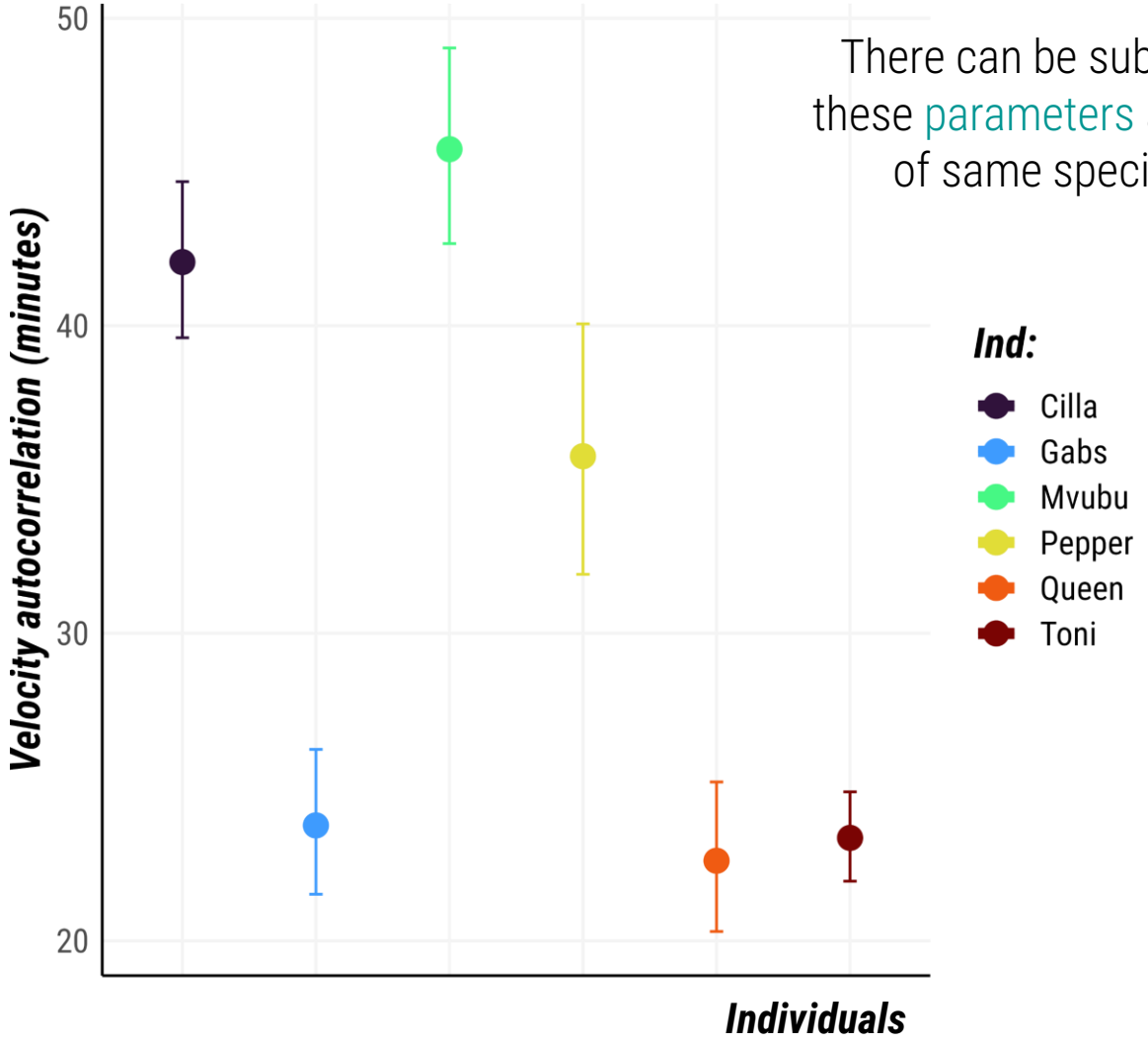
SYNCERUS CAFFER



LC

MONGOLIAN GAZELLE

PROCOPRA GUTTUROSA




There can be substantial variability in these parameters among the individuals of same species, or population.

# PARAMETERS



NT



AFRICAN BUFFALO

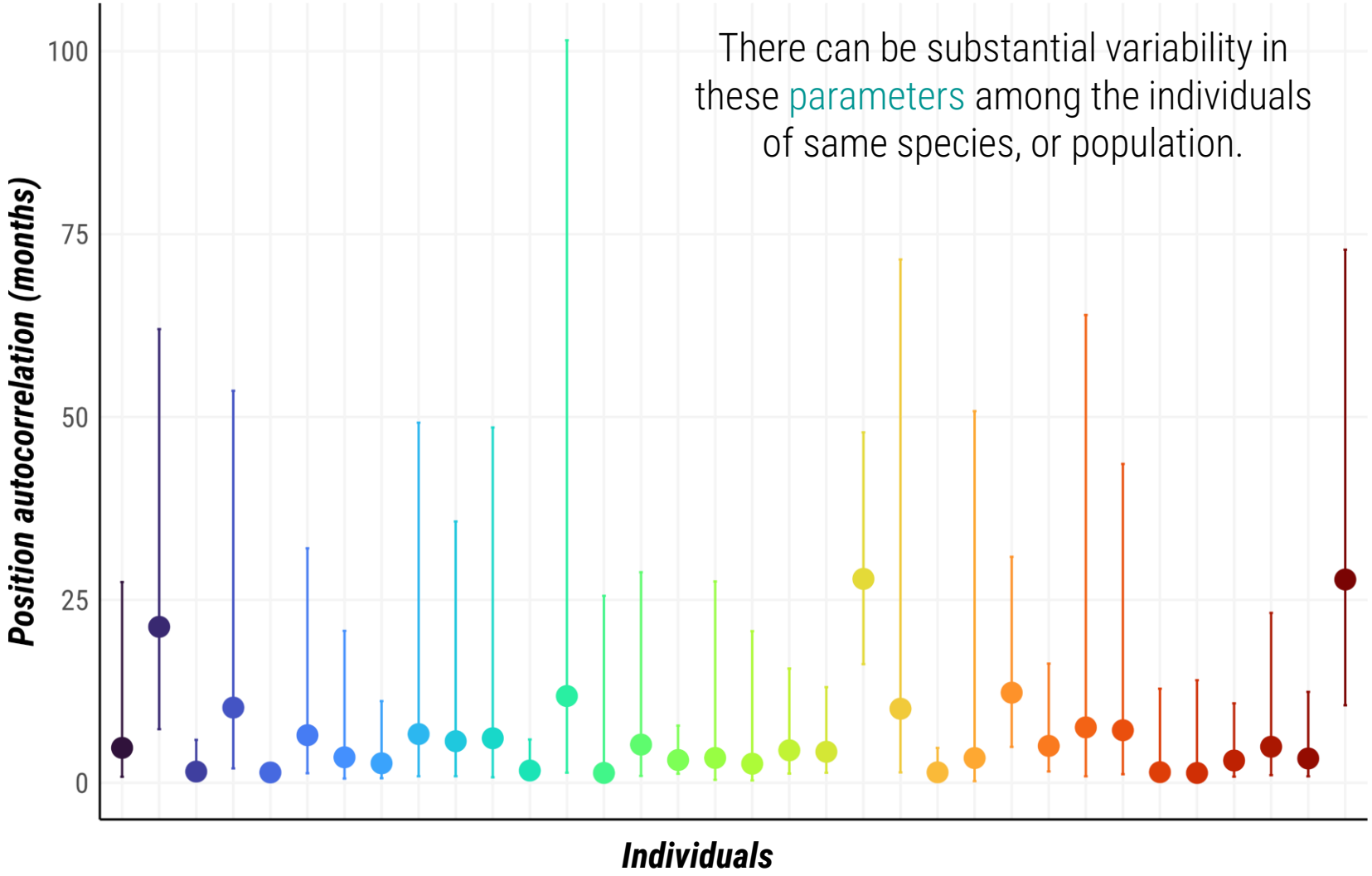
SYNCERUS CAFFER

LC

MONGOLIAN GAZELLE

PROCOPRA GUTTUROSA

There can be substantial variability in these parameters among the individuals of same species, or population.



# PARAMETERS



NT



AFRICAN BUFFALO

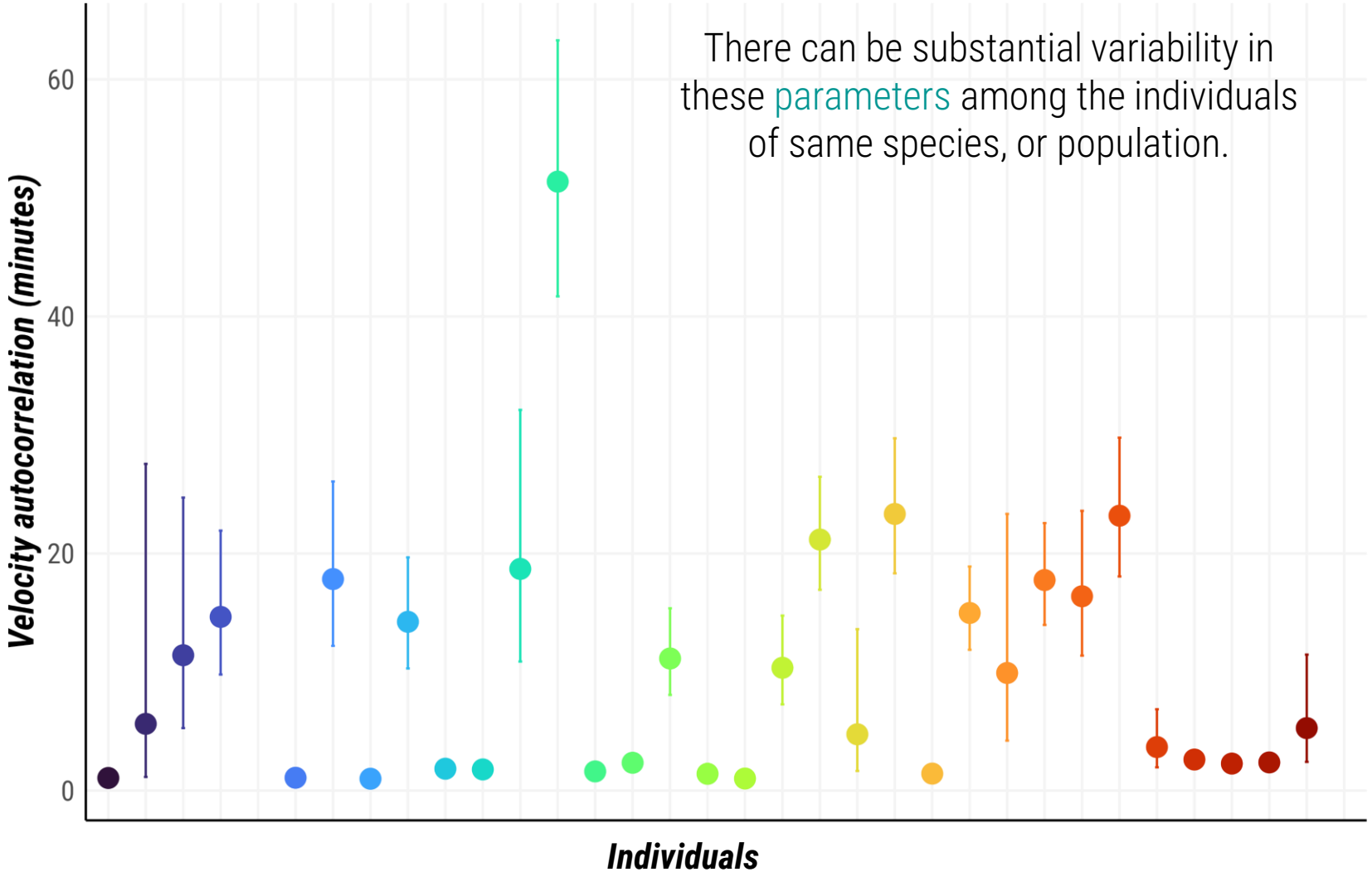
SYNCERUS CAFFER

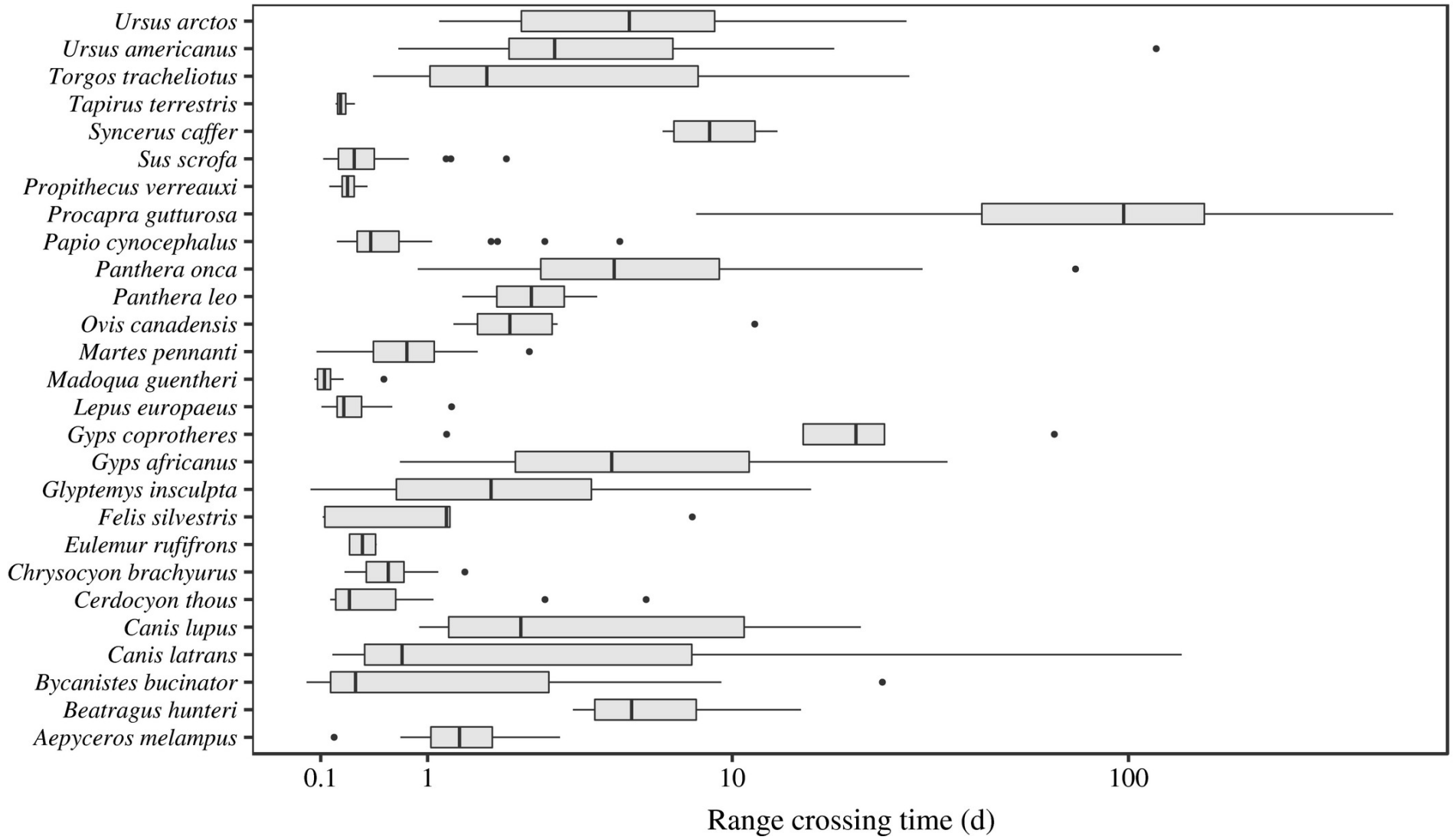
LC

MONGOLIAN GAZELLE

PROCOPRA GUTTUROSA

There can be substantial variability in these parameters among the individuals of same species, or population.




Ursus arctos
Ursus americanus
Torgos tracheliotus
Tapirus terrestris
Syncerus caffer
Sus scrofa
Propithecus verreauxi
Procapra gutturosa
Papio cynocephalus
Panthera onca
Panthera leo
Ovis canadensis
Martes pennanti
Madoqua guentheri
Lepus europaeus
Gyps coprotheres
Gyps africanus
Glyptemys insculpta
Felis silvestris
Eulemur rufifrons
Chrysocyon brachyurus
Cerdocyon thous
Canis lupus
Canis latrans
Bycanistes bucinator
Beatragus hunteri
Aepyceros melampus
0.1
1
10
100
Range crossing time (d)

## RANGE DISTRIBUTION

*extrapolate* space use into the future

Analyses that ignore position autocorrelation significantly underestimate home range areas.

≠

## OCCURRENCE DISTRIBUTION

*interpolate* between data points in the past

Analyses that ignore velocity autocorrelation typically overestimate occurrence areas.

# PARAMETERS



$\tau_p$

POSITION AUTOCORRELATION TIMESCALE

$\tau_v$

VELOCITY AUTOCORRELATION TIMESCALE

**SAMPLING DURATION**

*Duration of animal tracking data.*

**SAMPLING INTERVAL**

*Time between locations.*

*Sampling duration and interval* relative to *characteristic timescales* determine what you can see in the data.

$N_{AREA}$

*Effective sample size (area)*

$N_{SPEED}$

*Effective sample size (speed)*

# RECOMMENDATIONS

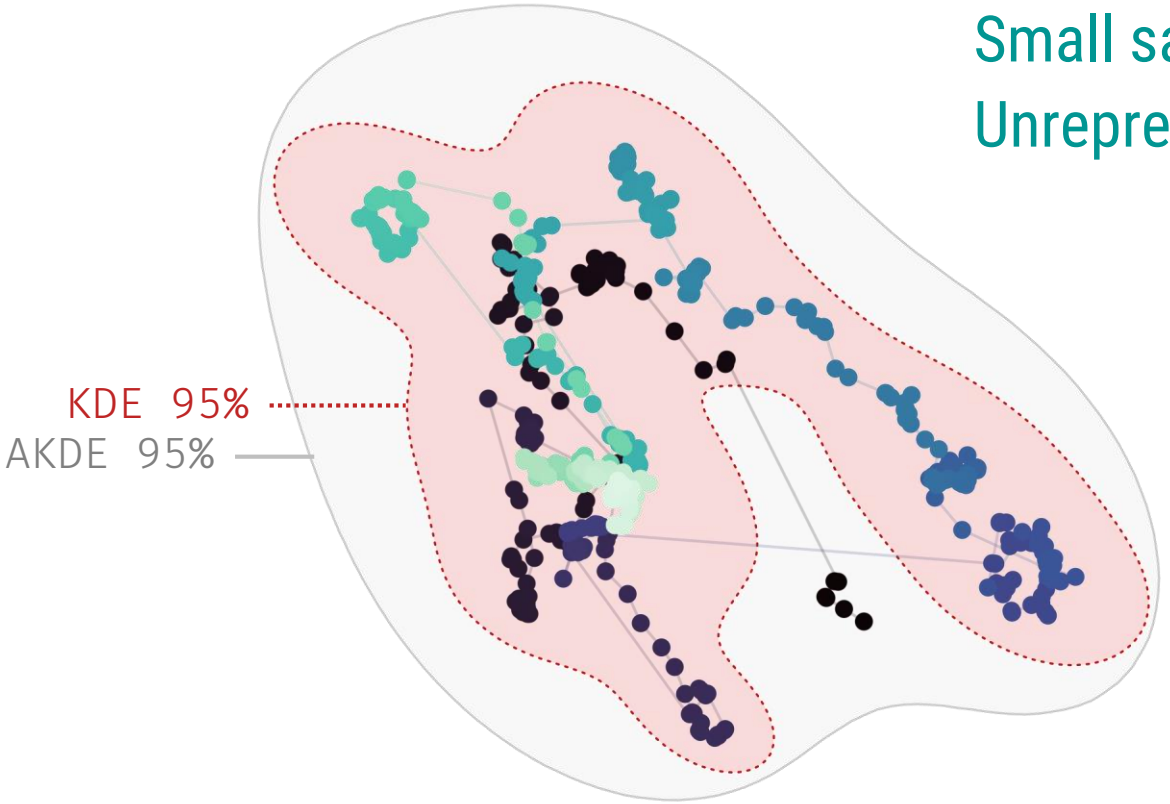


AKDE deals with biases such as

Autocorrelation,

Small sample sizes,

Unrepresentative sampling in time.



# RECOMMENDATIONS



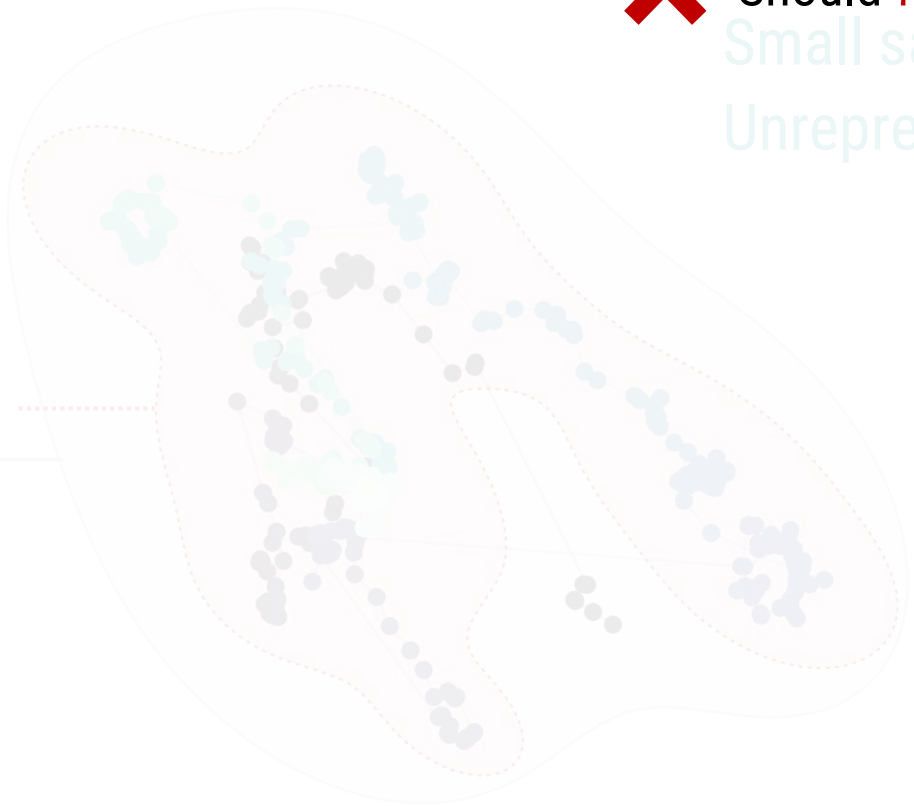
AKDE deals with biases such as

Autocorrelation,

Small sample sizes,

Unrepresentative sampling in time.

✗ Should *not* be a substitute for careful study design.

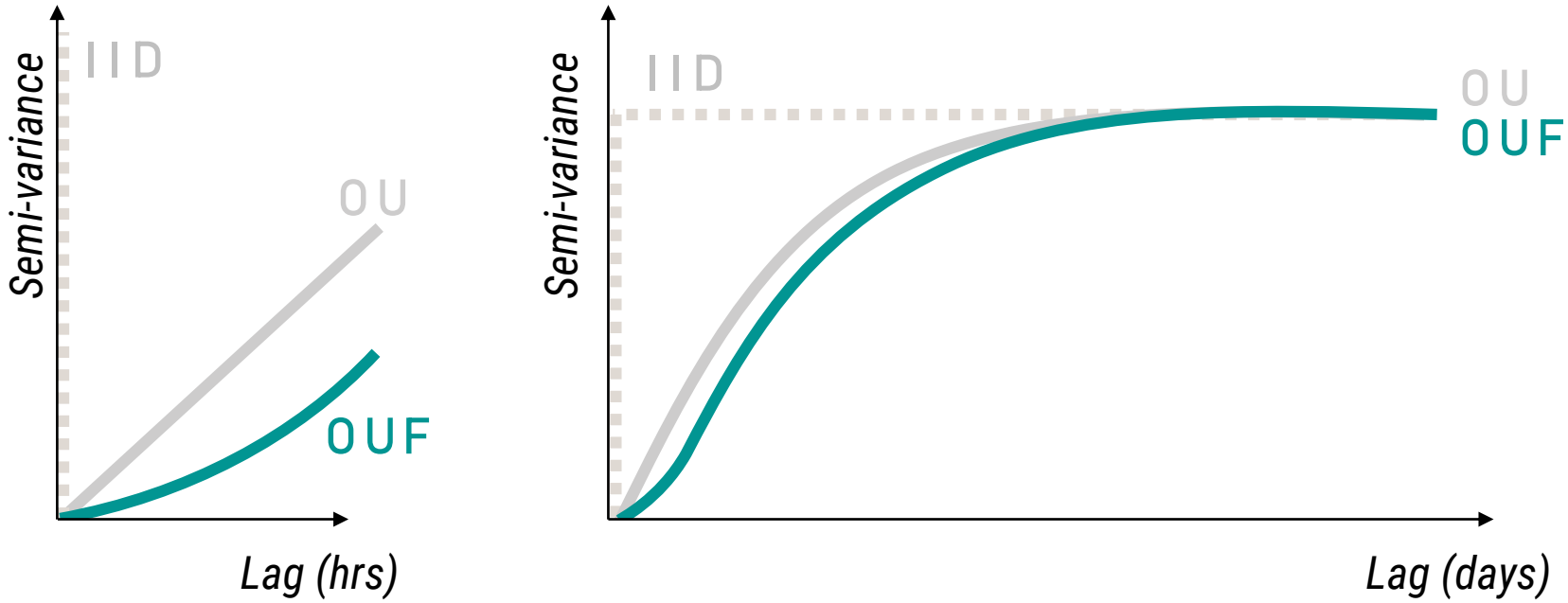


# Sampling *versus* characteristic scales

| | Tracking regime | |
|---|---|---|
| Model | Duration (T) | Sampling interval (Δt) |
| IID | $\tau_p < T$ | $\tau_v < \tau_p < \Delta t$ |
| BM | $T < \tau_p$ | $\tau_v < \Delta t < \tau_p$ |
| OU | $\tau_p < T$ | $\tau_v < \Delta t < \tau_p$ |
| IOU | $T < \tau_p$ | $\Delta t < \tau_v < \tau_p$ |
| OUF | $\tau_p < T$ | $\Delta t < \tau_v < \tau_p$ |

# RECOMMENDATIONS






The OUF model was best-fit for 240 of the 369 empirical GPS datasets,
with the IID model only being selected for 1 individual,
and the OU model for the remaining 128.

# Sampling *versus* characteristic scales

| | Tracking regime | |
|---|---|---|
| Model | Duration (T) | Sampling interval (Δt) |
| IID | $\tau_p < T$ | $\tau_v < \tau_p < \Delta t$ |
| BM | $T < \tau_p$ | $\tau_v < \Delta t < \tau_p$ |
| OU | $\tau_p < T$ | $\tau_v < \Delta t < \tau_p$ |
| IOU | $T < \tau_p$ | $\Delta t < \tau_v < \tau_p$ |
| **OUF** | $\tau_p < T$ | $\Delta t < \tau_v < \tau_p$ |

# RECOMMENDATIONS

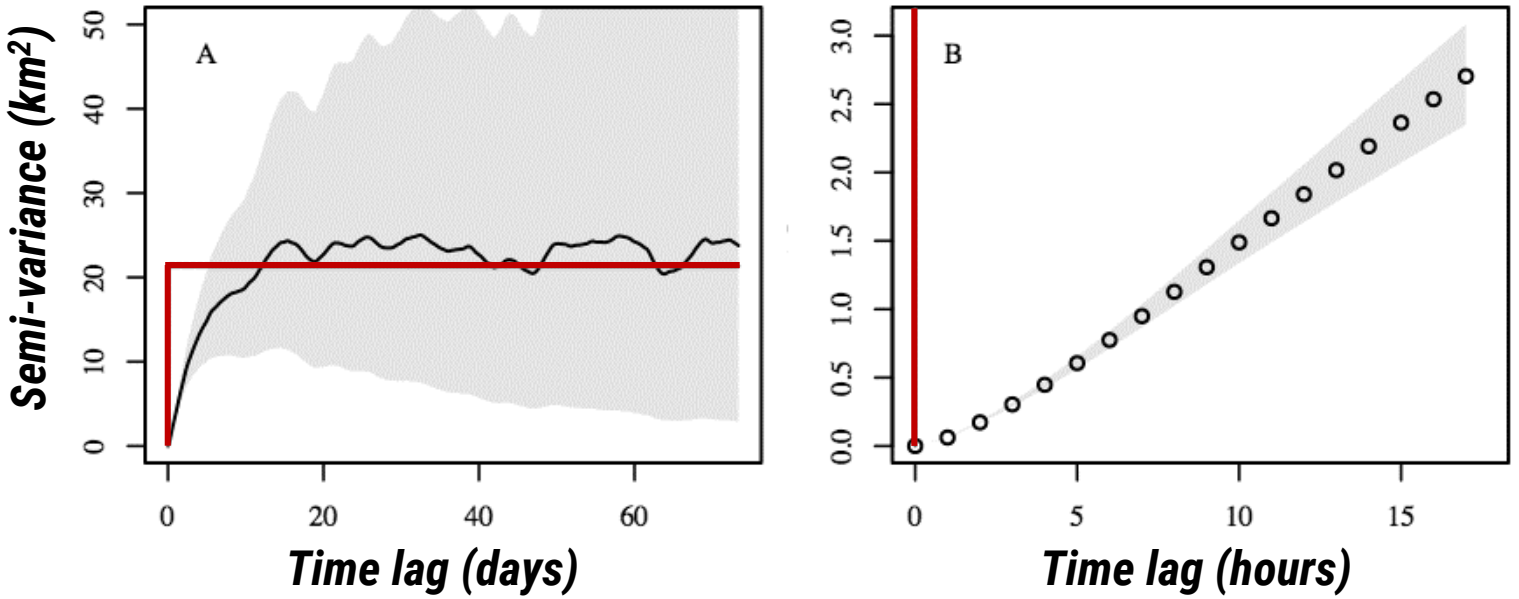


Sampling duration T ≈ 8 months

Sampling interval Δt = 1 hour

| Model | ΔAICc |
|---|---|
| OUF anisotropic | 0 |
| OUF isotropic | 85.4 |
| OU anisotropic | 869.7 |
| OUf anisotropic | 4963.0 |

IID process in red



# RECOMMENDATIONS

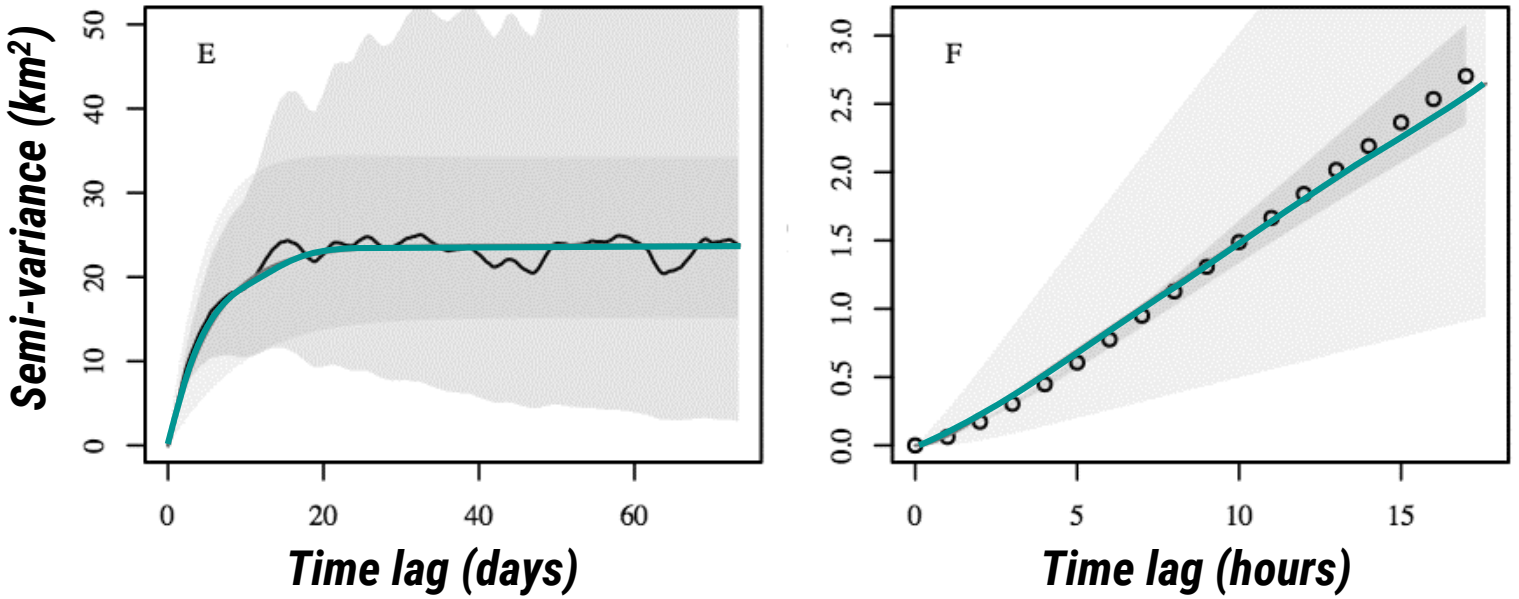


Sampling duration T ≈ 8 months
Sampling interval Δt = 1 hour

**OUF anisotropic process:**

n = 5,766
Narea ≈ 22.3
Nspeed ≈ 2933.9

$\tau_p$ = 11.2 days
$\tau_v$ = 23.4 minutes



# RECOMMENDATIONS

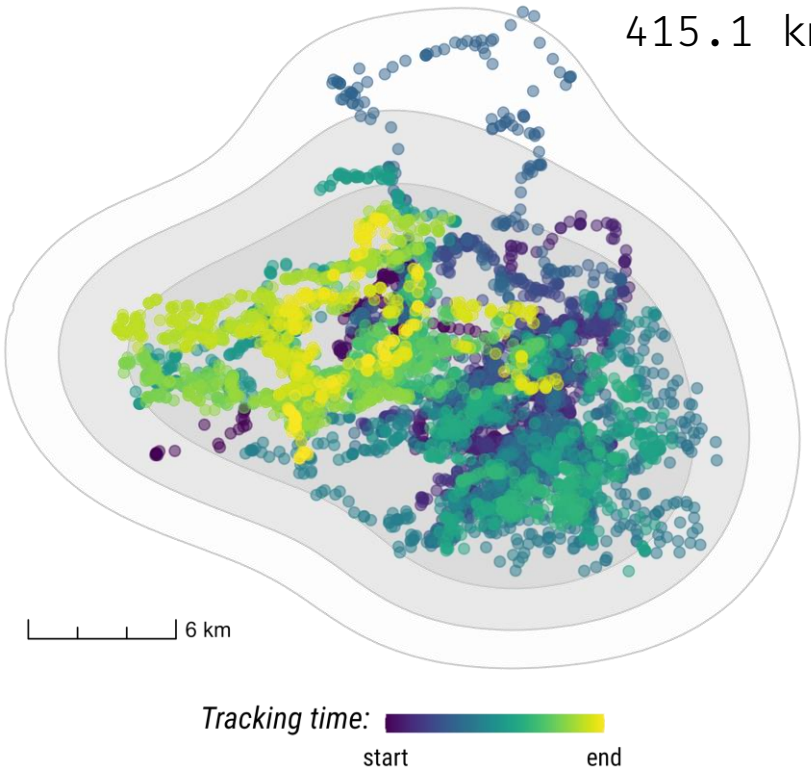


Sampling duration T ≈ 8 months
Sampling interval Δt = 1 hour

**OUF anisotropic process:**

n = 5,766
Narea ≈ 22.3
Nspeed ≈ 2933.9

$\tau_p$ = 11.2 days
$\tau_v$ = 23.4 minutes




**Home range**
415.1 km² (261.0 – 604.4)

**Speed**
12.1 km/day (11.9 – 12.4)

| | Tracking regime | |
|---|---|---|
| Model | Duration | Sampling interval |
| OUF | $\tau_p < T$ | $\Delta t < \tau_v < \tau_p$ |

# Sampling *versus* characteristic scales

| | Tracking regime | |
|---|---|---|
| Model | Duration (T) | Sampling interval (Δt) |
| IID | $\tau_p < T$ | $\tau_v < \tau_p < \Delta t$ |
| BM | $T < \tau_p$ | $\tau_v < \Delta t < \tau_p$ |
| OU | $\tau_p < T$ | $\tau_v < \Delta t < \tau_p$ |
| IOU | $T < \tau_p$ | $\Delta t < \tau_v < \tau_p$ |
| OUF | $\tau_p < T$ | $\Delta t < \tau_v < \tau_p$ |

# RECOMMENDATIONS



*What happens if we truncate the data?*

Sampling duration T ≈ ~~8 months~~ 16 days

Sampling interval Δt = 1 hour

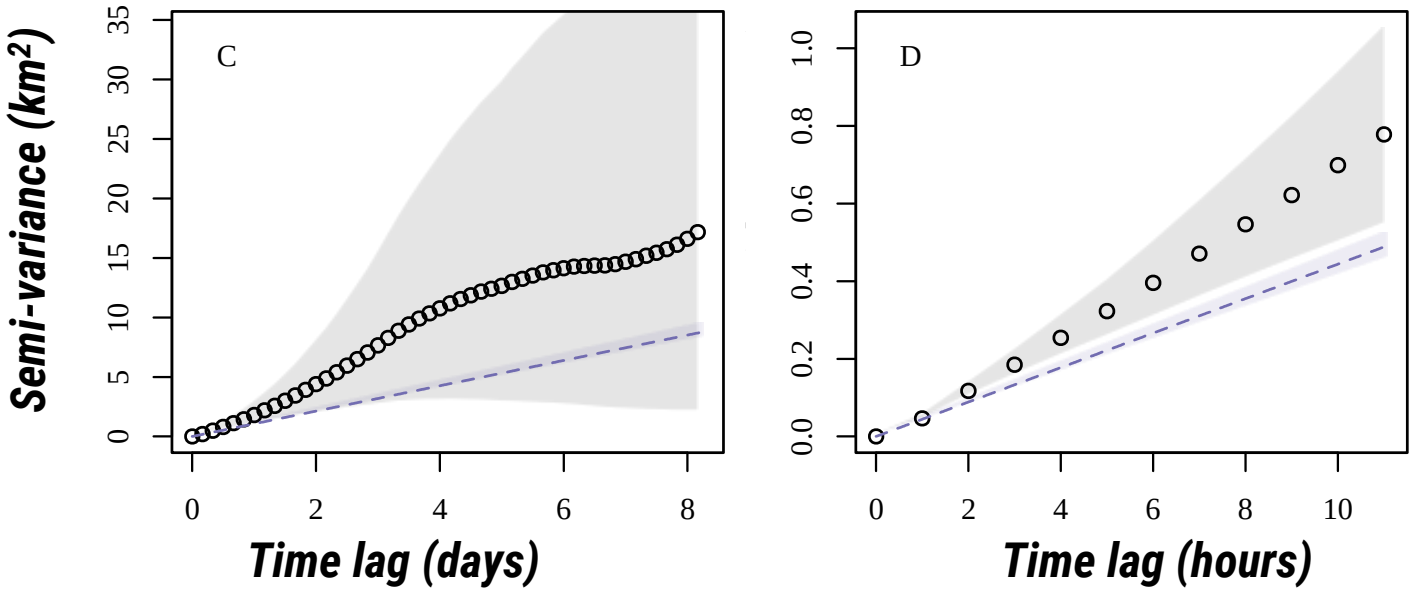
**BM process:**

n = 397

Narea ≈ 0

Nspeed ≈ 0

τ = Inf



# RECOMMENDATIONS



*What happens if we truncate the data?*

Sampling duration T ≈ ~~8 months~~ 16 days

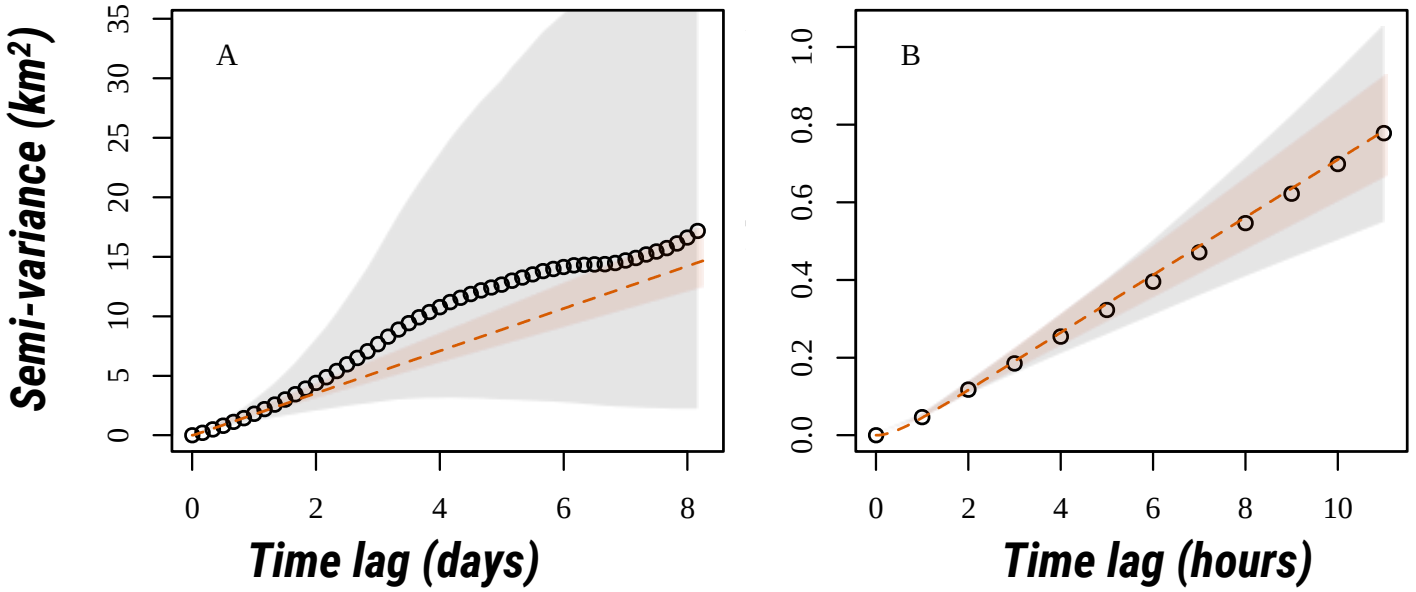
Sampling interval Δt = 1 hour

**IOU process:**

n = 397
Narea ≈ 0
Nspeed ≈ 335.3

$\tau_p$ = Inf
$\tau_v$ = 13.1 minutes

| Model | ΔAICc |
|---|---|
| IOU anisotropic | 0 |
| BM anisotropic | 199.5 |



# RECOMMENDATIONS



*What happens if we also thin the data?*

Sampling duration T ≈ ~~8 months~~ 16 days

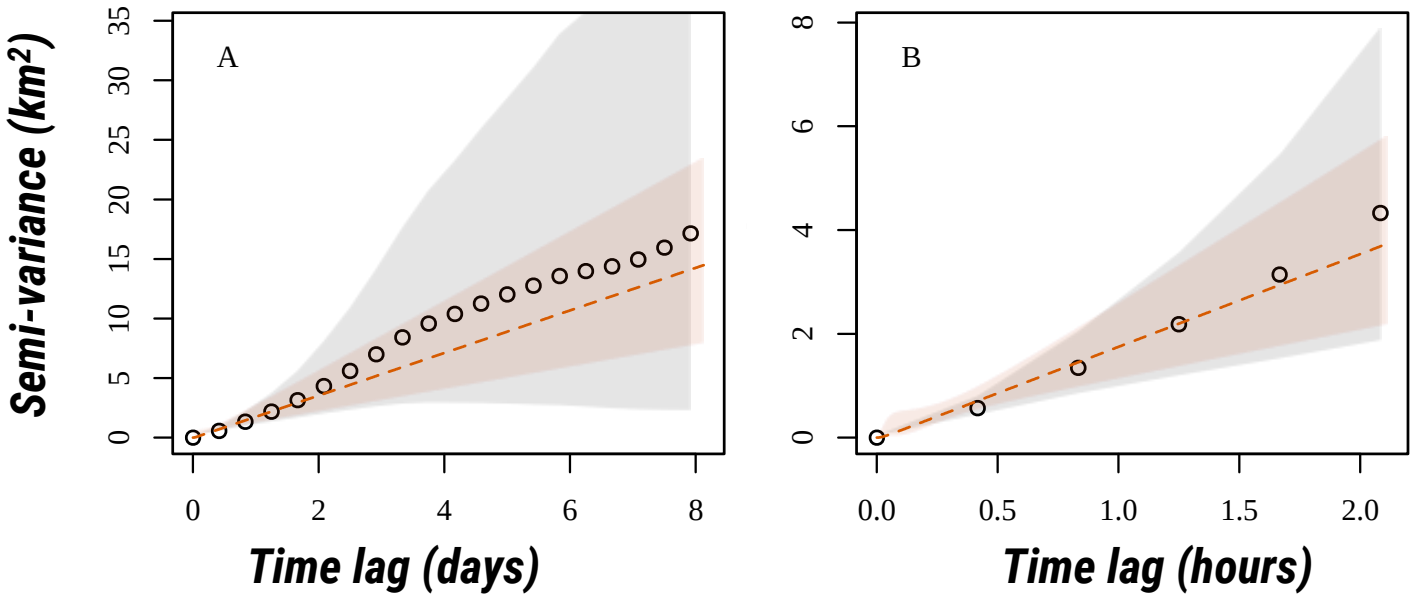
Sampling interval Δt = ~~1 hour~~ 19 hours

**IOU process:**

n = 40
Narea ≈ 0
Nspeed ≈ 2.86

$\tau_p$ = Inf
$\tau_v$ = 10.5 minutes

| Model | ΔAICc |
|---|---|
| BM anisotropic | 0 |
| IOU anisotropic | 1.269 |



# RECOMMENDATIONS



*What happens if we also thin the data?*

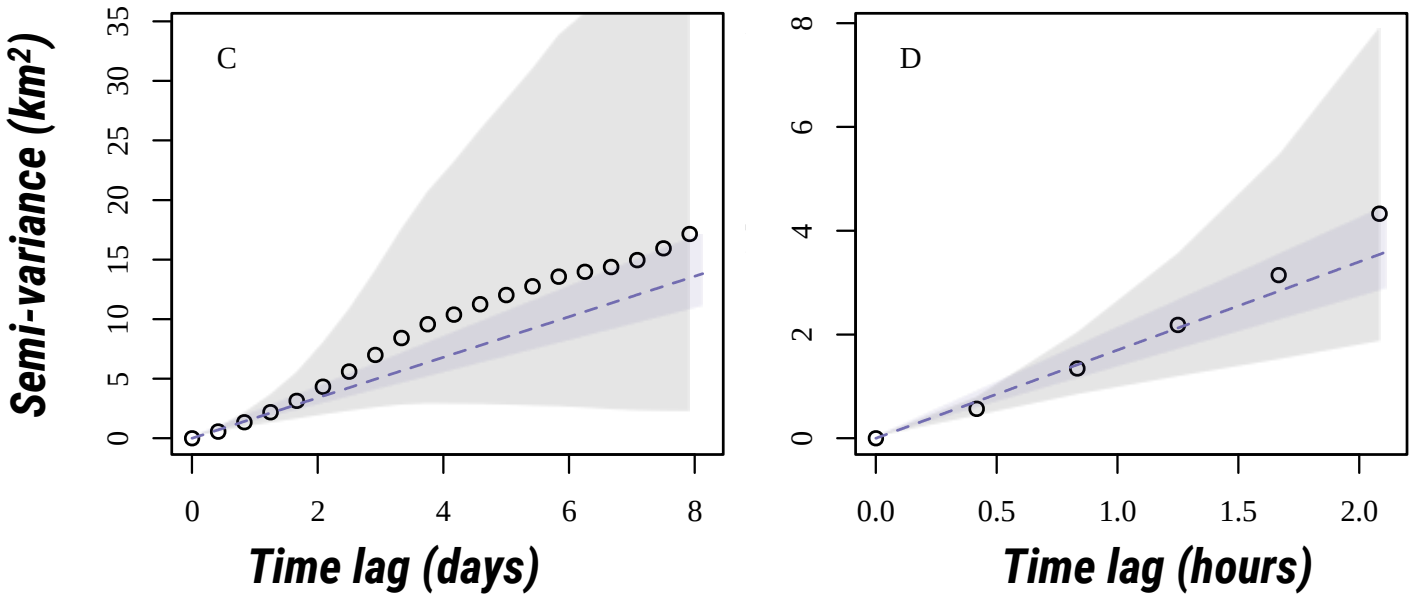
Sampling duration T ≈ ~~8 months~~ 16 days
Sampling interval Δt = ~~1 hour~~ 19 hours

**BM process:**

n = 397
Narea ≈ 0
Nspeed ≈ 0

τ = Inf

| Model | ΔAICc |
|---|---|
| BM anisotropic | 0 |
| IOU anisotropic | 1.269 |



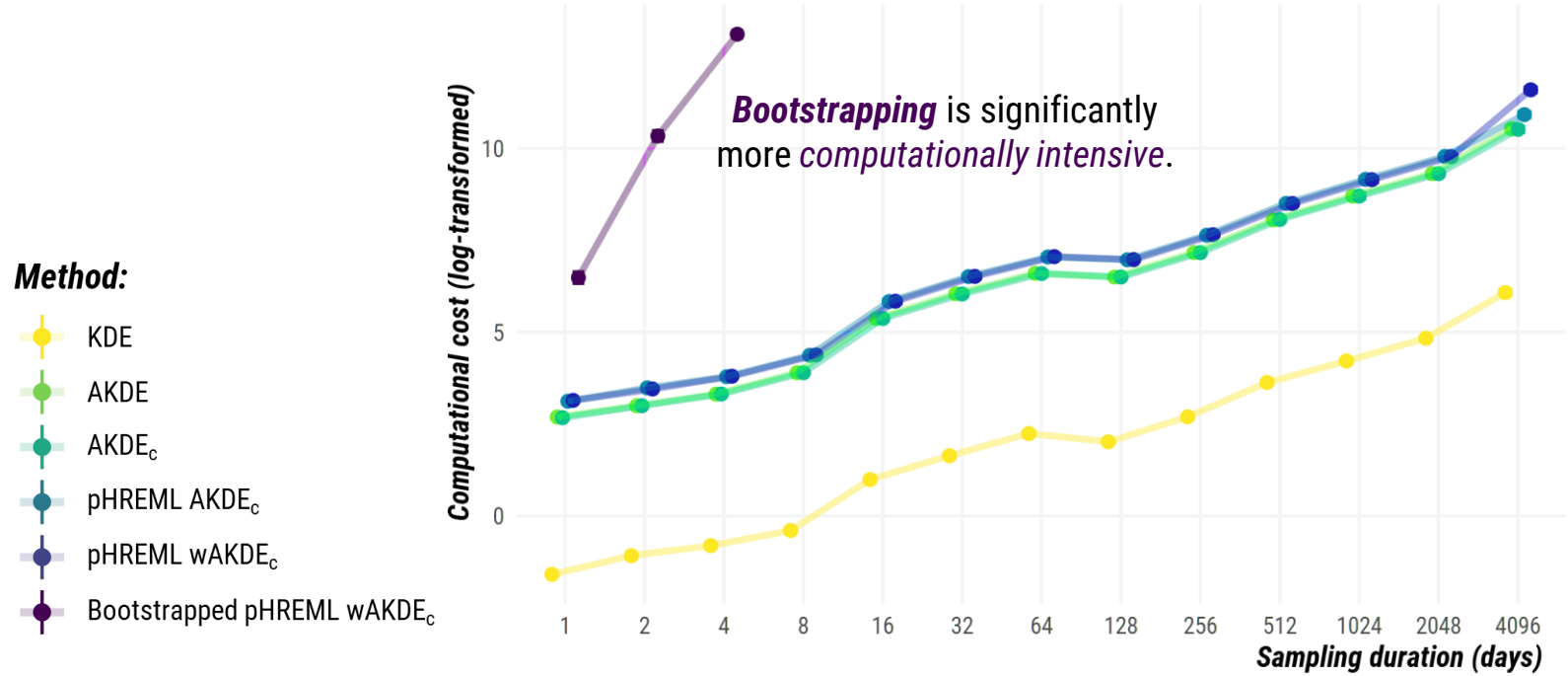

**So we couldn't use this data for home range *or* speed estimation.**

# RECOMMENDATIONS



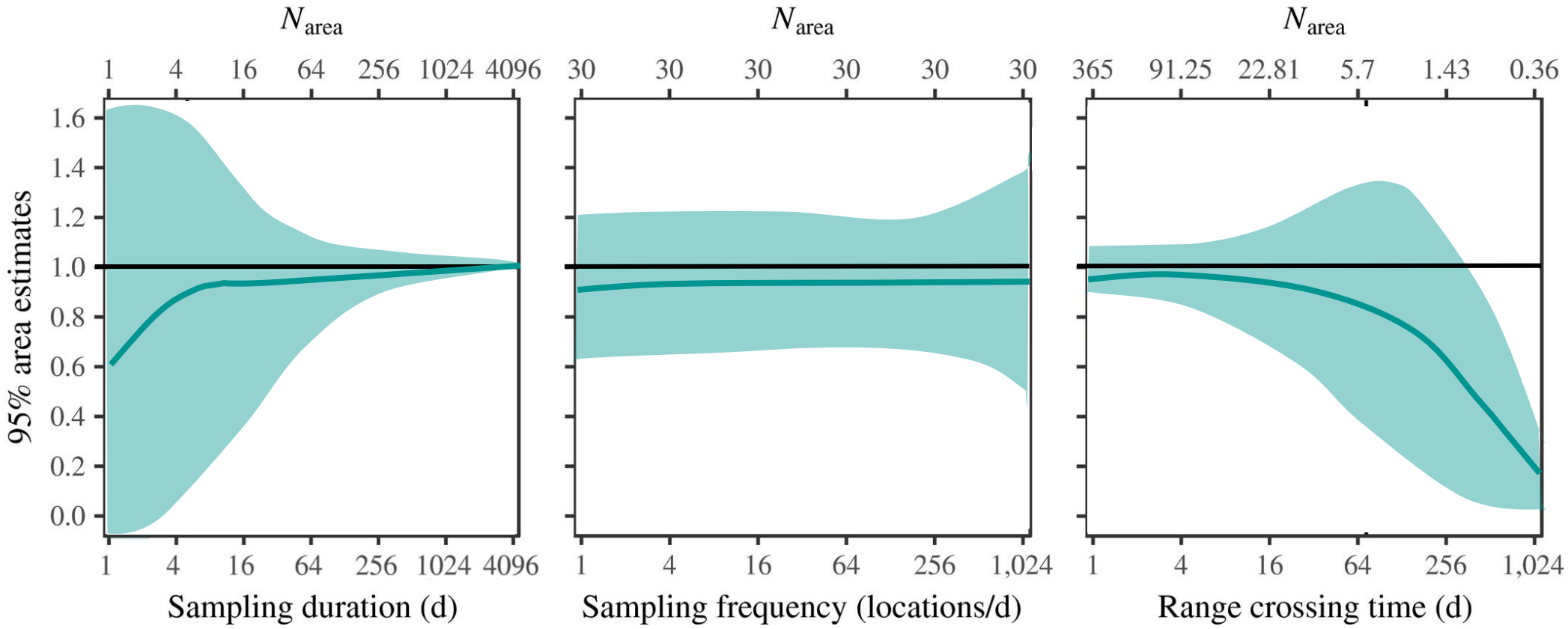
## 1. Home ranges — Autocorrelation Kernel Density Estimator (AKDE)

Given a relative target bias of ≈ 5%,
Minimum $N_{area}$ for ML is ≈ 20;
Minimum $N_{area}$ for pHREML is ≈ 4.5;
Minimum $N_{area}$ for bootstrapped pHREML is ≈ 2.7.



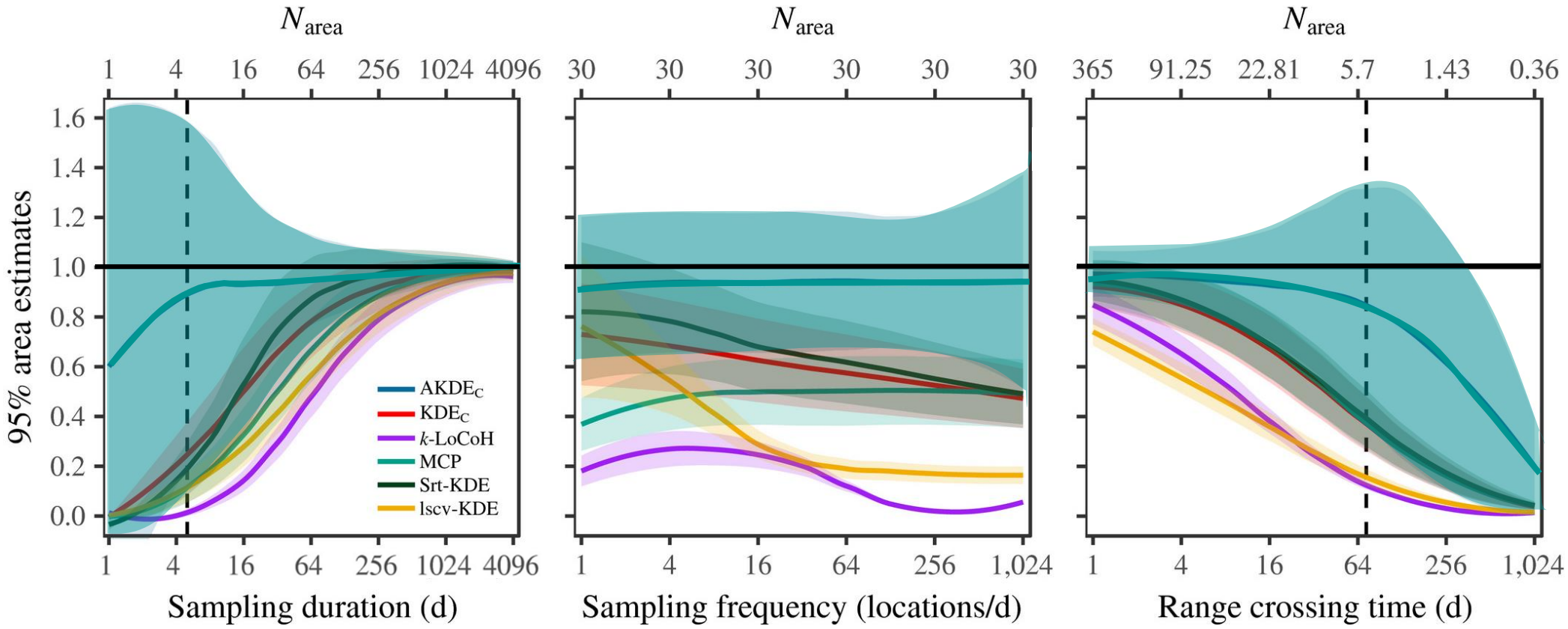
# RECOMMENDATIONS





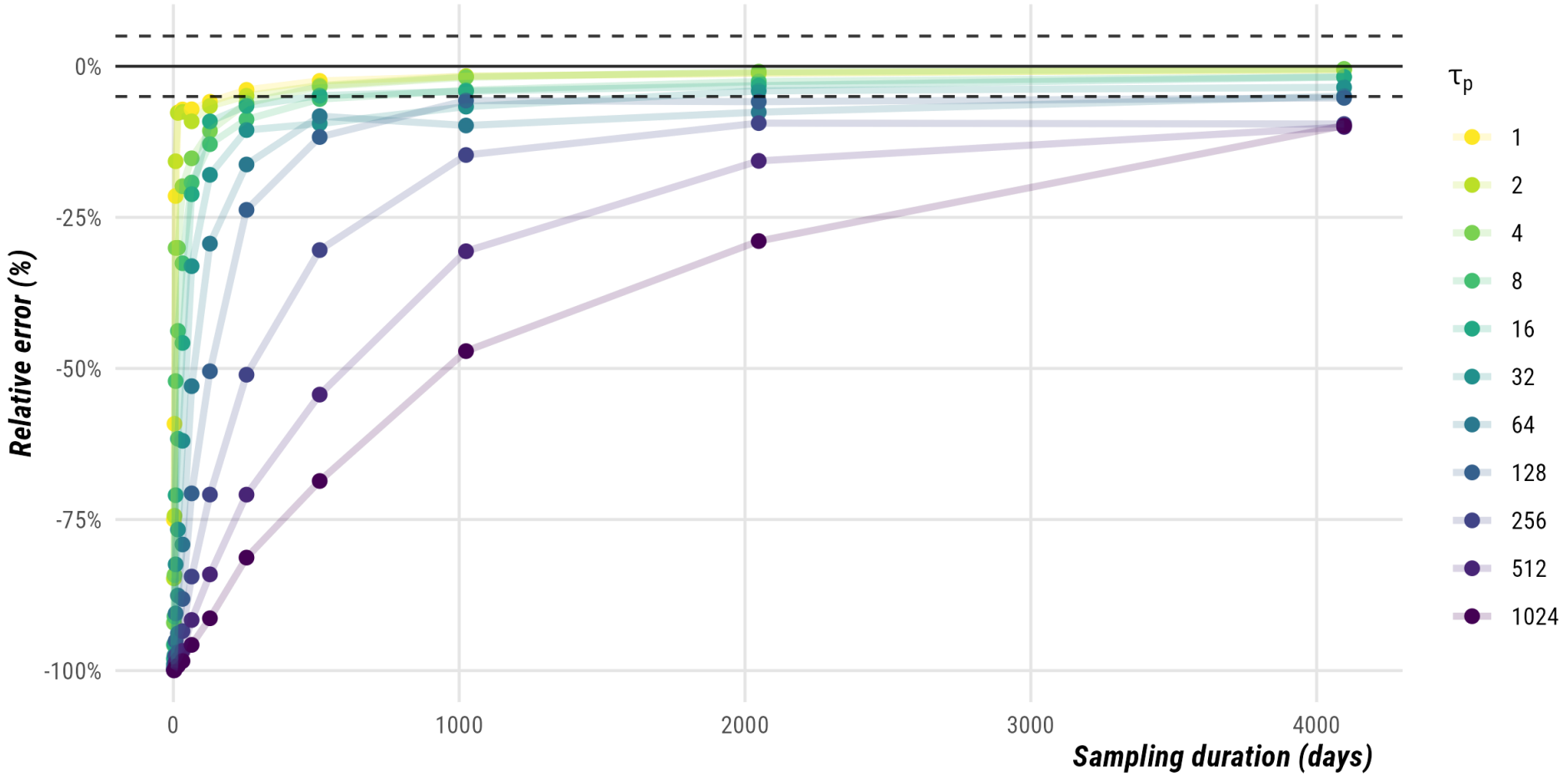
# RECOMMENDATIONS





# RECOMMENDATIONS





# RECOMMENDATIONS



## 1. Home ranges — Autocorrelation Kernel Density Estimator (AKDE)

Given a relative target bias of ≈ 5%,
Minimum $N_{area}$ for ML is ≈ 20;
Minimum $N_{area}$ for pHREML is ≈ 4.5;
Minimum $N_{area}$ for bootstrapped pHREML is ≈ 2.7.

If you have a tracking dataset with an **irregular sampling interval**.

**— Applying wAKDE;**
**— Subsampling data to exclude closest locations in time.**

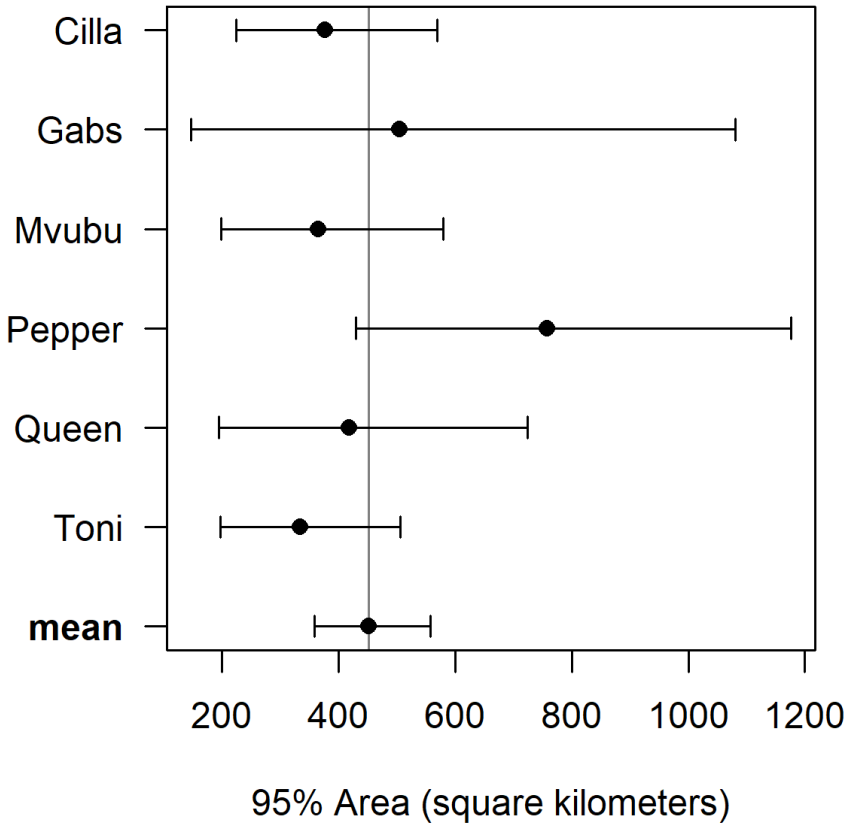
## General recommendation:

Adjust your sampling duration to ensure data is sufficient to detect $\tau_p$.

# RECOMMENDATIONS



## 1. Home ranges — Autocorrelation Kernel Density Estimator (AKDE)



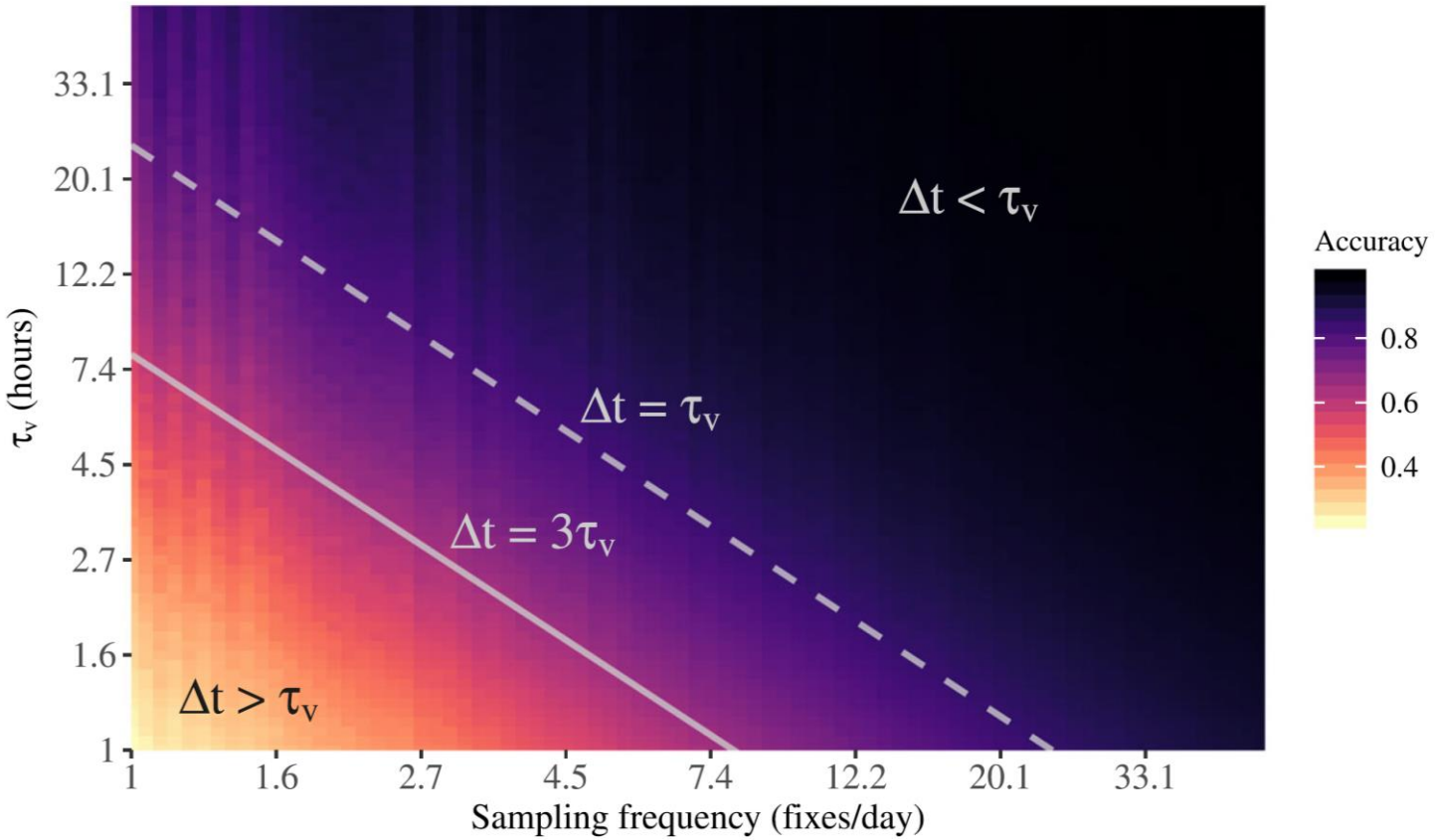

For home range ***comparisons***:

**> 2–3 observed home-range crossings ($\tau_p$)**
**> 2–3 representative individuals.**

# RECOMMENDATIONS



## 1. Speed & distance – Continuous-time speed and distance (CTSD)



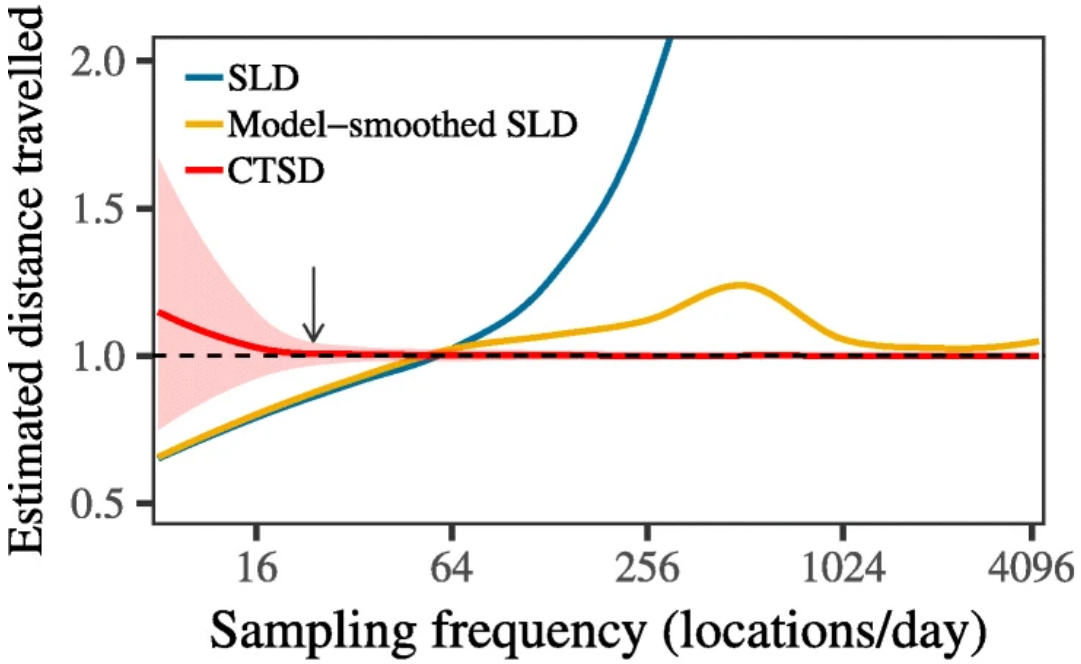

If **$\Delta t > 3\tau_v$**, no statistically significant signature of the animal's velocity will remain in the location data.

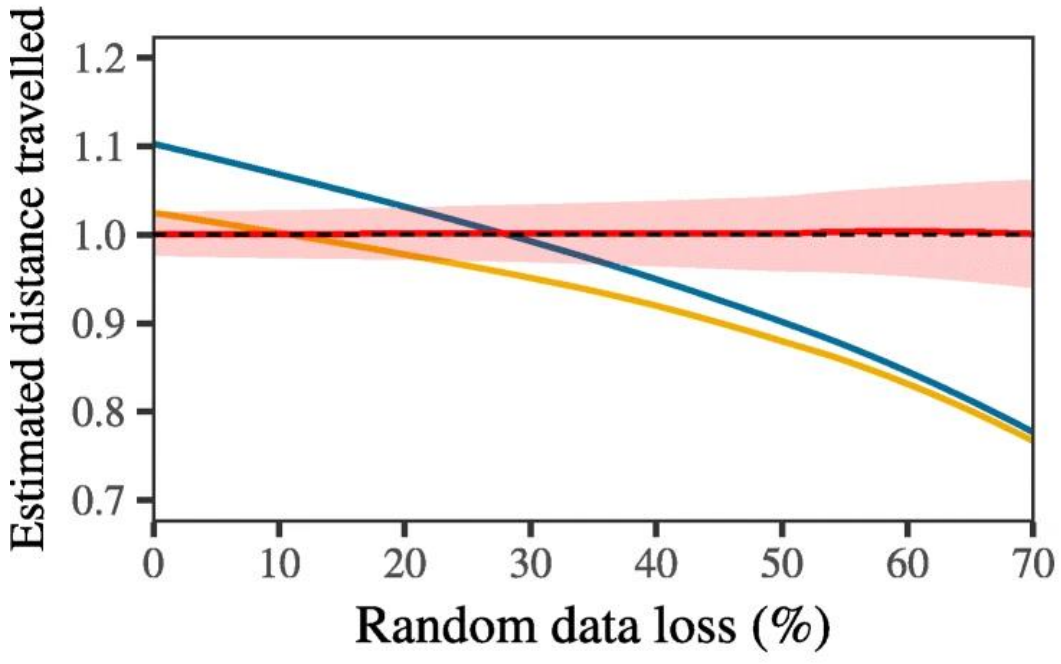
If **$3\tau_v > \Delta t > \tau_v$**, there will be some positive bias in the CTSD estimates ($\tau_v$ can not be accurately estimated).

# RECOMMENDATIONS



## 1. Speed & distance — Continuous-time speed and distance (CTSD)

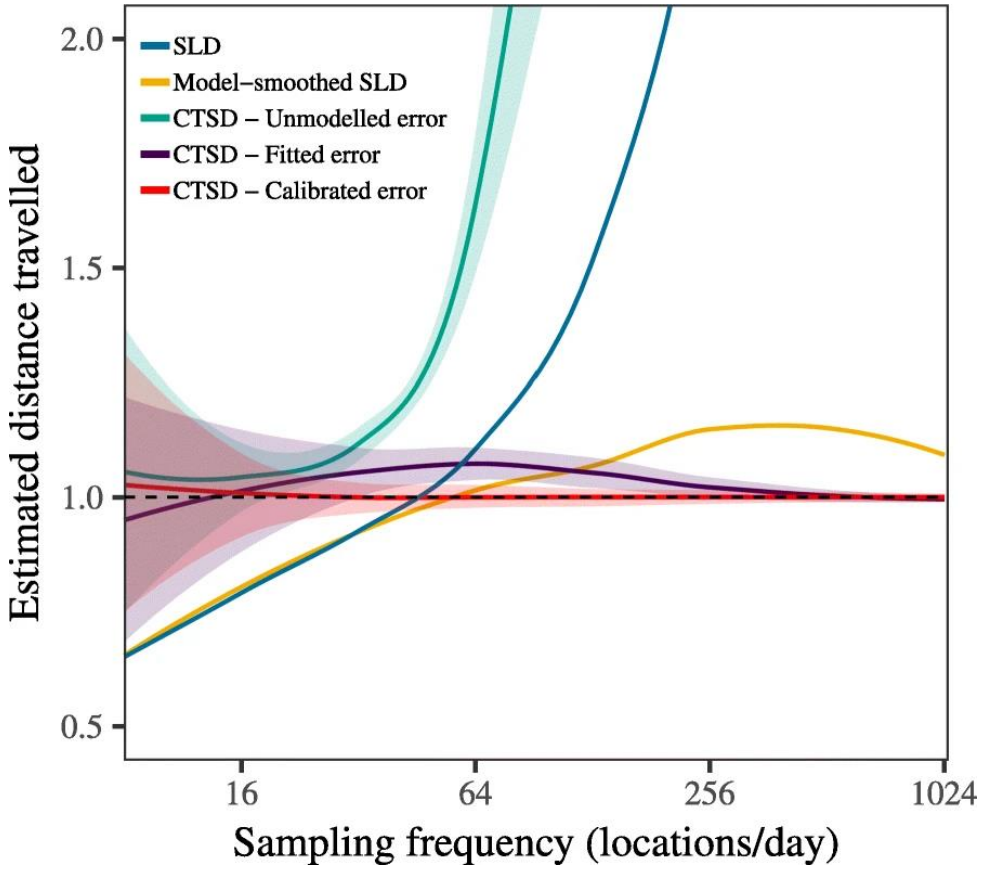


# RECOMMENDATIONS



## 1. Speed & distance – Continuous-time speed and distance (CTSD) ✓




We also need to account for ***measurement error*** (by collecting calibration data, or using pre-calibrated tracking devices).

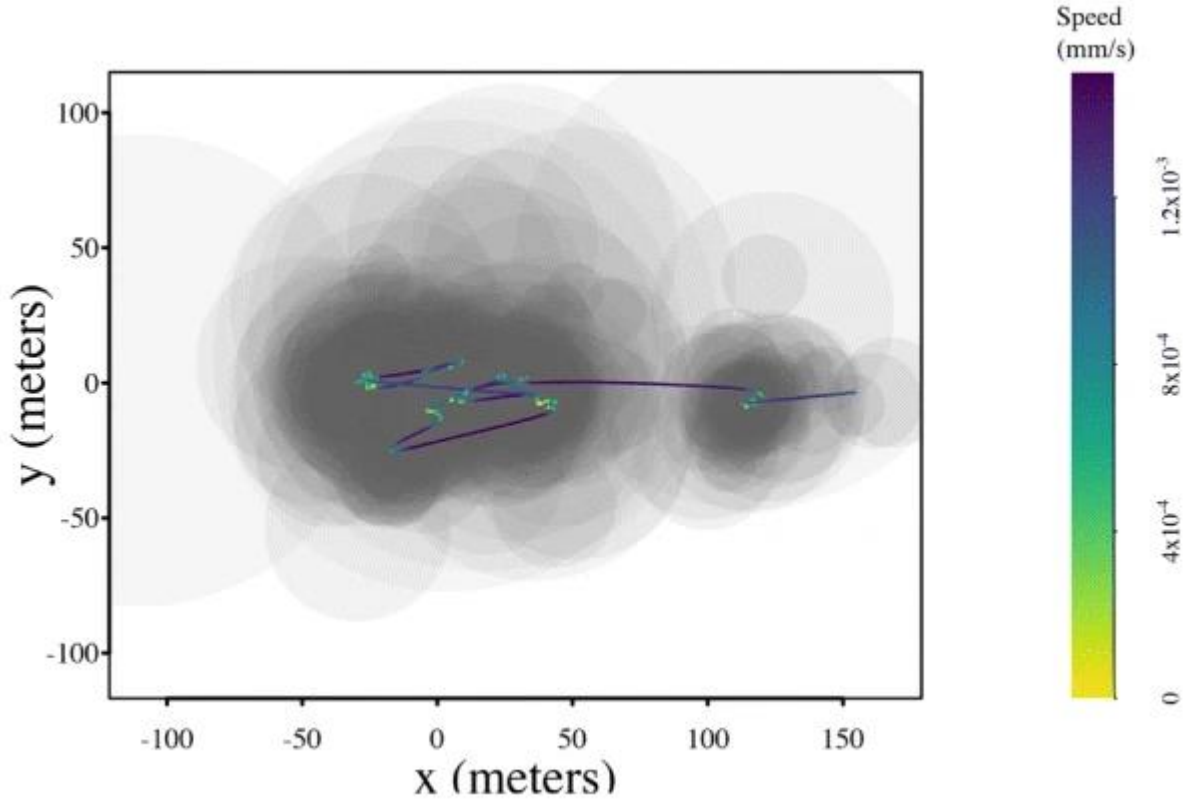
# RECOMMENDATIONS



## 1. Speed & distance — Continuous-time speed and distance (CTSD) ✓


**WOOD TURTLE**

GLYPTEMYS INSCULPTA



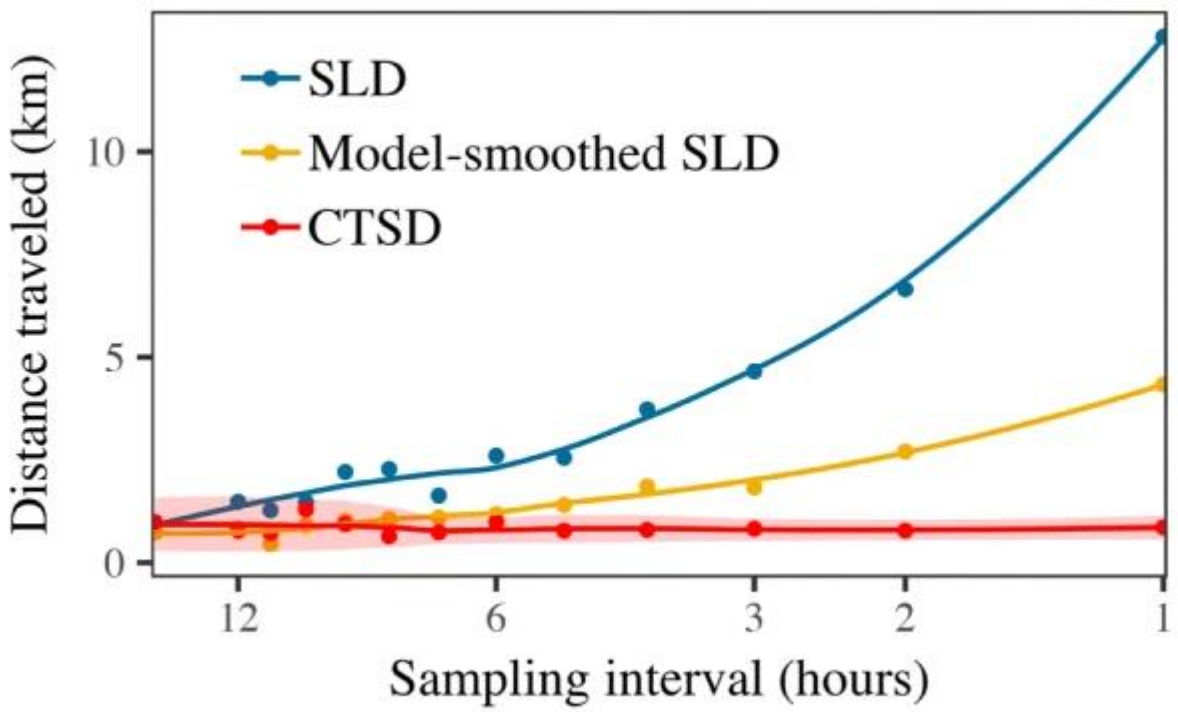
# RECOMMENDATIONS



## 1. Speed & distance — Continuous-time speed and distance (CTSD)


WOOD TURTLE

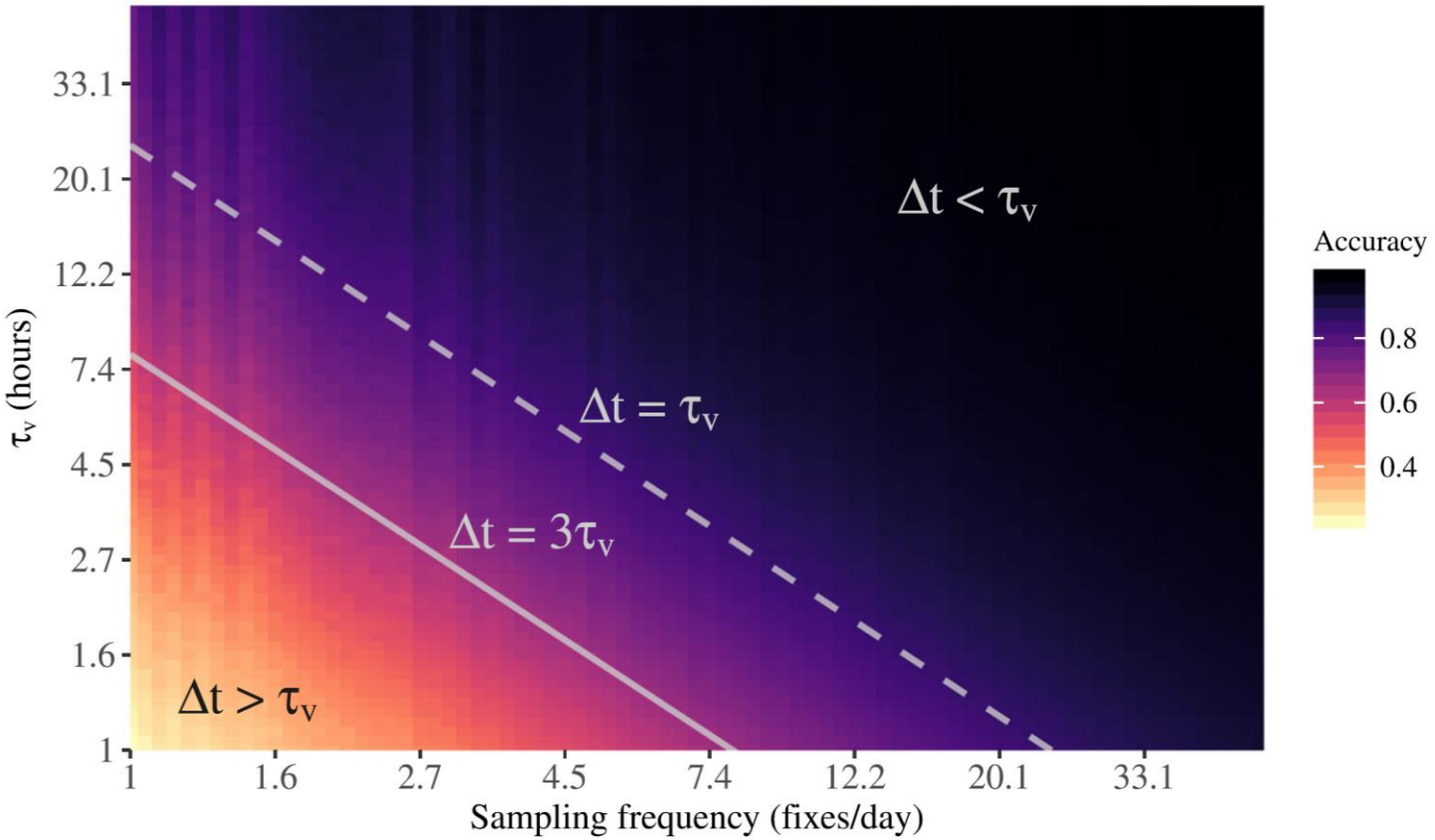
GLYPTEMYS INSCULPTA



# RECOMMENDATIONS



## 1. Speed & distance – Continuous-time speed and distance (CTSD)




**General recommendation:**

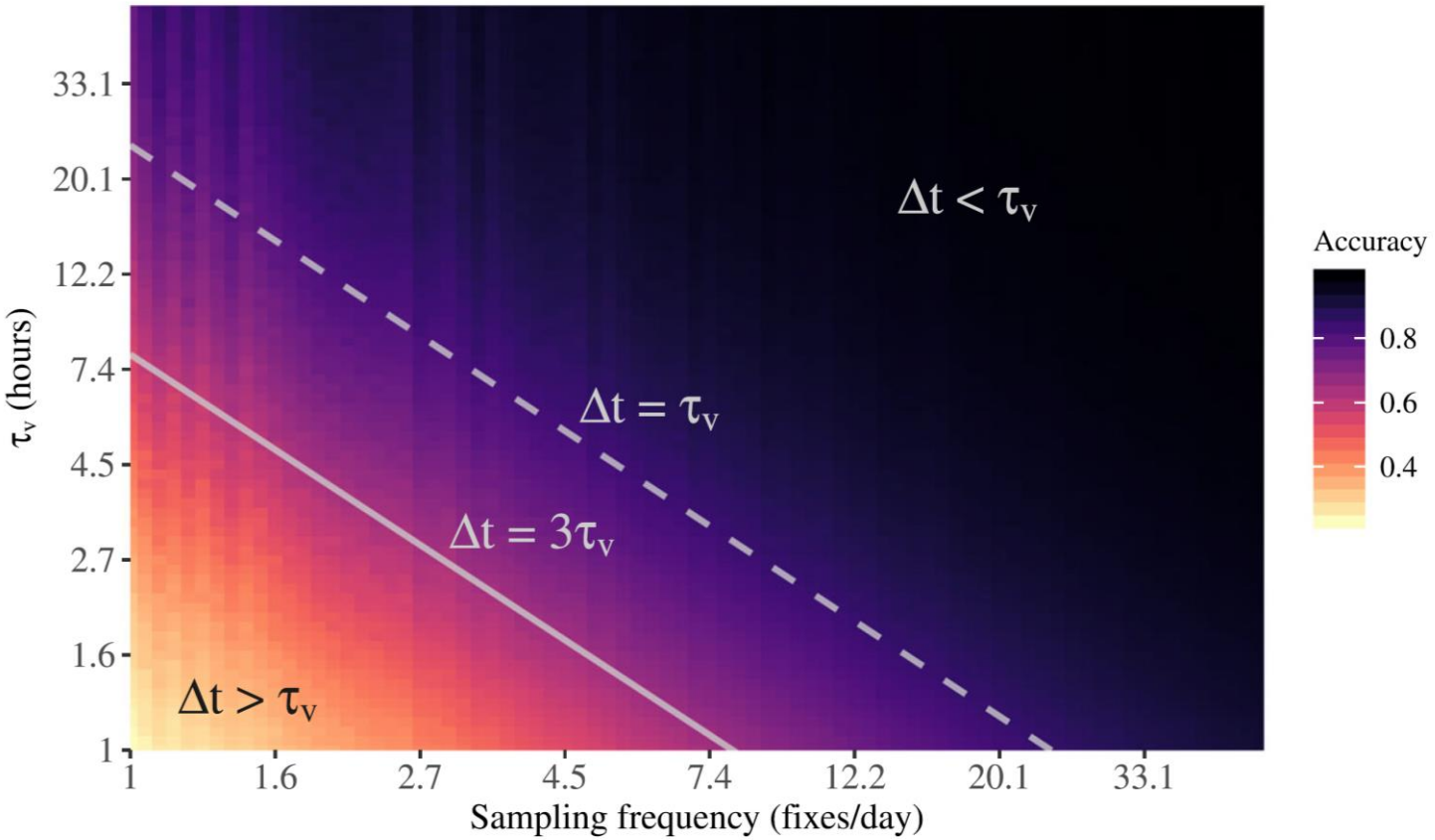
Adjust your sampling interval to ensure data is of sufficient resolution to detect $\tau_v$.

$$\Delta t \leq 3\tau_v$$

# RECOMMENDATIONS



## 1. Speed & distance – Continuous-time speed and distance (CTSD)




**General recommendation:**

Adjust your sampling interval to ensure data is of sufficient resolution to detect $\tau_v$.

$$\Delta t \leq 3\tau_v$$

# RECOMMENDATIONS



**Get informed:**
Collect pilot data, look at published studies & datasets.

**Make smart decisions:**
Consider spatiotemporal scales relative to questions.

**Keep it simple:**
Use *even sampling rates* if possible. Don't get too clever!

**Try before you buy:**
Check it first with simulated data.

**Divide and conquer:**
Use different individuals to answer different questions.

# RECOMMENDATIONS



**Get informed:**
Collect pilot data, look at published studies & datasets.

**Make smart decisions:**
Consider spatiotemporal scales relative to questions.

**Keep it simple:**
Use *even sampling rates* if possible. Don't get too clever!

**Try before you buy:**
Check it first with simulated data.

**Divide and conquer:**
Use different individuals to answer different questions.

move design

# RECOMMENDATIONS



## If you must sample *unevenly*…

**Keep it simple:**
Use no more than two different sampling rates.

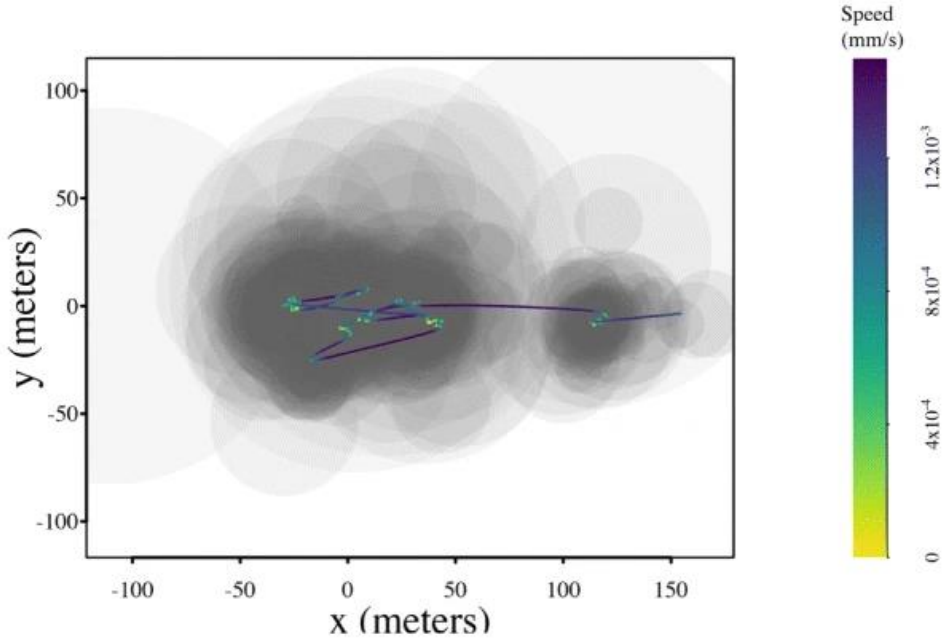
**Try before you buy:**
Simulate with uneven sampling first. Still works?

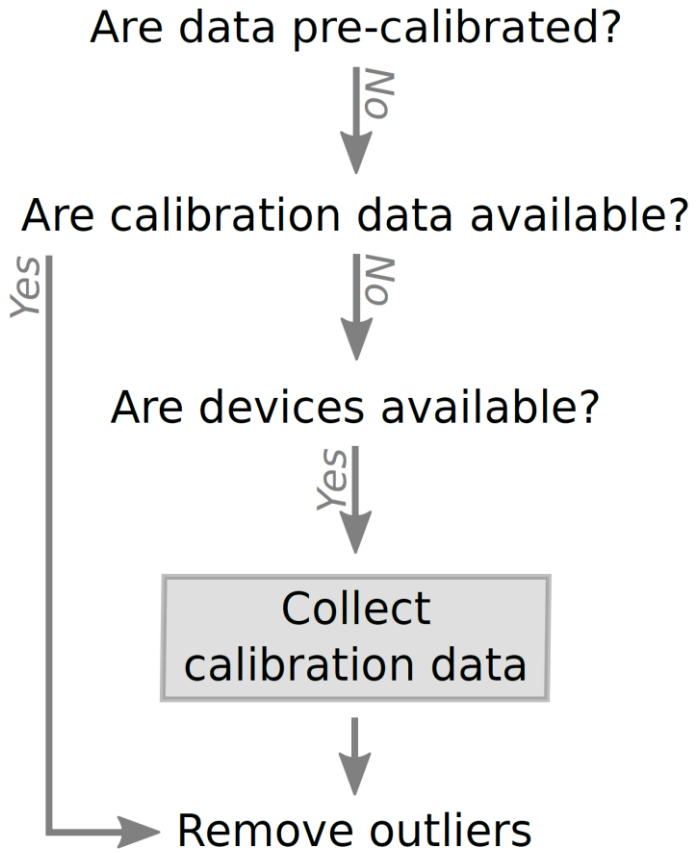
**Mind your math:**
Use *sampling rates* that are integer multiples.

**Be careful:**
Check for artifacts introduced by uneven sampling.

**Be realistic:**
No, you can’t have it all!

# Collect error calibration data!




Fleming *et al.* (2020)
DOI: 10.1101/2020.06.12.130195


R Studio®
+
move design