HELMHOLTZ-ZENTRUM DRESDEN-ROSSENDORF

CASUS CENTER FOR ADVANCED SYSTEMS UNDERSTANDING

AniMove 2024, June 17th to 28th

Optimizing animal tracking projects

Using the 'ctmm' R package

Inês Silva

i.simoes-silva@hzdr.de

TECHNISCHE UNIVERSITÄT DRESDEN

CBG Max Planck Institute of Molecular Cell Biology and Genetics

UFZ HELMHOLTZ Centre for Environmental Research

Uniwersytet Wrocławski

SPONSORED BY THE

Federal Ministry of Education and Research

STAATSMINISTERIUM FÜR WISSENSCHAFT KULTUR UND TOURISMUS

Freistaat SACHSEN

What can I do with the data I've already collected?"

What are your priorities?

I want to know everything!

i.simoes-silva@hzdr.de

Introduction


Define research questions


Identify spatiotemporal scales


Choose sampling design


Collect animal tracking data


Analyze data, mitigate biases


Assess conclusions

i.simoes-silva@hzdr.de

Introduction


Define research questions


Identify spatiotemporal scales


Choose sampling design


Collect animal tracking data


Analyze data, mitigate biases


Assess conclusions

Cost of devices (& data transfer), challenges during deployment, and technological limitations,

can all constrain study design.

"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."

i.simoes-silva@hzdr.de

Introduction

Fine-scale processes:

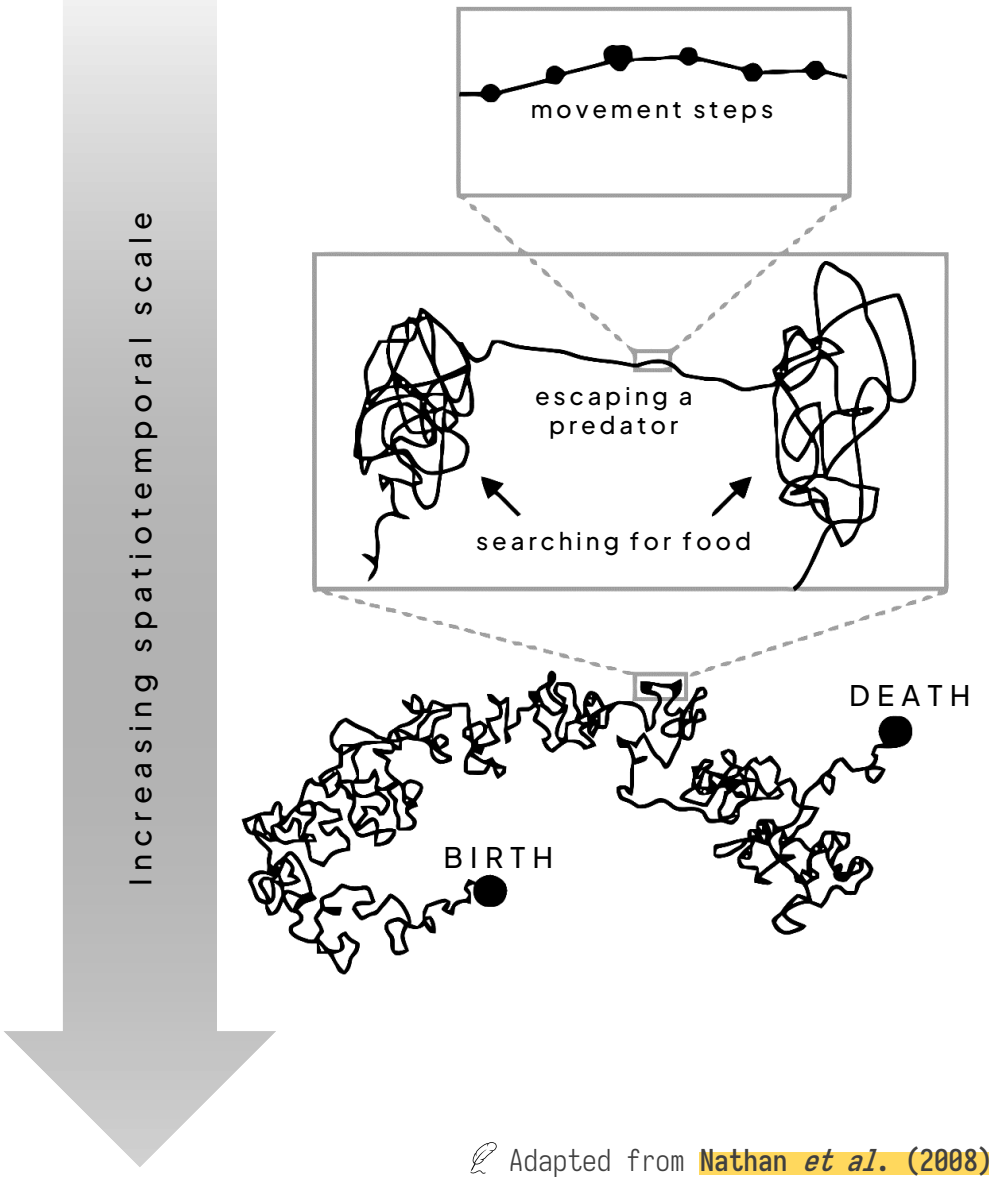
SPEED/DISTANCE — capture how far animals travel (and rate at which these distances are covered).

– To link behavior and energetics,
– As indicators of anthropogenic disturbance.

Large-scale processes:

HOME RANGE — capture the area repeatedly used throughout an animal's **lifetime**.

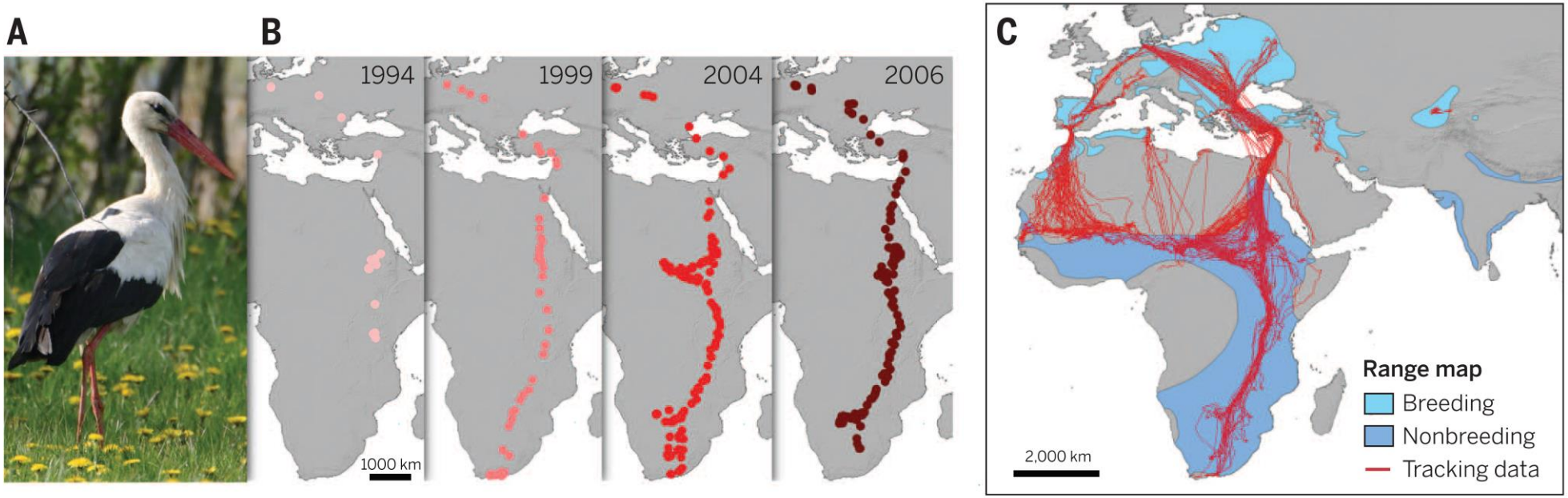
– For protected area delineation,
– To reduce human-wildlife conflict,
– To control the spread of infectious diseases.



Adapted from **Nathan *et al.* (2008)**

i.simoes-silva@hzdr.de

Introduction

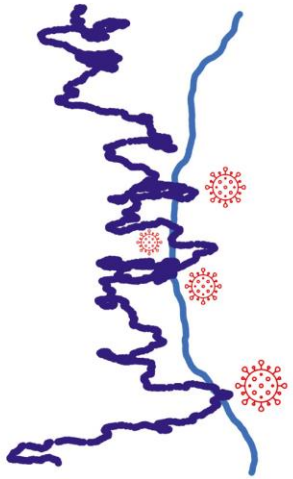


Adapted from Kays et al. (2015)

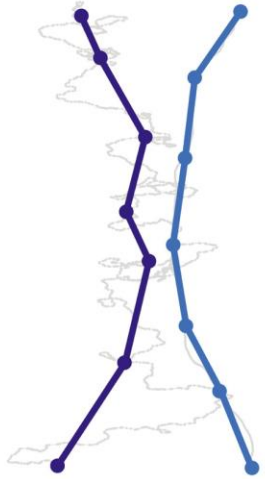
i.simoes-silva@hzdr.de

Nathan *et al.* (2022)

Higher resolution
(5 s intervals)



Lower resolution
(30 min intervals)



i.simoes-silva@hzdr.de

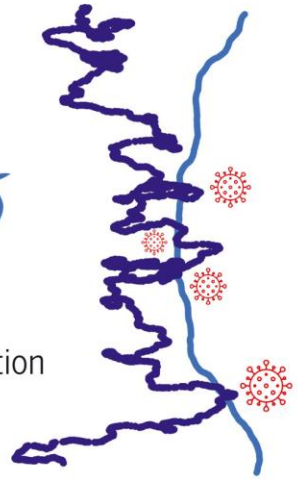
Nathan *et al.* (2022)

Higher resolution
(5 s intervals)

Exploration

Bold > Shy

Multiple interaction hotspots

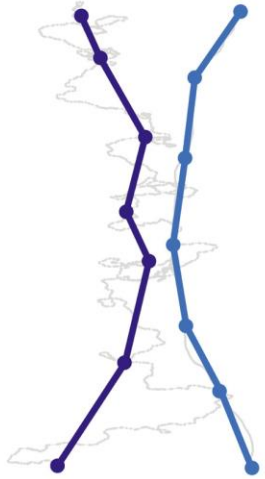


Lower resolution
(30 min intervals)

Exploration

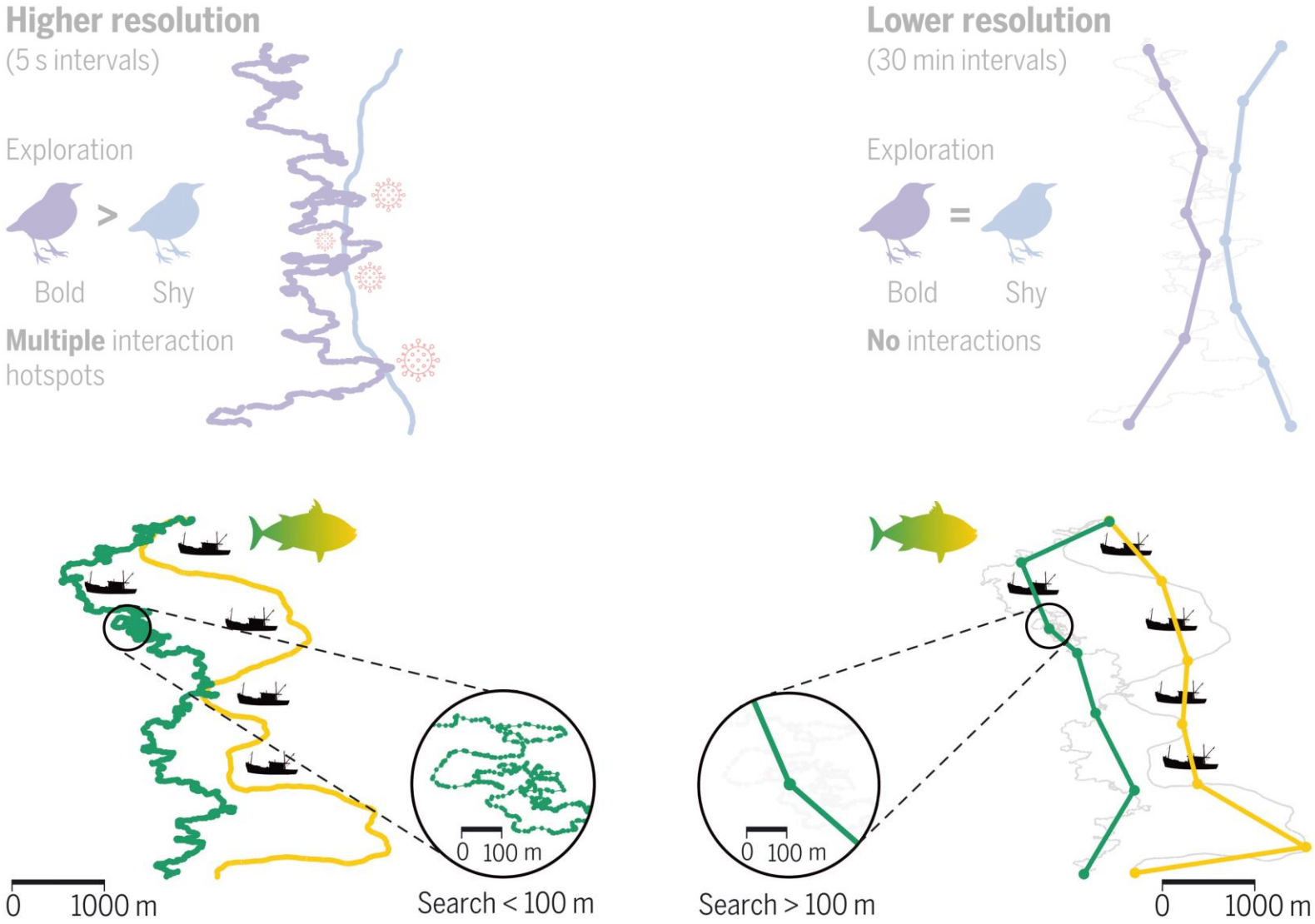
Bold = Shy

No interactions



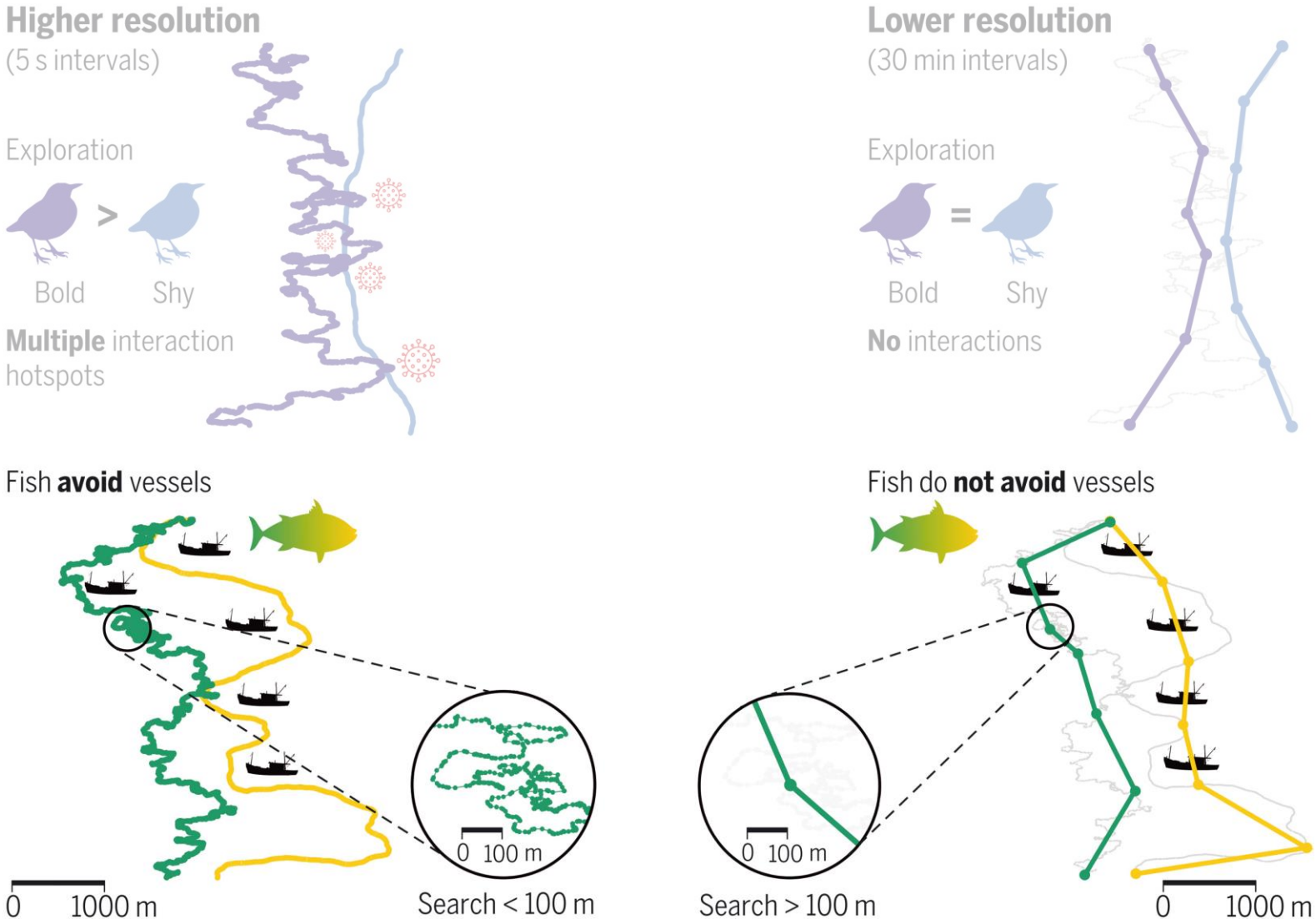
i.simoes-silva@hzdr.de

Nathan *et al.* (2022)



i.simoes-silva@hzdr.de

Nathan *et al.* (2022)

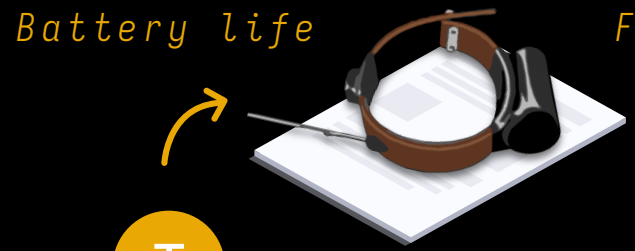


i.simoes-silva@hzdr.de

Introduction

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING

- Define research questions
- Identify spatiotemporal scales
- Choose sampling design
- Collect animal tracking data
- Analyze data, mitigate biases
- Assess conclusions



SAMPLING DURATION
How long is an animal tracked for?

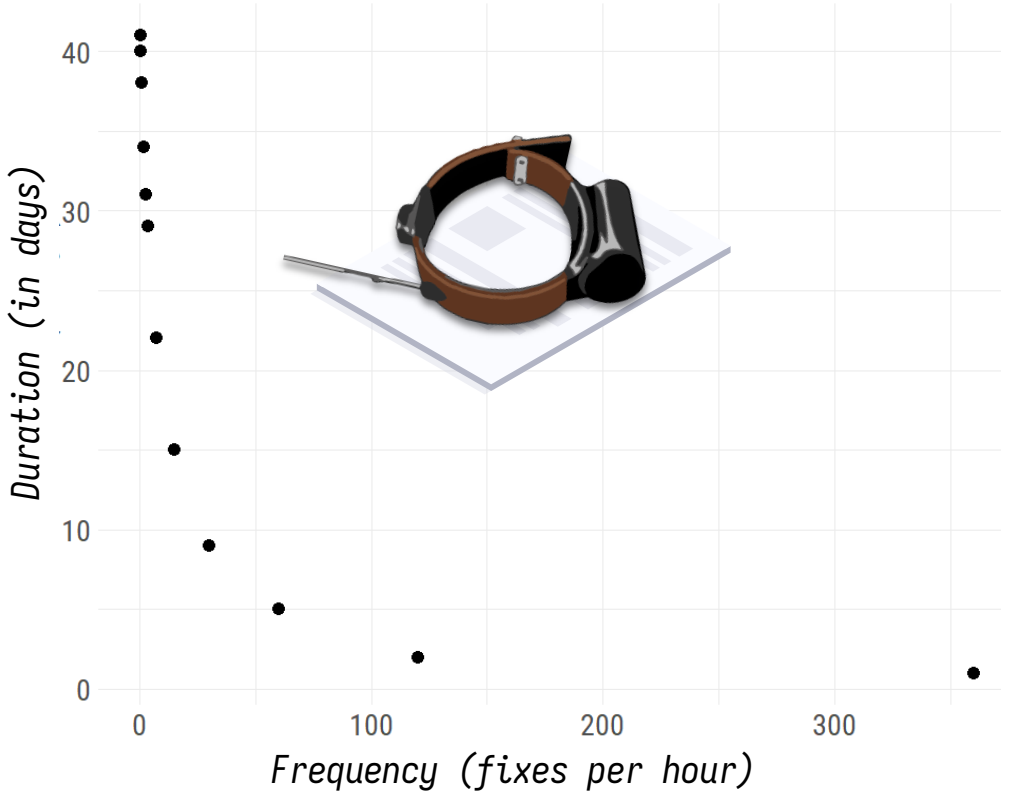
SAMPLING FREQUENCY
How frequently are locations collected?

i.simoes-silva@hzdr.de

Introduction

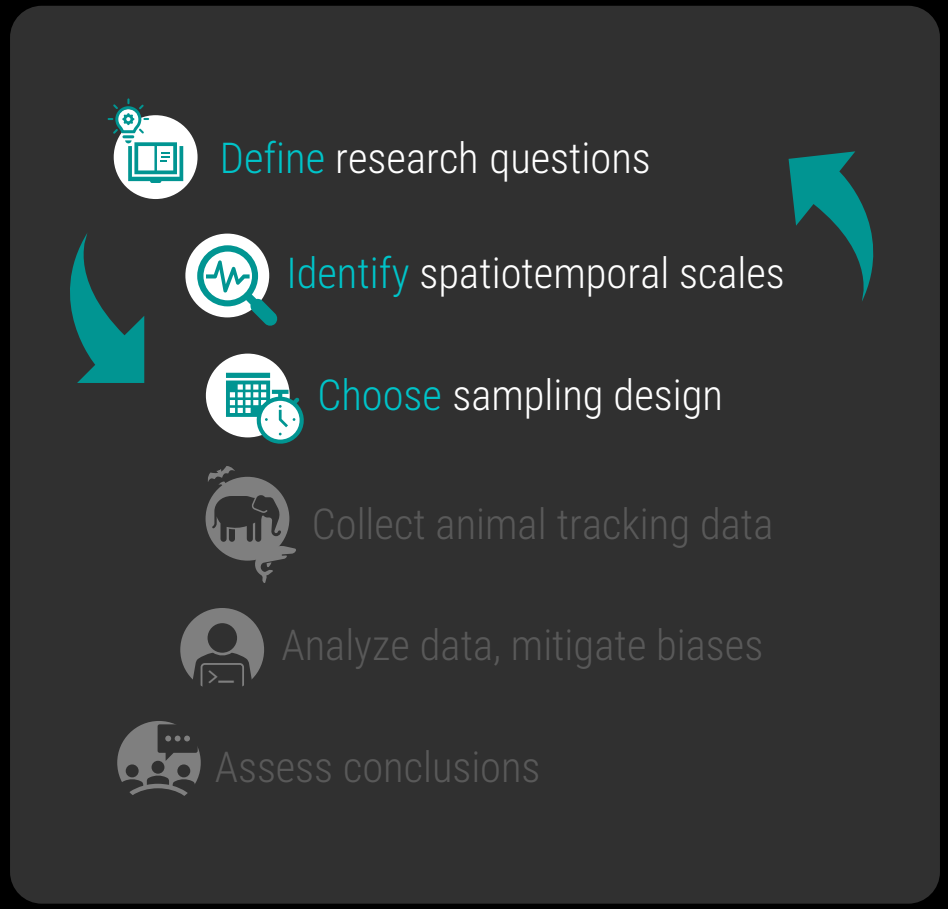
Trade-off between long battery life and high resolution of GPS devices.

Choosing a higher fix rate leads to lower battery life.



i.simoes-silva@hzdr.de

Introduction



1.

Animal movement paths are realizations of **continuous stochastic processes**,

2.

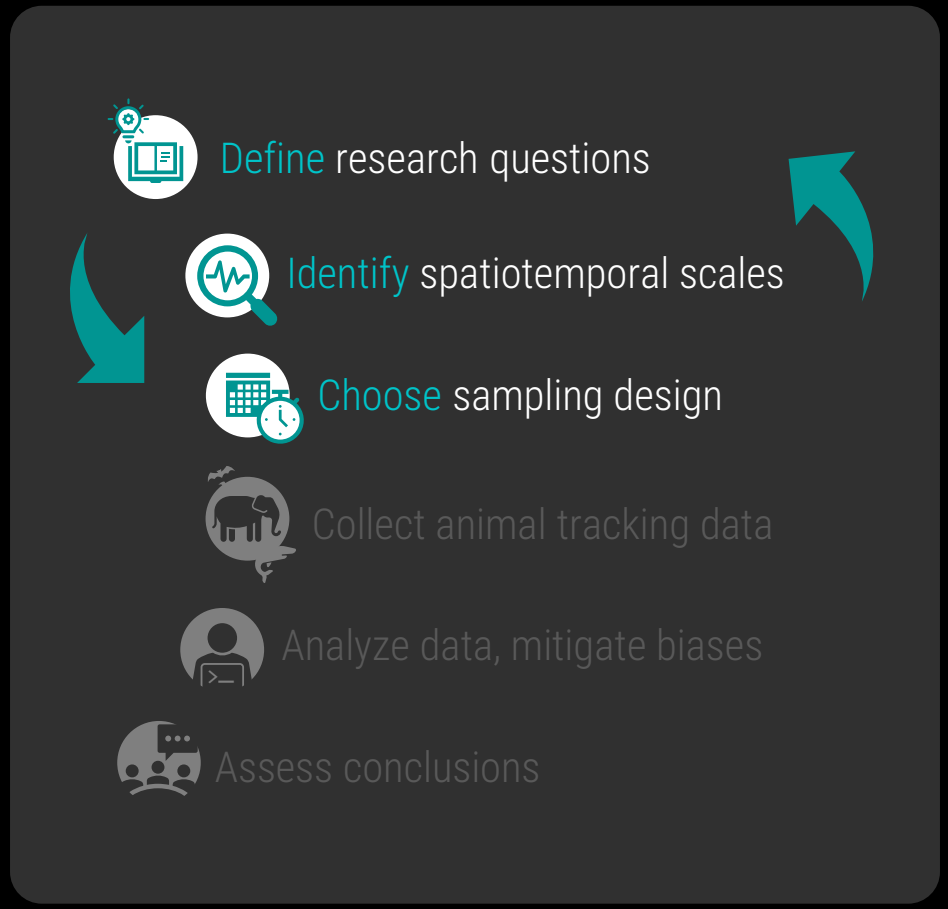
Summarize behavior using **characteristic timescales**,

τ_p — *Position autocorrelation timescale*

τ_v — *Velocity autocorrelation timescale*

i.simoes-silva@hzdr.de

Introduction



1.

Animal movement paths are realizations of **continuous stochastic processes**,

2.

Summarize behavior using **characteristic timescales**,

τ_p — *Position autocorrelation timescale*

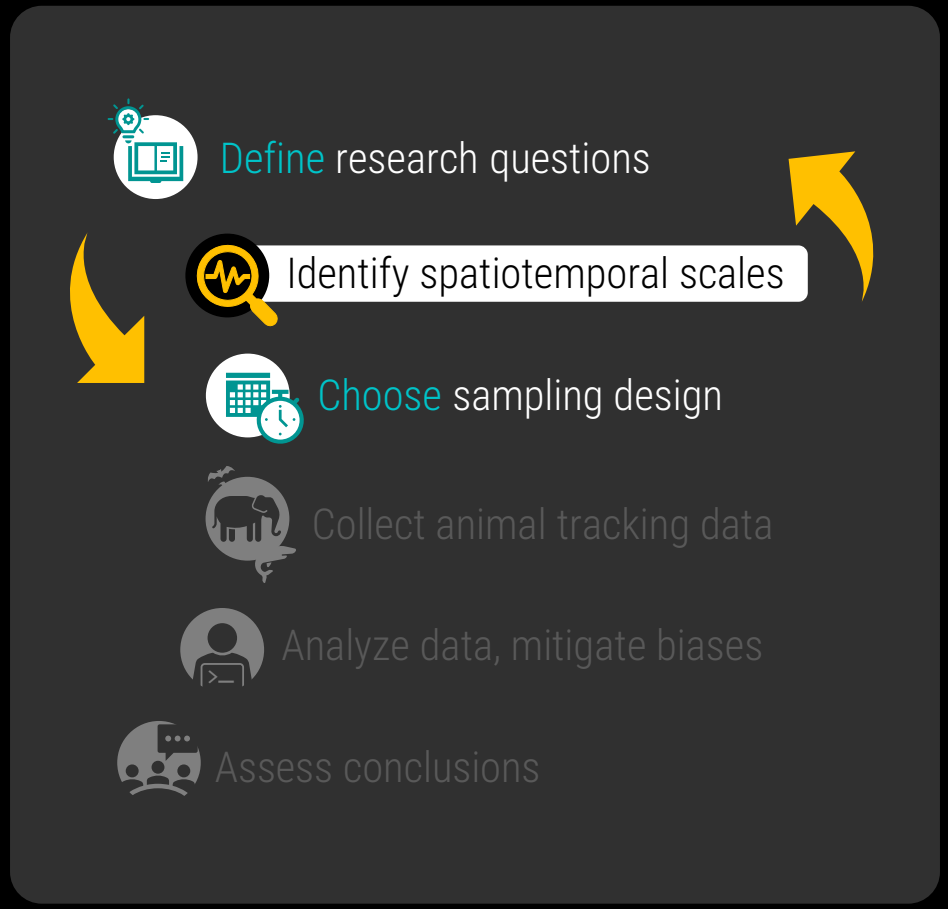
τ_v — *Velocity autocorrelation timescale*

3.

These **timescales** impose **constraints on sampling design** that *must* be met for sufficiently **large (*effective*) sample sizes**.

i.simoes-silva@hzdr.de

Introduction



1.

Animal movement paths are realizations of **continuous stochastic processes**,

2.

Summarize behavior using **characteristic timescales**,

τ_p — *Position autocorrelation timescale*

τ_v — *Velocity autocorrelation timescale*

3.

These **timescales** impose **constraints on sampling design** that *must* be met for sufficiently **large (*effective*) sample sizes**.

i.simoes-silva@hzdr.de

Introduction

‘movedesign’
Silva *et al.* (2023)

Objectives:
Develop a systematic approach, akin to statistical power analysis, to determine optimal sampling parameters in animal tracking projects.

Analytical targets:
We considered three common estimates —home range area, speed and distance traveled.

‘ctmm’ R package
Calabrese *et al.* (2016)

Like any statistical tool, these methods still require sufficiently large sample sizes to achieve high accuracy.

i.simoes-silva@hzdr.de

Conceptualization

A successful animal tracking project requires a sampling schedule that leads to sufficiently large (effective) sample sizes.

SAMPLE SIZES

ABSOLUTE SAMPLE SIZE

n

T + Δt

SAMPLING DURATION

How long is an animal tracked for?

SAMPLING FREQUENCY

How frequently are locations collected?

EFFECTIVE SAMPLE SIZE

N_{area}

number of statistically independent locations

N_{speed}

number of statistically independent velocities

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

i.simoes-silva@hzdr.de

A successful animal tracking project requires a sampling schedule that leads to sufficiently large (effective) sample sizes.

SAMPLE SIZES

EFFECTIVE SAMPLE SIZE

N_{area} — number of statistically independent locations

N_{speed} — number of statistically independent velocities

Out of 72% of 369 empirical datasets 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 $N \ll n$

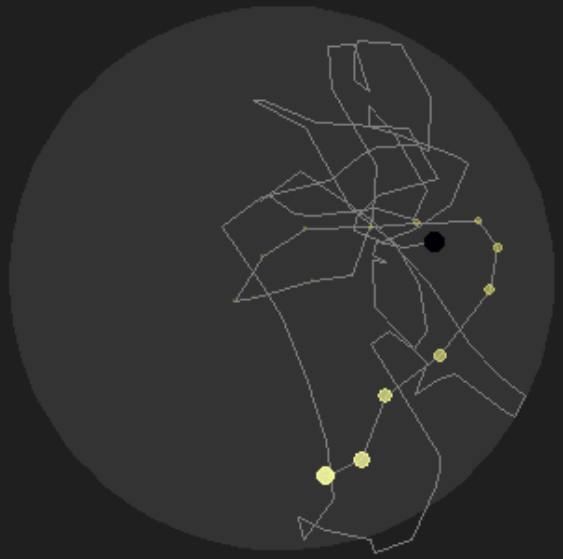
i.simoes-silva@hzdr.de

Conceptualization

τ_p

Position autocorrelation timescale

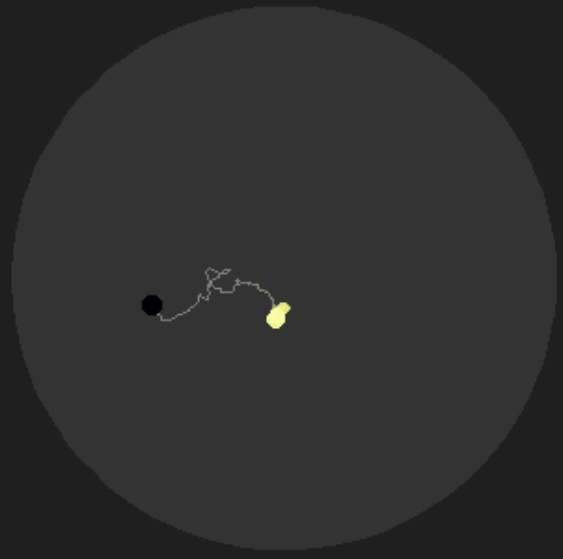
τ_p = 1 hour



τ_p = 1 day

τ_p = 5 days

τ_p = 10 days



T

SPACE-USE ◄ SAMPLING DURATION

☑ HOME RANGE

How long is an animal tracked for?

MOVEMENT BEHAVIOR

☐ SPEED & DISTANCE

i.simoes-silva@hzdr.de

Conceptualization

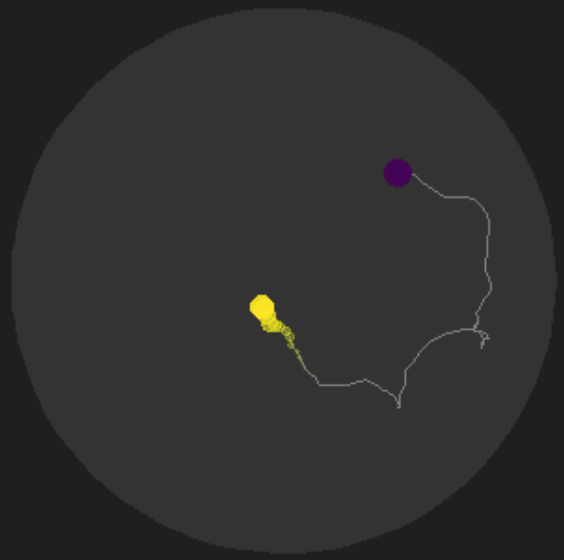
τ_v

Velocity autocorrelation timescale

τ_v = 1 minute

τ_v = 1 hour

τ_v = 12 hours



τ_v = 1 day

Δt

SPACE-USE

☐ HOME RANGE

SAMPLING FREQUENCY ⋯⋯▸ MOVEMENT BEHAVIOR

How frequently are locations collected?

☑ SPEED & DISTANCE

✉ i.simoes-silva@hzdr.de

Conceptualization

Define research questions

Identify spatiotemporal scales

Choose sampling design

Collect animal tracking data

Analyze data, mitigate biases

Assess conclusions

It is not physically possible for animal movement to be uncorrelated. Now, the questions are:

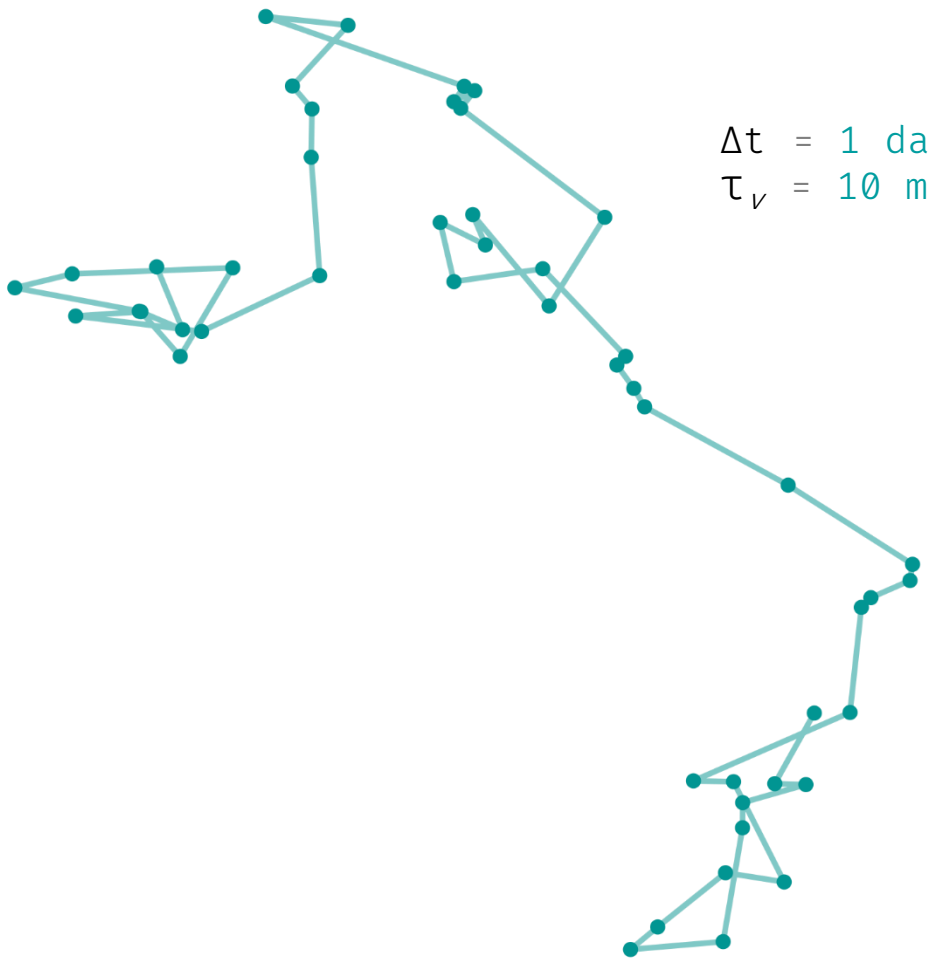
1. Can you detect a signature of these correlations in your data?

2. And is this data sufficient to answer specific research questions?

i.simoes-silva@hzdr.de

Conceptualization

Simulated tracking dataset with a new location once per day, τ_v > Δt.



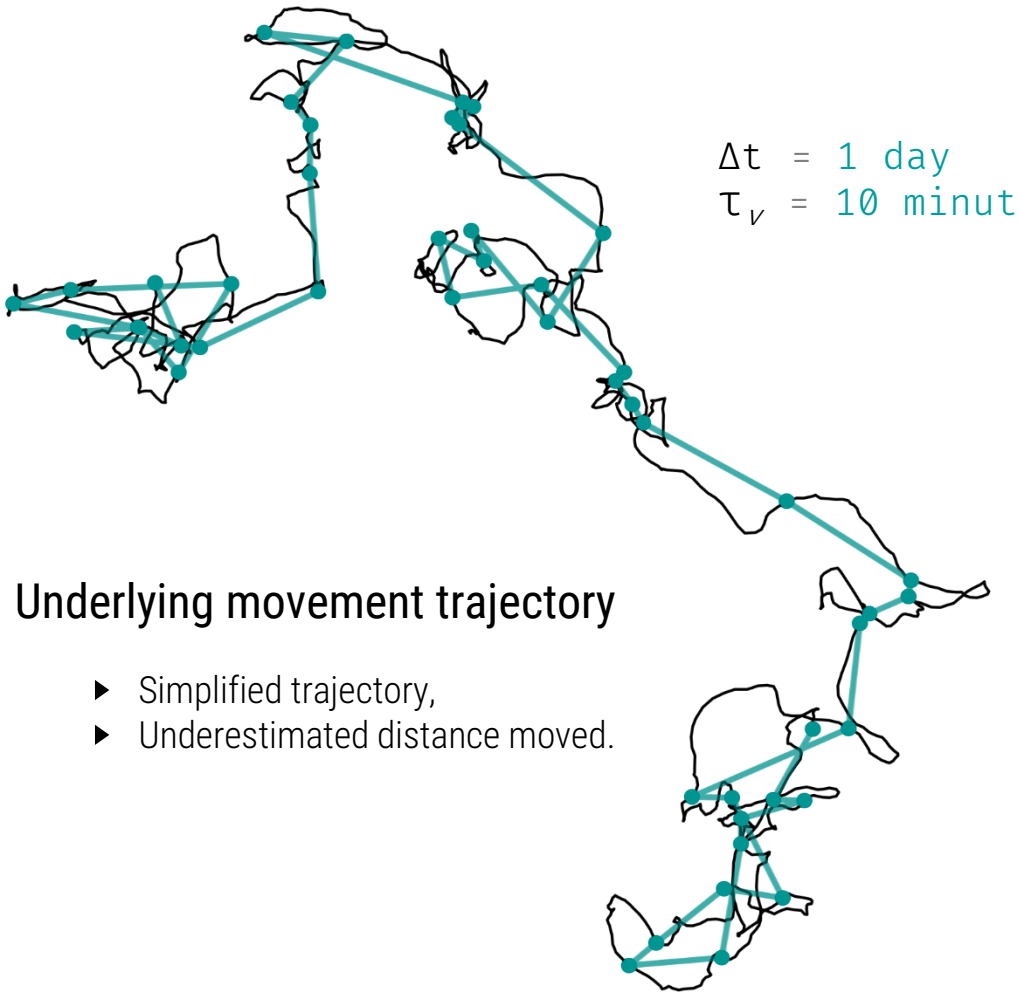
i.simoes-silva@hzdr.de

Conceptualization

Simulated tracking dataset with a new location once per day, $\tau_v > \Delta t$.

We must carefully consider the frequency of data collection!

For the same Δt, this bias will be greater for individuals with more tortuous movement (shorter τ_v).



i.simoes-silva@hzdr.de

Conceptualization

Simulated tracking dataset with a duration of 6 months, $\tau_p > T$.

T = 6 months

τ_p = 8 months

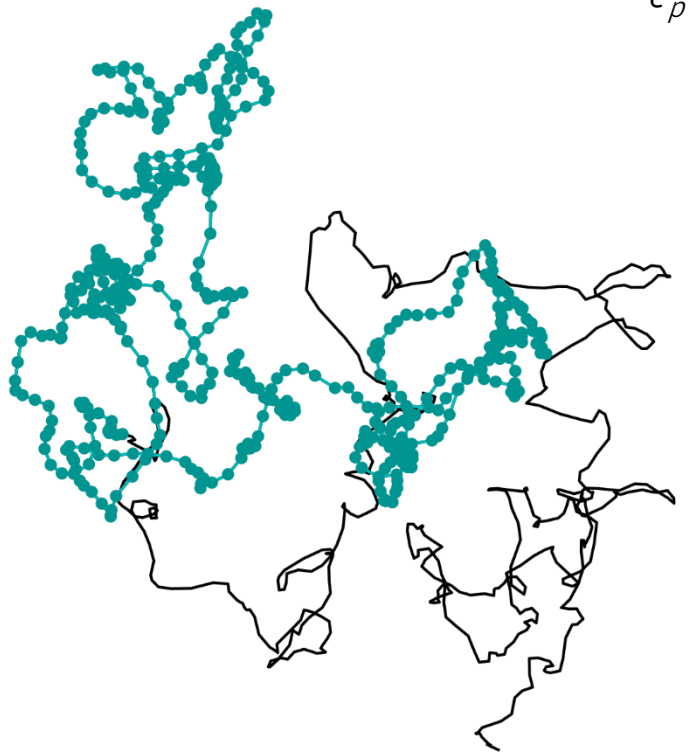
i.simoes-silva@hzdr.de

Conceptualization

Simulated tracking dataset with a duration of 6 months, $\tau_p > T$.

$T = 12$ months

$\tau_p = 8$ months



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

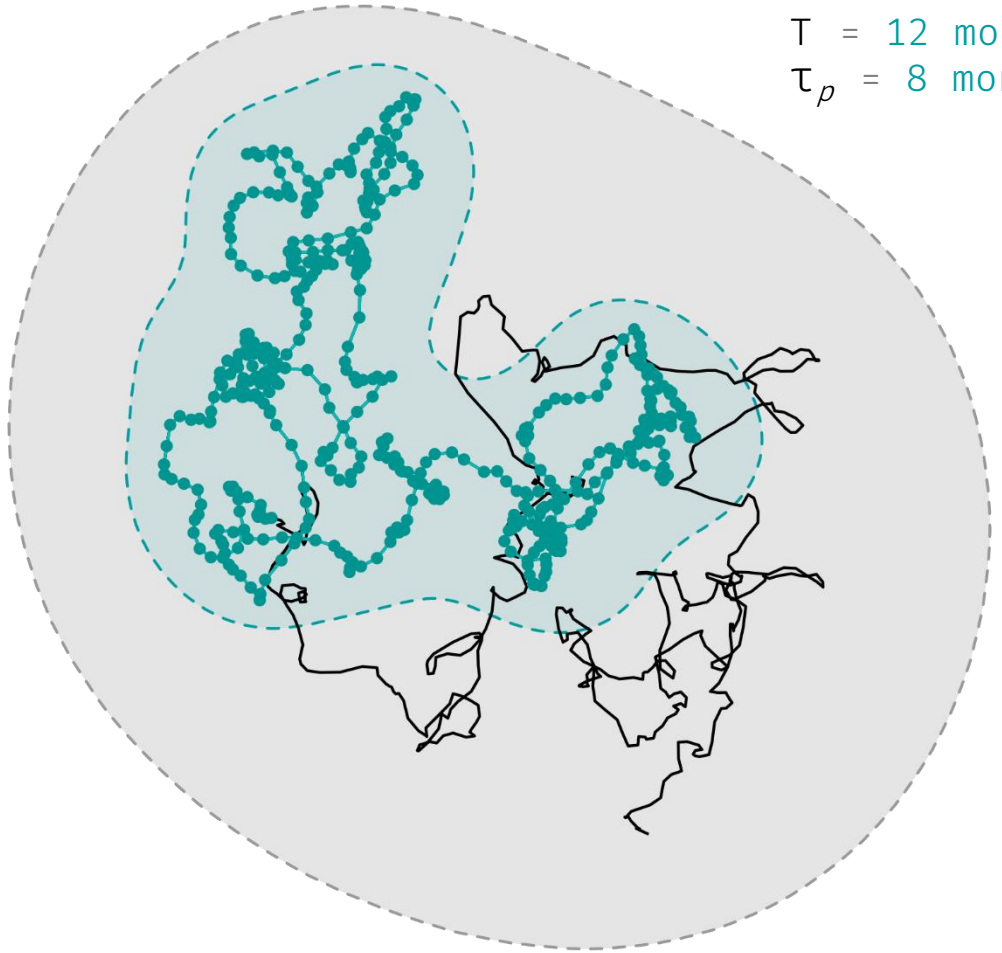
i.simoes-silva@hzdr.de

Conceptualization

Simulated tracking dataset with a duration of 6 months, $\tau_p > T$.

We must carefully consider the duration of data collection!

For the same T, the extent of this bias will be greater for individuals with longer crossing times (τ_p).

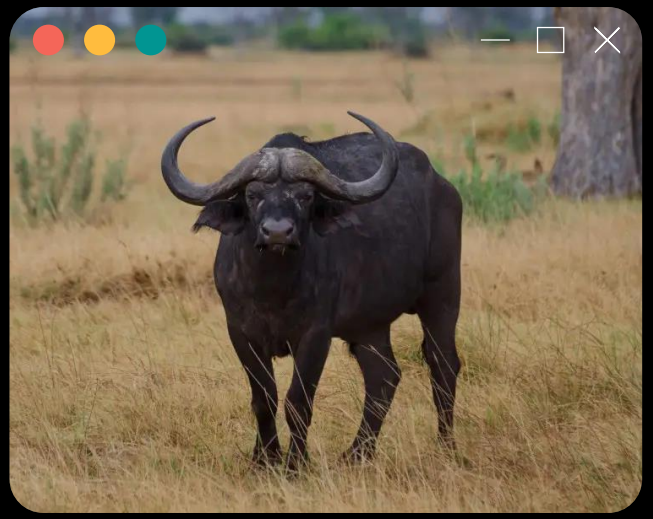


Movement trajectory for the following **6 months**.

- Sampling missed used areas.
- Underestimated home range.

i.simoes-silva@hzdr.de

Conceptualization



African buffalo
(Syncerus caffer)

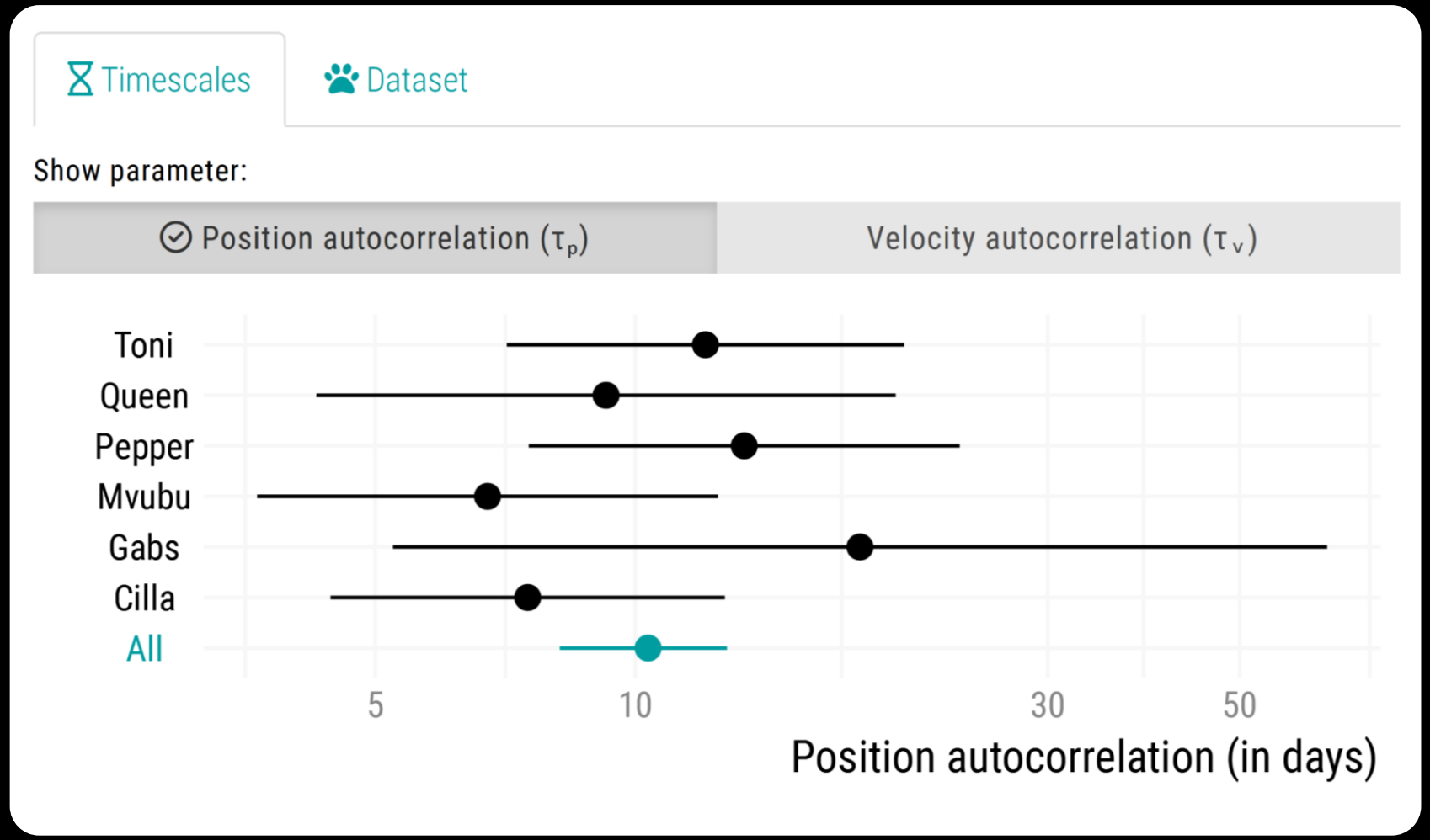
Position autocorrelation

10.3 days

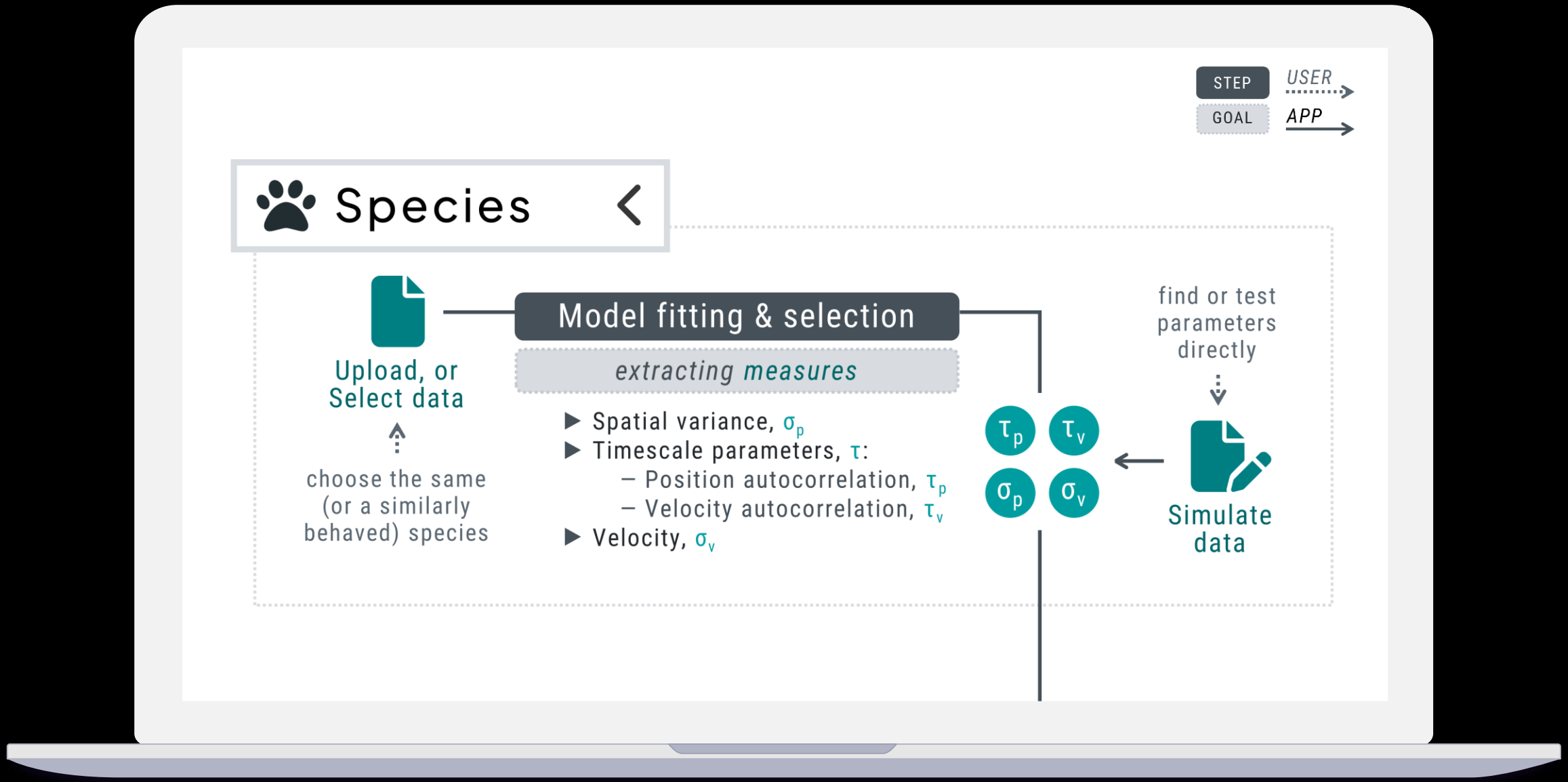
8.2 – 12.8

(τ_p)

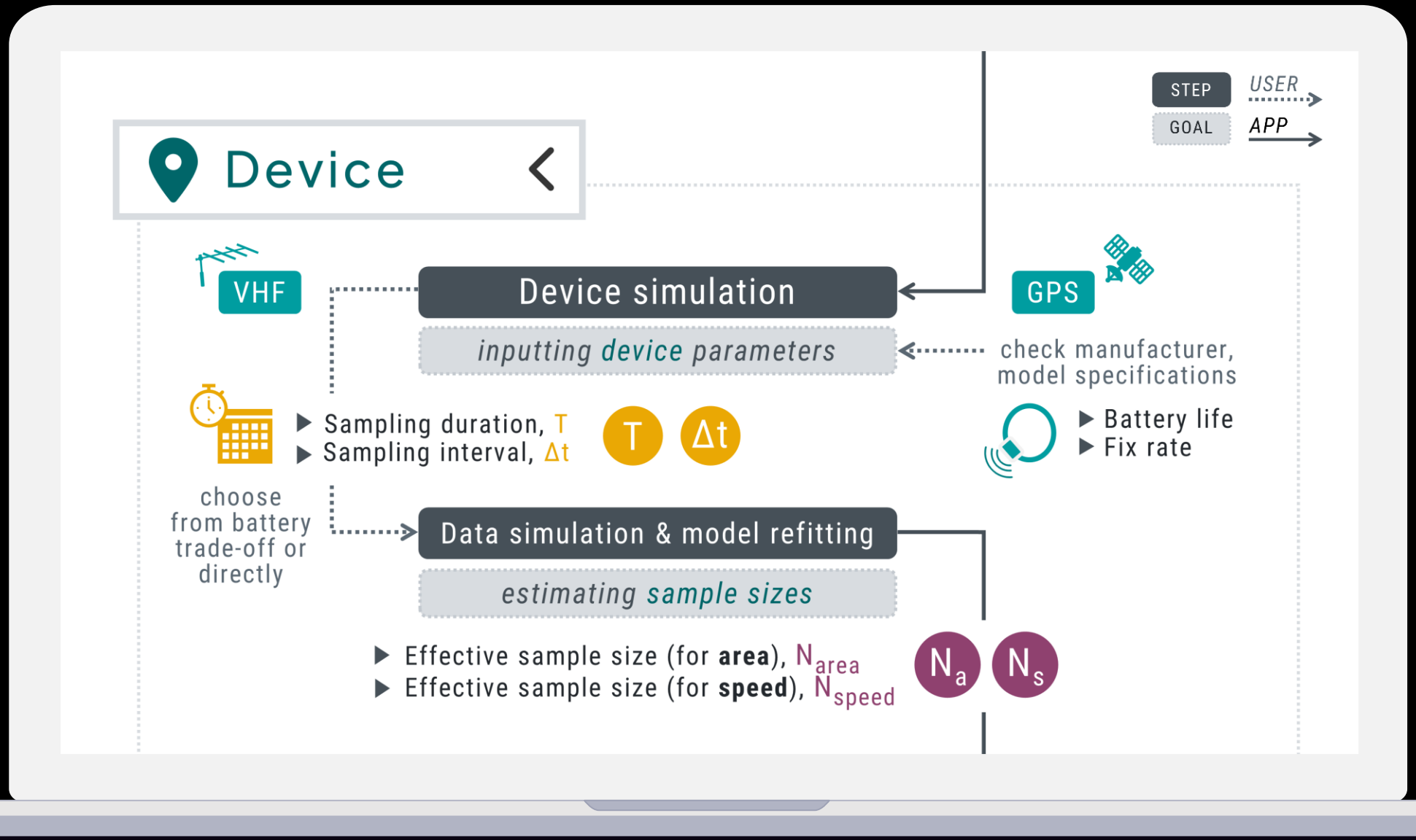
These parameters are fairly conservative at the species- and population-level.



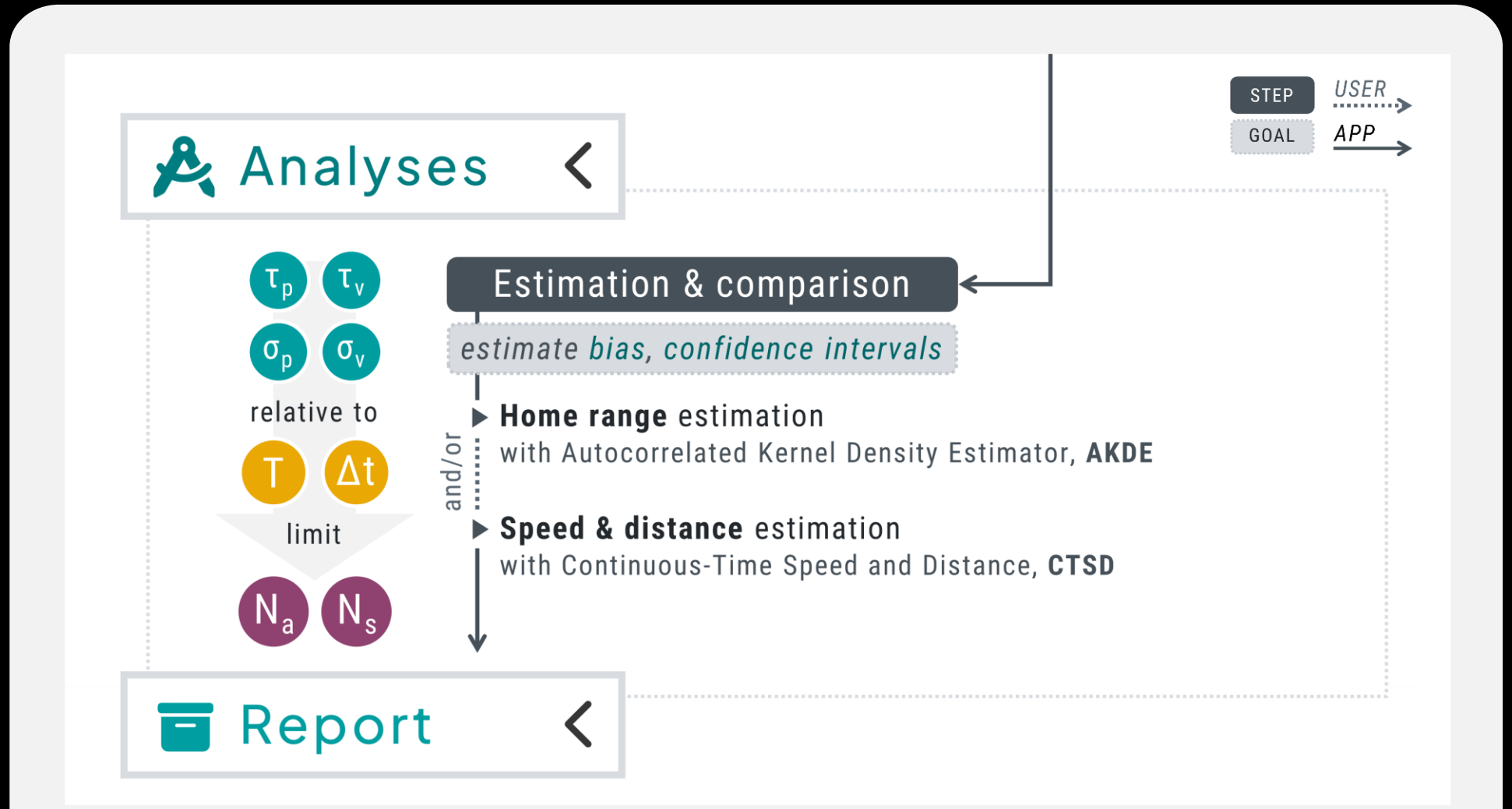
i.simoes-silva@hzdr.de



i.simoes-silva@hzdr.de

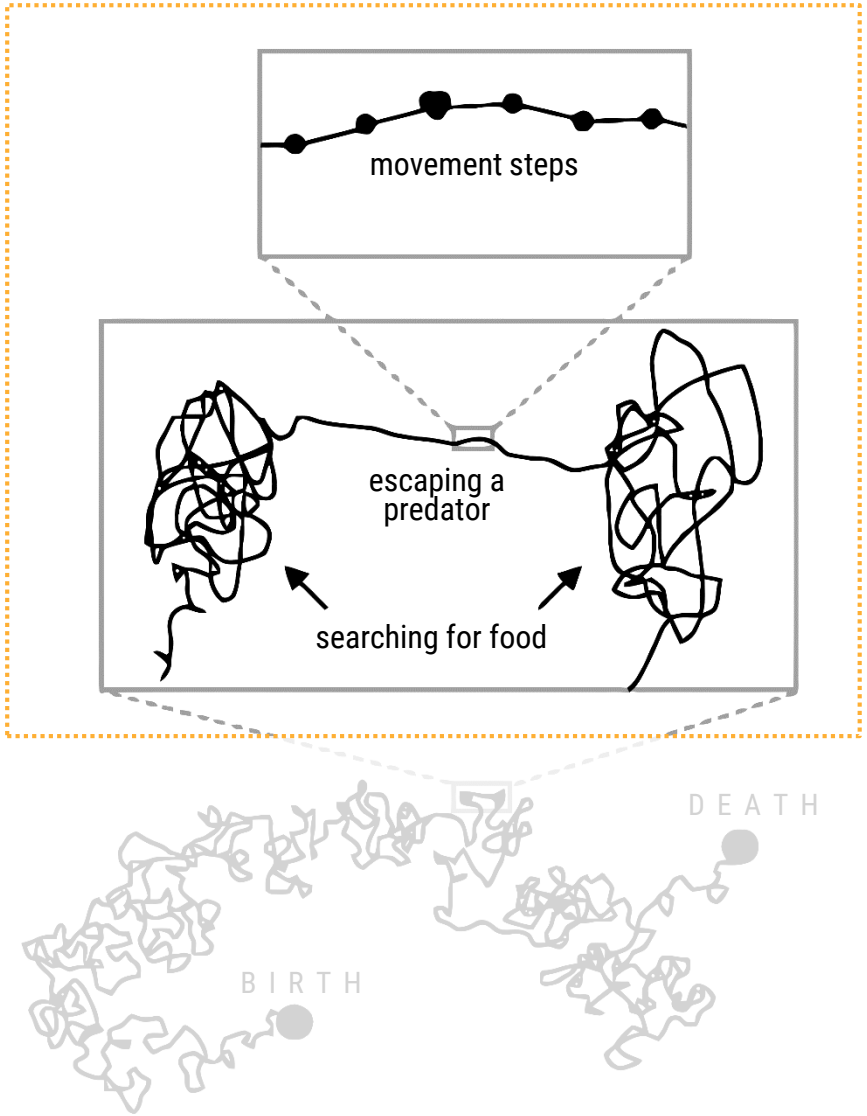


i.simoes-silva@hzdr.de



i.simoes-silva@hzdr.de

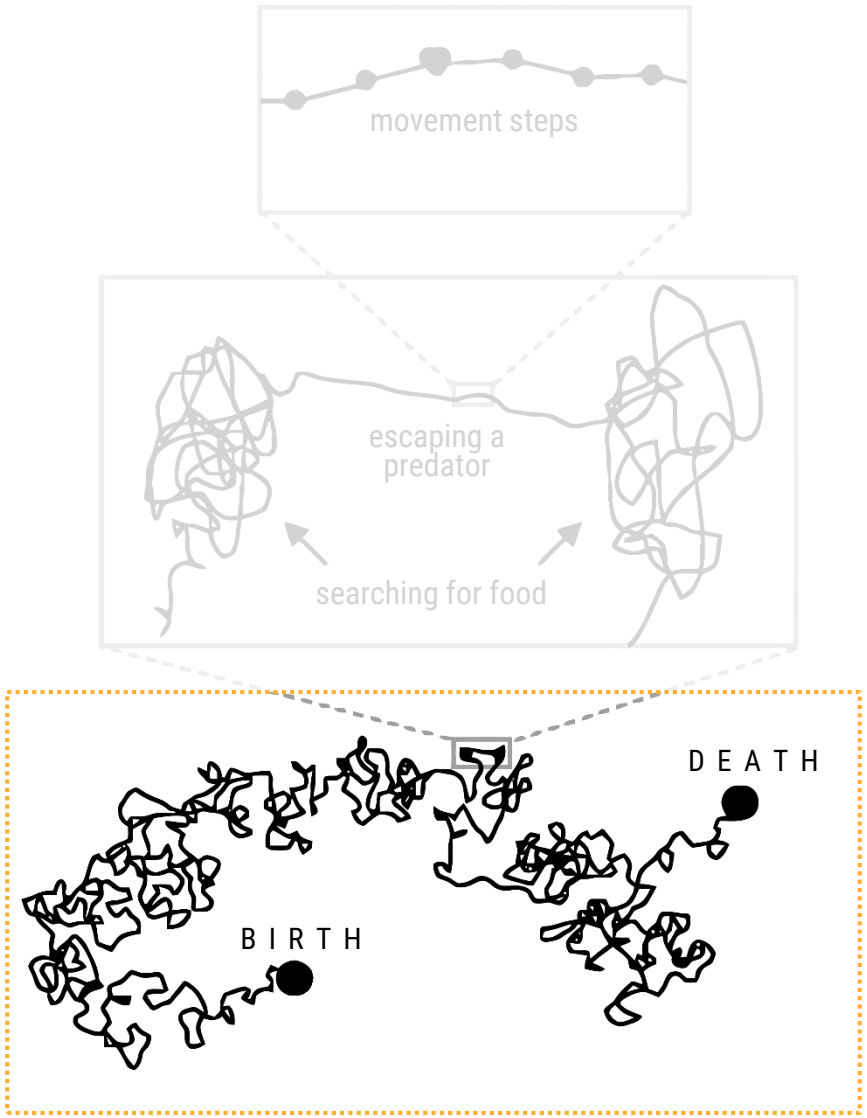
Nathan *et al.* (2008)



If the goal is speed & distance estimation, adjust your sampling interval (Δt) to ensure data is of sufficient resolution to detect τ_v.

i.simoes-silva@hzdr.de

Nathan *et al.* (2008)

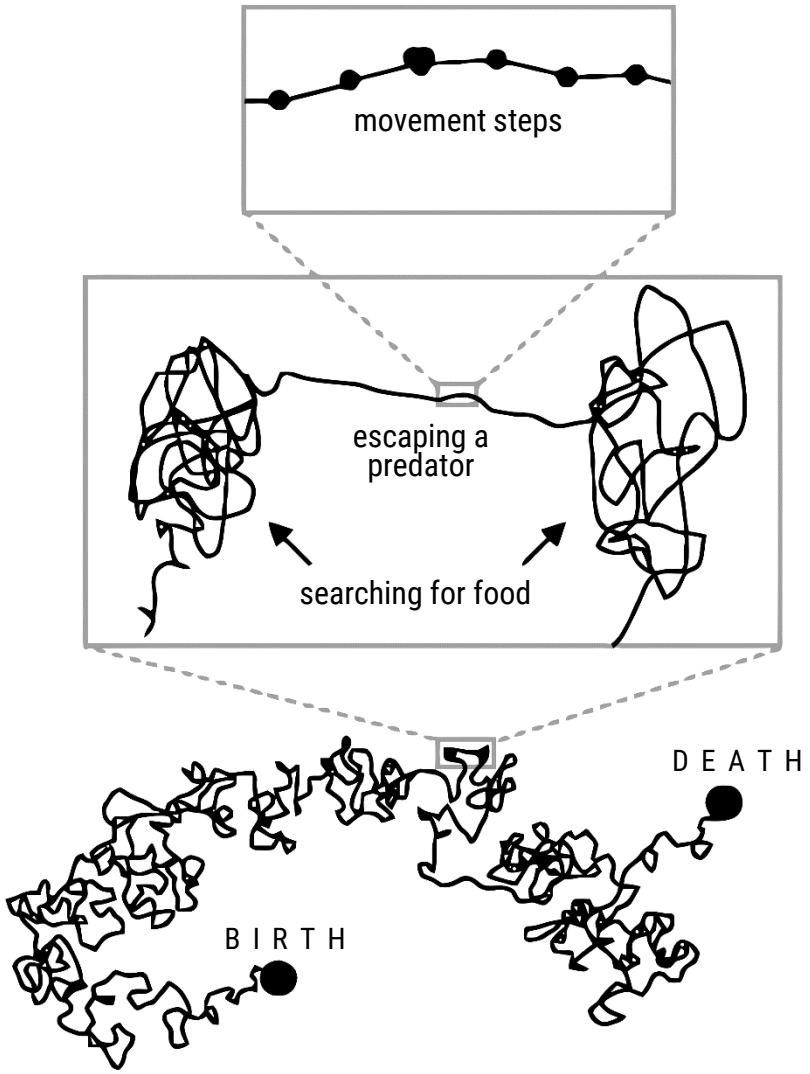


If the goal is speed & distance estimation, adjust your sampling interval (Δt) to ensure data is of sufficient resolution to detect τ_v.

If the goal is home range area estimation, adjust your sampling duration to ensure data is sufficient to detect τ_p.

i.simoes-silva@hzdr.de

Nathan *et al.* (2008)



If the goal is speed & distance estimation, adjust your sampling interval (Δt) to ensure data is of sufficient resolution to detect τ_v.

If the goal is home range area estimation, adjust your sampling duration to ensure data is sufficient to detect τ_p.

If both,
You may be able to address large-scale and fine-scale questions, but not always both concurrently.

i.simoes-silva@hzdr.de

move.design

Improved workflow (v0.3.0):

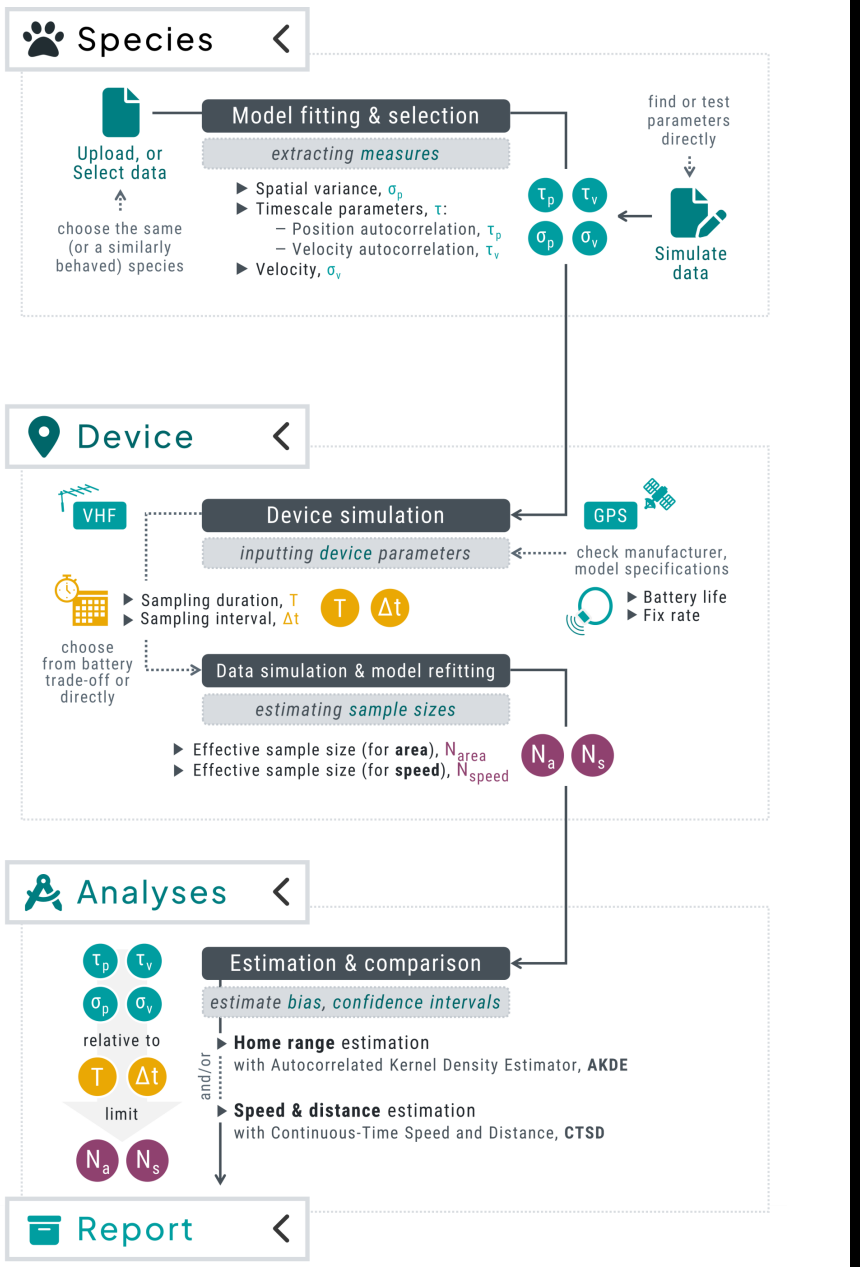
- Run **multiple simulations** for:
  - a predefined (*population*) sample size,
  - an iteratively higher sample size,

  until the estimate error is below a specified threshold.

- Run **meta-analyses**, and get estimates for:

  ↖ Population-level inference
  Fleming et al. (2022)

  - mean of sampled population,
  - compare means of two sampled populations.
    (*e.g.*, males/females)



i.simoes-silva@hzdr.de

General recommendations

Utilization distribution

Range distribution

extrapolate space use into the future

≠

Occurrence distribution

interpolate between data points in the past

Analyses that ignore ***position autocorrelation*** significantly *underestimate* home range areas.

Analyses that ignore ***velocity autocorrelation*** typically *overestimate* occurrence areas.

i.simoes-silva@hzdr.de

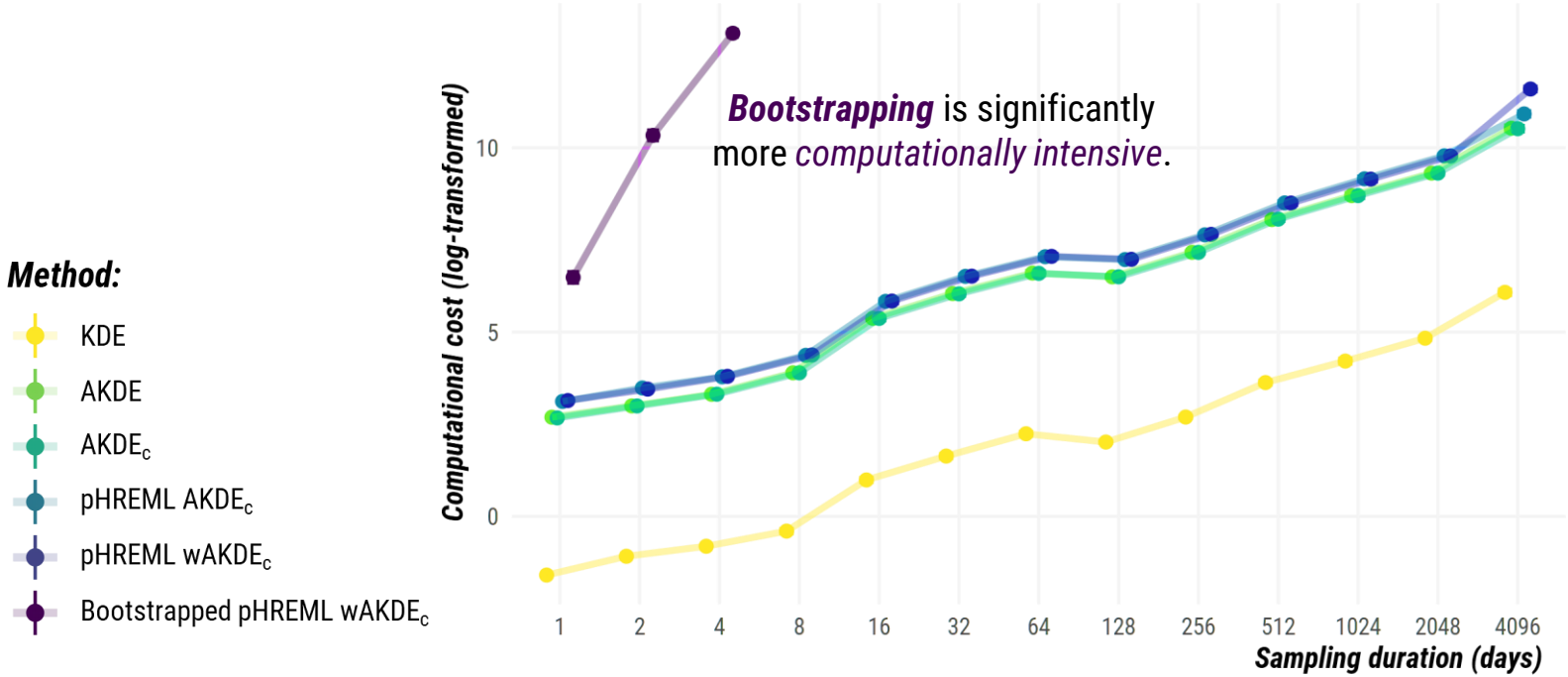
1. Home ranges — Autocorrelation Kernel Density Estimator (AKDE)

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.

i.simoes-silva@hzdr.de

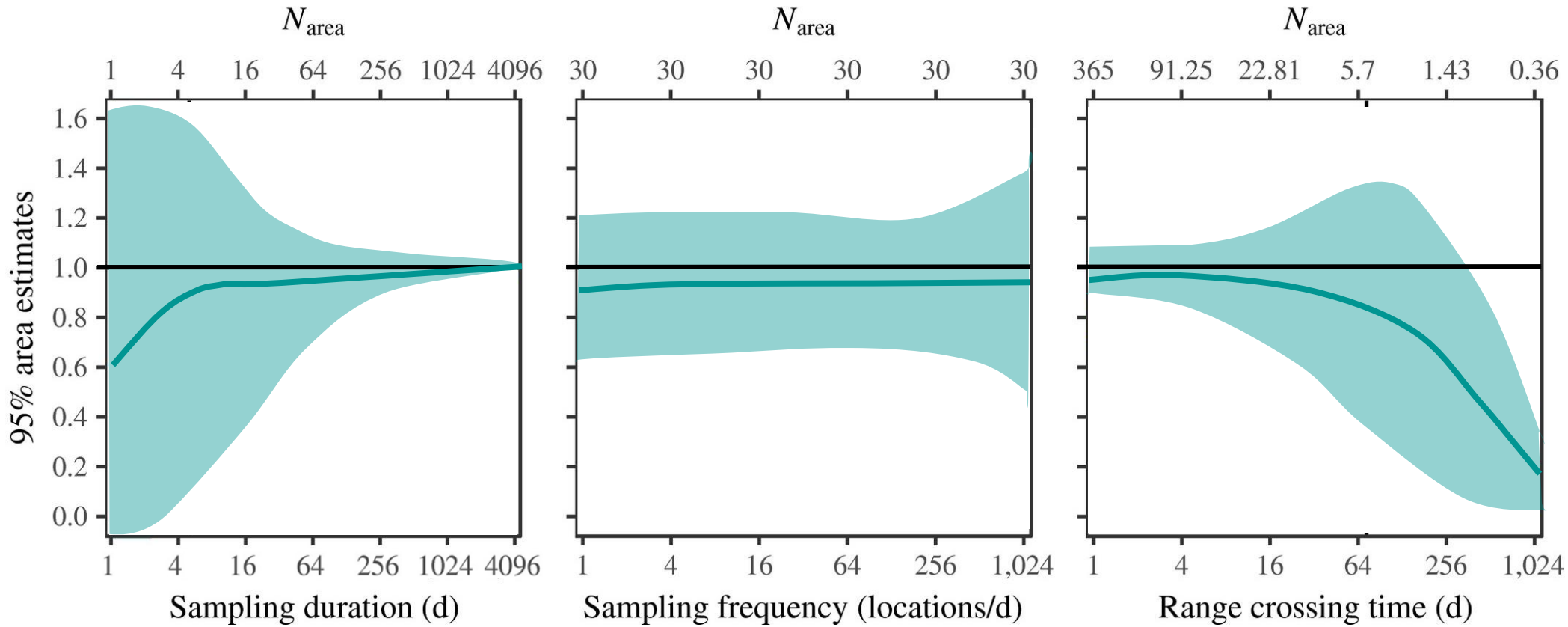
1. Home ranges — Autocorrelation Kernel Density Estimator (AKDE)

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.



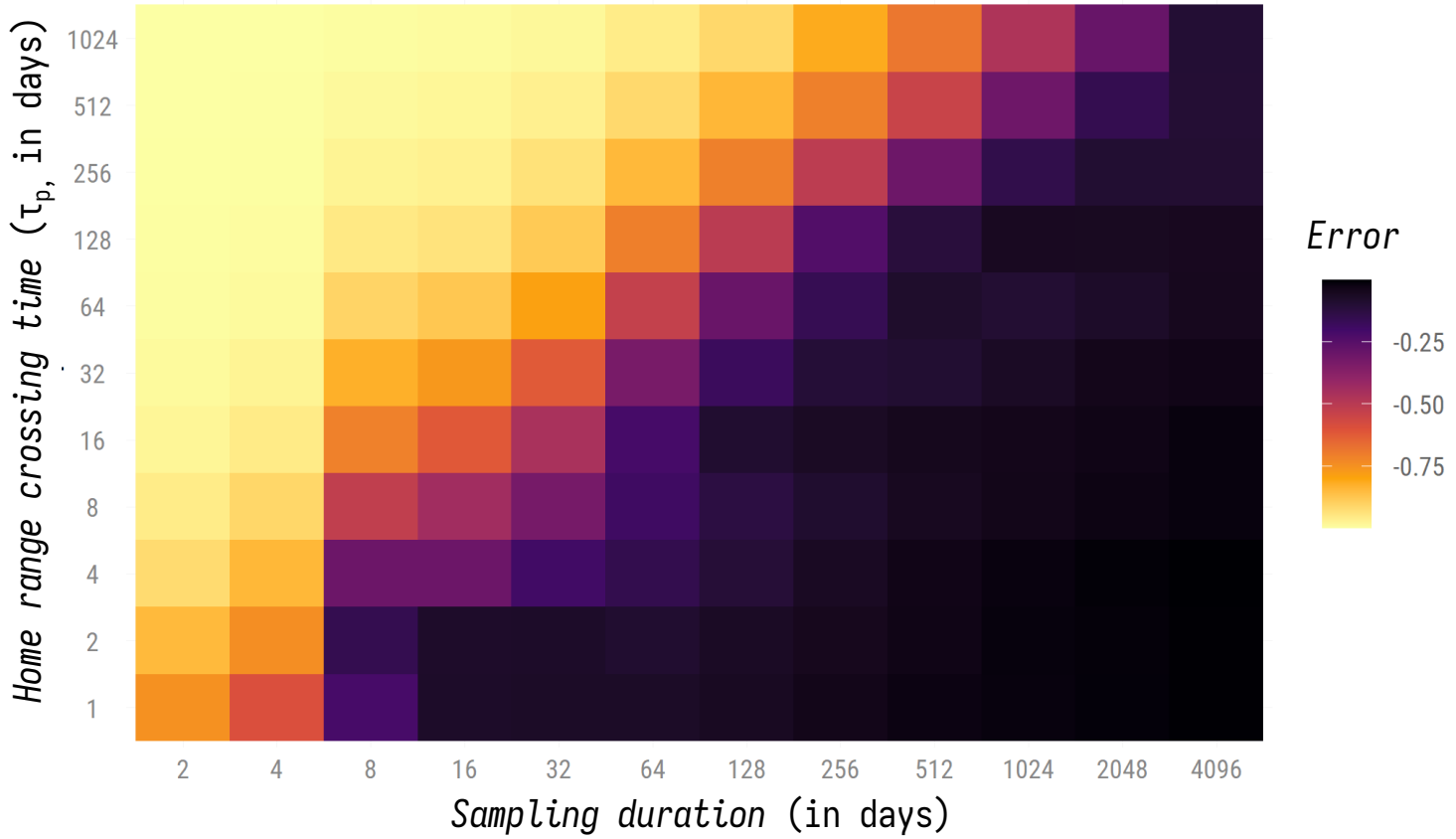
i.simoes-silva@hzdr.de

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



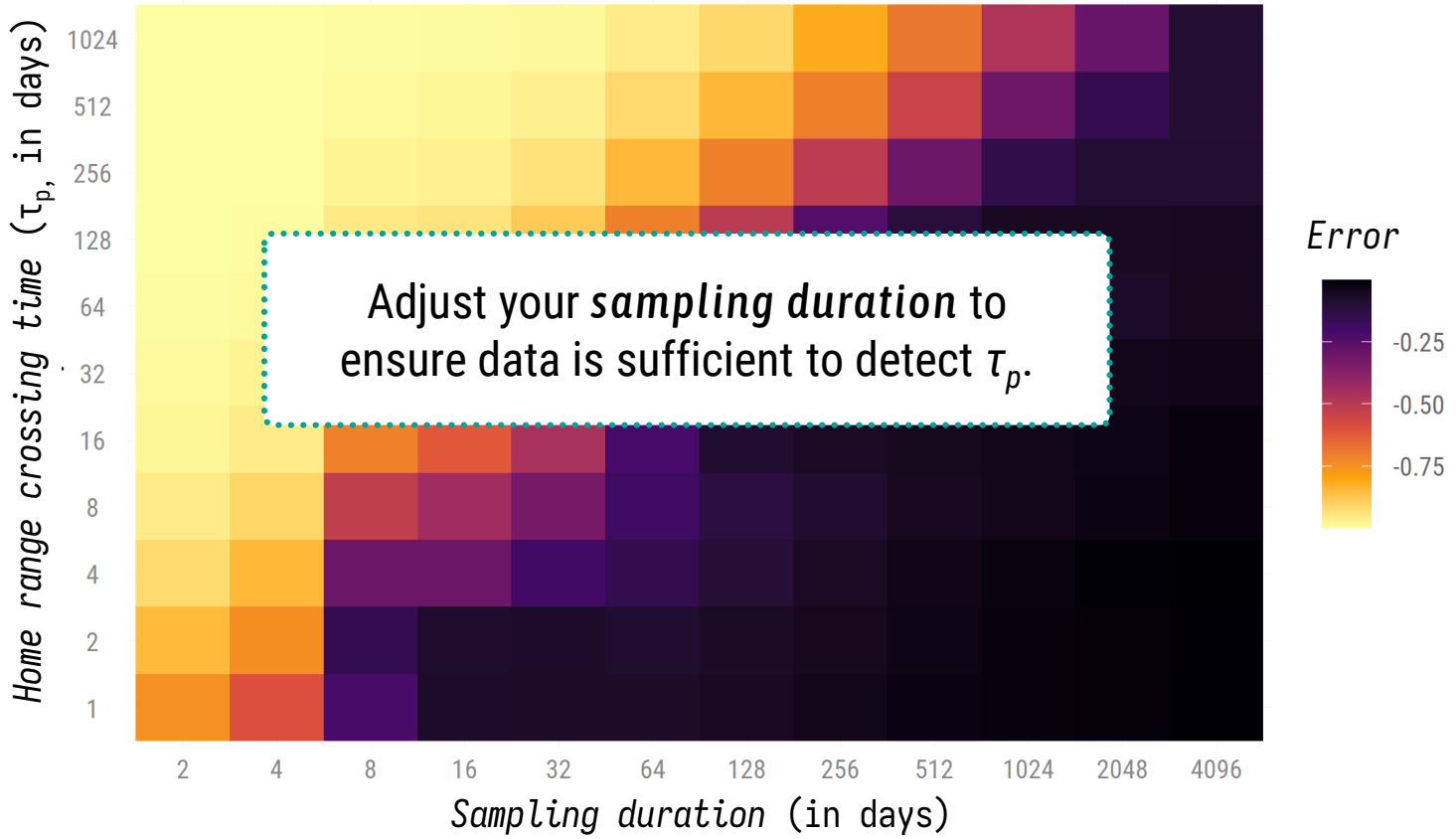
i.simoes-silva@hzdr.de

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



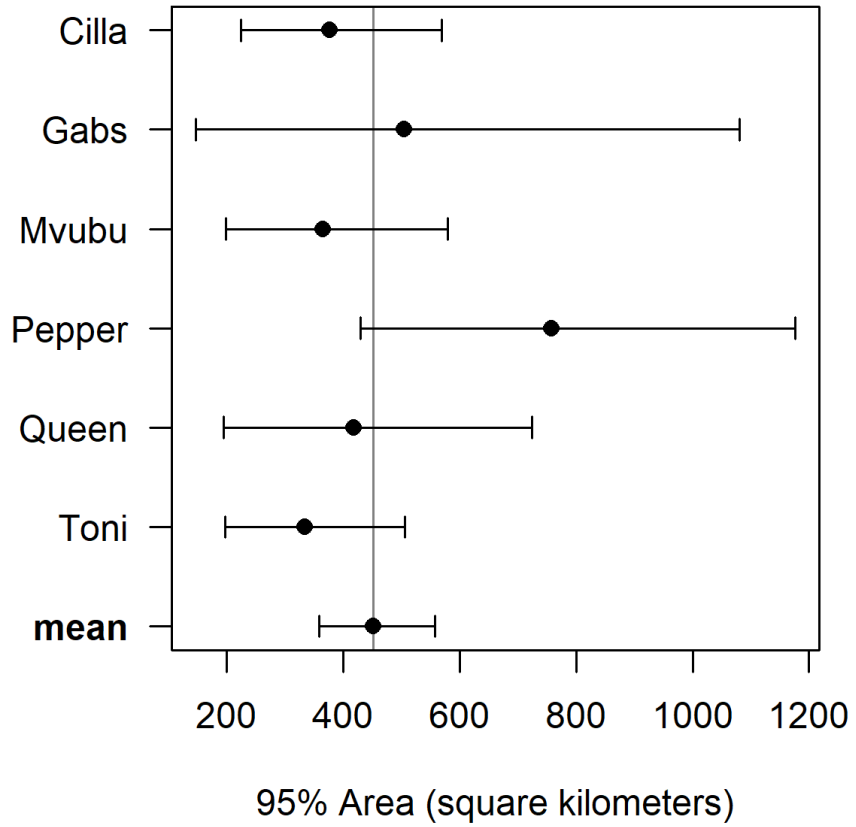
i.simoes-silva@hzdr.de

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



i.simoes-silva@hzdr.de

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

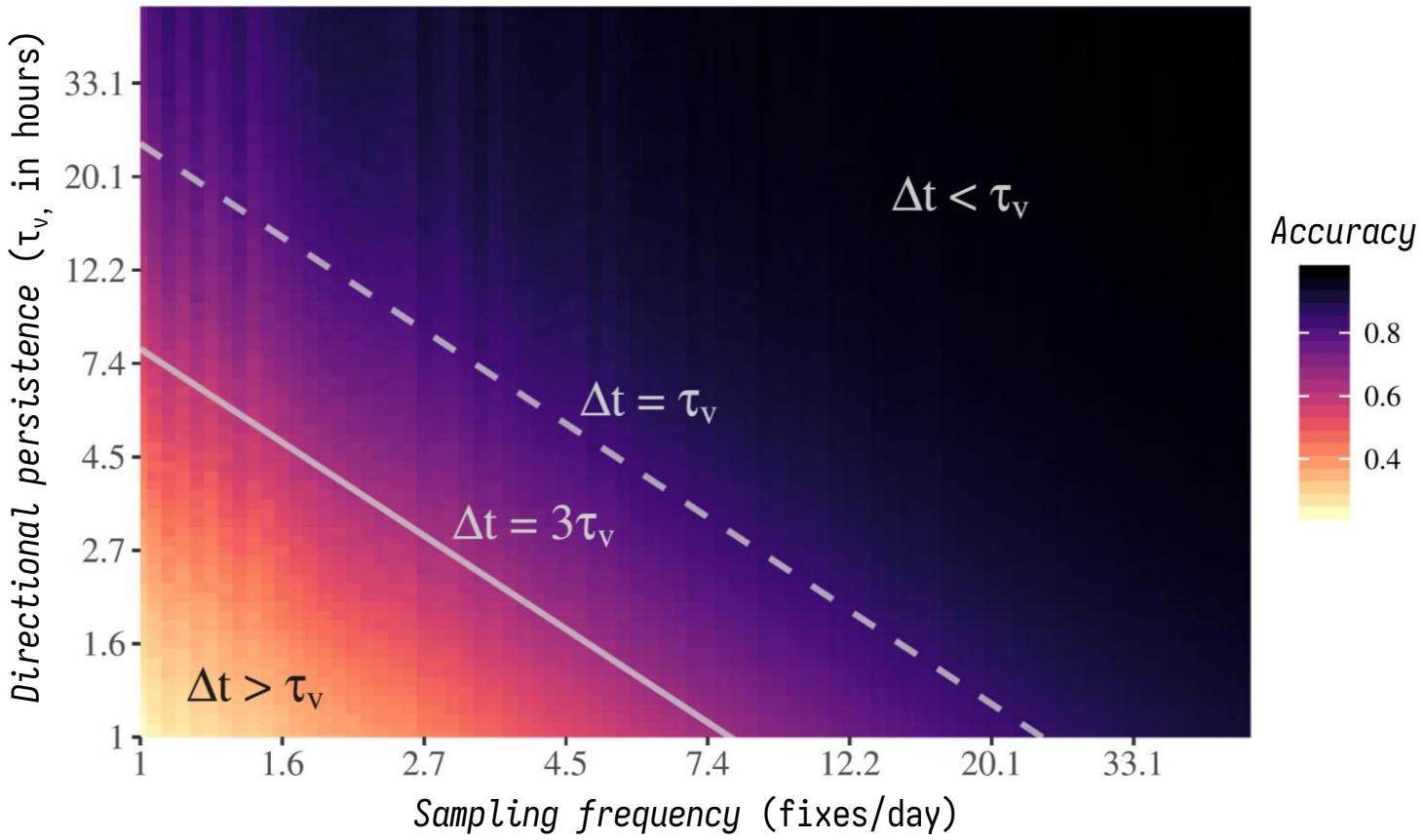


For *meta-analyses*:

> 2–3 observed home-range crossings (τ_p)
> 2–3 representative individuals.

i.simoes-silva@hzdr.de

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

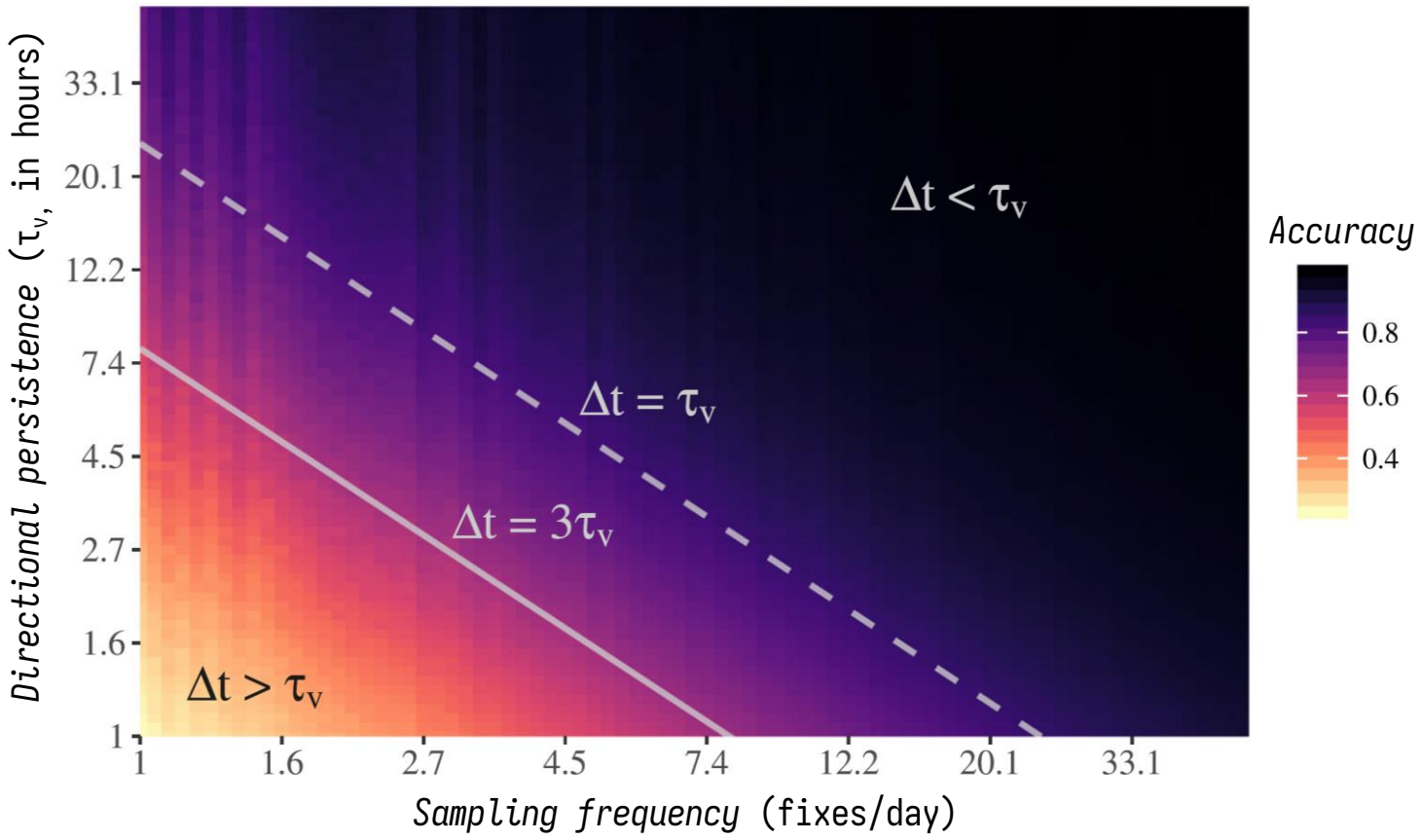


If $\Delta t > 3\tau_v$, no statistically significant signature of the animal's velocity will remain in the location data.

If $3\tau_v > \Delta t > \tau_v$, there will be some positive bias (τ_v can not be accurately estimated).

i.simoes-silva@hzdr.de

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

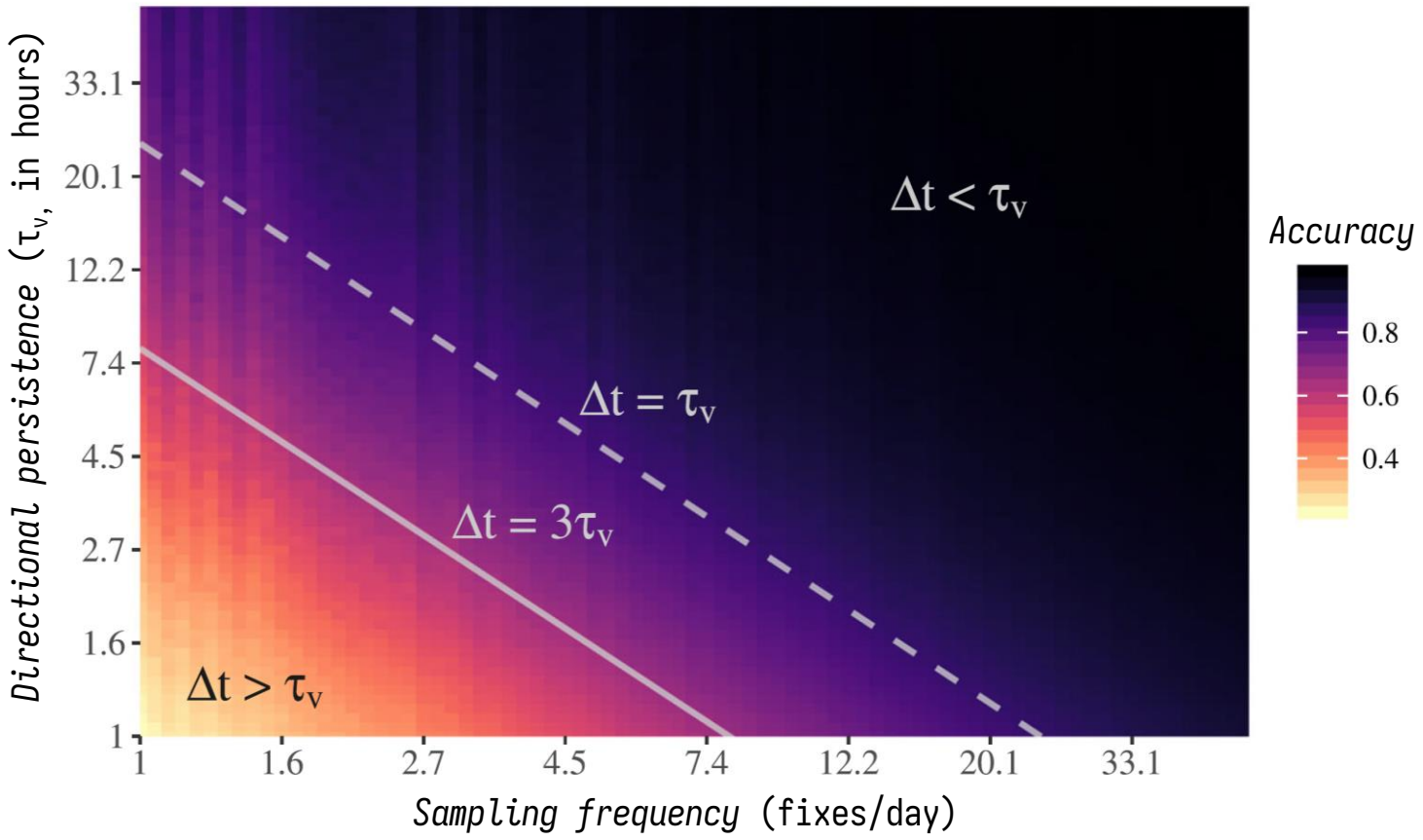


Adjust your sampling interval to ensure data is of sufficient resolution to detect τ_v.

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

i.simoes-silva@hzdr.de

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



Adjust your sampling interval to ensure data is of sufficient resolution to detect τ_v.

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

i.simoes-silva@hzdr.de

General 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. → move design

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

i.simoes-silva@hzdr.de

CASUS
CENTER FOR ADVANCED SYSTEMS UNDERSTANDING

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!

i.simoes-silva@hzdr.de

i.simoes-silva@hzdr.de