

MARE-Madeira 2025

Population-level inferences

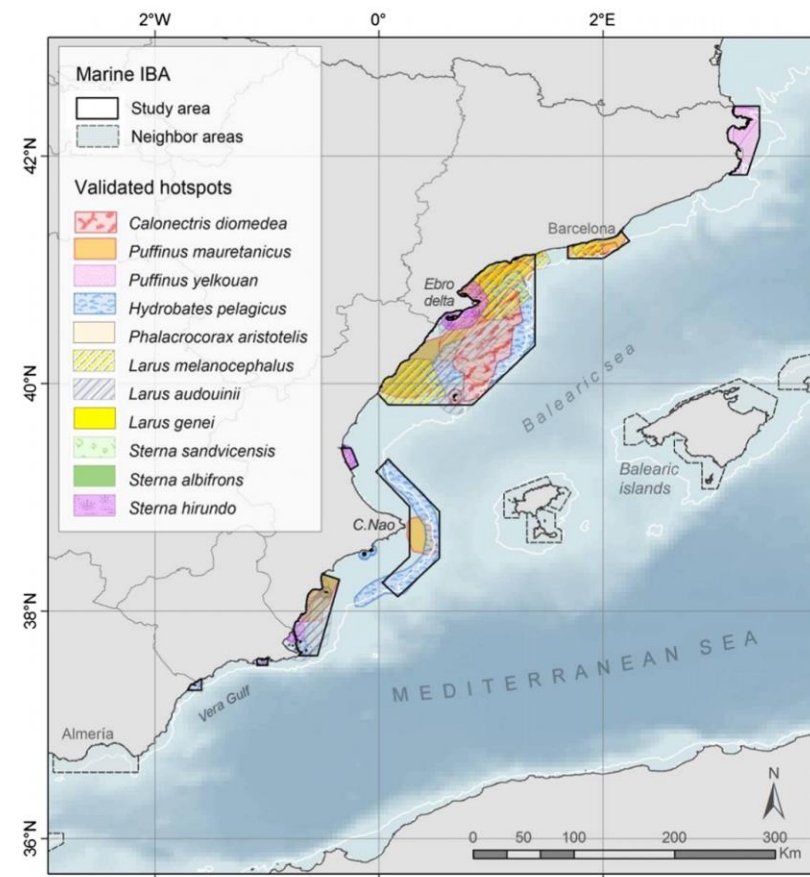
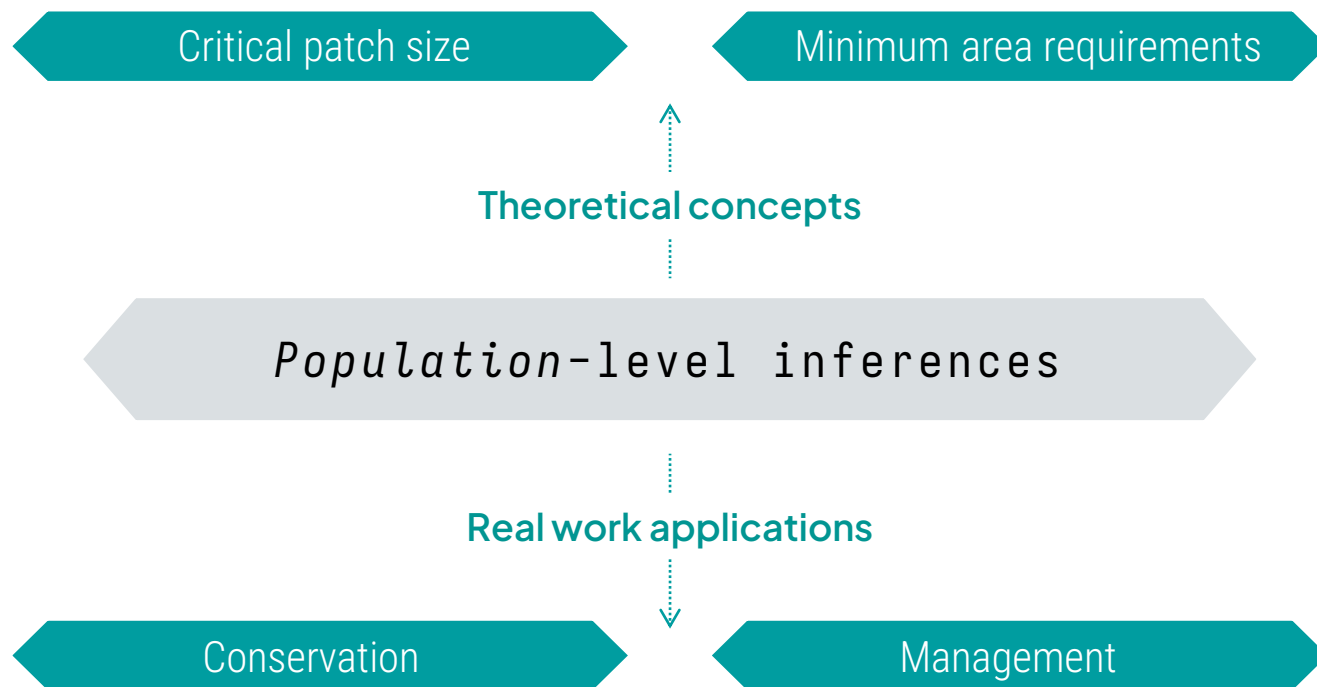
Using the 'ctmm' R package




Inês Silva

✉ i.simoes-silva@hzdr.de

Using foraging hotspots of pelagic seabirds to identify marine **Important Bird Areas (IBAs)** in Spain.



 **Arcos et al. (2012)**
DOI: 10.1016/j.biocon.2011.12.011



Analyses of ecological data should always account for the **uncertainty** in the process(es) that generated the data.



Accurately estimating **area requirements** is of utmost importance for conservation, from the **individual** to the **population level**.



We want to quantify the effect of covariates, such as **species**, **sex**, **body size**, **age**, **habitat**, **anthropogenic impact**, etc...



...even if we are comparing different populations with different **movement behaviors** or **sampling schedules**.

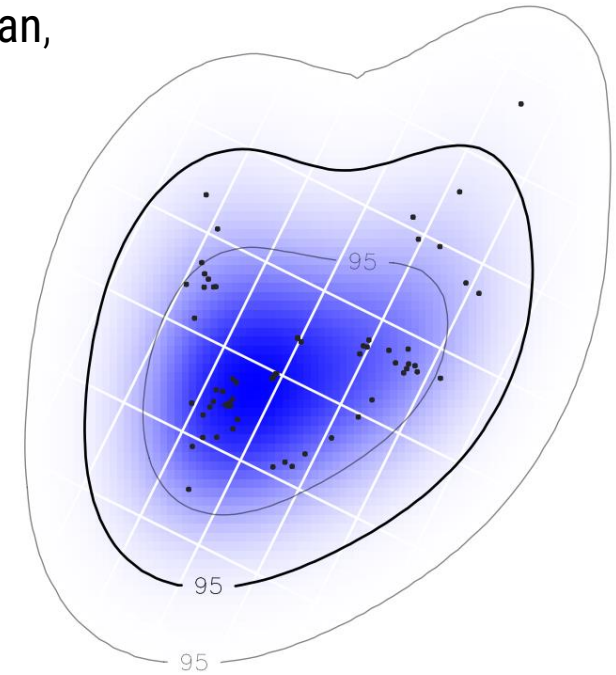
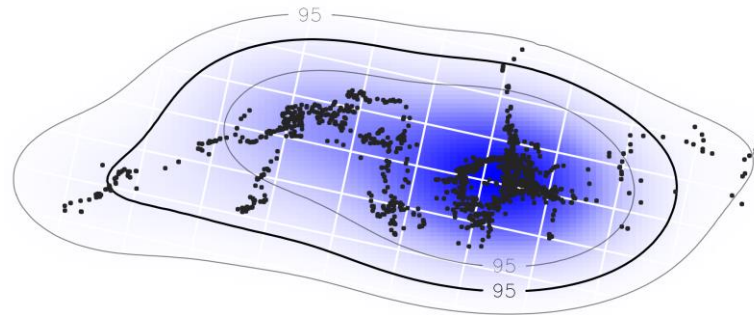




Conventionally,

Taking KDE and MCP home-ranged estimates and feed them into general purpose statistical analyses, which assumes that the individual home-range areas are measured exactly

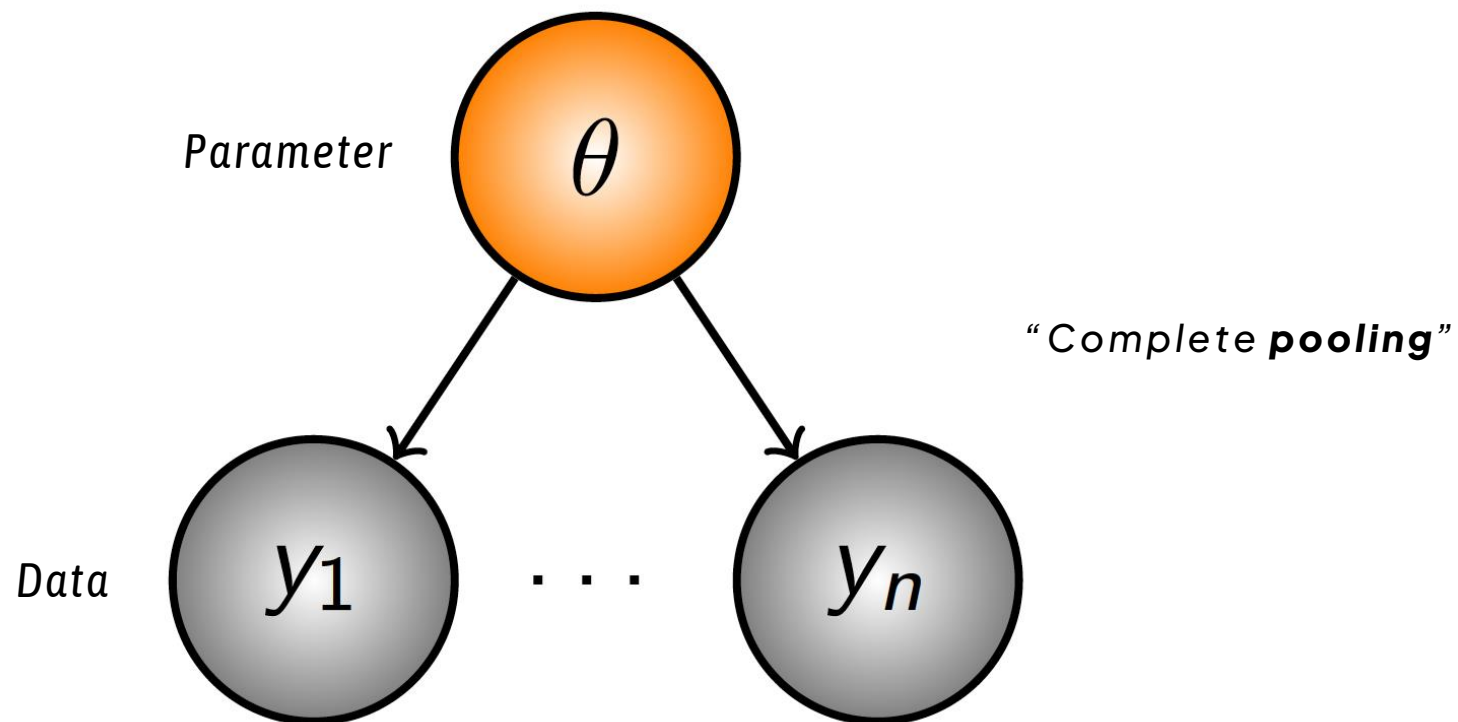
e.g., a single population would be described by its **sample mean**, and two populations would be compared with a **t-test**





NON-HIERARCHICAL MODELS

How does data inform parameters?

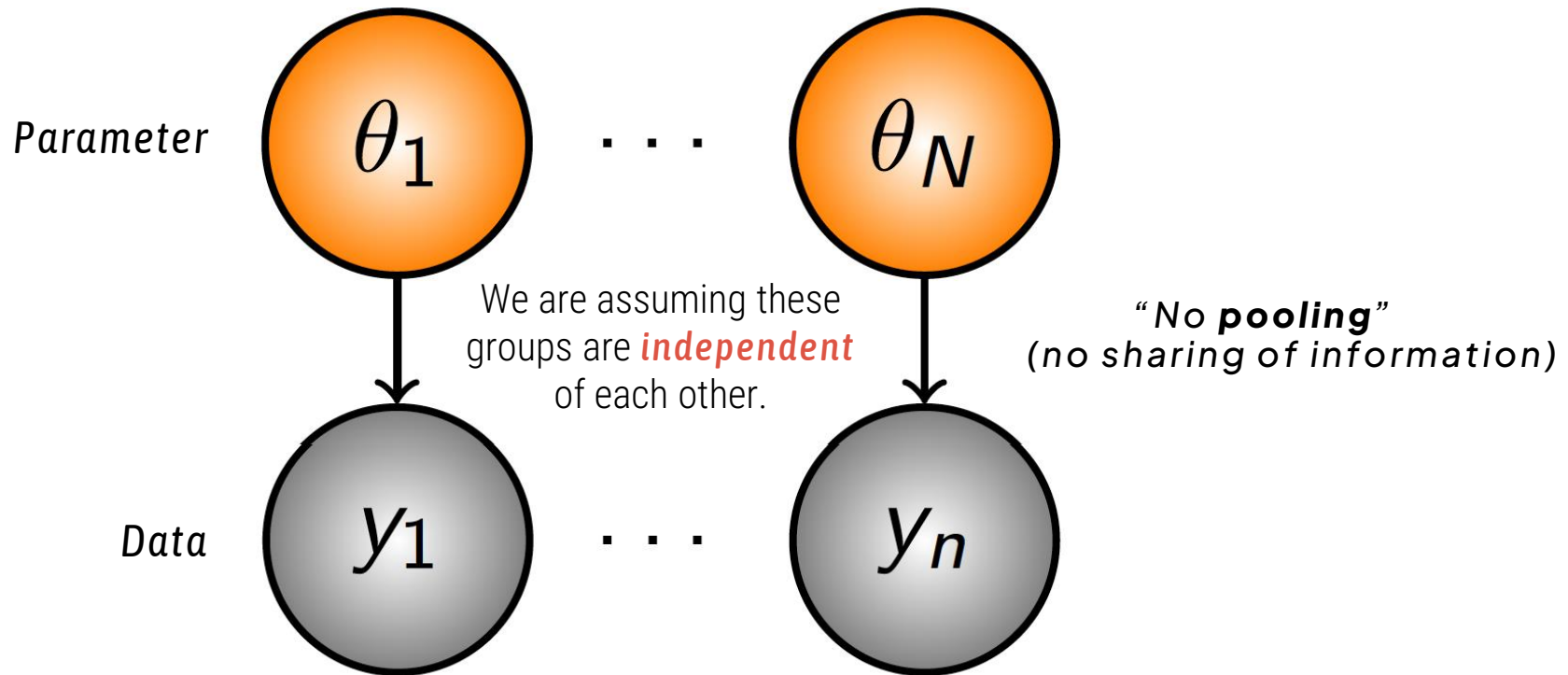


Adapted from Midway (2008)



NON-HIERARCHICAL MODELS

How does data inform parameters?

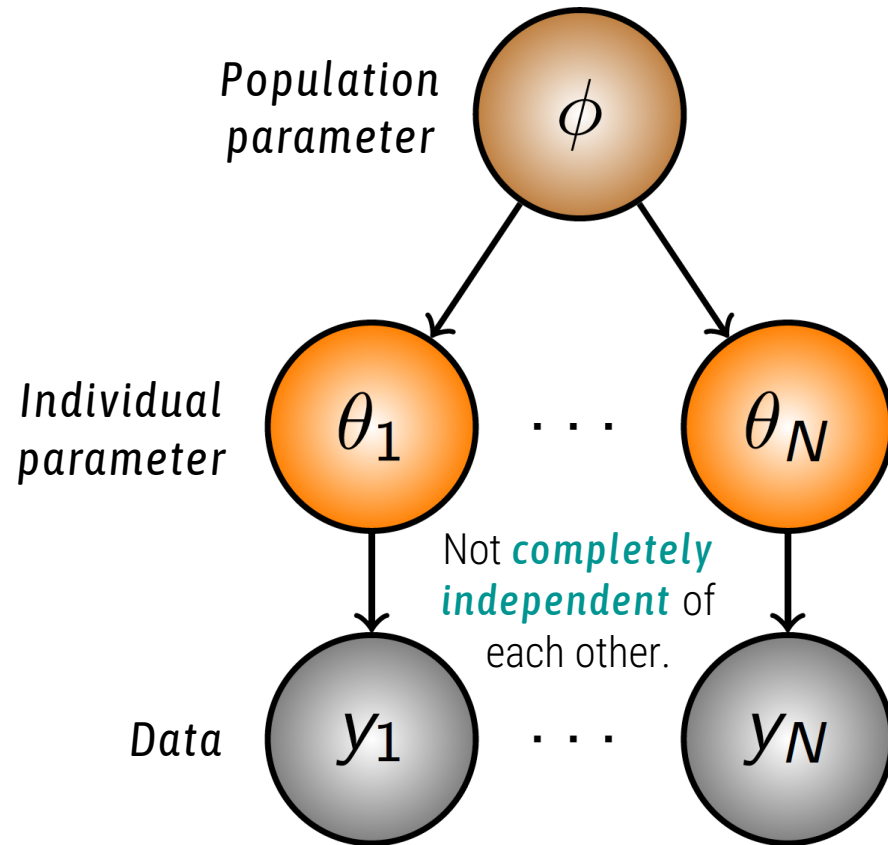


Adapted from Midway (2008)



HIERARCHICAL MODELS

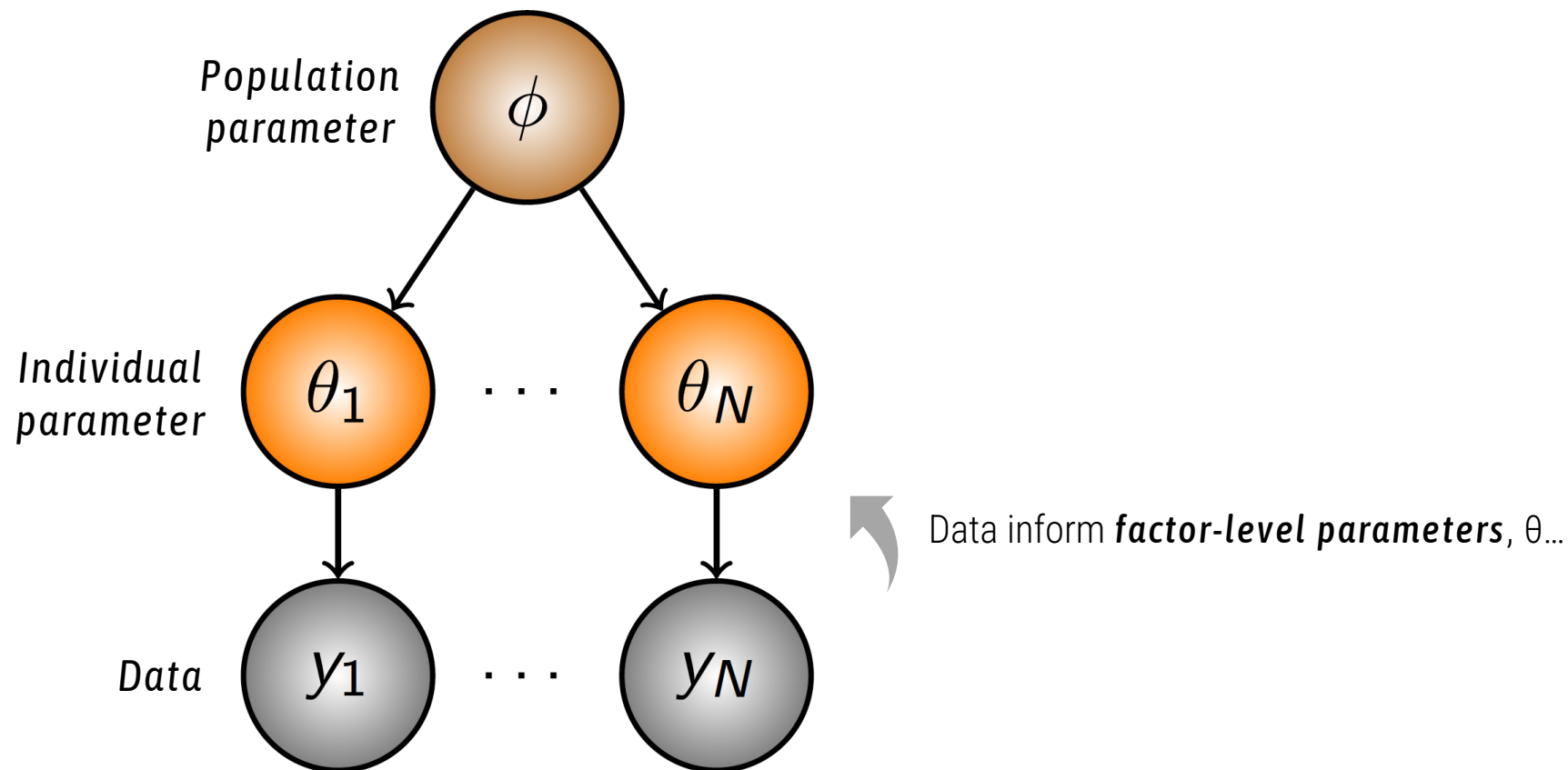
How does data inform parameters?



Adapted from Midway (2008)

HIERARCHICAL MODELS

How does data inform parameters?

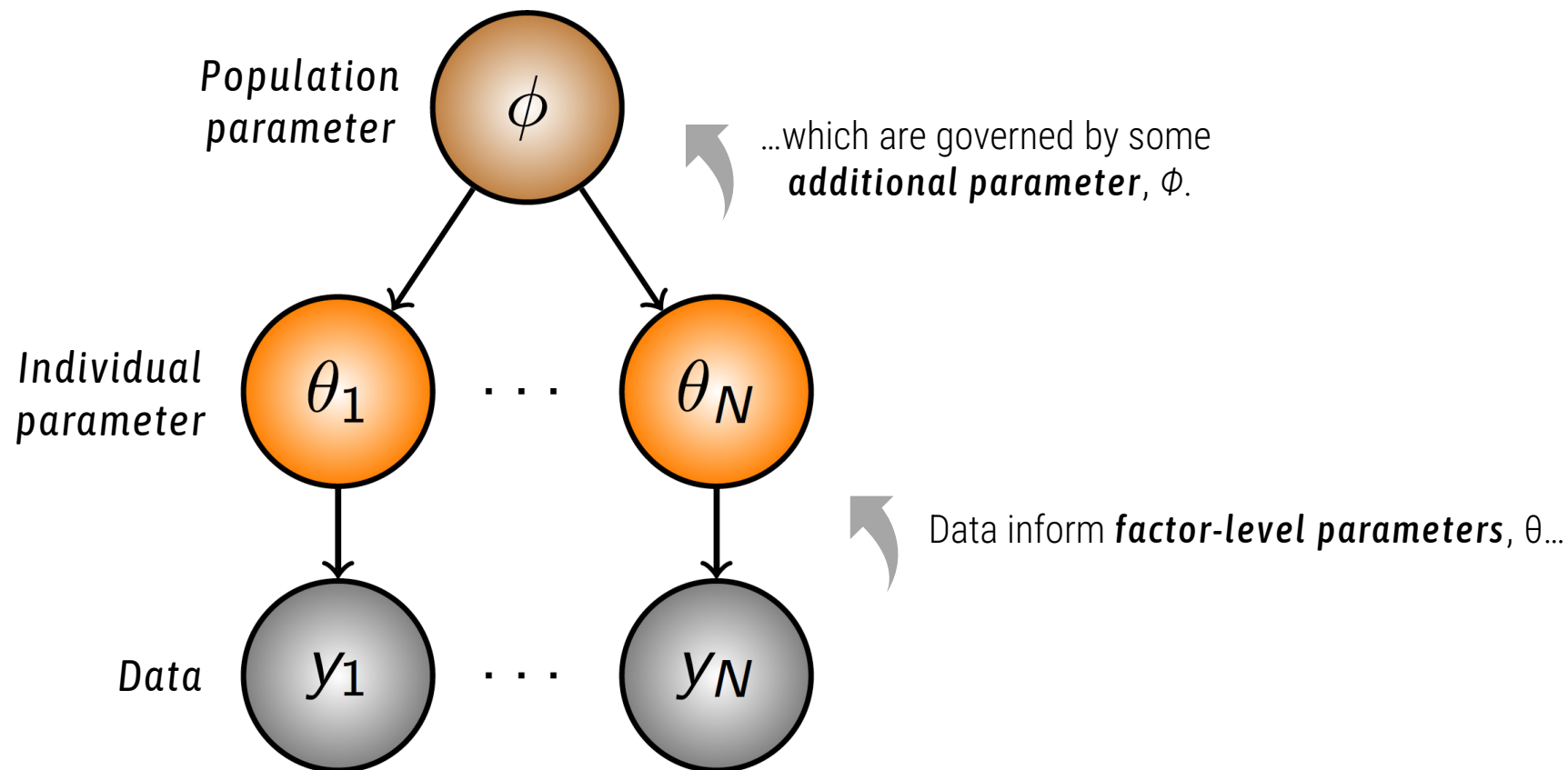


 Adapted from Midway (2008)



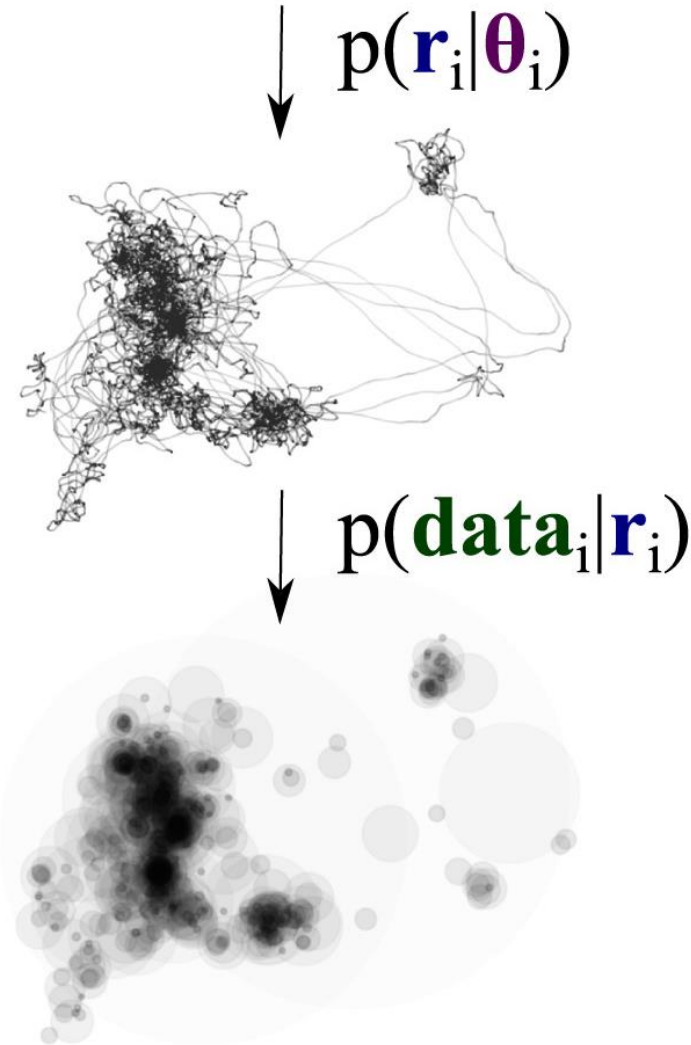
HIERARCHICAL MODELS

How does data inform parameters?



Adapted from Midway (2008)

Trajectory



Data

$$p(\mathbf{DATA} | \boldsymbol{\Theta}) = \prod_i^m \int d\mathbf{r}_i \int d\boldsymbol{\theta}_i \underbrace{p(\mathbf{data}_i | \mathbf{r}_i)}_{\text{observation}} \underbrace{p(\mathbf{r}_i | \boldsymbol{\theta}_i)}_{\text{movement}} \underbrace{p(\boldsymbol{\theta}_i | \boldsymbol{\Theta})}_{\text{population}},$$

DATA = tracking data from m individuals,

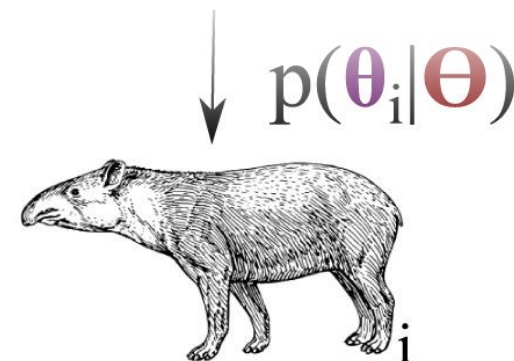
data_i = tracking data from individual i ,

r_i = trajectory of individual i ,

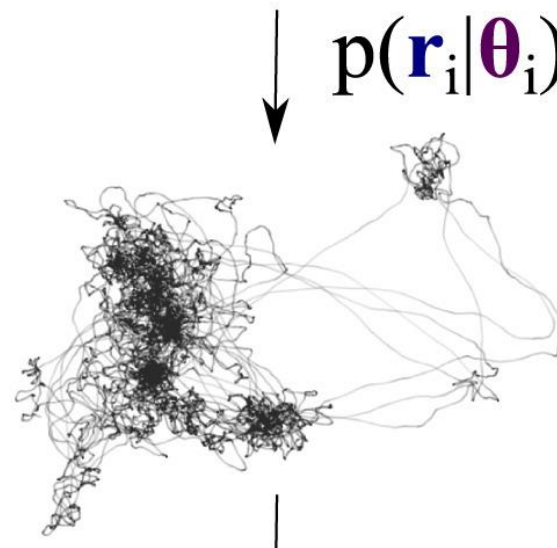
θ_i = movement characteristics (e.g., home range) of individual i ,

Θ = Population parameters (e.g., variation in home ranges).

Individual



Trajectory



$$p(\mathbf{DATA}|\Theta) = \prod_i^m \int d\mathbf{r}_i \int \boldsymbol{\theta}_i \underbrace{p(\mathbf{data}_i|\mathbf{r}_i)}_{\text{observation}} \underbrace{p(\mathbf{r}_i|\boldsymbol{\theta}_i)}_{\text{movement}} \underbrace{p(\boldsymbol{\theta}_i|\Theta)}_{\text{population}},$$

DATA = tracking data from m individuals,

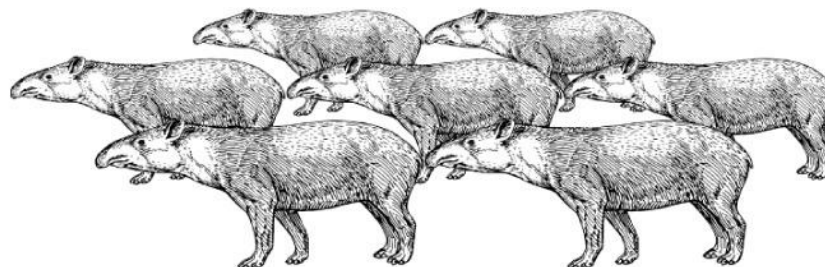
data_i = tracking data from individual i ,

r_i = trajectory of individual i ,

θ_i = movement characteristics (e.g., home range) of individual i ,

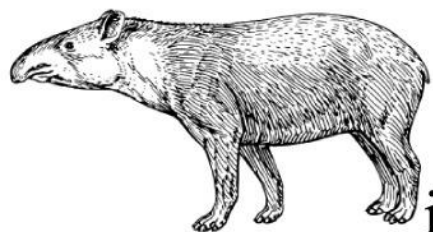
Θ = Population parameters (e.g., variation in home ranges).

Population



↓ $p(\theta_i | \Theta)$

Individual



$$p(\mathbf{DATA} | \Theta) = \prod_i^m \int d\mathbf{r}_i \int d\theta_i \underbrace{p(\mathbf{data}_i | \mathbf{r}_i)}_{\text{observation}} \underbrace{p(\mathbf{r}_i | \theta_i)}_{\text{movement}} \underbrace{p(\theta_i | \Theta)}_{\text{population}},$$

DATA = tracking data from m individuals,

data_i = tracking data from individual i ,

r_i = trajectory of individual i ,

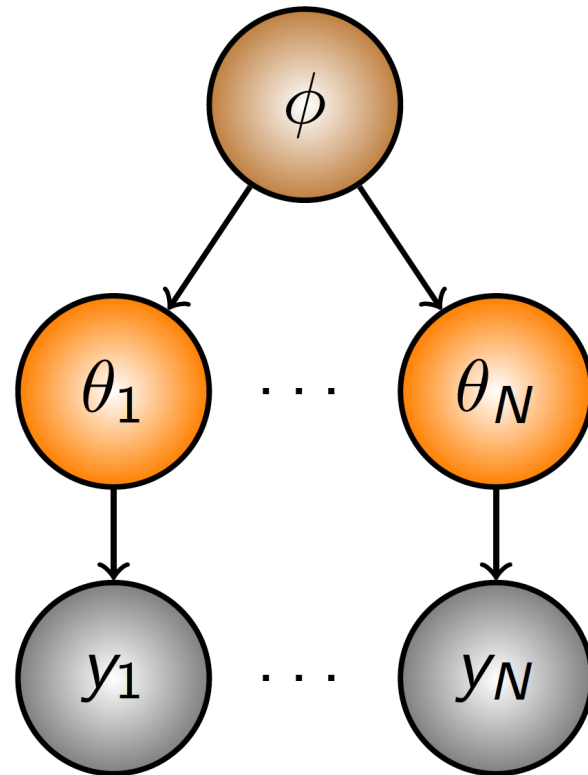
θ_i = movement characteristics (e.g., home range) of individual i ,

Θ = Population parameters (e.g., variation in home ranges).

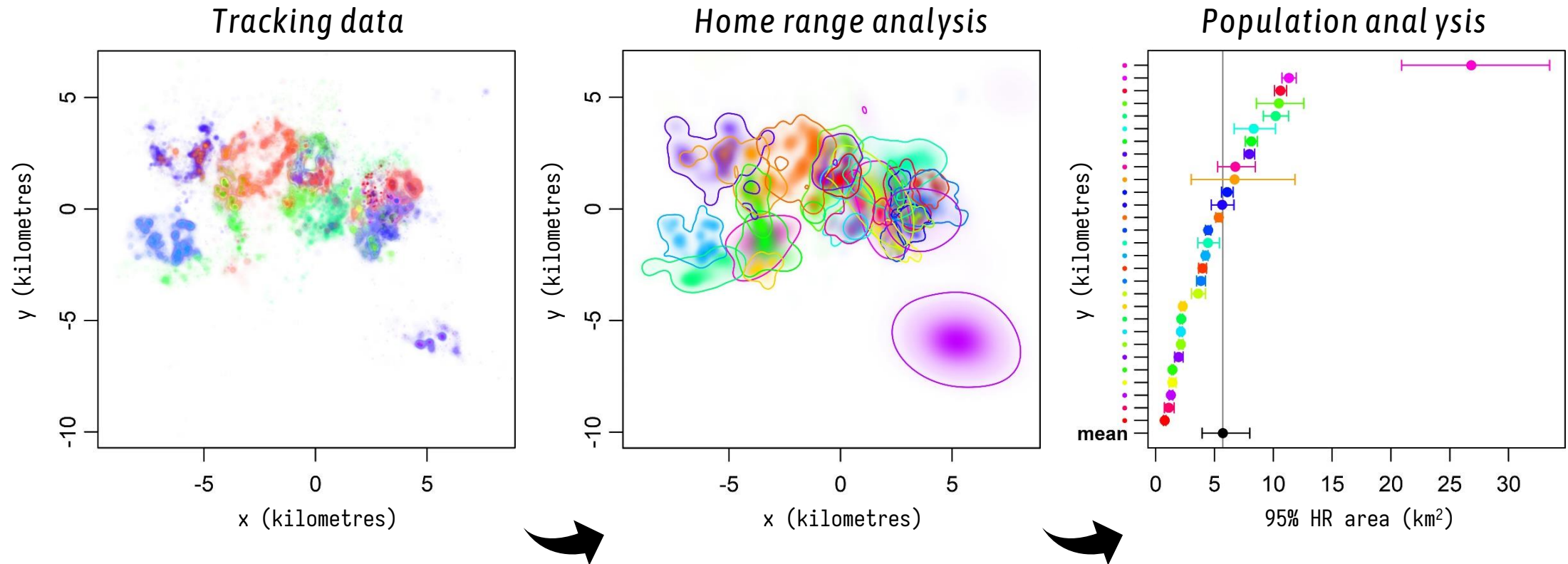
χ^2 inverse-Gaussian (χ^2 -IG) meta-analyses

$\mathcal{O}(1/N)$ bias

$\mathcal{O}(1/m)$ bias



Account for the uncertainties in individual home-range estimates by treating the home-range areas as **unknown latent variables** within a hierarchical model.



Home range area estimates
follow a χ^2 **sampling**
distribution

Population of home-range areas
follows an **inverse-Gaussian**
distribution

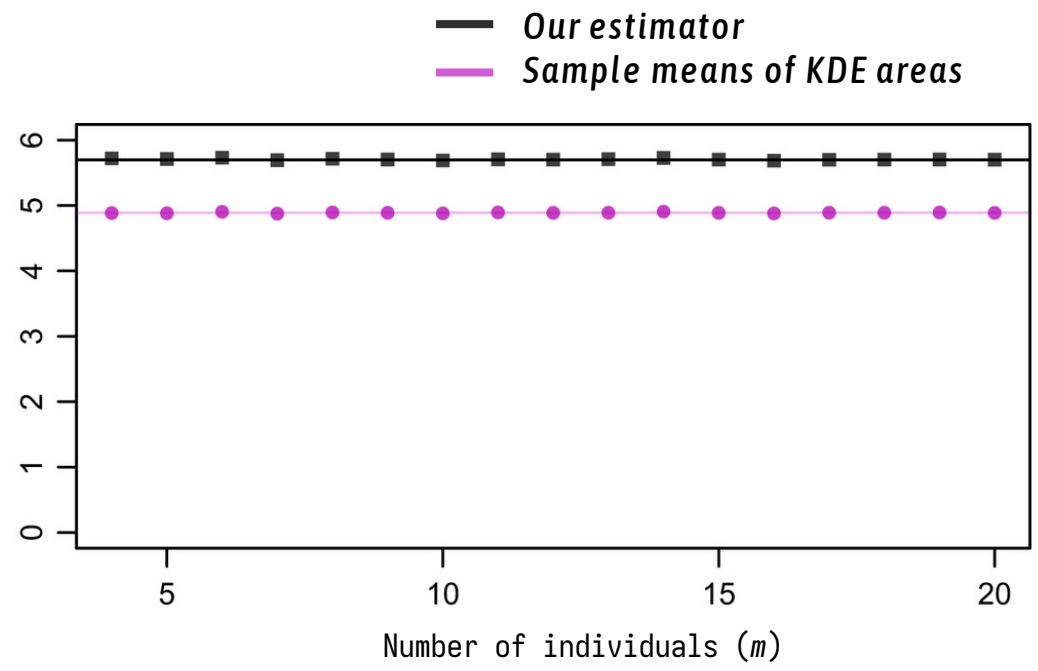
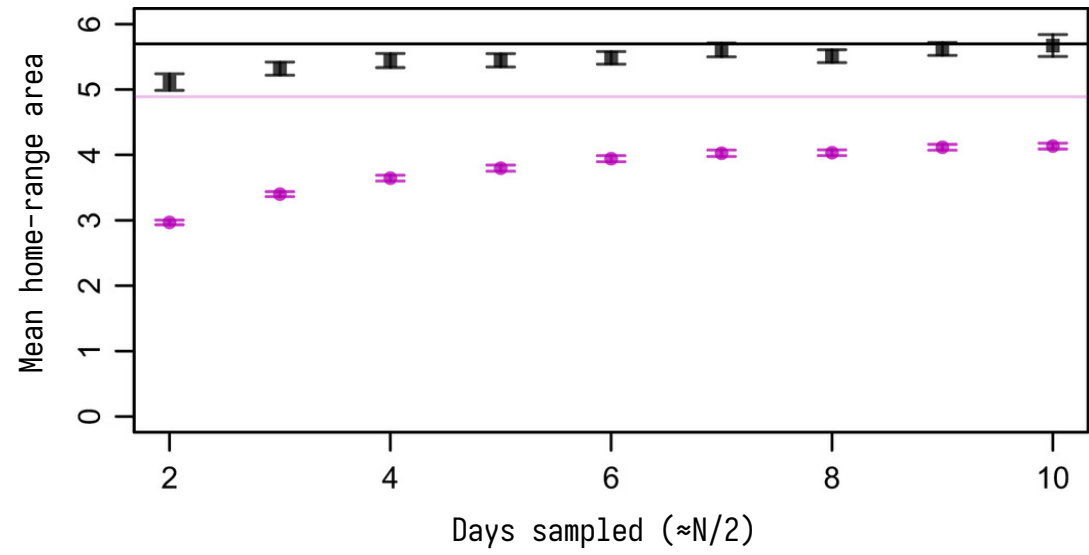
A statistically efficient estimator will **downweight** uncertain estimates relative to more certain estimates in such a way that the estimated mean has a smaller variance.



Lowland tapir
(*Tapirus terrestris*)

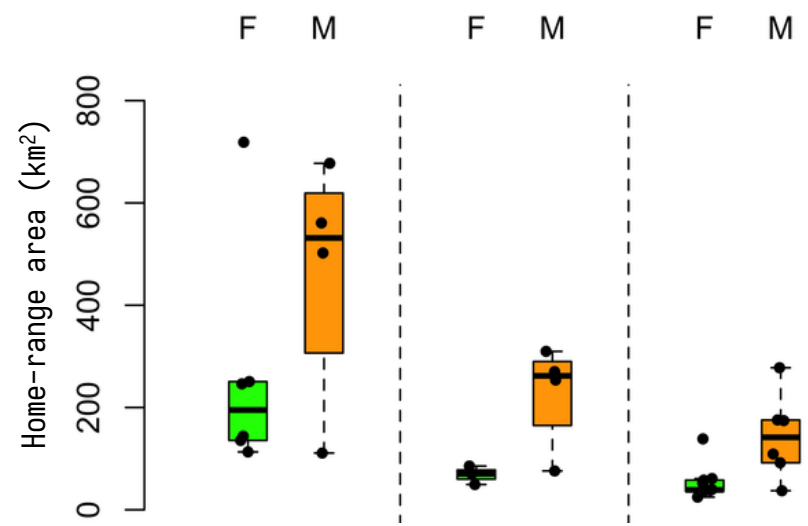


Tapirs have **home range crossing times** (τ_p) of 0.72 days,
(ranging from 0.05–12.8 days)





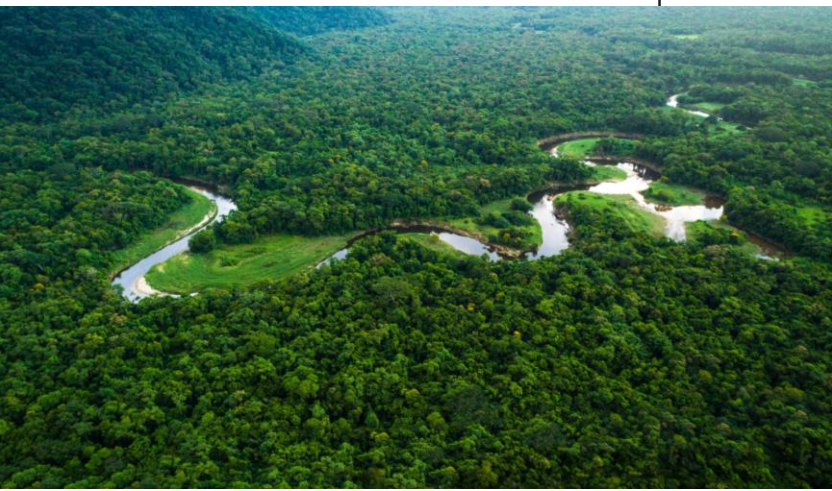
✍ **Morato *et al.* (2016)**
DOI: 10.1093/jmammal/gyab068



Atlantic forest

Amazon

Pantanal





Jaguar
(*Panthera onca*)

NT

Pete Wilcox

What's the **mean home range area**?

Average area used by individuals in a sample

What's the **population distribution**?

Spatial extent of the population as a whole

Methods:

- (A)KDE of population?
- Union of (A)KDEs?
- Mean of (A)KDEs?

Dual challenge:

Individual temporal autocorrelation
Population variation

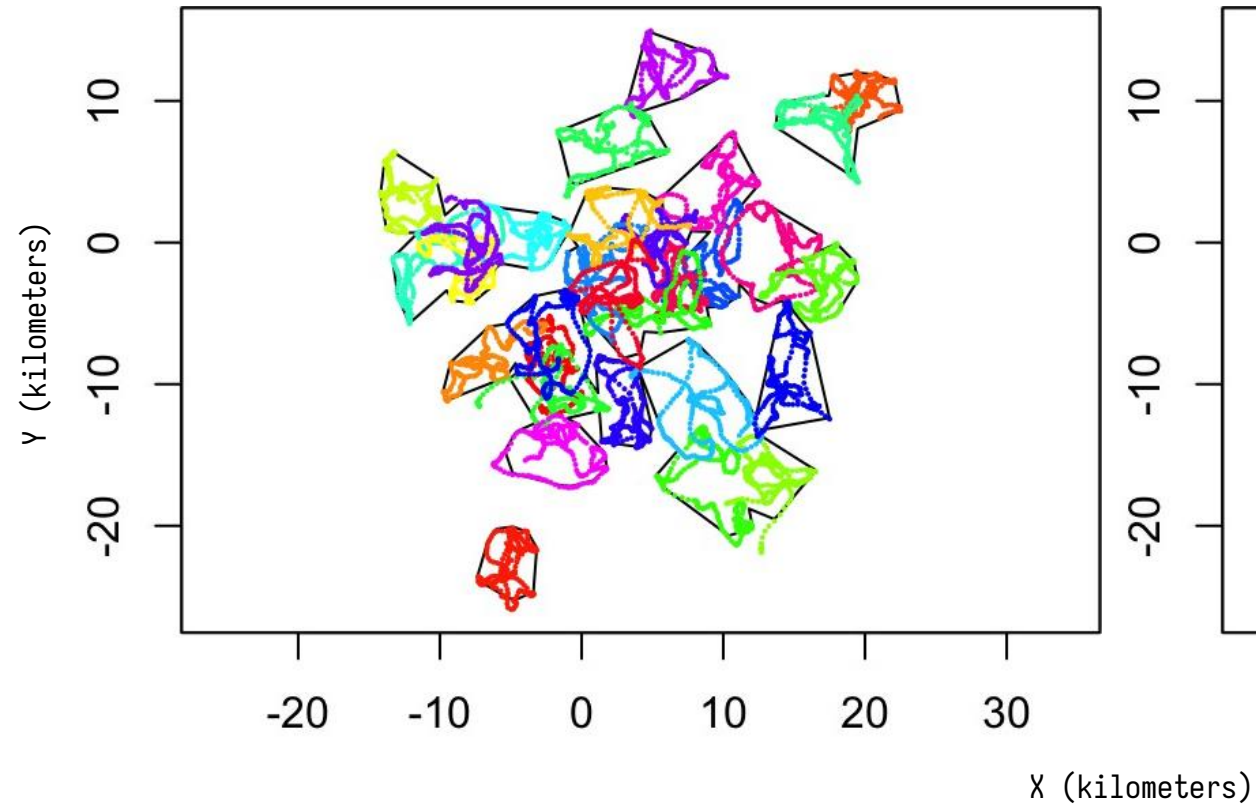
Sample
Tracked individuals

≠

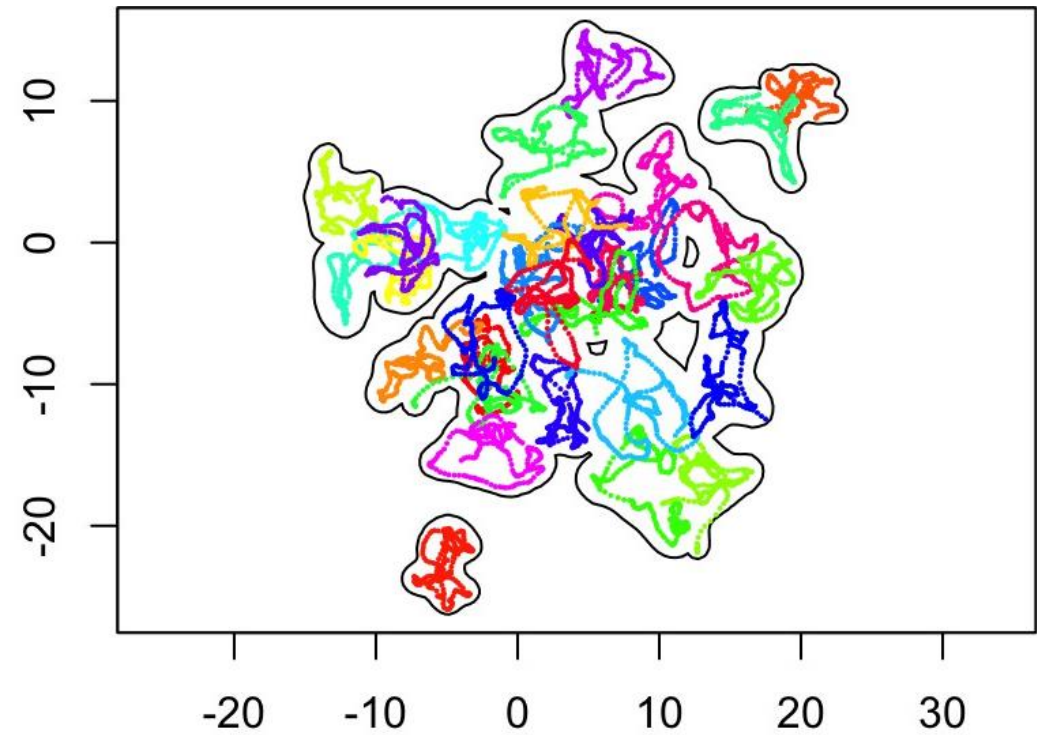
Population
Tracked + untracked individuals



Merged **MCP** estimate

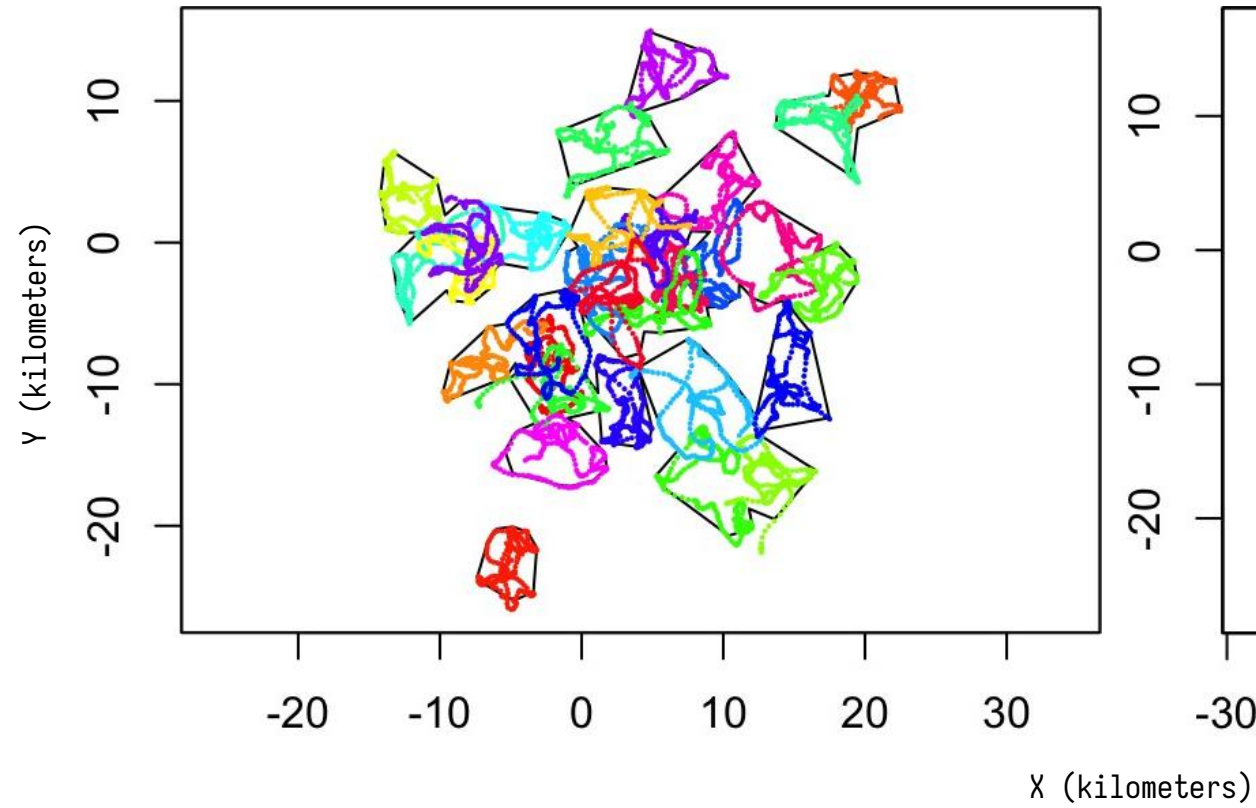


Merged **KDE** estimate

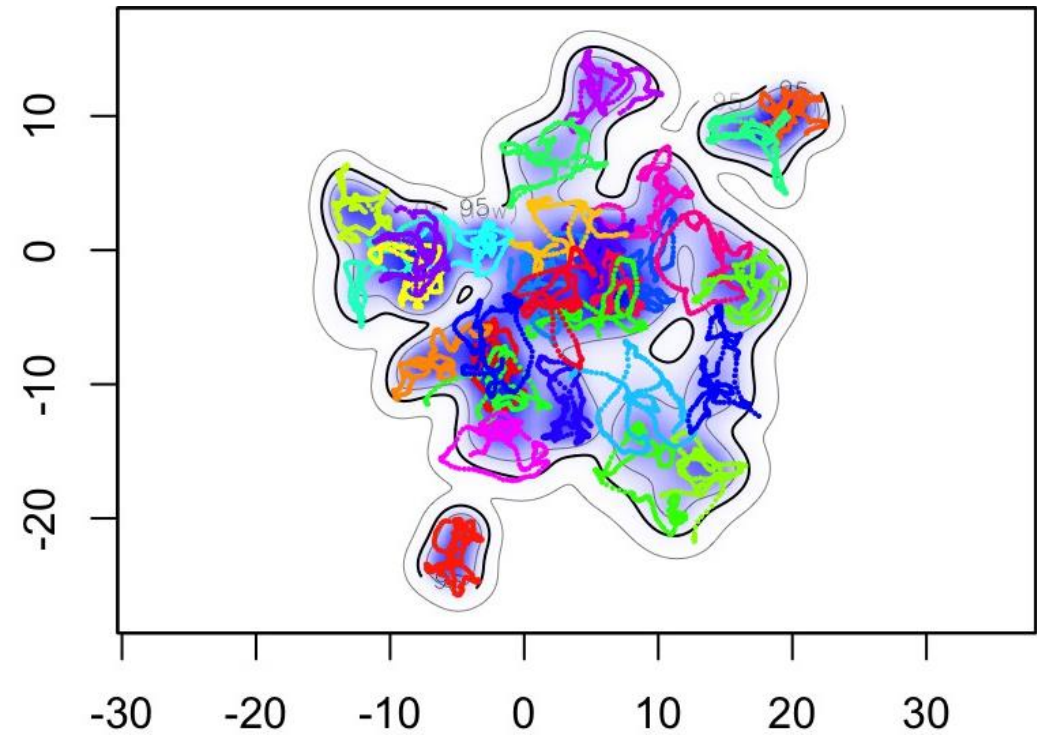




Merged **MCP** estimate



Merged **AKDE** estimate





(A)KDE is a weighted average of kernels, where the optimal H minimizes the MISE:

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

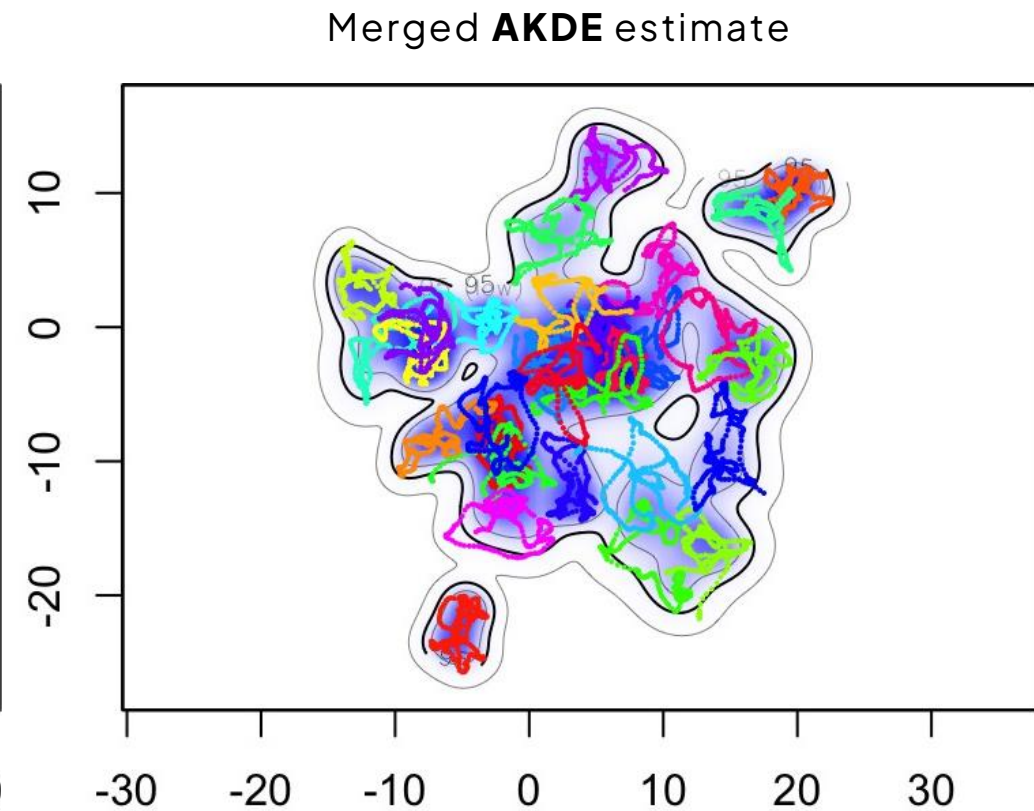
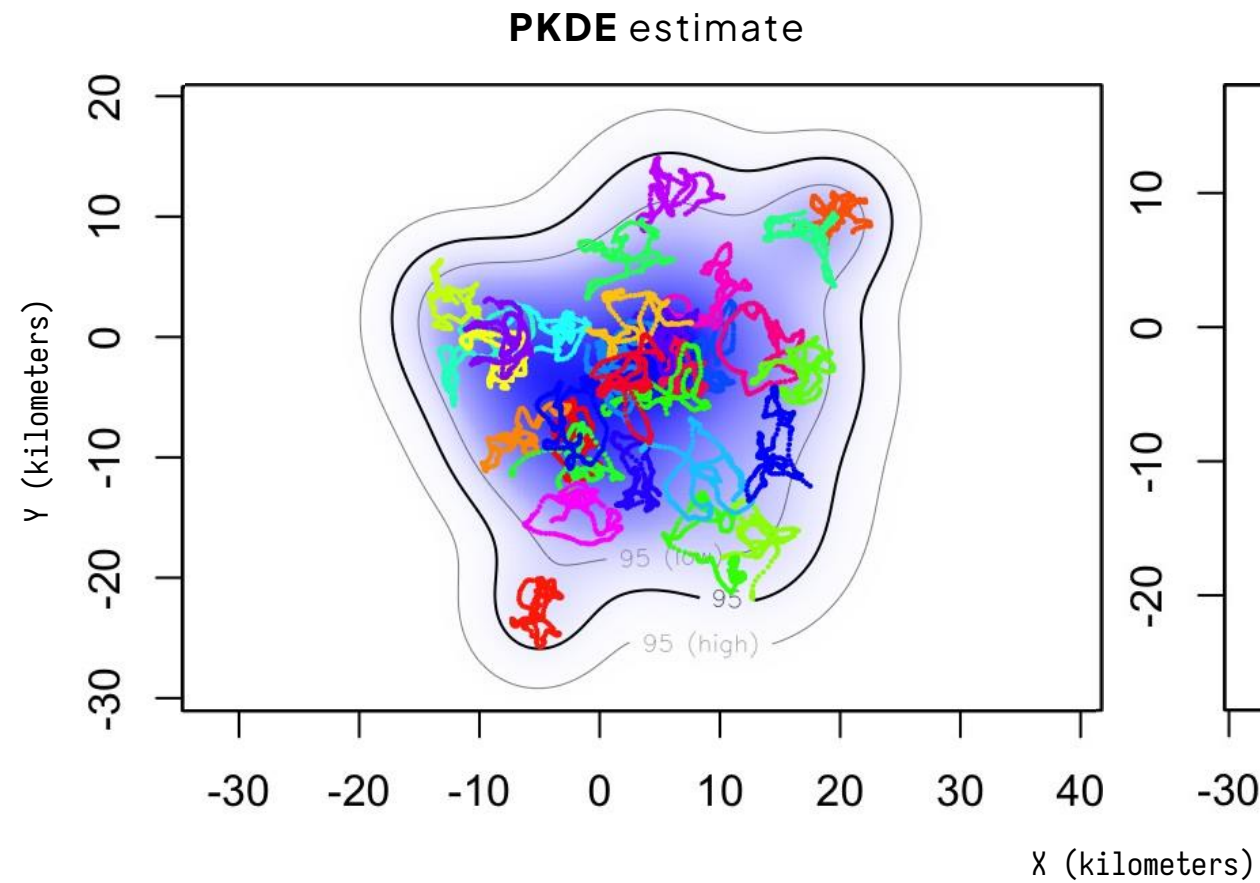
$p(\mathbf{x})$ = approximation (e.g., Gaussian reference function)

PKDE

$$\text{MISE}[\mathbf{H}] = \mathbb{E} \left[\iint (\hat{p}_{\text{pop.}}(\mathbf{x}|\mathbf{H}) - p_{\text{pop.}}(\mathbf{x}))^2 d\mathbf{x} \right]$$

$$\hat{p}_{\text{pop.}}(\mathbf{x}|\mathbf{H}) = \sum_{\text{ind.}} \sum_t w_{\text{ind.}}(t) \kappa(\mathbf{x} - \mathbf{x}_{\text{ind.}}(t) | \mathbf{H}_{\text{ind.}})$$

$p_{\text{pop}}(\mathbf{x})$ requires a hierarchical approximation





Saturation curves

