FACULTY OF SCIENCES

DEPARTMENT BIOLOGY
RESEARCH GROUP TERRESTRIAL ECOLOGY

(TEREC) CODING CLUB

Session 1 – 11/03/2024

GHENT UNIVERSITY

Terrestrial Ecology

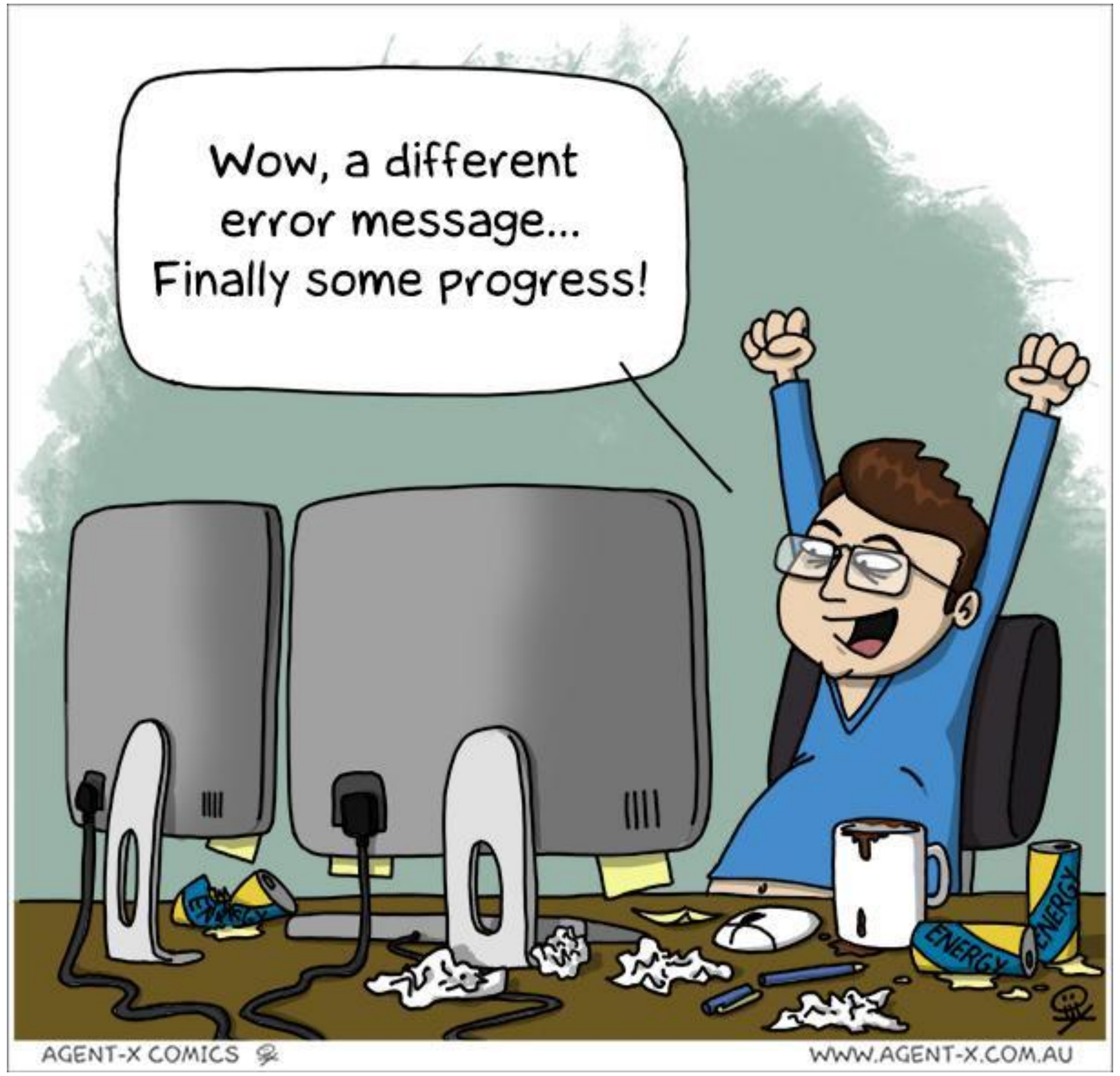
OVERVIEW

Charlotte

GHENT UNIVERSITY

OVERVIEW OF THIS SESSION

- Introduction
- Basics of Bayesian
- Brms
- INLA
- ABC
- Varia



GHENT
UNIVERSITY

BASICS OF BAYESIAN

Charlotte

GHENT UNIVERSITY

Basics of Bayesian

Who has ever worked with Bayesian statistics?

Basics of Bayesian

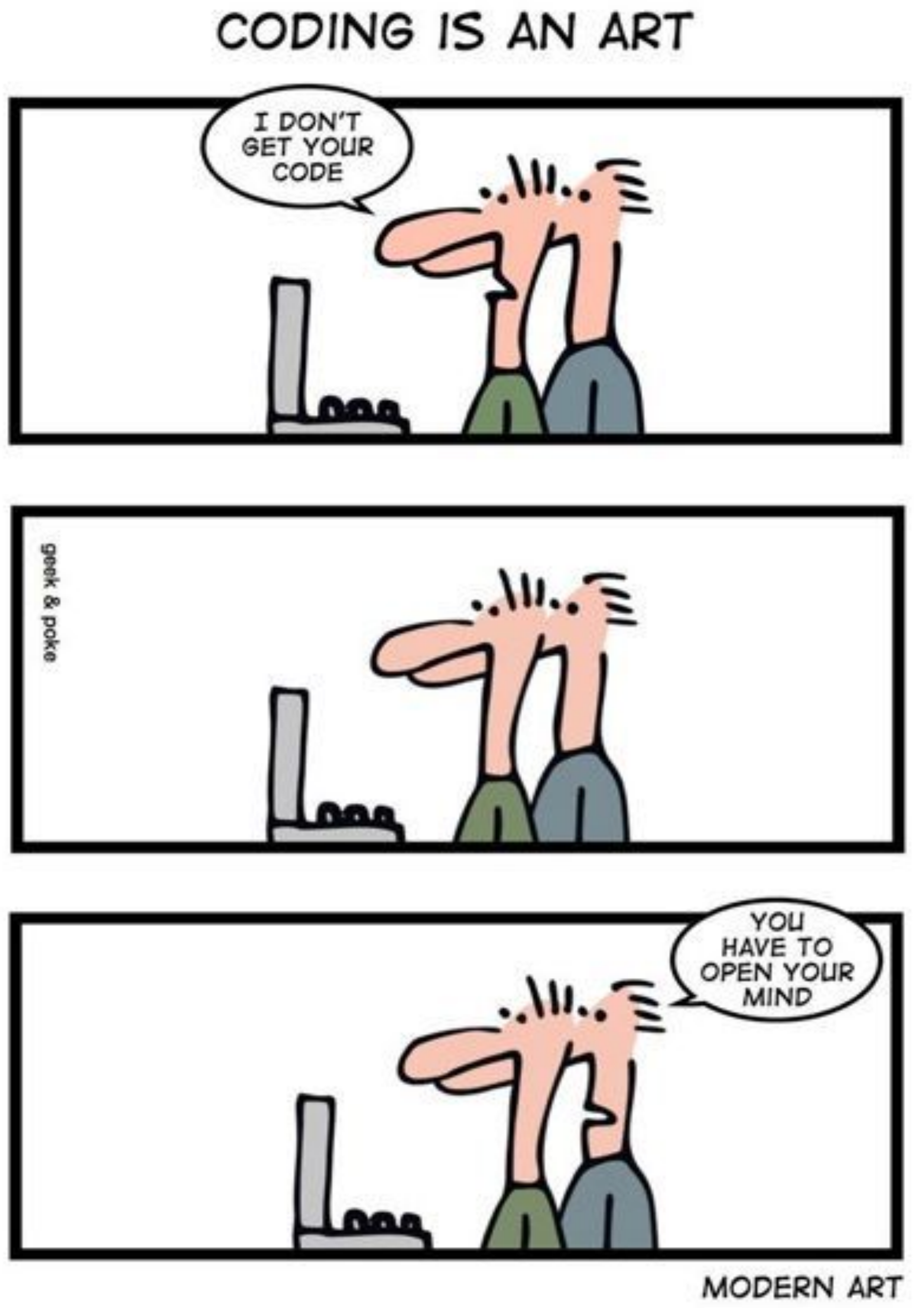
Who has ever worked with Bayesian statistics?

Who can explain what Bayesian is?

Basics of Bayesian

Who has ever worked with Bayesian statistics?

Who can explain what Bayesian is?



GHENT UNIVERSITY

Basics of Bayesian

When tossing a coin, we say “chances of *head* are 0,5”.
How to you interprete this?

-Raise hands on 3-

GHENT UNIVERSITY

Basics of Bayesian

When tossing a coin, we say “chances of *head* are 0.5”.
How to you interprete this?

-Raise hands on 3-

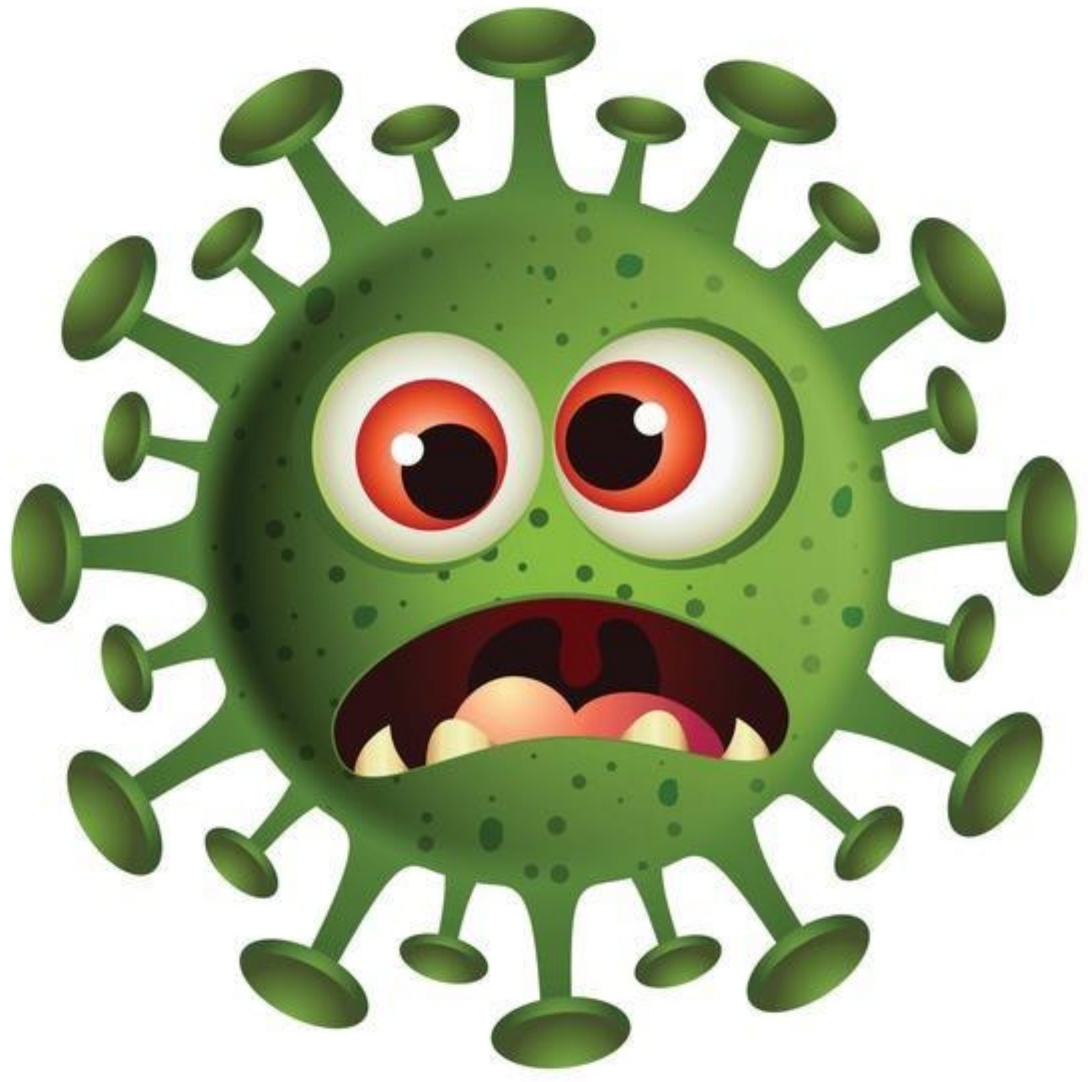
Left hand:
When I toss multiple times in a row, it will result in *head* in about 50% of the tosses.

Right hand:
Chances of getting *head* or *coin* are equally probable.

Basics of Bayesian

Imagine going to the doctors. You were tested positive for a very rare disease. You can ask only one question, which would it be?

-Raise hands on 3-



GHENT UNIVERSITY

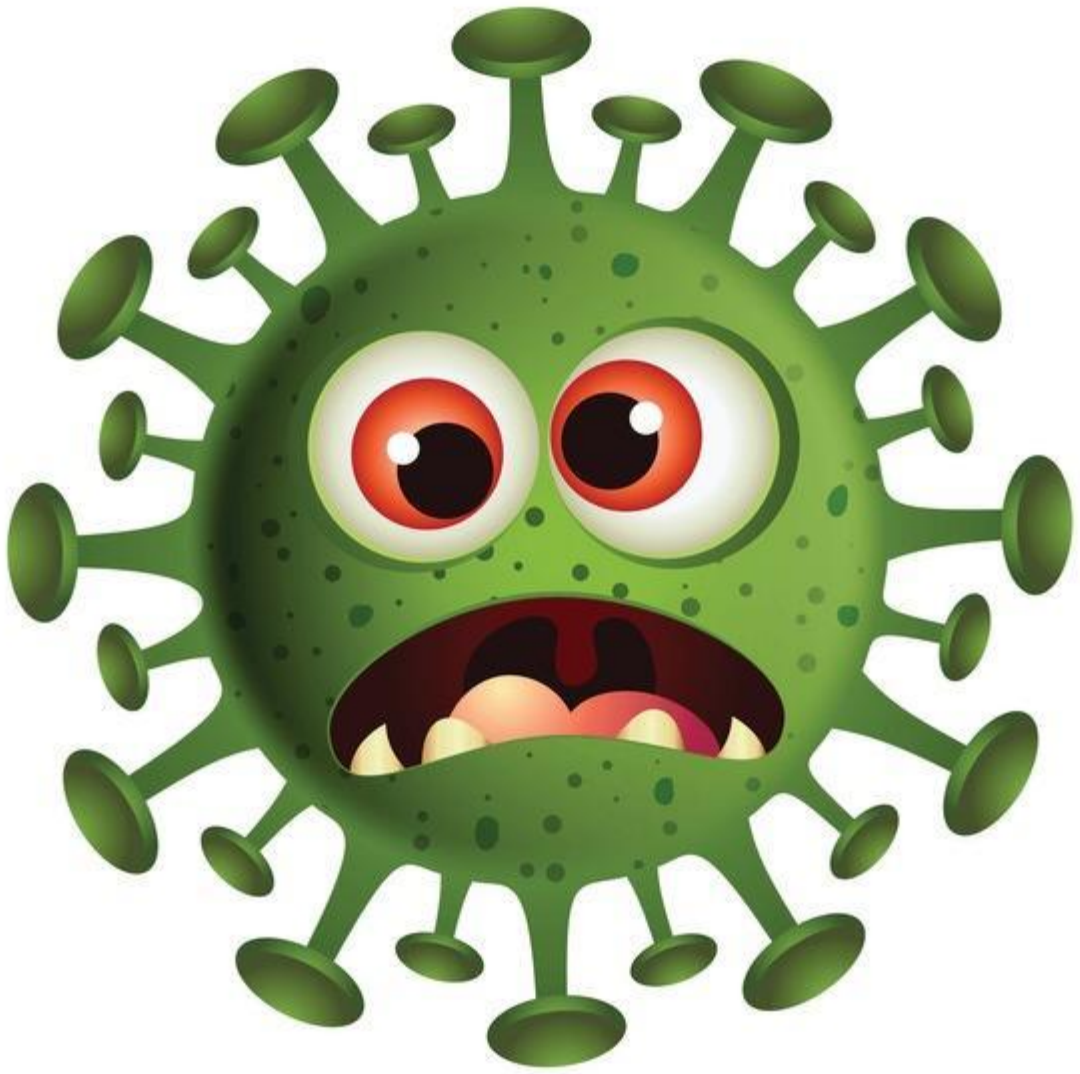
Basics of Bayesian

Imagine going to the doctors. You were tested positive for a very rare disease. You can ask only one question, which would it be?

-Raise hands on 3-

Left hand:
If I don’t have the disease, what would the chances be on a (false) postive result?

Right hand:
What are the chances I really have this disease?



GHENT UNIVERSITY

Basics of Bayesian

What is it?

“Bayesian statistics apply probabilities to statistical problems to update prior beliefs in light of the evidence of new data. The probability expresses a degree of belief in a specific event.”

Basics of Bayesian

What is it?

"Bayesian statistics apply probabilities to statistical problems to update prior beliefs in light of the evidence of new data. The probability expresses a degree of belief in a specific event."

→ Used in physics, cancer research, ecology, psychology…

→ Bayesian statistics vs. frequentist methods

What is it?

Bayesian statistics vs. frequentist methods

A chance or probability …

 Bayesian: …measures the relative probability of an event

Basics of Bayesian

What is it?

Bayesian statistics vs. frequentist methods

A chance or probability …

- *Bayesian: …measures the relative probability of an event*
- *Frequentist: … relative frequency on the long term of repeated events.*

Basics of Bayesian

What is investigated?

- *Bayesian: What is the chance that a hypothesis is correct, given the data we observe?*

Basics of Bayesian

What is investigated?

- *Bayesian: What is the chance that a hypothesis is correct, given the data we observe?*
- *Frequentist: If the hypothesis is incorrect, what is the chance that we observe this or more extreme data?*

Basics of Bayesian

What is investigated?

- *Bayesian: There is prior information and conclusions are continuously adapted as soon as new data is available.*

Basics of Bayesian

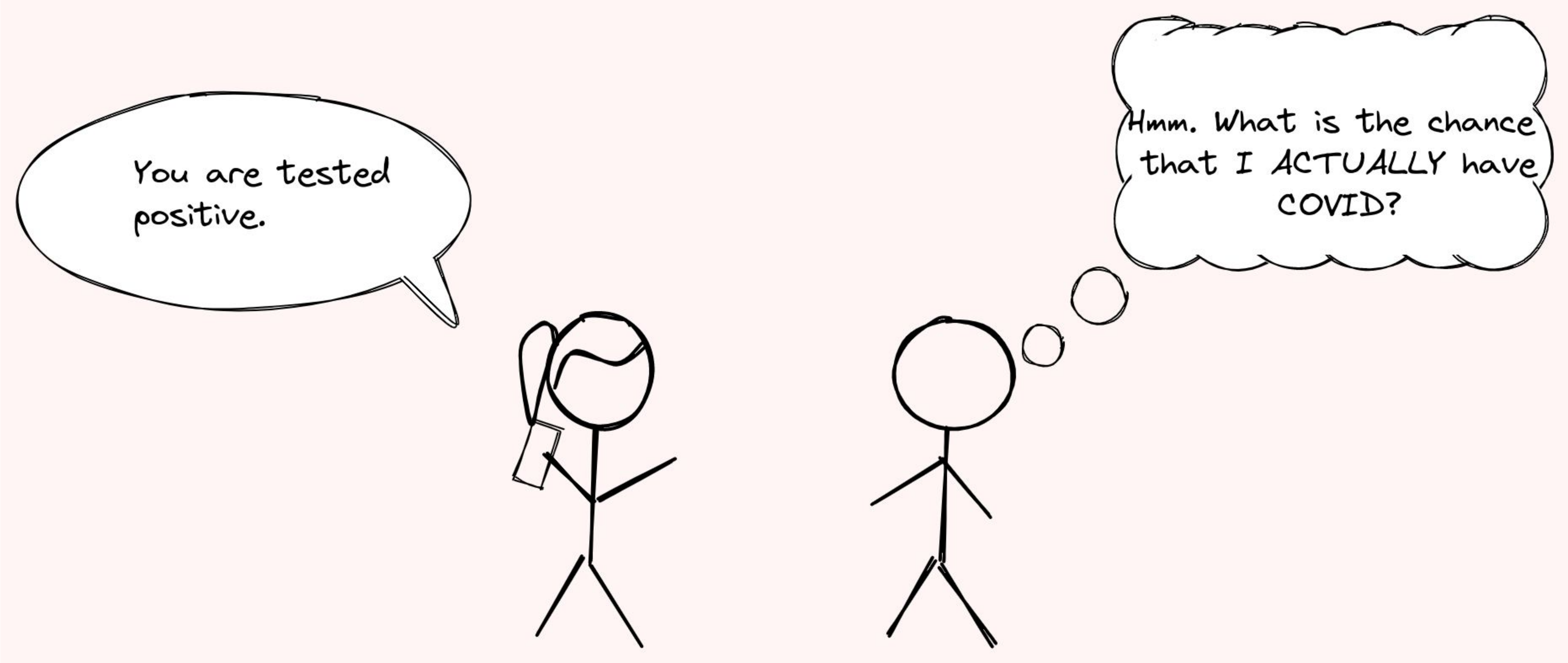
What is investigated?

- *Bayesian: There is prior information and conclusions are continuously adapted as soon as new data is available.*
- *Frequentist: Data is set and observed.*

Basics of Bayesian

Bayesian statistics vs. frequentist methods

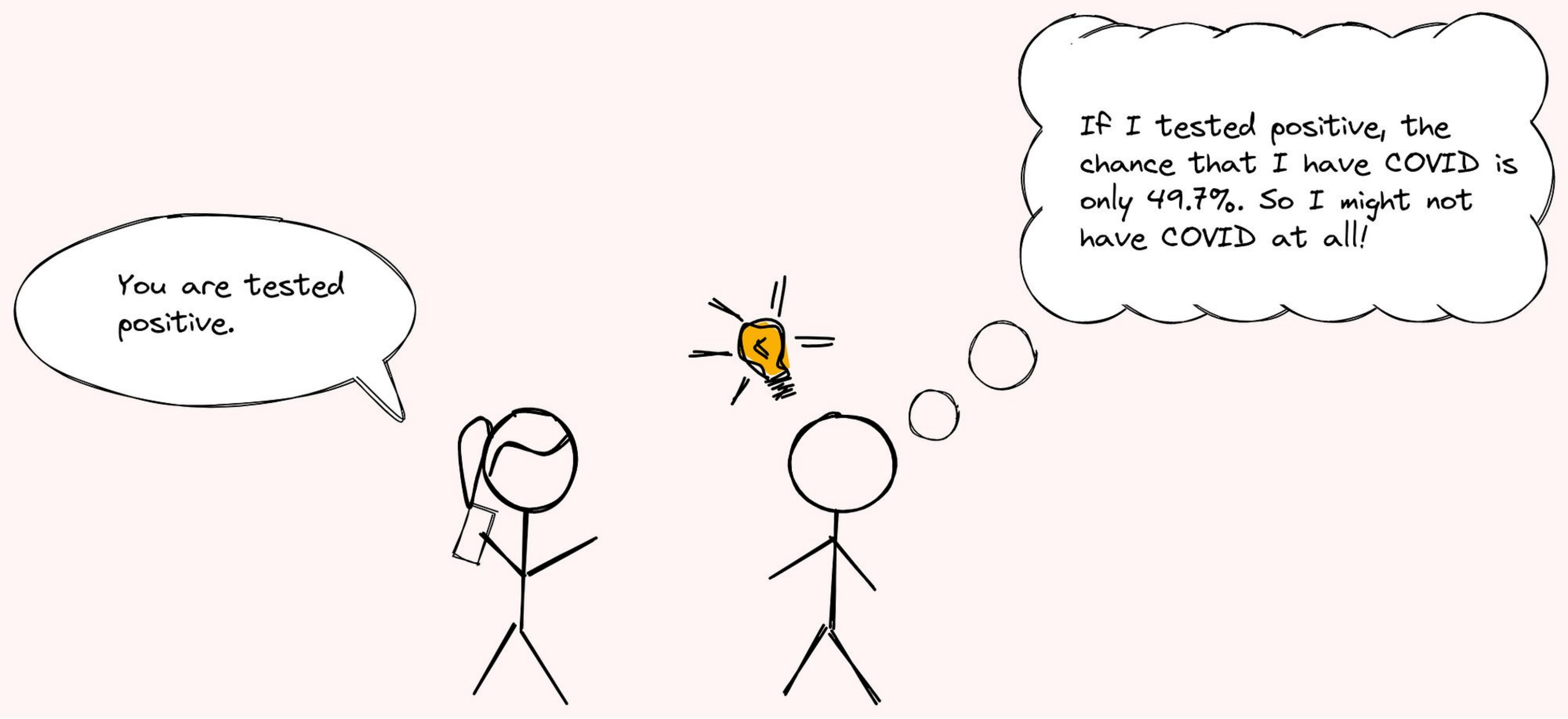
- Knowing these definitions
- Are our brains already Bayesian?
- Bayes' theorem



Basics of Bayesian

Bayesian statistics vs. frequentist methods

- Knowing these definitions
- Are our brains already Bayesian?
- Bayes' theorem



GHENT UNIVERSITY

Terminology

Frequentist

P-value

Confidence interval

Power

Significant

Basics of Bayesian

Terminology	Terminology
Frequentist	***Bayesian***
P-value	*Credible interval*
Confidence interval	*Prior*
Power	*Posterior*
Significant	*(Likelihood)*

Basics of Bayesian

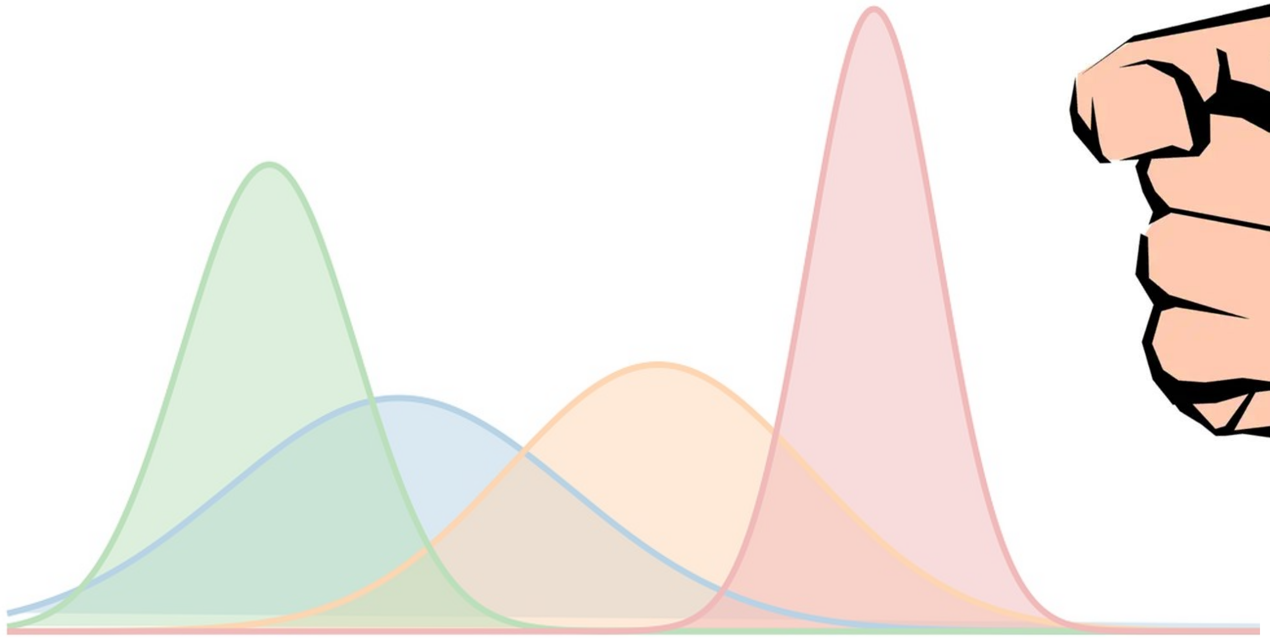
Terminology

Likelihood

Often associated with Bayesian

Helps to understand Bayesian

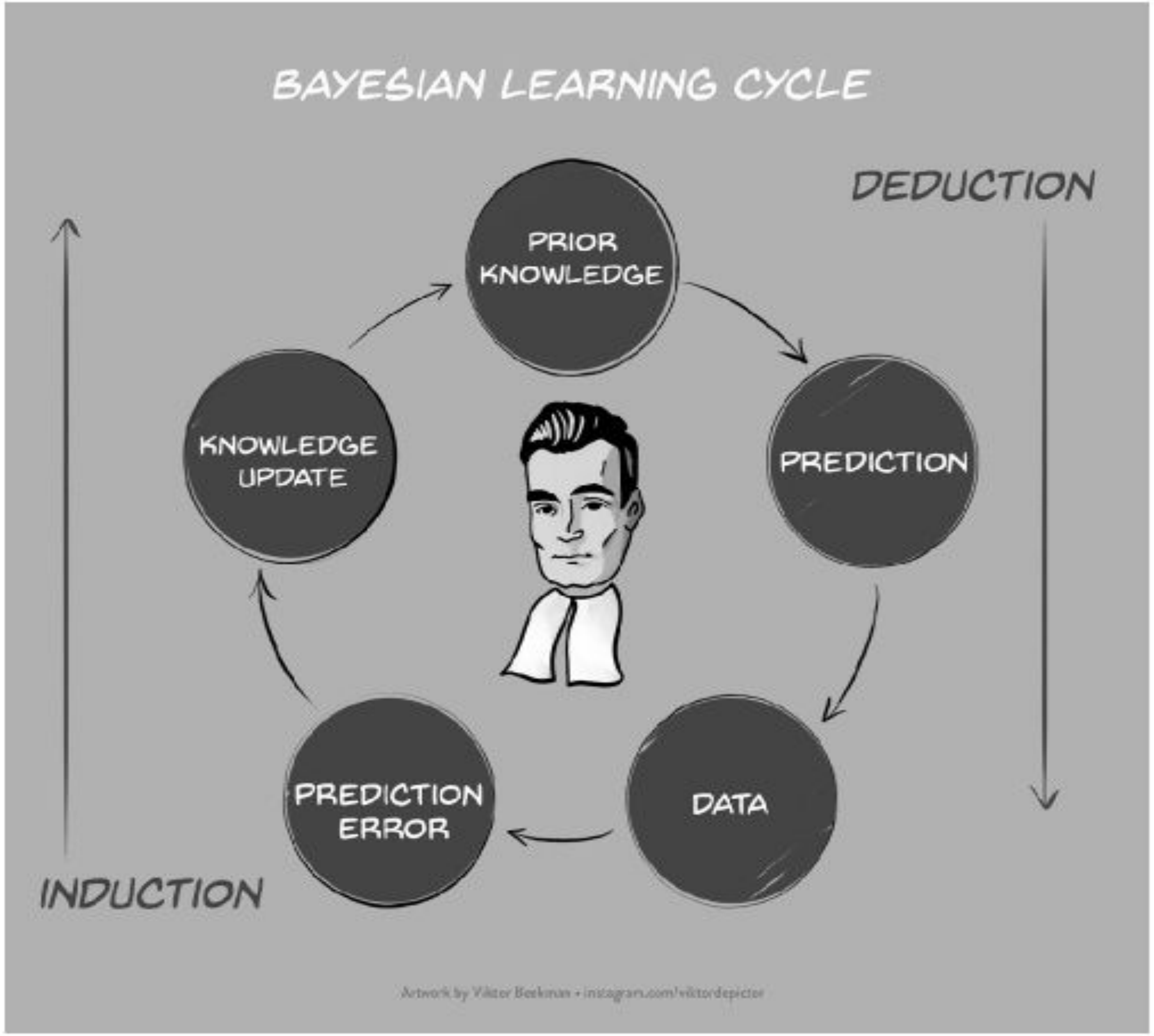
Not only Bayesian!



Basics of Bayesian

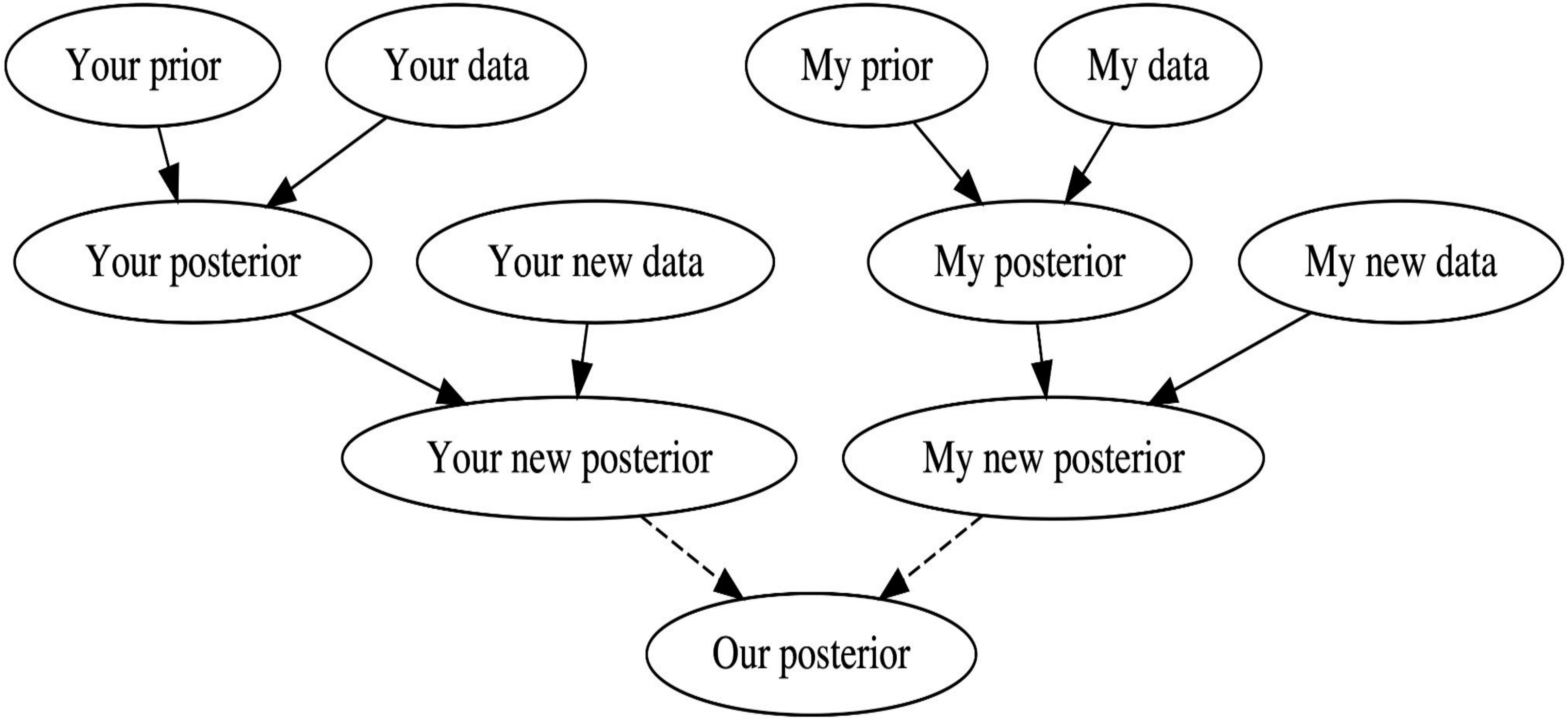
Terminology

Prior



GHENT UNIVERSITY

Basics of Bayesian



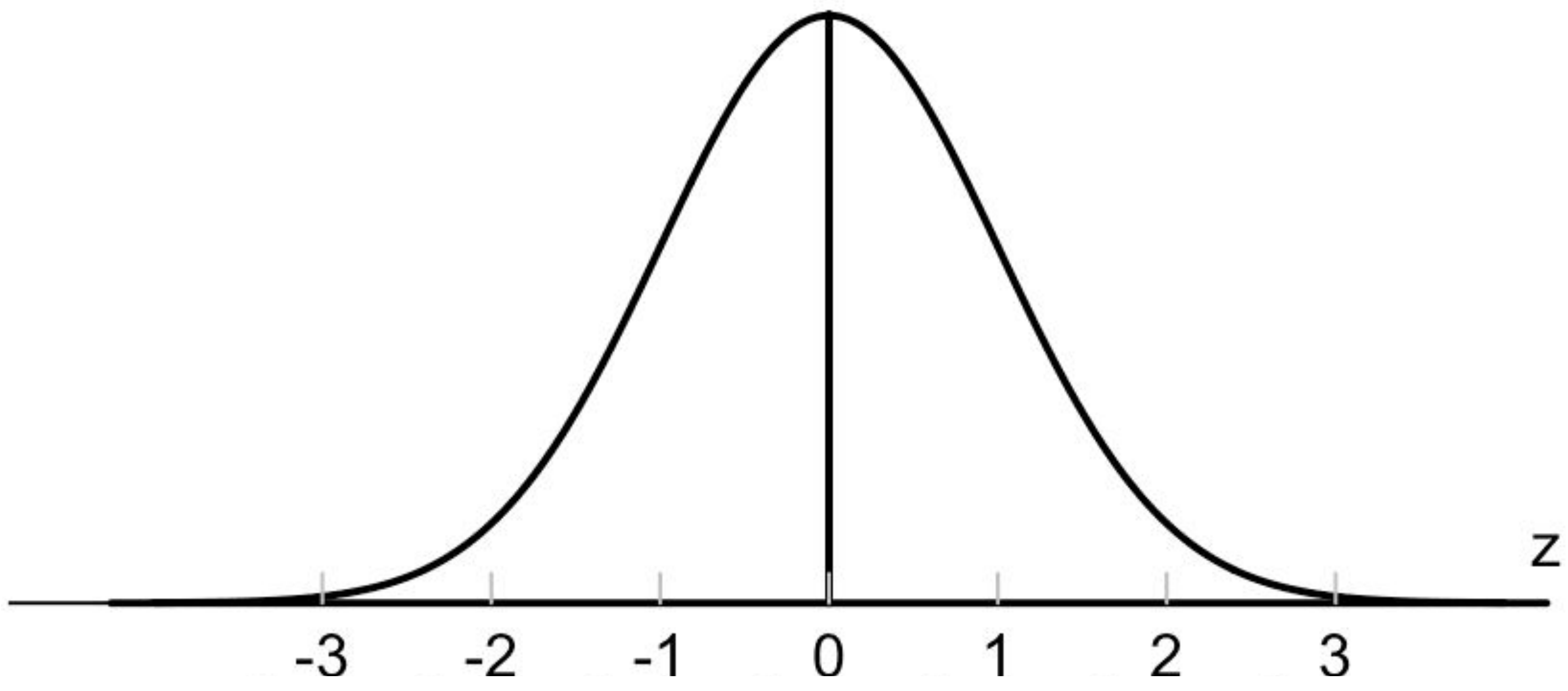
GHENT UNIVERSITY

BRMS

Frederik

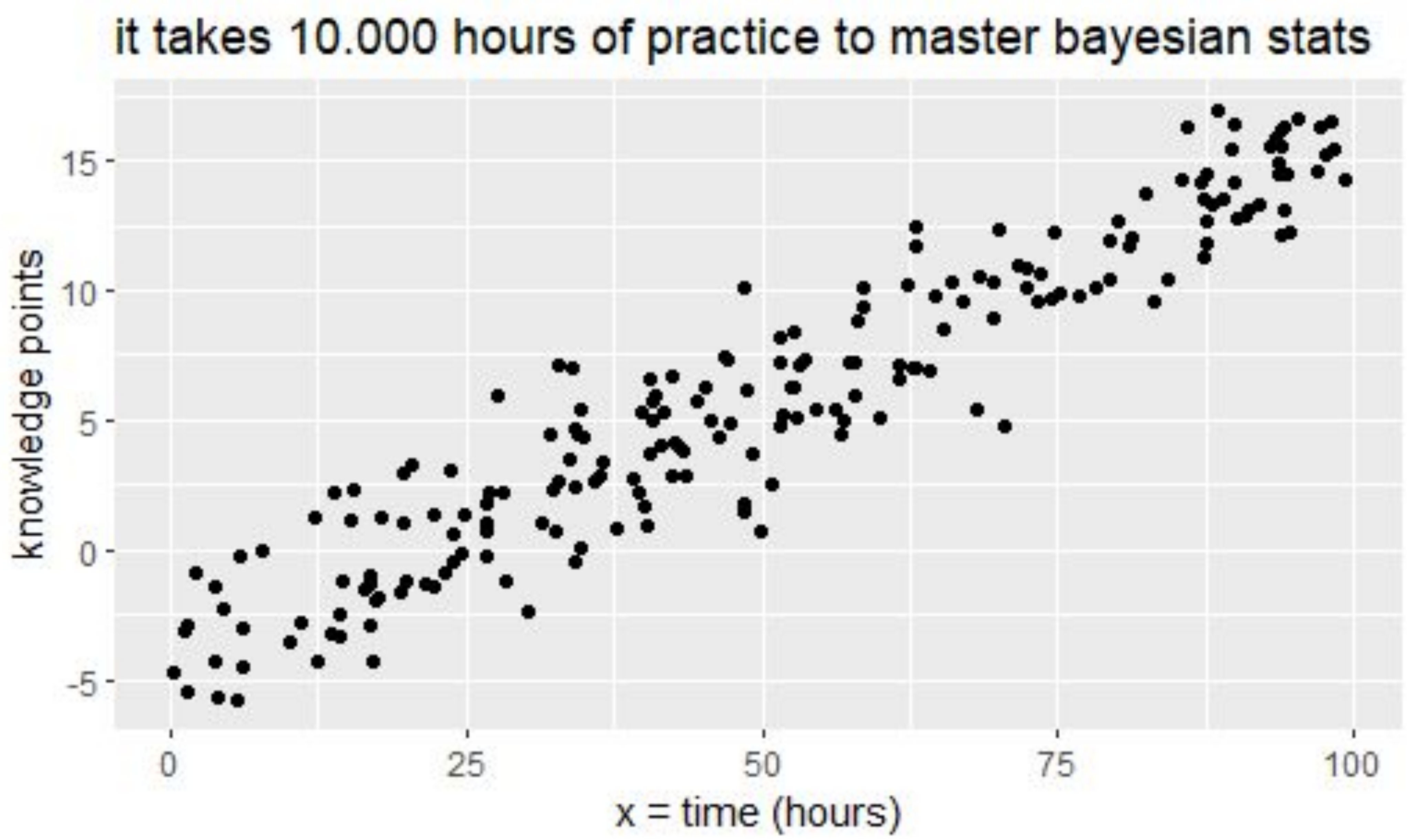
Bayesian Regression Modelling using Stan

– R package using Stan



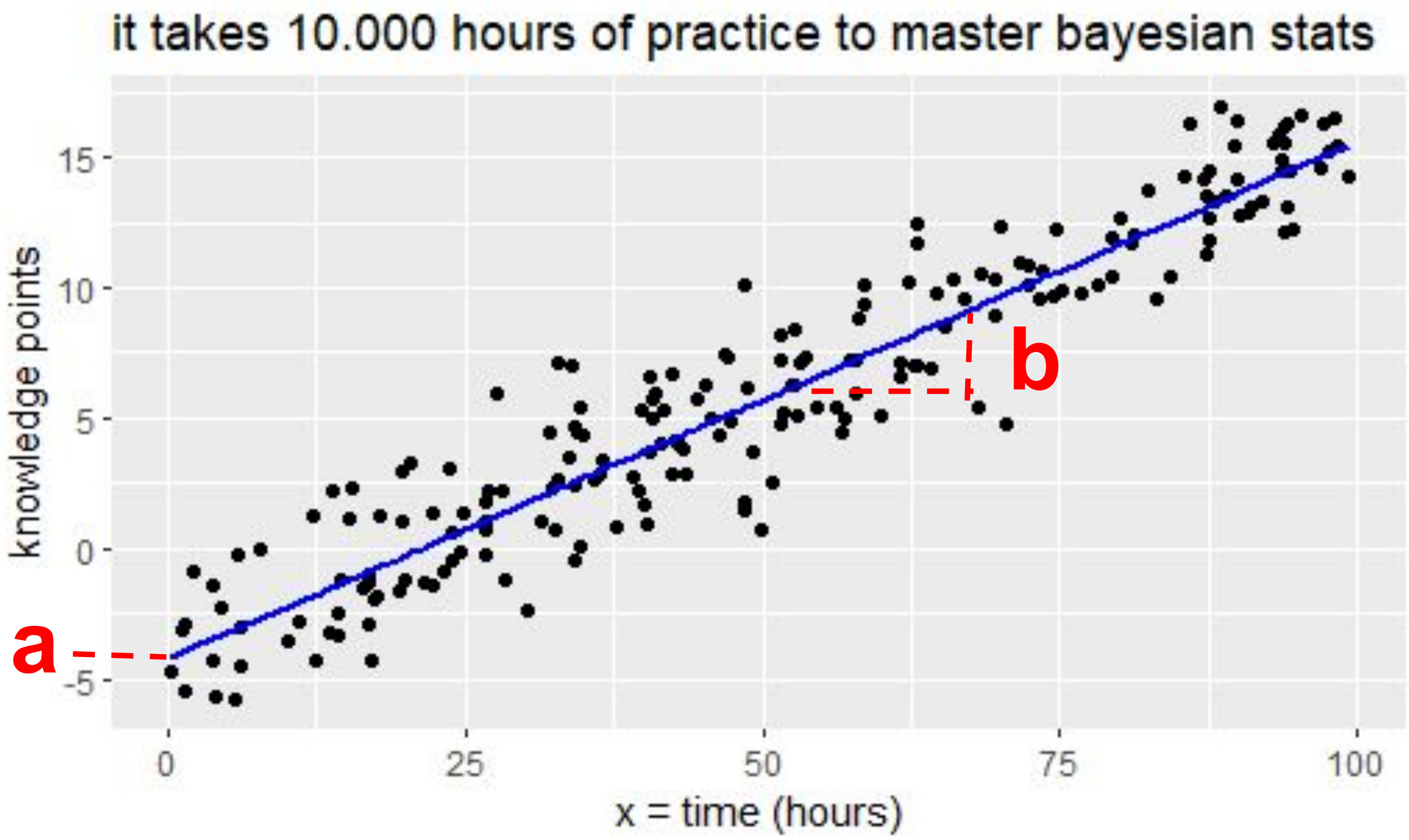
Regression Modelling

$$Y \sim a + b * X + \ldots$$



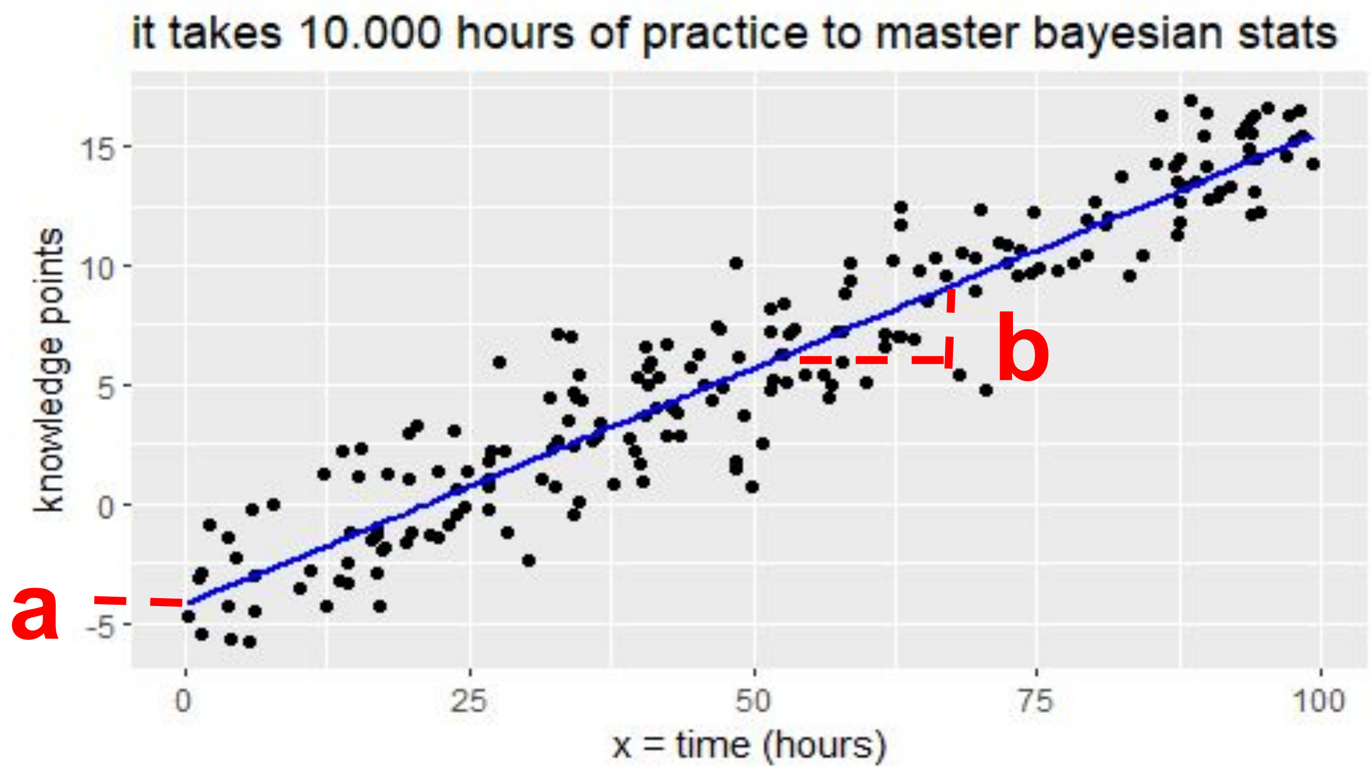
Bayesian Regression Modelling

$$Y \sim a + b * X + \ldots$$



Bayesian Regression Modelling

$$Y \sim a + b * X + \ldots$$

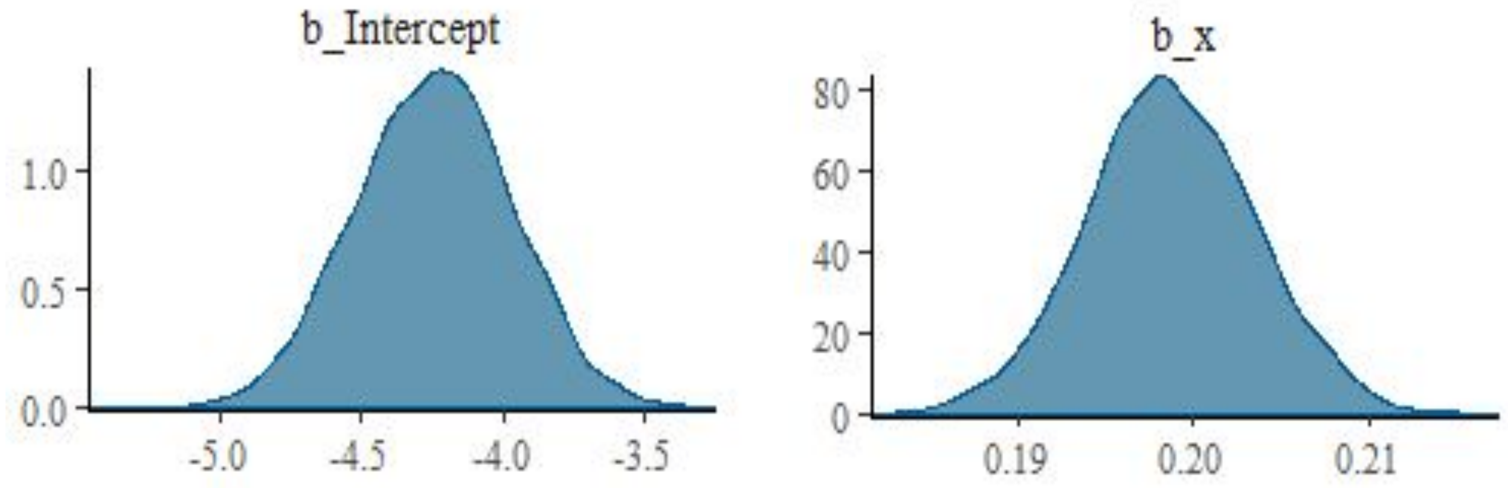


Frequentist model

```
Call:
lm(formula = y ~ x, data = sim)

Coefficients:
(Intercept)            x
    -4.1201       0.1987
```

Bayesian model



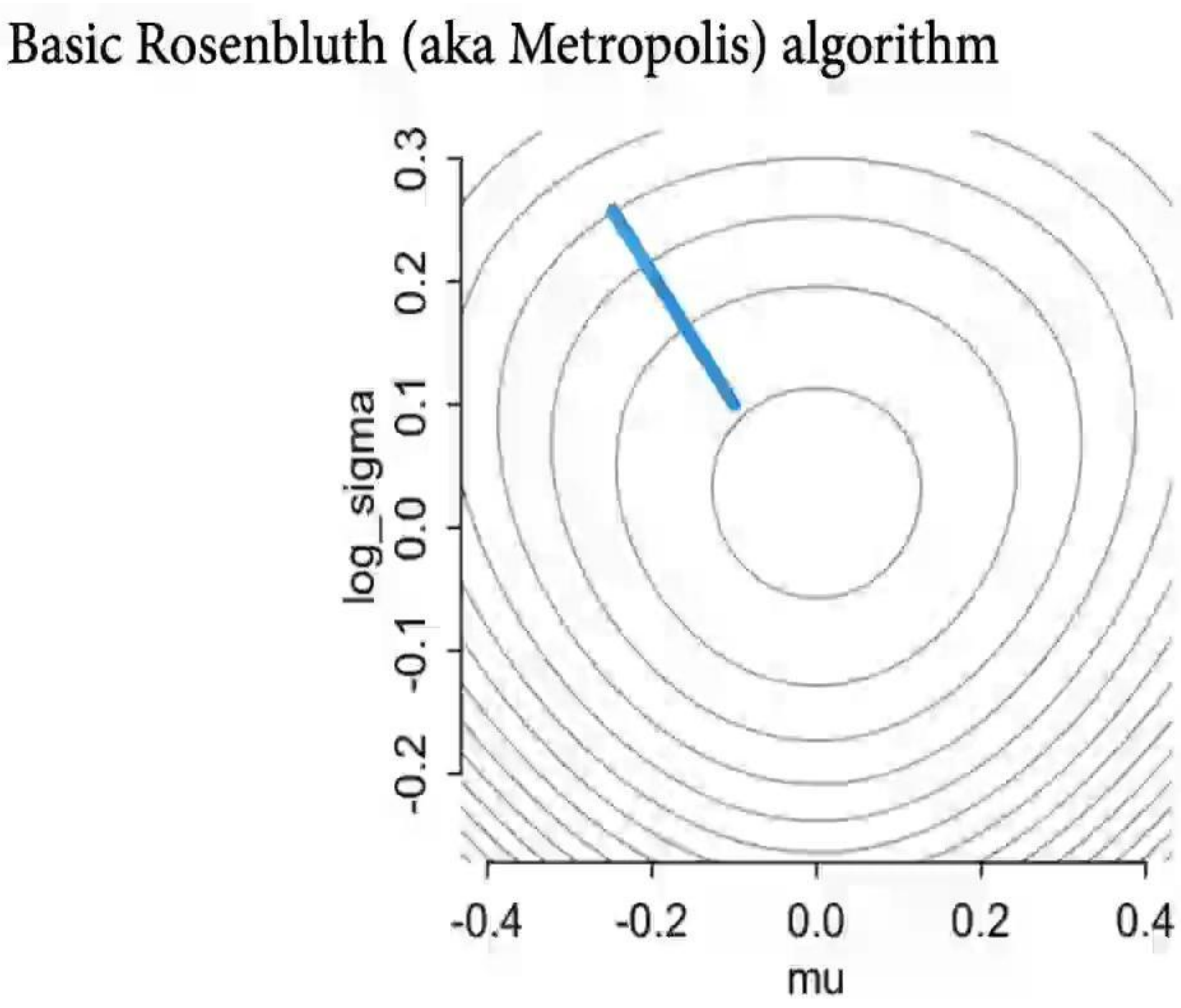
How to get a posterior distribution?

– Calculation of whole posterior distribution is hard

How to get a posterior distribution?

– Sample posterior distribution

→ Marchov Chain Monte Carlo algorithms



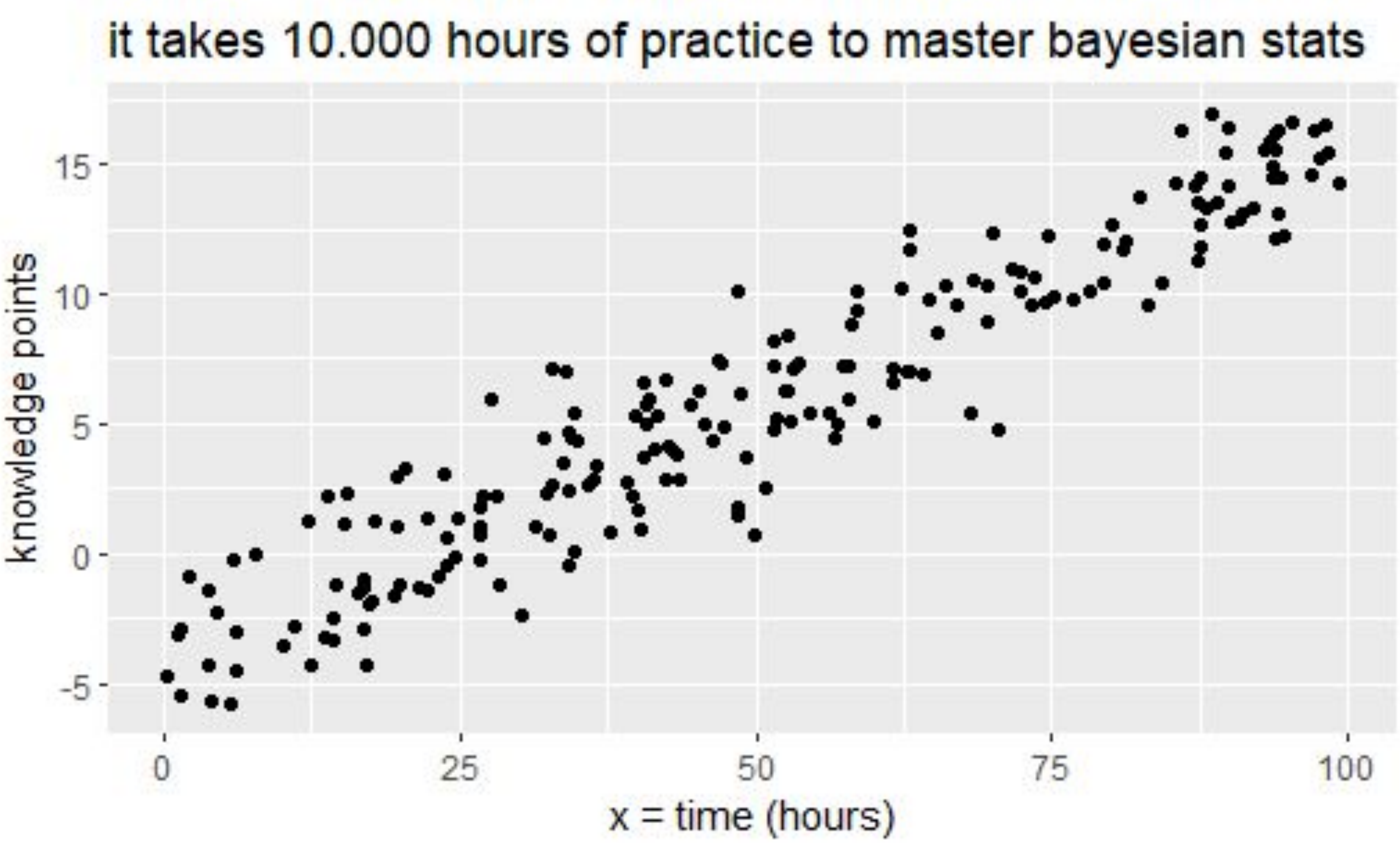
Richard McElreath

Example: learning Bayes

– It takes 10.000 hours to master a subject

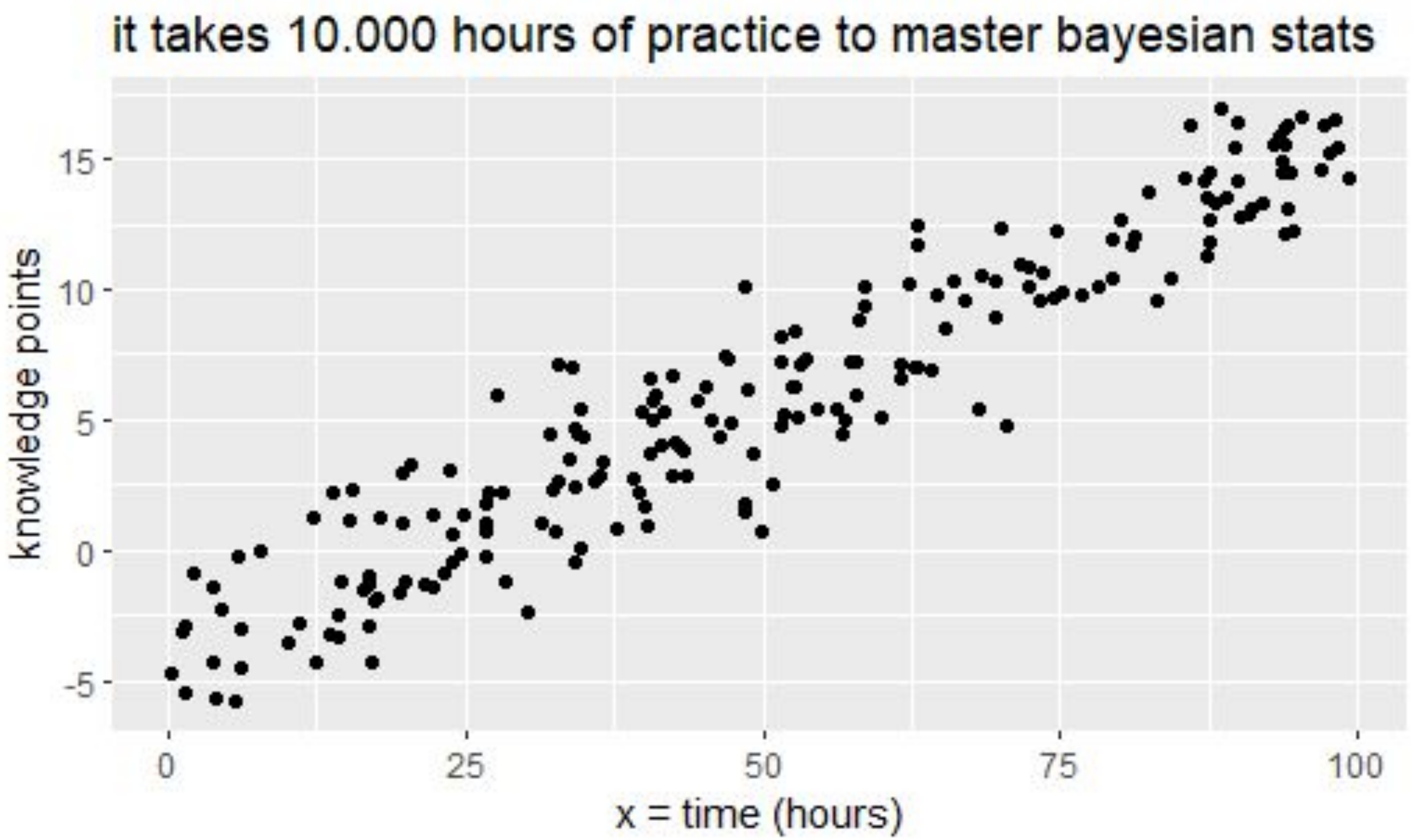
– Simulated example

learning Bayes: data model



GHENT UNIVERSITY

learning Bayes: data model



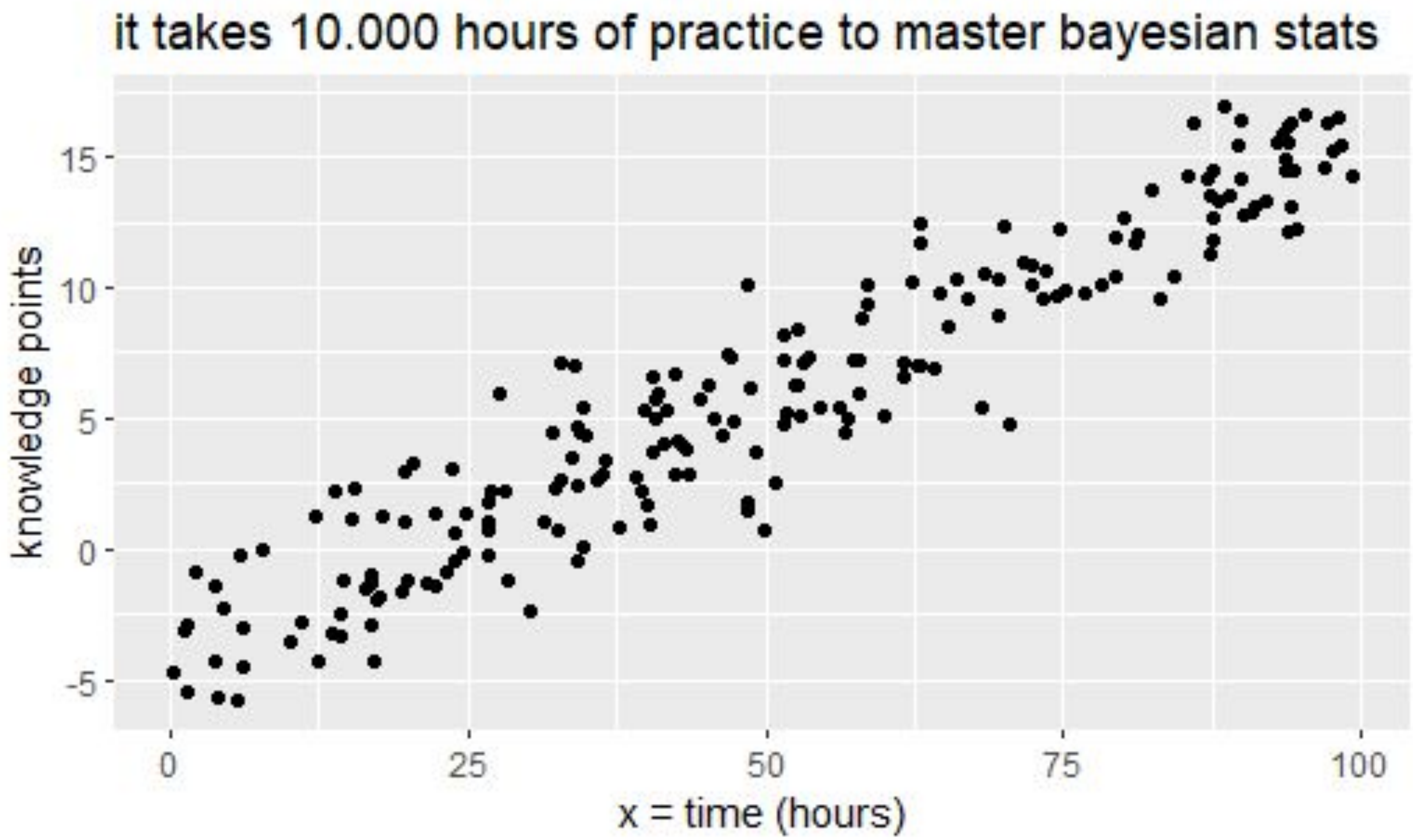
*Expected knowledge = a + b * time*

a = expected initial knowledge (-4)

b = expected rate of learning (0.2)

GHENT UNIVERSITY

learning Bayes: data model



Knowledge ~ normal(mean = expected knowledge, sd = variation (2))

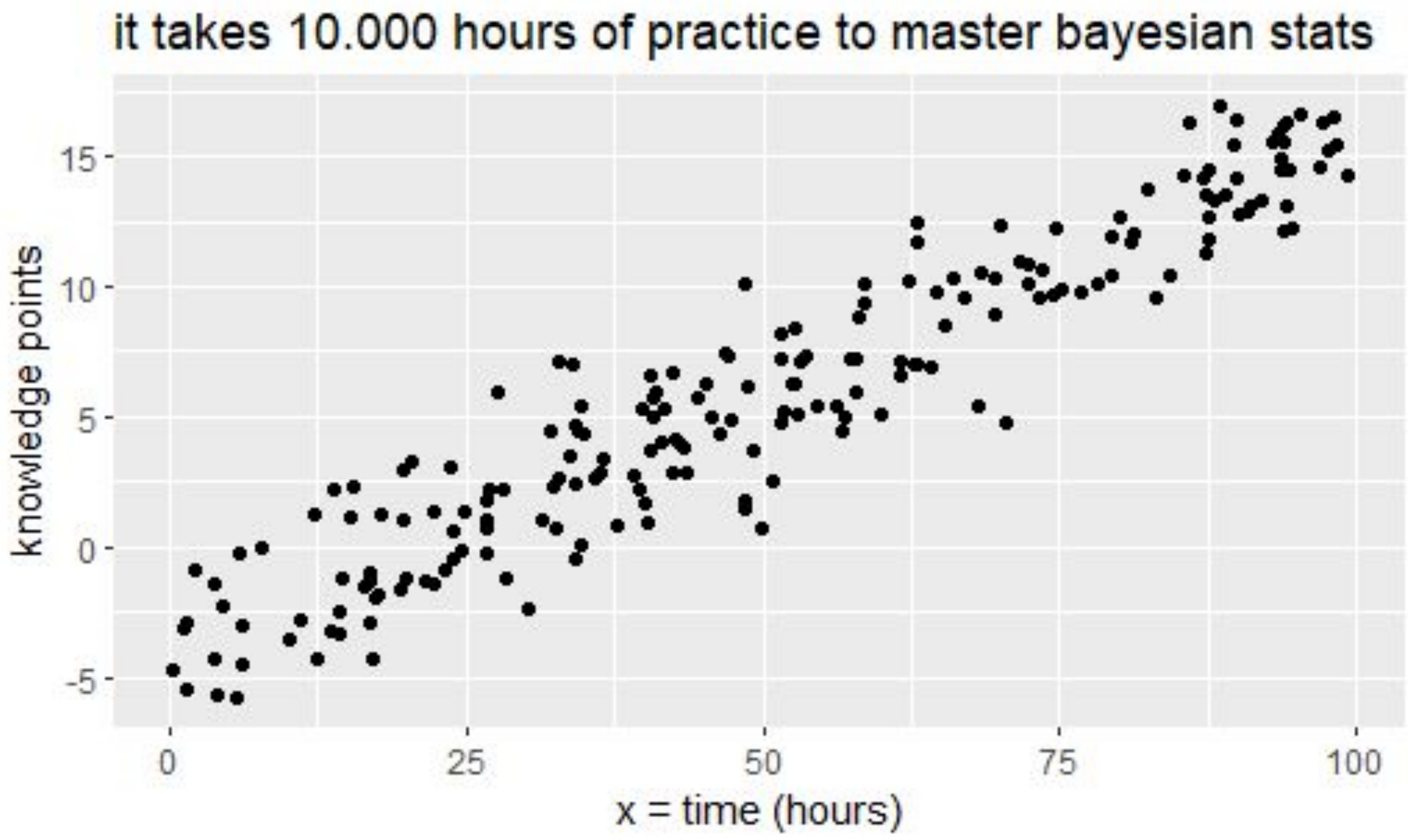
*expected knowledge = a + b * time*

a = expected initial knowledge (-4)

b = expected rate of learning (0.2)

GHENT UNIVERSITY

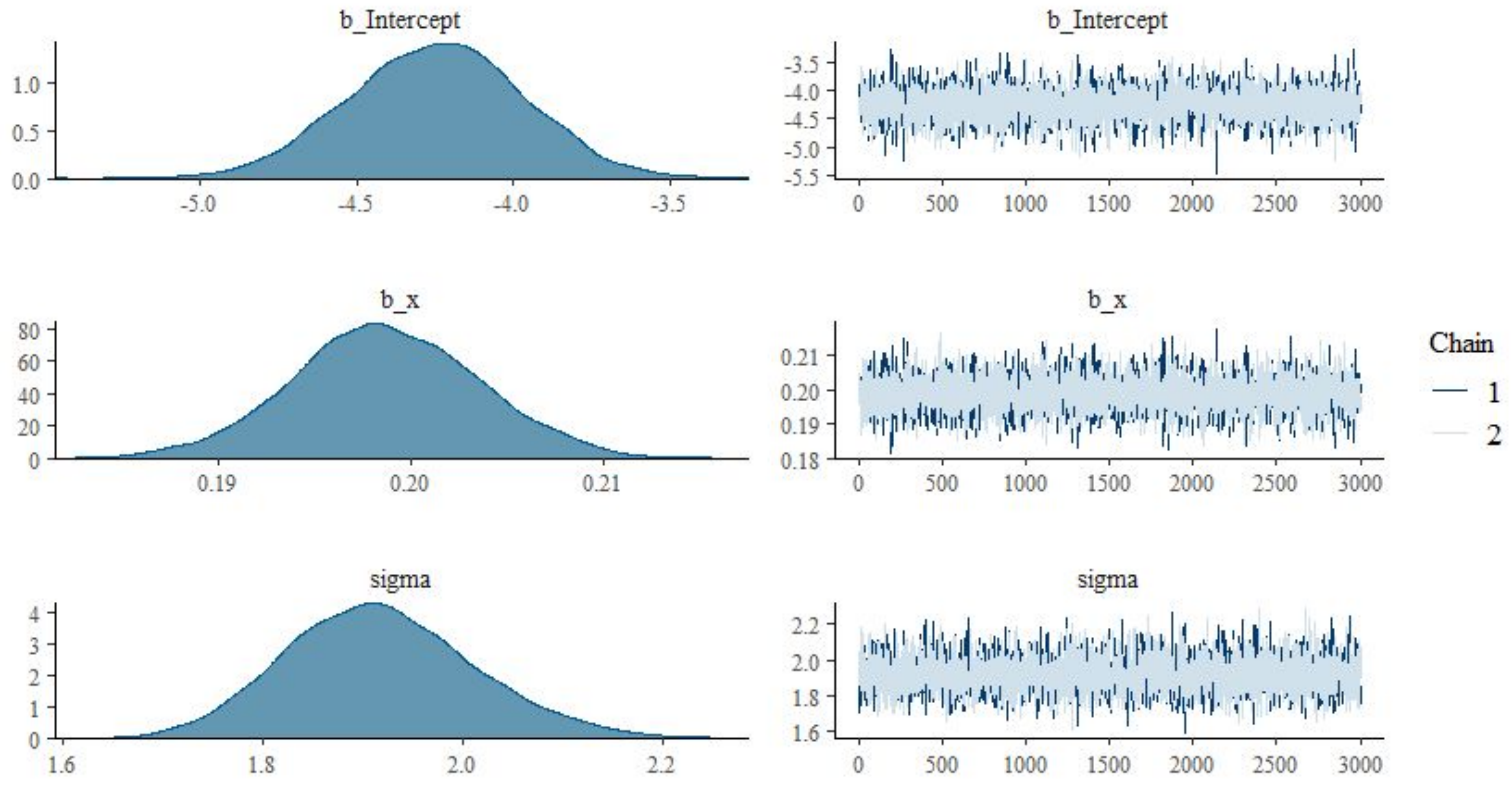
learning Bayes: formula




```
mod <- brm(formula = y ~ x, data = sim,
          family = “gaussian”
          prior = knowledge_prior,
          iter = 4000, warmup = 1000,
          chains = 2, …)
```

learning Bayes: posterior distributions

Posterior distributions | chains

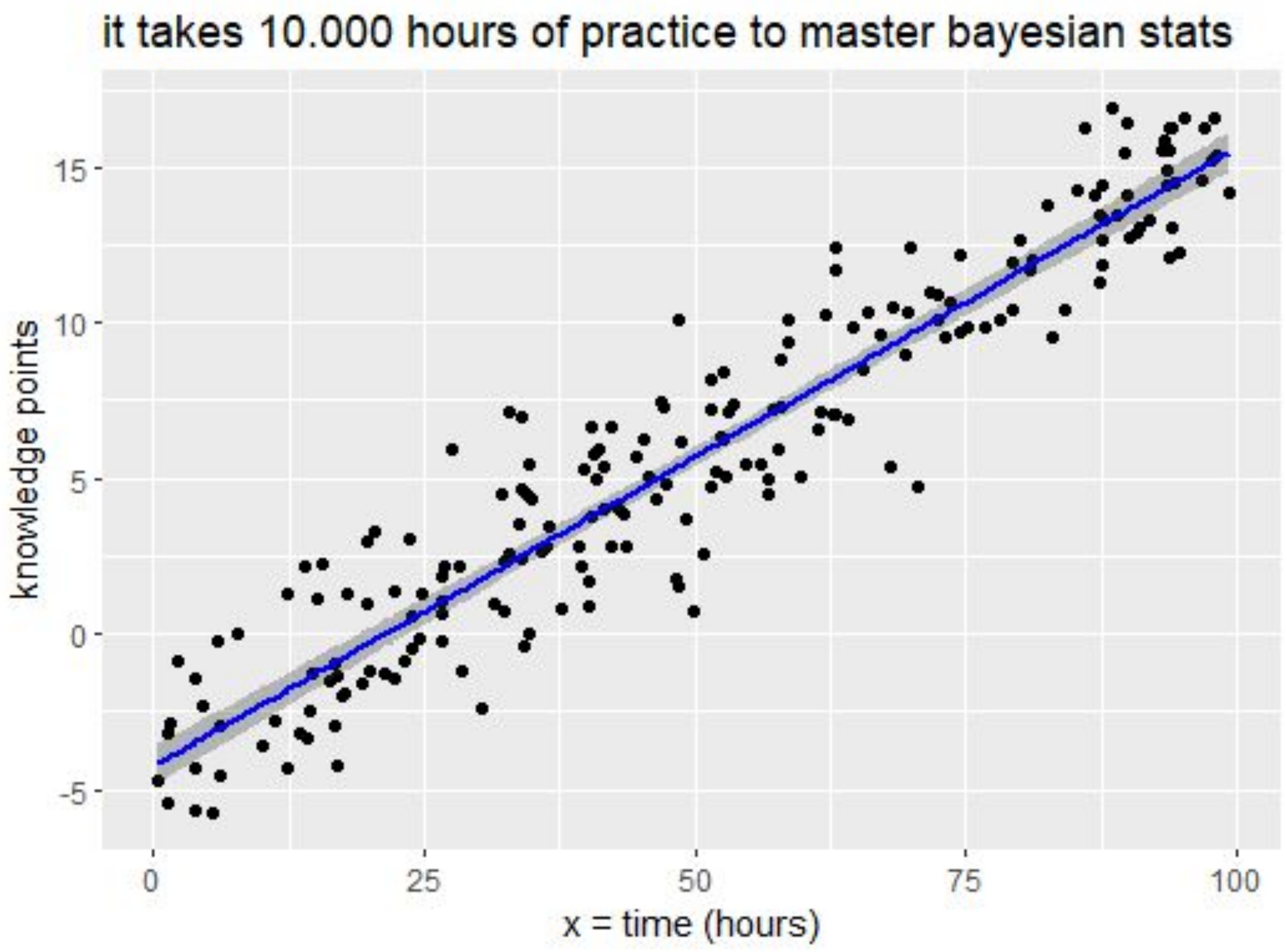


GHENT UNIVERSITY

(Posterior) distribution of trend lines

GHENT UNIVERSITY

Visualize average trend



GHENT UNIVERSITY

More than simple models

`bf(y ~ x*z + (1| group)`	Interactions and variable "random" effects
`bf(... family = binomial)`	Any error distribution you can define
`bf(... family = mixture(binomial, binomial))`	Mixture of data distributions
`bf(y ~ log(rc + or), rc ~ 1 + wodds), or ~ 1, nl = T)`	Non-linear models
`bf(y ~ 1 + line:mi(z),...)` `bf(z | mi() ~ 1)`	Missing data / latent variable

GHENT UNIVERSITY

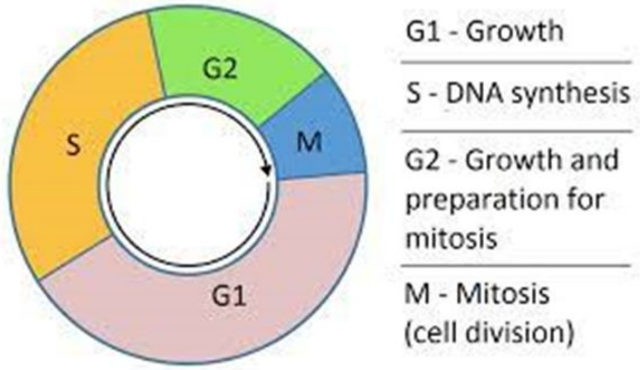
Example: difference in cell density

$$\frac{n_{4n}}{n_{2n}} = ?$$

Mass of 4n, 2n

Difference in cellular density

Cells in G2



GHENT UNIVERSITY

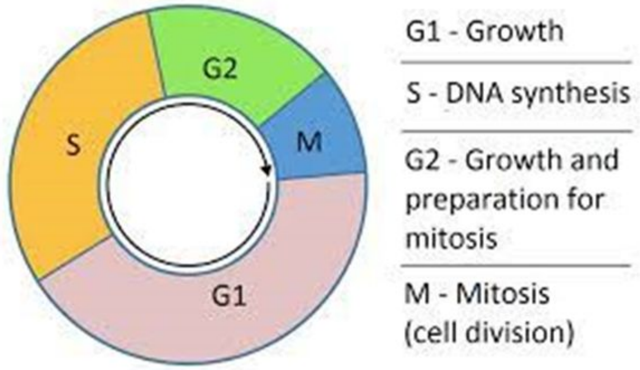
Example: difference in cell density

Mass of 4n, 2n

$$\frac{n_{4n}}{n_{2n}} = ?$$

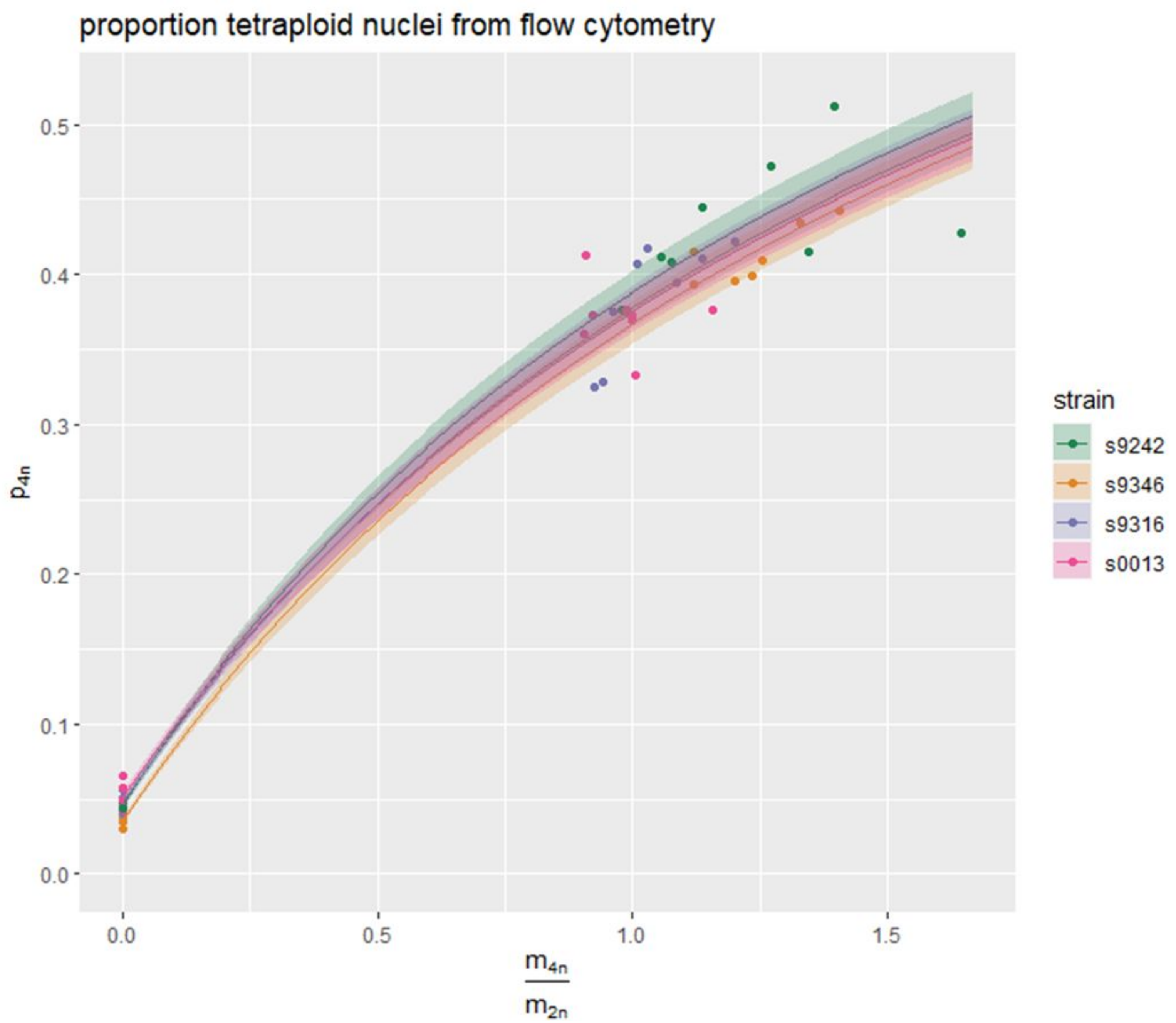
Difference in cellular density

Cells in G2



$$logit(p_{4n}) = \ln\left(\frac{m_{4n}}{m_{2n}} * \frac{cell\ density_{4n}}{cell\ density_{2n}} + \frac{p_{G2}}{1 - p_{G2}}\right)$$

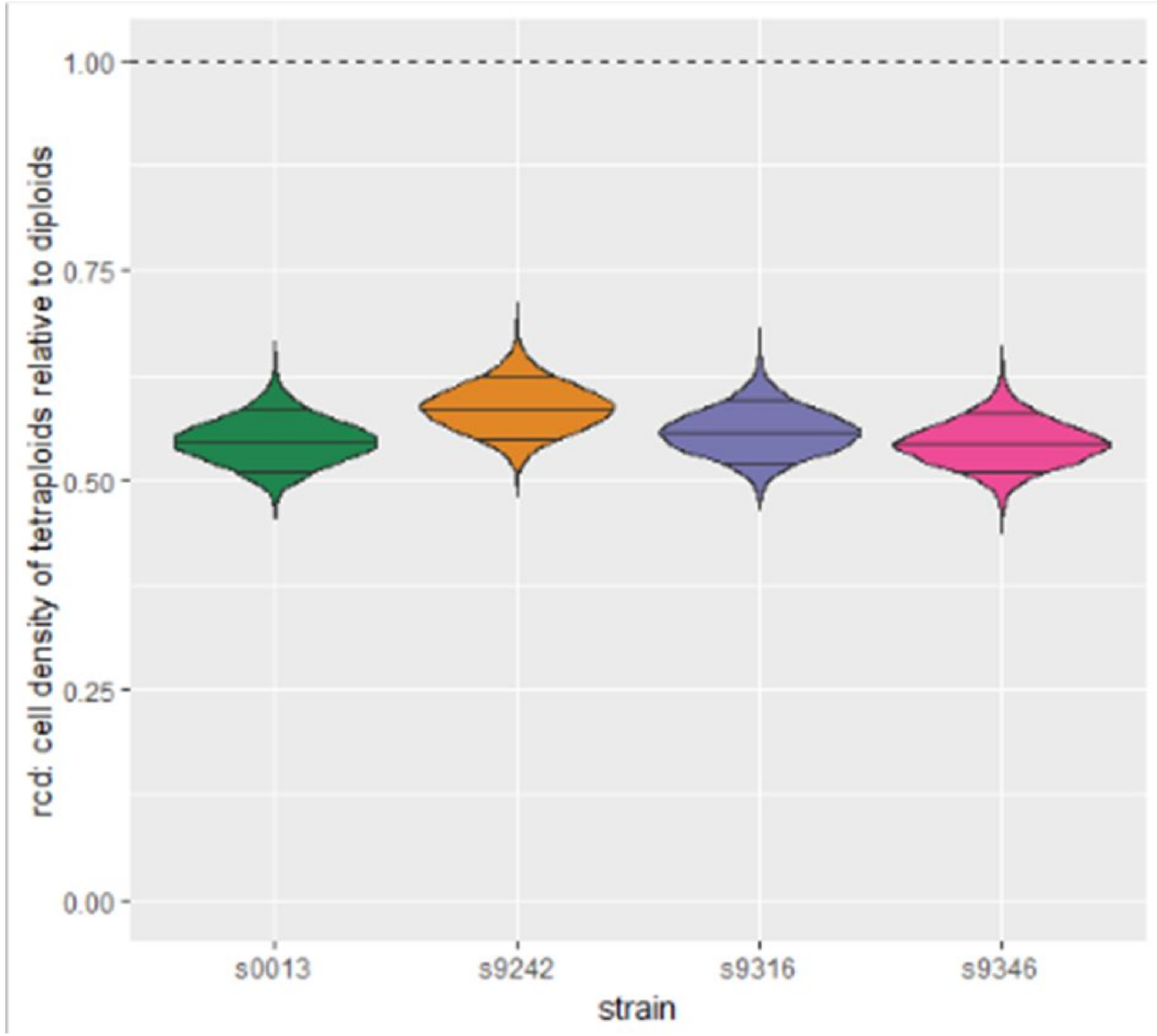
Example: difference in cell density



$$logit(p_{4n}) = \ln\left(\frac{m_{4n}}{m_{2n}} * \frac{cell\ density_{4n}}{cell\ density_{2n}} + \frac{p_{G2}}{1 - p_{G2}}\right)$$

GHENT UNIVERSITY

Difference in cell density estimation



$$logit(p_{4n}) = \ln\left(\frac{m_{4n}}{m_{2n}} * \frac{cell\ density_{4n}}{cell\ density_{2n}} + \frac{p_{G2}}{1 - p_{G2}}\right)$$

INLA

Femke

What is INLA?

- Integrated Nested Laplace Approximation
- Approximate Bayesian inference; alternative to MCMC, but ‘faster’
- ‘Restricted’ to GMRF (Gaussian Markov random field) models → posteriors of fixed effects are Gaussian → includes everything you are used to from frequentist ecology & more
- R-INLA project: https://www.r-inla.org/home

Applications of INLA

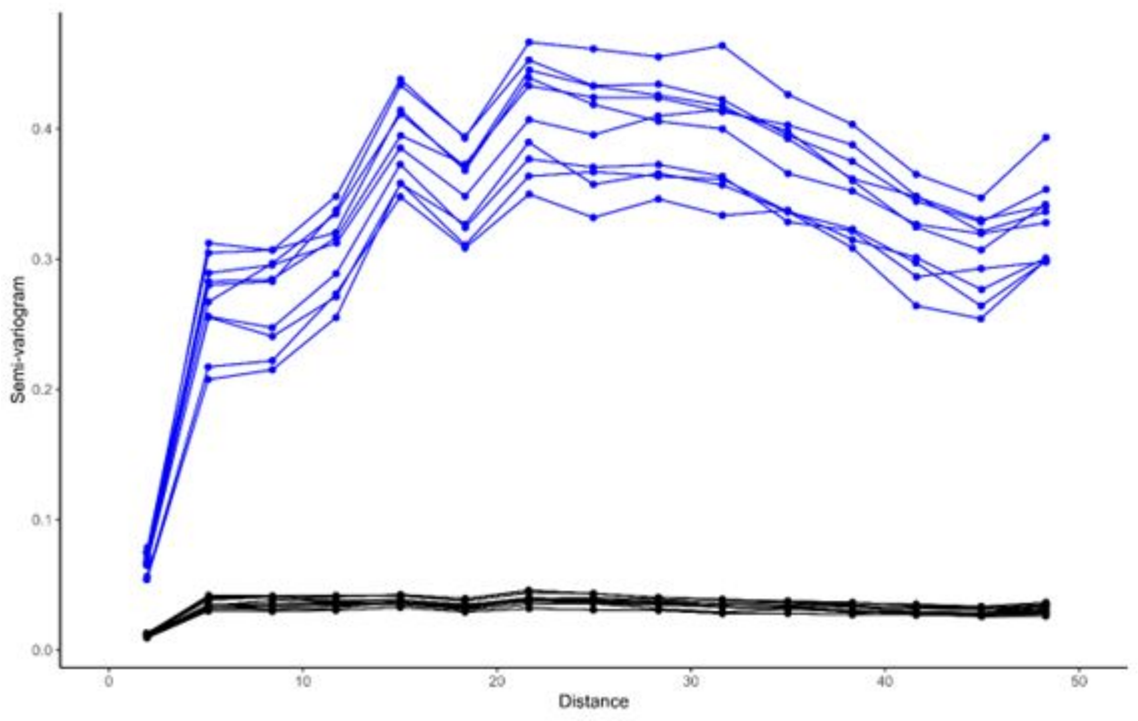
- Models correcting for spatial and/or temporal correlation
- Point processes (modelling preferential sampling)
- Modelling time series
- …
- All your (zero-inflated) GL(M)Ms, GAM(M)s with normal, (negative) binomial, Poisson, gamma, distributions (and much more! E.g. censored/truncated Poisson)

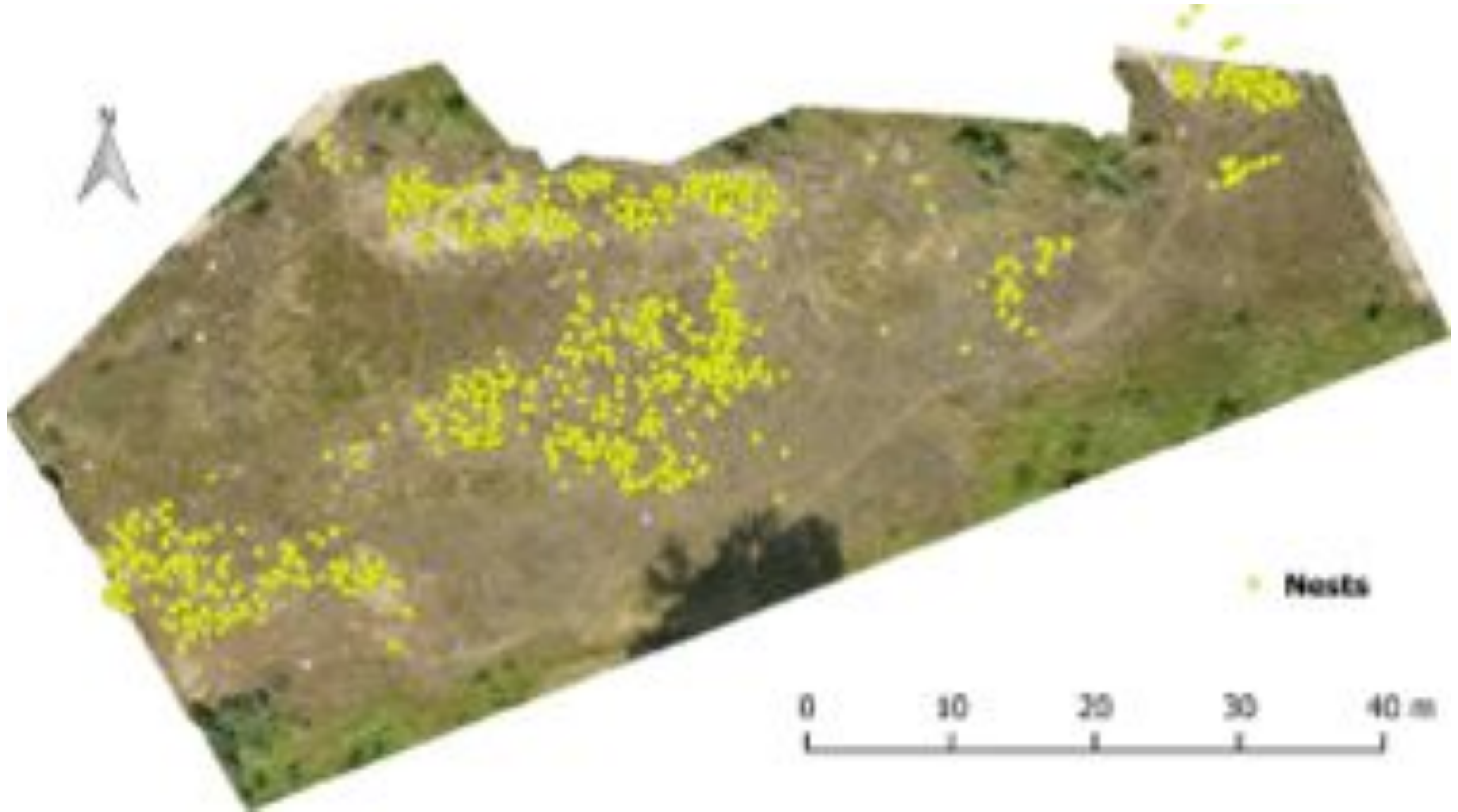
GHENT UNIVERSITY

Quick example

Batsleer, Maes & Bonte (2022) AmNat

- Microhabitat suitability model for Bembix rostrata
- Spatial autocorrelation!
- Presence ~ NDVI + Insolation + SPDE

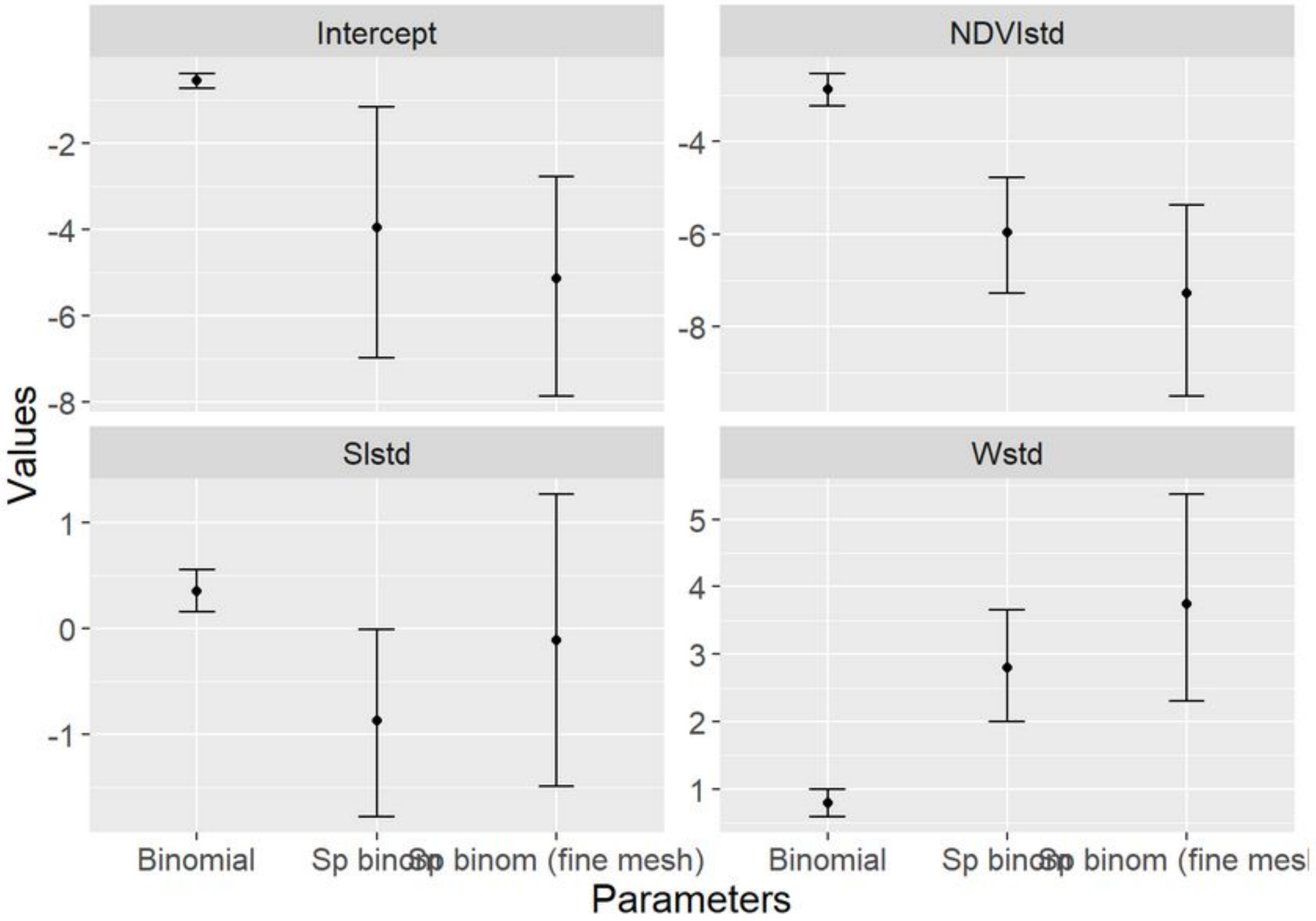


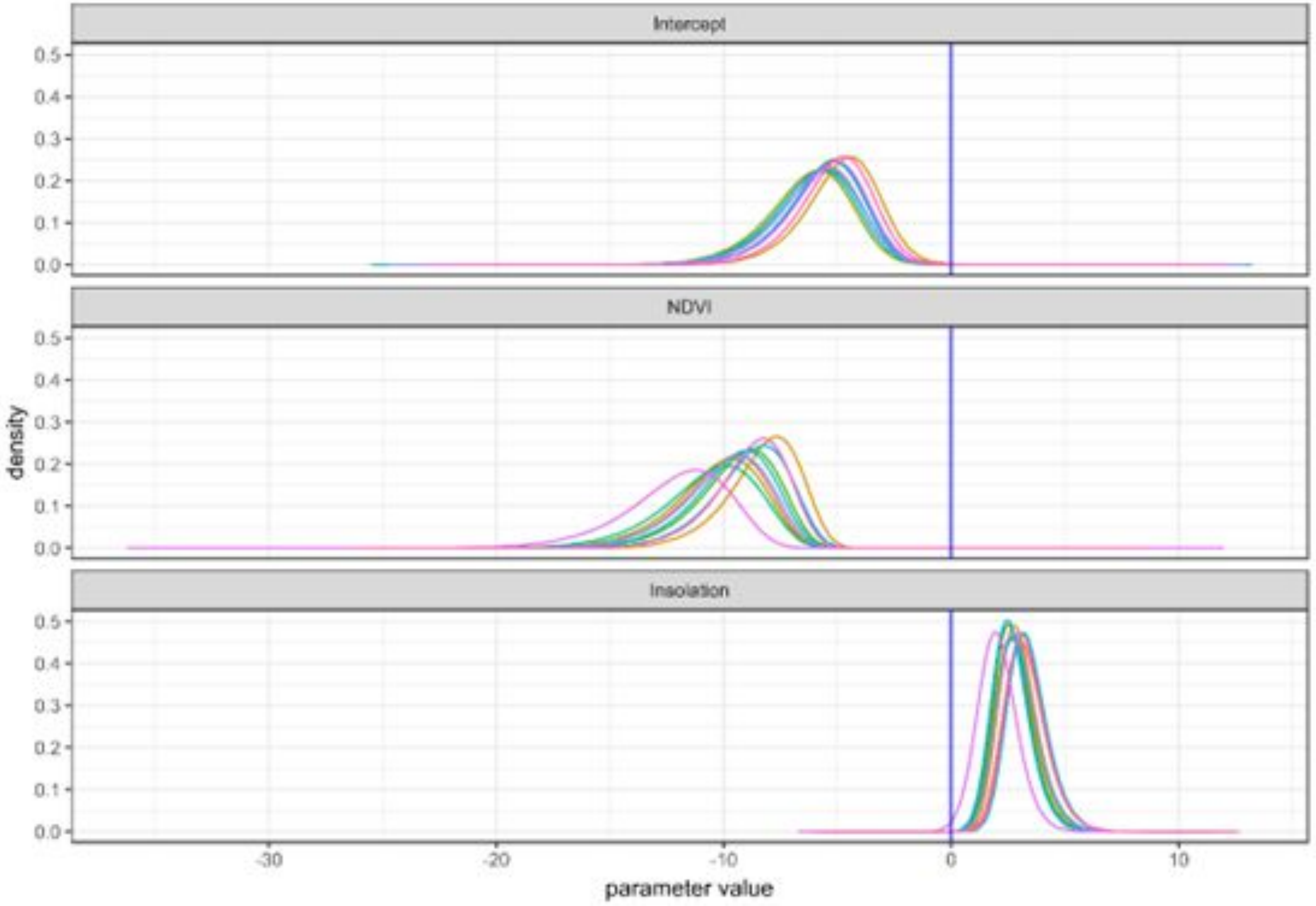


GHENT UNIVERSITY

Quick example

Posterior distributions





Better estimates, but more uncertainty!

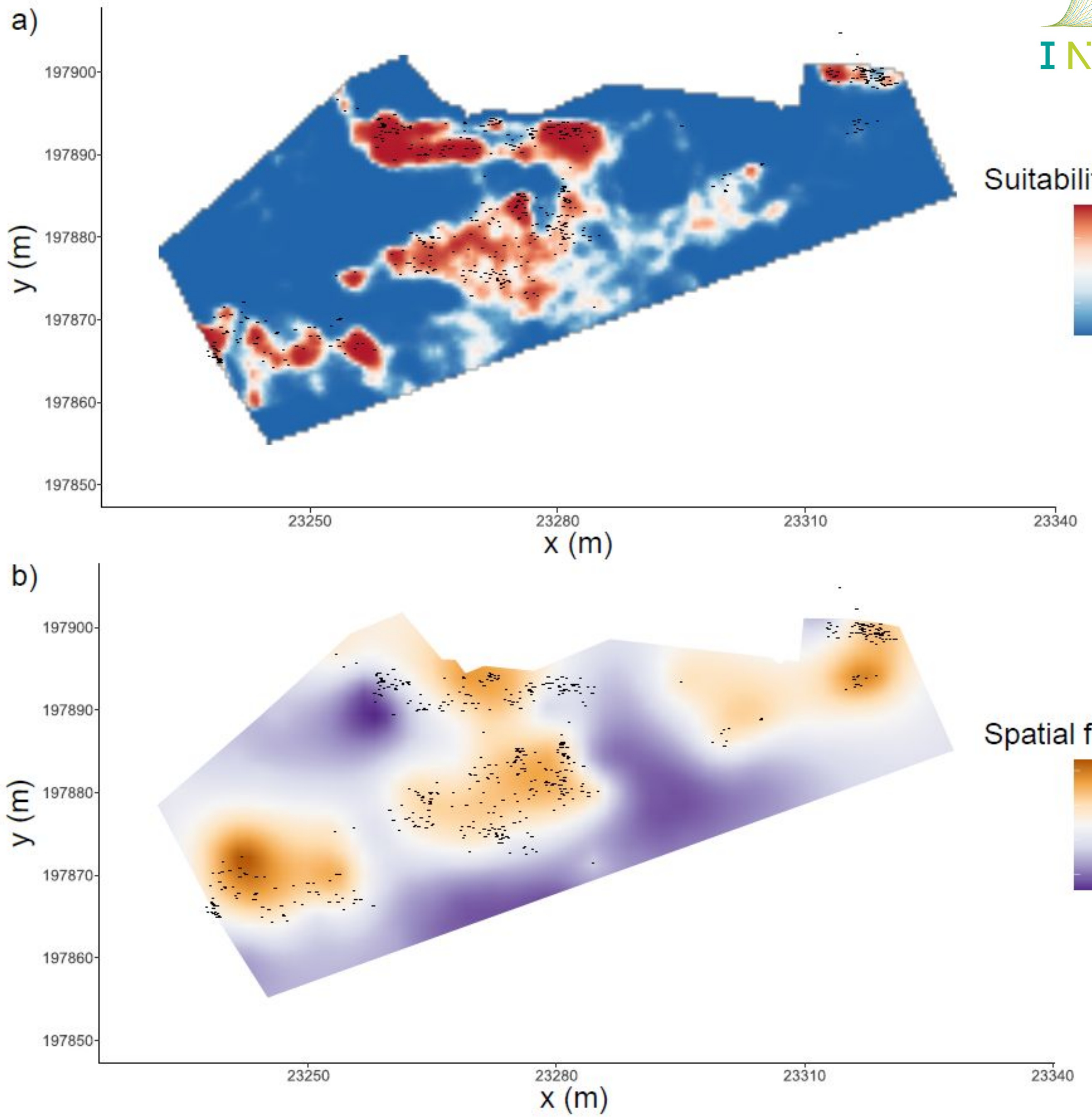
GHENT UNIVERSITY

Quick example

Models were compared using the Watanabe-Akaike information criterion (WAIC; Watanabe 2010; Gelman et al. 2014)

The plotting of the spatial field, the spatial residuals that INLA corrects for, shows whether the degree of clustering was higher (hot spots) or lower (cold spots) than expected on the basis of the covariates (NDVI and insolation) in the microhabitat model

For every run, zero was excluded from the 95% credibility intervals of the effect sizes, indicating that the signs of the effect sizes were clearly determined (this is a Bayesian approach to evaluating statistical significance at a specified level).



GHENT UNIVERSITY

Other applications from the lab

Biomorphogenic Feedbacks and the Spatial Organization of a Dominant Grass Steer Dune Development

Dries Bonte[1*] Femke Batsleer[1] Sam Provoost[2] Valérie Reijers[3] Martijn L. Vandegehuchte[1,4] Ruben Van De Walle[1] Sebastian Dan[5] Hans Matheve[1] Pieter Rauwoens[6] Glenn Strypsteen[6] Tomohiro Suzuki[5] Toon Verwaest[5] Jasmijn Hillaert[1,2]

Linking dune morphology to spatial configuration of marram grass

Comparing mechanistic model (IBM) with GIS derived field observations

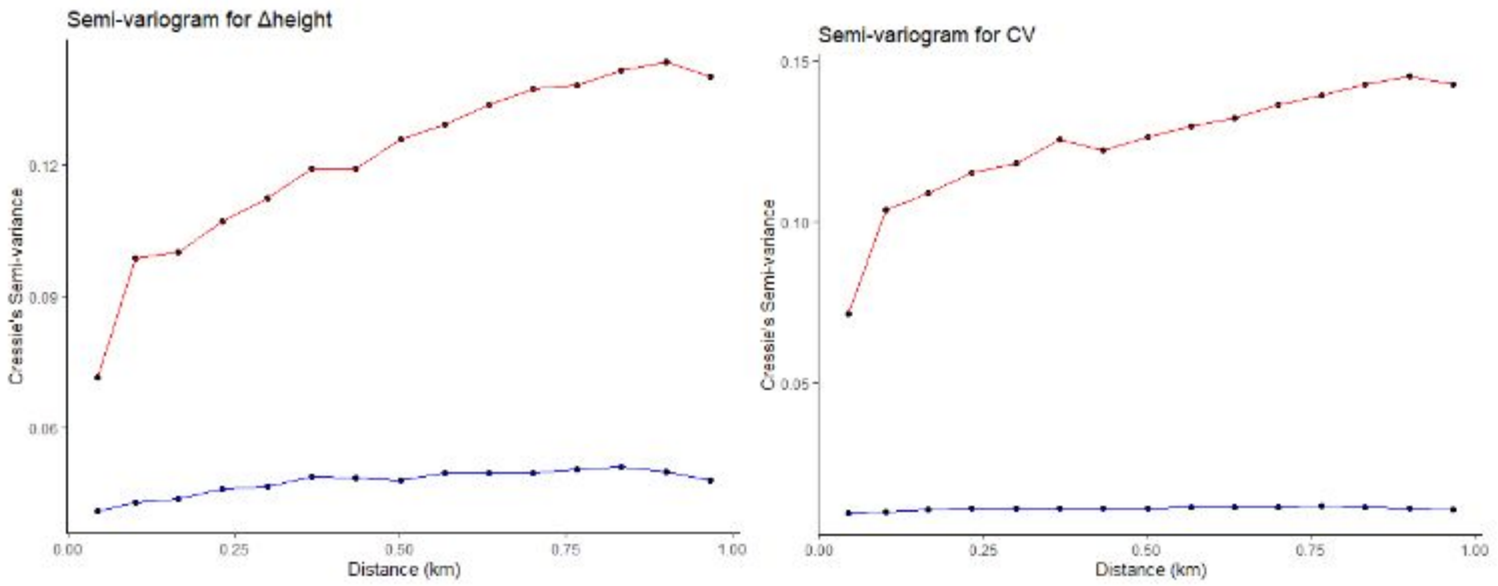
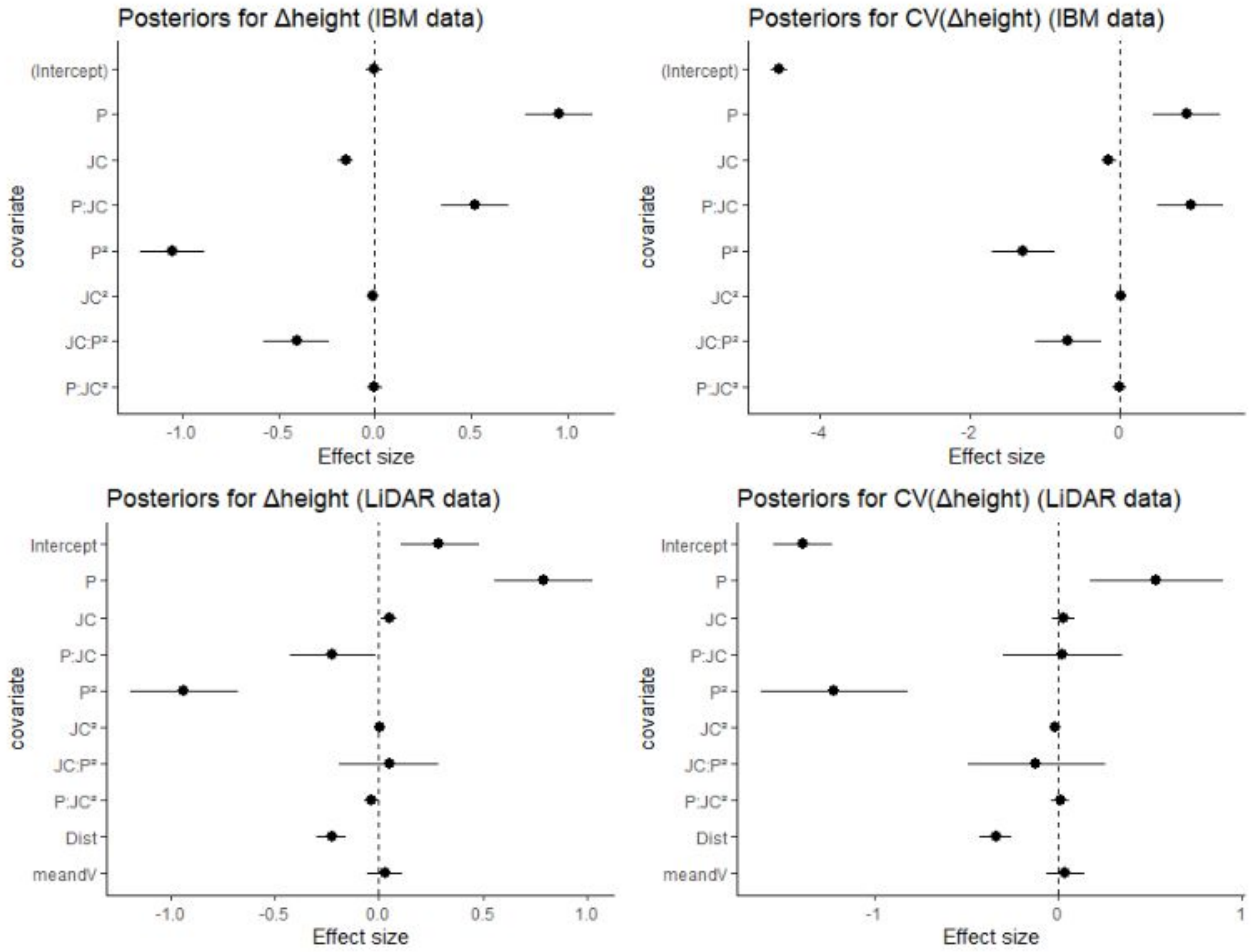


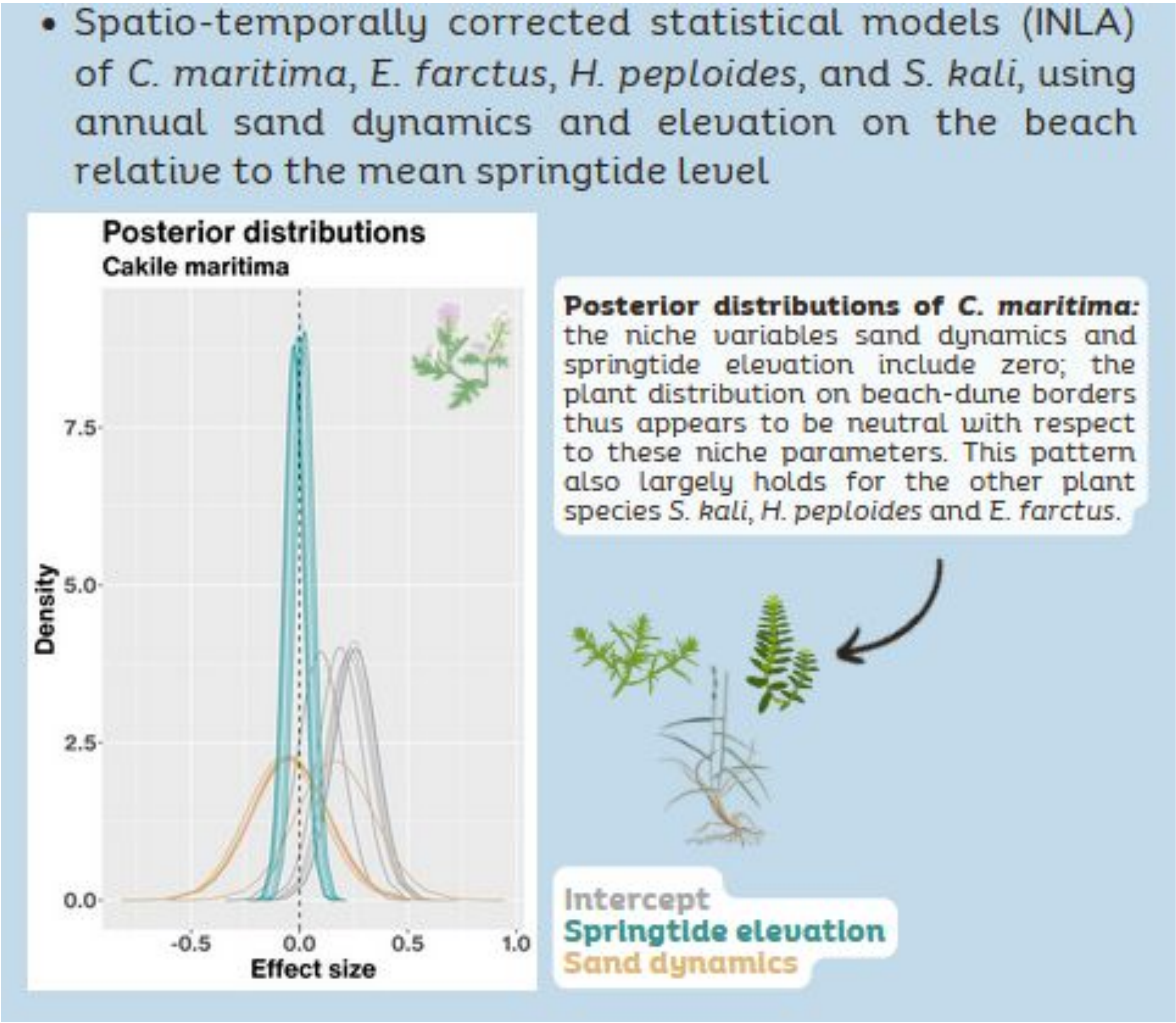
Fig S4.3: Semi-variogram of residuals for non-spatial model (red) and spatial model (blue) for *Δheight* (left) and *CV* (right).



GHENT UNIVERSITY

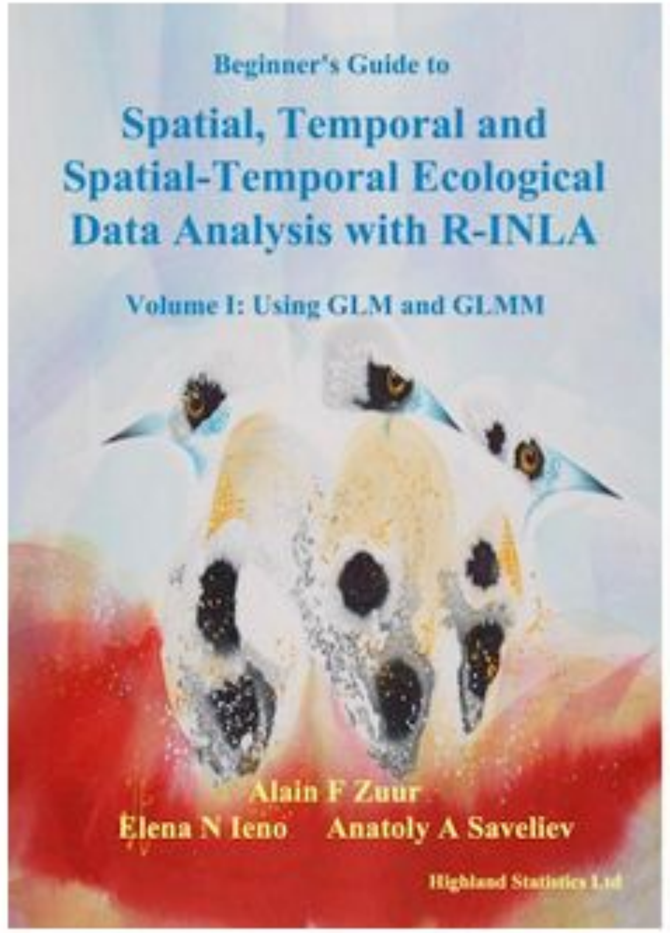
Other applications from the lab

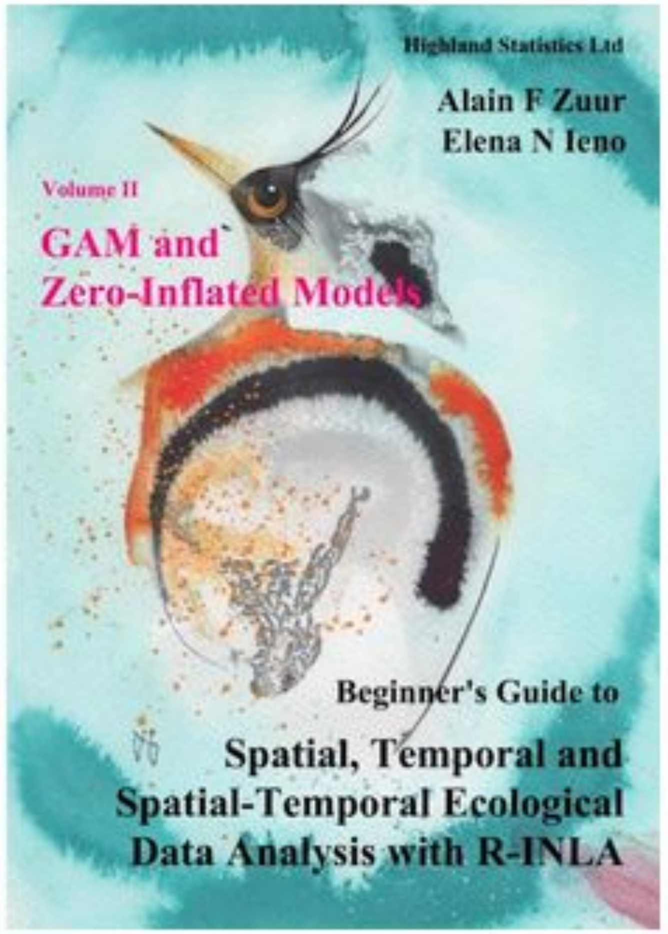
Charlotte Taelman



Want to learn how to implement INLA?

- Books we have at the lab
- Ask us for help!
- Coding club devoted to INLA examples?





GHENT UNIVERSITY

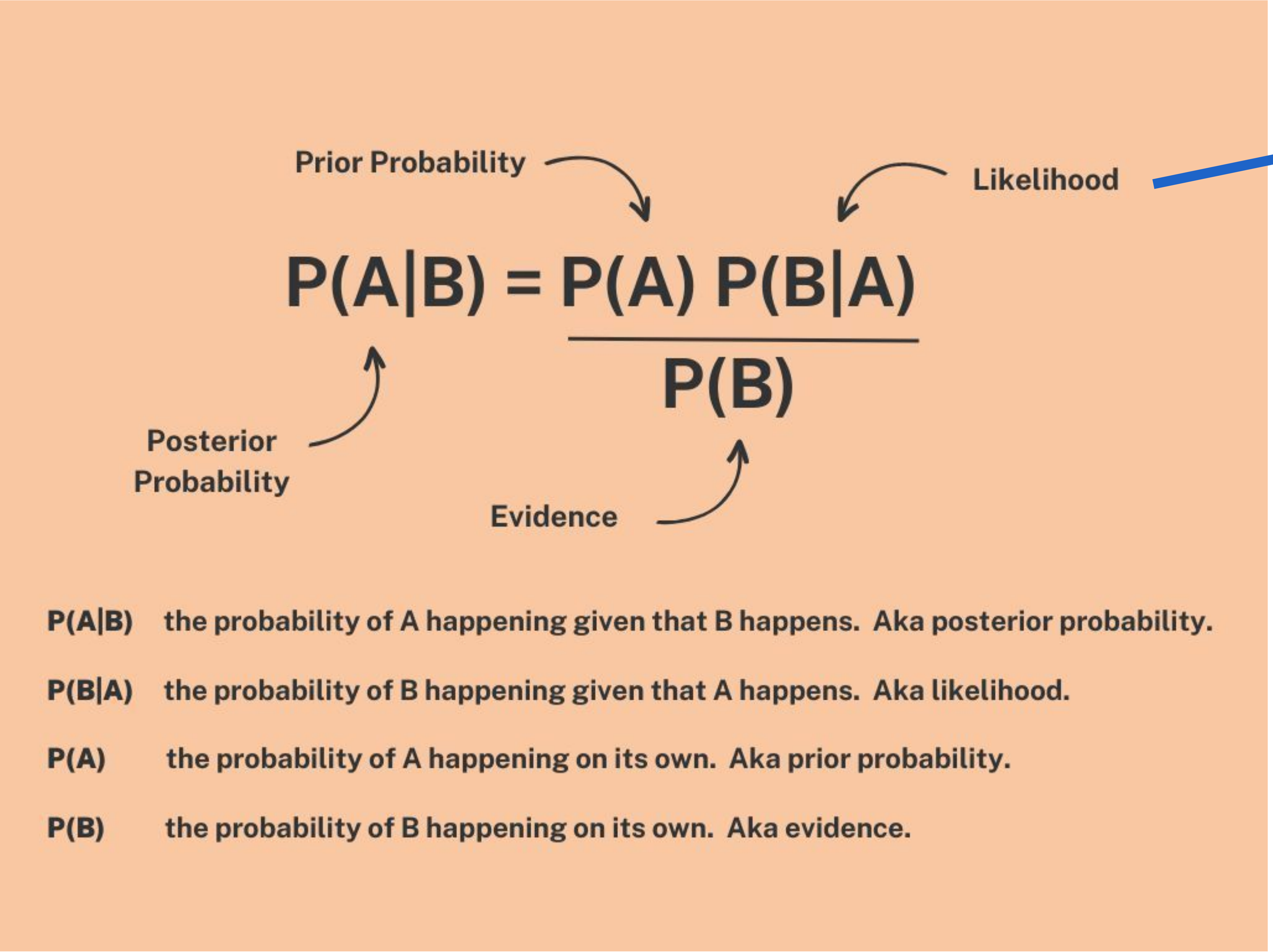
ABC

Femke

What is ABC?

- Approximate Bayesian Computation
- Bypass the evaluation of the likelihood function
- Compare different (complex) models and estimate parameter values
- Application: evaluation of complex models in population genetics, epidemiology, ecology
- E.g.: IBM/ABM

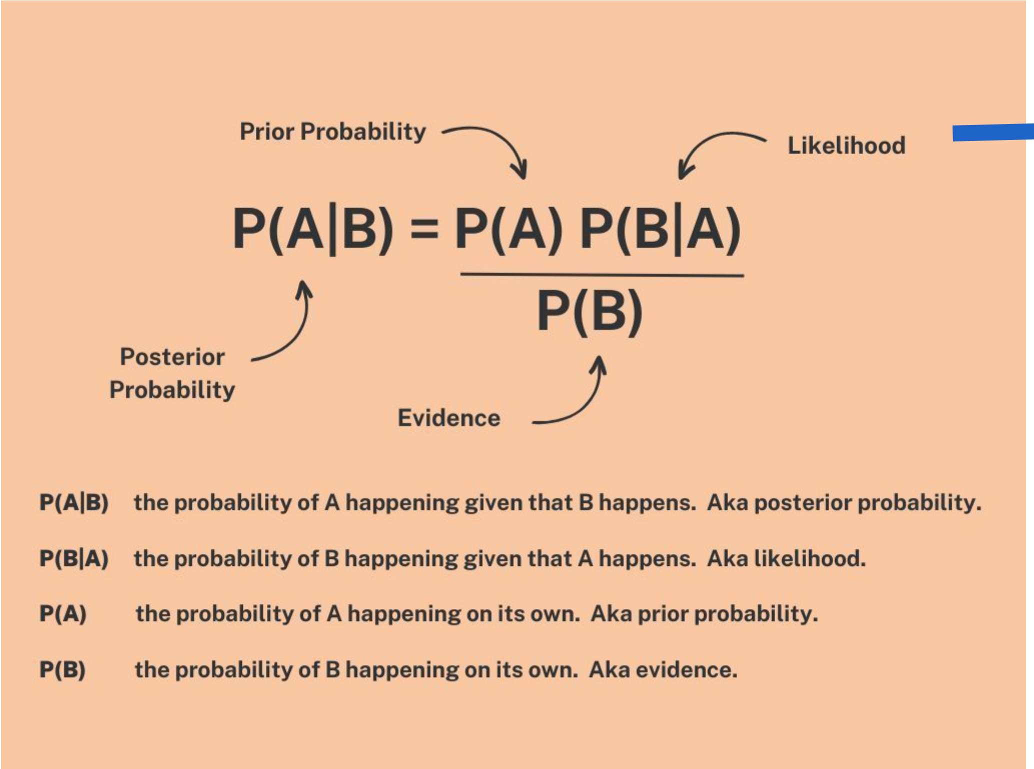
Bypassing evaluation of the likelihood?!



Derived from statistical model:
- Compatibility of the evidence (B) with the given hypothesis (A)
- Function of the evidence (B)

Probability of the data given the parameters

Bypassing evaluation of the likelihood?!



Probability of the data given the parameters

Probability function of your model formulation

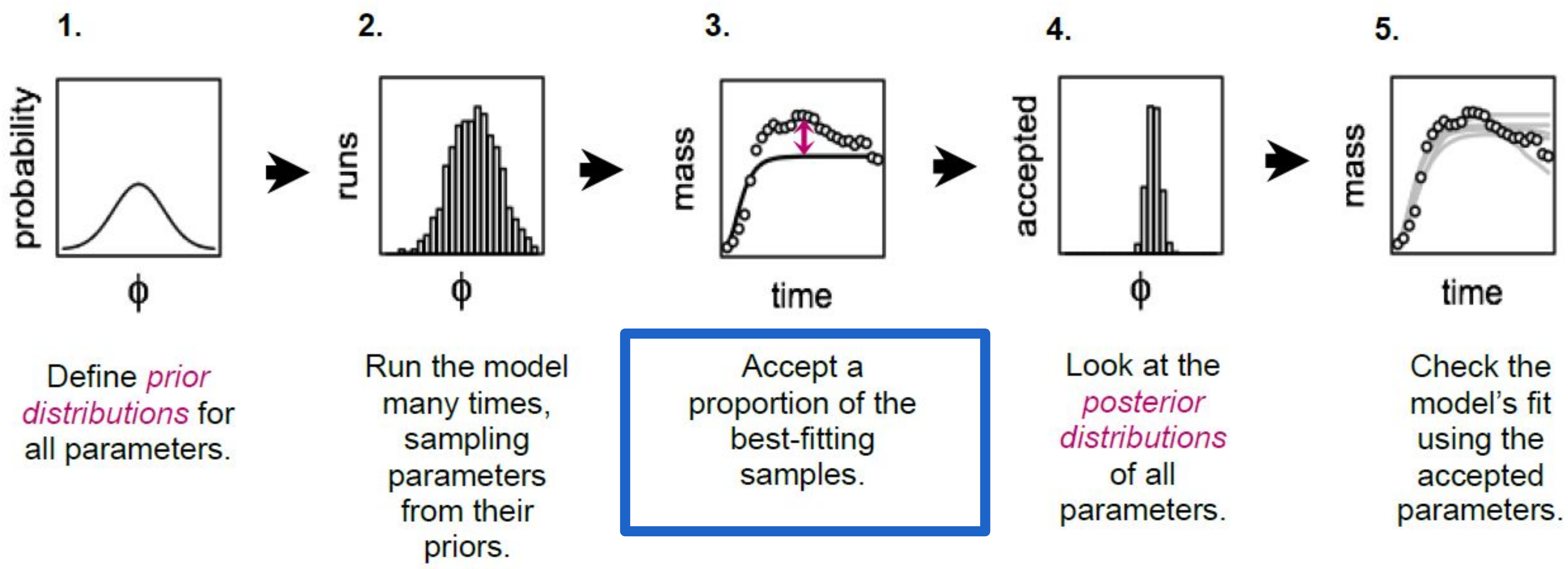
Deriving likelihood functions analytically is not possible for complex models like IBMs

→ ABC approximates likelihood using simulations

Here: approximation of likelihood
INLA: approximation of posteriors
(MCMC samples posterior)

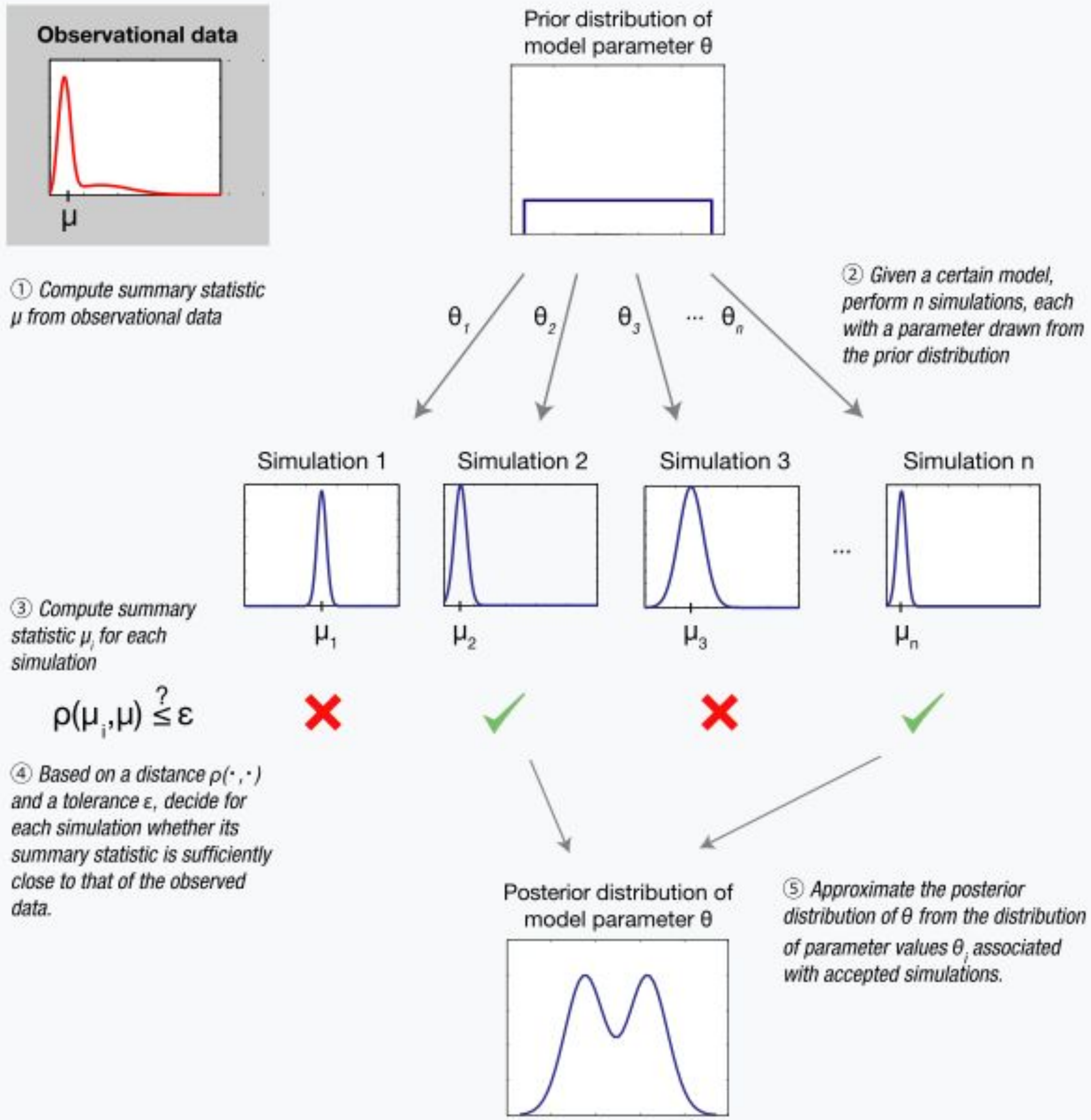
Parameter estimations with ABC

- Where ‘ϕ’ is a model parameter
- Where ‘mass’ [of an individual] is an example of a modelled property



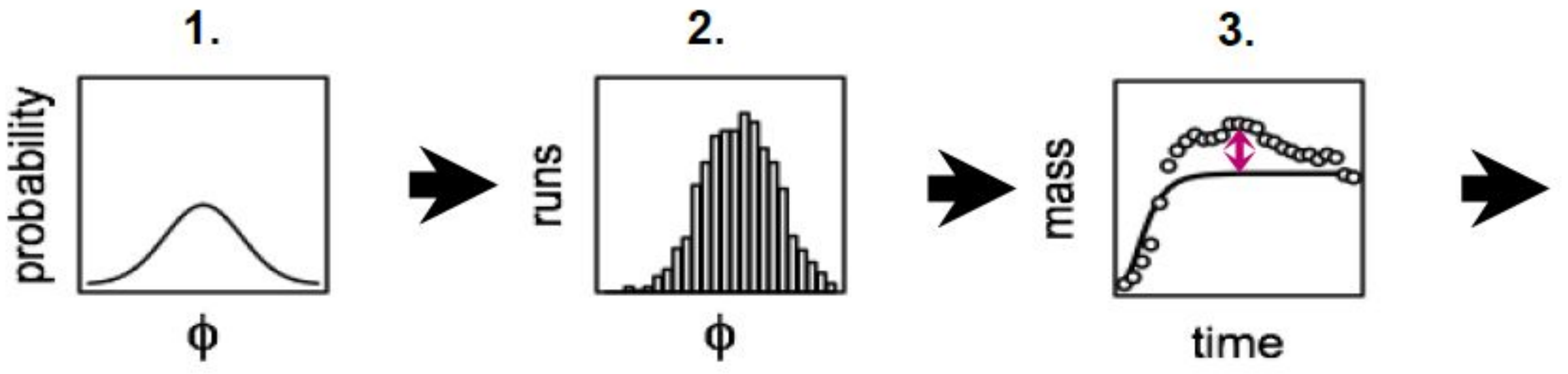
Based on summary statistics

Source: ABC4IBM workshop BES 2016; Elske van der Vaart & Richard Sibly



GHENT UNIVERSITY

Model selection with ABC



	Model A	Model B
Model A	-	19
Model B	0.05	-

1. Define *prior distributions* for all parameters in all models.
2. Run all models many times, sampling parameters from their priors.
3. Accept a proportion of the best-fitting samples.
4. Calculate how often each model is accepted relative to each other model - the *Bayes factors*.

Based on summary statistics

GHENT UNIVERSITY

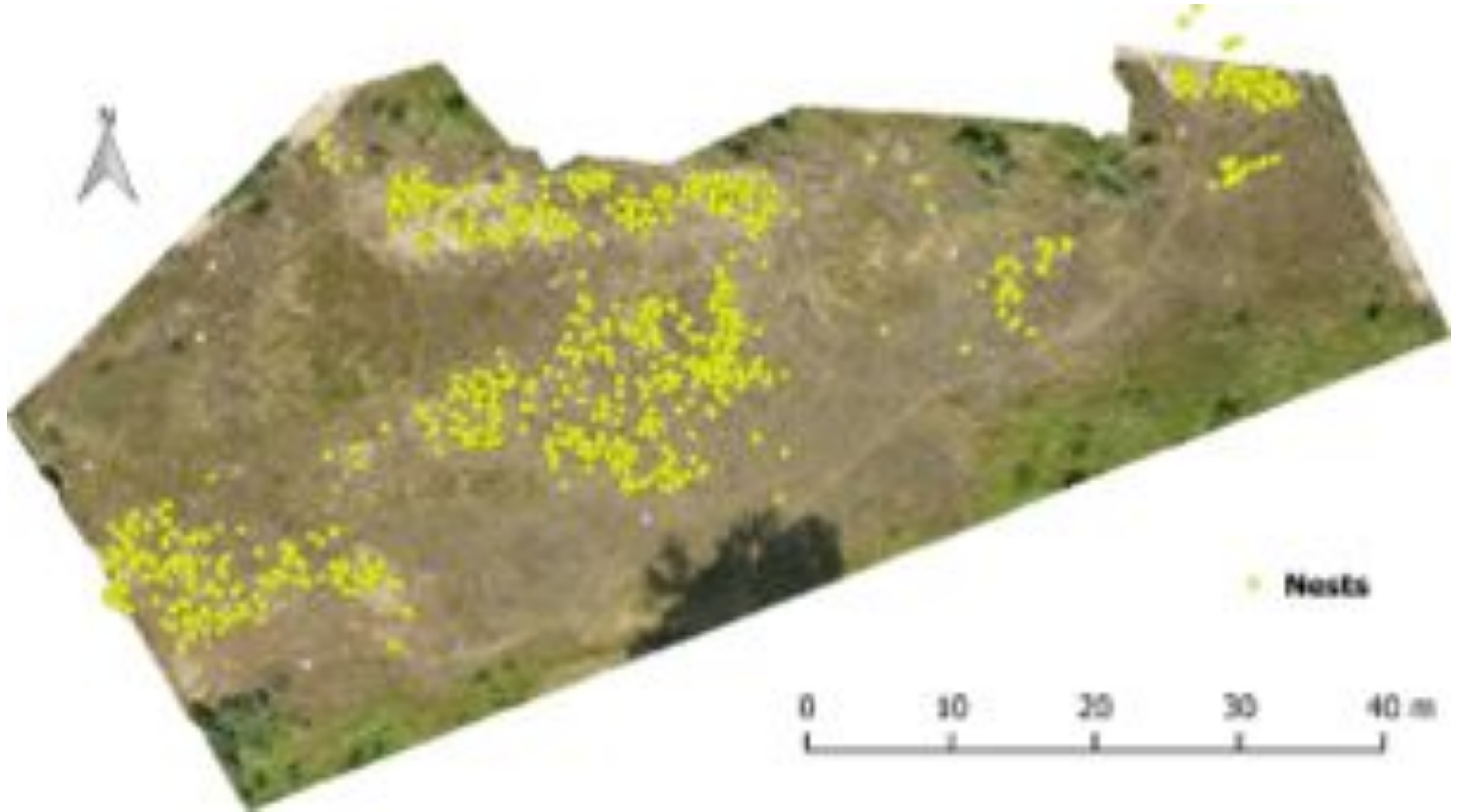
Source: ABC4IBM workshop BES 2016; Elske van der Vaart & Richard Sibly

Quick example

Batsleer, Maes & Bonte (2022) AmNat

- IBM of nest spatial pattern formation of *B. rostrata*
- Evaluate with ABC4IBM:

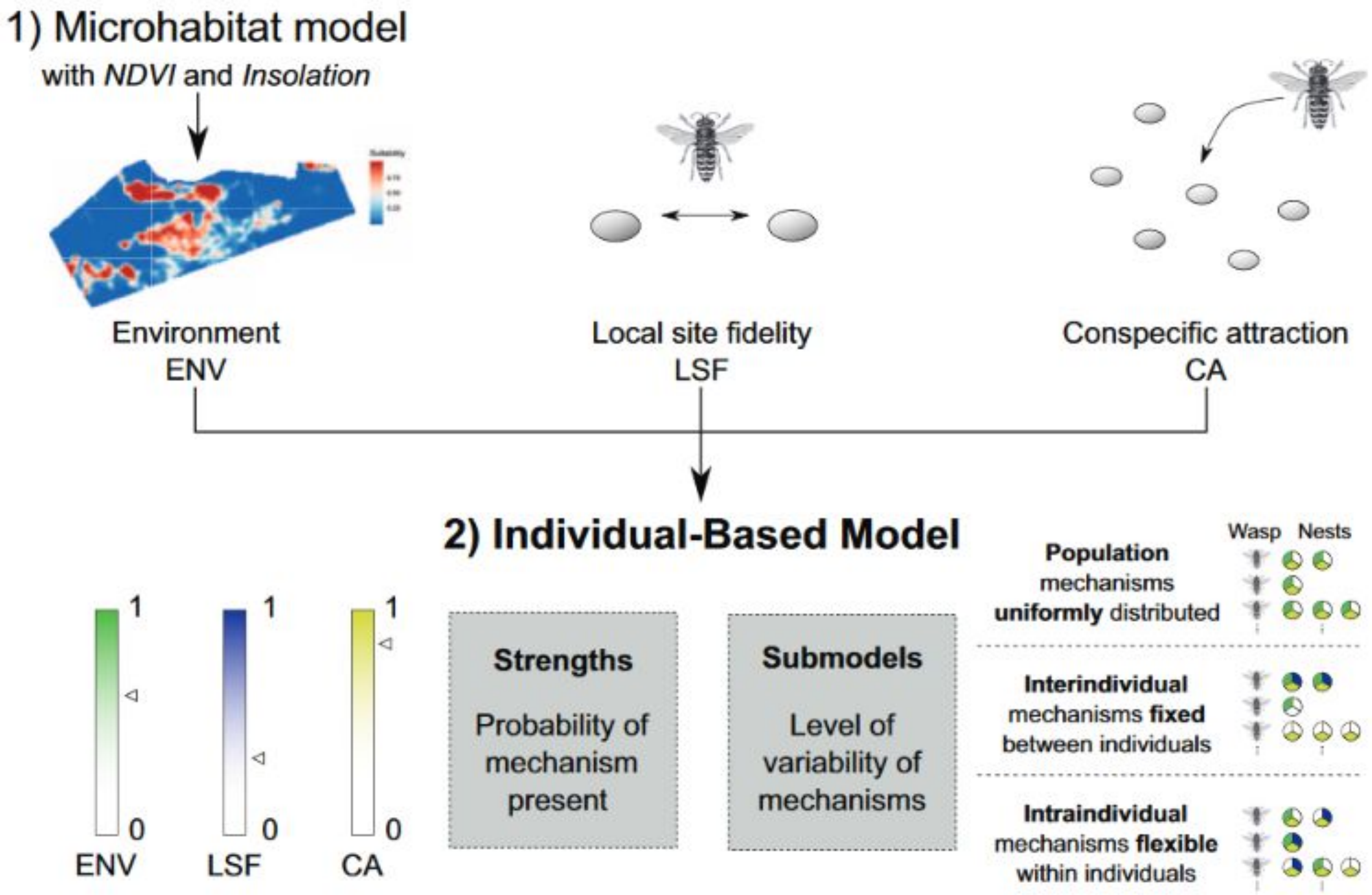
  spatial pattern modelled ⟵⟶ spatial pattern in the field
- Observation in the field: spatial patterns of nests



GHENT UNIVERSITY

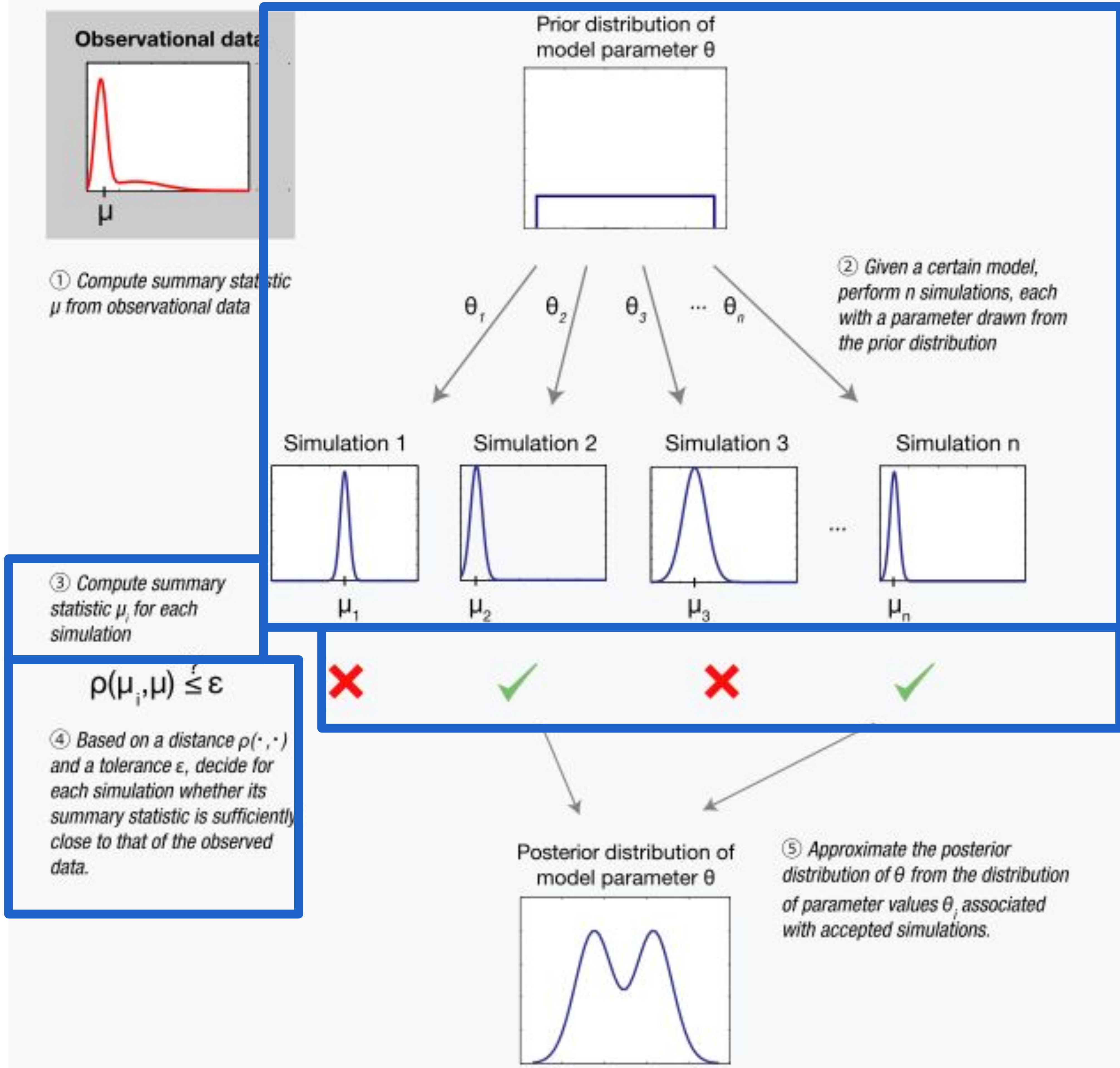
Quick example

- Model: parameters (3 strength of mechanisms + model specifics) + 3 hypotheses how these mechanism vary across the population
- Summary statistics: spatial clustering (standardized Ripley's K) + network metrics



Quick example

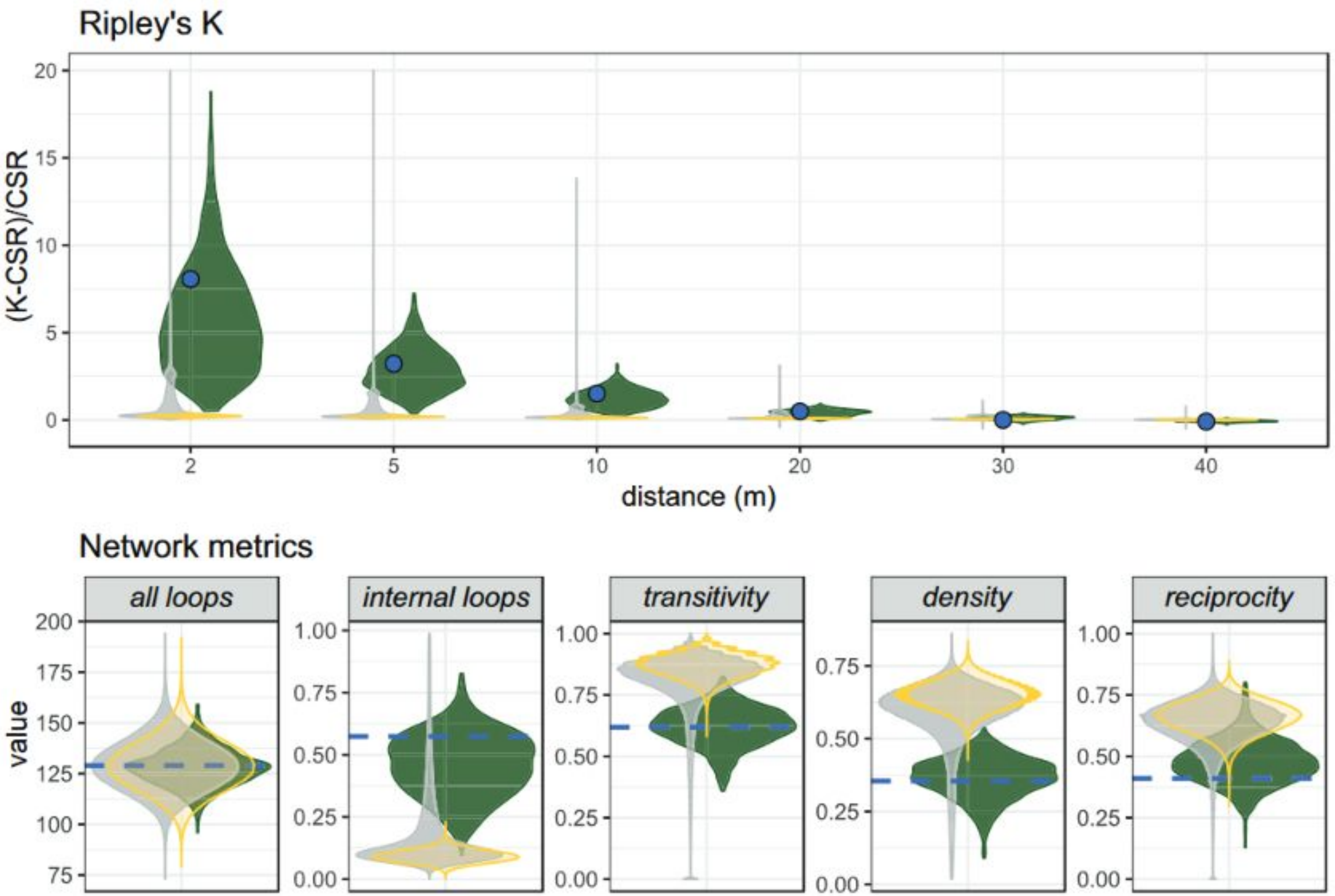
- Drawing from prior distribution of model parameters → run 1.000.000 simulations
- Calculate summary statistics
- Calculate distance of summary stats of each run tot the observational data
- Accepted 1000 (0.01%) best models



GHENT UNIVERSITY

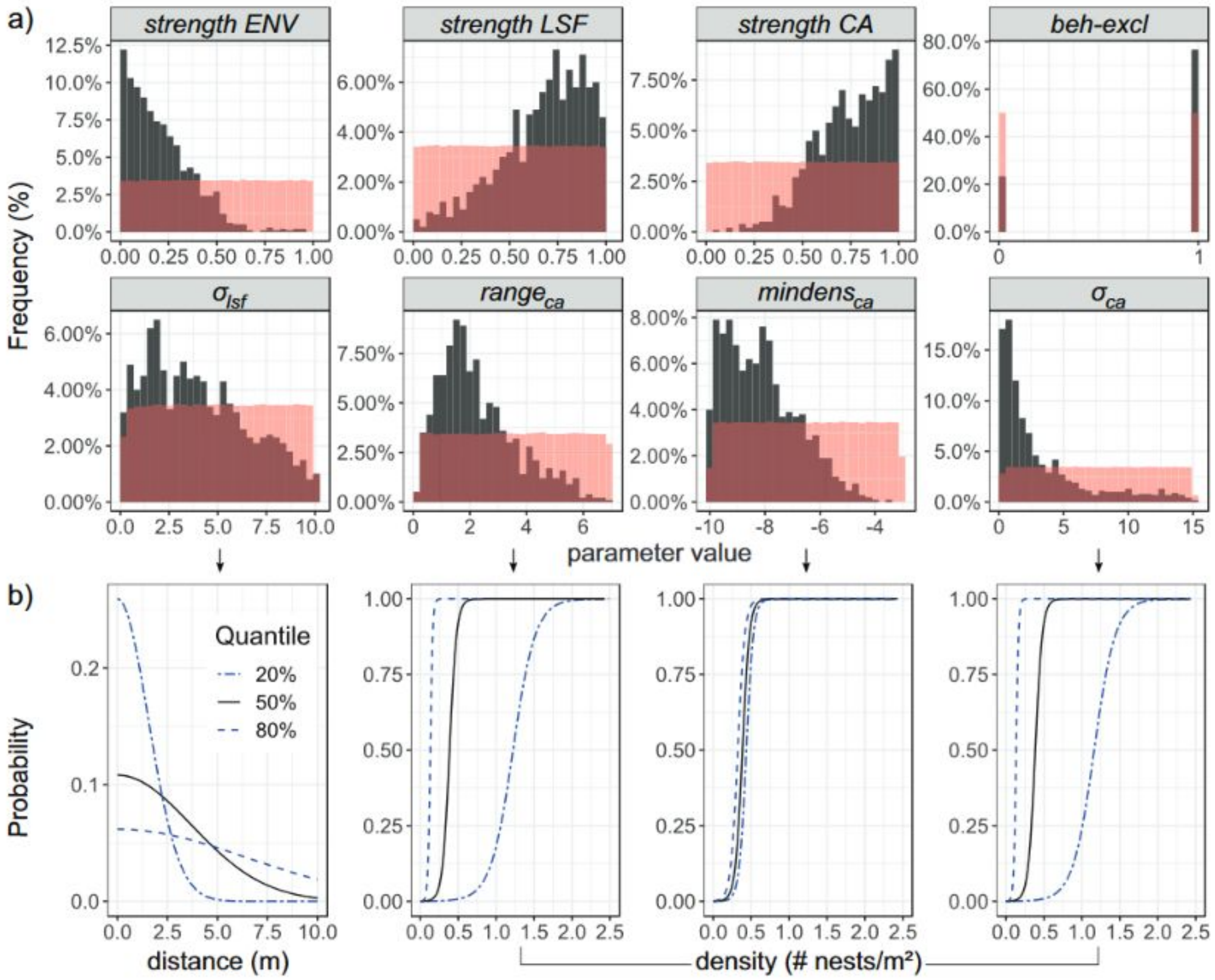
Quick example

Prior & posterior summary statistics



Quick example

Posteriors of model parameters



Quick example

Model selection

hypotheses on variation of mechanisms across population

Table 1: Bayes factors (BFs) and proportions of accepted models for model selection with approximate Bayesian computation (ABC) analysis

	Random	Population	Interindividual	Intraindividual	% accepted simulation
Random	–	.00	.00	.00	0
Population	∞	1.00	.31	.36	14.3
Interindividual	∞	3.20	1.00	1.14	45.7
Intraindividual	∞	2.80	.88	1.00	40.0

Note: The ABC analysis retained the 1,000 best simulations of 1,000,000 (0.1%). The submodels represent at which level the mechanisms can vary: population, interindividual, or intraindividual. BFs are the ratios of the posterior probabilities of two models, indicating the strength of evidence for model M_1 (rows) relative to model M_0 (columns), given the data. Evidence categories according to Kass and Raftery (1995) are as follows: $BF < 1$ indicates more evidence for M_0 than M_1, $1 < BF < 3$ indicates weak evidence for M_1 compared with M_0, and $3 < BF < 10$ indicates substantial evidence for M_1 compared with M_0.

ABC conclusions

The algorithm itself is actually easy to implement!

Comes with a lot of caveats and uncertainties

→ needed to perform prior predictive checks, weight summ stats,...

→ a lot of simulations needed!

Needs well-balanced hypotheses formulation

Want to learn how to implement ABC?

Van der Vaart et al. (2015, 2016) Ecological modelling

Ask for help!

One of the next coding clubs?

VARIA

Charlotte & Femke

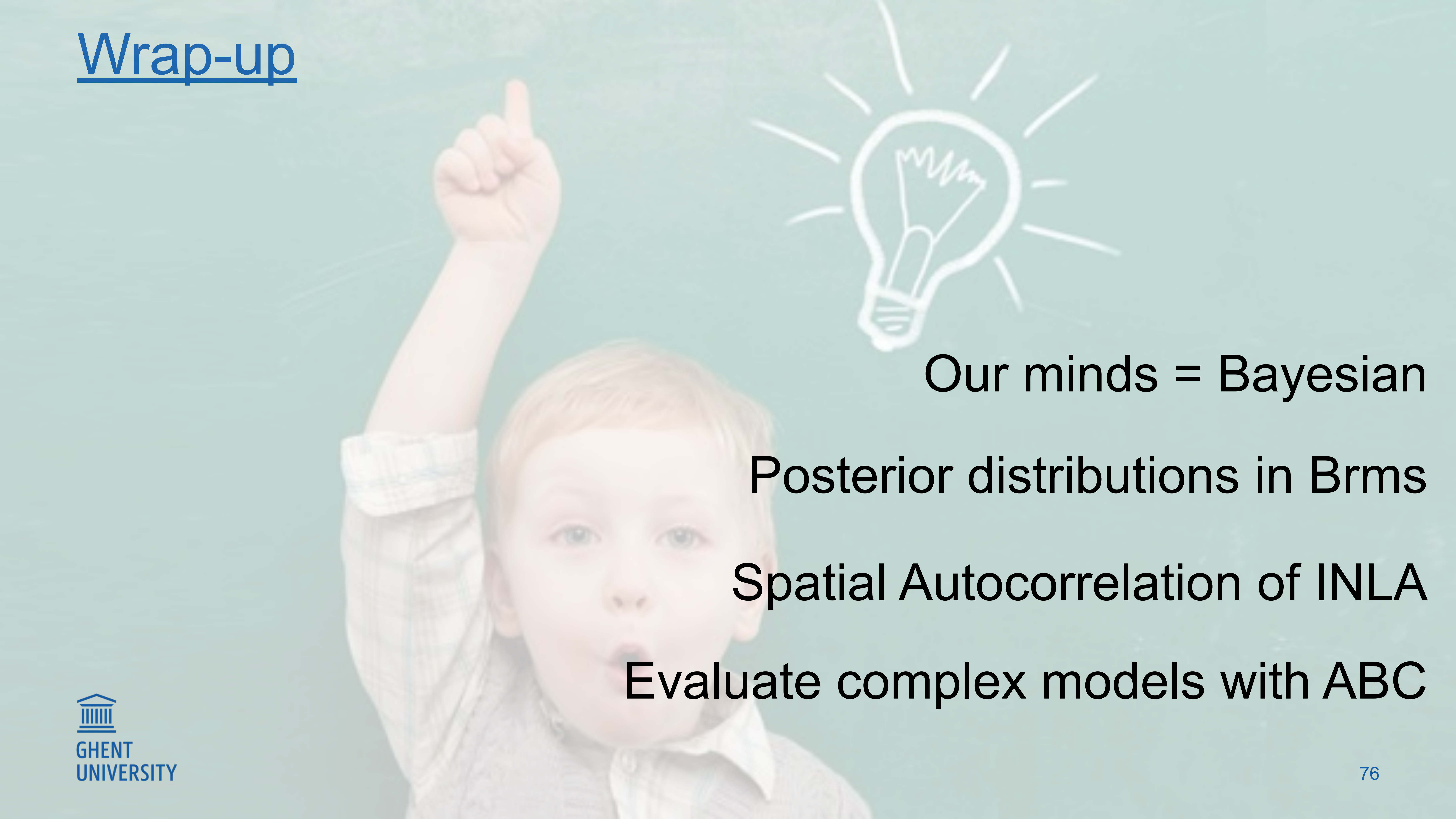
VARIA

- Next sessions: 22/4, 27/5, 17/6
- Knowledge assessment: Gitbhub?

GitHub

VARIA

- Next sessions: 22/4, 27/5, 17/6
- Knowledge assessment: Github?
- Next session theme: Github + Bayesian Deep Dive
- Other desired themes?
- Q&A
- Wrap-up



Wrap-up

Our minds = Bayesian

Posterior distributions in Brms

Spatial Autocorrelation of INLA

Evaluate complex models with ABC

GHENT UNIVERSITY

FACULTY OF SCIENCES

REMAINING QUESTIONS ON THIS SESSION?

E charlotte.vandriessche@ugent.be

E femke.batsleer@ugent.be

E fremorti.mortier@ugent.be

GHENT UNIVERSITY