10

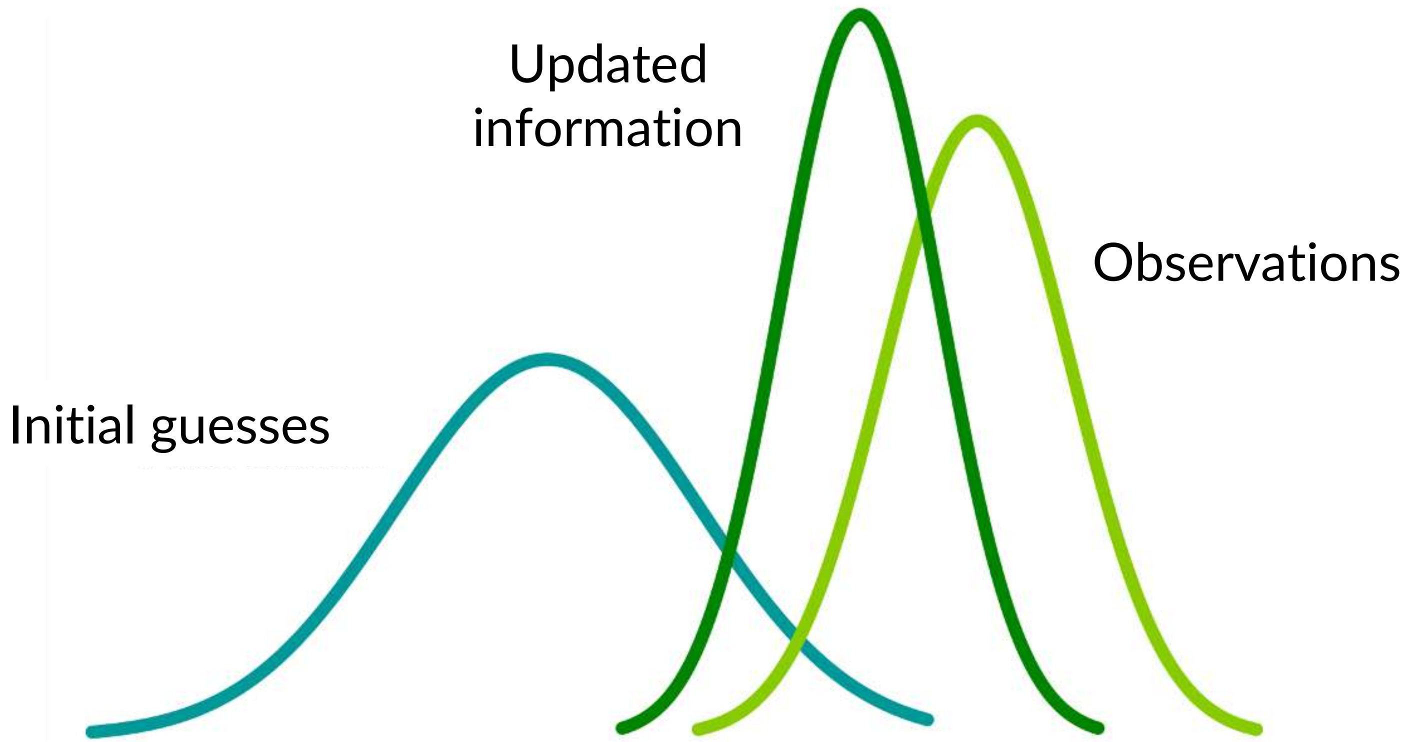
Bayesian model calibration

D. Nadal, S. Sabaté

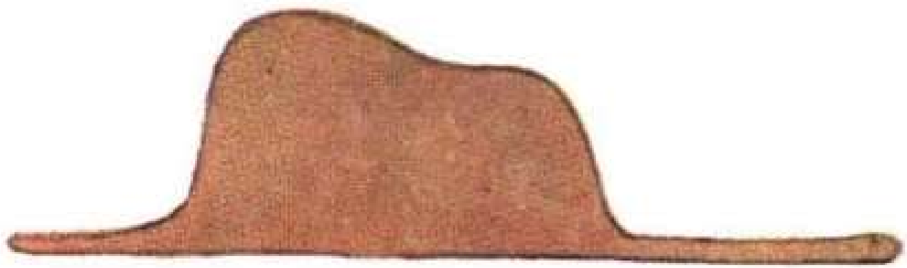
Department of Evolutionary Biology, Ecology and Environmental Sciences. University of Barcelona and CREAF

Ecology and Forest Management, 2024-25

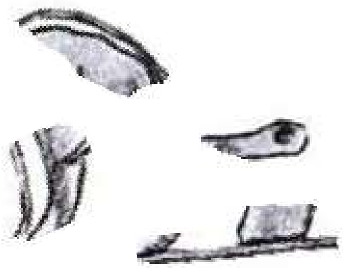
The Bayesian way of life



Integrating information iteratively

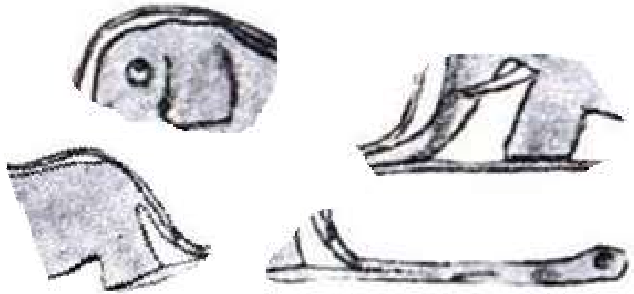


"**A priori**" guess: It can either be a hat or a snake that has eaten an elephant



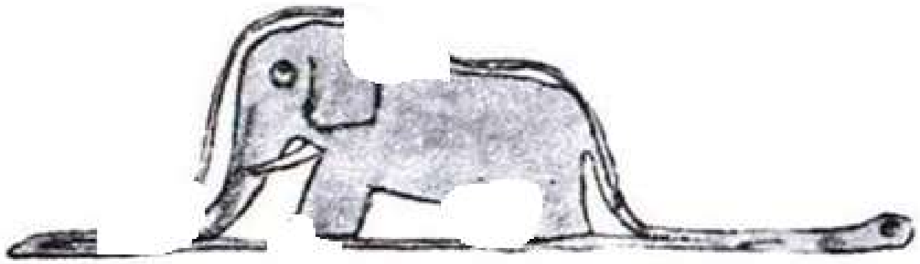
Initial observations

"a posteriori" knowledge



Updating observations

After repeating the process "n" times:



Our "**a posteriori**" knowledge, based on our "**a priori**" guess and fundamented on **observations**, is that a snake has eaten an elephant.

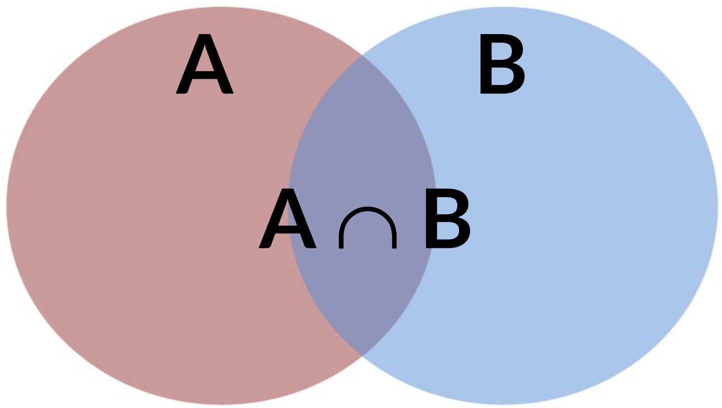
Thomas Bayes (1702-1761)

Inverse probability: The “true” distribution for a fact is achieved by iteratively integrating observations.

Bayes' Theorem

$$p\,(B \mid A)\quad p\,(A) \quad = \quad p\,(A \mid B)\quad p\,(B)$$

- $p\,(B \mid A)$: Probability of a **B** event giving **A**
- $p\,(A)$: Probability of a **A** event
- $p\,(A \mid B)$: Probability of a **A** event giving **B**
- $p\,(B)$: Probability of a **B** event



Example of application

After a test for a severe disease which only the 0.1% of population has, the doctor tells you that the test is positive, and its accuracy is 99%

Which are the odds that you actually have the illness?

It's a 99%, right?

NO!

Let's do some Bayesians…

Re-arranging Bayes' Theorem (I)

$$p(D \mid H+) = \frac{p(H+ \mid D) * p(D)}{p(H+)}$$

- $p(D \mid H+)$: Probability of having the disease **(D)** giving a positive result **(H+)** — **POSTERIOR**
- $p(H+ \mid D)$: Probability of having a positive result **(H+)** when having the disease **(D)** — **LIKELIHOOD**
- $p(H+)$: Probability of testing positive **(H+)**
- $p(D)$: "*A priori*" probability of having the disease **(D)** — **PRIOR**

Re-arranging Bayes' Theorem (II)

$$p(D \mid H+) = \frac{p(H+ \mid D) * p(D)}{p(D)*p(H+ \mid D) + p(D-)*p(H+ \mid D-)}$$

Probability of testing positive **(H+)** = Probability of testing positive **(H+)** having the disease **(D)** + Probability of testing positive **(H+)** without having the disease **(D-)**

So let's solve

$$p(D \mid H+) = \frac{0.99 * 0.001}{0.001 * 0.99 + 0.999 * 0.01}$$

0.09

Probability of testing positive **(H+)**	=	Probability of testing positive (H+) while having the disease	+	Probability of testing positive (H+) without having the disease

So, after a single positive test –i.e. without any PRIOR information, the chances we have the disease are 9%

What if a second test is also positive?

$$p(D \mid H+) = \frac{0.99 * 0.09}{0.09 * 0.99 + (1-0.09) * 0.01}$$

0.91

Probability of testing positive (H+) = Probability of testing positive (H+) while having the disease + Probability of testing positive (H+) without having the disease

So, after the second positive test, chances that we have the disease are 91%

What if the second test is negative?

$$p(D \mid H-) = \frac{0.99 * (1-0.09)}{(1-0.09) * 0.99 + 0.09 * 0.01}$$

0.9989

Probability of testing negative **(H-)** = Probability of testing negative (H-) while not having the disease + Probability of testing negative (H-) while having the disease

So, if the second test is negative, the probability of having the disease is 0.11%

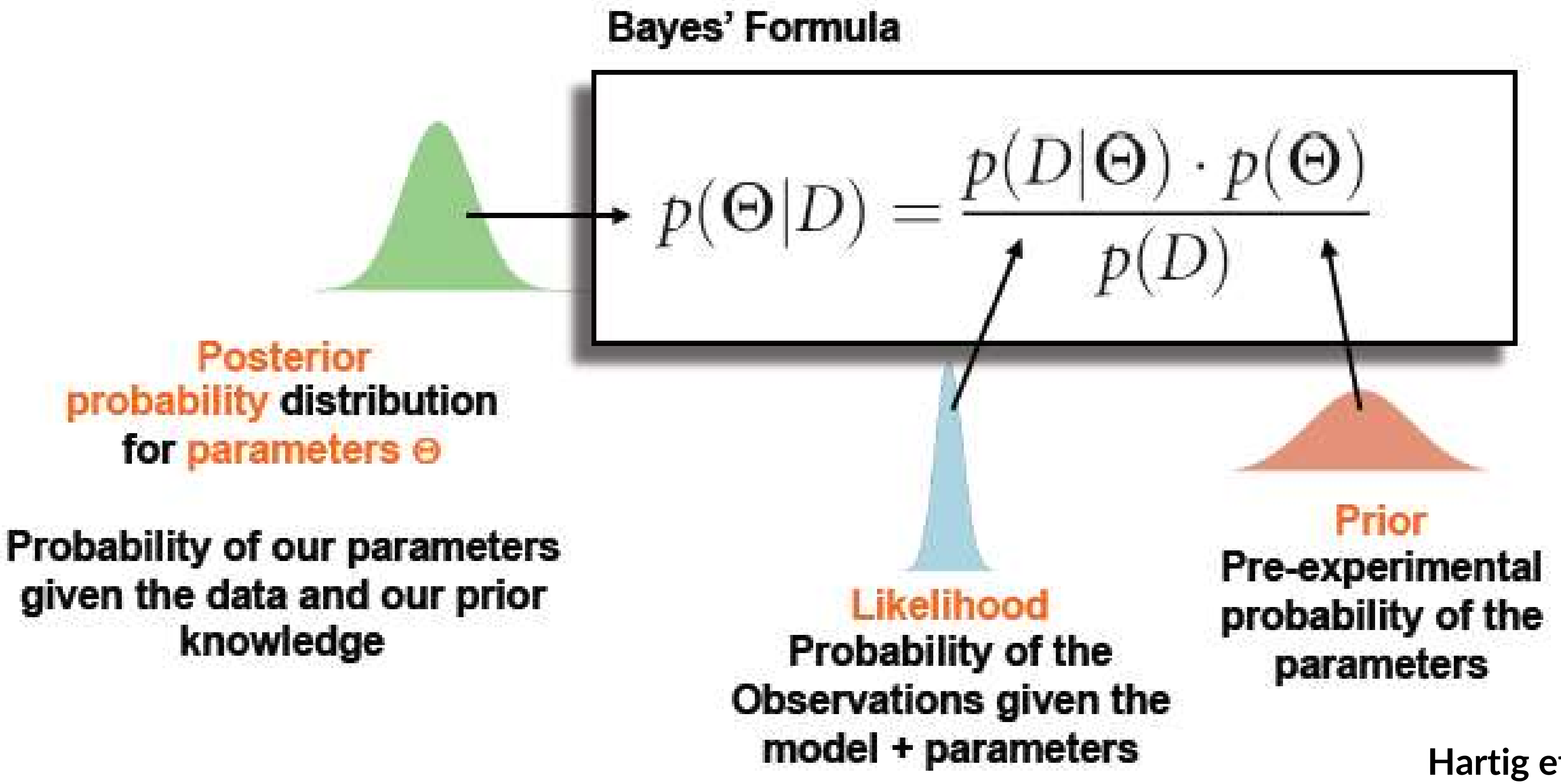
Summary

From our PRIOR knowledge, and providing the LIKELIHOOD of our observations, we can obtain a POSTERIOR probability estimate.

We build our knowledge by iteratively integrating observations.

Let’s follow the white rabbit

Model Bayesian inversion



Hartig et al., 2012

We can obtain our **POSTERIOR** parameter distribution from our **PRIOR** knowledge about parameters and model fit to observations (i.e. **LIKELIHOOD**) for a given set of parameters.

Advantages and disadvantages:

Advantages of Bayesian inverse modeling:

-Allows to include “a priori” observations in the model.

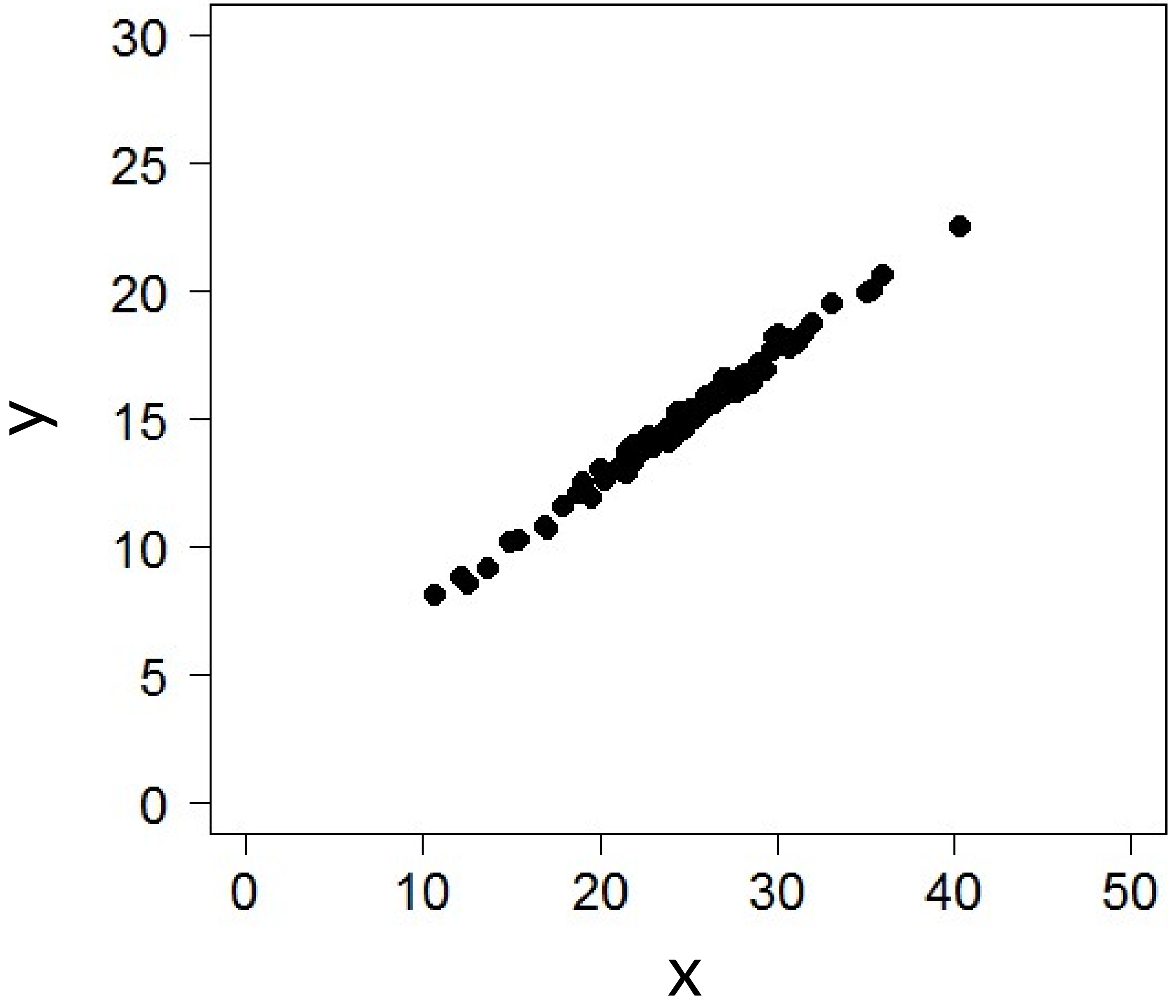
-Not only accounts for the best parameter estimates, but also includes parameter uncertainty.

-Robust parameter estimates in complex models.

Disadvantages:

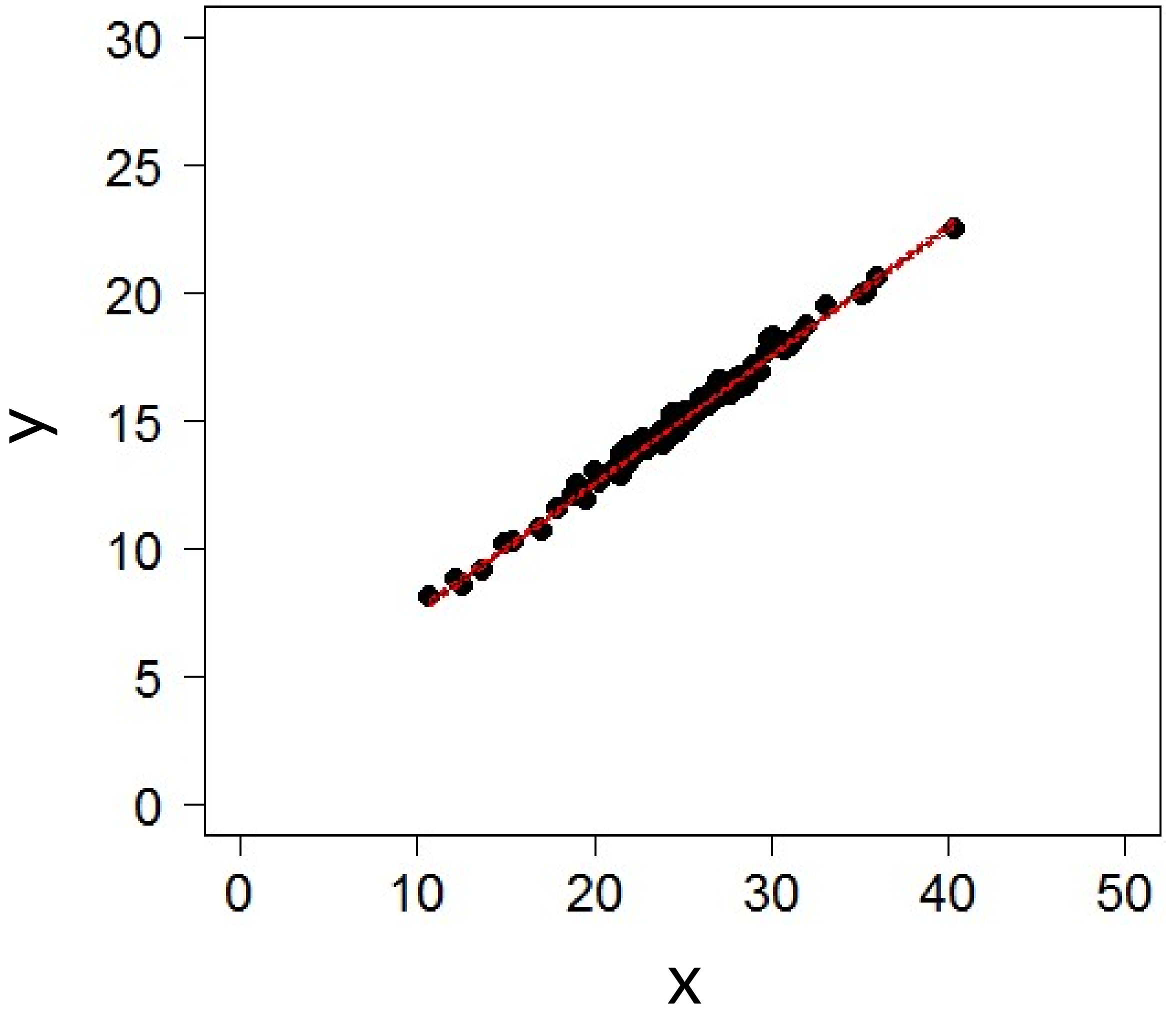
-Elevated computational costs.
-Dominancy of the frequentist approach.

Example: Simple model application



Dataset: n = 100
a = 2.5
b = 0.5
random = N(0,0.25)

By using Least Squares



Dataset: n = 100

$a_{estimate}$ = 2.48(0.12)

$b_{estimate}$ = 0.5(0.005)

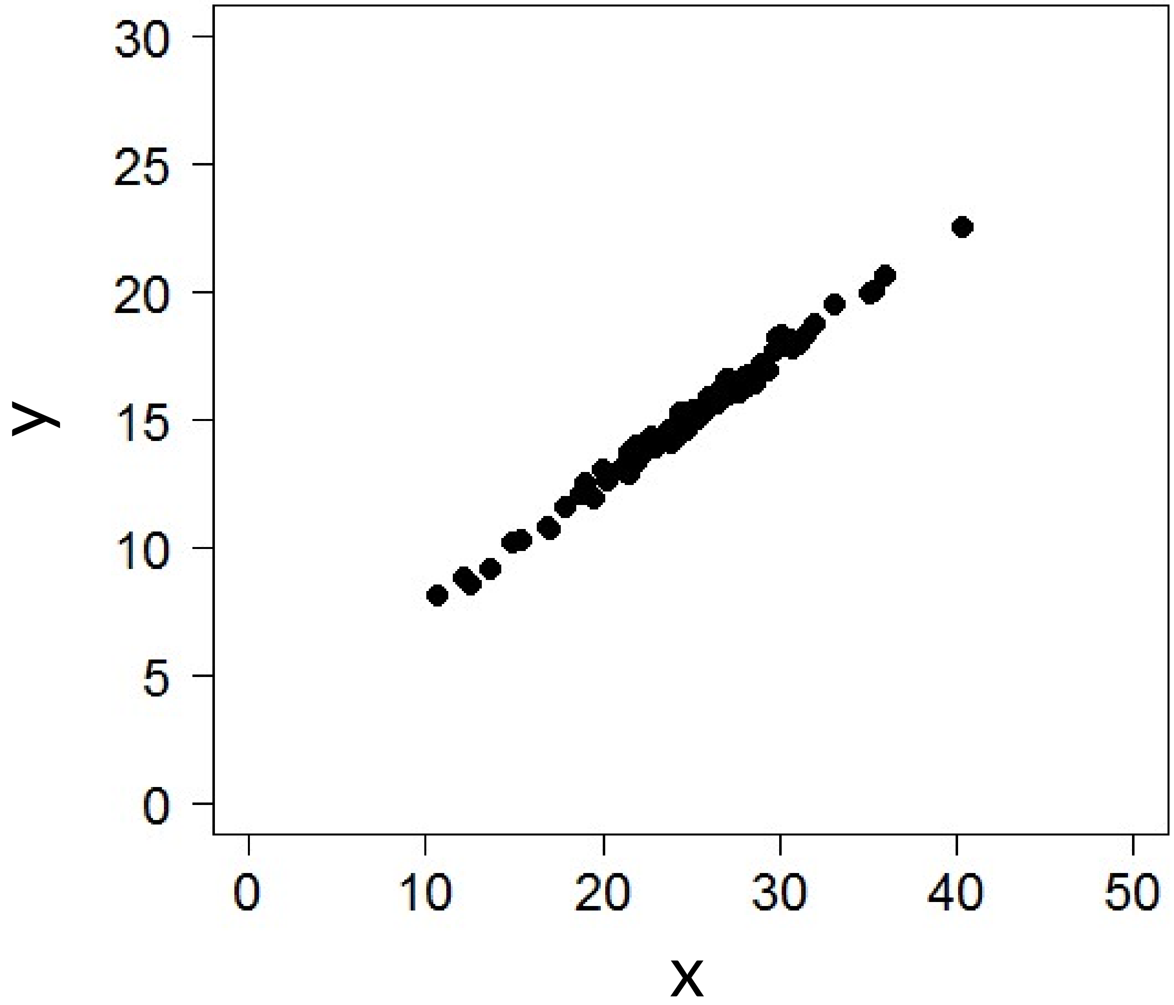
Least Squares assumptions

- Independence of observations.
- Homoscedasticity -> Homogeneity of variances
- Residuals distributed according to a $N(0,\sigma^2)$.

Assumptions Bayesian inversion

- Independence of the observations.
- Prior parameter distribution.
- Likelihood function dependent on a normal distribution.

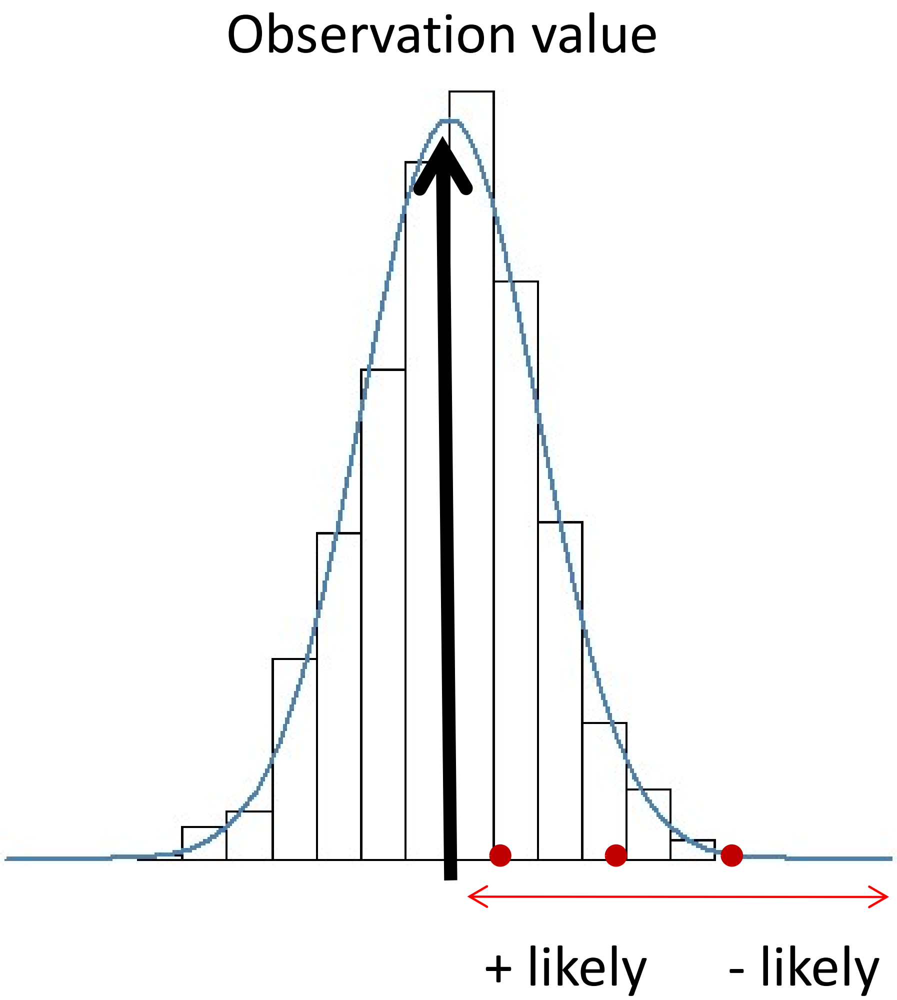
Prior parameter distribution:



Since y is always positive,
and y(10)~10
0<a<5
So b ~ U(0,5)

Similarly, while values are in the same order of magnitude, and at higher x values we find higher y values, we can assume that 0<b<10.
So b ~ U(0,10)

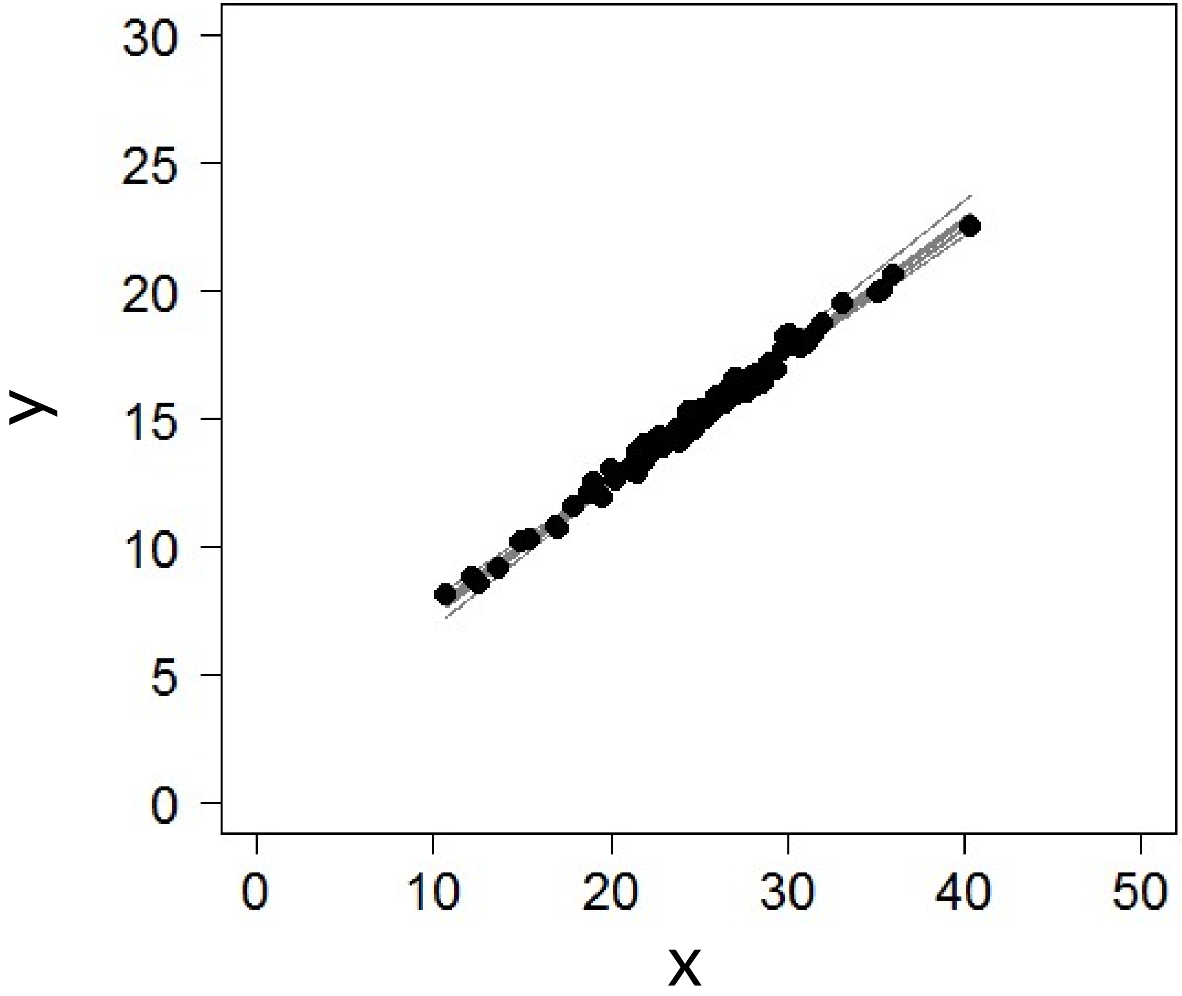
Likelihood function



For each observation, we assume a normal distribution centered in the observation, with a standard deviation according to 1/observation (in order to maintain homogeneity of the residuals).

Then, we multiply probabilities for all observations, as we are searching for the LIKELIHOOD of our model when representing the data.

Bayesian inversion



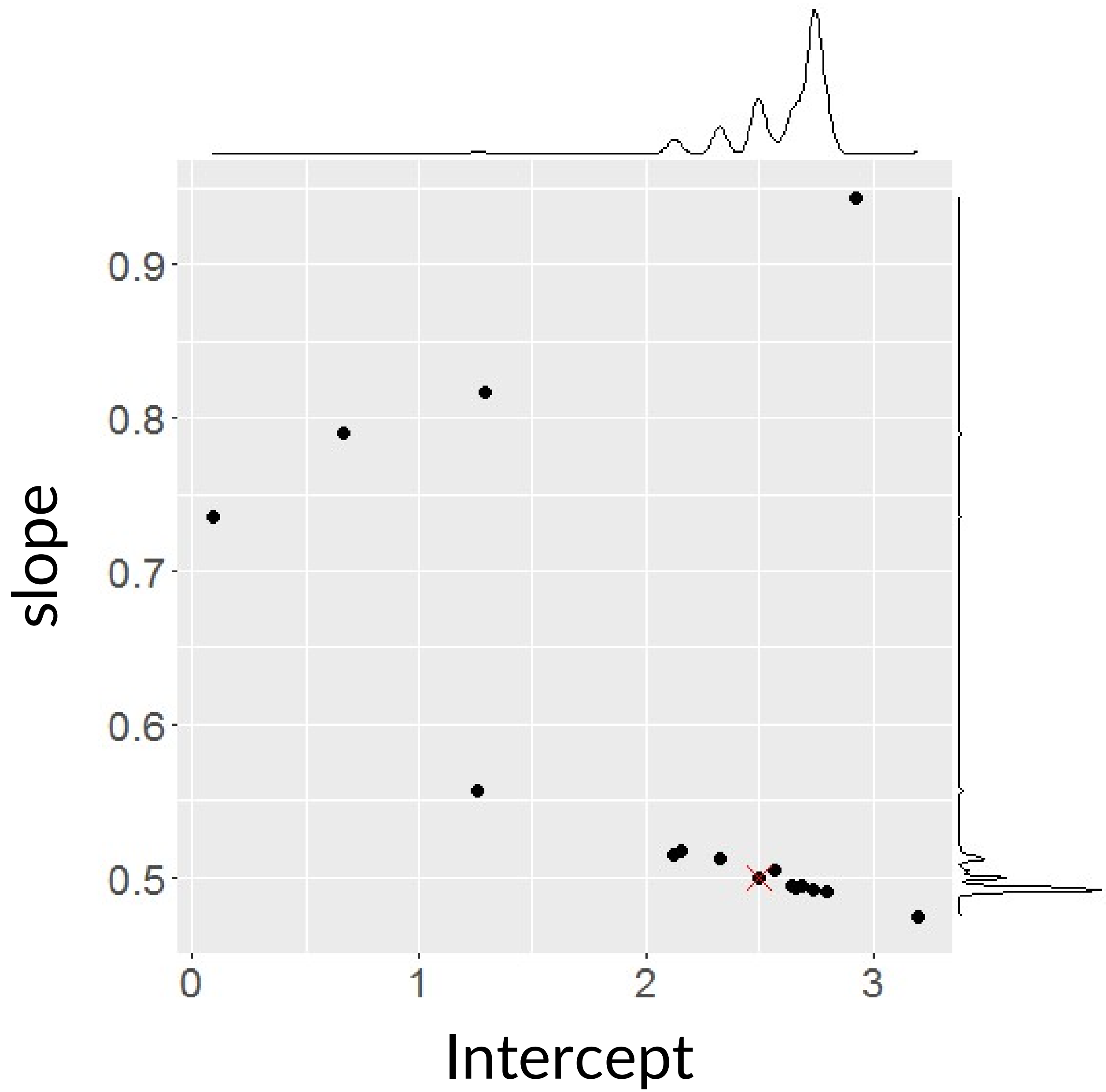
Dataset: n = 100

$a_{estimate} = p(\Phi_1 \mid D)$

$b_{estimate} = p(\Phi_2 \mid D)$

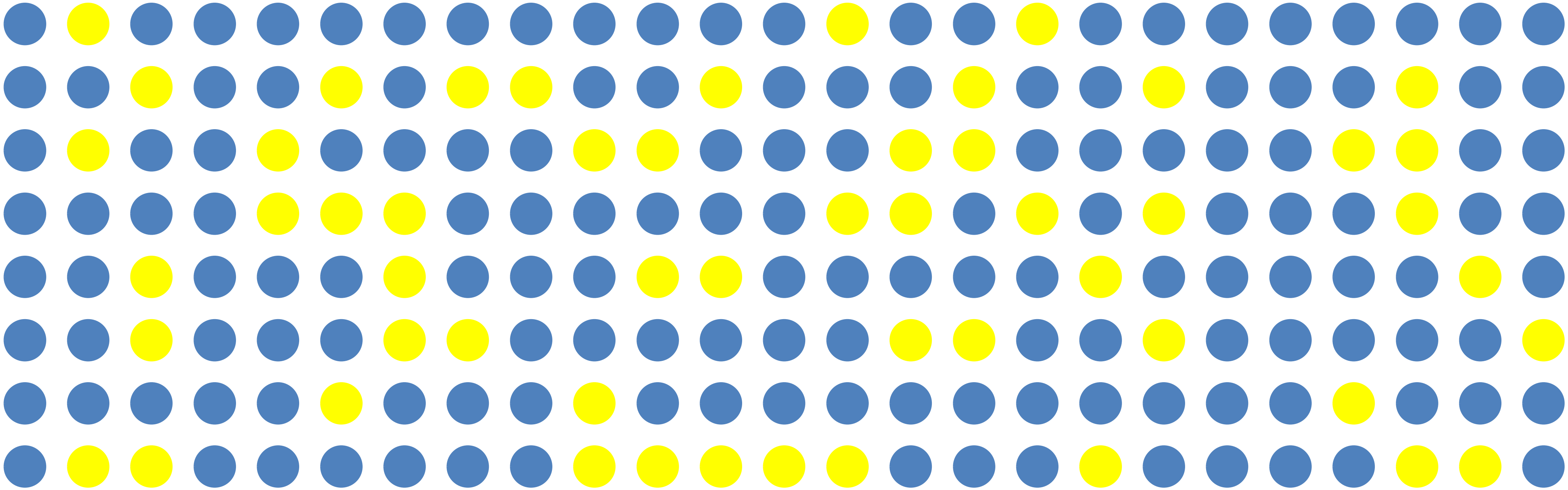
$\varepsilon_{estimate} = p(\Phi_3 \mid D)$

Posterior distribution



Example with binomial distribution

Two football teams in a city (n = 10000 people)



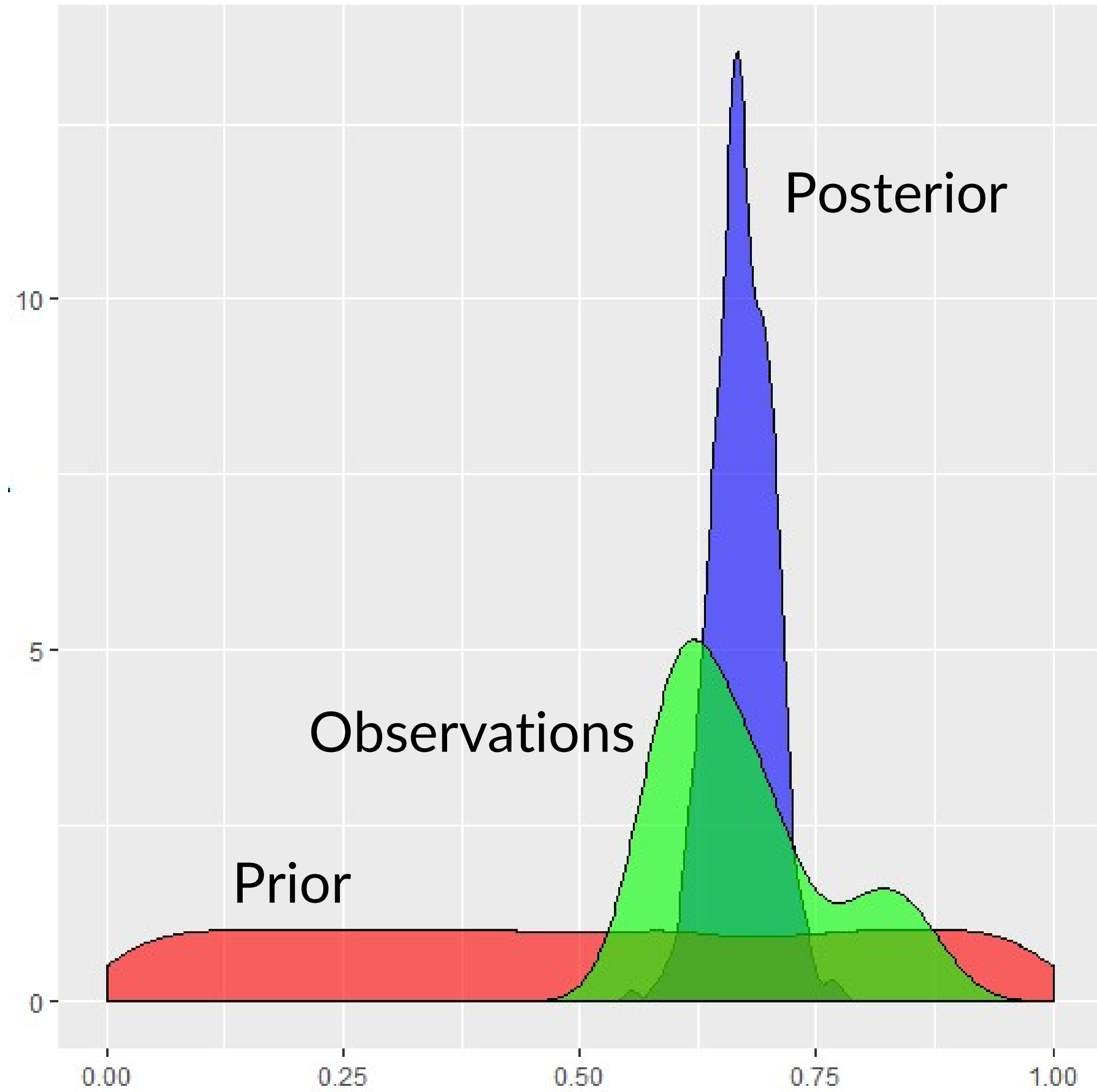
Probability for a given person being a supporter of Antidòping FC?

Binomial distribution likelihood

$$p(\text{Antidoping} \mid p, x) = \frac{n!}{(n-x)! * x!} * p^{x} * q^{(n-x)}$$

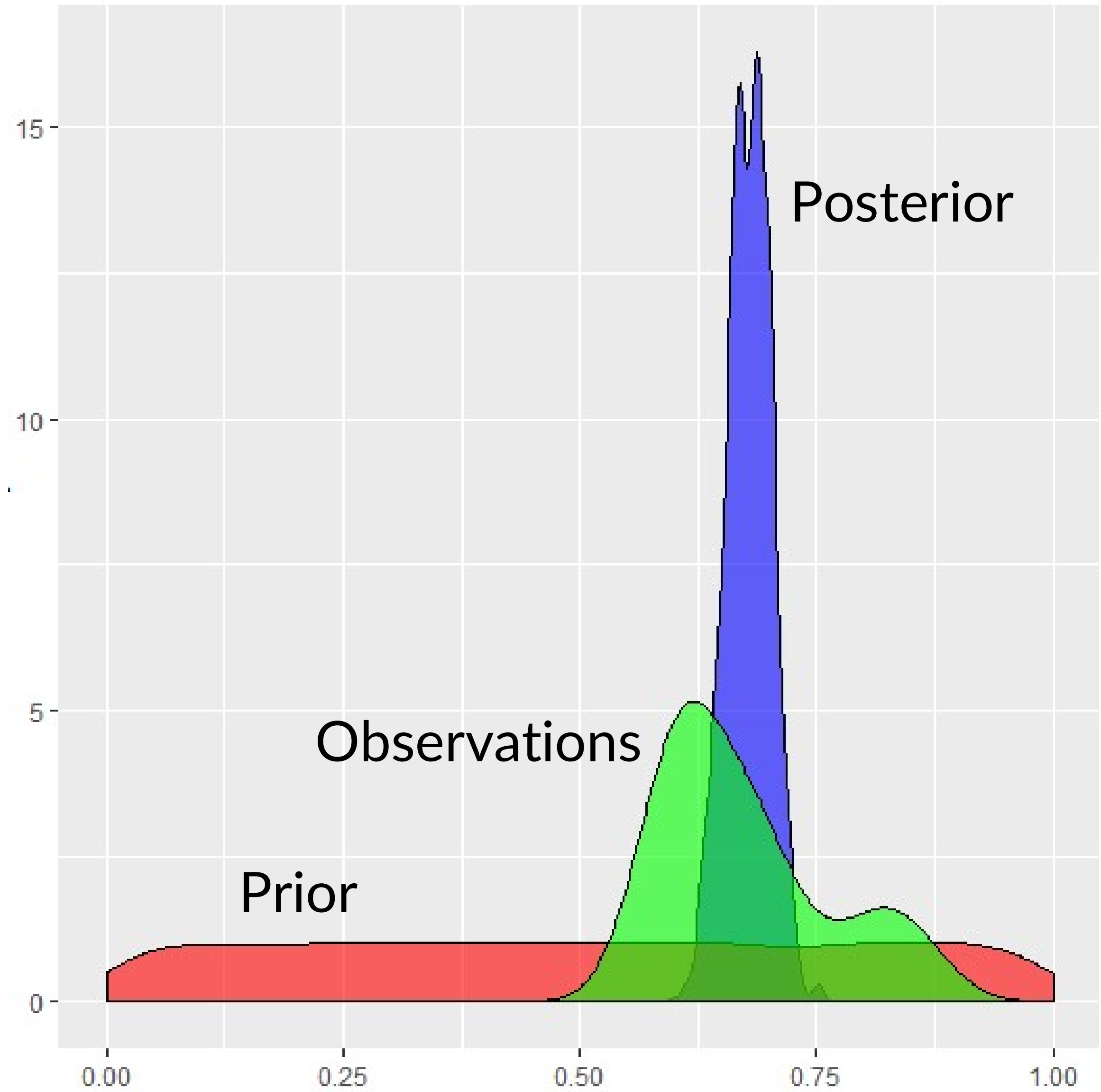
$$q = (1-p)$$

Asking 20 people 10 times:



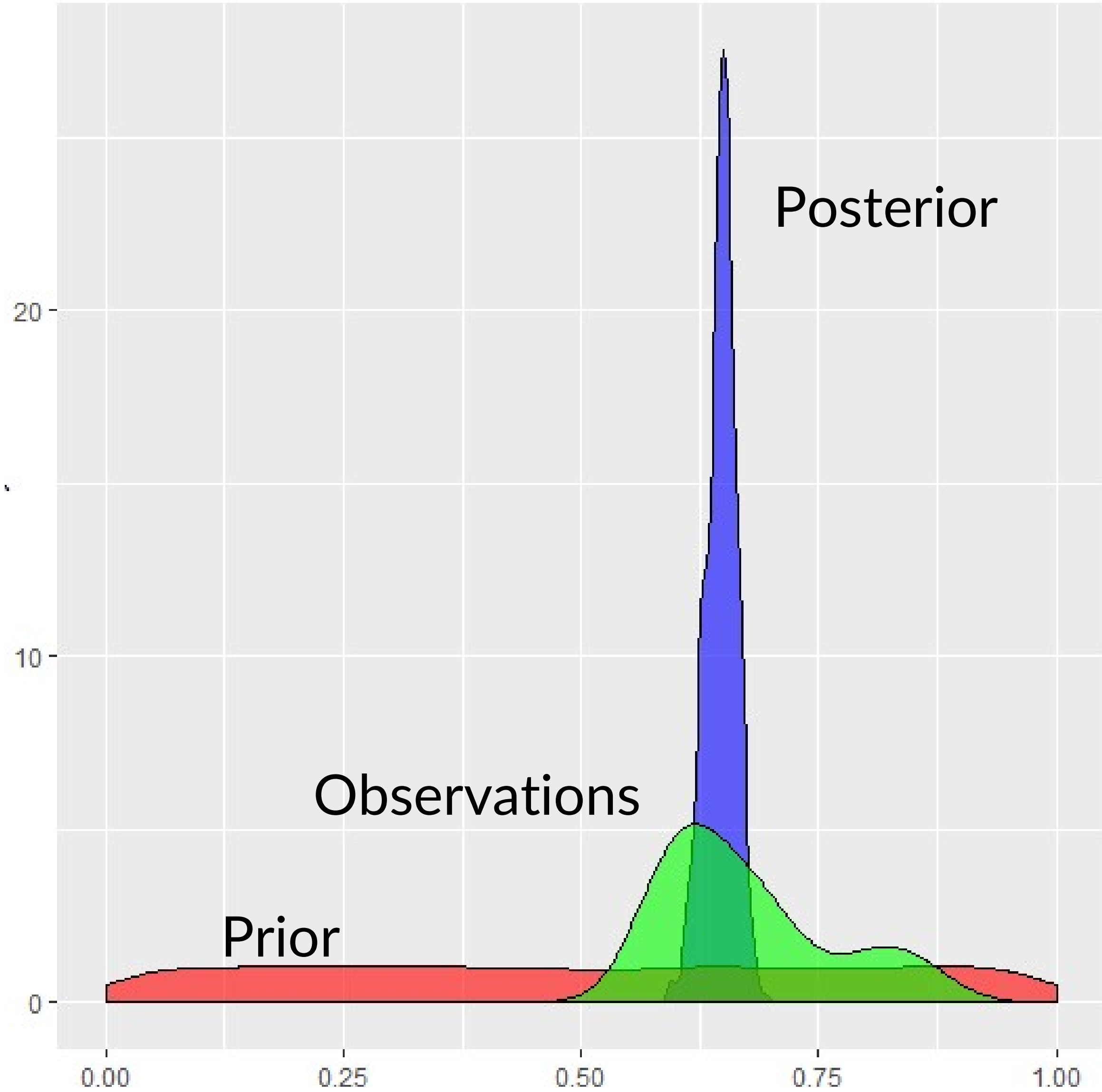
p = 0.672 (0.605-0.738)

Asking 20 people 20 times:



p = 0.68 (0.632-0.725)

Asking 20 people 50 times:



p = 0.648 (0.619-0.676)

The support to Antidòping FC is

61.9-67.6% chances of a given person supports Antidòping FC

Initial population chance: 0.65

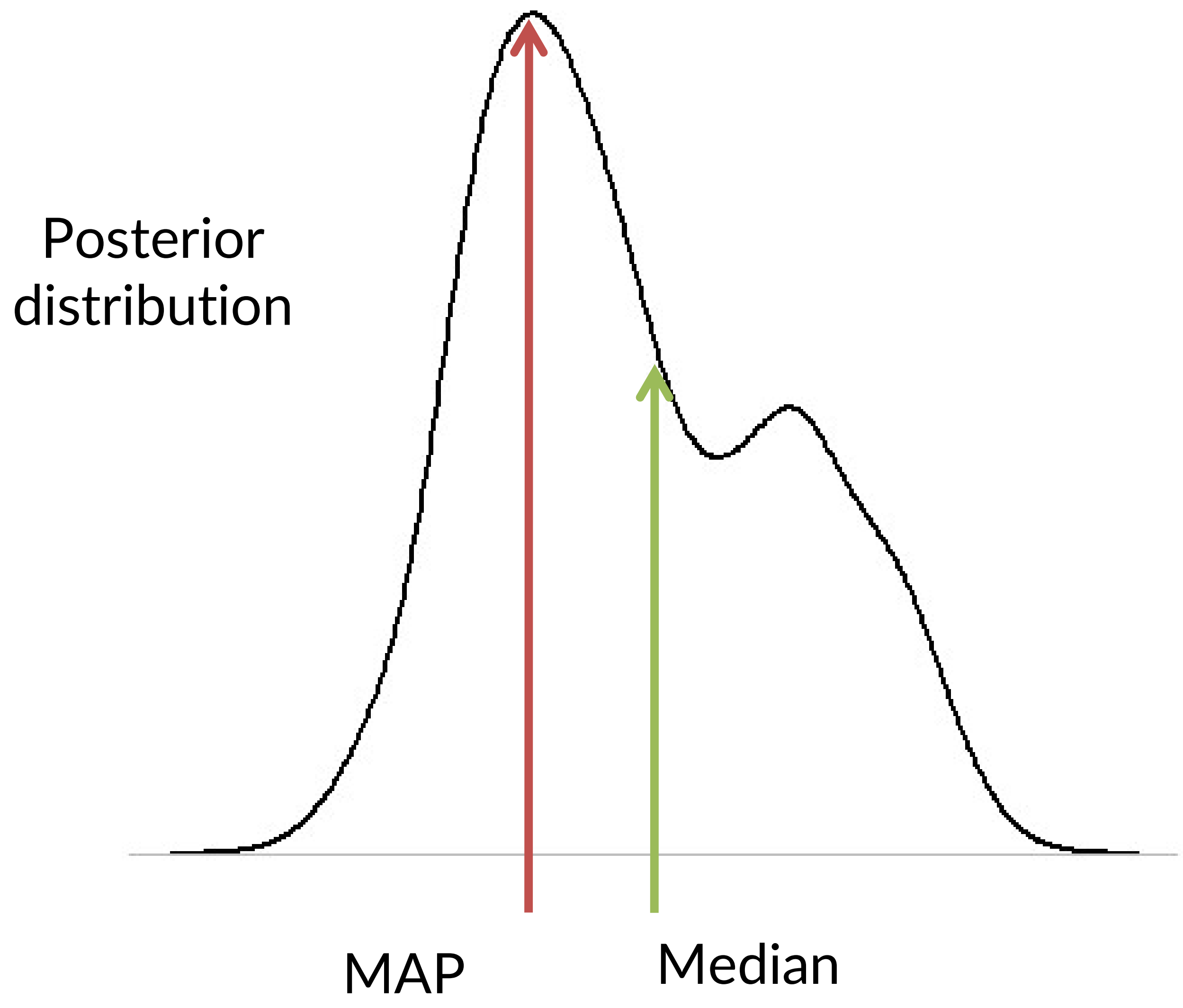
Why not simply using the MAP?

Why not simply using the parameter value that provides the most likely -i.e. less negative likelihood- value?

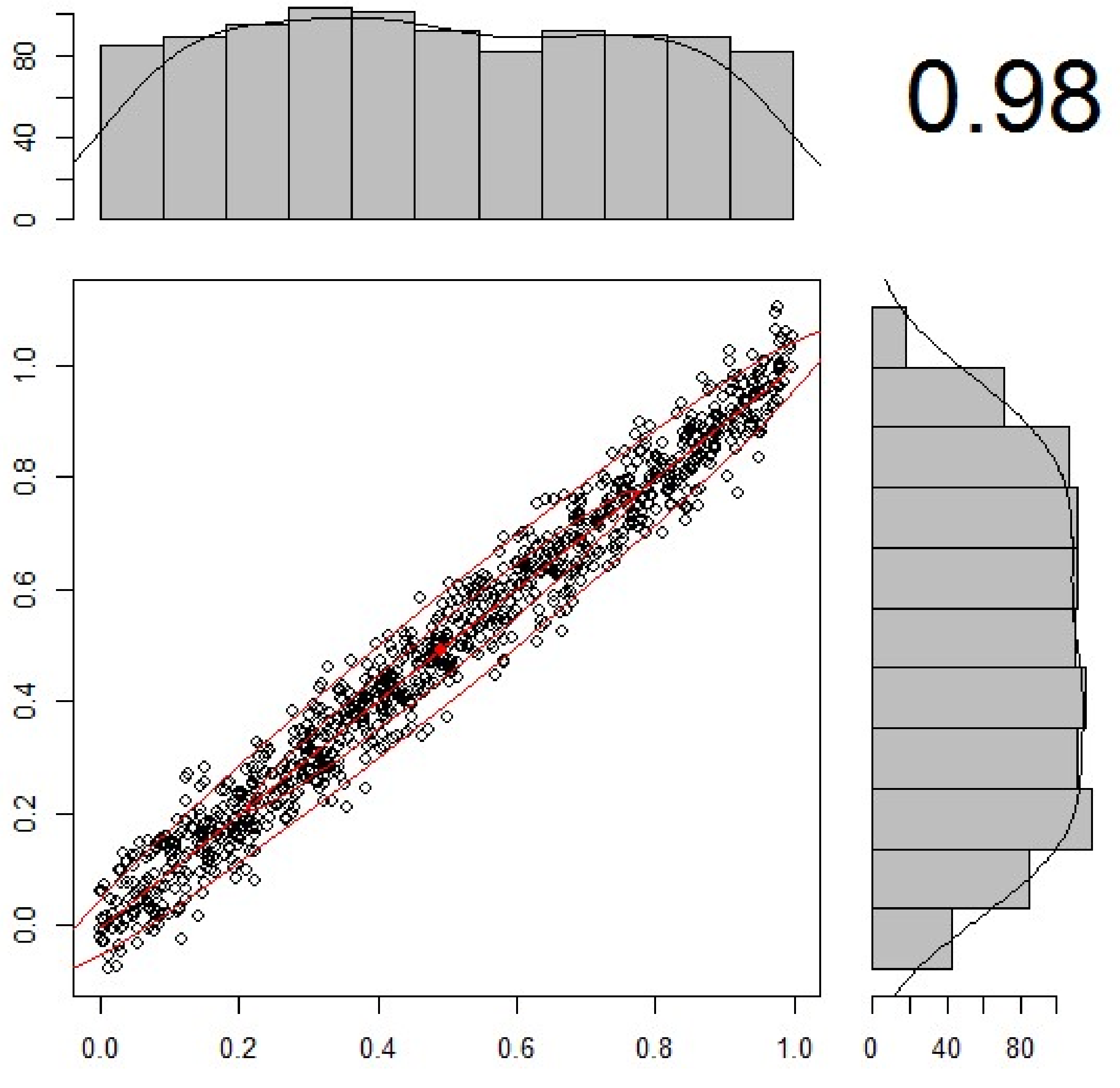
↓

This is called the **Maximum A Posteriori** likelihood (MAP)

MAP may not be representative:

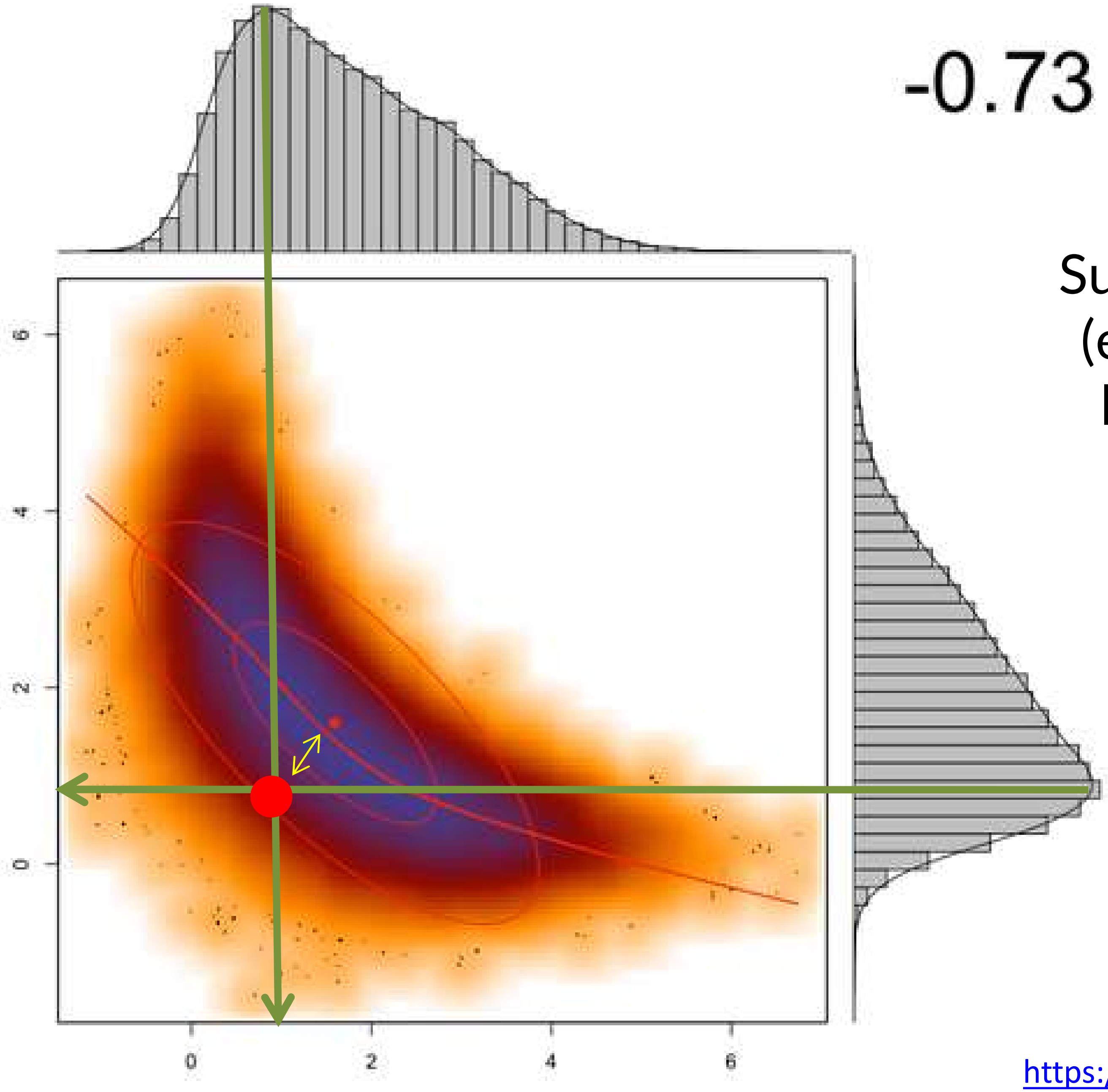


Correlations between marginal posteriors



https://github.com/florianhartig/LearningBayes

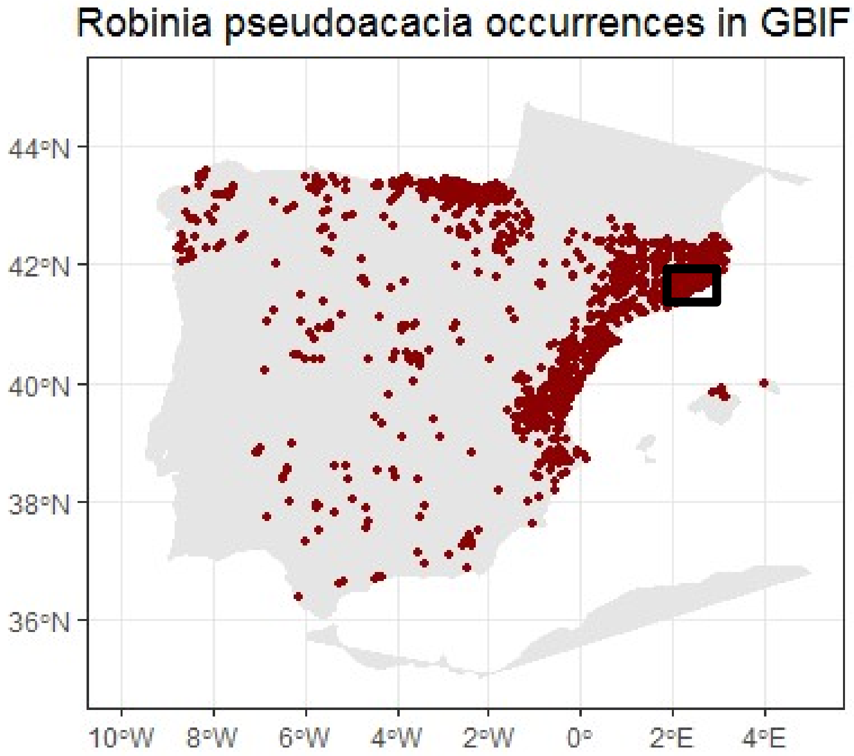
Non-linear correlations



Summarizing parameters (e.g. value at maximum likelihood) may move away model outputs from the maximum likelihood regions.

https://github.com/florianhartig/LearningBayes

Example of Bayesian inversion (I)



Nadal-Sala (PhD Thesis)

To evaluate the growth performance of two riparian tree species (*F. excelsior* and *R. pseudoacacia*) under climate change

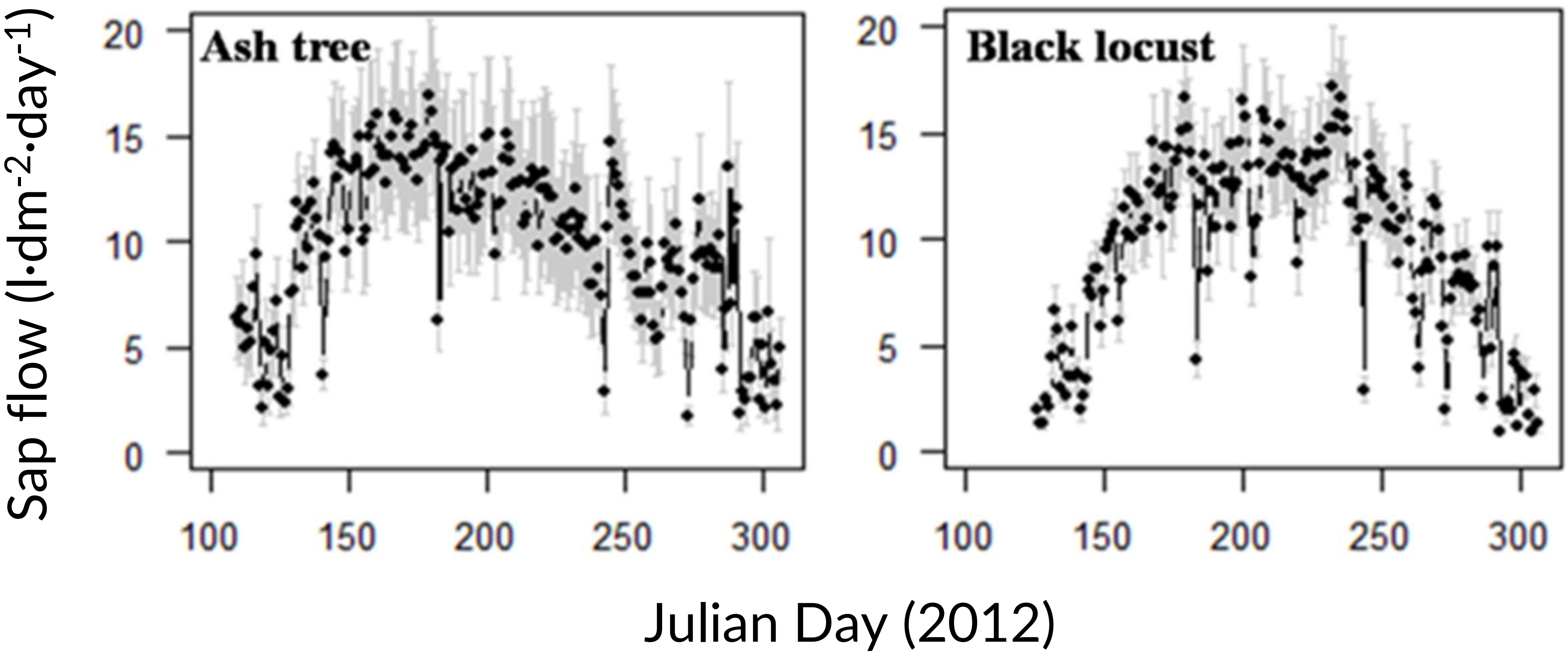
Example of Bayesian inversion (II)

Sap flow (Thermal Dissipation)

Basal Area Increment (Dendrometers)

Nadal-Sala (PhD Thesis)

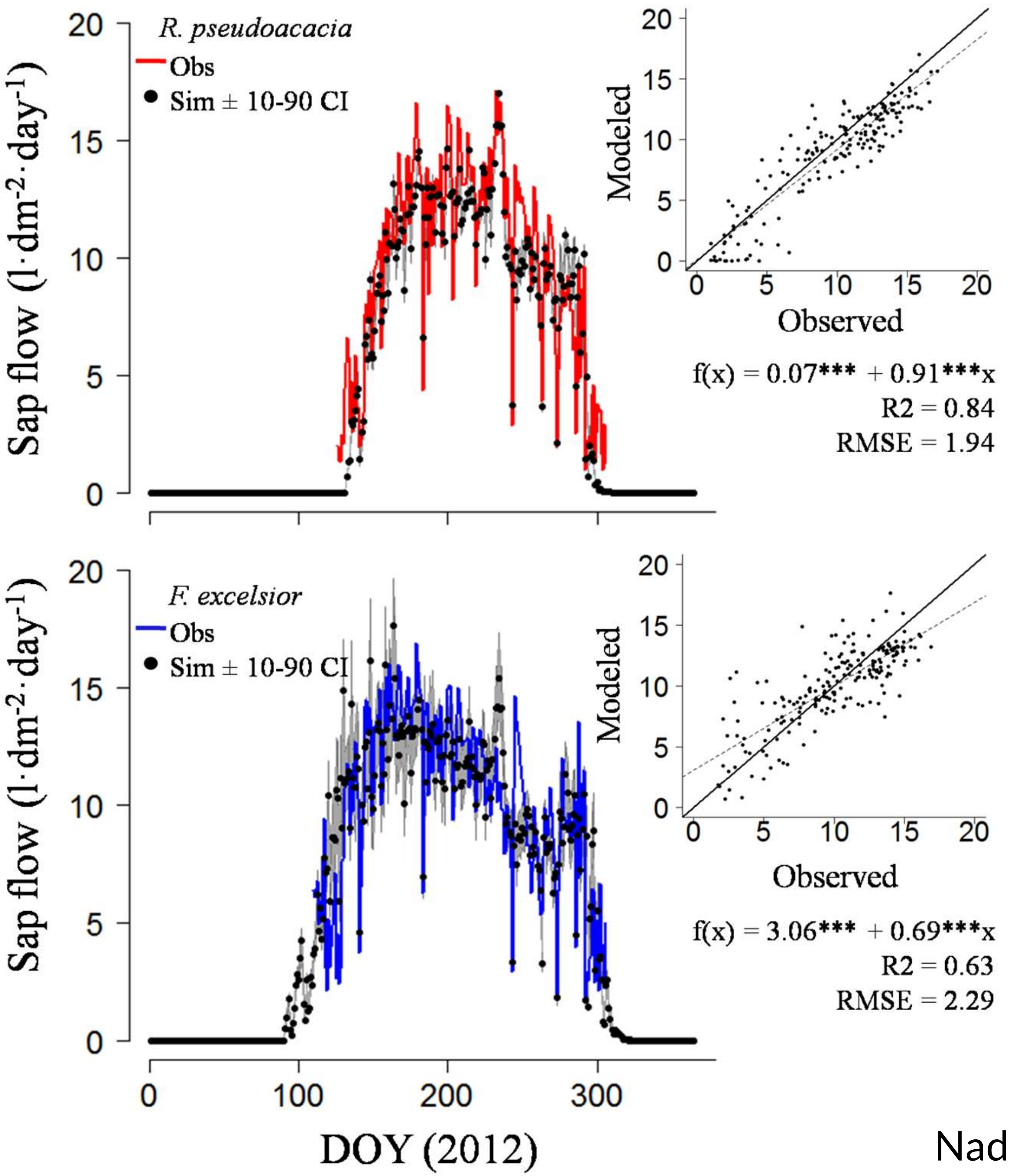
Field data



“In situ” sap flow observations in a Mediterranean riparian forest

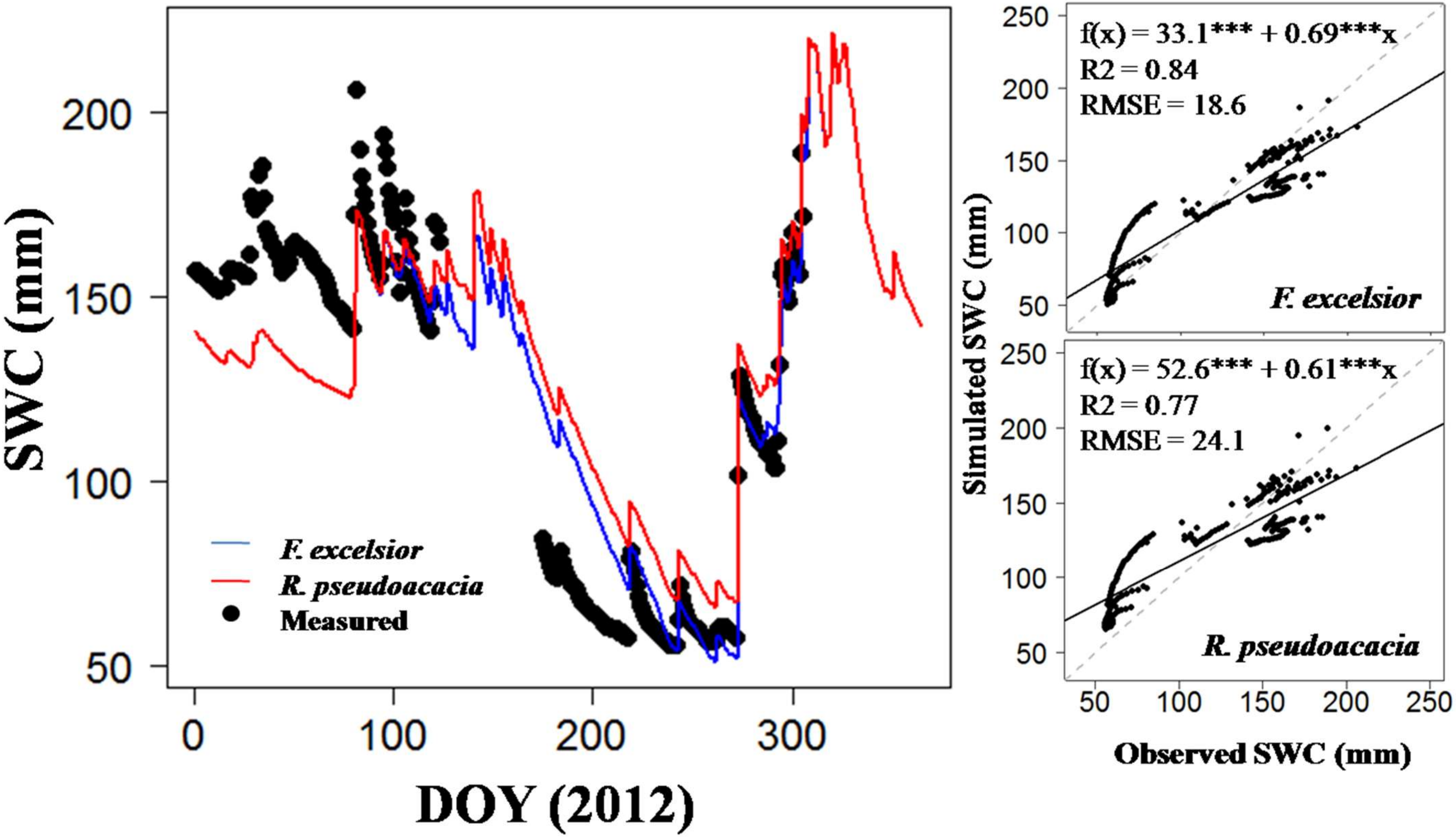
Nadal-Sala et al, (2017)

GOTILWA+ calibration



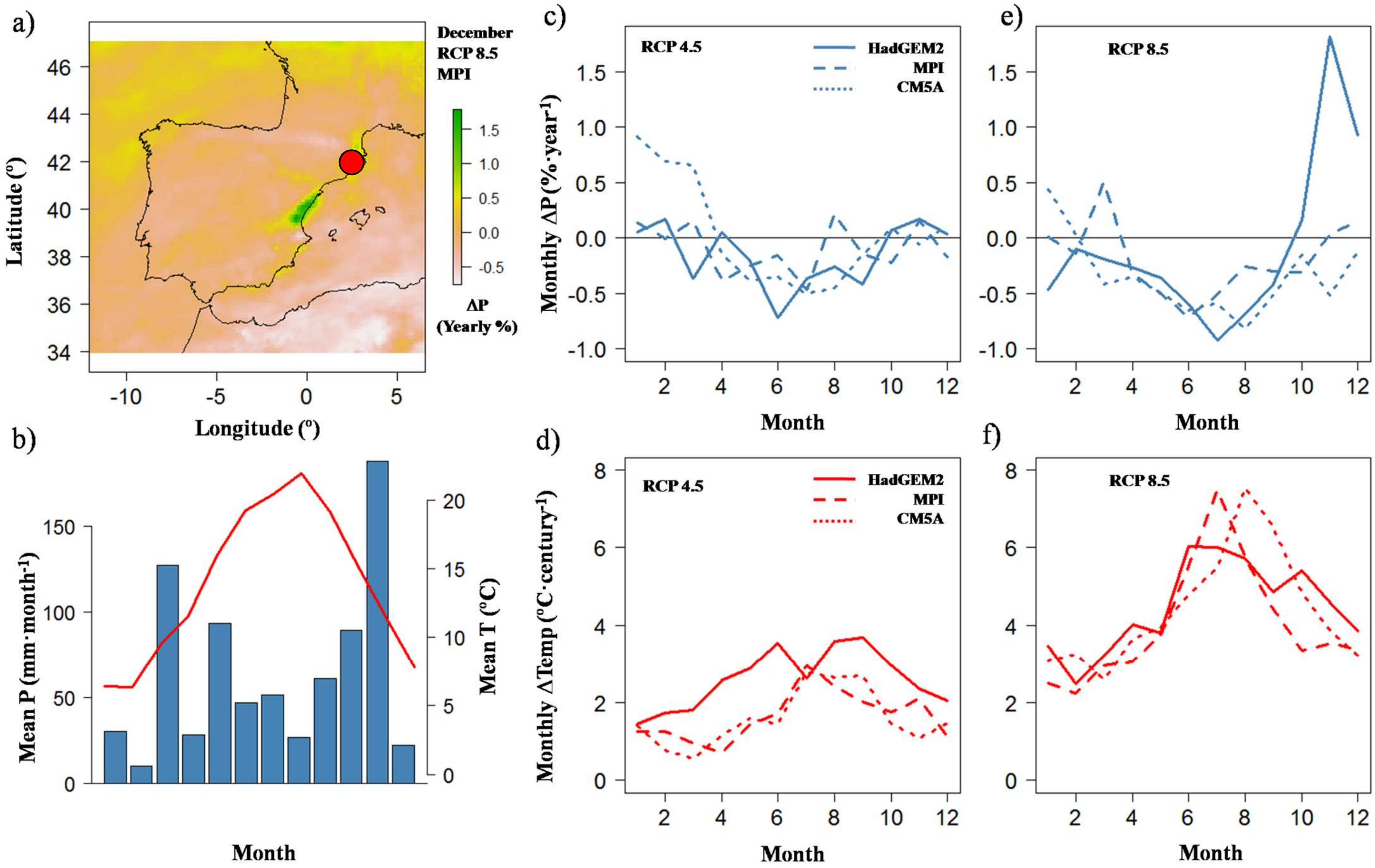
Nadal-Sala et al, (2019)

Simulation validation against SWC



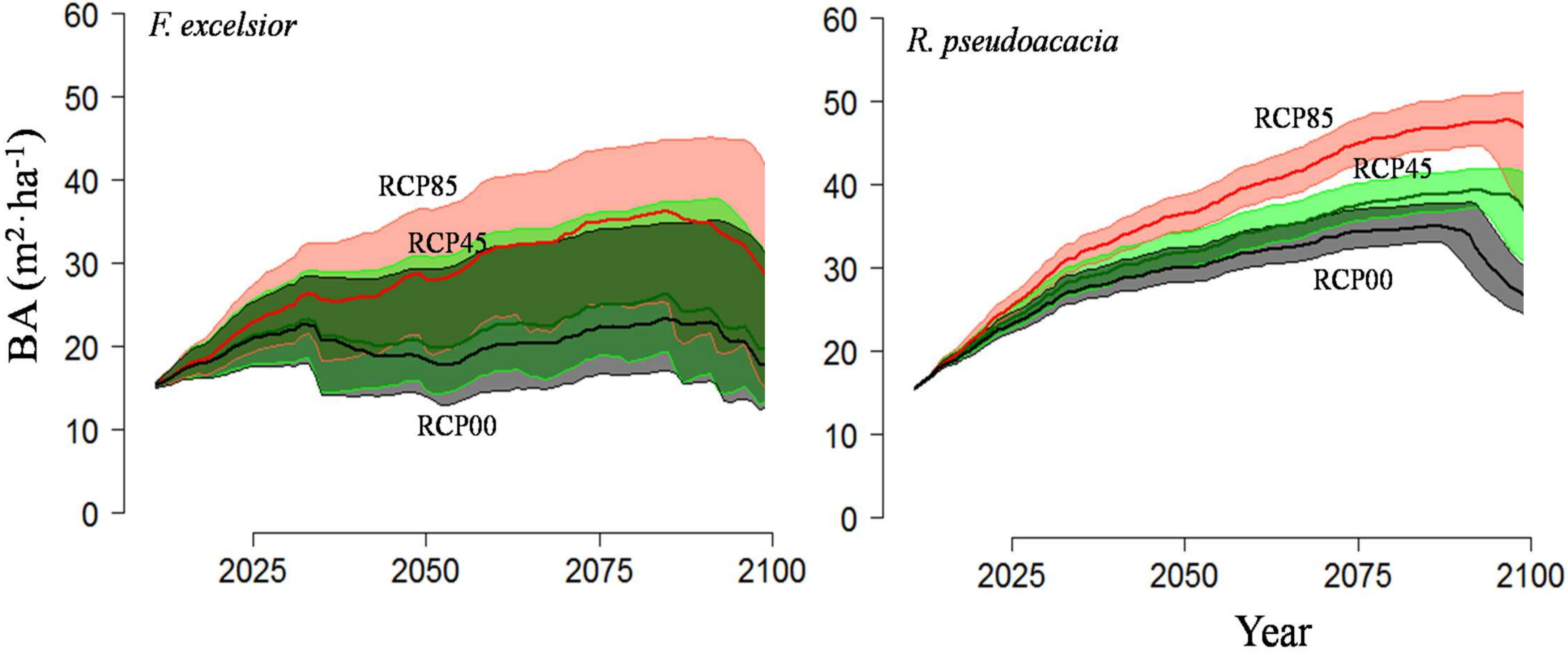
Nadal-Sala et al, (2019)

Climate change scenarios



Nadal-Sala et al, (2019)

Climate change projections



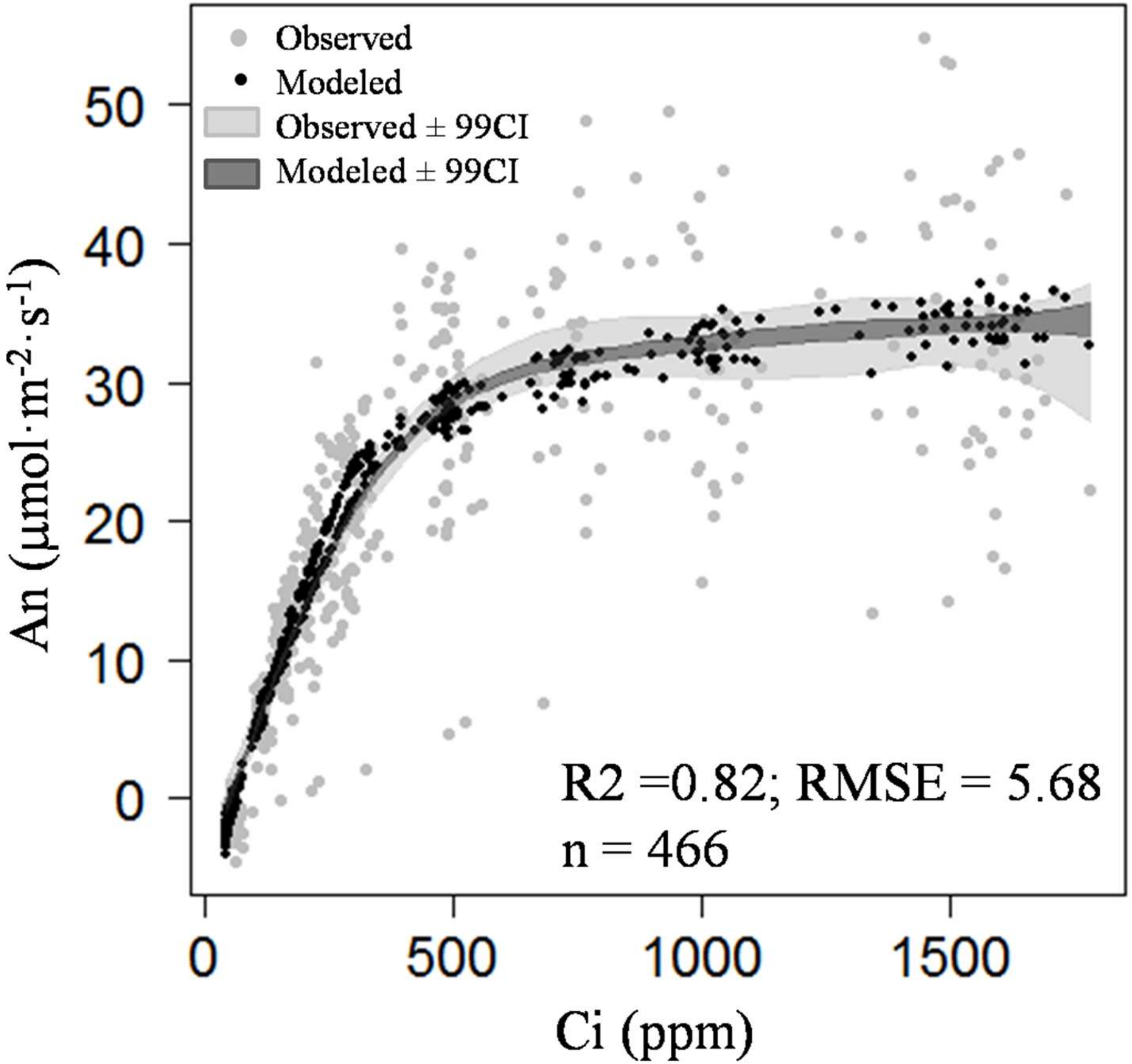
Higher increase in *R. pseudoacacia* growth than in *F. excelsior* under all climate change scenarios.

Nadal-Sala et al, (2019)

Projecting CC impacts on *E. saligna*

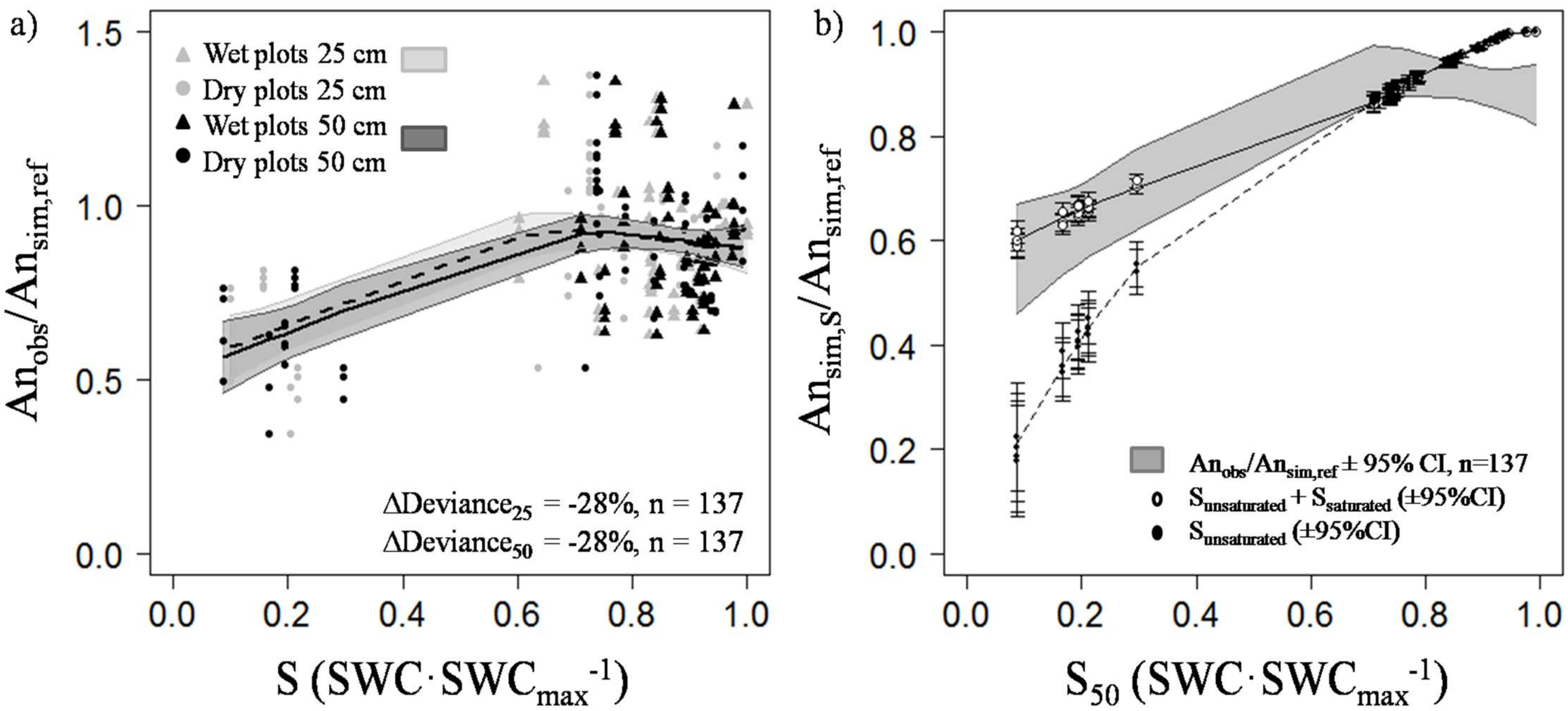
Nadal-Sala et al, (2021)

Bayesian calibration of FvCB model



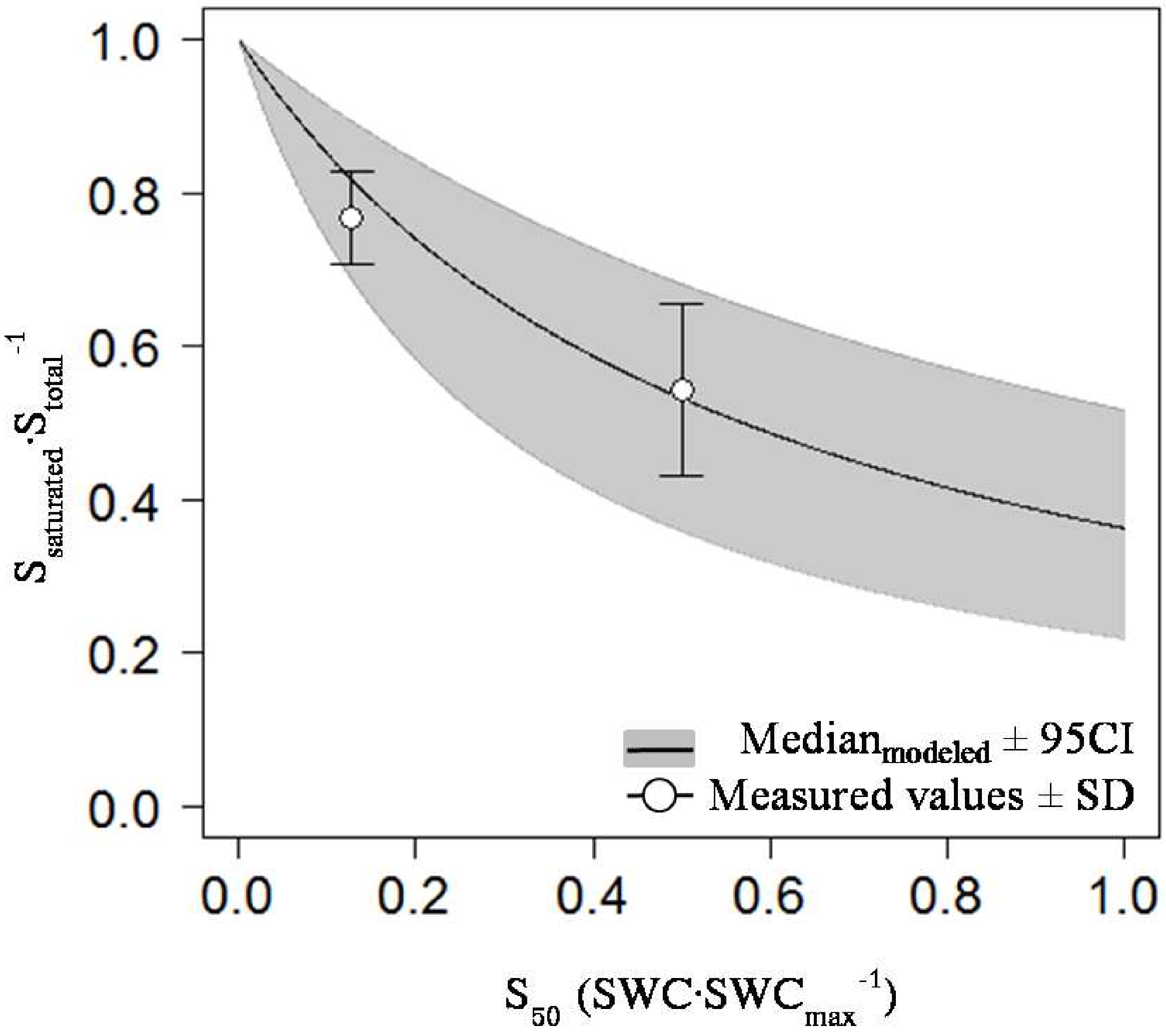
Nadal-Sala et al, (2021)

Bayesian calibration of β model

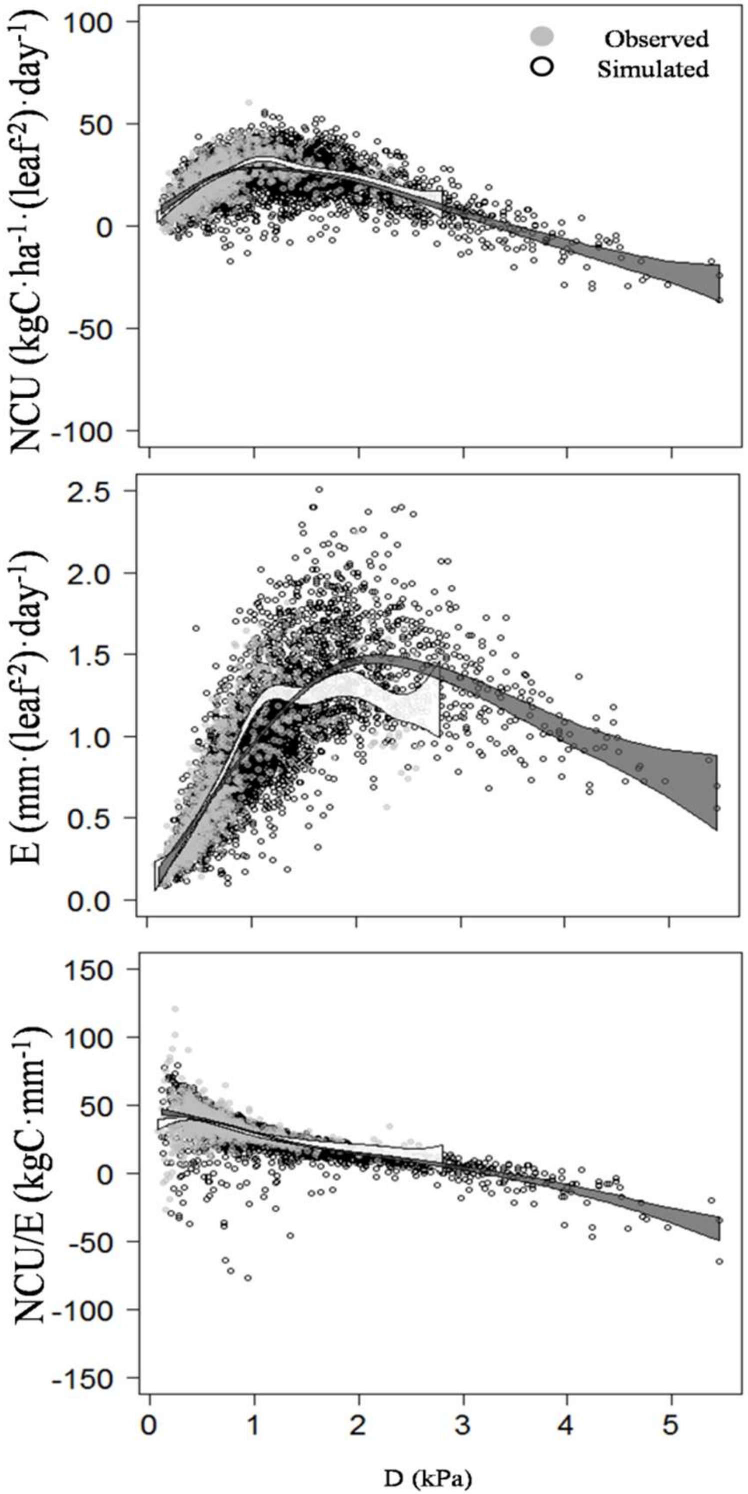


Nadal-Sala et al, (2021)

Comparison with observed water source



Nadal-Sala et al, (2021)

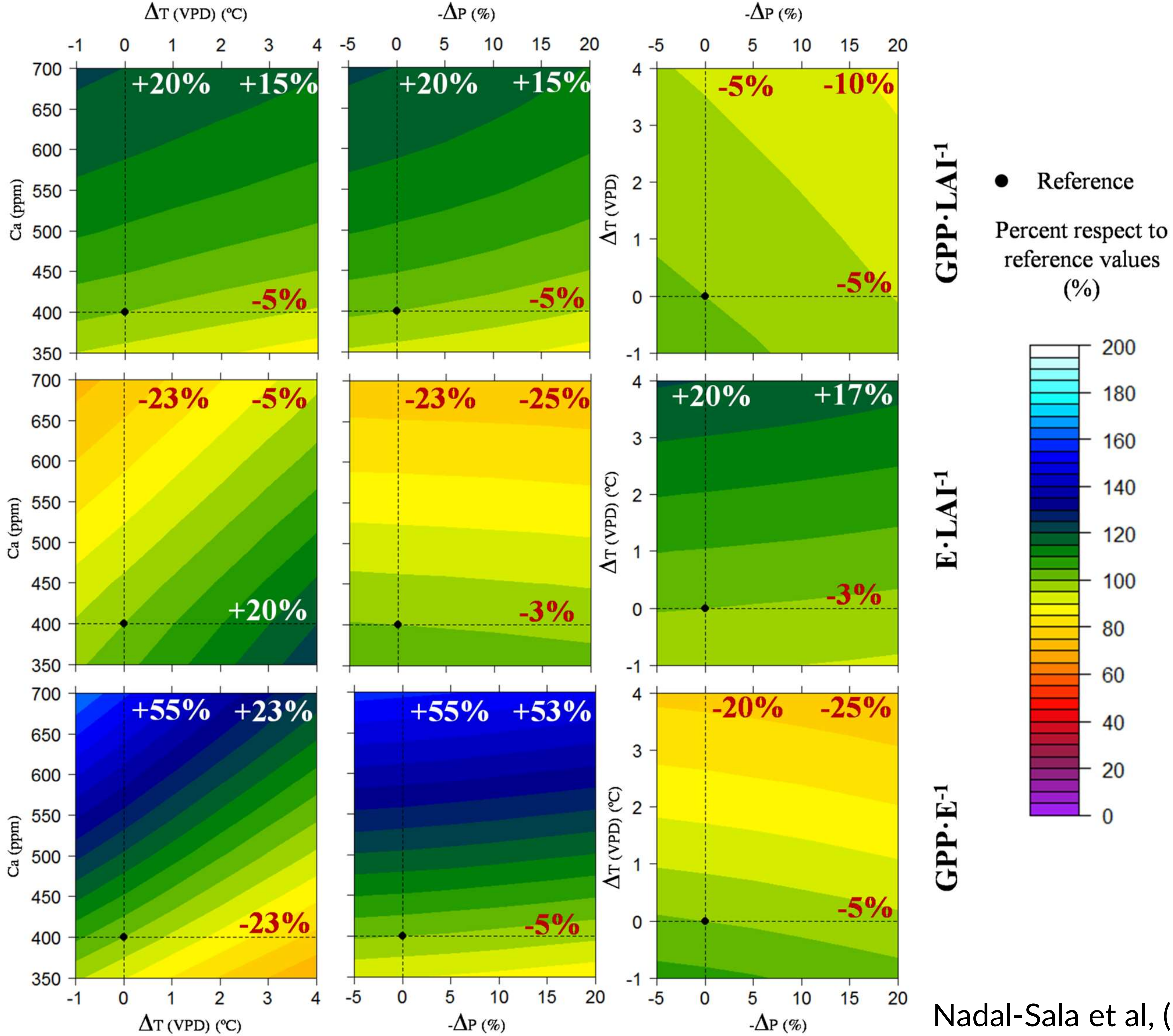


GOTILWA+ captures NCU, E and WUE observed responses to D increases.

↓

Model validated

Nadal-Sala et al, (2021)



Nadal-Sala et al, (2021)