« What’s new, what’s next in dynamic system modeling in agronomy »
September 7 2010
Paris

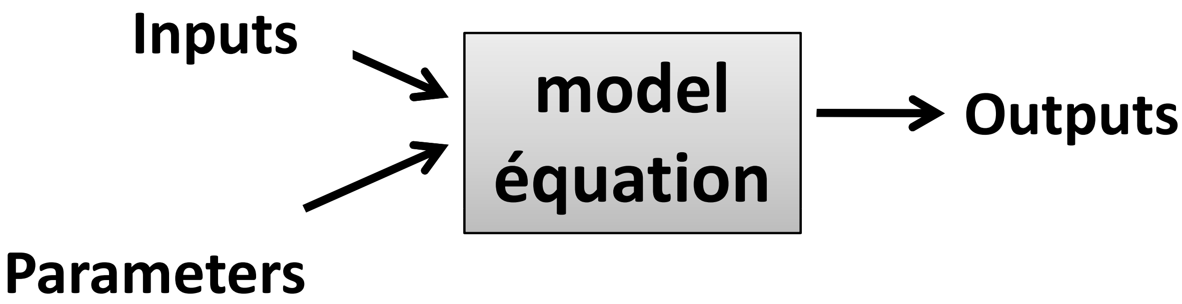
How to synthesize different types of knowledge using models?

David Makowski

INRA

makowski@grignon.inra.fr

Modeling as a tool to synthetize knowledge



Each model component can be defined using different types of knowledge about agricultural systems

Outline

1. **Different types of information about agricultural systems could be included in models**

2. **The Bayesian boom; an opportunity for exploring new avenues in agricultural modelling**

3. **Examples of application of Bayesian methods to combine several types of information**

4. **Perspectives for researchers and extension services**

1. Different types of information about agricultural systems could be included in models

- Experimental data
- Published papers
- Expert opinions
- Public statistics (Governments, FAO etc.)
- Other models

Experimental data

• Measurements of input variables (climatic data, soil characteristics etc.)

• Measurements of output variables (biomass, disease incidence, soil water content, N_20 emission etc.)

• Indirect measurements of input and output variables (reflectance, transmittance, PCR etc.)

HN tester, Yara

Challenge: Complex data structure (correlation, heteroscedasticity, missing data)

Expert opinions

- **Opinions about input and parameter values**

- **Opinions about farmers' practices**

- **Opinions about what model output values should be or should not be**

Challenge : Expert elicitation and quality of expert knowledge (e.g., conflict of interest)

Estimation of potential wheat yield by expert knowledge for two soil types

Individual answers

Wheat potential yield	Shallow soil 1st decile	Shallow soil 9th decile	Deep soil 1st decile	Deep soil 9th decile
Number of answers	*7*		*8*	
Mean ($q.ha^{-1}$)	55	71.4	77	95
STD ($q.ha^{-1}$)	12.9	14.4	10.2	8.5
Minimum ($q.ha^{-1}$)	40	55	60	80
Maximum ($q.ha^{-1}$)	80	100	91	105

Collective answers

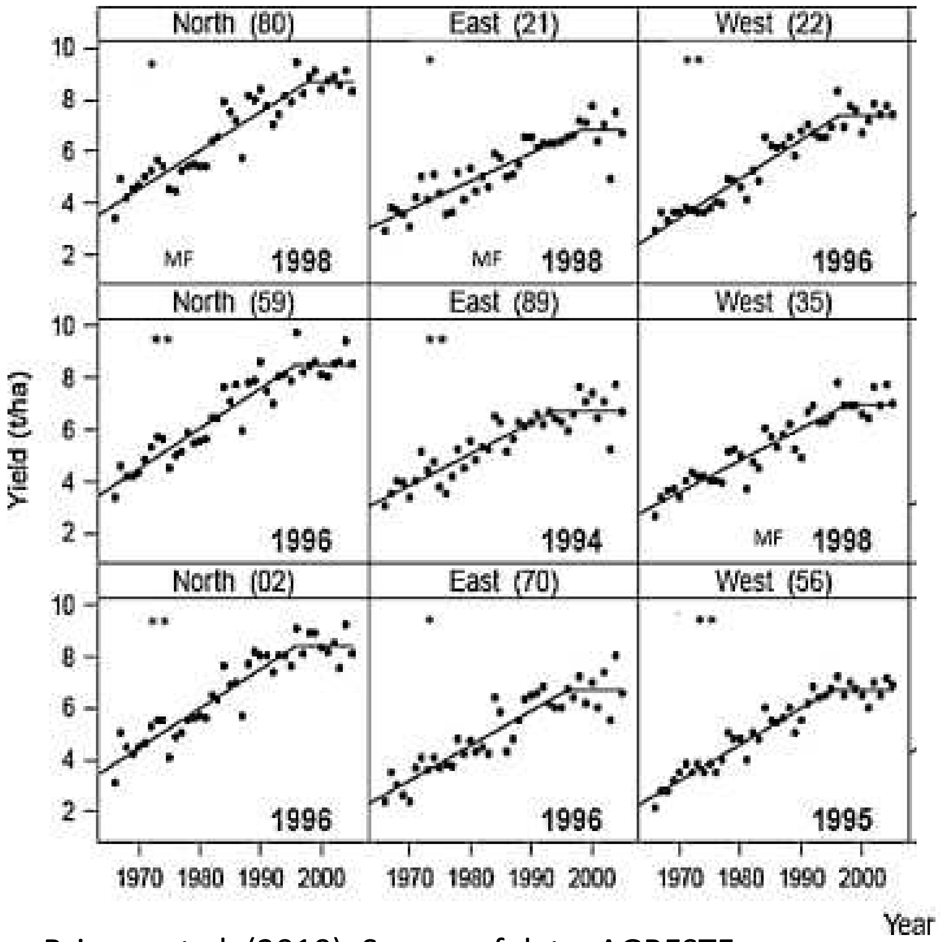
	Shallow soil 1st decile	Shallow soil 9th decile	Deep soil 1st decile	Deep soil 9th decile
	5		*6*	
	53	69	77	96
	5.7	5.5	6.9	5.8
	45	60	65	90
	60	75	85	105

Table 1: Parameter assessment for wheat potential yields in two soil types, a shallow soil (limestone, 40 cm) and a deep soil (loamy soil, 100 cm) (Poitou-Charentes, France). Minimum (maximum) is the min (max) value given among all experts answers. Left table presents the results obtained from the individual step of interviews while results after collective discussion are presented in the right table.

Guichard et al. (2010)

Public statistics

- **Statistics about farmers' practices (e.g., cultivated crops)**
- **Statistics about crop productions (e.g., time series of crop yields in a given region)**



Brisson et al. (2010). Source of data: AGRESTE

Challenges : Scale, data availability

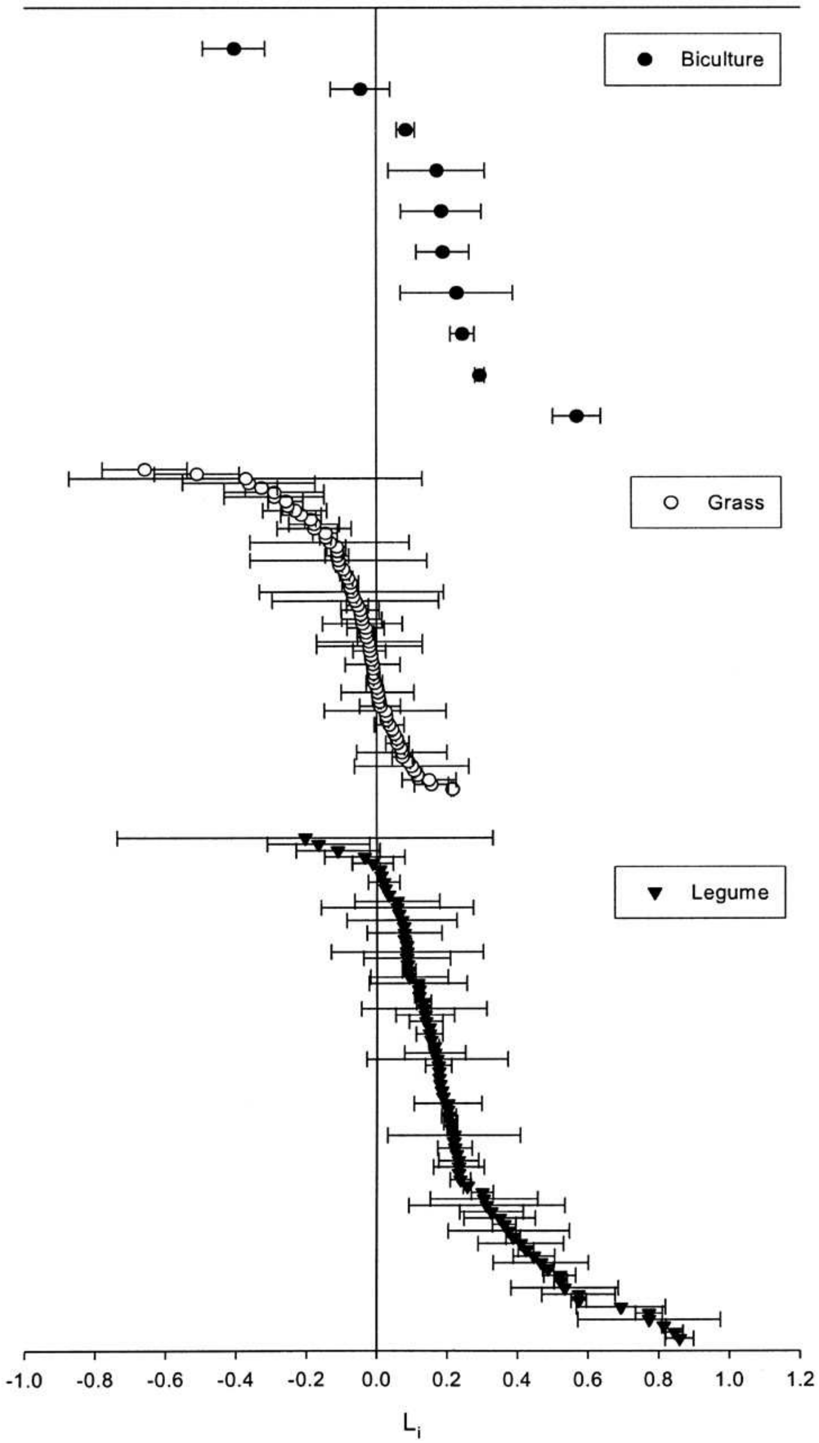
Published papers

Published papers may include:

- **Statistical analysis of experimental data (e.g., mean, standard deviation, correlations)**

- **Expert opinions**

- **They can be used for estimating parameters or assessing model performances**



Challenge: How to synthetize published data

Meta-analysis of 37 published manuscripts

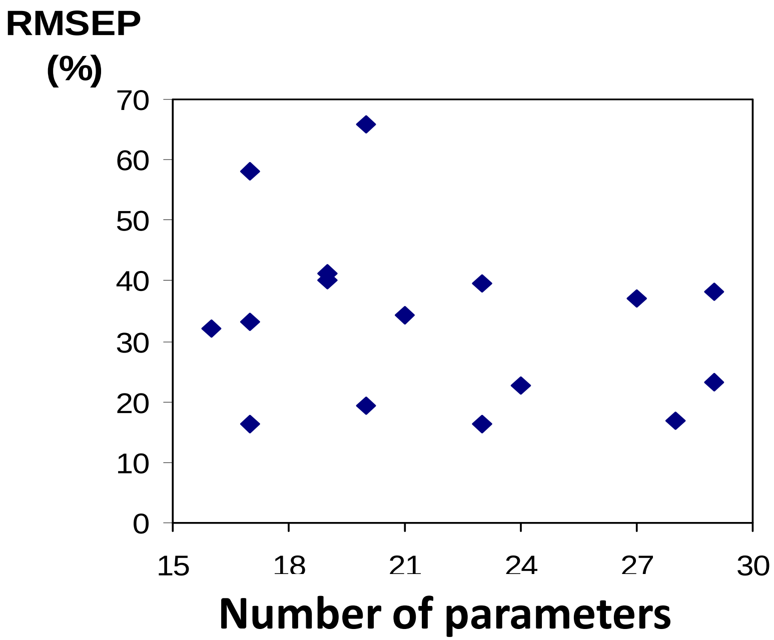
(Miguez & Bollero 2005)

Natural logarithm of the response ratio [ln(yield of corn following winter cover crops/yield of corn following no cover)](Li) for biculture (10 observations), grass (68 observations) and legume (82 observations) winter cover crops. The horizontal bars are the variance.

Several models for a given problem

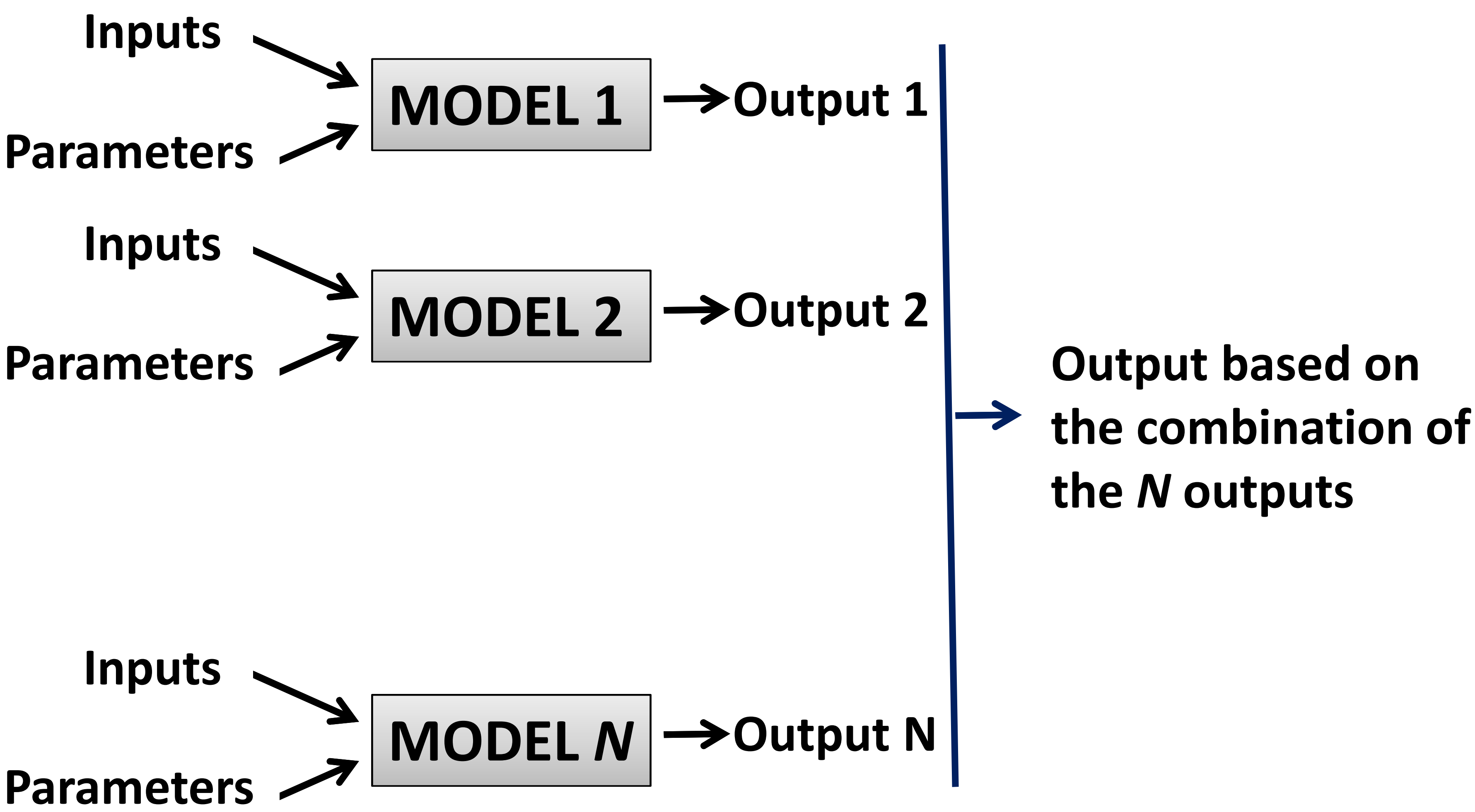
• Many models were developed in agricultural science

• Several models may be available for predicting a given variable (e.g., yield, disease incidence) or for optimizing a given management techniques (e.g., fertilizer dose)



16 models predicting incidence of take-all in wheat (Ennaifar, 2007)

Combining a model with other models



Challenge: model weighting

These types of information could be used to improve model performance

BUT

Only few possibilities have been explored so far in agricultural modelling

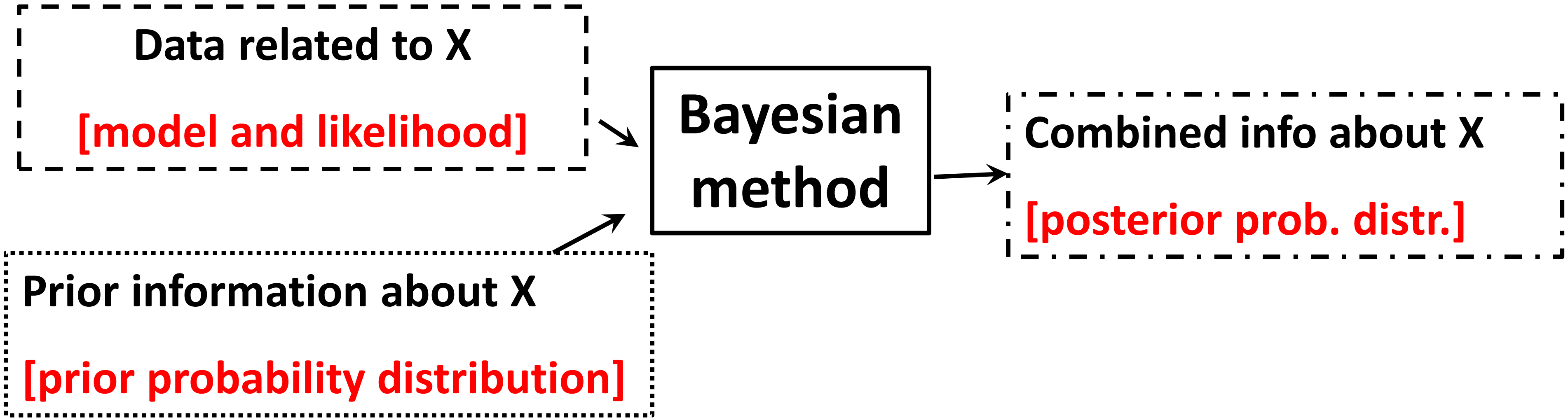
Several methodological challenges:

- **Complex data structure**
- **Expert elicitation and quality of expert knowledge**
- **Scale and availability of public statistics**
- **Meta-analysis of published data**
- **Model weighting**

2. The Bayesian boom

an opportunity for exploring new avenues in agricultural modelling

Combining models, data, and expert opinions using Bayesian methods

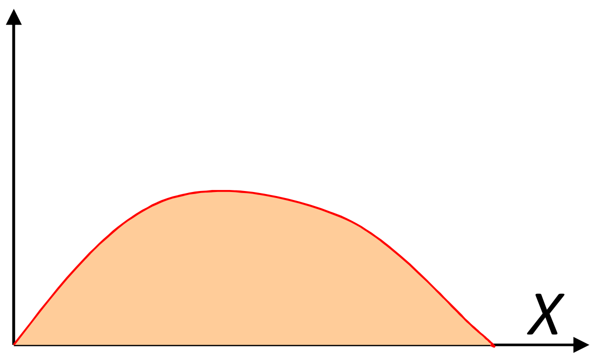


X = input variable
parameter
output variable

Bayesian methods aim at estimating posterior distributions

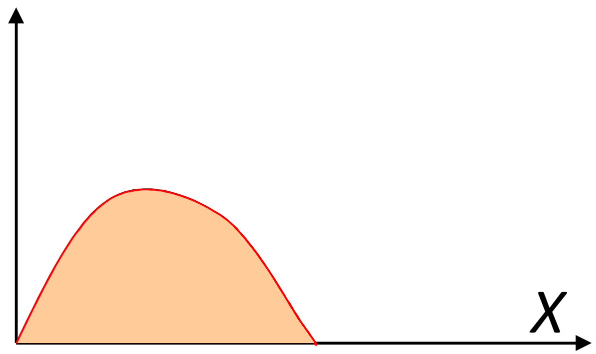
Prior

$P(X)$



Posterior

$P(X \mid data)$



The result is not a point estimate, but a probability distribution

The Bayesian boom

- New methods for estimating posterior probability distributions

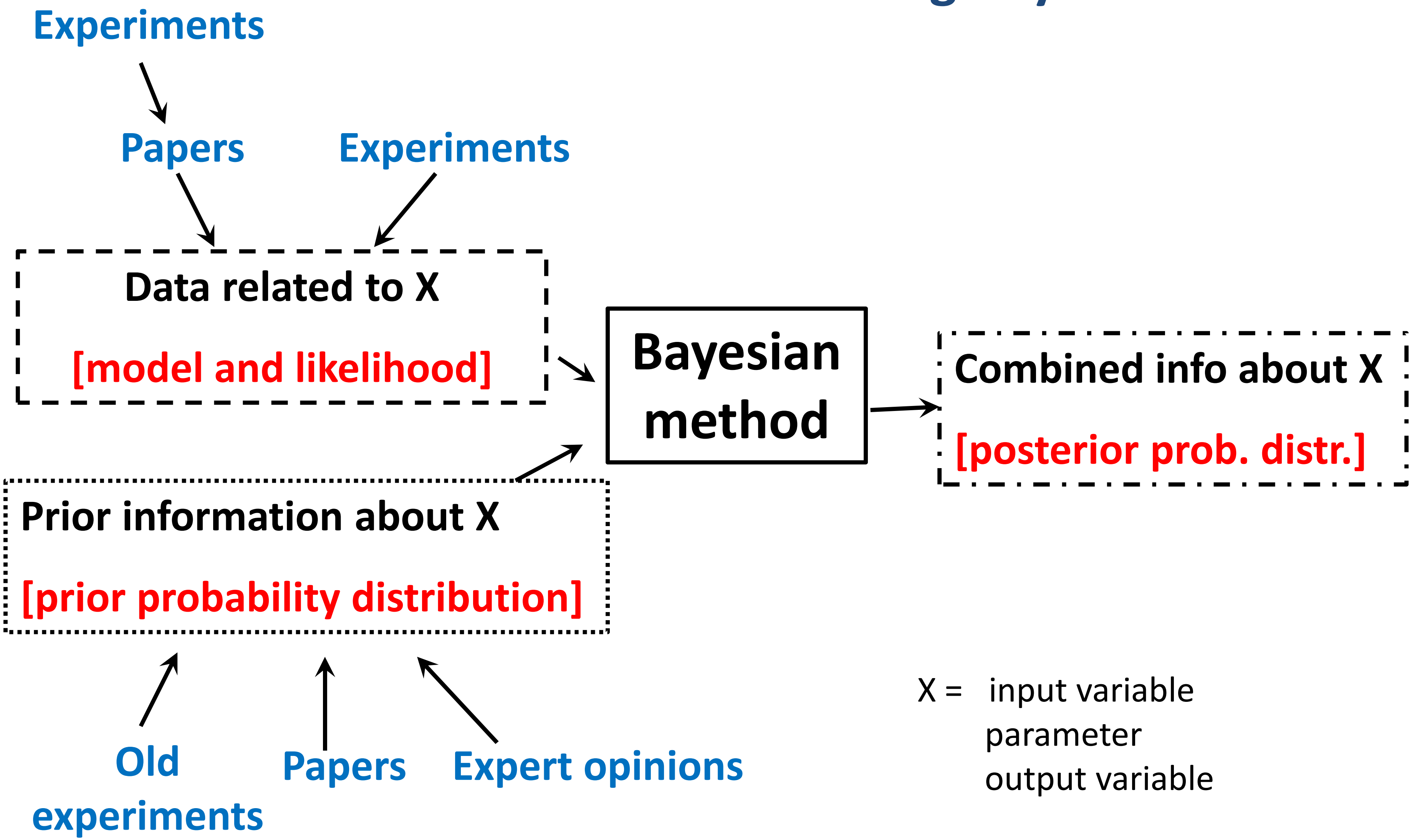
✓ Markov chain Monte Carlo (MCMC)	Late 1990s
✓ Importance sampling	Late 1990s
✓ Approximate Bayesian Computation (ABC)	Early 2000s
✓ Bayesian melding	Early 2000s
✓ Bayesian Model Averaging (BMA)	Early 2000s

- Strong decrease of computational time

- Dedicated softwares (BUGS, R packages etc.)

→ **It is now possible to apply Bayesian techniques to complex problems**

Many sources of information can be combined using Bayesian methods

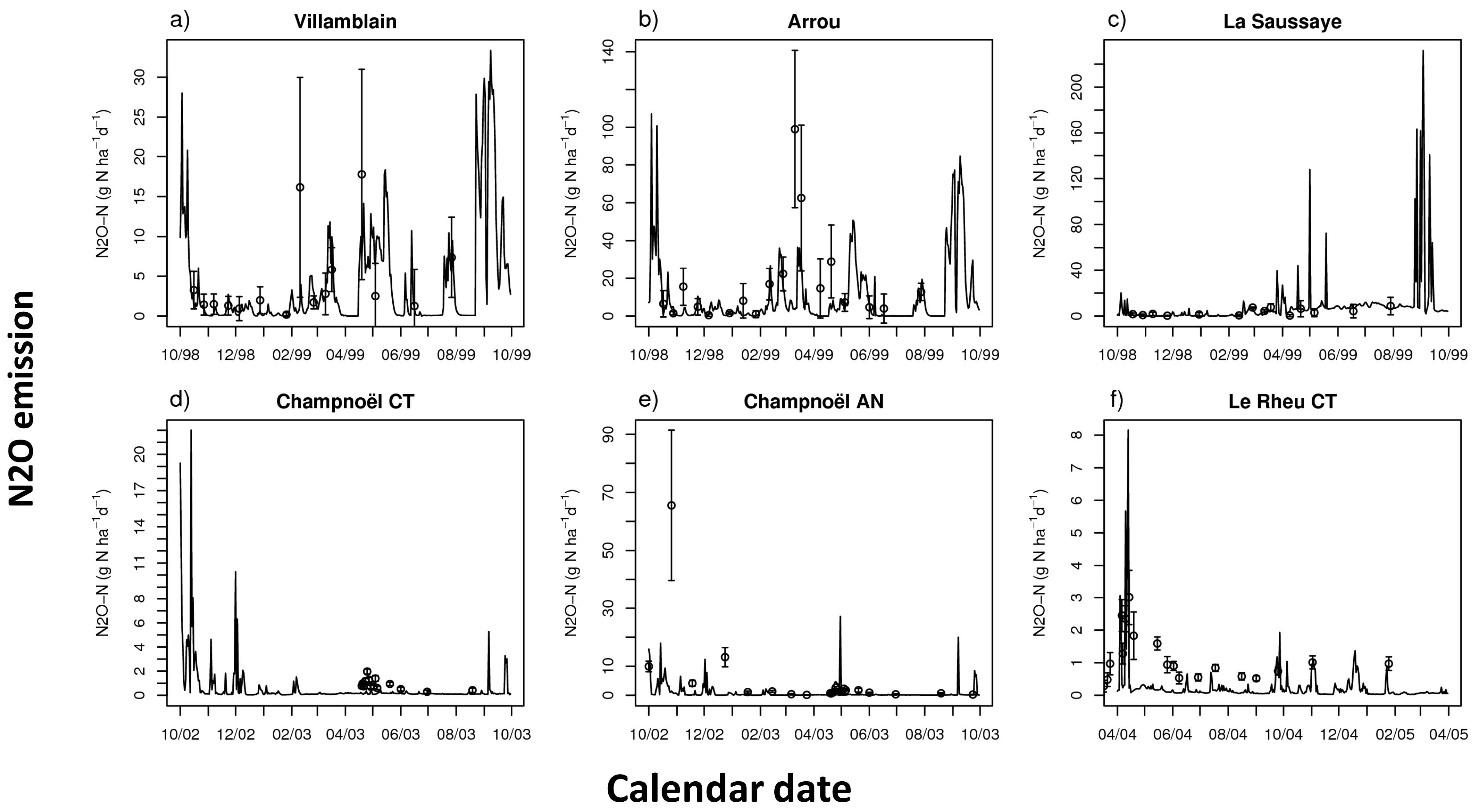


A specific application of Bayesian method is becoming more and more frequent:

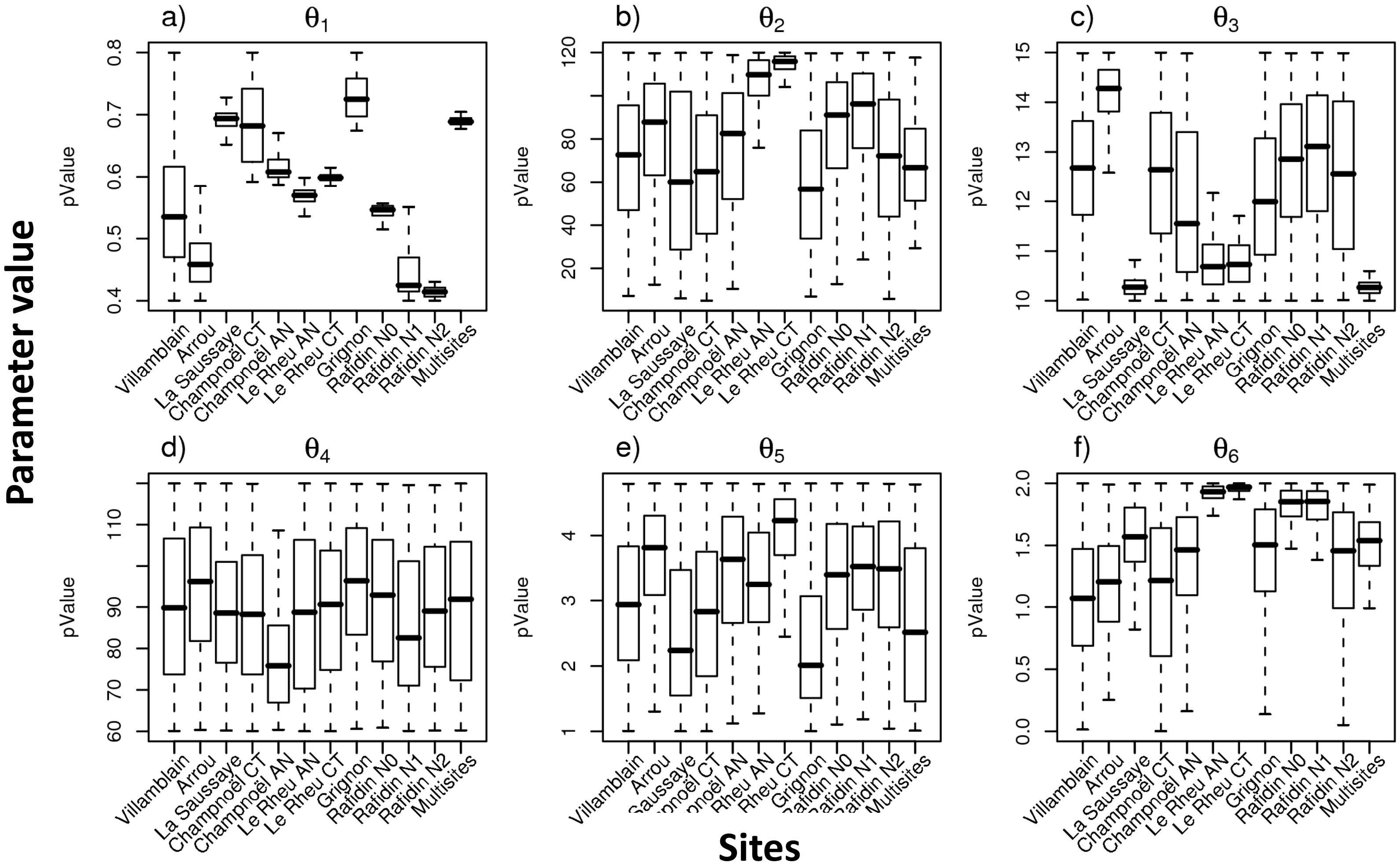
Estimation of model parameters from past measurements of output variables

Types of information	Results
Prior about parameters	Posterior about parameters
Past measurements of model output variables	Posterior about outputs
Likelihood function	

Example : Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model (Lehuger et al. 2009)



Example : Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model (Lehuger et al. 2009)



Bayesian estimation of model parameters becomes more and more frequent

- but, we still have problems to take into account complex data structure (correlation, heteroscedasticity, missing data etc.)

- and we have many other possibilities!

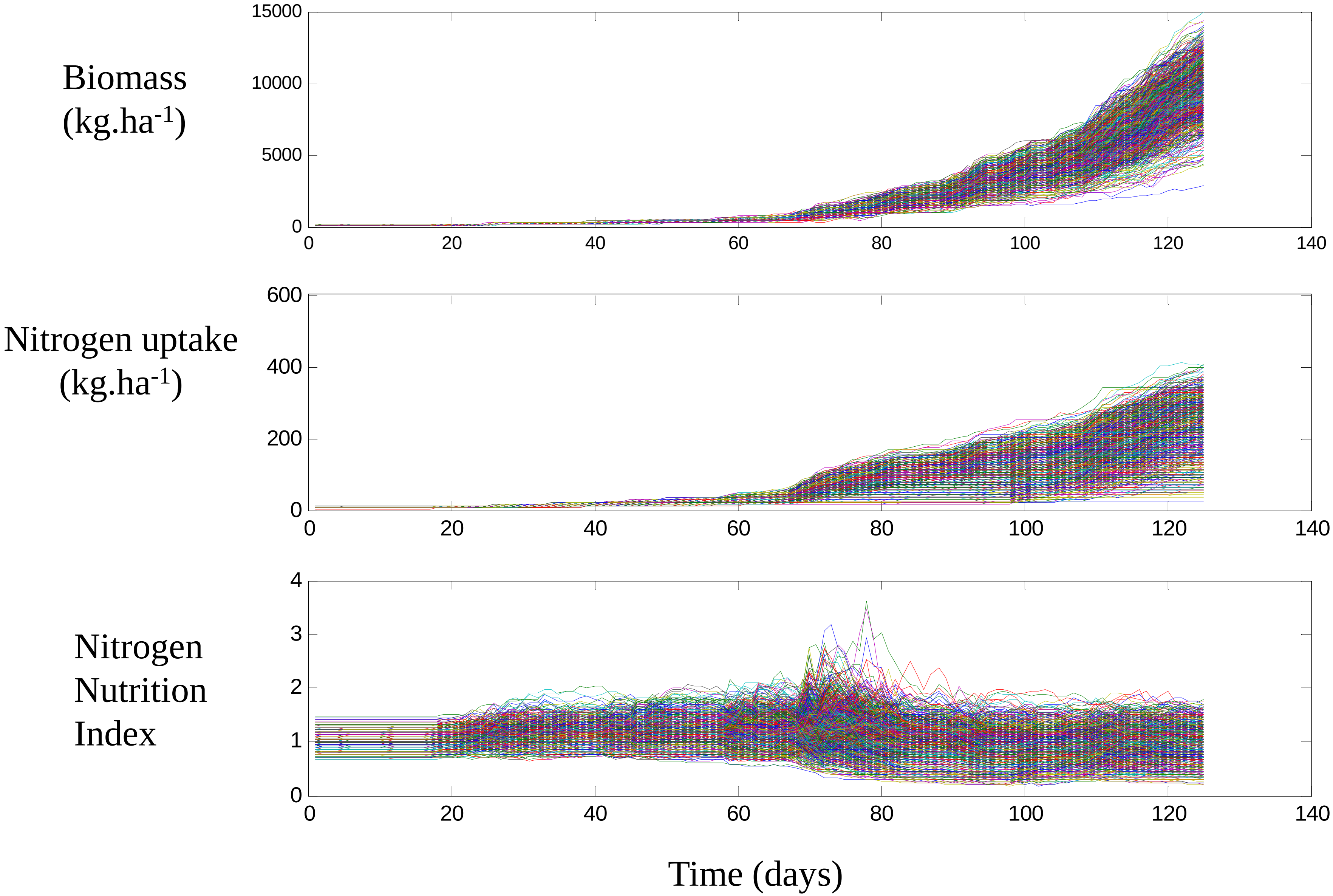
3. Examples of application of recent Bayesian methods to combine different types of information

3.1. Particle filter (Doucet et al., 2001)

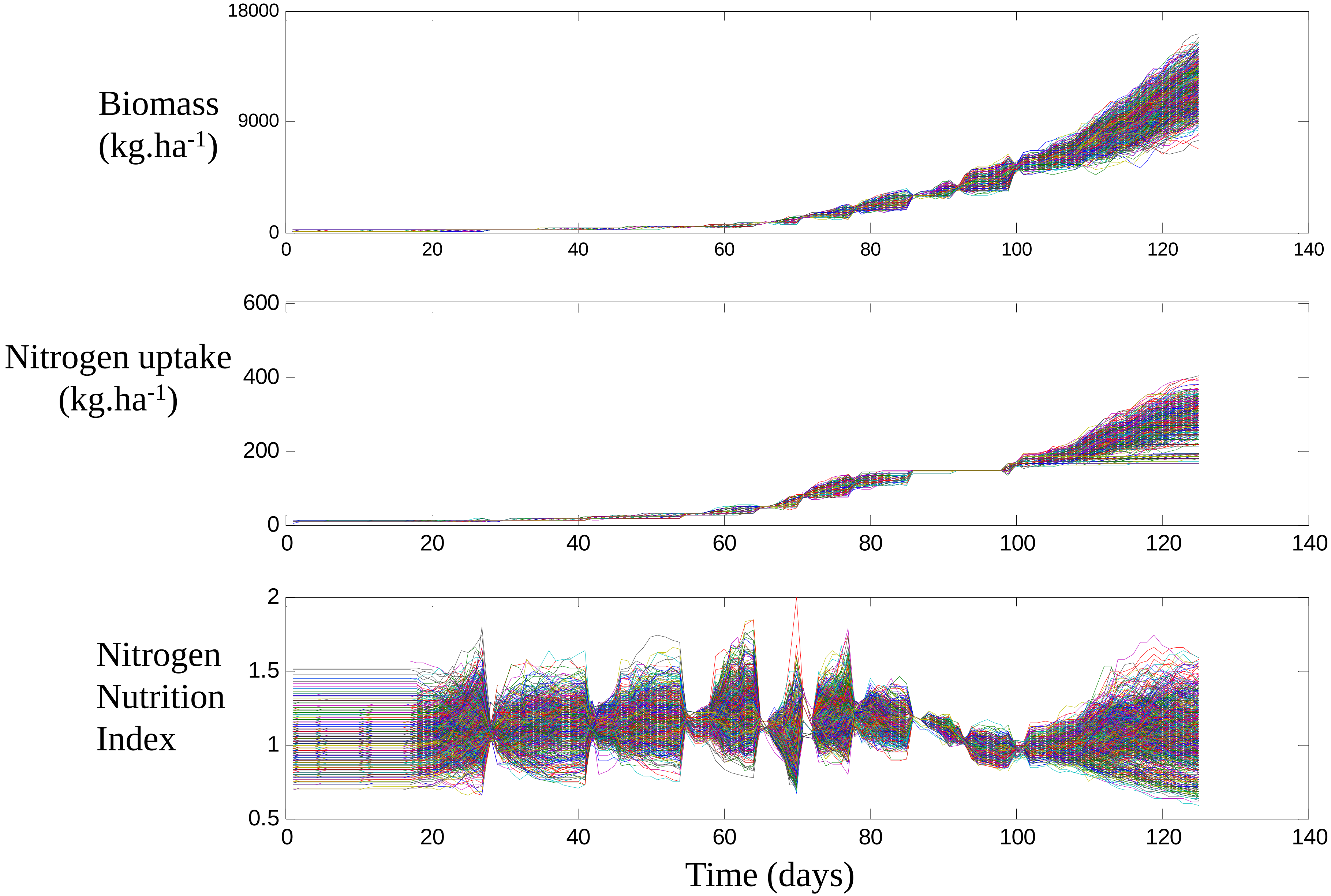
Sequential estimation of posterior distribution by importance sampling

Types of information	Results
Prior about dynamic ouput variables	Posterior about dynamic outputs
Real-time measurements of dynamic output variables	
Likelihood function	

Simulations without correction for one wheat plot (site-year)



Results for the same wheat plot obtained with a filter using biomass and nitrogen uptake measurements (Naud et al., 2007)

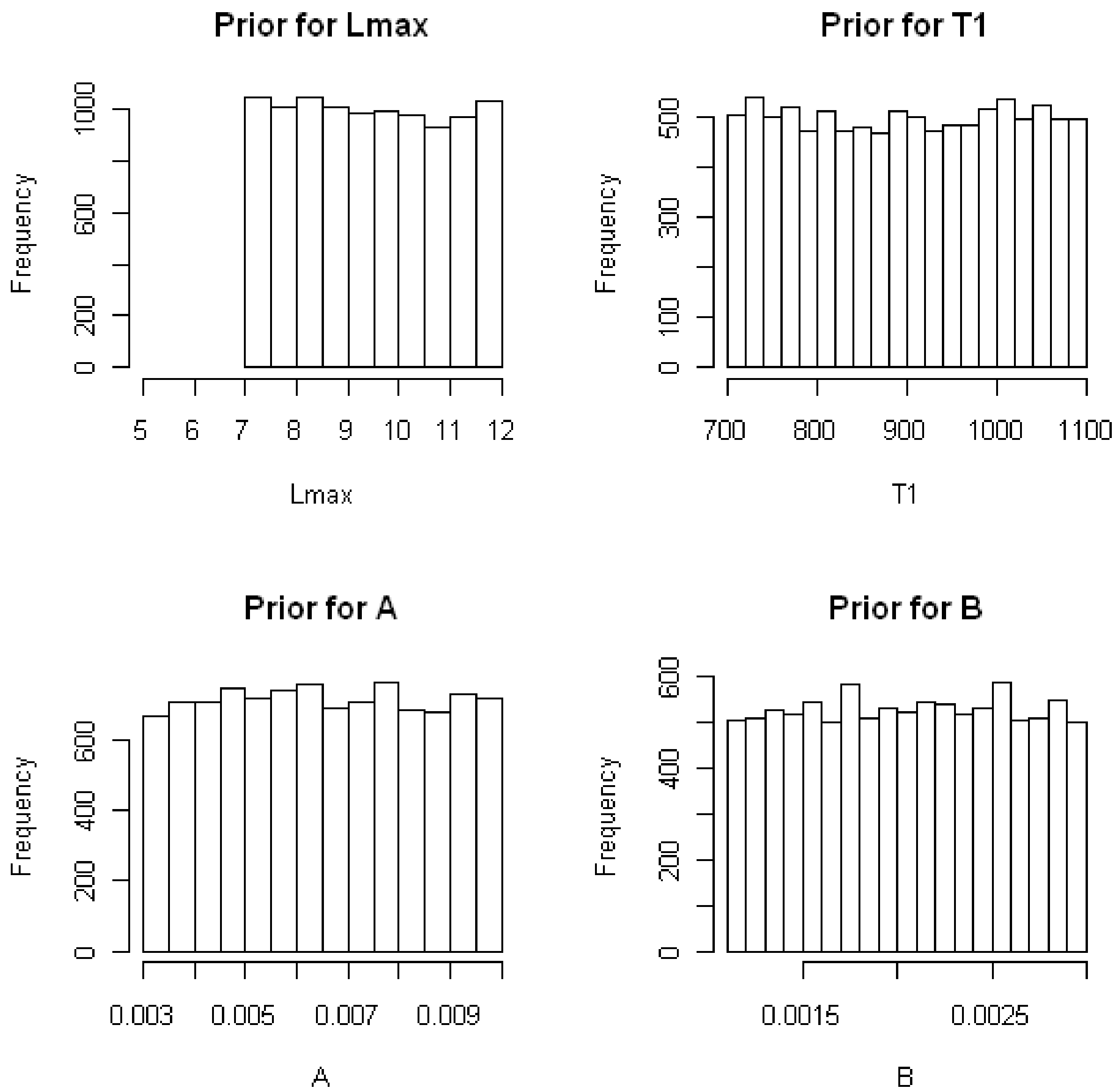


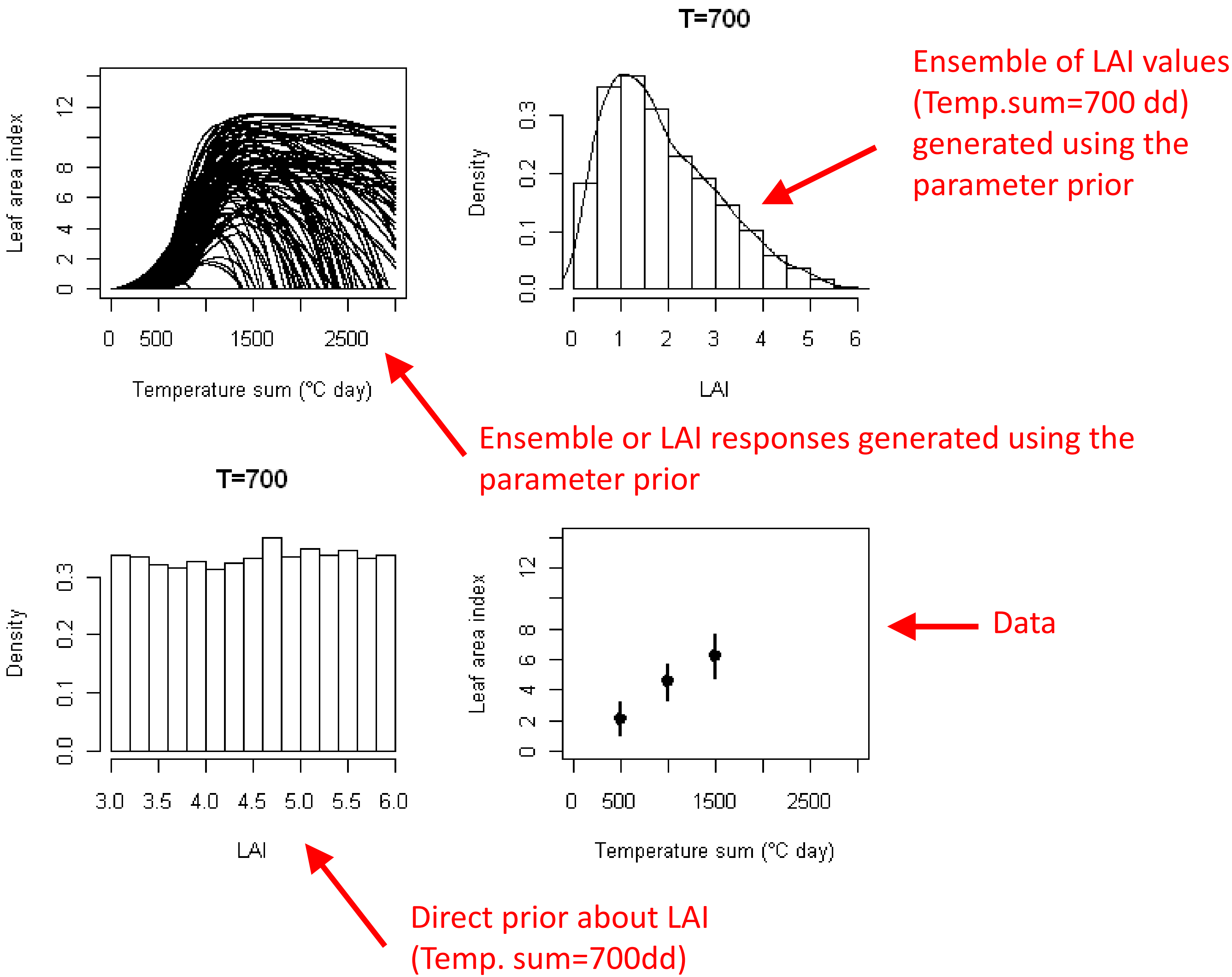
3.2. Bayesian melding (Poole and Raftery, 2000)

Combining model prediction, data, and different types of expert knowledge

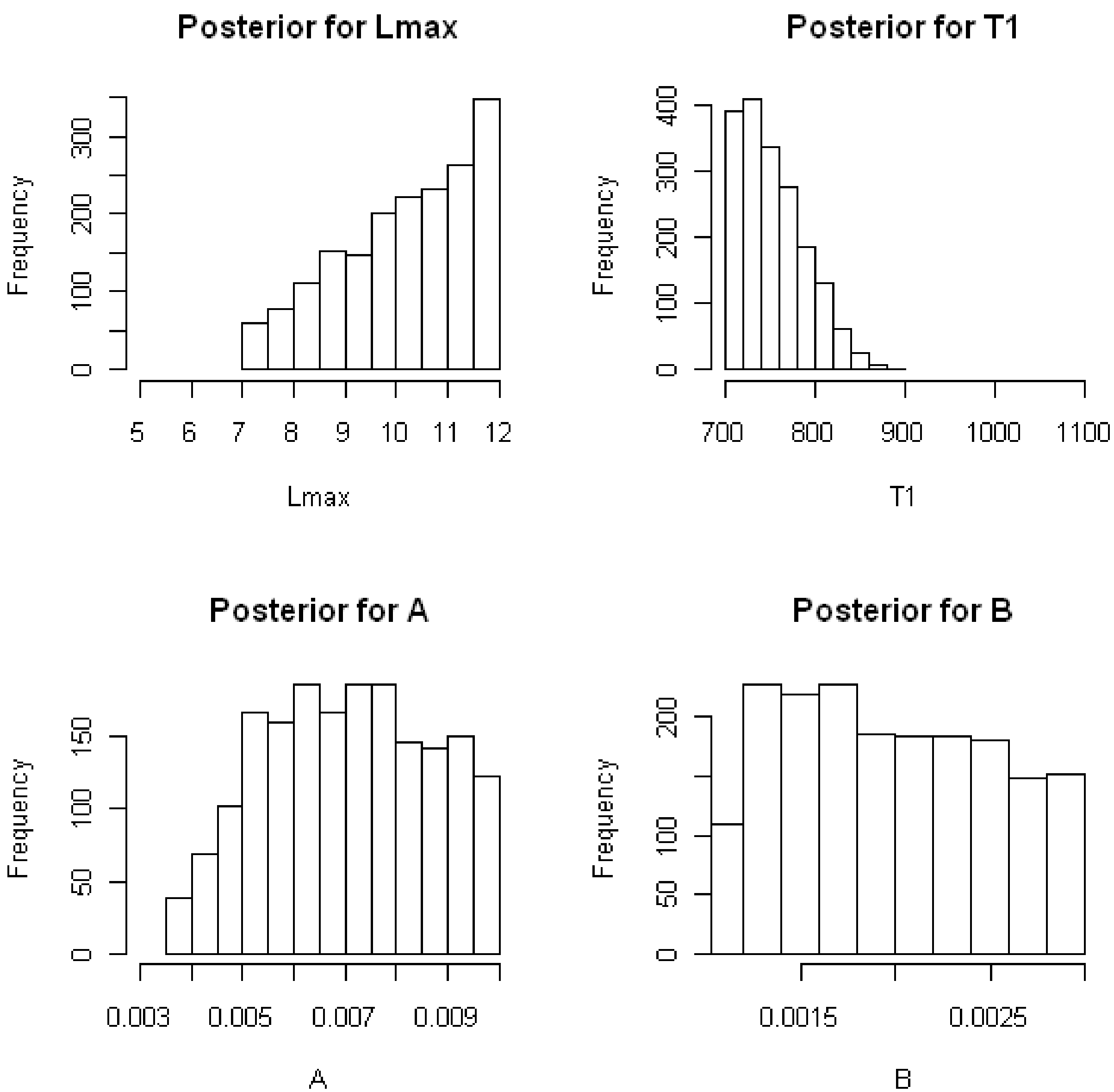
Types of information	Results
Prior about parameters	Posterior about parameters
Prior about model outputs	Posterior about model outputs
Output measurements	
Likelihood function	

Estimation of the parameters of a leaf area index (LAI) model by Bayesian melding





Leaf area index
0 2 4 6 8 12
0 500 1500 2500
Temperature sum (°C day)
T=700
Density
0.0 0.1 0.2 0.3
0 1 2 3 4 5 6
LAI
Ensemble of LAI values (Temp.sum=700 dd) generated using the parameter prior
Ensemble or LAI responses generated using the parameter prior
T=700
Density
0.0 0.1 0.2 0.3
3.0 3.5 4.0 4.5 5.0 5.5 6.0
LAI
Leaf area index
0 2 4 6 8 12
0 500 1500 2500
Temperature sum (°C day)
Data
Direct prior about LAI (Temp. sum=700dd)



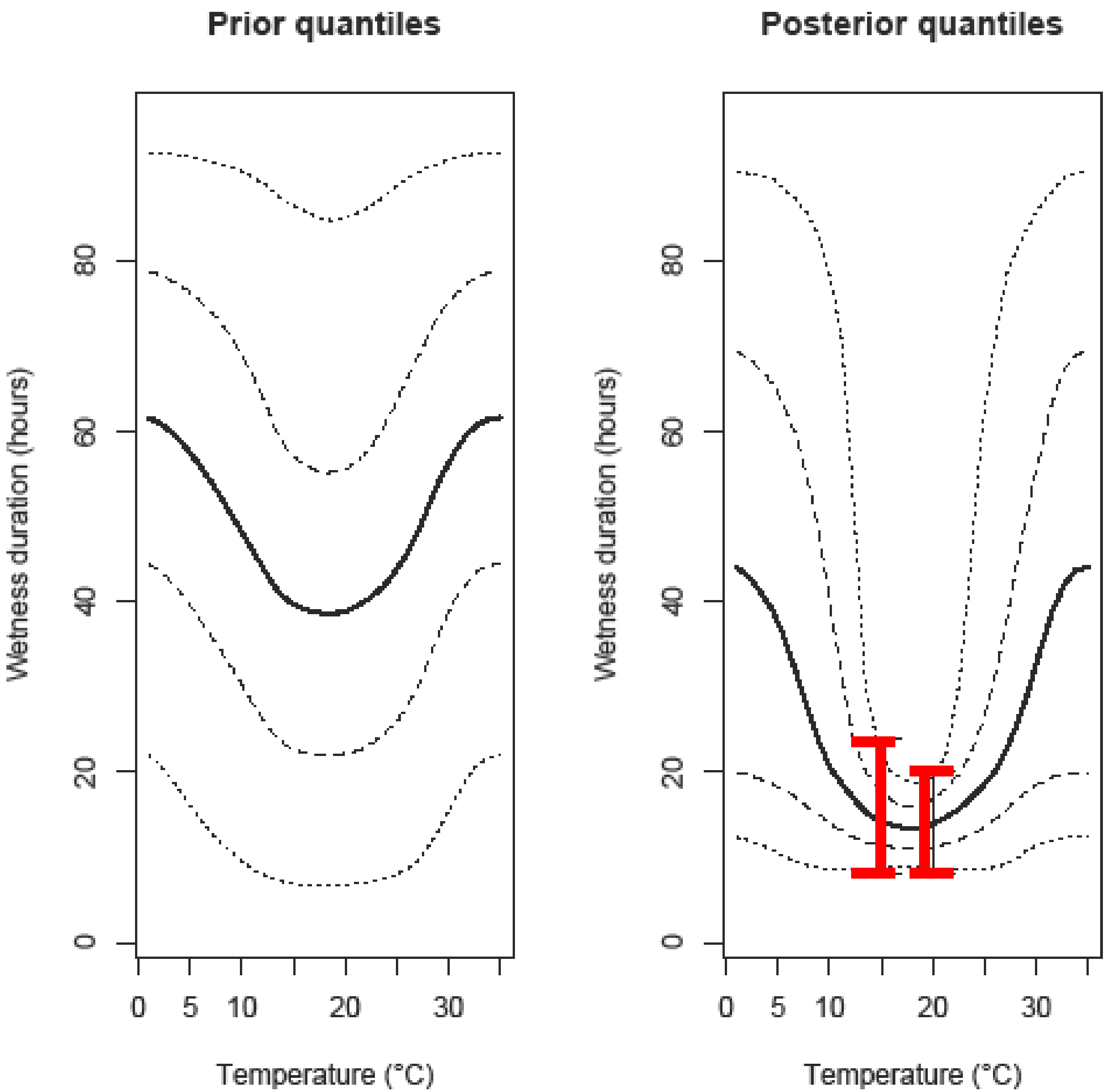
Posterior for Lmax
Frequency
0
100
200
300
5
6
7
8
9
10
11
12
Lmax
Posterior for T1
Frequency
0
100
200
300
400
700
800
900
1000
1100
T1
Posterior for A
Frequency
0
50
100
150
0.003
0.005
0.007
0.009
A
Posterior for B
Frequency
0
50
100
200
0.0015
0.0025
B

3.3. Approximate Bayesian Computation (ABC)

Computing posterior distribution without likelihood

Types of information	Results
Prior about parameters	Posterior about parameters
Possible ranges of values for the model outputs	Posterior about model outputs

Response of wetness duration requirement for citrus infection by a fungus in function of temperature



3.4. Bayesian Model Averaging (BMA)

Combination of several models for predicting a given variable

Types of information	Results
Prior about models	Posterior about models
Prior about parameters	Posterior about parameters
Measurements of model outputs	Posterior about model outputs
Likelihood function	

Recent applications in crop science with linear and logistic models

(Barbottin et al. 2008; Prost et al., 2008; Casagrande et al. 2009; Barbottin et al., 2010)

$$Final\ prediction = \sum_{k=1}^{K} weight\_k \times prediction\_model\_k$$

Data **Prior**

R packages: BMA, MMIX

4. Perspectives for research institutes and extension services

- **Different types of information can be combined using Bayesian methods**

- **Further applications of Bayesian methods in the future**

  - Real-time predictions (e.g., water budget, epidemiology)

  - Optimization of agricultural practices using ensemble of models (e.g., fertilization and pesticide applications, choice of genotype)

  - Risk analysis (e.g., yield loss, biological invasion)

- **Methodological researchs**

  - Taking into account complex data structure

  - Defining new experimental design

  - Methods for expert elicitation

  - Methods for using public statistics and published data (meta-analysis)

  - Combination of complex models