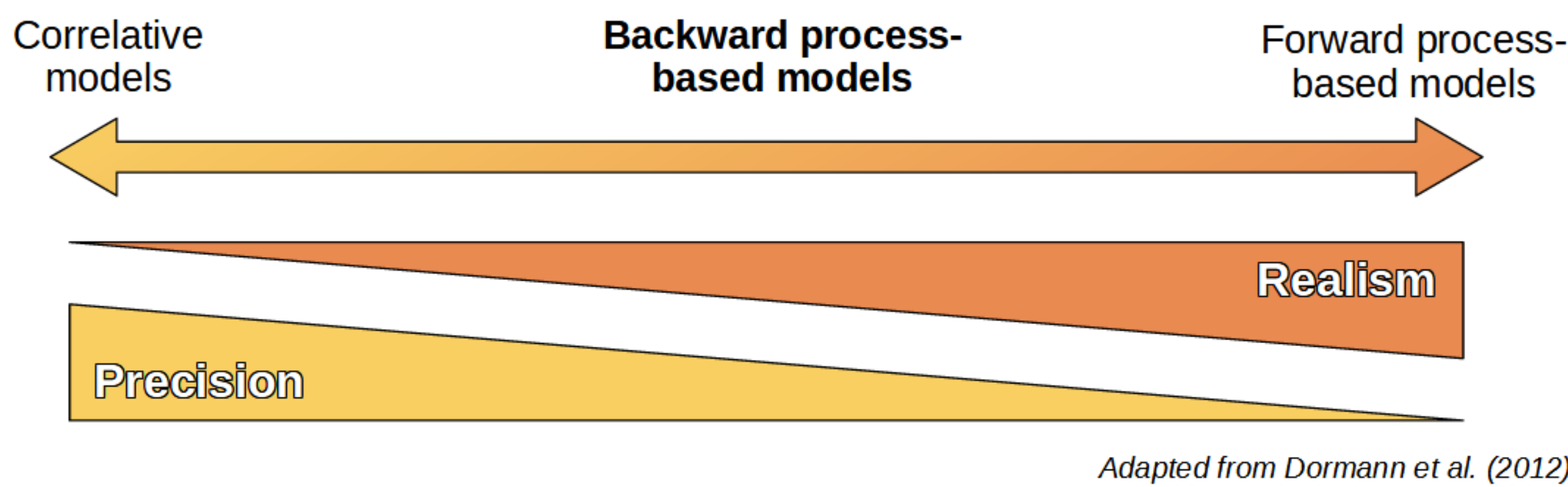# A novel method to estimate process-based model parameters from species distribution data

Victor Van der Meersch, Isabelle Chuine

CEFE, CNRS, Université de Montpellier - `<victor.vandermeersch@cefe.cnrs.fr>`

1. Context and objective

- Two types of species distribution models (SDMs): **correlative niche-based models** and **process-based models**[1]

Correlative models ← **Backward process-based models** → Forward process-based models

Realism / Precision

Adapted from Dormann et al. (2012)

- Process-based models: translate **physiological and ecological processes** into **mathematical equations**. Take more time to develop, **calibration is challenging** and **time-consuming**

Our goal: exploration of the feasibility of **calibrating process-based models using species occurrence data** as correlative SDMs, i.e. **backward calibration**

2. Models and data

Tree occurrence data

- **EU-Forest** dataset
- Completed with records from **GBIF**, filtered with **Atlas Flora Europeae** and **EuroVegMap** species ranges
- For *Quercus ilex*, added records from **WOODIV** database

Climate and soil data

- Climatic variables derived from **ERA5-Land** (1970-2000), 0.1° resolution
- Soil variables extracted from **EU-SoilHydroGrids** and **SoilGrids250m**
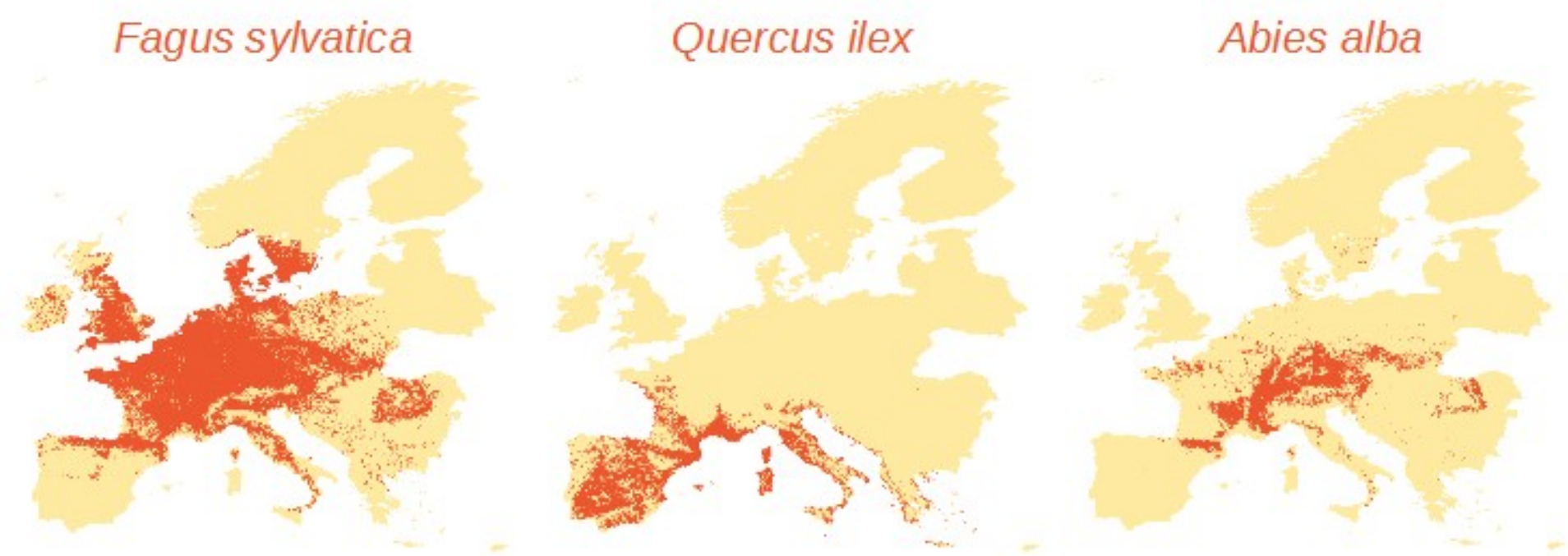

Fagus sylvatica   *Quercus ilex*   *Abies alba*

Fig. 3: Species distribution

Process-based models

- **PHENOFIT**[3]: focuses on phenology and stress resistance, and simulates the fitness (survival and reproductive success) of an average individual (27 to 37 parameters)
- **CASTANEA**[4]: simulates carbon and water fluxes in forest, with several processes such as photosynthesis, stomatal conductance, or respiration (77 parameters)

3. Covariance matrix adaptation, evolution strategy

CMA-ES in a nutshell

- Robust optimization algorithm for **non-linear optimization problems** in continuous domain developed by **Nikolaus Hansen**[2]

- Inspired by **evolutionary biology**: recombination, mutation and selection

In practice

- **Parallel implementation** on two computing clusters (40 to 100 CPUs)

- Calibration runtime for 2000 points and 30 years:
  - PHENOFIT: ~ 24 hours
  - CASTANEA: ~ 20 days

4. Improvement to model calibration

PHENOFIT   CASTANEA

CMA-ES calibration   Forward calibration   CMA-ES calibration   Forward calibration

F. sylvatica: AUC = 0.91   AUC = 0.79   AUC = 0.92   AUC = 0.73

Q. ilex: AUC = 0.96   AUC = 0.86   AUC = 0.95   AUC = 0.7

A. alba: AUC = 0.9   AUC = 0.72   AUC = 0.89   AUC = 0.8

Fig. 4: Species distribution maps obtained with PHENOFIT and CASTANEA forward calibration and backward calibration with CMA-ES. Optimal threshold to dichotomize model outputs is the Youden Index-based cut-off point. Forward calibration is based on expert knowledge, observations and measurements.

5. Non-identifiability of parameters

- Focus on the leaf unfolding submodel in PHENOFIT: **high variability in the parameter estimates** because of **structural correlations**
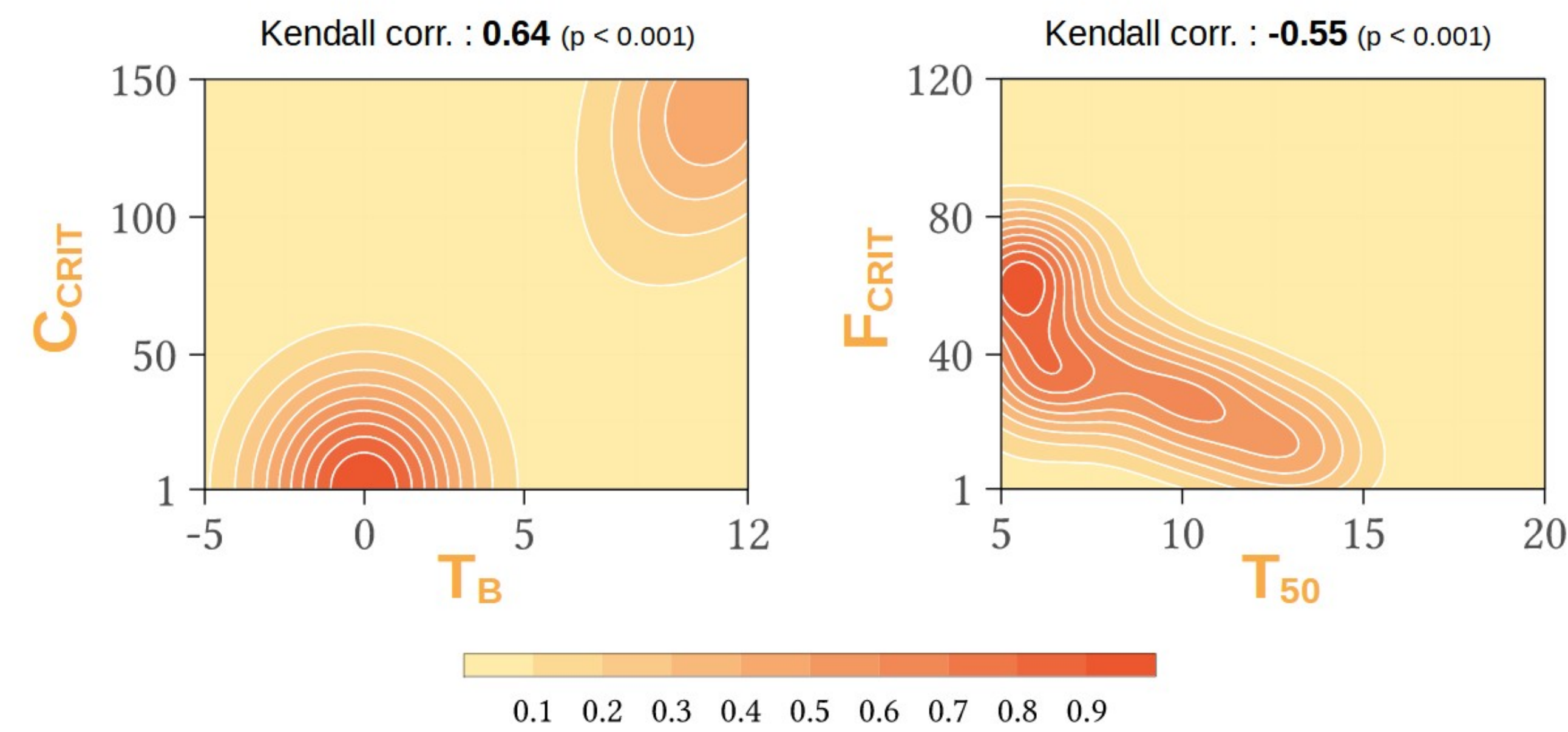

Kendall corr. : **0.64** (p < 0.001)   Kendall corr. : **-0.55** (p < 0.001)

0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9

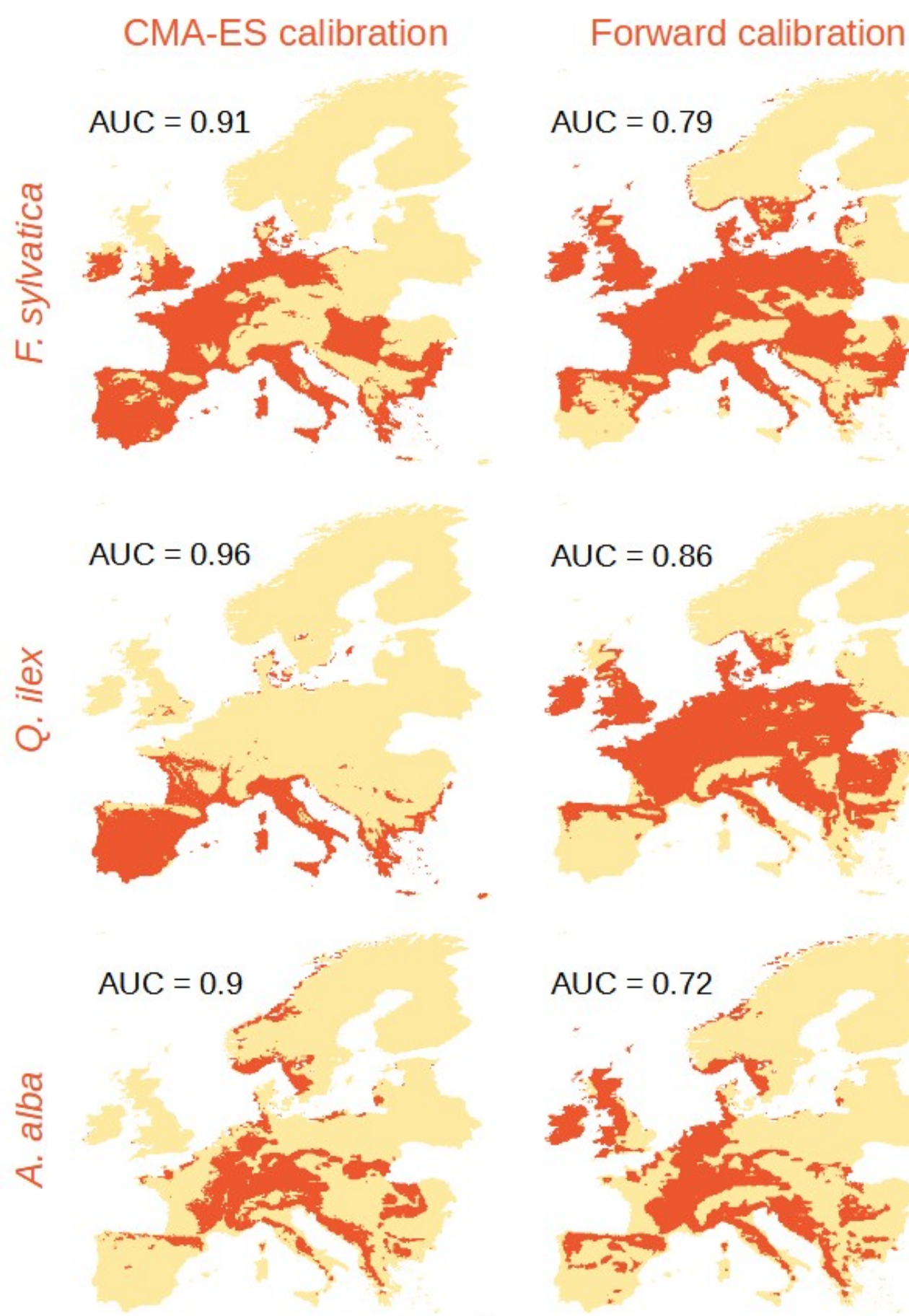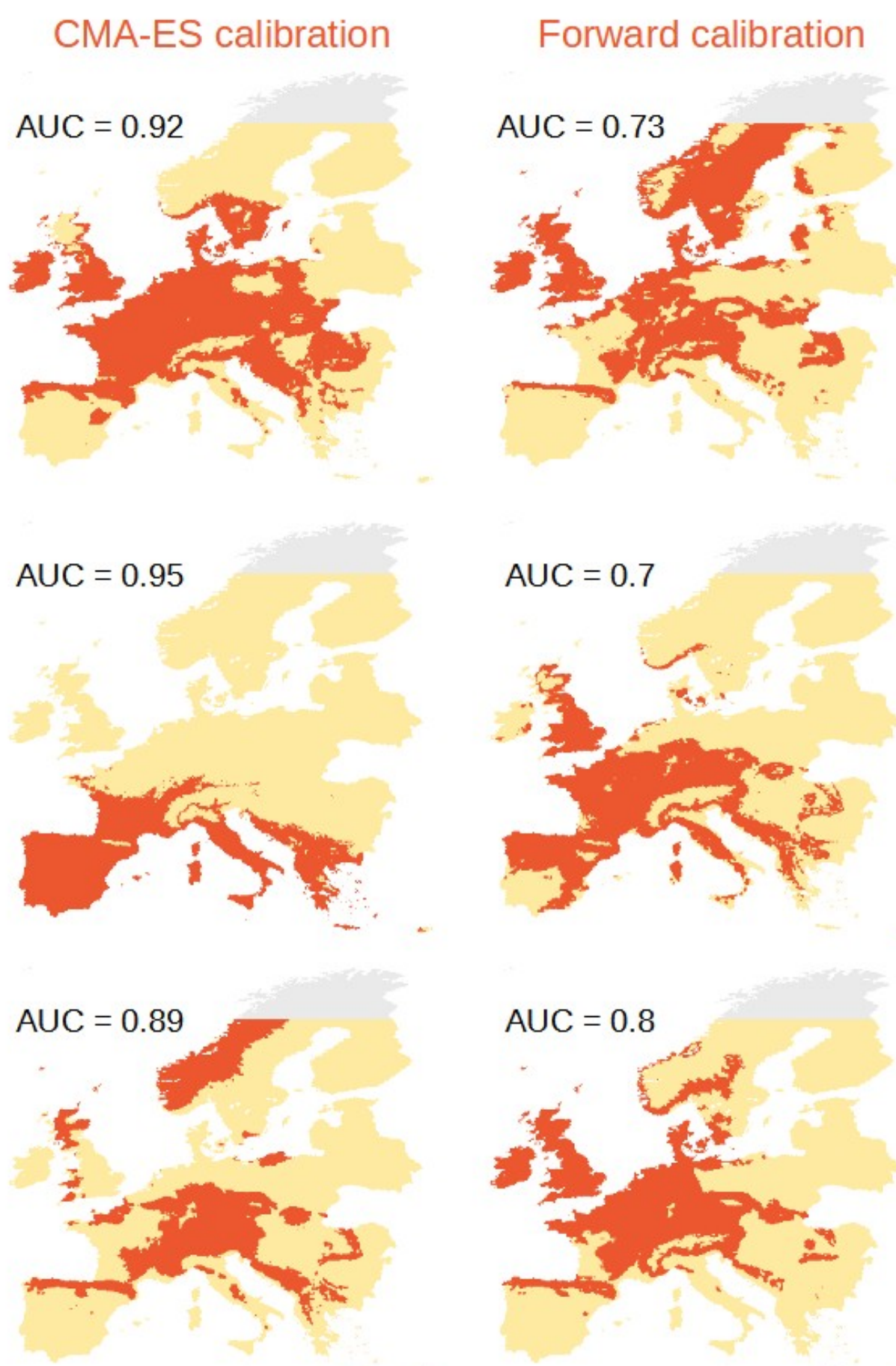*Fig. 2: Density of the main parameters of the leaf unfolding date submodel, after 100 calibrations for F. sylvatica*

- Parameter values **not necessarily ecologically relevant**

6. CMA-ES rocks!

Calibration performance

- **Great increase in AUC** compared to classical (forward) parametrization

- **Easy to use**: does not require an extensive tuning

- Efficient to deal with **ill-conditioned and non-separable problems**

- Can be a **useful technique** to calibrate submodels whose parameter values can hardly be experimentally measured

Comparison with ABC

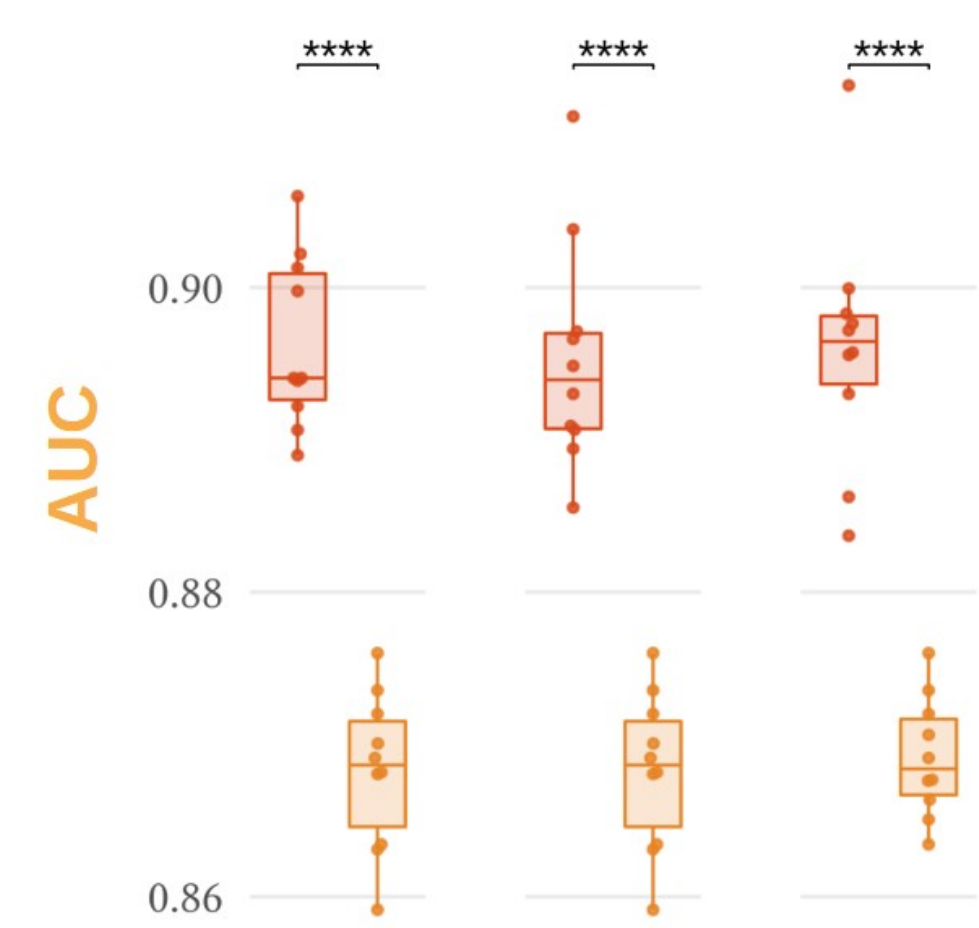CMA-ES slighty **outperforms** approximate bayesian computation (ABC)

Fig. 5: Comparison of CMA-ES (red) and ABC-rejection (orange) performances for F. sylvatica

7. Take-home message

CMA-ES is a **powerful algorithm to calibrate complex ecological models**.

References

1. Dormann, C. F. et al. 2012. *Correlation and process in species distribution models: Bridging a dichotomy.* Journal of Biogeography 39: 2119–2131.
2. Hansen, N. and Ostermeier, A. 2001. *Completely derandomized self-adaptation in evolution strategies.* Evolutionary Computation 9: 159–195.
3. Chuine, I. and Beaubien, E. G. 2001. *Phenology is a major determinant of tree species range.* Ecology Letters 4: 500–510.
4. Dufrêne, E. et al. 2005. *Modelling carbon and water cycles in a beech forest: Part I: Model description and uncertainty analysis on modelled NEE.* - Ecological Modelling 185: 407–436.

This poster was entirely created using `RMarkdown` and `drposter` R package.