

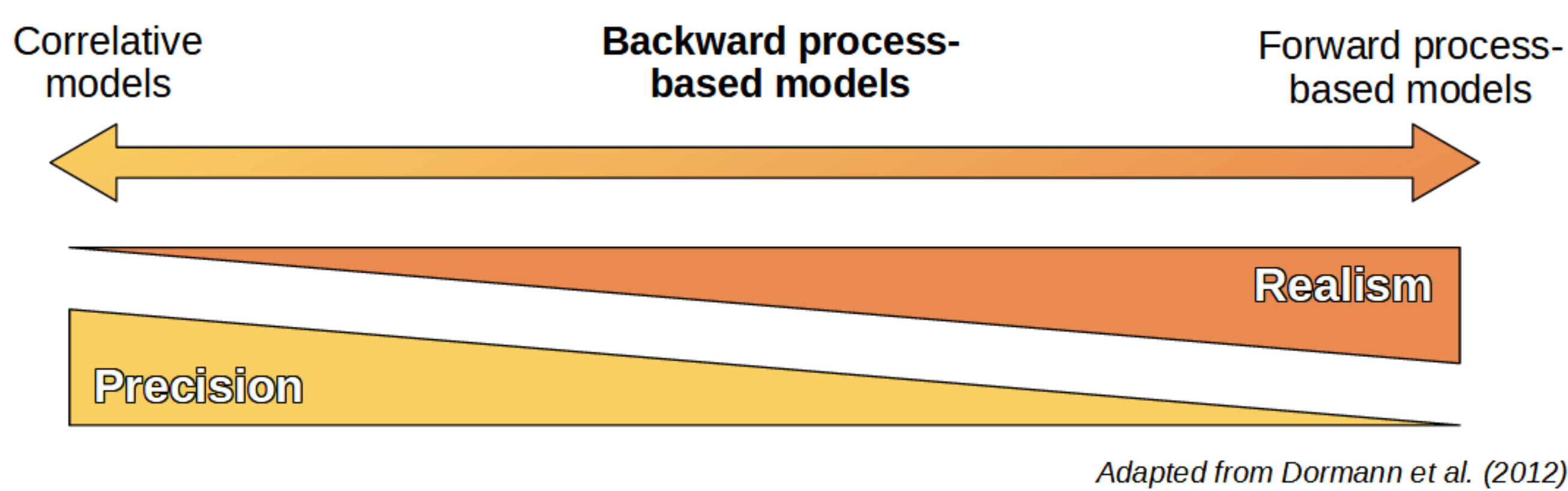
A novel method to estimate process-based model parameters from species distribution data

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



- 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**

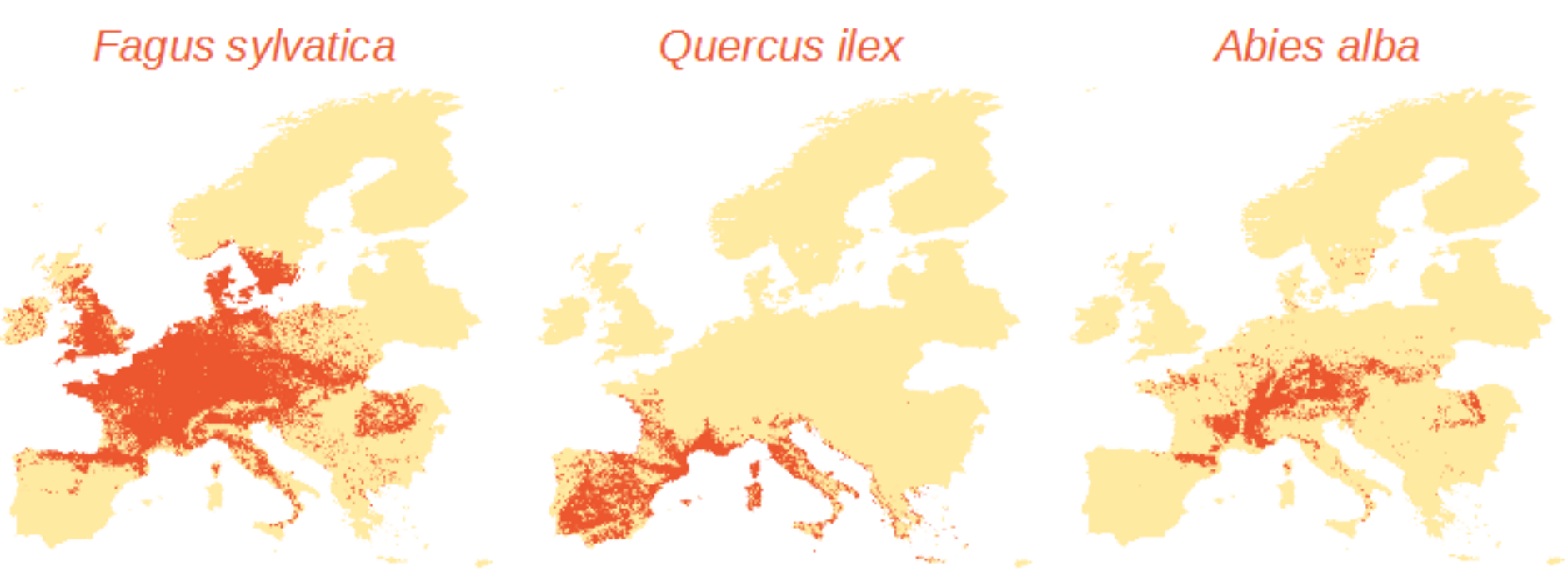


Fig. 3: Species distribution

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

5. Non-identifiability of parameters

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

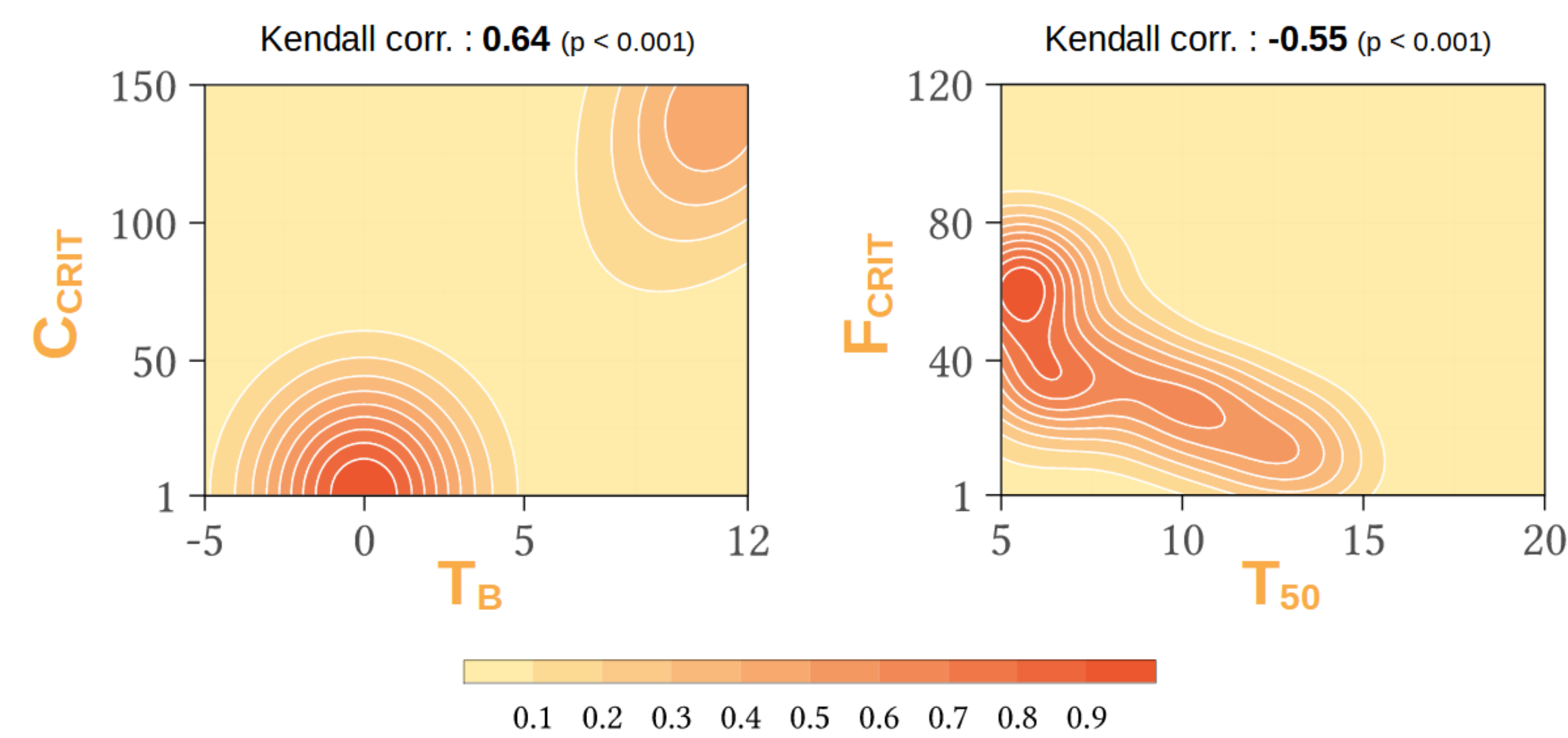


Fig. 2: Density of the main parameters of the leaf unfolding date submodel, after 100 calibrations for *F. sylvatica*

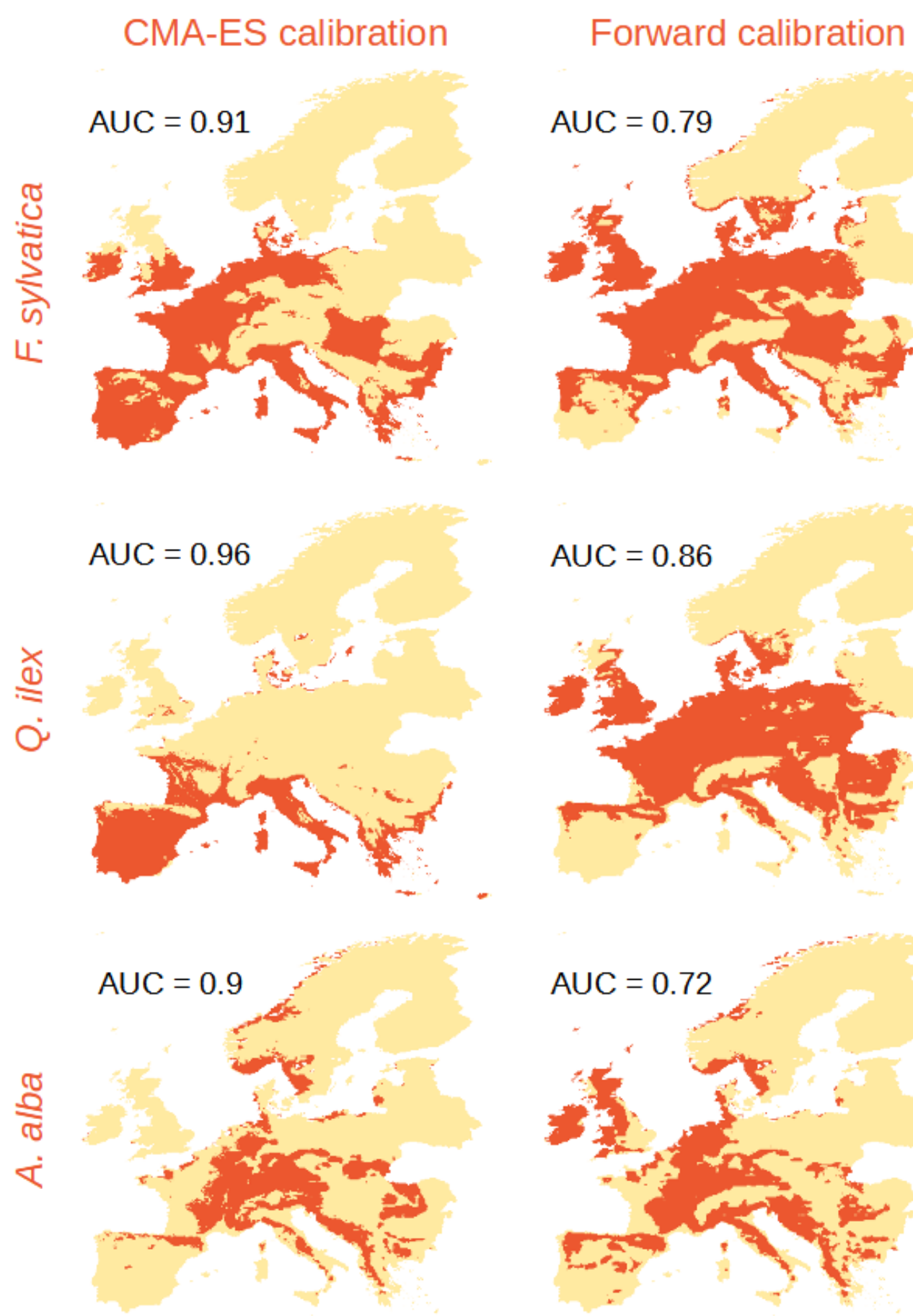
- Parameter values **not necessarily ecologically relevant**

7. Take-home message

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

4. Improvement to model calibration

PHENOFIT



CASTANEA

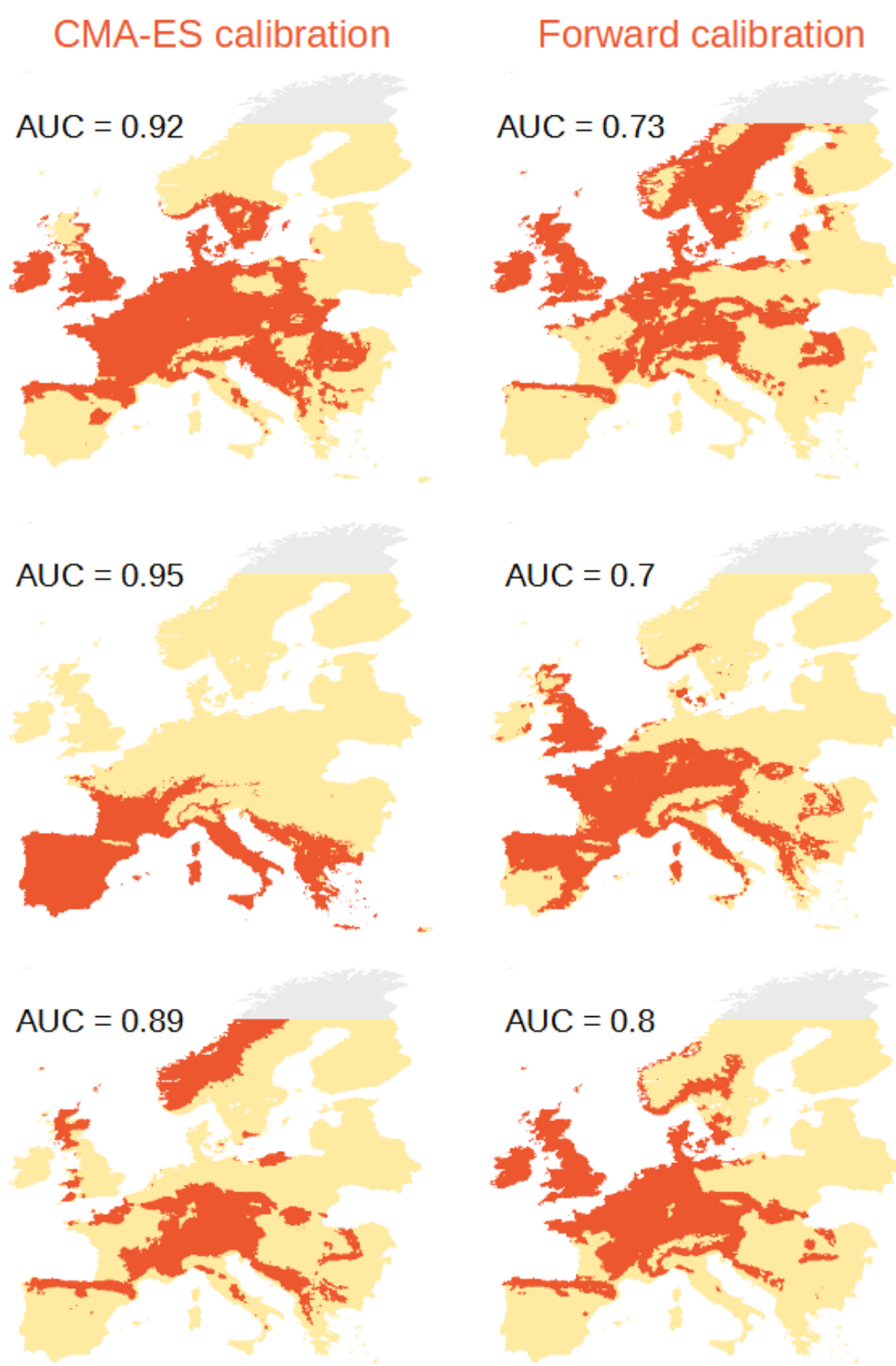


Fig. 4: Species distribution maps obtained with PHENOFIT and CASTANEA forward calibration and backward calibration with CMA-ES. Optimal threshold to dichotomize model outputs is the Youden Index-based cut-off point. Forward calibration is based on expert knowledge, observations and measurements.

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 slightly **outperforms** approximate bayesian computation (ABC)

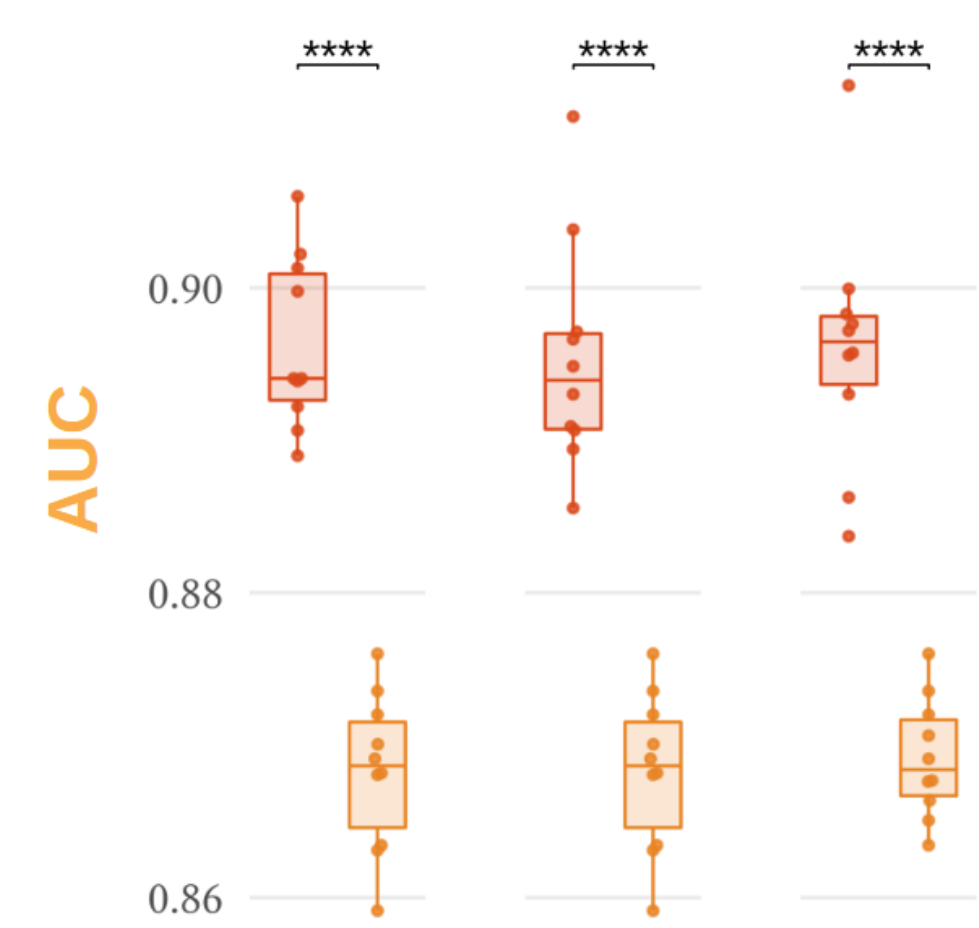


Fig. 5: Comparison of CMA-ES (red) and ABC-rejection (orange) performances for *F. sylvatica*

References

- Dormann, C. F. et al. 2012. *Correlation and process in species distribution models: Bridging a dichotomy*. Journal of Biogeography 39: 2119–2131.
- Hansen, N. and Ostermeier, A. 2001. *Completely derandomized self-adaptation in evolution strategies*. Evolutionary Computation 9: 159–195.
- Chuine, I. and Beaubien, E. G. 2001. *Phenology is a major determinant of tree species range*. Ecology Letters 4: 500–510.
- Dufrène, 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.