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

Victor Van der Meersch, Isabelle Chuine

CEFE, CNRS, Université de Montpellier

1. Introduction

- Two types of species distribution models (SDMs): **correlative niche-based models** and **process-based models**
- Process-based models aim to translate into **mathematical equations** our knowledge about the **physiological and ecological processes**
- They take more time to develop: in particular, **calibration** is **challenging** and **time-consuming**

Correlative models — **Backward process-based models** — Forward process-based models

Ecological knowledge

Species distribution
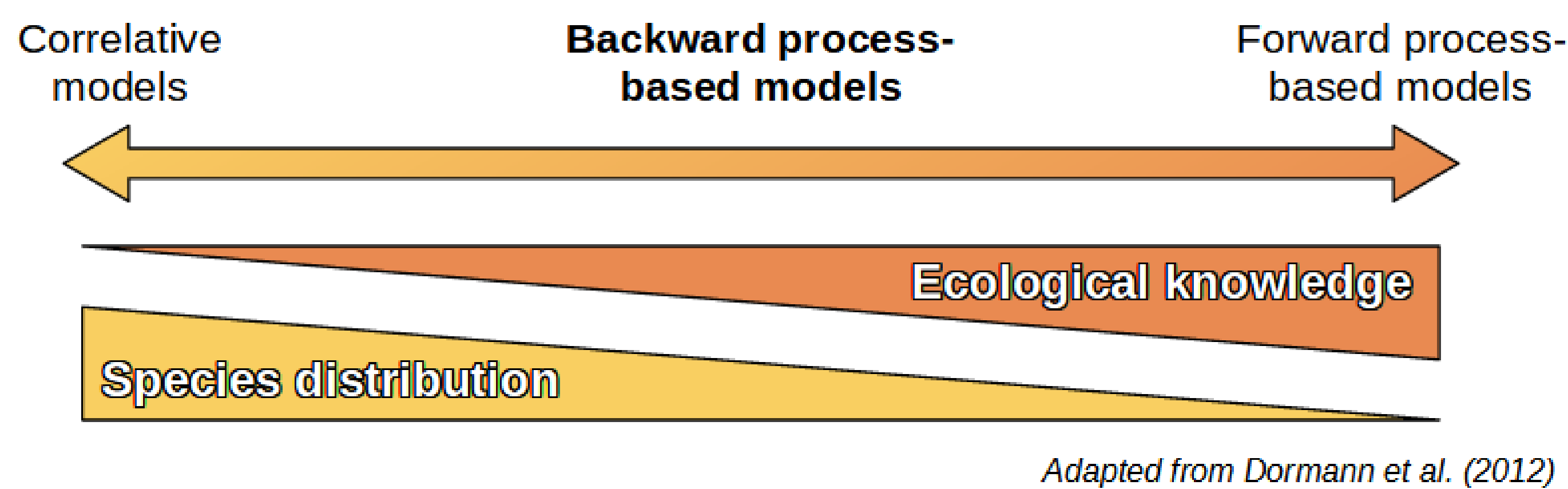
Adapted from Dormann et al. (2012)

- We explored the feasibility and interests of **calibrating process-based models using species occurrence data** as correlative SDMs do (using the geographic distributions of species), i.e. **backward calibration**

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**[1]
- Inspired by **evolutionary biology**: recombination, mutation and selection

In practice

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

- Runtime :
  - PHENOFIT: ~ 24 hours
  - CASTANEA: ~ 20 days

5. Non-identifiability of parameters

- We focus here on the leaf unfolding submodel in PHENOFIT. We found a **high variability in the parameter estimates**

C_{CRIT} T_B F_{CRIT} T_{50}
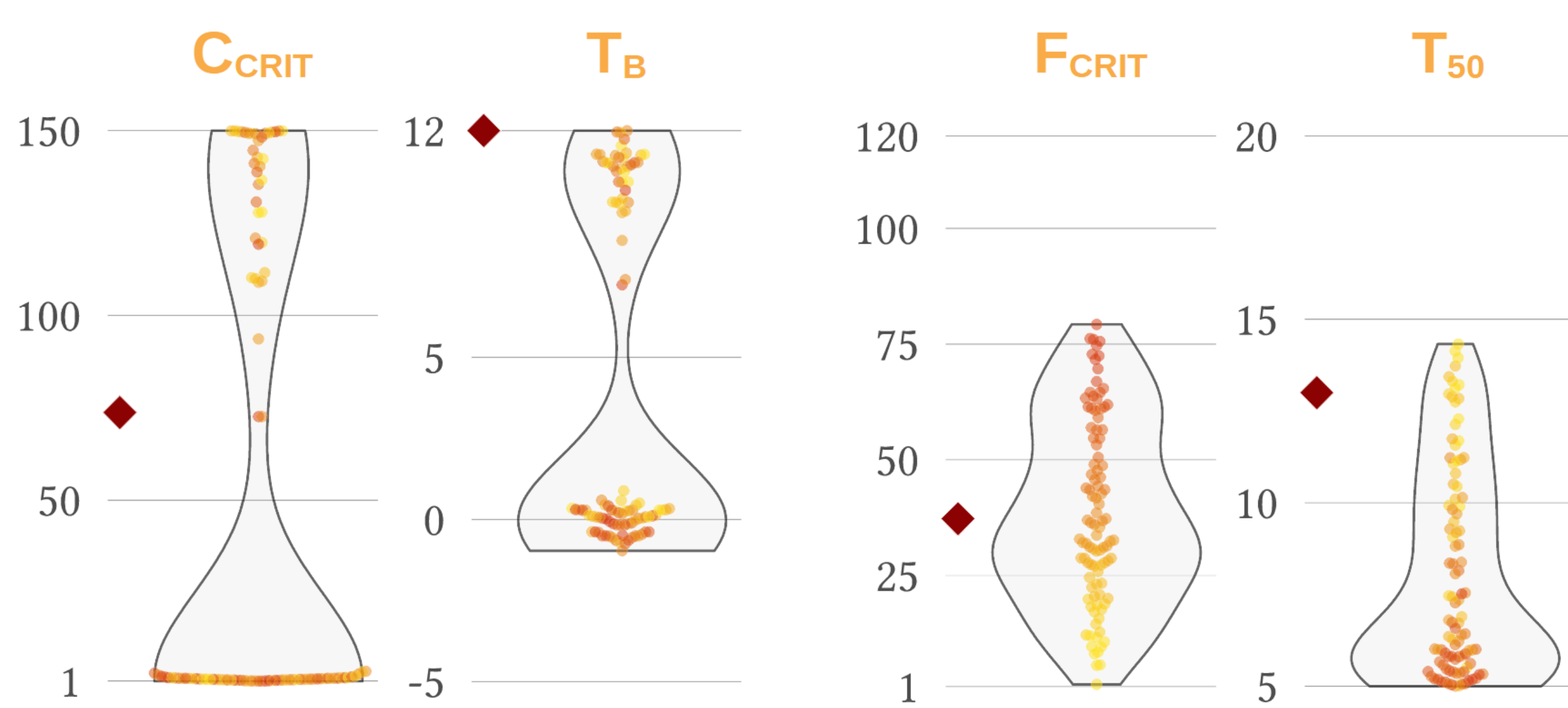
Fig. 1: Effects of stochasticity of CMA-ES calibration on PHENOFIT leaf unfolding submodel parameter

- Parameters are strongly dependent because of **structural correlations**

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

C_{CRIT} T_B F_{CRIT} T_{50}

0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9

Fig. 2: Leaf unfolding submodel parameter density

- Thus, parameter values are **not necessarily ecologically relevant**

2. Data and models

Tree occurrence data

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

Climate and soil data

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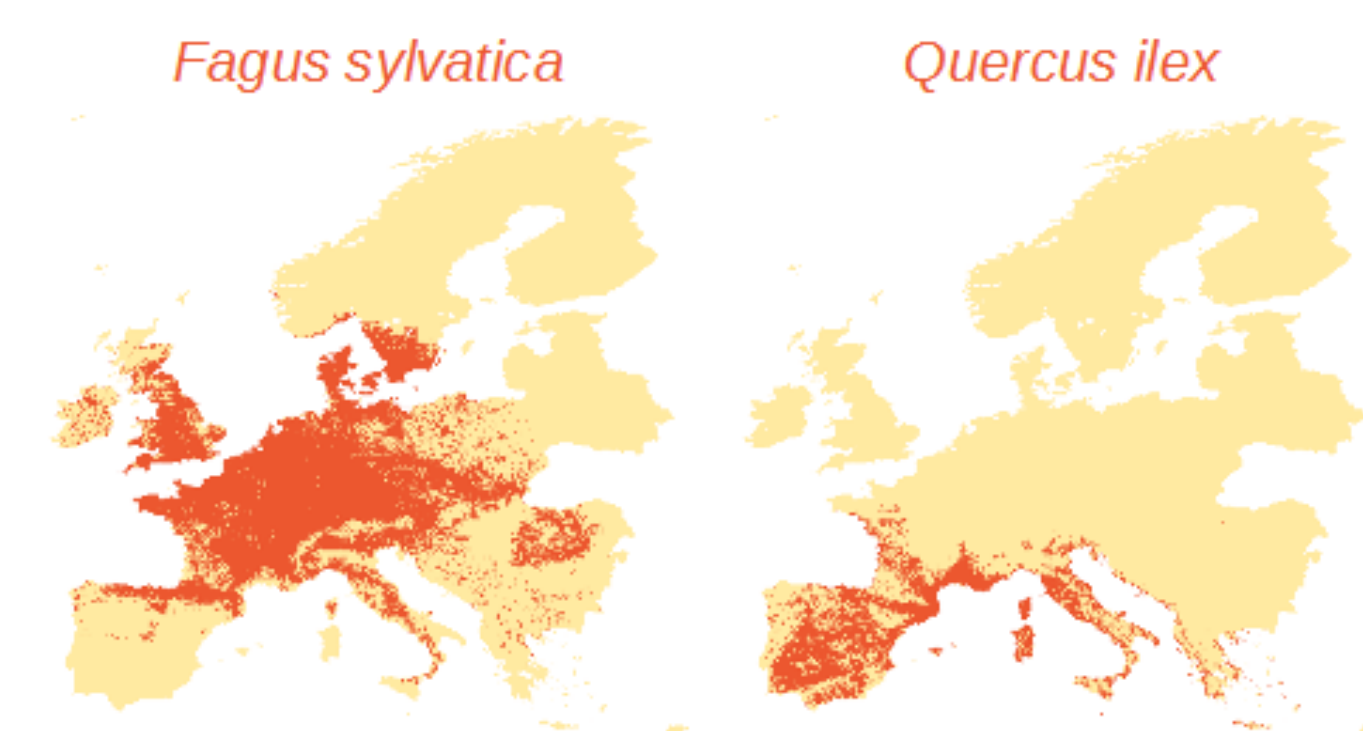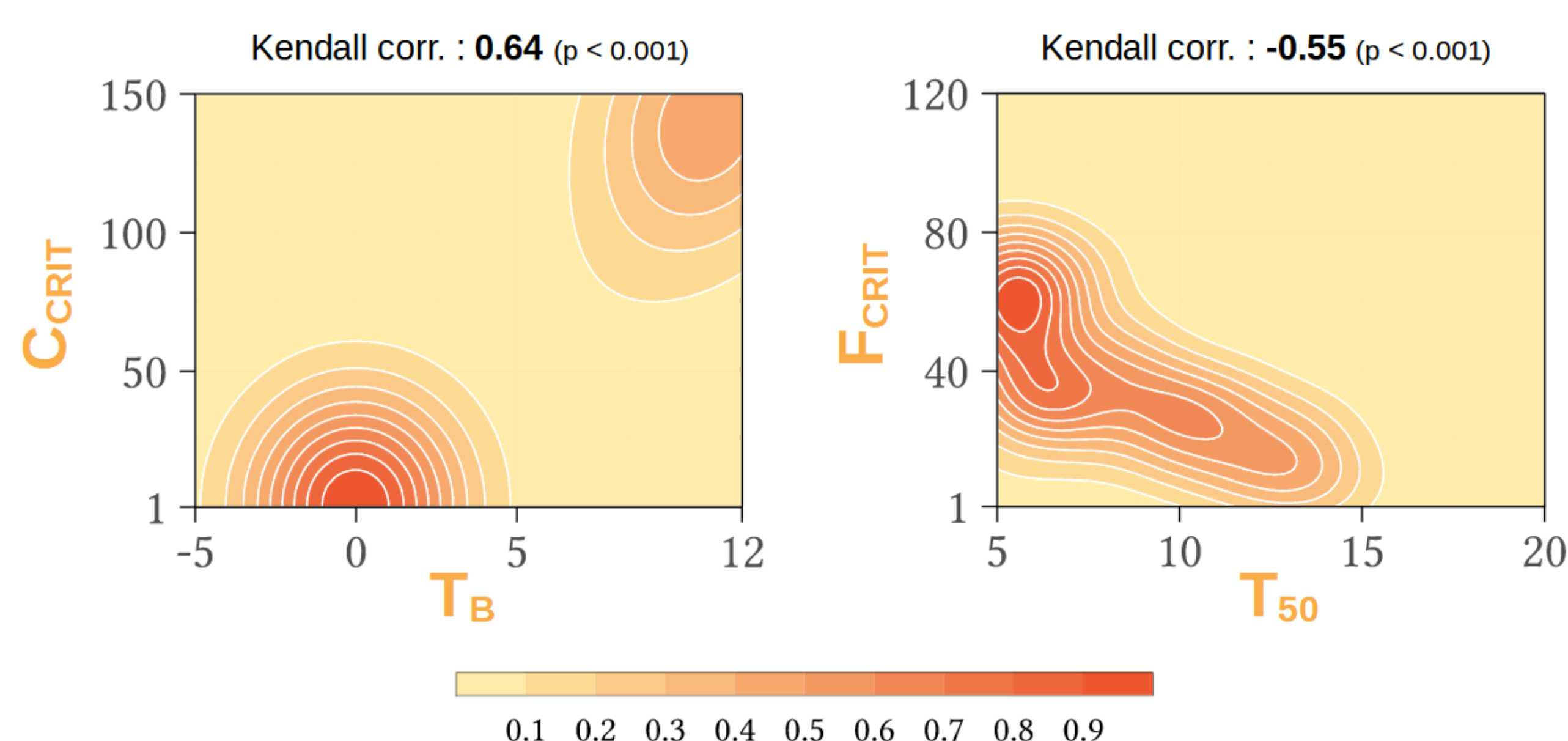- Soil variables were extracted from **EU-SoilHydroGrids** and **SoilGrids250m**

Fagus sylvatica   *Quercus ilex*

Fig. 3: Species distribution

Process-based models

- **PHENOFIT**[2]: focuses on phenology and simulates the fitness (survival and reproductive success) of an average individual
- **CASTANEA**[3]: simulates carbon and water fluxes in forest, with several processes (photosynthesis, stomatal conductance, respiration…)

4. Calibration results

PHENOFIT

Forward calibration / CMA-ES calibration

F. sylvatica: AUC = 0.79 / **AUC = 0.91**
Q. ilex: AUC = 0.86 / **AUC = 0.96**

CASTANEA

Forward calibration / CMA-ES calibration

F. sylvatica: AUC = 0.73 / **AUC = 0.92**
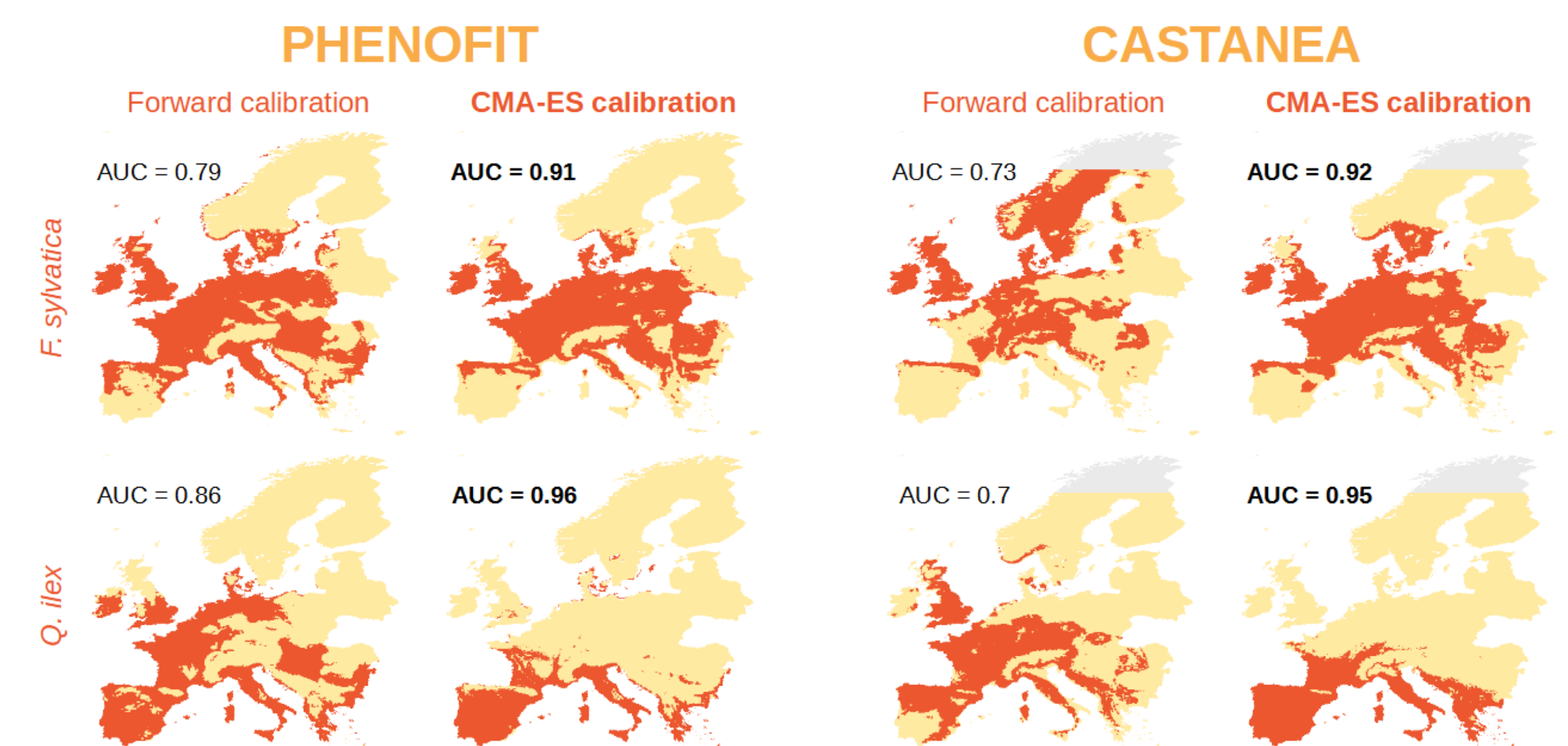*Q. ilex*: AUC = 0.7 / **AUC = 0.95**

Fig. 4: Species distribution maps obtained with PHENOFIT and CASTANEA forward and backward parametrization. Optimal threshold to dichotomize model outputs is the Youden Index-based cut-off point.

6. Advantages of CMA-ES

Comparison with ABC

CMA-ES slighty **outperforms** a widely used approximate bayesian computation (ABC) optimization method

**** **** ****

0.90

AUC

0.88

0.86

Fig. 5: Comparison of CMA-ES (red) and ABC-rejection (orange) methods, with F. sylvatica

Calibration performance

- AUC is **sharply improved** compared to classical (forward) parametrization

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

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

- CMA-ES backward optimization can be a **powerful technique** to calibrate submodels whose parameter values can hardly be experimentally measured
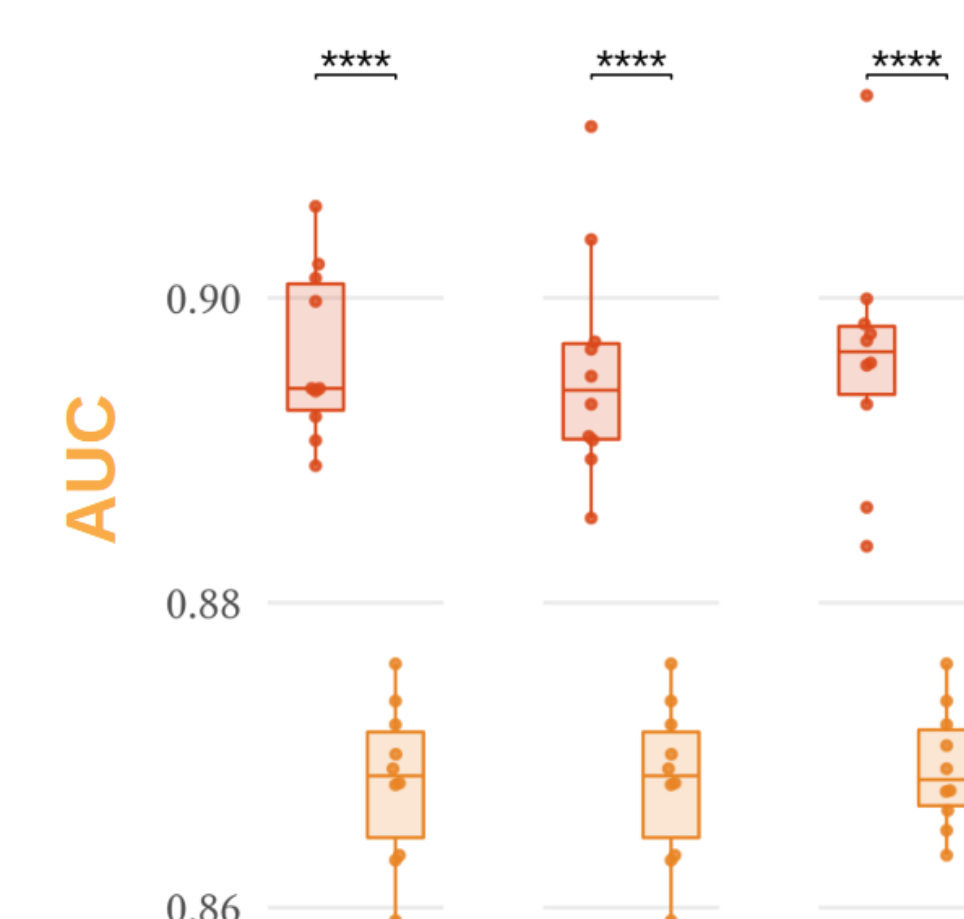
References

1. Hansen, N. and Ostermeier, A. 2001. *Completely derandomized self-adaptation in evolution strategies.* Evolutionary Computation 9: 159–195.
2. Chuine, I. and Beaubien, E. G. 2001. *Phenology is a major determinant of tree species range.* Ecology Letters 4: 500–510.
3. 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.