



Estimating process-based model parameters from species distribution data using the evolutionary algorithm CMA-ES

Victor van der Meersch, Isabelle Chuine

► To cite this version:

Victor van der Meersch, Isabelle Chuine. Estimating process-based model parameters from species distribution data using the evolutionary algorithm CMA-ES. *Methods in Ecology and Evolution*, 2023, 14 (7), pp.1808-1820. <10.1111/2041-210X.14119>. <hal-04271909>

HAL Id: hal-04271909

<https://hal.science/hal-04271909v1>

Submitted on 6 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Estimating process-based model parameters from species distribution data using the evolutionary algorithm CMA-ES

Victor Van der Meersch^{*}, Isabelle Chuine *CEFE, Université de Montpellier, CNRS, EPHE, IRD, Montpellier, France*

1. Two main types of species distribution models are used to project species range shifts in future climatic conditions: correlative and process-based models. Although there is some continuity between these two types of models, they are fundamentally different in their hypotheses (statistical relationships vs mechanistic relationships) and their calibration methods (SDMs tend to be occurrence data-driven while PBMs tend to be prior-driven).
2. One of the limitations to the use of process-based models is the difficulty to parameterize them for a large number of species compared to correlative SDMs. We investigated the feasibility of using an evolutionary algorithm (called covariance matrix adaptation evolution strategy, CMA-ES) to calibrate process-based models using species distribution data. This method is well established in some fields (robotics, aerospace research, ...), but has never been used, to our knowledge, in ecology, despite its ability to deal with very large space dimensions. Using tree species occurrence data across Europe, we adapted the CMA-ES algorithm to find appropriate values of model parameters. We estimated simultaneously 27 to 77 parameters of two process-based models simulating forest tree's ecophysiology for three species with varying range sizes and geographical distributions.
3. CMA-ES provided parameter estimates leading to better prediction of species distribution than parameter estimates based on experts knowledge. Our results also revealed that some model parameters and processes were strongly dependent, and different parameter combinations could therefore lead to high model accuracy.
4. We conclude that CMA-ES is an efficient state-of-the-art method to calibrate process-based models with a large number of parameters using species occurrence data. Inverse modelling using CMA-ES is a powerful method to calibrate process-based parameters which can hardly be measured. However, the method does not warranty that parameter estimates are correct because of several sources of bias, similarly to correlative models, and expert knowledge is required to validate results.

Keywords: calibration, evolutionary algorithm, cma-es, species distribution model, process-based model, trees

Introduction

The speed and magnitude of projected climate changes are profoundly affecting species distributions, ecological communities and ecosystem processes, and numerous ecological systems are now approaching

^{*}Corresponding author - victor.vandermeersch@cefe.cnrs.fr

tipping points (Lenton *et al.* 2008; Barnosky *et al.* 2012; Steffen *et al.* 2018; but see Brook *et al.* 2013). Large uncertainties on the persistence and the resilience of ecosystems exist. Ecological forecasting has now become a critical tool for managers and decision-makers (Urban 2015), and robust predictive approaches are necessary to provide reliable projections of species geographic range shifts and ecosystem functioning (Mouquet *et al.* 2015). Forecasting the dynamics of ecological systems for the upcoming decades and centuries is very difficult, because ecological systems are extremely complex, influenced by a lot of factors and processes, and climatic conditions with no analogues in the recent past are forecasted to become common (Williams *et al.* 2007; Radeloff *et al.* 2015; Fitzpatrick *et al.* 2018). Ecological models have thus increased in complexity over the last 50 years, incorporating more and more processes described with various degrees of complexity depending on their objectives.

Nowadays, two main types of species distribution models (SDM) are used to project species range shifts in future climatic conditions: correlative and process-based models (Dormann *et al.* 2012). The vast majority of currently used SDMs are correlative: they seek to find statistical relationships between various environmental descriptors and species presence and absence. They assume there is an equilibrium between species distribution and environment (equilibrium postulate, Guisan & Thuiller 2005), that there is no adaptive responses within a generation (no trait plasticity, Berzaghi *et al.* 2020), and that species niche is stable over macroevolutionary time (niche conservatism, Pearman *et al.* 2008). Most of them include a fairly large number of predictors (particularly in machine-learning approaches), and consider flexible transformations (linear, quadratic...) and interactions between them (Merow *et al.* 2014). Even though some authors advocate for “putting more biology into SDMs” (Higgins *et al.* 2012), parameters have no a priori defined ecological meaning (Dormann *et al.* 2012) and shape of response curves to environmental variables is generally not constrained based on biological considerations. Although these models are not always used correctly (Araújo *et al.* 2019; Santini *et al.* 2021), their flexibility makes them an important tool in predictive ecology (Mouquet *et al.* 2015). They have been widely used especially to generate species range projections under current and future climates (e.g. Guisan & Thuiller 2005). Nevertheless, their ability to accurately describe the effects of climate on species distributions has recently been questioned (e.g. Fourcade *et al.* 2018; Journé *et al.* 2020; Warren *et al.* 2021).

For all these reasons, another kind of models has been developed. Process-based models aim to translate into mathematical equations our knowledge about the physiological and ecological processes involved in an organism’s life, such as growth, reproduction, survival, movement, and interactions with other livebe-

ings. Process-based models take more time to develop and are more challenging to use, but they might provide a greater comprehension of the complexity of ecosystem dynamics and more robust projections in novel conditions (Evans 2012; Zurell *et al.* 2016; Singer *et al.* 2016; Urban *et al.* 2016). A wide variety of process-based models exists, from quite simple models (e.g. Kleidon & Mooney 2000) to much more complex ones (e.g. Dufrêne *et al.* 2005). They all rely on an explicit representation of processes at stake, with a direct biological interpretation - at least in principle (Connolly *et al.* 2017). Process-based models may also include some phenomenological relationships on lower-level processes, but never on the pattern itself. The choices about the specific processes to include into the model are made based on theory, empirical observations and the objectives of the research, and modeler subjectivity may play an important role. One of the challenges is to build a model with the appropriate amount of complexity: a too simplistic model might be unrealistic whereas a very complex model could be far beyond our ability to understand it (because of interconnected mechanisms) and calibrate it. Each model relies on different hypotheses with its own balance of complexity, accuracy and parsimony - and thus different numbers of unknown parameters to calibrate. Parameter values of this kind of models are obtained with different methods. Some of these parameters can be individually estimated with field observations or experimental data, or are already available in the literature. When this is not possible, groups of parameters defining a relationship between a process variable and environmental variables or other processes are estimated jointly using inverse modelling methods and data on the processes modelled (e.g. Cailleret *et al.* 2020; Asse *et al.* 2020).

Calibration (i.e. parameter setting and estimation) is a fundamental step in the modelling process. It allows the model to reproduce the reality with more or less success. The result of the calibration provides insights on the ability of the model to reproduce and explain reality (model predictive power). Calibration of complex models such as process-based SDMs is time-consuming, and modelers are often challenged by the dimension of the parameter space, the complexity of the possible correlations among parameters, and the scarcity of observed experimental data to calibrate them. Therefore, although the philosophy of such models is to measure parameters, statistical inference might be useful when data are not yet available to infer parameter values. Parameter inference can be achieved through several methods which have been developed in the last decades. Most of them fall into two categories: informal and statistical calibration. On the one hand, statistical calibration assumes a data-generating model, and a likelihood function, which quantifies the probability of the observed data given the model parameters, is expressed mathematically. This likelihood is the foundation of *maximum likelihood estimation* and *Bayesian inference*, two commonly

used methods for parameter estimation. In practice, deriving the likelihood function can be quite challenging for complex models with many parameters and thresholding. One approach is to use numerical methods such as *simulation-based inference* or \emph{approximate Bayesian computation} (e.g. [Hartig et al. 2014](#)). However, obtaining a reliable estimate of the likelihood of a complex process-based model can require a large number of simulated samples, which can be computationally expensive. On the other hand, informal calibration uses an informal objective function to measure the discrepancy between the model predictions and the observed data. By minimizing the objective function, one is able to identify the parameter values that best fit the observed data, even though it may not have the same statistical rigor as statistical calibration. This optimization step can be achieved with several methods, either deterministic (e.g. Nelder-Mead method) or stochastic (e.g. simulated annealing, evolutionary algorithms).

Recently, an approach belonging to the evolutionary algorithm family, called Covariance Matrix Adaptation Evolution Strategy (CMA-ES), has been proposed ([Hansen & Ostermeier 2001](#)). One of the advantages of CMA-ES is its ability to cumulate information over iterations in order to adapt its own parameters (in particular the covariance matrix), which makes it more robust to noise. CMA-ES is especially performant for non-separable problems (i.e. when the model parameters are dependent) and large search space. This method has been successfully applied in various fields such as aerospace (e.g. [Collange et al. 2010](#)), optics (e.g. [Gagné et al. 2008](#)), and robotics (e.g. [Hill et al. 2020](#)). CMA-ES is acknowledged to be one of the most efficient approaches in continuous black-box optimization ([Hansen et al. 2010](#)) but to our knowledge has never been used in ecology.

Here we explored the feasibility and interests of calibrating process-based SDMs with CMA-ES using species occurrence data as correlative SDMs do (fitted process-based models *sensu* [Dormann et al. 2012](#)). We focused on two forest process-based models of varying levels of complexity to evaluate the ability of CMA-ES to calibrate such models. The two models are PHENOFIT (27 to 36 parameters, [Chuine & Beaubien 2001](#)) and CASTANEA (77 parameters, [Dufrêne et al. 2005](#)). Each model also emphasizes different ecological processes: while PHENOFIT focuses on phenology and how it relates to survival and reproduction, CASTANEA focuses on carbon and water cycles. We focused on three European common tree species, with different range extent and ecological preferences in order to evaluate the algorithms performance in various geographical and climatic conditions. European beech (*Fagus sylvatica* L.) is one of the most widely distributed broadleaved tree in Europe (from southern Sweden to Sicily and from Spain to northwest Turkey), holm oak (*Quercus ilex* L.) is an evergreen broadleaved tree native of the Mediter-

ranean region, and silver fir (*Abies alba* Mill.) is a coniferous tree which mainly occurs in mountain forests of Central Europe and some parts of Southern and Eastern Europe.

1. Material and methods

1.1. Process-based models

All versions of the models used for this study are coded in Java and distributed by the CAPSIS platform.

PHENOFIT is a process-based species distribution model for forest tree species which focuses on phenology. It relies on the principle that the distribution of a tree species depends mainly on the synchronization of its timing of development to the local climatic conditions (Chuine & Beaubien 2001). It is composed of several submodels, including phenology models for leaves, flowers and fruits, and stress resistance models. It simulates the fitness (survival and reproductive success) of an average individual using daily meteorological data, soil water holding capacity and species specific parameters (see Appendix A for details). PHENOFIT has been validated for several North American and European species by comparing their known distribution to the modelled fitness (e.g. Morin *et al.* 2007; Saltré *et al.* 2013; Duputié *et al.* 2015; Gauzere *et al.* 2020).

CASTANEA is an ecophysiological process-based model which simulates carbon and water fluxes in forests (Dufrêne *et al.* 2005). The model simulates the ecosystem as an average tree with six compartments (leaves, branches, stem, coarse roots, fine roots and reserves). It is much more complex than PHENOFIT, with several processes described and computed, such as photosynthesis, stomatal opening, maintenance and growth respiration, transpiration, and carbon allocation (see Appendix A for details). CASTANEA requires daily meteorological variables and soil characteristics. The model has been initially validated at stand scale for beech (Davi *et al.* 2005), and was then successfully applied to other European species (e.g. Davi *et al.* 2006; Delpierre *et al.* 2012; Davi & Cailleret 2017).

Both models, in their standard version (called here after expert calibration), are parameterized using various sources of information. Some parameters are directly measured or found in the literature such as leaf life span, specific leaf area, LT50 (freezing temperature causing 50% mortality) of leaves, leaf reflectance, and so on. Other parameters cannot be measured at all, or their measurement require an enormous effort that cannot be deployed for a large number of species. These parameters are thus inferred by inverse modelling using either Bayesian methods or optimizing methods and data on the specific process

they are describing. This is for example the case for phenology models for which all parameters cannot be measured, especially those describing bud dormancy break regulation, since no method allow to measure dormancy break precisely so far (except for a few fruit tree species). Finally, a few parameters are prescribed based on expert knowledge as no data to estimate them exist. For this study, the standard versions of both models were run for the three species using the same climatic data used to do the inverse calibration (see 1.2.1. Climate data), in order to compare the correctness of the predictions obtained with the two types of calibration.

1.2. Data for the calibration

1.2.1. Climate and soil data

Raw climatic variables were extracted from ERA5-Land hourly dataset (Muñoz Sabater 2019, 2021) from 1970 to 2000, at a spatial resolution of 0.1 degree in latitude and longitude. We calculated the daily mean values of the following variables used by PHENOFIT and CASTANEA: minimum, mean and maximum daily temperatures, mean dewpoint temperature, daily precipitation, daily global radiation and daily mean wind speed. We computed the daily relative humidity with the ratio of vapor pressure and saturation vapor pressure (both calculated with Clausius-Clapeyron equation) using *humidity* R package (Cai 2019). Daily potential evapotranspiration was calculated with Penman–Monteith equation (FAO standard of hypothetical grass reference surface) using a slightly modified version of the $ET()$ function in *Evapotranspiration* R package (Guo *et al.* 2016).

Water content at field capacity and wilting point data were extracted from EU-SoilHydroGrids (Tóth *et al.* 2017) which is at 1km resolution. Percentage of sand, silt and clay particles, percentage of coarse fragments, bulk density and soil depth were extracted from SoilGrids250m (Hengl *et al.* 2017) at a 250m resolution. These data (except for soil depth) are provided at seven soil depths, so we summarized them (weighted sum or weighted mean) taking into account each layer width and total soil depth. Finally, all variables were upscaled at the ERA5-Land spatial resolution 0.1° using bilinear interpolation.

1.2.2. Tree occurrence data used for the calibration

Sources of occurrence data are known to differ even for common European trees (Duputié *et al.* 2014) and this makes it quite challenging to gather comprehensive data at a sufficient spatial resolution all over

Europe. The occurrence data we used essentially rely on the EU-Forest dataset (Mauri *et al.* 2017) which benefits from inventory and monitoring programmes implemented in most European countries. As EU-Forest is limited to forest ecosystems, we combined it with presence records extracted from the Global Biodiversity Information Facility (GBIF 2022, see Appendix B for all download links) but removing observations outside natural species ranges as defined by Atlas Flora Europae (AFE, Jalas & Suominen 1972–2005) and EuroVegMap (Bohn *et al.* 2003). By doing so, we also included occurrences of isolated native trees living outside forests, excluding records from arboreta or gardens where the species would have been planted as an exotic. For holm oak, we also added occurrence records in the Mediterranean Basin from the WOODIV database (Monnet *et al.* 2021), leaving out EU-Forest and GBIF records we had already gathered. We upscaled all species records at the ERA5-Land resolution (0.1°, see 1.2.1. Climate data), i.e. the species is considered to be present in the cell if there is at least one record. We finally obtained 21458 occurrence cells for beech, 6653 for holm oak and 5385 for silver fir (see Appendix B for details).

All the datasets described above are presence-only data. Therefore, we generated cells where species are supposed to be absent, i.e. pseudo-absence cells. In order to avoid as far as possible creating false absence data, we used EU-Forest cells where the species is not reported present as pseudo-absence cells. We assumed that national forest inventories were exhaustive (which is not true since only specific forest plots in a 0.1° cell are monitored). We obtained 25423 absence cells for beech, 37931 for holm oak and 38365 for silver fir (see Appendix C).

We selected subsets of 2000 points (1000 presences and 1000 pseudo-absences) in order to reduce computational costs. For each species, we generated ten presence clusters (k-means algorithm) of similar bioclimatic conditions based on annual climate normals computed with R package *dismo* (Hijmans *et al.* 2021) and ERA5-Land variables. In each cluster, we randomly sampled a number of cells where the species is present proportional to the total number of a number of cells where the species is present in the cluster. The aim of this stratified random sampling was to make sure that all species environmental preferences were proportionally represented. We then randomly sampled the same number of pseudo-absence cells (see Appendix B for details).

Regarding PHENOFIT model, we calibrated ten times each species parameter set, with 5 repetitions on 2 random subsets of presences/pseudo-absences, except for beech. In the latter case, we ran 10 repetitions on 10 subsets (i.e. 100 calibrations) to investigate both the effect of subsampling and the effect of stochasticity. Since CASTANEA computing time was much higher (see Table 1), we ran only two calibration for

each species (on 2 different random subsets).

1.3. Model calibration

1.3.1. Covariance Matrix Adaptation Evolution Strategy principles

Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is widely accepted as a robust optimization algorithm for non-linear, non-convex, as well as non-separated optimization problems in continuous domain (Hansen & Ostermeier 1996; Hansen & Ostermeier 2001; Hansen 2006). It is based on the principle of evolutionary biology, via recombination, mutation and selection of the most fit candidate solutions (i.e. parameter sets providing the best predictions). At each iteration:

- λ candidate solutions are evaluated, i.e. model runs λ times with λ different parameter sets and the objective function is evaluated
- the best μ candidate solutions are selected
- the weighted mean candidate solution m is computed (mean of the best μ parameter sets weighted by their objective function value)
- covariance matrix C and step size σ are updated (with information accumulated over several consecutive iterations)
- new λ candidate solutions are sampled in a normal distribution $\mathcal{N}(m, \sigma C)$, with both recombination (via the favorite solution m) and mutations (via the perturbations σC)

One of the strengths of this approach lies in the combination of rank- μ -update, where prior information from previous generations is exploited (mean of the previous covariance matrices, with a higher weight for recent generations), and cumulation, where correlations between generations are retained in an evolution path (sum of consecutive steps), to update the covariance matrix at each step (see Hansen 2016 for a detailed description of the algorithm).

1.3.2. CMA-ES in practice

One of the advantages of CMA-ES is that it does not require a complex parameter tuning: as best parameter values at a given time of the optimization process might no longer be efficient later, CMA-ES implements an internal adaptation of its parameters. We only chose the number of candidate solutions λ , depending on the optimization problem complexity (μ was set to $\lambda/2$). The default recommended value for λ is

$4 + 3\ln(N)$, where N is the number of parameters to calibrate (i.e. $\lambda \in [14, 17]$ in our case). We set $\lambda = 20$, in order to improve the global search capability (Hansen & Kern 2004) and take advantage of the computation power at our disposal. All model parameters were linear scaled into $[0; 10]$ so that the same standard deviation can be applied to all parameters: here we chose $\sigma = 2$ (see Nikolaus Hansen personal website for practical hints on variable encoding). Our stopping criterion for the optimization procedure was the budget, i.e. the number of model runs.

For an easier use and the sake of reproducibility, we chose to use a pure R implementation of CMA-ES available in the R package *cmaes* (Trautmann *et al.* 2011). The function *cma_es()* enables us to do λ function evaluations in parallel so as to substantially reduce computation time. It also allows us to define lower and upper bound constraints, by penalising the objective function value of the candidate solution if it violates the boundaries. We customized the *cma_es()* function to add an option to define death penalty constraints (rejection of the infeasible candidate solution who is sampled again), in order to define a range of ecologically possible solutions in terms of inequality constraints between parameters (see Appendix D for details about boundaries and constraints handling). Death penalty is the easiest way to handle constraints when the feasible region is fairly large, but it is not perfect as there is no use of information from infeasible points (i.e. points which violate the constraints).

The objective function for the calibration was the area under the receiver operating characteristic curve (AUC), evaluated against a subsets of 2000 points (see 1.2.2. Tree occurrence data). Although AUC has been criticized as an imperfect measure of model performance (Lobo *et al.* 2008; Leroy *et al.* 2018), we used it as objective function because our goal here was only to calibrate models by maximizing discriminating capacity (i.e. potential to correctly classify presences and absences) with a threshold-independent measure. We used the *AUC* R package (Ballings & Van den Poel 2013), and chose the two following model output variables as proxies of classification probabilities (i.e. used to determine if the species can be present or not): fitness index for PHENOFIT and carbon reserves for CASTANEA (see Appendix A).

We implemented the CMA-ES calibration on two computing clusters: GenOuest from IRISA-INRIA (genouest.org) and TGCC (*Très Grand Centre de Calcul*) from CEA (hpc.cea.fr). As the models are coded in Java (see 1.1 Process-based models), they need a process of deallocating memory handled by a *garbage collector*. For PHENOFIT, each function evaluation (i.e. each model simulation) was run on a 2-core computing unit in order to have enough computing resources for both simulation and garbage collection. We thus needed twice as many cores as functions evaluated in parallel. CASTANEA model requires a fairly

Table 1: Summary of model calibration settings. Average runtime was assessed on the GenOuest cluster.

Model	Output variable of interest	Number of parameters calibrated	Number of candidate solutions λ	Number of cores	Total memory	Number of model evaluations	Average runtime for one calibration
PHENOFIT	Fitness index	[27; 36]	20	40	80 GB	6000	~ 24 hours
CASTANEA	Carbon reserves	77	20	100	120 GB	4000	~ 20 days

high computation time, so we used a nested parallelism distribution, where each parallel simulation was distributed on 4 computing units. We thus used 4 times as many cores as functions evaluated in parallel, plus some extra cores for garbage collection. We used R package *future* (Bengtsson 2021) for parallel processing.

2. Results

2.1. Calibration results

Calibrations using species distribution data are hereafter called inverse calibrations, and calibrations based on expert knowledge, observations and measurements of the processes modelled are called expert calibrations.

CMA-ES calibration of PHENOFIT model allows an average 17.2% increase of AUC across the three species compared to expert calibration (Figure 1). The maximum increase is obtained for silver fir, from 0.72 to 0.9 (25%).

CMA-ES calibration of CASTANEA allows an average 23.7 % increase of AUC compared to expert calibration (Figure 2), and a maximum increase obtained for holm oak (34.7%).

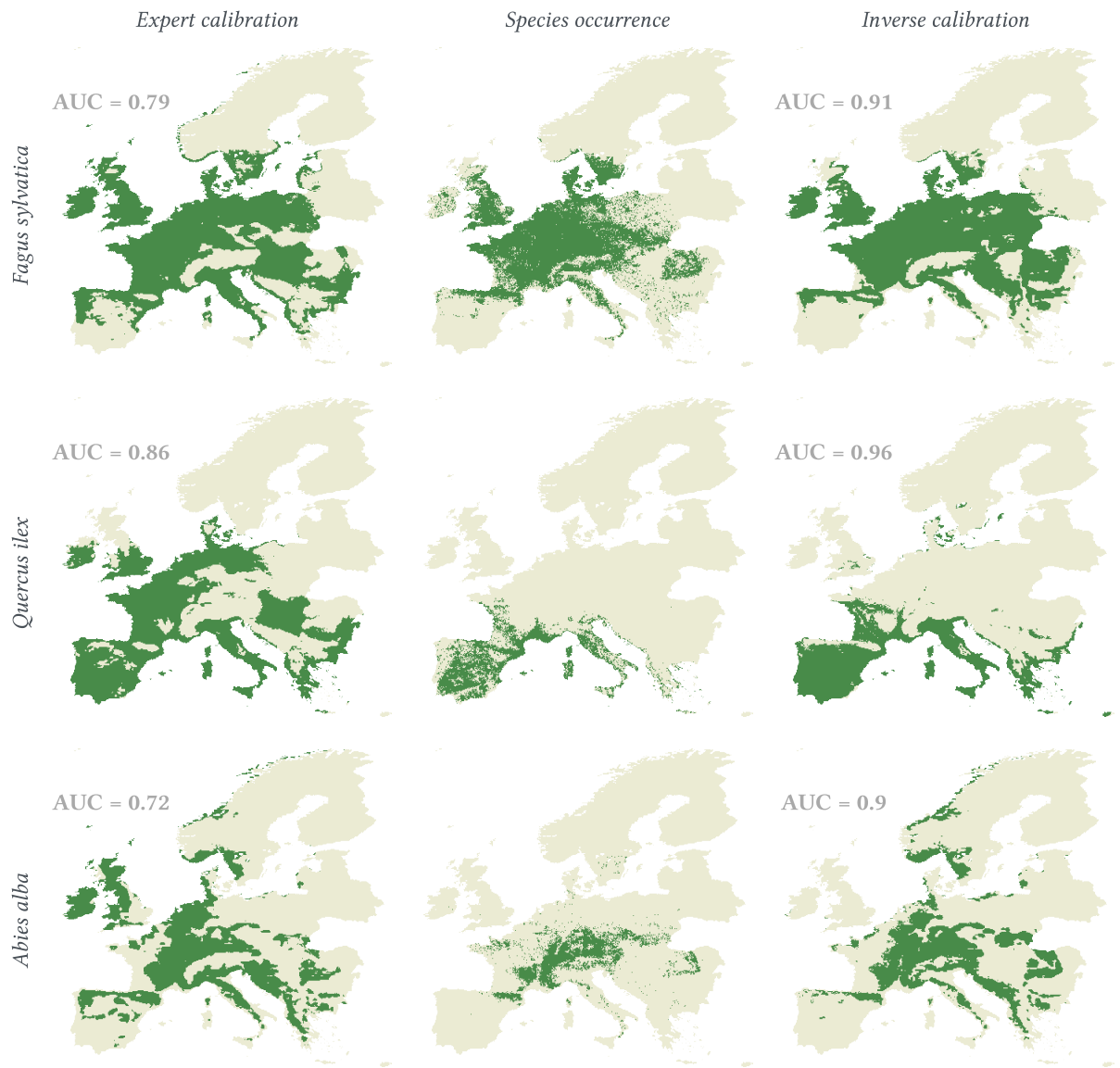


Figure 1: Species distribution maps obtained with PHENOFIT expert and inverse calibrations, compared with observed species occurrences. Optimal threshold to dichotomize model predicted fitness index in presence/absence is the Youden index-based cut-off point. Note that models predict species climatic niche which is larger than the realized niche that corresponds to species presence map.

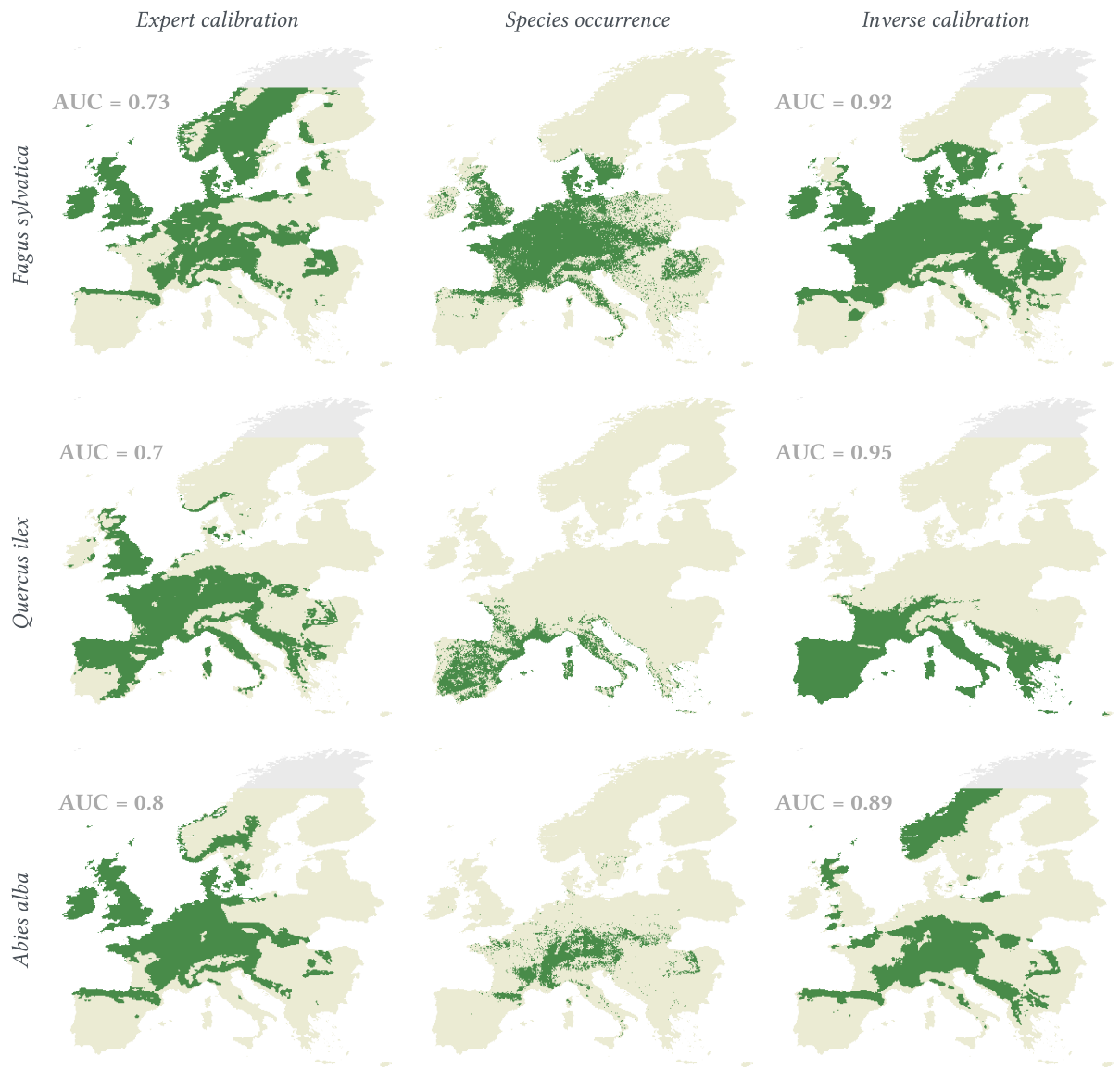


Figure 2: Species distribution maps obtained with CASTANEA expert and inverse calibrations, compared with observed species occurrences. Optimal threshold to dichotomize model predicted carbon reserves in presence/absence is the Youden index-based cut-off point. Note that models predict species climatic niche which is larger than the realized niche that corresponds to species presence map. Note also that CASTANEA cannot be used in high-latitude regions (grey area).

2.2. Impacts of subsampling and calibration stochasticity

2.2.1. Variability of calibration performance

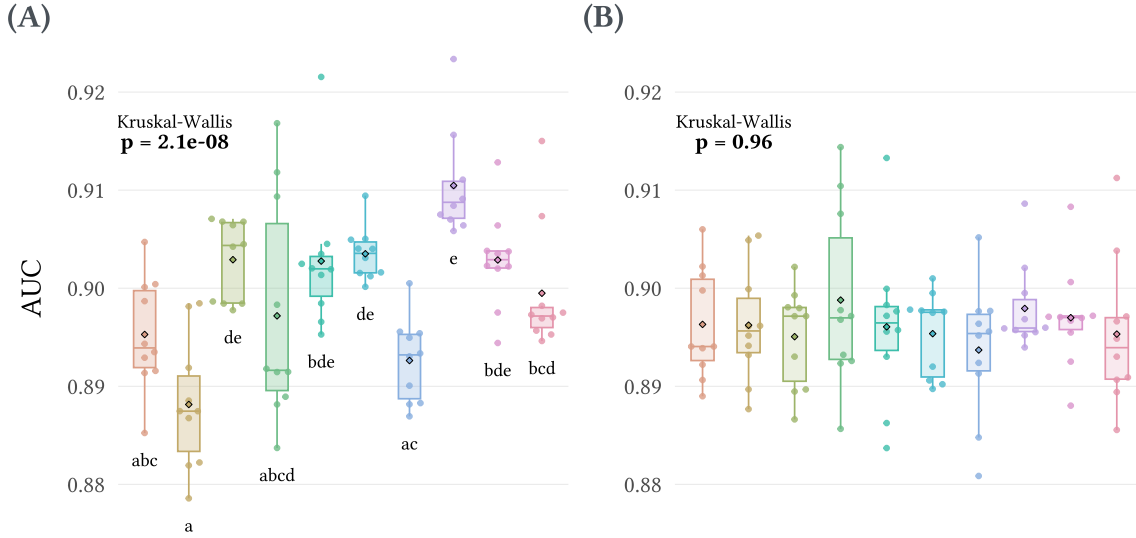


Figure 3: Effects of data sub-sampling and stochasticity on CMA-ES calibration using the PHENOFIT model for beech: **(A)** calibration AUC (calculated only with calibration cells) and **(B)** total AUC (calculated with all presence/absence cells). Each color is a different sub-sampling of occurrence data, each point is a calibration run. Diamonds (with black border) are mean AUC values. On **(A)**, the grouping letters represent the multiple comparisons with pairwise Dunn's tests.

The 100 calibrations of the PHENOFIT model realized for beech showed that random data subsampling had an effect on the final objective function value (i.e. the AUC computed on the 2000 calibration points). Kruskal-Wallis test was significant ($p = 2.1e-08$), meaning that at least one subset provided better AUC during calibration. According to Dunn's tests, 11 pairwise comparisons out of 45 were significant (Figure 3.A.). The calibration AUC ranged from 0.879 to 0.923 over all subsets, with a mean value of 0.9.

However, more importantly, the repetition of calibrations on different subsets had no significant impact on the total AUC computed on all presence/absence points (Kruskal-Wallis test, $p = 0.96$). Thus, no subset led to an overall better prediction of the species distribution (see Figure 3.B.). The total AUC ranged from 0.881 to 0.914, with a mean value of 0.896.

2.2.2. Non-identifiability of parameters

To illustrate the variability in the parameter estimates that can be obtained with the inverse calibration, we focused on the leaf unfolding date submodel of PHENOFIT (see [Appendix A](#) and [Appendix G](#)). The parameter values found by CMA-ES varied greatly across the 100 calibrations ([Figure 4](#)). For example, the critical amount of chilling C_{crit} required to break bud dormancy and the critical amount of forcing F_{crit} required to break bud ranged from 1.02 to 149.96 and from 1.5 to 79.26 respectively, with a mean value of 51.52 and 38.78. Their coefficient of variations were 126.7% and 51.9% respectively. Kendall correlation coefficient between C_{crit} and the threshold temperature of the response function to temperature during dormancy T_b is 0.64 ($p < 0.001$). Kendall correlation coefficient between F_{crit} and the mid-response temperature T_{50} is -0.55 ($p < 0.001$).

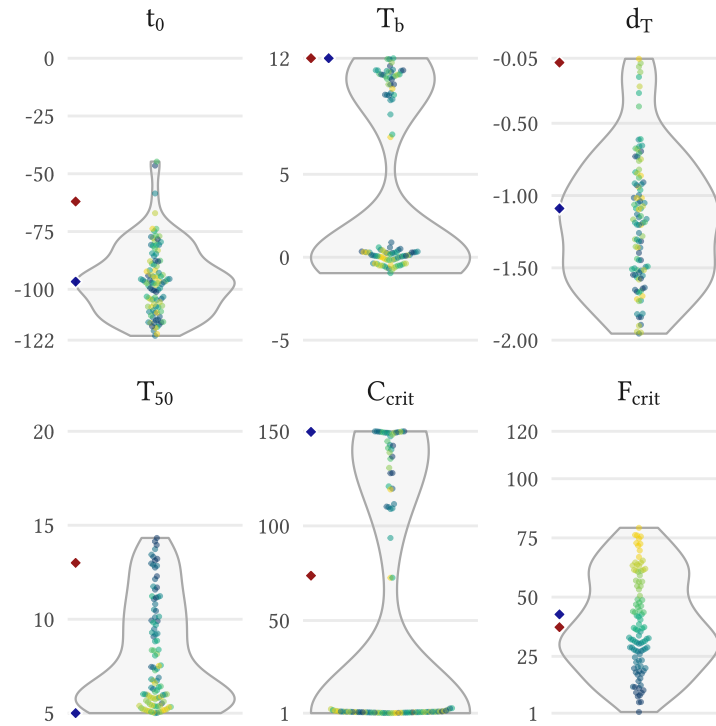


Figure 4: Effects of stochasticity of CMA-ES calibration on PHENOFIT leaf unfolding model parameter values for beech. Each panel is a parameter. Y-axis limits are lower and upper bounds used during calibration. Each point is a calibrated parameter value, color gradient is based on F_{crit} values. Red diamonds are parameter values obtained with expert calibration, blue ones are parameter values obtained with the best inverse calibration.

3. Discussion

Our results showed that CMA-ES is an efficient optimizer for inverse calibration of complex ecological models. The algorithm was able to find parameter sets that significantly improved the predictions of the calibrated models compared to the expert parametrization. However, our study also highlighted the issue of non-identifiability of parameter values due to the data limitation and the dependence between process-based model parameters, which may result in diverging parameter values even though the calibrated models describe the observed species distribution well.

3.1. Performance and advantages of CMA-ES to calibrate complex models in ecology

Here we demonstrated that inverse calibration with CMA-ES is feasible and provide good results for complex and runtime-expensive ecological models.

With a subsampling of species occurrence data, the algorithm succeeds in finding parameter sets which provide higher AUC values. The predictions of the calibrated models are sharply improved compared to expert parametrization (Figure 1 and Figure 2). Two striking examples are the increase in the performance of PHENOFIT model for silver fir, from 0.72 to 0.9, and of CASTANEA model for holm oak, from 0.7 to 0.95. Moreover, CMA-ES performed equally well regardless of the species occurrence subset used during calibration (Figure 3.B.), and thus permitted to find a good compromise between computational cost and calibration efficiency.

CMA-ES is a “generic” optimizer which can be applied to various problems. It is easy to use as it does not require an extensive tuning to efficiently explore the parameter space. We only had to choose the number of candidate solutions λ , and the initial search region (initial starting point and step size σ). As well as being quasi parameter-free, CMA-ES has several structural advantages particularly useful for complex optimization problems. First, the algorithm’s covariance matrix enables the learning of second-order information, which provides insights into pairwise dependencies between parameters. Second, the covariance matrix adaptation of CMA-ES is highly efficient in handling ill-conditioned and non-separable problems (Hansen *et al.* 2011). Last, CMA-ES’s update mechanism of step size σ (i.e. the mutation force) helps prevent premature convergence (Hansen & Ostermeier 2001), allowing the algorithm to explore more of the search space. CMA-ES has been shown to outperform several other optimization algorithms (Hansen *et al.* 2010), and is usually the most efficient method when the target cost (i.e. the number of

objective function evaluations) is about $100 * N$ (N being the dimension of the parameter search space, [Bäck et al. 2013](#)).

3.2. Non-identifiability of parameter values

Equifinality is a common problem encountered during model calibration, where multiple parameter sets can produce equally good fits to the observed data. This issue can arise due to several factors, including limitations in the quantity or quality of available data, competing processes within the model, and parameter interactions that affect the model output in complex ways.

In both models used in this study, biological mechanisms are explicitly calculated in several submodels (e.g. a leaf unfolding submodel or a stomatal opening submodel). A submodel output has inevitably a significant influence on the other submodels as biological processes can be highly dependent with feedbacks: in CASTANEA, for example, the stomatal opening affects the photosynthesis, and *vice versa*. Within each submodel, parameters are also strongly dependent because of structural correlations. To illustrate this problem, we focused on the beech leaf unfolding submodel of PHENOFIT (see [Appendix A](#)). This model has 6 parameters ([Chuine 2000](#)): a starting date of the processes (t_0), one parameter describing the response function to temperature during the dormancy phase (T_b), two parameters describing the response function to temperature during the phase of bud growth (d_T , T_{50}), and two parameters representing the sums of the daily responses to temperature during bud dormancy (C_{crit}) and during bud growth (F_{crit}) that respectively determine the date of bud dormancy break and the date of leaf unfolding (see [Appendix G](#) for details). Since no information on the date of bud dormancy break is available for the calibration, a first structural negative correlation exists between C_{crit} and F_{crit} : the same leaf unfolding date can be obtained with either a long dormancy phase and short bud growth phase or a short dormancy phase and a long bud growth phase. Other structural correlations exist between T_b and C_{crit} on the one hand and d_T/T_{50} and F_{crit} on the other hand: for example, a rapid accumulation of chilling units with a high critical chilling requirement could yield identical results as a slow accumulation with a low critical chilling requirement (i.e. the threshold temperature T_b and the critical chilling requirement C_{crit} are dependent, see Figure G.1.A. in [Appendix G](#)).

Consequently, several parameter sets may be statistically equivalent and parameters non-identifiable. In fact, calibration repetitions gave diverging parameter values ([Figure 4](#)) while being efficient in distinguishing between species presence and absence (i.e. AUC ~ 0.9 , [Figure 3](#)). Thus, even if the calibrated model

describes the observed species distribution very well, it does not necessarily mean that parameter values are ecologically relevant. This concern is similar to the criticisms against correlative SDMs, in which parameter values and correlations that well reproduce species ranges do not necessarily describe a complex biological reality. In our case, the constraints imposed by the explicit mathematical equations embedded in the models we used were not sufficient to ensure calibration convergence towards similar solutions that would have suggested a high biological realism. However, it is worth noting that we deliberately chose large parameter ranges (although biologically realistic, i.e. corresponding to the observations made on the different processes modelled across different species) in order to give free rein to the optimization algorithm. As our goal was to assess the performance of CMA-ES objectively, we did not attempt to minimize this non-identifiability issue by restricting the parameter space. To deal with equifinality issues, an avenue to explore could be the use of multiple objective functions during model calibration to assess different aspects of the model performance. Additionally, if a closed-form likelihood can be derived, one could use a Bayesian framework to combine prior knowledge and inverse estimation of parameters to constrain the parameter space and study the nature of trade-offs between parameters (Hartig *et al.* 2012; Cailleret *et al.* 2020).

3.3. Methodological issues and perspectives

Our goal here was to investigate the performance of CMA-ES to calibrate quite complex process-based species distribution models using species occurrence data. We did not attempt to validate our parametrizations using temporally or spatially independent data, and AUC was only used to determine if model outputs were consistent with species distributions. However, AUC is scale-invariant: it measures how well predictions are ranked rather than their absolute values. For example, with PHENOFIT, a species with a fitness of 0.8 could be considered as absent while another one with the same fitness could be considered as present. Therefore, when it is used as an objective function for model calibration, we probably lack some precision and consistency among species' parameters estimation. Further work could thus be conducted to examine the effects of choosing a different objective function.

It would also be valuable to use a significantly higher computing power, with an adapted version of CMA-ES. To improve the global search performance of CMA-ES, we slightly increased the number of candidate solutions λ (Hansen & Kern 2004) and used a computing cluster to evaluate λ functions in parallel. We were able to use between 40 and 120 cores, which is far from the computing power of some GPUs

(> 2000 cores). In this case, choosing a very large number of candidate solutions might not be the best choice. To use efficiently this large parallel computing power, one could rather use a CMA-ES restart strategy (e.g. IPOP-CMA-ES, [Auger & Hansen 2005](#)), where the number of candidate solutions is successively increased (by a factor of 2), and run these calibrations in parallel. Moreover, when a model requires a high computation time and thus only a small budget can be afforded, the original fitness function could be approximated with a surrogate model in order to reduce the number of original function evaluations required (e.g. [Auger et al. 2004](#) ; [Loshchilov et al. 2013](#)).

There are several issues regarding the process-based models we used which can impact their calibration and bias their parameter estimates. First, like any model, although they have a certain level of complexity, they are not a perfect representation of the reality, and their inherent structural errors increase the probability of finding parameter values that deviate from the true values of the underlying processes (see [Oberpriller et al. 2021](#)). Second, they do not necessarily include all the environmental factors at stake. For example, pedologic variable in PHENOFIT only involve the water holding capacity. In this model, other variables such as pH or soil texture are not considered. Third, and more importantly, they are used here to represent the fundamental niche, and to estimate the potential distribution of the species using pedoclimatic variables. When calibrated against observed distributions, which represent the realized niche, they face the same issues as correlative models, and their parameter estimated can be distorted because compensating for processes not represented in the model (e.g. biotic interactions). In addition, in our case here, land use management probably also play an important role in shaping tree realized distribution, while not being addressed in the models. We included GBIF occurrence data, and especially as much as possible isolated native tree records outside forests, to help correct this problem, but it is impossible to be exhaustive at the spatial resolution of 0.1°. At this scale, local variations of soil characteristics and of competitive interactions among trees (e.g. along an altitudinal gradient) can also not be considered.

Finally, several authors advocate for process-based modeling approaches relying upon species response functions that are a priori defined (e.g. [Higgins et al. 2020](#)). However, the main limitation of such models is the data availability to infer their parameters ([Urban et al. 2016](#)). Expert parameterization is often long and arduous. One possible way to facilitate parameter value estimation would be to use inverse calibration, and we demonstrated here that CMA-ES can be a powerful optimizer to this end. For example, CMA-ES driven by species occurrence data could be used to calibrate submodels whose parameter values cannot be measured or are too hard to measure experimentally. However, when a structural correlation exists

(i.e. trade-off between processes that are inherently present and interconnected in the model, as in the leaf unfolding submodel), inverse calibration might not provide the right parameter estimates. In such a case, expert knowledge, observations and measurements are necessary to determine *a posteriori* which estimates are the most realistic. This is possible in the case of process-based SDMs, and usually not feasible in the case of correlative SDMs. A combination of both expert and inverse calibrations might offer a new perspective for spreading the use of process-based models in predictive ecology, especially for climate change impact studies.

Acknowledgements

The authors would like to thank Hendrik Davi for helping us in using the CASTANEA model. We are also deeply grateful for many helpful comments from Florence Tauc. We would also like to thank François de Coligny, manager of the CAPSIS platform, Gilles Le Moguedec, and the GenOuest and TGCC teams for their support. Finally, we would like to thank Florian Hartig and another anonymous reviewer whose comments and suggestions helped us improve and clarify this manuscript.

V.V. was supported by a GAIA doctoral school PhD Fellowship.

Conflict of interest

The authors have no conflicts of interest to declare.

Author contributions

I.C. devised the main conceptual ideas. V.V. worked out the technical details, performed the numerical calculations and wrote the first draft of the manuscript. The two authors discussed the analyses and the results, and contributed to the final manuscript.

Data availability

ERA5-Land dataset is available on the [Copernicus Climate Change Service website](#). EU-SoilHydroGrids is available on the [European Soil Data Centre website](#). SoilGrids250m is available on the [International Soil Reference and Information Centre website](#). EU-Forest database is available on [FigShare](#). The R code associated with this work is available on [this GitHub repository](#), as well as on Zenodo ([10.5281/zenodo.7774981](https://doi.org/10.5281/zenodo.7774981)).

References

- Araújo, M.B., Anderson, R.P., Márcia Barbosa, A., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R.B., Zimmermann, N.E. & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, **5**. [10.1126/sciadv.aat4858](https://doi.org/10.1126/sciadv.aat4858)
- Asse, D., Randin, C.F., Bonhomme, M., Delestrade, A. & Chuine, I. (2020). Process-based models out-compete correlative models in projecting spring phenology of trees in a future warmer climate. *Agricultural and Forest Meteorology*, **285–286**, 107931. [10.1016/j.agrformet.2020.107931](https://doi.org/10.1016/j.agrformet.2020.107931)
- Auger, A. & Hansen, N. (2005). A restart CMA evolution strategy with increasing population size. *2005 IEEE Congress on Evolutionary Computation*, pp. 1769–1776 Vol. 2.
- Auger, A., Schoenauer, M. & Vanhaecke, N. (2004). LS-CMA-ES: A Second-Order Algorithm for Covariance Matrix Adaptation. *Parallel Problem Solving from Nature - PPSN VIII* (eds X. Yao, E.K. Burke, J.A. Lozano, J. Smith, J.J. Merelo-Guervós, J.A. Bullinaria, J.E. Rowe, P. Tino, A. Kabán & H.-P. Schwefel), pp. 182–191. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg.
- Bäck, T., Foussette, C. & Krause, P. (2013). Empirical Analysis. (eds T. Bäck, C. Foussette & P. Krause), pp. 55–83. Natural Computing Series. Springer, Berlin, Heidelberg.
- Ballings, M. & Van den Poel, D. (2013). *AUC: Threshold independent performance measures for probabilistic classifiers*.
- Barnosky, A.D., Hadly, E.A., Bascompte, J., Berlow, E.L., Brown, J.H., Fortelius, M., Getz, W.M., Harte, J., Hastings, A., Marquet, P.A., Martinez, N.D., Mooers, A., Roopnarine, P., Vermeij, G., Williams, J.W., Gillespie, R., Kitzes, J., Marshall, C., Matzke, N., Mindell, D.P., Revilla, E. & Smith, A.B. (2012). Approaching a state shift in Earth's biosphere. *Nature*, **486**, 52–58. [10.1038/nature11018](https://doi.org/10.1038/nature11018)
- Bengtsson, H. (2021). A unifying framework for parallel and distributed processing in r using futures.
- Berzaghi, F., Wright, I.J., Kramer, K., Oddou-Muratorio, S., Bohn, F.J., Reyer, C.P.O., Sabaté, S., Sanders, T.G.M. & Hartig, F. (2020). Towards a New Generation of Trait-Flexible Vegetation Models. *Trends in Ecology & Evolution*, **35**, 191–205. [10.1016/j.tree.2019.11.006](https://doi.org/10.1016/j.tree.2019.11.006)

- Bohn, U., Neuhäusl, R., Gisela Gollub, Hettwer, C., Neuhäuslová, Z., Raus, T., Schlüter, H. & Weber, H. (2003). *Map of the natural vegetation of europe - scale 1:2500000*.
- Brook, B.W., Ellis, E.C., Perring, M.P., Mackay, A.W. & Blomqvist, L. (2013). Does the terrestrial biosphere have planetary tipping points? *Trends in Ecology & Evolution*, **28**, 396–401. [10.1016/j.tree.2013.01.016](https://doi.org/10.1016/j.tree.2013.01.016)
- Cai, J. (2019). *Humidity: Calculate water vapor measures from temperature and dew point*.
- Cailleret, M., Bircher, N., Hartig, F., Hülsmann, L. & Bugmann, H. (2020). Bayesian calibration of a growth-dependent tree mortality model to simulate the dynamics of European temperate forests. *Ecological Applications*, **30**, e02021. [10.1002/eap.2021](https://doi.org/10.1002/eap.2021)
- Chaine, I. (2000). A Unified Model for Budburst of Trees. *Journal of Theoretical Biology*, **207**, 337–347. [10.1006/jtbi.2000.2178](https://doi.org/10.1006/jtbi.2000.2178)
- Chaine, I. & Beaubien, E.G. (2001). Phenology is a major determinant of tree species range. *Ecology Letters*, **4**, 500–510. [10.1046/j.1461-0248.2001.00261.x](https://doi.org/10.1046/j.1461-0248.2001.00261.x)
- Collange, G., Reynaud, S. & Hansen, N. (2010). Covariance matrix adaptation evolution strategy for multidisciplinary optimization of expendable launcher family. *13th AIAA/ISSMO multidisciplinary analysis optimization conference*.
- Connolly, S.R., Keith, S.A., Colwell, R.K. & Rahbek, C. (2017). Process, Mechanism, and Modeling in Macroecology. *Trends in Ecology & Evolution*, **32**, 835–844. [10.1016/j.tree.2017.08.011](https://doi.org/10.1016/j.tree.2017.08.011)
- Davi, H. & Cailleret, M. (2017). Assessing drought-driven mortality trees with physiological process-based models. *Agricultural and Forest Meteorology*, **232**, 279–290. [10.1016/j.agrformet.2016.08.019](https://doi.org/10.1016/j.agrformet.2016.08.019)
- Davi, H., Dufrêne, E., Francois, C., Le Maire, G., Loustau, D., Bosc, A., Rambal, S., Granier, A. & Moors, E. (2006). Sensitivity of water and carbon fluxes to climate changes from 1960 to 2100 in European forest ecosystems. *Agricultural and Forest Meteorology*, **141**, 35–56. [10.1016/j.agrformet.2006.09.003](https://doi.org/10.1016/j.agrformet.2006.09.003)
- Davi, H., Dufrêne, E., Granier, A., Le Dantec, V., Barbaroux, C., François, C. & Bréda, N. (2005). Modelling carbon and water cycles in a beech forest: Part II.: Validation of the main processes from organ to stand scale. *Ecological Modelling*, **185**, 387–405. [10.1016/j.ecolmodel.2005.01.003](https://doi.org/10.1016/j.ecolmodel.2005.01.003)

- Delpierre, N., Soudani, K., François, C., Le Maire, G., Bernhofer, C., Kutsch, W., Misson, L., Rambal, S., Vesala, T. & Dufrêne, E. (2012). Quantifying the influence of climate and biological drivers on the interannual variability of carbon exchanges in European forests through process-based modelling. *Agricultural and Forest Meteorology*, **154-155**, 99–112. [10.1016/j.agrformet.2011.10.010](https://doi.org/10.1016/j.agrformet.2011.10.010)
- Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B. & Singer, A. (2012). Correlation and process in species distribution models: Bridging a dichotomy. *Journal of Biogeography*, **39**, 2119–2131. [10.1111/j.1365-2699.2011.02659.x](https://doi.org/10.1111/j.1365-2699.2011.02659.x)
- Dufrêne, E., Davi, H., François, C., Maire, G. le, Dantec, V.L. & Granier, A. (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. [10.1016/j.ecolmodel.2005.01.004](https://doi.org/10.1016/j.ecolmodel.2005.01.004)
- Duputié, A., Rutschmann, A., Ronce, O. & Chuine, I. (2015). Phenological plasticity will not help all species adapt to climate change. *Global Change Biology*, **21**, 3062–3073. [10.1111/gcb.12914](https://doi.org/10.1111/gcb.12914)
- Duputié, A., Zimmermann, N.E. & Chuine, I. (2014). Where are the wild things? Why we need better data on species distribution. *Global Ecology and Biogeography*, **23**, 457–467. [10.1111/geb.12118](https://doi.org/10.1111/geb.12118)
- Evans, M.R. (2012). Modelling ecological systems in a changing world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 181–190. [10.1098/rstb.2011.0172](https://doi.org/10.1098/rstb.2011.0172)
- Fitzpatrick, M.C., Blois, J.L., Williams, J.W., Nieto-Lugilde, D., Maguire, K.C. & Lorenz, D.J. (2018). How will climate novelty influence ecological forecasts? Using the Quaternary to assess future reliability. *Global Change Biology*, **24**, 3575–3586. [10.1111/gcb.14138](https://doi.org/10.1111/gcb.14138)
- Fourcade, Y., Besnard, A.G. & Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, **27**, 245–256. [10.1111/geb.12684](https://doi.org/10.1111/geb.12684)
- Gagné, C., Beaulieu, J., Parizeau, M. & Thibault, S. (2008). Human-competitive lens system design with evolution strategies. *Applied Soft Computing*, **8**, 1439–1452. [10.1016/j.asoc.2007.10.018](https://doi.org/10.1016/j.asoc.2007.10.018)

- Gauzere, J., Teuf, B., Davi, H., Chevin, L.-M., Caignard, T., Leys, B., Delzon, S., Ronce, O. & Chuine, I. (2020). Where is the optimum? Predicting the variation of selection along climatic gradients and the adaptive value of plasticity. A case study on tree phenology. *Evolution Letters*, **4**, 109–123. [10.1002/evl3.160](https://doi.org/10.1002/evl3.160)
- GBIF. (2022). The global biodiversity information facility. <https://www.gbif.org> [accessed 24 January 2022]
- Guisan, A. & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009. [10.1111/j.1461-0248.2005.00792.x](https://doi.org/10.1111/j.1461-0248.2005.00792.x)
- Guo, D., Westra, S. & Maier, H.R. (2016). An R package for modelling actual, potential and reference evapotranspiration. *Environmental Modelling & Software*, **78**, 216–224. [10.1016/j.envsoft.2015.12.019](https://doi.org/10.1016/j.envsoft.2015.12.019)
- Hansen, N. (2016). The CMA evolution strategy: A tutorial.
- Hansen, N. (2006). The CMA Evolution Strategy: A Comparing Review. (eds J.A. Lozano, P. Larrañaga, I. Inza & E. Bengoetxea), pp. 75–102. *Studies in Fuzziness and Soft Computing*. Springer, Berlin, Heidelberg.
- Hansen, N., Auger, A., Ros, R., Finck, S. & Posik, P. (2010). Comparing Results of 31 Algorithms from the Black-Box Optimization Benchmarking BBOB-2009. *ACM-GECCO Genetic and Evolutionary Computation Conference*. Portland, United States.
- Hansen, N. & Kern, S. (2004). Evaluating the CMA Evolution Strategy on Multimodal Test Functions. *Parallel Problem Solving from Nature - PPSN VIII* (eds X. Yao, E.K. Burke, J.A. Lozano, J. Smith, J.J. Merelo-Guervós, J.A. Bullinaria, J.E. Rowe, P. Tino, A. Kabán & H.-P. Schwefel), pp. 282–291. *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg.
- Hansen, N., Niederberger, A.S.P., Guzzella, L. & Koumoutsakos, P. (2009). A Method for Handling Uncertainty in Evolutionary Optimization With an Application to Feedback Control of Combustion. *IEEE Transactions on Evolutionary Computation*, **13**, 180–197. [10.1109/TEVC.2008.924423](https://doi.org/10.1109/TEVC.2008.924423)
- Hansen, N. & Ostermeier, A. (1996). Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. *Proceedings of IEEE International Conference on Evo-*

- lutionary Computation*, pp. 312–317.
- Hansen, N. & Ostermeier, A. (2001). Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, **9**, 159–195. [10.1162/106365601750190398](https://doi.org/10.1162/106365601750190398)
- Hansen, N., Ros, R., Mauny, N., Schoenauer, M. & Auger, A. (2011). Impacts of invariance in search: When CMA-ES and PSO face ill-conditioned and non-separable problems. *Applied Soft Computing*, **11**, 5755–5769. [10.1016/j.asoc.2011.03.001](https://doi.org/10.1016/j.asoc.2011.03.001)
- Hartig, F., Dislich, C., Wiegand, T. & Huth, A. (2014). Technical Note: Approximate Bayesian parameterization of a process-based tropical forest model. *Biogeosciences*, **11**, 1261–1272. [10.5194/bg-11-1261-2014](https://doi.org/10.5194/bg-11-1261-2014)
- Hartig, F., Dyke, J., Hickler, T., Higgins, S.I., O’Hara, R.B., Scheiter, S. & Huth, A. (2012). Connecting dynamic vegetation models to data – an inverse perspective. *Journal of Biogeography*, **39**, 2240–2252. [10.1111/j.1365-2699.2012.02745.x](https://doi.org/10.1111/j.1365-2699.2012.02745.x)
- Hengl, T., Jesus, J.M. de, Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotic, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S. & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, **12**, e0169748. [10.1371/journal.pone.0169748](https://doi.org/10.1371/journal.pone.0169748)
- Higgins, S.I., Larcombe, M.J., Beeton, N.J., Conradi, T. & Nottebrock, H. (2020). Predictive ability of a process-based versus a correlative species distribution model. *Ecology and Evolution*, **10**, 11043–11054. [10.1002/ece3.6712](https://doi.org/10.1002/ece3.6712)
- Higgins, S.I., O’Hara, R.B. & Römermann, C. (2012). A niche for biology in species distribution models. *Journal of Biogeography*, **39**, 2091–2095. [10.1111/jbi.12029](https://doi.org/10.1111/jbi.12029)
- Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. (2021). *Dismo: Species distribution modeling*.
- Hill, A., Laneurit, J., Lenain, R. & Lucet, E. (2020). Online gain setting method for path tracking using CMA-ES: Application to off-road mobile robot control. *IROS 2020, International Conference on Intelligent Robots and Systems*. Las Vegas, United States.

- Jalas, J. & Suominen, J. (1972–2005). *Atlas florae europaeae*. Committee for Mapping the Flora of Europe; Societas Biologica Fennica Vanamo, Helsinki, Finland.
- Journé, V., Barnagaud, J., Bernard, C., Crochet, P. & Morin, X. (2020). Correlative climatic niche models predict real and virtual species distributions equally well. *Ecology*, **101**. [10.1002/ecy.2912](https://doi.org/10.1002/ecy.2912)
- Kleidon, A. & Mooney, H.A. (2000). A global distribution of biodiversity inferred from climatic constraints: Results from a process-based modelling study. *Global Change Biology*, **6**, 507–523. [10.1046/j.1365-2486.2000.00332.x](https://doi.org/10.1046/j.1365-2486.2000.00332.x)
- Lenton, T.M., Held, H., Kriegler, E., Hall, J.W., Lucht, W., Rahmstorf, S. & Schellnhuber, H.J. (2008). Tipping elements in the Earth’s climate system. *Proceedings of the National Academy of Sciences*, **105**, 1786–1793. [10.1073/pnas.0705414105](https://doi.org/10.1073/pnas.0705414105)
- Leroy, B., Delsol, R., Hugueny, B., Meynard, C.N., Barhoumi, C., Barbet-Massin, M. & Bellard, C. (2018). Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography*, **45**, 1994–2002. [10.1111/jbi.13402](https://doi.org/10.1111/jbi.13402)
- Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151. [10.1111/j.1466-8238.2007.00358.x](https://doi.org/10.1111/j.1466-8238.2007.00358.x)
- Loshchilov, I., Schoenauer, M. & Sèbag, M. (2013). Bi-population CMA-ES algorithms with surrogate models and line searches. *Proceedings of the 15th annual conference companion on Genetic and evolutionary computation*, pp. 1177–1184. GECCO ’13 Companion. Association for Computing Machinery, New York, NY, USA.
- Mauri, A., Strona, G. & San-Miguel-Ayanz, J. (2017). EU-Forest, a high-resolution tree occurrence dataset for Europe. *Scientific Data*, **4**, 160123. [10.1038/sdata.2016.123](https://doi.org/10.1038/sdata.2016.123)
- Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., McMahon, S.M., Normand, S., Thuiller, W., Wüest, R.O., Zimmermann, N.E. & Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography*, **37**, 1267–1281. [10.1111/ecog.00845](https://doi.org/10.1111/ecog.00845)

- Monnet, A.-C., Cilleros, K., Médail, F., Albassatneh, M.C., Arroyo, J., Bacchetta, G., Bagnoli, F., Barrina, Z., Cartereau, M., Casajus, N., Dimopoulos, P., Domina, G., Doxa, A., Escudero, M., Fady, B., Hampe, A., Matevski, V., Misfud, S., Nikolic, T., Pavon, D., Roig, A., Barea, E.S., Spanu, I., Strid, A., Vendramin, G.G. & Leriche, A. (2021). WOODIV, a database of occurrences, functional traits, and phylogenetic data for all Euro-Mediterranean trees. *Scientific Data*, **8**, 89. [10.1038/s41597-021-00873-3](https://doi.org/10.1038/s41597-021-00873-3)
- Morin, X., Augspurger, C. & Chuine, I. (2007). Process-Based Modeling of Species' Distributions: What Limits Temperate Tree Species' Range Boundaries? *Ecology*, **88**, 2280–2291. [10.1890/06-1591.1](https://doi.org/10.1890/06-1591.1)
- Mouquet, N., Lagadeuc, Y., Devictor, V., Doyen, L., Duputié, A., Eveillard, D., Faure, D., Garnier, E., Gimenez, O., Huneman, P., Jabot, F., Jarne, P., Joly, D., Julliard, R., Kéfi, S., Kergoat, G.J., Lavorel, S., Le Gall, L., Meslin, L., Morand, S., Morin, X., Morlon, H., Pinay, G., Pradel, R., Schurr, F.M., Thuiller, W. & Loreau, M. (2015). REVIEW: Predictive ecology in a changing world. *Journal of Applied Ecology*, **52**, 1293–1310. [10.1111/1365-2664.12482](https://doi.org/10.1111/1365-2664.12482)
- Muñoz Sabater, J. (2021). ERA5-land hourly data from 1950 to 1980. <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land>
- Muñoz Sabater, J. (2019). ERA5-land hourly data from 1981 to present. <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land>
- Oberpriller, J., Cameron, D.R., Dietze, M.C. & Hartig, F. (2021). Towards robust statistical inference for complex computer models. *Ecology Letters*, **24**, 1251–1261. [10.1111/ele.13728](https://doi.org/10.1111/ele.13728)
- Pearman, P.B., Guisan, A., Broennimann, O. & Randin, C.F. (2008). Niche dynamics in space and time. *Trends in Ecology & Evolution*, **23**, 149–158. [10.1016/j.tree.2007.11.005](https://doi.org/10.1016/j.tree.2007.11.005)
- Radeloff, V.C., Williams, J.W., Bateman, B.L., Burke, K.D., Carter, S.K., Childress, E.S., Cromwell, K.J., Gratton, C., Hasley, A.O., Kraemer, B.M., Latzka, A.W., Marin-Spiotta, E., Meine, C.D., Munoz, S.E., Neeson, T.M., Pidgeon, A.M., Rissman, A.R., Rivera, R.J., Szymanski, L.M. & Usinowicz, J. (2015). The rise of novelty in ecosystems. *Ecological Applications*, **25**, 2051–2068. [10.1890/14-1781.1](https://doi.org/10.1890/14-1781.1)
- Saltré, F., Saint-Amant, R., Gritti, E.S., Brewer, S., Gaucherel, C., Davis, B.A.S. & Chuine, I. (2013). Climate or migration: What limited European beech post-glacial colonization? *Global Ecology and*

Biogeography, **22**, 1217–1227. [10.1111/geb.12085](https://doi.org/10.1111/geb.12085)

Santini, L., Benítez-López, A., Maiorano, L., Cengic, M. & Huijbregts, M.A.J. (2021). Assessing the reliability of species distribution projections in climate change research. *Diversity and Distributions*, **27**, 1035–1050. [10.1111/ddi.13252](https://doi.org/10.1111/ddi.13252)

Singer, A., Johst, K., Banitz, T., Fowler, M.S., Groeneveld, J., Gutiérrez, A.G., Hartig, F., Krug, R.M., Liess, M., Matlack, G., Meyer, K.M., Pe'er, G., Radchuk, V., Voinopol-Sassu, A.-J. & Travis, J.M.J. (2016). Community dynamics under environmental change: How can next generation mechanistic models improve projections of species distributions? *Ecological Modelling*, **326**, 63–74. [10.1016/j.ecolmodel.2015.11.007](https://doi.org/10.1016/j.ecolmodel.2015.11.007)

Steffen, W., Rockström, J., Richardson, K., Lenton, T.M., Folke, C., Liverman, D., Summerhayes, C.P., Barnosky, A.D., Cornell, S.E., Crucifix, M., Donges, J.F., Fetzer, I., Lade, S.J., Scheffer, M., Winkelmann, R. & Schellnhuber, H.J. (2018). Trajectories of the Earth System in the Anthropocene. *Proceedings of the National Academy of Sciences*, **115**, 8252–8259. [10.1073/pnas.1810141115](https://doi.org/10.1073/pnas.1810141115)

Tóth, B., Weynants, M., Pásztor, L. & Hengl, T. (2017). 3D soil hydraulic database of Europe at 250 m resolution. *Hydrological Processes*, **31**, 2662–2666. [10.1002/hyp.11203](https://doi.org/10.1002/hyp.11203)

Trautmann, H., Mersmann, O. & Arnu, D. (2011). *Cmaes: Covariance matrix adapting evolutionary strategy*.

Urban, M.C. (2015). Accelerating extinction risk from climate change. *Science*, **348**, 571–573. [10.1126/science.aaa4984](https://doi.org/10.1126/science.aaa4984)

Urban, M.C., Bocedi, G., Hendry, A.P., Mihoub, J.-B., Pe'er, G., Singer, A., Bridle, J.R., Crozier, L.G., De Meester, L., Godsoe, W., Gonzalez, A., Hellmann, J.J., Holt, R.D., Huth, A., Johst, K., Krug, C.B., Leadley, P.W., Palmer, S.C.F., Pantel, J.H., Schmitz, A., Zollner, P.A. & Travis, J.M.J. (2016). Improving the forecast for biodiversity under climate change. *Science*, **353**, aad8466. [10.1126/science.aad8466](https://doi.org/10.1126/science.aad8466)

Warren, D.L., Dornburg, A., Zapfe, K. & Iglesias, T.L. (2021). The effects of climate change on Australia's only endemic Pokémon: Measuring bias in species distribution models. *Methods in Ecology and Evolution*, **12**, 985–995. [10.1111/2041-210X.13591](https://doi.org/10.1111/2041-210X.13591)

- Williams, J.W., Jackson, S.T. & Kutzbach, J.E. (2007). Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences*, **104**, 5738–5742. [10.1073/pnas.0606292104](https://doi.org/10.1073/pnas.0606292104)
- Zurell, D., Thuiller, W., Pagel, J., Cabral, J.S., Münkemüller, T., Gravel, D., Dullinger, S., Normand, S., Schiffrers, K.H., Moore, K.A. & Zimmermann, N.E. (2016). Benchmarking novel approaches for modelling species range dynamics. *Global Change Biology*, **22**, 2651–2664. [10.1111/gcb.13251](https://doi.org/10.1111/gcb.13251)

Supplementary Appendix A: Insights on the models

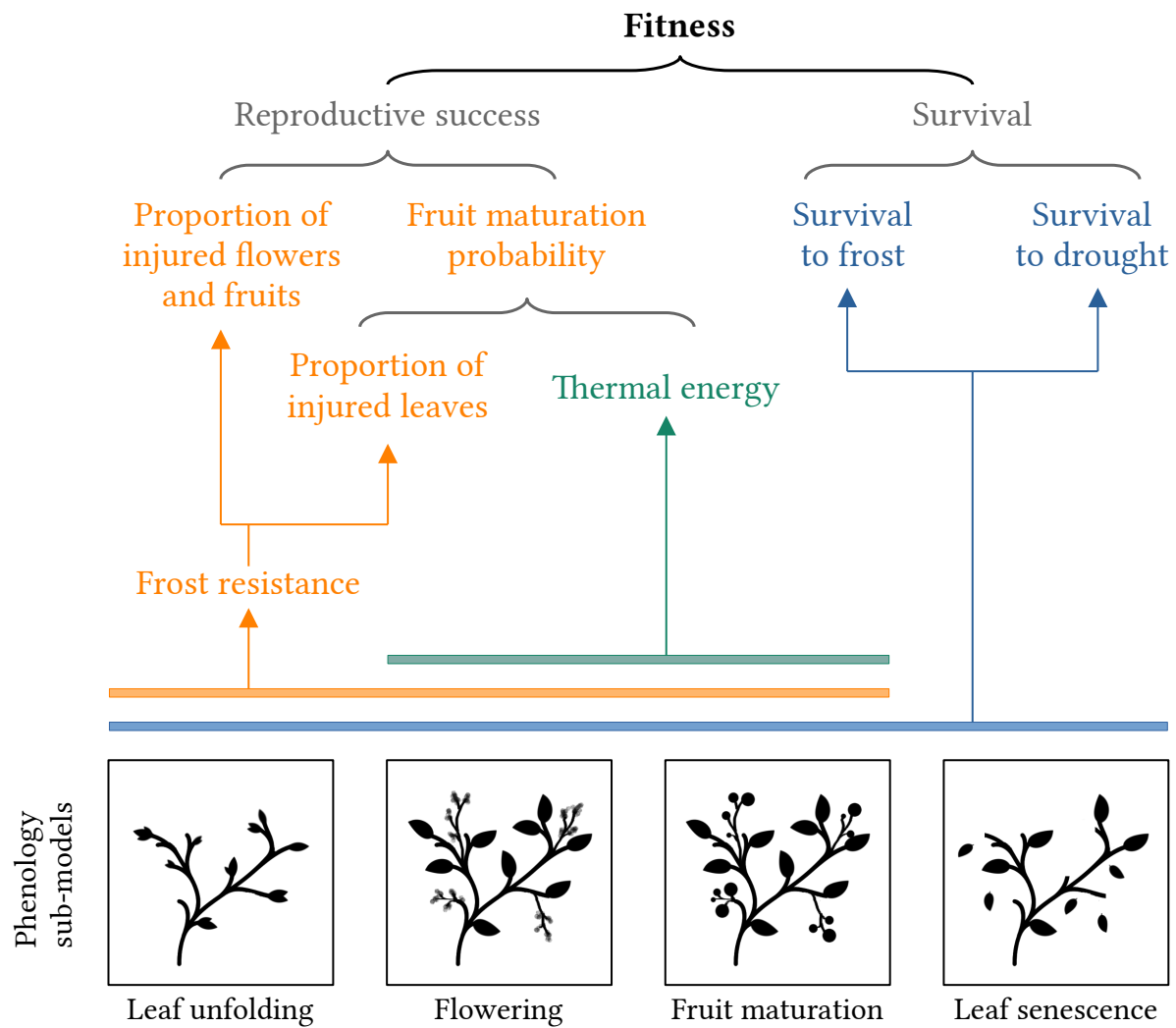


Figure A.1: PHENOFIT model in a nutshell.

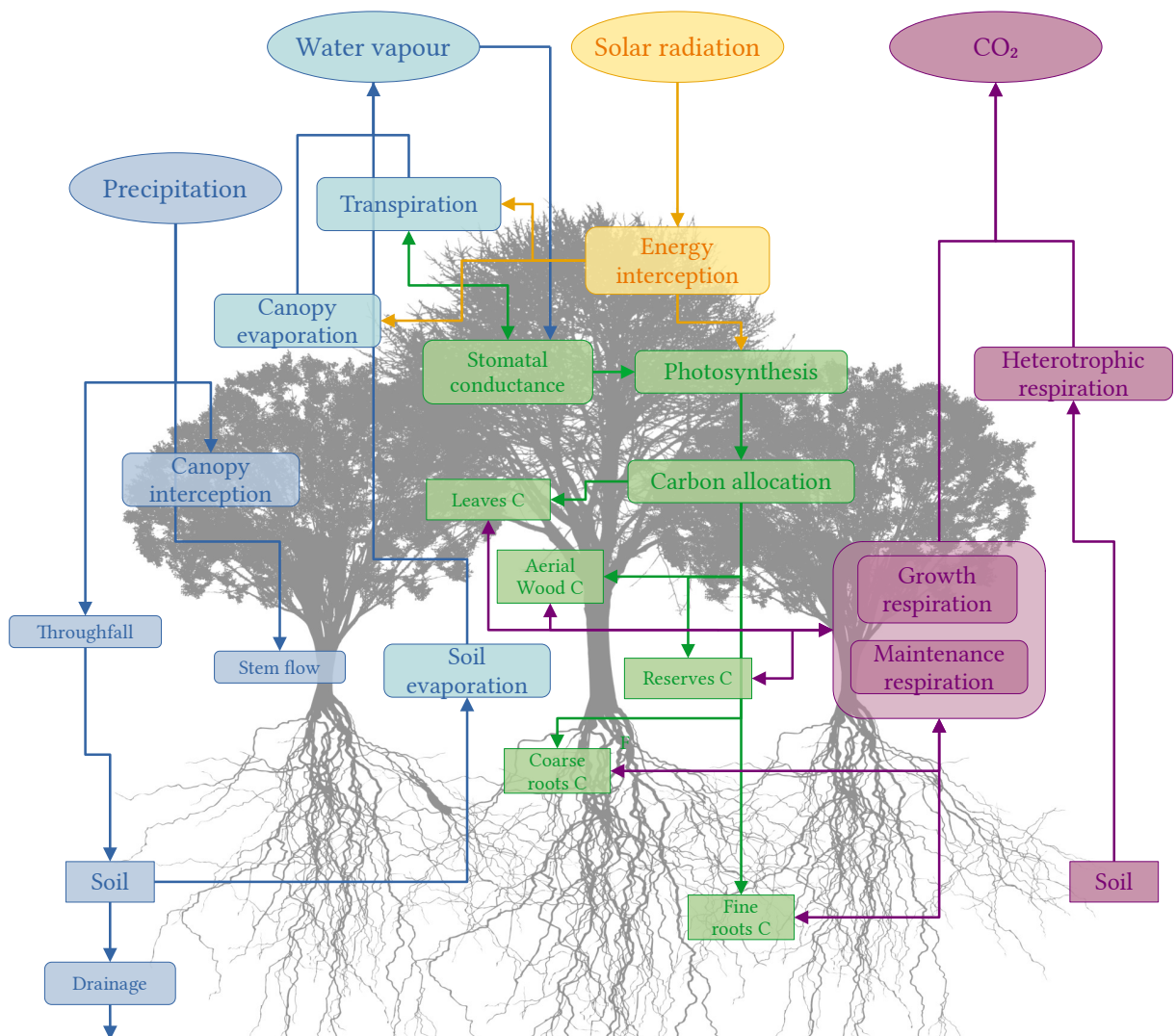


Figure A.2: CASTANEA model in a nutshell.

Supplementary Appendix B: Processing of occurrence data

Table B.1: GBIF download links

Species	Number of occurrences	Download link
<i>Fagus sylvatica</i>	718.898	https://doi.org/10.15468/dl.e9wasa
<i>Quercus ilex</i>	78.979	https://doi.org/10.15468/dl.2a4haw
<i>Abies alba</i>	119.891	https://doi.org/10.15468/dl.my6c9t

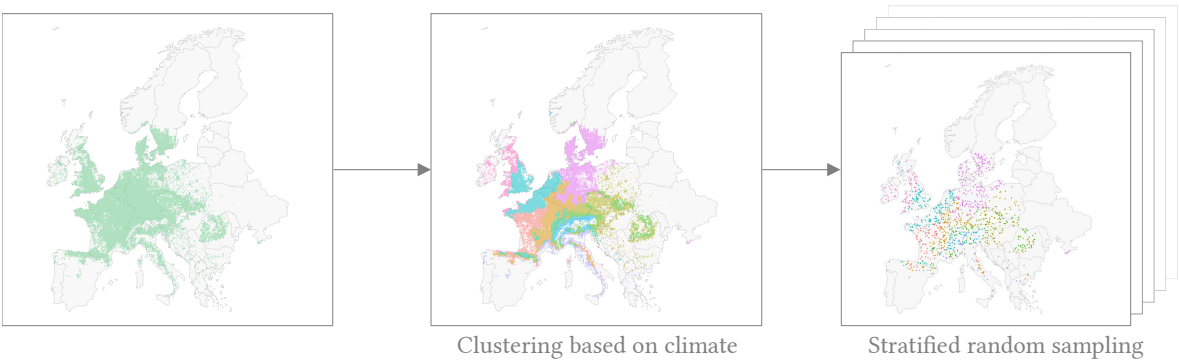


Figure B.1: Stratified random sampling of beech presence records based on climate clusters.

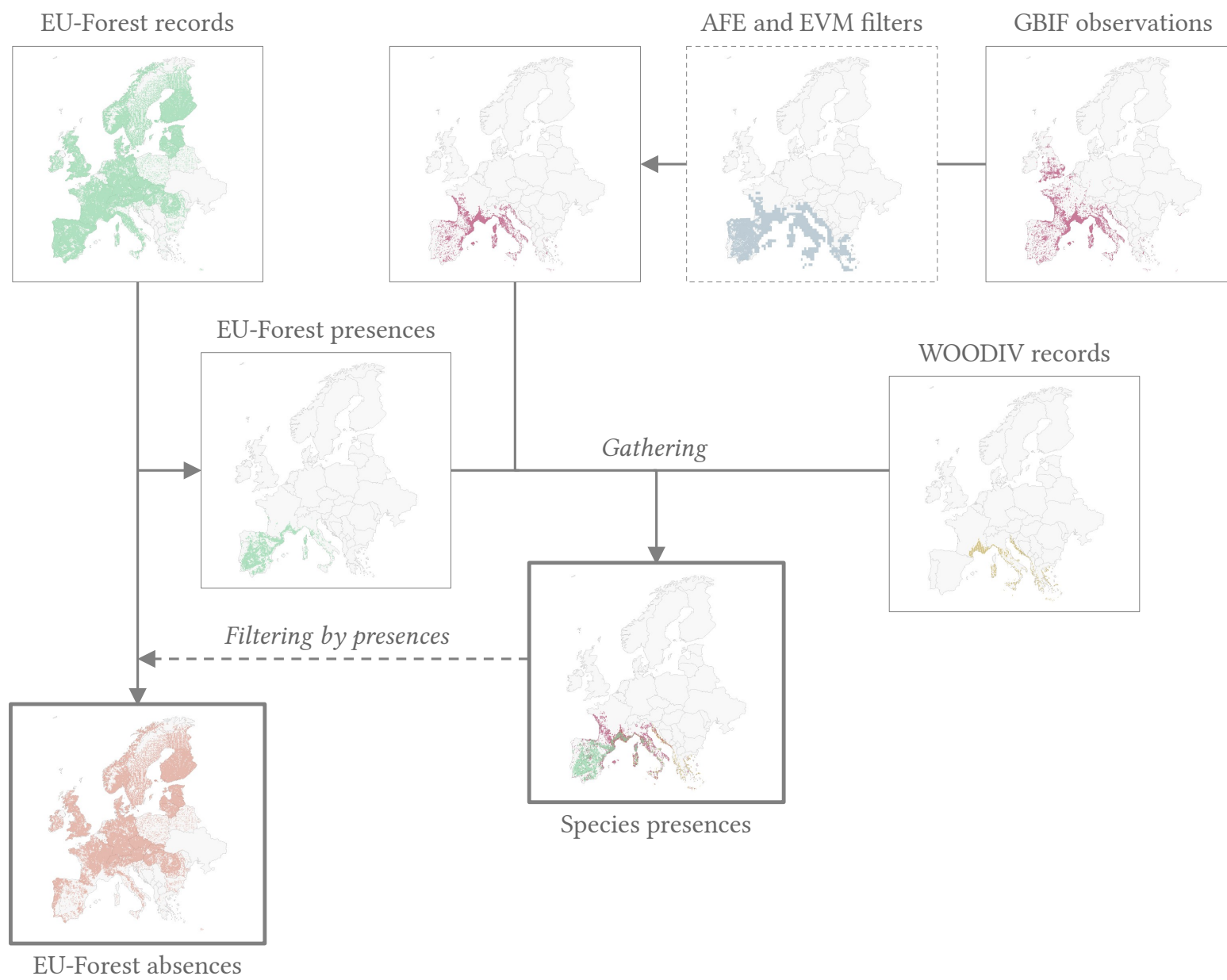


Figure B.2: Processing of holm oak occurrence records. GBIF: Global Biodiversity Information Facility, AFE: Atlas Flora Europae, EVM: EuroVegMap.

Supplementary Appendix C: Species distributions

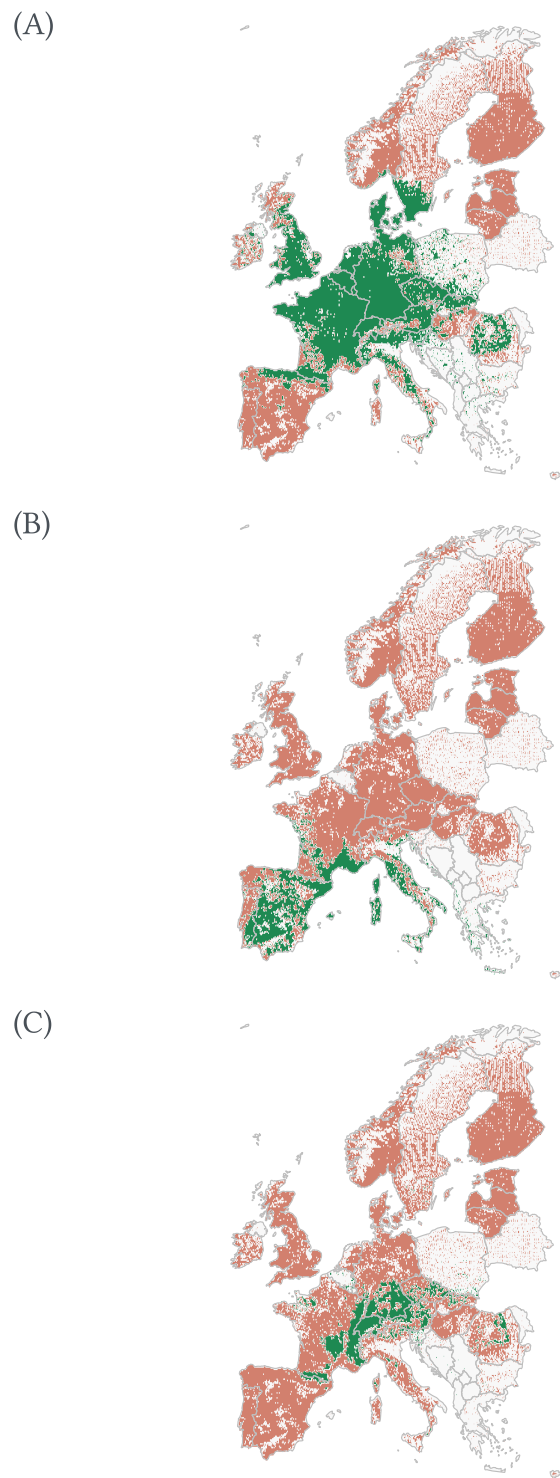


Figure C.1: Species distributions of (A) beech, (B) holm oak and (C) silver fir. Green cells are 0.1° cells where species is present, orange cells where species is supposed to be absent.

Supplementary Appendix D: Objective function and constraint handling in CMA-ES

The AUC as an objective function

The AUC is the area under the receiver operating characteristic curve, which plots sensitivity (correctly predicted positive fraction) as a function of commission error (falsely predicted positive fraction), as the probability threshold discriminating presence/absence varies. It is a discrimination metric, which has been widely used in the species distribution modelling literature.

Box constraint handling

With this constraint handling - implemented by default in the R package *cmaes* (Trautmann *et al.* 2011) - each evaluated solution is guaranteed to lie within the feasible space. Let's say we have a parameter vector x . For each parameter x_i , we have a lower bound lb_i and an upper bound ub_i . If a parameter x_i violates one of this bound, we set x_i to a new value $x_i^{repaired}$ equal to the closest boundary value (lb_i or ub_i). We thus obtained a new parameter set $x^{repaired}$, with a minimal $\|x - x^{repaired}\|$ value. This new feasible solution $x^{repaired}$ is used for the evaluation of the objective function $AUC_{model}(x^{repaired})$, and to compute a penalty term $pen = \sum_i (x_i - x_i^{repaired})^2 = \|x^{repaired} - x\|^2$. Then $x^{repaired}$ is discarded, and the algorithm computes the penalized objective function of $x^{repaired}$ as follows: $AUC_{model}(x) = AUC_{model}(x^{repaired}) + pen$. This boundary handling could be improved with adaptive weights (see Hansen *et al.* 2009).

Ecological infeasibility constraint

We added a simple way to handle ecological constraint (e.g. unfolding before flowering in beech mixed bud) with a death penalty. When a parameter vector x violates a constraint, it is rejected and generated again. The main drawback of this approach is that CMA-ES does not use information from unfeasible

points. An other approach could be to set $AUC_{model}(x) = 0$. However, as our feasible space was large, the death penalty constraint worked well in our case.

We applied an inequality constraint on both *Fagus sylvatica* and *Quercus ilex*, which have mixed buds (leaves and flowers within the same bud): unfolding must happen before flowering. On the contrary, we did not apply any inequality constraint on *Abies alba* simple bud phenology parameters.

Supplementary Appendix E: Holm oak and silver fir calibrations

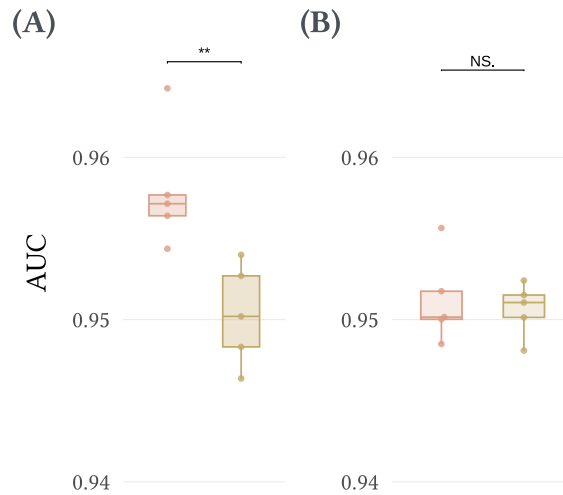


Figure E.1: CMA-ES calibration using the PHENOFIT model and holm oak: **(A)** calibration AUC (only calibration cells) and **(B)** total AUC (every presence/absence cells). Each color is a different sub-sampling of occurrence data, each point is a calibration run. The black horizontal bars represent the pairwise Mann-Whitney tests between the two subsets.

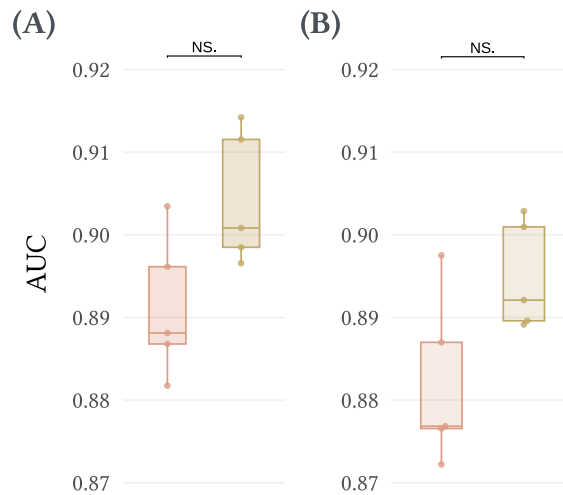


Figure E.2: CMA-ES calibration using the PHENOFIT model and silver fir: **(A)** calibration AUC (only calibration cells) and **(B)** total AUC (every presence/absence cells). Each color is a different sub-sampling of occurrence data, each point is a calibration run. The black horizontal bars represent the pairwise Mann-Whitney tests between the two subsets.

Supplementary Appendix F: Raw model outputs

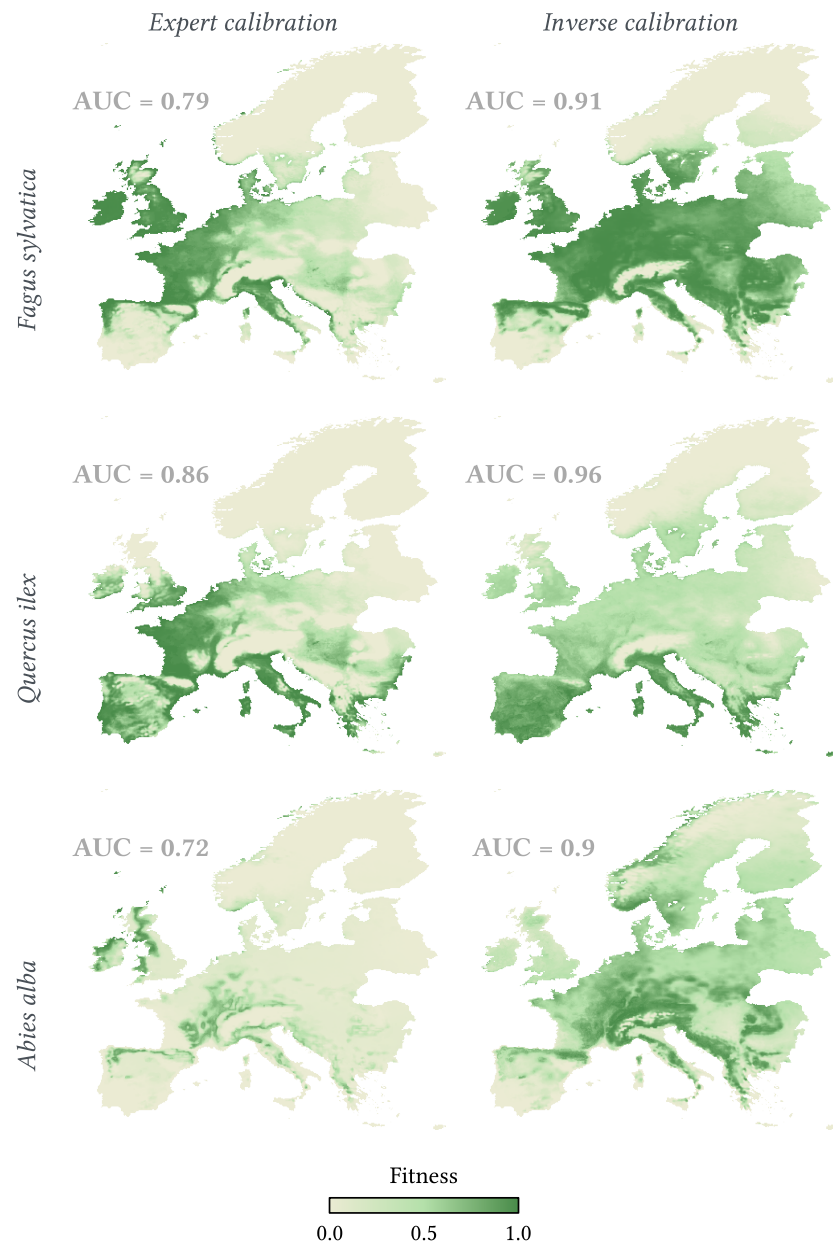


Figure F.1: Fitness index predicted by PHENOFIT with the expert and the inverse calibrations.

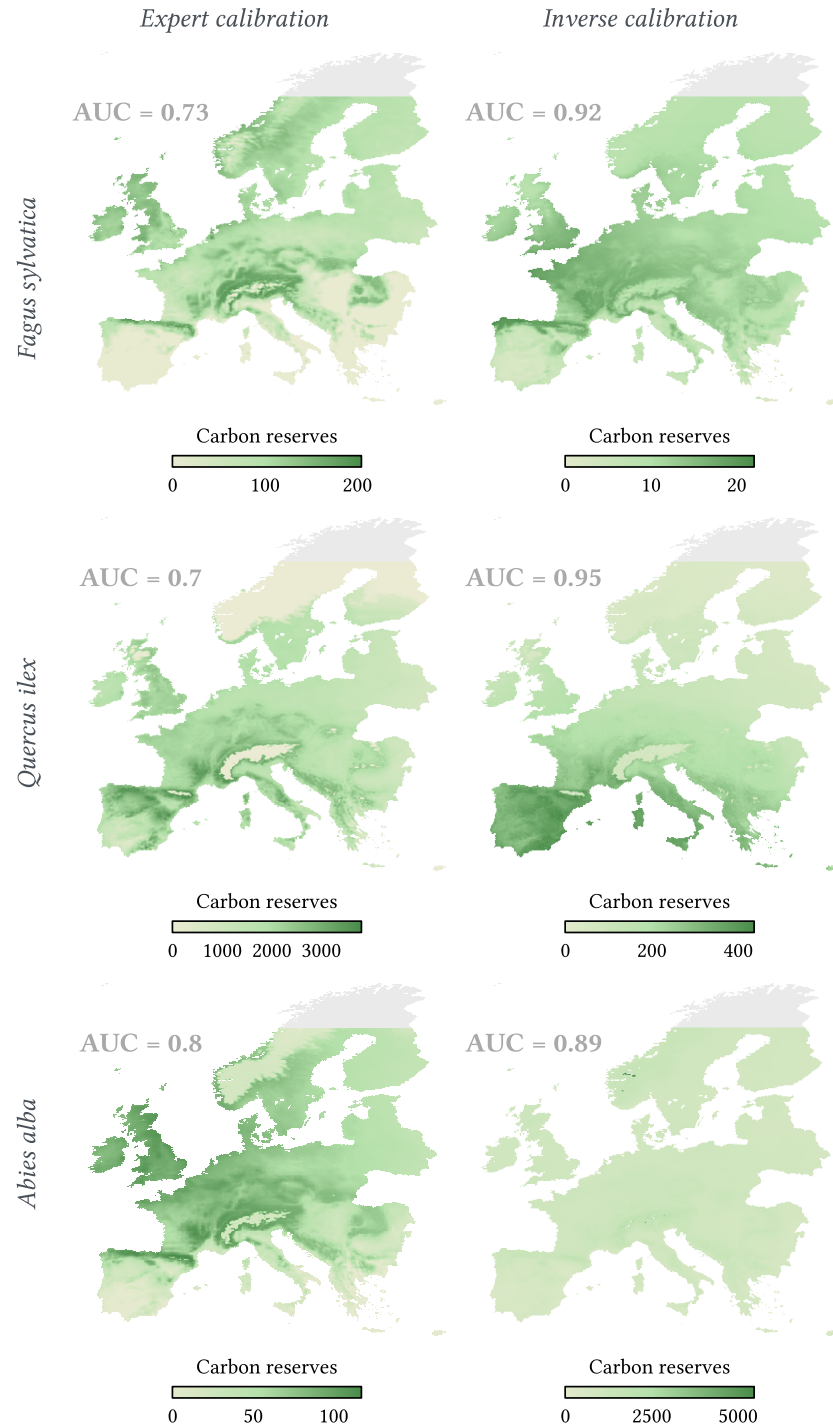


Figure F.2: Carbon reserves predicted by CASTANEA with the expert and the inverse calibrations. Note that CASTANEA cannot be used in high-latitude regions (grey area).

Supplementary Appendix G: Leaf unfolding submodel

Fagus sylvatica leaf unfolding submodel

This model, called UniChill (Chuine 2000), is a sequential two-phase model (endodormancy and ecodormancy phases).

The endodormancy phase begins at day t_0 . The daily rate of chilling R_c is defined as a threshold function of the daily mean temperature T_d :

$$R_c(T_d) = \begin{cases} 0 & T_d \geq T_b \\ 1 & T_d < T_b \end{cases}$$

where T_b is the threshold temperature below which the bud accumulates chilling units.

The endodormancy releases at day t_c when the accumulated rate of chilling has reached the level C_{crit} :

$$\sum_{t_0}^{t_c} R_c(T_d) \geq C_{crit}$$

Then, the ecodormancy phase begins. The daily rate of forcing R_f is defined as a sigmoid function of the daily mean temperature T_d :

$$R_f(T_d) = \frac{1}{1 + e^{-d_T(T_d - T_{50})}}$$

where d_T is the slope and T_{50} the mid-response temperature. Bud break occurs at day t_f when the accumulated rate of forcing has reached the level F_{crit} :

$$\sum_{t_c}^{t_f} R_f(T_d) \geq F_{crit}$$

Thus, the UniChill model has 6 parameters: t_0 , T_b and C_{crit} for the first phase, d_T , T_{50} and F_{crit} for the second phase.

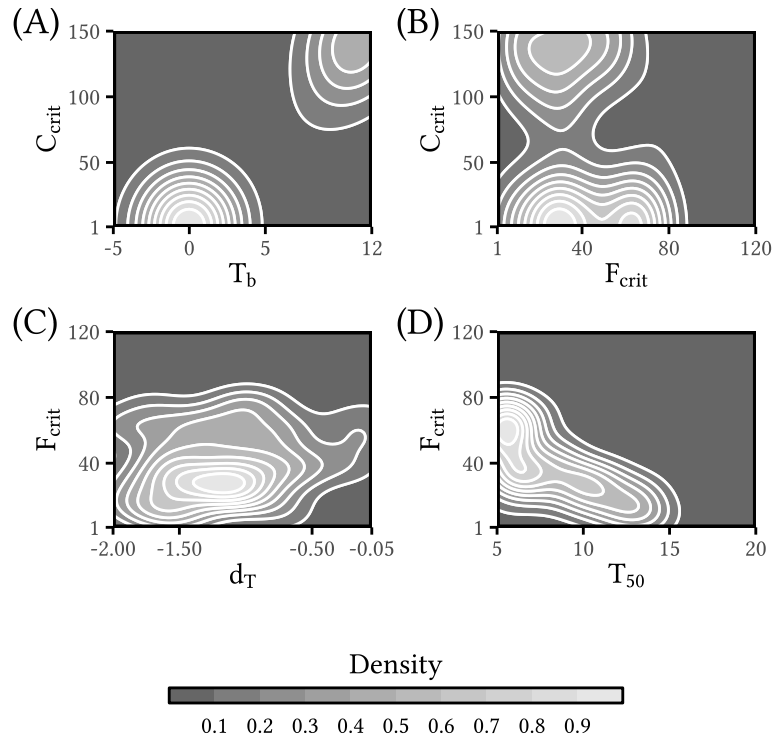


Figure G.1: Beech leaf unfolding model parameter density. Y-axis and X-axis limits are lower and upper bounds used during calibration.

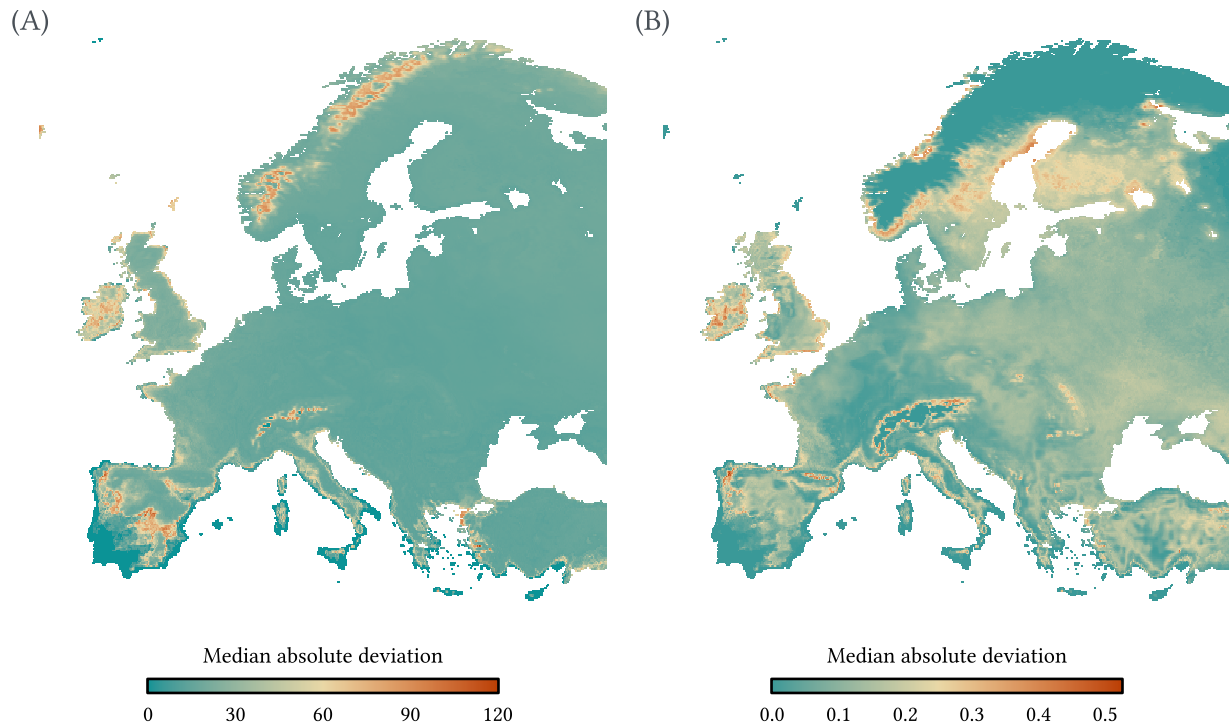


Figure G.2: Median absolute deviation of beech (A) leaf unfolding date and (B) fitness, predicted with 100 calibrated parameter sets of PHENOFIT.

The median standard deviation of unfolding date across Europe was about 15.4 days. On beech presence points, it was about 16.2 days. Nearly 90.3% of cells had a median absolute deviation lower than 30 days (Figure F.2.A.). The median standard deviation of fitness across Europe was about 0.148. On beech presence points, it was about 0.153. Nearly 46.4% and 91.5% of total cells had a median absolute deviation lower than 0.1 and 0.2 respectively (Figure F.2.B.).

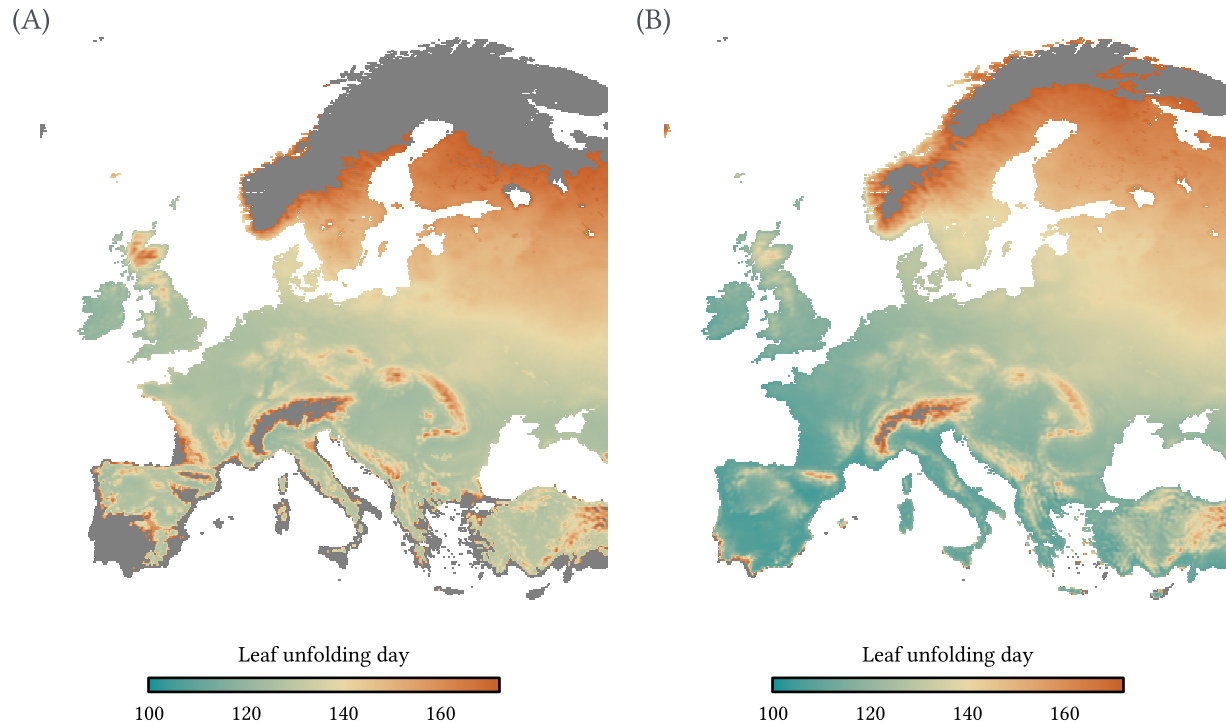


Figure G.3: Mean leaf unfolding day of beech with (A) best CMA-ES calibrated parameters and (B) expert parameters. Values above June solstice day (167) are in grey. Note that PHENOFIT model assign a value of 365 when unfolding has not happened at all due to climate conditions.