Can inverse calibration help improving process-explicit species distribution models?

Victor Van der Meersch[a,*], Isabelle Chuine[a]

[a] *CEFE, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France*

Abstract

Process-explicit models (PEMs) are expected to provide more reliable projections of species range shifts because they explicitly model biological mechanisms that govern species response to climate. However, this kind of approach requires detailed and diverse data, which are available only for few species. Inverse calibration has been identified as an avenue to help calibrating PEMs for many species—yet it is still unclear whether it can yield correctly specified parameters. Here, we explored the potential of such inverse calibration techniques to enhance the accuracy of PEMs. We calibrated a species distribution PEM which makes a strong focus on phenology and stress resistance using species distribution data, following two strategies: (i) calibrating all parameters simultaneously or (ii) focusing only on critical parameters. We then evaluated the realism of the parameter estimates obtained by comparing them to measurements and comparing simulated phenology to observed phenology across Europe. We showed that model structure alone do not sufficiently constrain the calibration process, which may produce unrealistic parameter values. However, using inverse calibration parsimoniously can improve model performance while still allowing to simulate realistic processes.

Keywords: calibration, process-based model, process-explicit model, mechanistic model, species distribution

*Corresponding author

1. Introduction

There have been repeated calls for focusing efforts on the development of process-explicit models (PEMs) in ecology (Urban et al., 2016; Singer et al., 2016; Pilowsky et al., 2022). Because they describe cause-to-effect relationships, they may provide more robust projections in novel climatic conditions (Van der Meersch et al., 2024). However, PEMs are still applicable to few species (Evans et al., 2016), and their widespread use would first require to address and to quantify their error and uncertainty.

We can distinguish two main source of errors in ecological models: the model's structure and the value of the parameters. The structure of a model is defined by its underlying assumptions and level of simplification. Our understanding of the regulation of the ecophysiological processes at the scale of organisms—that ultimately drive ecosystem dynamics—has grown substantially in the last decades. This has led to significant improvements in how we represent these processes in models, e.g. plant hydraulics (Ruffault et al., 2022) , although some remains challenging, e.g. bud dormancy (Chuine et al., 2016) or carbon allocation. These advancements enable researchers to explore complex interacting mechanisms, providing useful insights and tools for determining conservation and management strategies in a changing world (Urban et al., 2016). However, the incorporation of new detailed processes can become a trap (Franklin et al., 2020), as it often increases the number of poorly known parameters and limits the number of species for which these models can be applied to.

Process-explicit modeling depends a great deal on the calibration and the availability of data (Cabral et al., 2017). In a classical framework, parameters are typically determined from experiments, measurements, and expert knowledge. When it is too difficult or impossible, a cost-effective solution to the problem of estimating many parameters is the use of inverse modeling to bridge observations and simulations (Evans et al., 2016). This involves adjusting the parameters until the model outputs closely match the observed data—often using an optimization algorithm with an informal objective function, or a likelihood-based statistical method (Malchow and Hartig, 2024). Resulting parameter values are therefore conditional on the structure of the model and the observed data: the model structure explicitly constrains how climatic factors determine species performance, and the parameters and simulation outputs depend on the observations in a similar way to statistical models (Zhang et al., 2024). Inverse modeling is increasingly used in PEMs to infer model parameters or at least some of them. Process-related data can be used to fit some part of the models separately or sequentially, e.g. tree ring series have been used to calibrate a sapwood growth submodel (De Cáceres et al., 2023), and phenological records are frequently

used to calibrate phenological submodels (Chuine et al., 2013). Several data sources can also be combined to refine models (Benito Garzón et al., 2019) and to calibrate several processes simultaneously in a model-data fusion fashion (e.g. Trotsiuk et al., 2020). This allows to integrate data at multiple spatiotemporal scales (Hartig et al., 2012; Niu et al., 2014), and reduce potential overfitting issues (Bacour et al., 2023).

However, being able to reproduce observed data, even from several sources, does not guarantee convergence towards biologically sound parameter estimates. First, models are a simplification of reality for several reasons. They often target particular processes that are important for addressing specific questions within a specific spatiotemporal context, and some processes may be unknown or not completely understood. Whether intentional or not, this simplification may miss important processes (Forrester et al., 2021), and inverse calibration may thus lead to parameter values that compensate for these missing processes. Second, even in a structurally correct model, measurements are not likely to coincide precisely with what the model simulates (Zhang et al., 2024). Third, some parameters might also be model-specific (i.e. conceptual), and not correspond to something observable nor measurable.

Process-explicit species distribution models have rarely used inverse modeling with species distribution data available across large scales (Higgins et al., 2012; Van der Meersch and Chuine, 2023), probably because this does not align well with the general idea of this modeling approach. Traditionally, these data have only been used by correlative niche models. Yet, the large and constantly increasing volume of such data (Feng et al., 2022) represents a vast amount of information that could help improve the accuracy of PEMs (Evans et al., 2016). While using only species occurrence data to infer the values of many parameters can be seen as a brute-force approach, such inverse calibrated PEMs may outperform correlative models and better reproduce the distributional patterns of the species on different spatio-temporal situations (Higgins et al., 2020; Van der Meersch et al., 2024). Moreover, such inverse calibration could also help making PEMs easier to apply to a greater number of species and on a larger scale (e.g. Conradi et al., 2024). However, a good model fit to observations, even long-term data series, does not necessarily imply a good estimation of the processes truly responsible for the observations.

Here, we investigate the differences between parameter estimates of a PEM obtained either with a classical calibration or with inverse calibration, and what it implies in terms of realism of the simulated processes and of model performance. More precisely, we seek to understand whether and to what extent inverse calibration using species occurrence (i) can lead to realistic parameter values and processes and (ii) may help us improving PEMs. To do this, we focus on PHENOFIT, a PEM which has been used to study species range determinants and to forecast species

distribution shifts with past and future climate change in North America and Europe (Morin et al., 2007; Saltré et al., 2013, 2015; Cheaib et al., 2012). We first examine in detail 100 calibrations obtained for European beech (*Fagus sylvatica* L.), and analyze the compensations between parameters and the realism of the simulated processes against observations all over Europe. We then investigate whether inverse calibration could improve parameter value estimation of some processes for which data are lacking for eight European forest tree species.

2. Materials and Methods

2.1. Process-explicit modeling with PHENOFIT

PHENOFIT is a process-explicit model developed for temperate tree species which has been used to project their distributions using climatic conditions. It estimates the probability of presence of an adult tree (i.e. fitness), defined as the product of the probability to survive frost and drought events (survival), the proportion of flowers/fruits not killed by frost events (fruit index), and the probability that fruits reach full ripening (maturation index) at a yearly time step (Figure S1).

Each phenological event (leaf unfolding, flowering, fruit maturation and leaf senescence) is simulated with daily climate forcing, and the model assumes that a tree species range depends mainly on the synchronization of its timing of development to the local abiotic conditions and especially the occurrence of some abiotic stresses. Thus, for example, the fitness can be reduced when a severe drought event occurs between budburst and leaf senescence, or when a substantial proportions of leaves and flowers that take part to the development of the fruits are killed by frost. In the following, we will provide a more detailed description of three important processes of PHENOFIT, which will be discussed further in this article. A precise description of the submodels and the response functions can also be found in Supplementary Material for beech.

2.1.1. Leaf unfolding and flowering submodels

Dates of leaf unfolding (date at which 50% of leaves are unfold) and flowering (date at which 50% of flowers are mature) are calculated using mechanistic phenology models (Chuine and Régnière, 2017). Organ development is represented by a state variable, S for development state, which is the integration of development rates (R) over time (in daily steps) from a start date t_0. The general structure of mechanistic phenology models for one specific development phase is the following: t_n such that $S_{n,t} = \sum_{t=t_{n-1}}^{t_n} R_{n,t} = S_n^*$ where n is a development phase, $S_{n,t}$ is the state of development on day t in phase n; t_n is the end of phase n and t_{n-1} the end of phase $n-1$,

$R_{n,t}$ is the rate of development during phase n on day t which is a function of temperature, and S_n^* is the critical state required to reach t_n. Several phases of development can be modeled in a single model composed of several submodels, each one describing a specific phase such as dormancy induction, endodormancy, ecodormancy, etc. In such case, phases either follow each other sequentially or can overlap depending on the development phase and the species. Leaf unfolding and flowering dates have been modelled here using 2-phases models describing both the bud endodormancy phase (bud remain dormant despite meteorological conditions that could sustain cell growth) and the ecodormancy phase (bud cell growth depends on the meteorological conditions). Development rates (R) are response functions to daily temperature that can be linear or nonlinear (usually with a single optimum), and vary with the species and the development phase.

2.1.2. Frost hardiness submodel

Frost hardiness of vegetative and reproductive organs is modelled according to Leinonen (1996), as a function of the additive effect of photoperiod and temperature, and depends on the state of development (or phenological state): hardiness varies dynamically between a maximum value reached during bud winter dormancy and a minimum value reached at bud break and flowering. The model has several parameters, including the minimum level of frost hardiness (FH_{min}) and the maximum increase of frost hardiness (ΔFH_{max}) conveyed both by photoperiod and temperature. The maximum potential level of frost hardiness is thus $FH_{min} + \Delta FH_{max}$ (see Supplementary Material for a detailed description of the model).

2.1.3. Fruit maturation submodel

Fruits development is modeled with a two-phases mechanistic phenology model: a first phase of cell multiplication and growth, and a second phase of photosynthetic assimilate accumulation. This second phase is the most important, and depends on several components including the proportion of leaves that resisted frost, water availability and photosynthetic activity. The latter varies according to temperature conditions following the unimodal function of Wang and Engel (1998) which involves an optimal temperature parameter T_{opt}. Fruit maturation is integrated over time at the scale of the tree crown, following a normal distribution. The mean date of fruit maturation corresponds to the date when fruit development reaches Mat_{moy}, and correspond to the stage 50% of fruits are ripe, and fruit maturation starts at Mat_{moy} -3σ and ends at Mat_{moy} $+3\sigma$.

2.1.4. Leaf senescence submodel

Leaf senescence date (date at which 50% of leaves have changed color or have fallen) is modelled using a 1-phase mechanistic phenology model with different response functions to temperature and to day length depending on the species (Delpierre et al., 2009).

2.2. Model calibration

2.2.1. Expert calibration

PHENOFIT has been calibrated for several European tree species, and validated by comparing their historical and Holocene distribution to the simulated fitness used as a proxy of species probability of presence (Saltré et al., 2013; Duputié et al., 2015; Gauzere et al., 2020; Van der Meersch et al., 2024). Some parameters were directly measured or found in the literature, e.g. the frost hardiness submodel parameters. Phenology submodel parameters were inferred by inverse modelling using phenological data across Europe (provided through the TEMPO data portal `data.pheno.fr`, and the PEP725 database `pep725.eu`). This expert calibration of the model does not involve species occurrence data at any point.

2.2.2. Inverse calibration with CMA-ES algorithm and species occurrence data

Following Van der Meersch and Chuine (2023), we calibrated PHENOFIT using the covariance matrix adaptation evolution strategy (CMA-ES), which is a robust algorithm for complex optimization problems (Hansen and Ostermeier, 2001). It is inspired by Darwin's theory of evolution to find the most fit parameter sets. We ran the CMA-ES calibration on the multicore cluster GenOuest (`genouest.org`).

The objective function for the calibration was the area under the receiver operating characteristic curve (AUC), to maximize model discriminating capacity (i.e. potential to distinguish between species presences and absences). To compute the AUC, we used the same occurrence data as in Van der Meersch and Chuine (2023). These were mainly extracted from the EU-Forest dataset (Mauri et al., 2017), completed with presence records extracted from the Global Biodiversity Information Facility (`gbif.org`) to account for tree occurrences outside forests. We removed GBIF occurrences outside natural species ranges as defined by Atlas Flora Europeae (Jalas and Suominen, 1972–2005) and EuroVegMap (Bohn et al., 2003). The EU-Forest cells where the species is not reported present were considered as (pseudo-)absences. In order to reduce calibration computational costs, we selected subsets of 1000 presences and 1000 absences. Presences were sampled based on a k-means clustering to make sure that all species environmental preferences were proportionally represented (see Van der Meersch and Chuine (2023) for details).

First, we ran one hundred calibrations for *Fagus sylvatica*, using species occurrences and the same parameter bounds as in Van der Meersch and Chuine (2023) in order to remain in realistic parameter ranges. These calibrations are called *full* inverse calibrations in the following (all parameters optimized at once, Figure S2). Second, we ran a second set of inverse calibrations, for eight different species (*Abies alba*, *Betula pendula*, *Fagus sylvatica*, *Fraxinus excelsior*, *Larix decidua*, *Picea abies*, *Quercus pubescens* and *Quercus robur*), to optimize only a subset of parameters. These parameters corresponded to processes that we identified as responsible for false absence errors in the predictions of the expert calibration version of the model—i.e. cases where simulated fitness was low despite the species being present. To identify these processes, we calculated the relative contribution of the three sub-components of fitness: survival (S), fruit index (Fr) and maturation index (M) to the simulated fitness (F). The relative contribution of a factor was calculated as the ratio of the product of the two other factors over the total sum of all possible two-factor products. For example, the contribution of survival S was calculated as: $\frac{Fr*M}{S*Fr+S*M+Fr*M}$. Other parameters—corresponding to processes that we did not identified as responsible for false absence errors—were fixed at the expert values. For each species, we ran 5 calibrations on 2 different occurrence subsets (i.e. 10 repetitions). These calibrations are called *partial* inverse calibrations in the following.

2.2.3. Climate and soil data used to calibrate and run the model

Simulations and calibrations were run with climate variables extracted from the ERA5-Land hourly dataset (Muñoz-Sabater et al., 2021), for the period 1970-2000. As in Van der Meersch and Chuine (2023), we computed daily mean values of several variables: temperatures (minimum, mean and maximum), dewpoint temperature, precipitation, global radiation and wind speed. Daily potential evapotranspiration was calculated using the standard FAO Penman–Monteith equation (Allen et al., 1998). Soil water holding capacity was calculated with the field capacity and wilting point data from EU-SoilHydroGrids (Tóth et al., 2017) and the percentage of coarse fragments from SoilGrids250m (Hengl et al., 2017).

To assess the similarity between the different calibrations in climatic conditions that differ significantly from those used for calibration, we also ran paleosimulations, using the same climatic forcing as in Van der Meersch et al. (2024). For this, daily data were generated with the GWGEN wheather generator (Sommer and Kaplan, 2017), from the monthly simulations of HadCM3B-M2.1 coupled general circulation model (Armstrong et al., 2019).

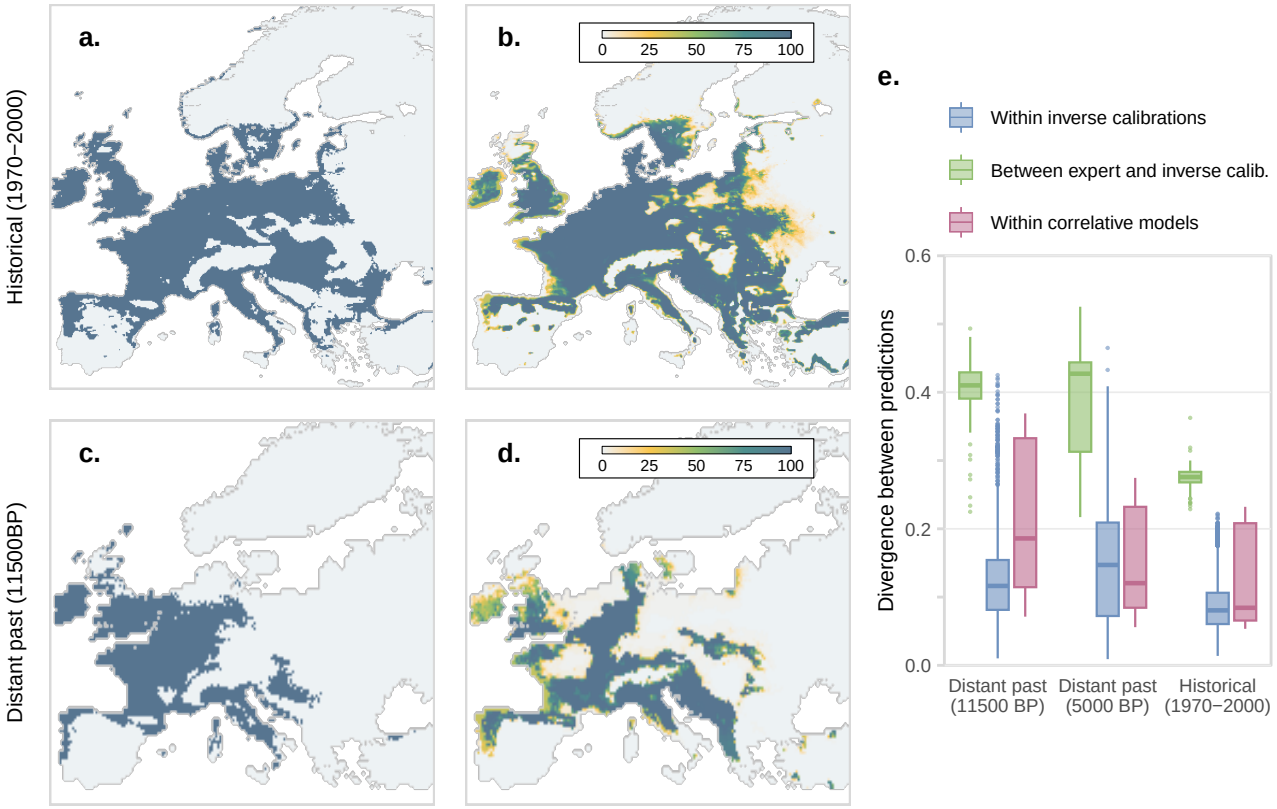

Figure 1: Simulated potential presence of *F. sylvatica* with **(a,c)** the expert parametrization and **(b,d)** the set of 100 *full* inverse calibrations, in the historical climatic conditions **(a,b)** and in the paleoclimatic conditions **(c,d)**. Simulated fitness is converted to presence/absence (blue/grey) using the optimal threshold that maximizes the true skill statistic (TSS). **(e)** Sørensen dissimilarity between inverse calibrations, and between expert calibration and inverse calibrations. Sørensen dissimilarity between 5 different correlative species distribution models (from Van der Meersch et al., 2024) is shown for comparison. BP stand for "Before Present" (i.e. 1950).

2.3. Parameter estimates' evaluation

We partitioned the 100 *full* inverse calibrations of *Fagus sylvatica* using two k-means clustering procedures in a row (Figure 2a). The first clustering was based on the simulated leaf dormancy break and leafout dates. Then, within each cluster, the second clustering was computed based on the fruit maturation and leaf senescence dates.

In order to verify that the parameter values after calibration lead to realistic processes, we calculated root-mean-square errors (RMSE) between simulated phenological dates and observed dates. The latter were extracted from two databases, PEP725 and TEMPO, covering the period 1970-2000, essentially in Central Europe (Figure S3). For *Fagus sylvatica*, we used 59484 observations of leafout, 10449 of flowering, 23606 of fruit maturation and 71469 of leaf senescence (Figure 3). Moreover, we used 16329 obs. of *Betula pendula* flowering, 21813 obs. of *Picea abies* flowering and 49236 of *Quercus robur* fruit maturation (Figure 5).

3. Results

3.1. Coherent and stable predictions despite process discrepancies

The similarity between the simulated ranges of European beech obtained with the 100 *full* inverse calibrations was relatively stable over the last 12k years. The disagreements were mostly at the margins of the distribution (Figures 1b and 1d). The Sørensen dissimilarity between the inverse calibration projections was lower in historical conditions (median = 0.0805 [IQR = 0.0605 - 0.106]) than in the distant past, e.g. 11500 years ago (0.116 [0.0813 - 0.154], Figure 1e). However it increased relatively little while moving back to the past, i.e. to more dissimilar climatic conditions, compared to the dissimilarity between expert calibration and inverse calibrations, from 0.276 [0.268 - 0.283] in the present to 0.410 [0.391 - 0.429] at 11500 BP.

The relatively similar projections of the 100 *full* inverse calibrations, however, resulted from simulated biological processes that differed substantially (Figure 2). Regarding bud development, 65 calibrations (clusters blue and green, Figure 2a) had a short endodormancy phase (ending in November or December of the previous year, Figure 2b) and a longer ecodormancy phase (with an average duration of 187 days ± 56.9). On the contrary, the remainder of the 34 calibrations (clusters yellow and orange, Figure 2a) had a longer endodormancy phase (ending in February or March) and a shorter ecodormancy phase (with an average duration of 89.4 days ± 38.7). Note that one of the calibrations stood out due to both short endodormancy

and short ecodormancy (Figures 2b and 2c). Similarly, two behaviors can be identified in terms of fruit maturation: 83 calibrations (clusters green and yellow) led to early maturation (August/September, Figure 2d), and 16 led to late maturation (October/November, Figure 2d).

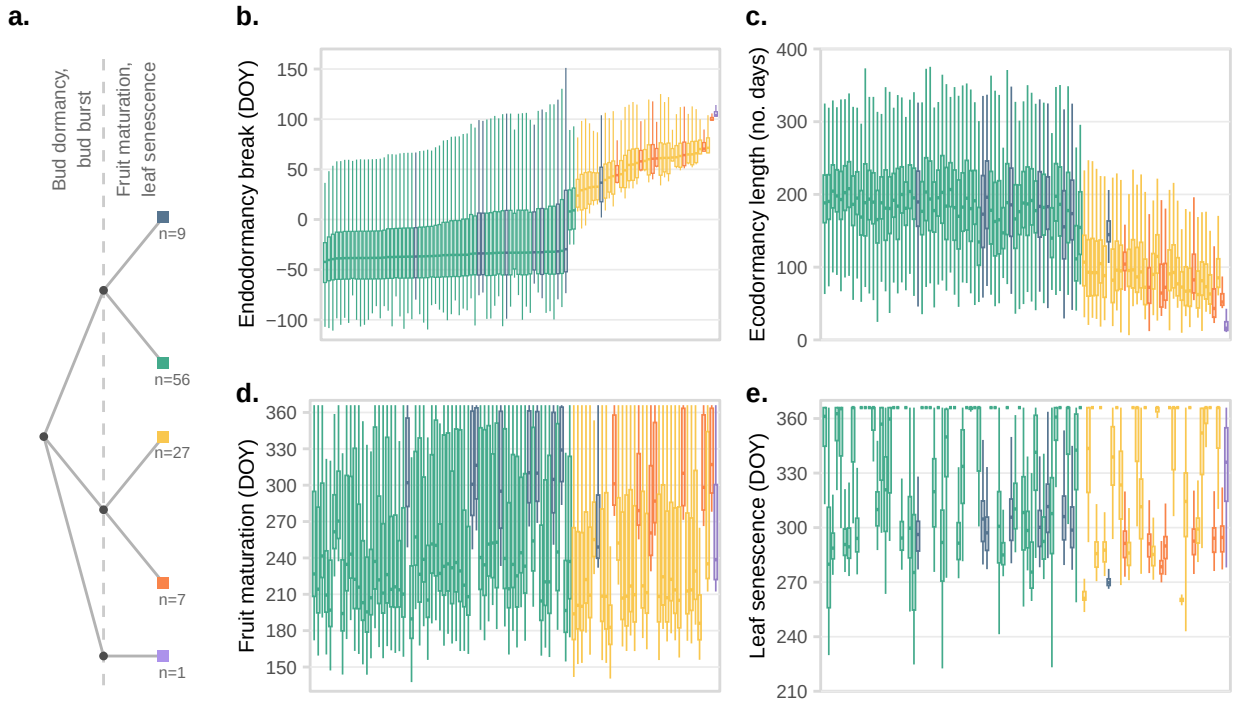

Figure 2: **(a)** Partition of the 100 *full* inverse calibrations for *Fagus sylvatica* after a two-step clustering. First clustering was based on the simulated leaf dormancy break and leafout dates. Second clustering was computed based on the fruit maturation and leaf senescence dates. **(b,c,d,e)** Simulated dates (day of the year, DOY) of **(b)** bud endodormancy, **(c)** bud ecodormancy, **(d)** fruit maturation and **(e)** leaf senescence. Colors correspond to the different clusters in **(a)**. Note that we do not show flowering as it occurs almost simultaneously with leaf unfolding.

3.2. Errors in the simulated processes

These strong discrepancies between simulated processes inevitably led to differences in terms of agreement with phenological observations across Europe (Figure 3). The first group of calibrations (short endodormancy/long ecodormancy) had a median error of 28.3 days for the budburst date, and up to 120 days for the worst calibration. The second group (long endodormancy/short ecodormancy) had a smaller

median error of 16.3 days [7.73 - 27.4]. Both performed worse than the expert version of the model (6.95 days [3.36 - 12.1], Figure 3a) whose parameters were calibrated using some of these observations. Regarding fruit maturation, the late-maturation cluster (blue and orange) got a median error of 17 days [8 - 29], closer to the expert version (13 days [6 - 24]). Moreover, inverse calibrations showed almost no year where fruit maturation could not occur unlike the expert version, whose parameters did not allow for ripe fruits in more than 50% of the cases (Figure 3c). For leaf senescence, contrary to the expert version which always predicted a senescence date, 43% of the calibrations led to no senescence date for at least 50% of the cases (Figure 3d), and 34% to no senescence at all. These calibrations belonged exclusively to the early-maturation cluster (green and yellow), whereas the late-maturation cluster got a median error of senescence date of 14.6 days [8.75 - 22], slightly higher than the expert version (10.8 days [6 - 16]).

3.3. Partial calibrations on selected parameters

The previous results showed significant differences between the calibrations in terms of simulated processes and of agreements with the available observations. However, some of the calibrations simulated processes that were relatively consistent with observations, and this was confirmed in the following when attempting to inversely calibrate the processes causing false absence errors with the expert version of the model. Most partial calibrations resulted in an AUC higher than 0.8 (Figure 4), and in the best-case scenario an increase of more than 0.4 for spruce *(Picea abies*), where we transitioned from a model worst than random to one with a good discriminatory ability.

The lack of data, particularly for fruit maturation, prevented us from checking the realism of partial calibrations for all species. We were able to evaluate the inverse calibrations of the fruit maturation date submodel for beech (Fagus sylvatica) and pedunculate oak (*Quercus robur*) (Figure 5a), of the flowering date submodels for birch (Betula pendula) and spruce (Picea abies) (Figure 5b), and the frost hardiness submodel for fir (Abies alba), beech and birch (Figure 5c). For beech, most calibrated parameter sets converged towards a median error of the fruit maturation date ranging from 11 to 16 days (except one with a median error of 24 days), close to the expert version error of 13 days (Figure 5a). Similarly to the full calibrations, they also corrected the non-fruit maturation issue of the expert version. For pedunculate oak , half of the calibrations led to errors in the fruit maturation date similar to the expert version, around 16 days, and only 9.67 days [4.33 - 17] for the best parameter set (Figure 5a). Regarding the flowering date of birch, 2 out of 10 partial calibrations resulted in a lower median error (5.91 days [2.55 - 11.6] and 5.91 days [2.72 - 11.4])

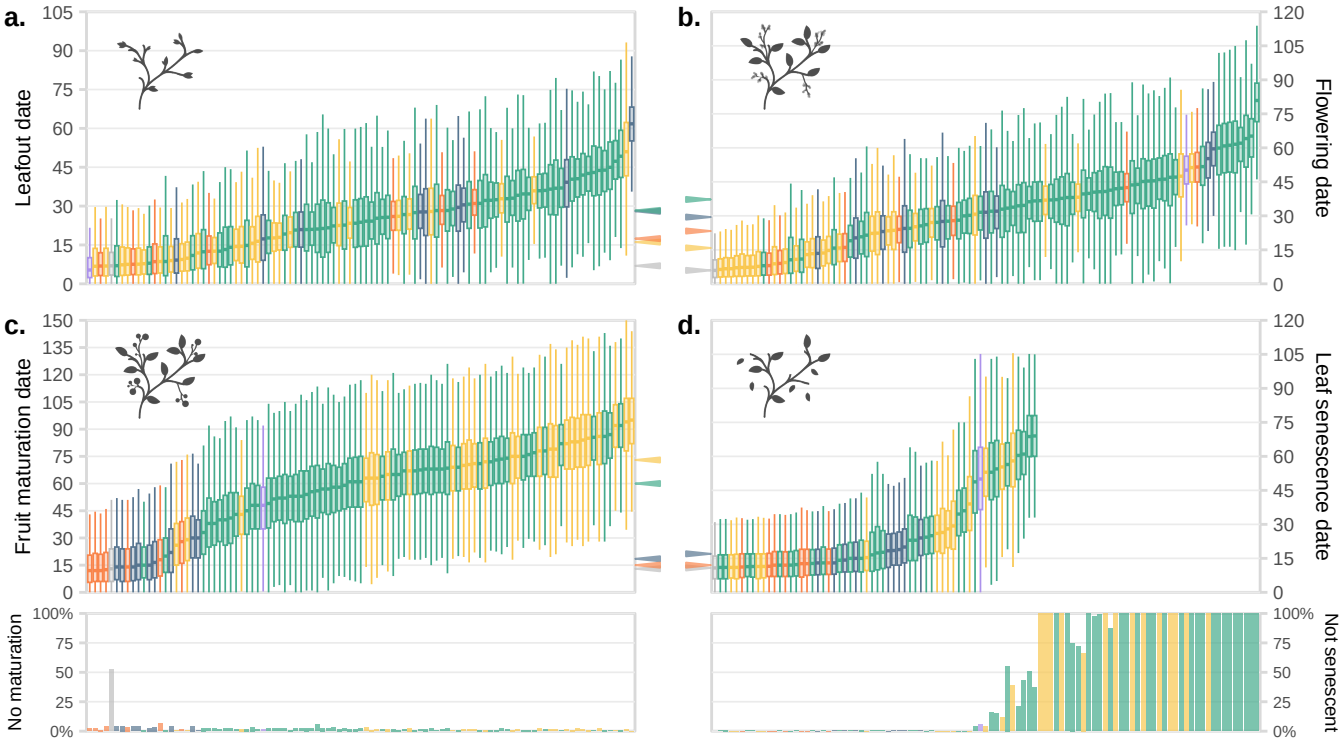

Figure 3: Root-mean-square error (RMSE) between phenological dates that were observed for *Fagus sylvatica* between 1970 and 2000 and those that were predicted by the 100 *full* inverse calibrations, for **(a)** leaf unfolding, **(b)** flowering, **(c)** fruit maturation and **(d)** leaf senescence. For fruit maturation and leaf senescence, the lower panels show the percentage of cases (year x site) for which the parameter sets predict that the event does not occur (i.e. no maturation or no leaf senescence). Colors are based on the clustering shown in Fig. 2, grey corresponds to the expert parameter set. Arrows in the middle indicate the median RMSE for each group.

than the expert parameter set (8.87 days [4.21 - 14.9]). However, some had a much higher median error, up to 50.2 days (Figure 5b). For spruce, the best calibration in terms of RMSE simulated no flowering in 37.3% of the cases. Apart from this one, the other 9 calibrations had a higher median error (between 18.3 days [10.3 - 27.9] and 29.7 days [20.0 - 41.7]) than the expert version of the model (11.5 [5.68 - 18.9]). Finally, the frost resistance submodel parameter estimates were very close to the expert version value based on the literature in only one third of the cases (Figure 5c).

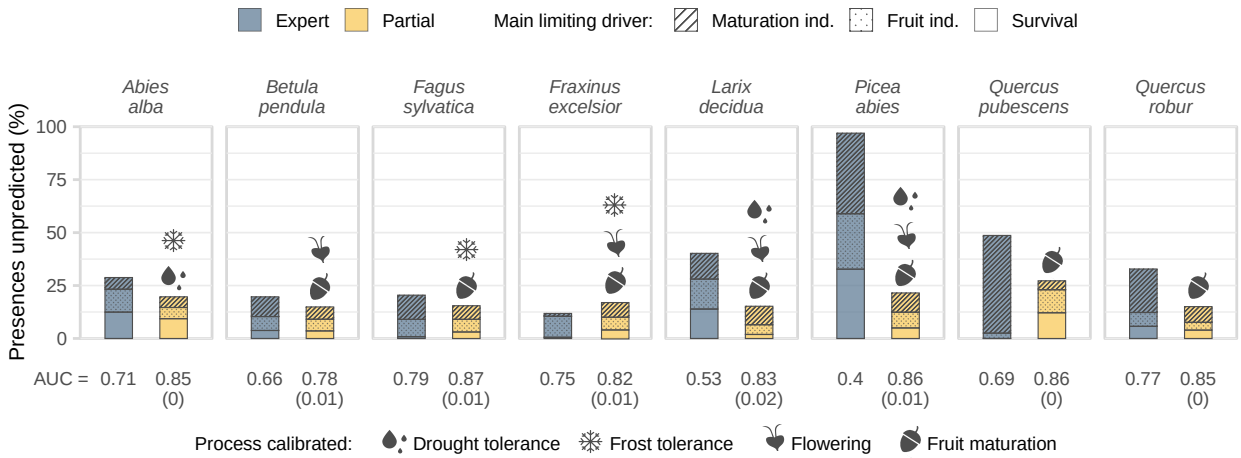

Figure 4: Results of the *partial* calibrations for the eight species considered, where only some of the parameters were optimized (pictograms indicate which processes have been recalibrated using species distribution data). The y-axis shows the percentage of observed presences that are predicted as absences by the model (i.e. *false absences*), and the bar patterns represent the main simulated processes explaining these errors. The bottom row shows the average AUC (a classic discrimination performance metric) and its standard deviation in parenthesis.

4. Discussion

Our results show that inverse calibration of process-explicit models (PEMs) using species distribution data does not necessarily lead to realistic parameters and simulated processes. They further demonstrate that inverse calibration can be highly useful as a model diagnostic tool, and, more importantly, to calibrate selected parts of a model where specific data and expert knowledge are lacking, albeit important cautions.

4.1. Inverse calibration can lead to accurate species range predictions despite unrealistic parameter estimates

Few parameter sets outperformed the expert calibrations used for this study (especially when few data are available for the expert calibration), but most resulted in larger errors when predicting the dates of the phenological events—up to 2 months in the worst case (Figure 3). For some other processes, inverse calibration induced even unrealistic functional traits. For example, a large portion of the calibrations considered beech as an evergreen species, or at least not senescent, in some regions (Figure 3d).

Despite similar predictions of species distribution (Figure 1b), we observed that simulated processes could strongly diverge between calibrated models. Some calibrated models led for example to a short endodormancy phase, followed by a longer ecodormancy phase, and *vice-versa* (Figures 2a and 2b). This arises because the compensation between these two processes has little effect on the functional trait they regulate, the leaf unfolding date, and thus on the predicted distributions. The non-identifiability of parameter values obtained with inverse calibration—i.e. when different sets of parameters may result in equivalent model outputs, also called equifinality—is a known issue (He et al., 2017; Cameron et al., 2022; Van der Meersch and Chuine, 2023; Malchow and Hartig, 2024). Here we show that compensations can occur between components of a same process or between different processes. For example, errors between simulated and observed leafout dates vary greatly across calibrated models (Figure 3a), suggesting that these errors are compensated by another process—for example, leaf frost hardiness or fruit development.

Unexpectedly, these discrepancies in the simulated individual processes did not cause a sharp increase in the dissimilarity between the predictions of the different calibrated models over the Holocene (Figures 1c, 1d and 1e), even in the very different climatic conditions of the Early Holocene (Figure S4). In other words, while inverse calibration led sometimes to very different parameter sets and thus very different "phenotypes" (early/late budburst, deciduous/evergreen...), the predictions nevertheless remained consistent across long time scales and novel climatic conditions. Therefore, the optimization algorithm manages to find consistently a similar relationship between climatic conditions and the higher-level model output (fitness), regardless of the divergent lower-level processes (frost hardiness, fruit development, etc.). The link between climate and species distribution as captured in the occurrence data thus seems to constrain the model's behavior without necessarily capturing the realistic underlying processes. In other words, the optimization algorithm appears to accommodate the model structure, making it flexible enough to fit the data well. However, this strength of the optimization algorithm is also its weakness:

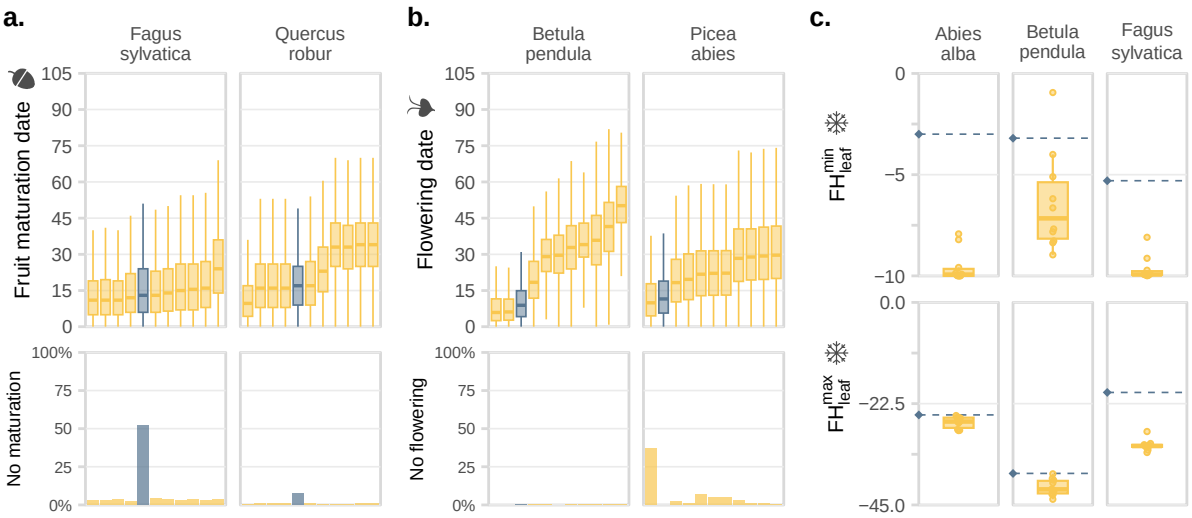

Figure 5: Root-mean-square error (RMSE) between phenological dates that were observed for the different species between 1970 and 2000 in Europe and those that were predicted by the partial calibrations (yellow) and the expert calibration (blue), for **(a)** fruit maturation and **(b)** flowering. **(c)** Distribution of the frost hardiness parameter estimates after partial calibration. The blue dotted line shows the value in the expert version of the model.

the model structure and the mathematical functions used to describe the processes are not sufficient to constrain the parameter estimates. Thus, contrarily to what is usually assumed (Higgins et al., 2020), here we find that PEMs are not necessarily less flexible than correlative models because of their structure. This highlights the importance of not focusing solely on some metrics of (apparent) performance, even in novel climatic conditions, but rather investigating the intermediate outputs of the model.

Finally, and similarly to correlative models (Barbet-Massin et al., 2010; Duputié et al., 2014), PEMs are affected by the bias in the occurrence data used for the calibration. For example, all simulations predict a low fitness of *Fagus sylvatica* in south-western France contrary to the expert model predictions (Figure 1b), whereas its absence is potentially more attributable to the legacy of forest management (vast maritime pine plantations) than to truly limiting climatic conditions as the presence of some old relict beech forests in the region suggests (de Lafontaine et al., 2014).

4.2. Inverse calibration can help identifying model limitations and opportunities for improvement

Our results reveal both the power and the shortcomings of inverse calibration in ecological modeling, and call for a careful use of the method when one aims at providing projections in very different conditions from the calibration conditions.

They further suggest that inverse calibration can be an effective diagnostic tool to improve PEMs, both in terms of model hypotheses and parametrization. Using species occurrence data to calibrate such models might be fruitful, if a detailed evaluation of the realism of the parameter estimates and the values of the functional traits (phenotypes) they produce is carried out.

Inverse calibration can help identify processes that are not accounted for, or incompletely accounted for in the model, even though they may be important in the context of the study. Reversely, inverse calibration can help identify processes taken into account in the model which have little impacts on the targeted outputs. For example, the discrepancies between leaf senescence date predictions across the different calibrations (Figures 2d and 3d) highlight that, in the model, leaf senescence date is more weakly constrained by the calibration data compared to other traits because it has more limited impact on fitness. More precisely, the senescence date affects fitness if it occurs before the end of fruit development but late senescence date does not affect fitness because there is no nutrient remobilization in the model that would confer an advantage in loosing leaves before the first frost event. In this example inverse calibrations indicate that it might be opportune to refine the model by adding nutrient remobilization.

Using inverse calibration with species occurrence data can further improve PEMs in several ways. First, it can help improving the modeling of processes that have a significant impact on the model outputs but for which we have limited measurements and observations. For example, for forest tree species, observations of fruit maturation are often sparse and limited to some areas. It is therefore difficult to find the correct parameter values for the fruit maturation submodel, which can cause errors (e.g. beech, Figure 3c). Relaxing the fruit maturation submodel parameters in the expert version of the model for several species, and calibrating them using inverse calibration and species occurrence data (Figure 4), not only increased the model goodness-of-fit (i.e AUC)—as expected—but also resulted in more accurate predictions of fruit maturation dates on average (Figure 5a).

Second, as pointed by Harrison et al. (2021), the estimation of PEM parameters are sometimes *ad hoc* or rely on outdated data. Thus *partial* inverse calibration can also be an opportunity to identify the parameters where the disagreement between the expert value and the calibrated values is systematically significant. For example, the maximum frost resistance of beech buds during winter ($FH_{min} + \Delta FH_{max}$) is -25.3°C in the expert version of the model, while inverse calibration yields an average value of -41.4°C (Figure 5c), which is notably lower and suggests the need to reassess the expert value. Recent studies indeed indicate values closer to -32 to -37°C (Delaporte, 2015; Kreyling et al., 2014; Lenz et al., 2016; Baffoin et al., 2021; Charra-

Vaskou et al., 2012).

However, we should keep in mind that the parameter values inferred with such *partial* calibrations are conditioned by the rest of the fixed processes and by the structural errors and hypotheses of the model. For example, the minimum frost resistance of leaves at leaf unfolding, FH^{min}, converged towards an unrealistically low value of -10°C for both beech and fir (the lower bound constraining the optimization of this parameter, Figure 5c). The parameters of the leaf unfolding submodel might actually be not valid for the entire study area (e.g. existence of local adaptation of populations to climatic conditions; Kreyling et al., 2014; Lazic et al., 2024) and the calibration algorithm compensated too early leaf unfolding date with a higher frost resistance of leaves. Local adaptation could explain some variation in parameter values between expert and inversely calibrated versions of the model but not of this magnitude. More probably, this difference points to a weakness in the model, which does not account for possible different resistance to frost between a freshly unfold leaf and a mature leaf, which may have biased the estimation of FH^{min}. Parameter estimates obtained by inverse calibration using species distributions might thus also pinpoint potential improvement of the models.

4.3. Inverse calibration can provide a more comprehensive assessment of model uncertainties

In many applications of PEMs, parameter uncertainties are not considered and simulations are run with a single parameter sets (Niu et al., 2014; Lobell and Burke, 2010). Typically, in previous studies using PHENOFIT, all parameters had a chosen fixed value, and model deterministic outputs did not account for parameter uncertainties. To avoid being overconfident with individual model projections, multiple inverse calibrations could be used to generate an ensemble of model projections (Figure 1b). The spread of model outputs across the ensemble may then allow to assess the uncertainty associated with the parameter values (Simmonds et al., 2024), and to provide a more comprehensive range of projections rather than single deterministic outcomes.

However, these projections will always be contingent on model hypotheses and structure. The representation of processes in models may indeed represent a large source of uncertainty, as processes can usually be modeled with various equations determining their functional form (Keenan et al., 2011). One could test a set of alternative mathematical structure to quantify the impacts of this structural uncertainty (Huber et al., 2020). An efficient way could be to include them during the calibration procedure, i.e. by also optimizing similarly plausible process formulations rather than just the parameter values, or by using highly flexible functions where

the parameter values determine the function type.

Overall, inverse calibration using species distribution data show promise as a method to enhance specific model components in the absence of more precise data. However, process-explicit model structure and mathematical functions alone do not sufficiently constrain the calibration process. To avoid producing unrealistic parameter estimates, inverse calibration thus necessitates a careful application and a thorough evaluation to ensure realistic modeling outcomes.

A more robust application of inverse modeling would involve integrating diverse types of data simultaneously, in a multi-objective fashion, to leverage complementary information from various sources (Cameron et al., 2022). In the near future, the increasing availability of high-resolution data—such as remote sensing and LiDAR—combined with an effective inverse-calibration framework will probably enable a better integration of large-scale data and ultimately a more accurate model parametrization. This could even pave the way for real-time continuous data integration and model recalibration, transforming ecological models into digital twins (Koning et al., 2023).

Appendix A. PHENOFIT processes and parameters

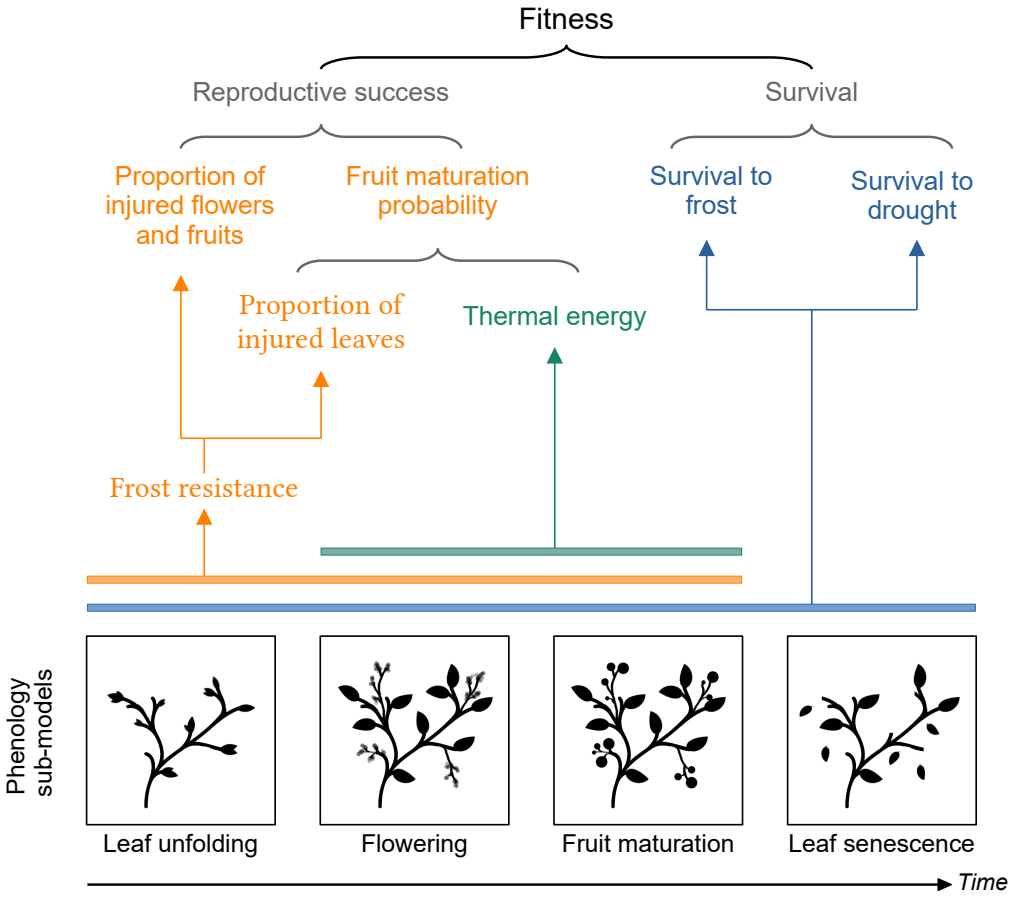

Figure A.6: **Schematic description of PHENOFIT.**

Table A.1: **Description of PHENOFIT parameters for *Fagus sylvatica*.**

Process	Subprocess	Function	Parameter	Definition
Leaf unfolding	Endodormancy	Threshold inferior	t_0	Starting date of the endodormancy phase
			T_b	Threshold temperature of the Threshold inferior function
			C_{crit}	Critical state of development corresponding to dormancy break
	Ecodormancy	Sigmoid	d_T	Slope at the inflection point of the sigmoid function
			T_{50}	Mid-responsee temperature of the sigmoid function
			F_{crit}^{leaf}	Critical state of development corresponding to leaf unfolding
Flowering	Endodormancy	Threshold inferior	t_0	Starting date of the endodormancy phase
			T_b	Threshold temperature of the sigmoid function
			C_{crit}	Critical state of development corresponding to dormancy break
	Ecodormancy	Sigmoid	d_T	Slope at the inflection point of the sigmoid function
			T_{50}	Mid-response temperature of the sigmoid function
			F_{crit}^{flower}	Critical state of development corresponding to flowering
Fruit maturation	Cell multiplication and growth	Sigmoid	aa	Slope at the inflection point of the sigmoid function
			bb	Mid-response temperature of the sigmoid function
			F_{crit}^{fruit}	Accumulation threshold above which accumulation starts
	Photosynthetic assimilate accumulation	Wang and Engel (1998)	T_{opt}	Optimal temperature of the Wang function
			Mat_{moy}	Accumulation threshold above which maturation occurs
			σ	Standard deviation
			pfe_{50}	Prop. of injured leaves reducing by 50% the flux of photosynthetic assimilates
Leaf senescence		Delpierre et al. (2009)	P_b	Photoperiod threshold below which photothermal accumulation starts
			T_b	Max. temperature above which there no photothermal accumulation
			α	Effect of temperature on photothermal accumulation α
			β	Effect of photoperiod on photothermal accumulation β
			S_{crit}	Accumulation threshold above which senescence occurs
			σ^{senes}	Standard deviation
Frost hardening	Leaf bud	Leinonen (1996)	FH_{min}^{leaf}	Minimum level of frost hardiness
			$\Delta FHT_{max}^{leaf}/\Delta FHP_{max}^{leaf}$	Max. level of increase of frost hardiness induced by temperature/photoperiod
	Flower bud		FH_{min}^{flower}	Minimum level of frost hardiness
			$\Delta FHT_{max}^{flower}$, $\Delta FHP_{max}^{flower}$	Max. level of increase of frost hardiness induced by temperature/photoperiod
	Thermal control		Te_1	Upper limit of the effective range of temperature
			Te_2	Lower limit of the effective range of temperature
	Photoperiod control		NL_1	Lower limit of the effective range of photoperiod
			NL_2	Upper limit of the effective range of photoperiod
	Fruit	Piecewise linear	FHfrmax1	Minimum level of frost hardiness
			FHfrmax2	Maximum level of frost hardiness
Water stress	Precipitation range		pp_{extlow}	1^{th} percentile
			pp_{low}	5^{th} percentile
			pp_{high}	95^{th} percentile
			$pp_{exthigh}$	99^{th} percentile

Appendix B. Leinonen frost hardiness model

The model developed by Leinonen (1996) predicts the frost hardiness of tree buds, i.e. their ability to withstand frost, and frost damages. The model uses daily temperature, photoperiod, and bud development state (phenological state) as the drivers of bud hardening and dehardening. In particular, a tree frost hardiness undergoes acclimation in response to two factors : decreasing temperatures and shorter days. The damage caused by freezing thus depends on the minimum temperature of day t and the actual frost hardiness FH_t reach on day t:

$$FH_t = FH_{min} + CR_t * (\Delta FHT_t + \Delta FHP_t) \tag{B.1}$$

FH_{min} is the minimum level of frost hardening of a bud expressed in degree celsius and it corresponds to the level after bud break when leaves have just unfold (see table above), CR_t is the hardening competence (which depends on the phenological state provided by the phenology model), ΔFHT_t is the increase of frost hardiness induced by temperature and ΔFHP_t is the increase of frost hardiness induced by photoperiod.

Exposure to cold temperatures. The temperature component ΔFHT_t is modelled by a piece-wise linear function depending on the daily minimum temperature T_t:

$$\Delta FHT_t = \begin{cases} \Delta FHT_{max} & \text{if } T_t < Te_2 \\ \Delta FHT_{max} * (1 - \frac{T_t - Te_2}{Te_1 - Te_2}) & \text{if } Te_2 \leq T_t \leq Te_1 \\ 0 & \text{if } T_t > Te_1 \end{cases} \tag{B.2}$$

ΔFHT_{max}, Te_2, and Te_1 are described in the table above. ΔFHT_{max} correspond to the maximum level of increase in frost hardiness that temperature can induce when the bud is fully dormant during winter.

Photoperiod sensitivity. The photoperiod component ΔFHP_t is modelled by a piece-wise linear function depending on the daily night length NL_t:

$$\Delta FHP_t = \begin{cases} \Delta FHP_{max} & \text{if } NL_t > NL_2 \\ \Delta FHP_{max} * \frac{NL_t - NL_1}{NL_2 - NL_1} & \text{if } NL_1 \leq NL_t \leq NL_2 \\ 0 & \text{if } NL_t < NL_1 \end{cases} \tag{B.3}$$

ΔFHP_{max}, NL_2, and NL_1 are described in the table above. ΔFHP_{max} correspond to the maximum level of increase in frost hardiness that photoperiod can induce when the bud is fully dormant during winter. $FH_{min} + \Delta FHT_{max} + \Delta FHP_{max}$ corresponds to the maximum level of frost hardiness that the bud can reach when it is fully dormant during winter.

Appendix C. Fitted parameter values for *Fagus sylvatica*

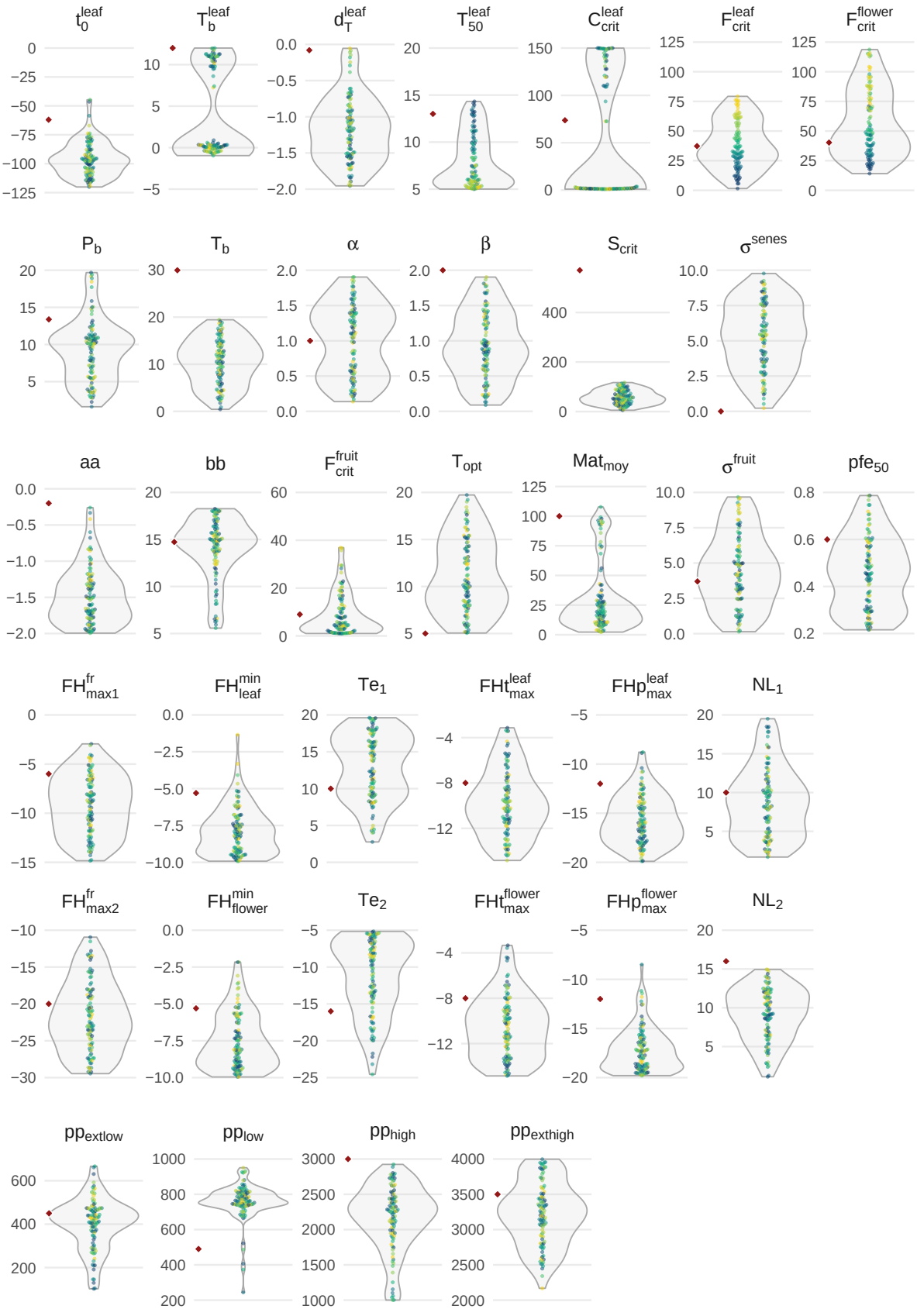

Figure C.7: **CMA-ES calibration of PHENOFIT parameters for *Fagus sylvatica*.** One panel per parameter. Y-axis limits are lower and upper bounds used during calibration. Each point is a calibrated parameter value and colour gradient represents the values of parameter F_{crit}^{leaf} . Red diamonds are parameter values used in the expert version of the model.

Appendix D. Phenological records in Europe

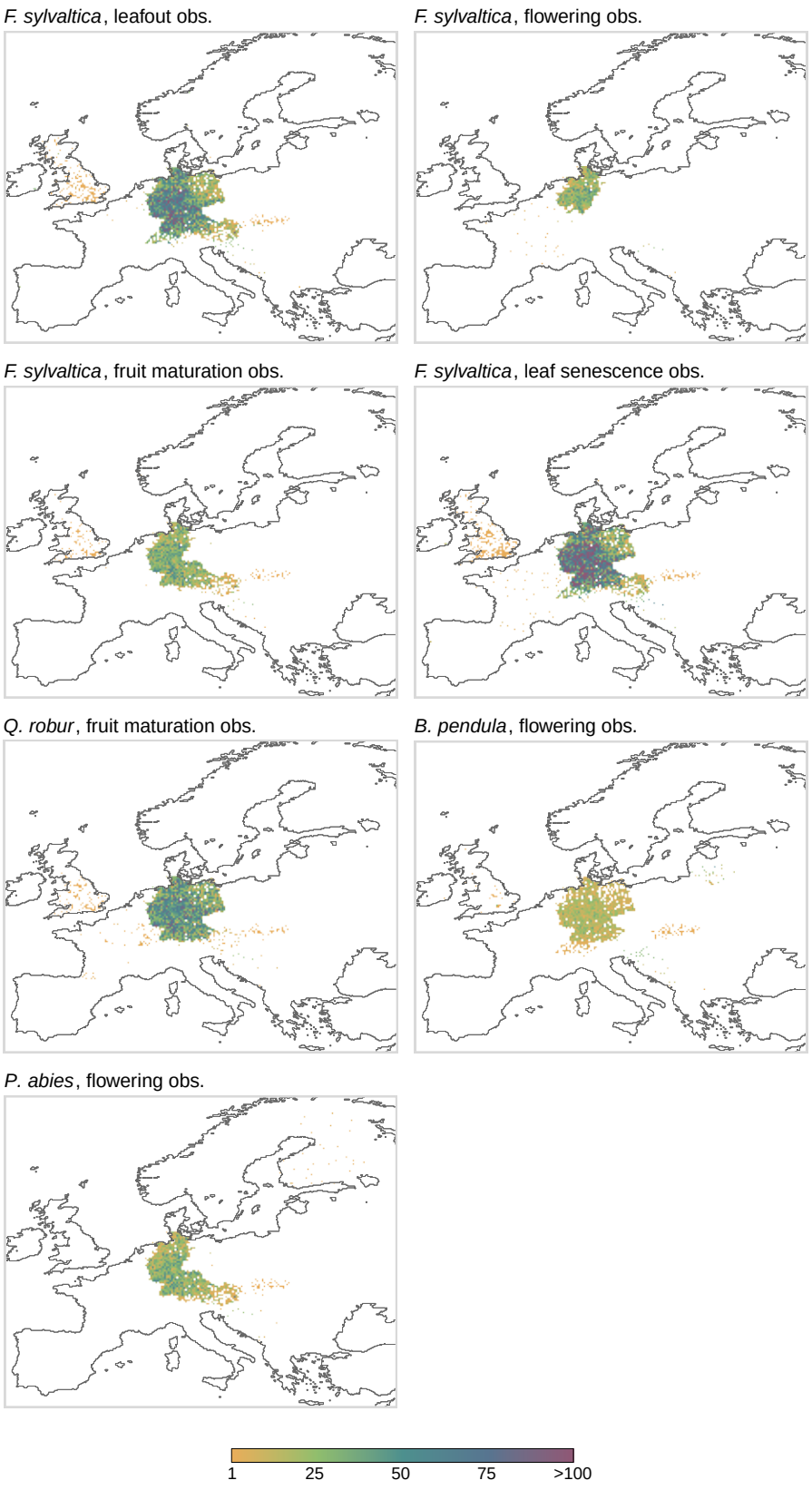

Figure D.8: **Observations used to calculate RMSE on simulated phenological dates.** These observations were extracted from the PEP725 (pep725.eu) and the TEMPO (data.pheno.fr) databases, for the period 1970-2000. Note that they were aggregated to 0.2° spatial resolution, to enhance the readability of the figure.

Appendix E. Climatic dissimilarity over the Holocene

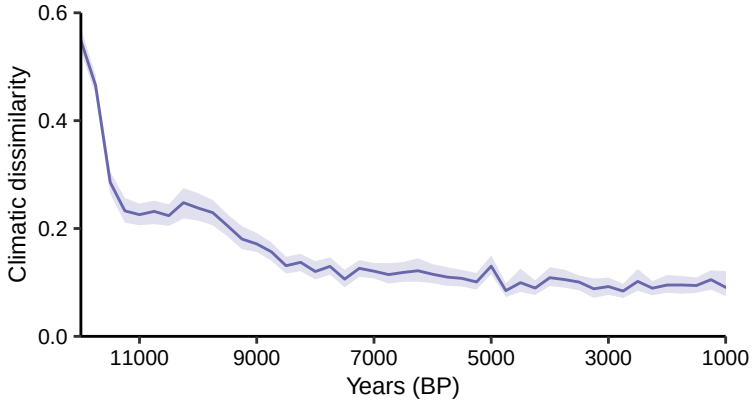

Figure E.9: **Climatic dissimilarity between past conditions and historical conditions.** The metric was calculated as the Sørensen dissimilarity between climatic hypervolumes (a metric of overlap in multidimensional space). See Van der Meersch et al. (2024) for details.

References

Allen, R., Pereira, L., Raes, D., Smith, M., 1998. Crop evapotranspiration: Guidelines for computing crop water requirements. FAO Irrigation and Drainage Papers 56.

Armstrong, E., Hopcroft, P.O., Valdes, P.J., 2019. A simulated Northern Hemisphere terrestrial climate dataset for the past 60,000 years. Scientific Data 6, 265. URL: https://www.nature.com/articles/s41597-019-0277-1, doi:10.1038/s41597-019-0277-1.

Bacour, C., MacBean, N., Chevallier, F., Léonard, S., Koffi, E.N., Peylin, P., 2023. Assimilation of multiple datasets results in large differences in regional- to global-scale nee and gpp budgets simulated by a terrestrial biosphere model. Biogeosciences 20, 1089–1111. URL: https://bg.copernicus.org/articles/20/1089/2023/, doi:10.5194/bg-20-1089-2023.

Baffoin, R., Charrier, G., Bouchardon, A.E., Bonhomme, M., Améglio, T., Lacointe, A., 2021. Seasonal changes in carbohydrates and water content predict dynamics of frost hardiness in various temperate tree species. Tree Physiology 41, 1583–1600. URL: https://doi.org/10.1093/treephys/tpab033, doi:10.1093/treephys/tpab033.

Barbet-Massin, M., Thuiller, W., Jiguet, F., 2010. How much do we overestimate future local extinction rates when restricting the range of occurrence data in climate suitability models? Ecography 33, 878–886. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0587.2010.06181.x, doi:10.1111/j.1600-0587.2010.06181.x.

Benito Garzón, M., Robson, T.M., Hampe, A., 2019. ?TraitSDMs: species distribution models that account for local adaptation and phenotypic plasticity. New Phytologist 222, 1757–1765. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.15716, doi:10.1111/nph.15716.

Bohn, U., Neuhäusl, R., Gollub, G., Hettwer, C., Neuhäuslová, Z., Raus, T., Schlüter, H., Weber, H., 2003. Map of the natural vegetation of europe - scale 1:2500000. URL: https://www.synbiosys.alterra.nl/eurovegmap/.

Cabral, J.S., Valente, L., Hartig, F., 2017. Mechanistic simulation models in macroecology and biogeography: state-of-art and prospects. Ecography 40, 267–280. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ecog.02480, doi:10.1111/ecog.02480.

Cameron, D., Hartig, F., Minnuno, F., Oberpriller, J., Reineking, B., Van Oijen, M., Dietze, M., 2022. Issues in calibrating models with multiple unbalanced constraints: the significance of systematic model and data errors. Methods in Ecology and Evolution 13, 2757–2770. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14002, doi:10.1111/2041-210X.14002.

Charra-Vaskou, K., Charrier, G., Wortemann, R., Beikircher, B., Cochard, H., Ameglio, T., Mayr, S., 2012. Drought and frost resistance of trees: a comparison of four species at different sites and altitudes. Annals of Forest Science 69, 325–333. URL: https://annforsci.biomedcentral.com/articles/10.1007/s13595-011-0160-5, doi:10.1007/s13595-011-0160-5.

Cheaib, A., Badeau, V., Boe, J., Chuine, I., Delire, C., Dufrêne, E., François, C., Gritti, E.S., Legay, M., Pagé, C., Thuiller, W., Viovy, N., Leadley, P., 2012. Climate change impacts on tree ranges: model intercomparison facilitates understanding and quantification of uncertainty. Ecology Letters 15, 533–544. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1461-0248.2012.01764.x, doi:10.1111/j.1461-0248.2012.01764.x.

Chuine, I., Bonhomme, M., Legave, J.M., García de Cortázar-Atauri, I., Charrier, G., Lacointe, A., Améglio, T., 2016. Can phenological models predict tree phenology accurately in the future? The unrevealed hurdle of endodormancy break. Global Change Biology 22, 3444–3460. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.13383, doi:10.1111/gcb.13383.

Chuine, I., de Cortazar-Atauri, I.G., Kramer, K., Hänninen, H., 2013. Plant Development Models. Springer Netherlands, Dordrecht. pp. 275–293. URL: https://doi.org/10.1007/978-94-007-6925-0$_1$5, doi:10.1007/978-94-007-6925-0$_1$5.

Chuine, I., Régnière, J., 2017. Process-Based Models of Phenology for Plants and Animals. Annual Review of Ecology, Evolution, and Systematics 48, 159–182. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-ecolsys-110316-022706, doi:10.1146/annurev-ecolsys-110316-022706.

Conradi, T., Eggli, U., Kreft, H., Schweiger, A.H., Weigelt, P., Higgins, S.I., 2024. Reassessment of the risks of climate change for terrestrial ecosystems. Nature Ecology & Evolution 8, 888–900. URL: https://www.nature.com/articles/s41559-024-02333-8, doi:10.1038/s41559-024-02333-8.

De Cáceres, M., Molowny-Horas, R., Cabon, A., Martinez-Vilalta, J., Mencuccini, M., Garcia-Valdés, R., Nadal-Sala, D., Sabaté, S., Martin-StPaul, N.,

Morin, X., D'Adamo, F., Batllori, E., Améztegui, A., 2023. MEDFATE 2.9.3: a trait-enabled model to simulate mediterranean forest function and dynamics at regional scales. Geoscientific Model Development 16, 3165–3201. URL: https://gmd.copernicus.org/articles/16/3165/2023/, doi:doi:10.5194/gmd-16-3165-2023.

Delaporte, A., 2015. Vers une compréhension fonctionnelle des dépérissements forestiers : étude du cas du hêtre (Fagus sylvatica L.) en forêt de Fontainebleau. Theses. Université Paris Sud - Paris XI. URL: https://theses.hal.science/tel-01216464.

Delpierre, N., Dufrêne, E., Soudani, K., Ulrich, E., Cecchini, S., Boé, J., François, C., 2009. Modelling interannual and spatial variability of leaf senescence for three deciduous tree species in France. Agricultural and Forest Meteorology 149, 938–948. URL: https://www.sciencedirect.com/science/article/pii/S0168192308003353, doi:10.1016/j.agrformet.2008.11.014.

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. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.12914, doi: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. doi:10.1111/geb.12118.

Evans, M.E.K., Merow, C., Record, S., McMahon, S.M., Enquist, B.J., 2016. Towards Process-based Range Modeling of Many Species. Trends in Ecology & Evolution 31, 860–871. URL: https://www-cell-com.inee.bib.cnrs.fr/trends/ecology-evolution/abstract/S0169-5347(16)30132-X, doi:10.1016/j.tree.2016.08.005.

Feng, X., Enquist, B.J., Park, D.S., Boyle, B., Breshears, D.D., Gallagher, R.V., Lien, A., Newman, E.A., Burger, J.R., Maitner, B.S., Merow, C., Li, Y., Huynh, K.M., Ernst, K., Baldwin, E., Foden, W., Hannah, L., Jørgensen, P.M., Kraft, N.J.B., Lovett, J.C., Marquet, P.A., McGill, B.J., Morueta-Holme, N., Neves, D.M., Núñez-Regueiro, M.M., Oliveira-Filho, A.T., Peet, R.K., Pillet, M., Roehrdanz, P.R., Sandel, B., Serra-Diaz, J.M., Šímová, I., Svenning, J.C., Violle, C., Weitemier, T.D., Wiser, S., López-Hoffman, L., 2022. A review of the heterogeneous landscape of biodiversity databases: Opportunities and challenges for a syn-

thesized biodiversity knowledge base. Global Ecology and Biogeography 31, 1242–1260. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.13497, doi:10.1111/geb.13497.

Forrester, D.I., Hobi, M.L., Mathys, A.S., Stadelmann, G., Trotsiuk, V., 2021. Calibration of the process-based model 3-PG for major central European tree species. European Journal of Forest Research 140, 847–868. URL: https://doi.org/10.1007/s10342-021-01370-3, doi:10.1007/s10342-021-01370-3.

Franklin, O., Harrison, S.P., Dewar, R., Farrior, C.E., Brännström, A., Dieckmann, U., Pietsch, S., Falster, D., Cramer, W., Loreau, M., Wang, H., Mäkelä, A., Rebel, K.T., Meron, E., Schymanski, S.J., Rovenskaya, E., Stocker, B.D., Zaehle, S., Manzoni, S., van Oijen, M., Wright, I.J., Ciais, P., van Bodegom, P.M., Peñuelas, J., Hofhansl, F., Terrer, C., Soudzilovskaia, N.A., Midgley, G., Prentice, I.C., 2020. Organizing principles for vegetation dynamics. Nature Plants 6, 444–453. URL: https://www-nature-com.inee.bib.cnrs.fr/articles/s41477-020-0655-x, doi:10.1038/s41477-020-0655-x.

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. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/evl3.160, doi:10.1002/evl3.160.

Hansen, N., Ostermeier, A., 2001. Completely Derandomized Self-Adaptation in Evolution Strategies. Evolutionary Computation 9, 159–195. doi:10.1162/106365601750190398.

Harrison, S.P., Cramer, W., Franklin, O., Prentice, I.C., Wang, H., Brnnström, A., de Boer, H., Dieckmann, U., Joshi, J., Keenan, T.F., Lavergne, A., Manzoni, S., Mengoli, G., Morfopoulos, C., Peñuelas, J., Pietsch, S., Rebel, K.T., Ryu, Y., Smith, N.G., Stocker, B.D., Wright, I.J., 2021. Eco-evolutionary optimality as a means to improve vegetation and land-surface models. New Phytologist 231, 2125–2141. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.17558, doi:10.1111/nph.17558.

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. URL: https://onlinelibrary-wiley-com.inee.bib.cnrs.fr/doi/abs/10.1111/j.1365-2699.2012.02745.x, doi:10.1111/j.1365-2699.2012.02745.x.

He, D., Wang, E., Wang, J., Robertson, M.J., 2017. Data requirement for effective calibration of process-based crop models. Agricultural and Forest Meteorology 234-235, 136–148. URL: https://www.sciencedirect.com/science/article/pii/S0168192316307420, doi:10.1016/j.agrformet.2016.12.015.

Hengl, T., Jesus, J.M.d., 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. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0169748, doi: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. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.6712, doi:10.1002/ece3.6712.

Higgins, S.I., O'Hara, R.B., Bykova, O., Cramer, M.D., Chuine, I., Gerstner, E.M., Hickler, T., Morin, X., Kearney, M.R., Midgley, G.F., Scheiter, S., 2012. A physiological analogy of the niche for projecting the potential distribution of plants. Journal of Biogeography 39, 2132–2145. URL: https://onlinelibrary-wiley-com.inee.bib.cnrs.fr/doi/abs/10.1111/j.1365-2699.2012.02752.x, doi:10.1111/j.1365-2699.2012.02752.x.

Huber, N., Bugmann, H., Lafond, V., 2020. Capturing ecological processes in dynamic forest models: why there is no silver bullet to cope with complexity. Ecosphere 11, e03109. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ecs2.3109, doi:10.1002/ecs2.3109.

Jalas, J., Suominen, J., 1972–2005. Atlas Florae Europaeae. Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki, Finland.

Keenan, T.F., Carbone, M.S., Reichstein, M., Richardson, A.D., 2011. The model–data fusion pitfall: assuming certainty in an uncertain world. Oecologia 167, 587–597. URL: https://doi.org/10.1007/s00442-011-2106-x, doi:10.1007/s00442-011-2106-x.

Koning, K.d., Broekhuijsen, J., Kühn, I., Ovaskainen, O., Taubert, F., Endresen, D., Schigel, D., Grimm, V., 2023. Digital twins: dynamic model-data fusion for ecol-

ogy. Trends in Ecology & Evolution 38, 916–926. URL: https://www.cell.com/trends/ecology-evolution/abstract/S0169-5347(23)00090-3, doi:10.1016/j.tree.2023.04.010.

Kreyling, J., Buhk, C., Backhaus, S., Hallinger, M., Huber, G., Huber, L., Jentsch, A., Konnert, M., Thiel, D., Wilmking, M., Beierkuhnlein, C., 2014. Local adaptations to frost in marginal and central populations of the dominant forest tree Fagus sylvatica L. as affected by temperature and extreme drought in common garden experiments. Ecology and Evolution 4, 594–605. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.971, doi:10.1002/ece3.971.

de Lafontaine, G., Amasifuen Guerra, C.A., Ducousso, A., Petit, R.J., 2014. Cryptic no more: soil macrofossils uncover Pleistocene forest microrefugia within a periglacial desert. New Phytologist 204, 715–729. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.12833, doi:10.1111/nph.12833.

Lazic, D., Geßner, C., Liepe, K.J., Lesur-Kupin, I., Mader, M., Blanc-Jolivet, C., Gömöry, D., Liesebach, M., González-Martínez, S.C., Fladung, M., Degen, B., Müller, N.A., 2024. Genomic variation of European beech reveals signals of local adaptation despite high levels of phenotypic plasticity. Nature Communications 15, 8553. URL: https://www.nature.com/articles/s41467-024-52933-y, doi:10.1038/s41467-024-52933-y.

Leinonen, I., 1996. A Simulation Model for the Annual Frost Hardiness and Freeze Damage of Scots Pine. Annals of Botany 78, 687–693. URL: https://doi.org/10.1006/anbo.1996.0178, doi:10.1006/anbo.1996.0178.

Lenz, A., Hoch, G., Vitasse, Y., 2016. Fast acclimation of freezing resistance suggests no influence of winter minimum temperature on the range limit of European beech. Tree Physiology 36, 490–501. URL: https://doi.org/10.1093/treephys/tpv147, doi:10.1093/treephys/tpv147.

Lobell, D.B., Burke, M.B., 2010. On the use of statistical models to predict crop yield responses to climate change. Agricultural and Forest Meteorology 150, 1443–1452. URL: https://www.sciencedirect.com/science/article/pii/S0168192310001978, doi:10.1016/j.agrformet.2010.07.008.

Malchow, A.K., Hartig, F., 2024. Calibration, sensitivity and uncertainty analysis of ecological models – a review URL: https://www.authorea.com/users/852871/

articles/1238623-calibration-sensitivity-and-uncertainty-analysis-of-ecological-models-a-review.

Mauri, A., Strona, G., San-Miguel-Ayanz, J., 2017. EU-Forest, a high-resolution tree occurrence dataset for Europe. Scientific Data 4, 160123. URL: https://www.nature.com/articles/sdata2016123, doi:10.1038/sdata.2016.123.

Van der Meersch, V., Armstrong, E., Mouillot, F., Duputié, A., Davi, H., Saltré, F., Chuine, I., 2024. Biological mechanisms are necessary to improve projections of species range shifts. bioRxiv URL: https://www.biorxiv.org/content/early/2024/05/08/2024.05.06.592679, doi:10.1101/2024.05.06.592679, arXiv:https://www.biorxiv.org/content/early/2024/05/08/2024.05.06.592679.full.pdf.

Van der Meersch, V., Chuine, I., 2023. Estimating process-based model parameters from species distribution data using the evolutionary algorithm CMA-ES. Methods in Ecology and Evolution 14, 1808–1820. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14119, doi:10.1111/2041-210X.14119.

Morin, X., Augspurger, C., Chuine, I., 2007. Process-Based Modeling of Species' Distributions: What Limits Temperate Tree Species' Range Boundaries? Ecology 88, 2280–2291. URL: https://onlinelibrary.wiley.com/doi/abs/10.1890/06-1591.1, doi:10.1890/06-1591.1.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., 2021. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth system science data 13, 4349–4383.

Niu, S., Luo, Y., Dietze, M.C., Keenan, T.F., Shi, Z., Li, J., Iii, F.S.C., 2014. The role of data assimilation in predictive ecology. Ecosphere 5, art65. URL: https://onlinelibrary-wiley-com.inee.bib.cnrs.fr/doi/abs/10.1890/ES13-00273.1, doi:10.1890/ES13-00273.1.

Pilowsky, J.A., Colwell, R.K., Rahbek, C., Fordham, D.A., 2022. Process-explicit models reveal the structure and dynamics of biodiversity patterns. Science Advances 8, eabj2271. URL: https://www.science.org/doi/10.1126/sciadv.abj2271, doi:10.1126/sciadv.abj2271.

Ruffault, J., Pimont, F., Cochard, H., Dupuy, J.L., Martin-StPaul, N., 2022. Sureau-ecos v2.0: a trait-based plant hydraulics model for simulations of plant water

729 status and drought-induced mortality at the ecosystem level. Geoscientific Model
730 Development 15, 5593–5626. URL: https://gmd.copernicus.org/articles/15/
731 5593/2022/, doi:10.5194/gmd-15-5593-2022.

732 Saltré, F., Duputié, A., Gaucherel, C., Chuine, I., 2015. How climate, mi-
733 gration ability and habitat fragmentation affect the projected future distri-
734 bution of European beech. Global Change Biology 21, 897–910. URL:
735 https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.12771, doi:10.1111/
736 gcb.12771.

737 Saltré, F., Saint-Amant, R., Gritti, E.S., Brewer, S., Gaucherel, C., Davis,
738 B.A.S., Chuine, I., 2013. Climate or migration: what limited European
739 beech post-glacial colonization? Global Ecology and Biogeography 22, 1217–
740 1227. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.12085,
741 doi:10.1111/geb.12085.

742 Simmonds, E.G., Adjei, K.P., Cretois, B., Dickel, L., González-Gil, R., Laver-
743 ick, J.H., Mandeville, C.P., Mandeville, E.G., Ovaskainen, O., Sicacha-Parada,
744 J., Skarstein, E.S., O'Hara, B., 2024. Recommendations for quantitative un-
745 certainty consideration in ecology and evolution. Trends in Ecology & Evo-
746 lution 39, 328–337. URL: https://www.sciencedirect.com/science/article/
747 pii/S0169534723002793, doi:10.1016/j.tree.2023.10.012.

748 Singer, A., Johst, K., Banitz, T., Fowler, M.S., Groeneveld, J., Gutiérrez,
749 A.G., Hartig, F., Krug, R.M., Liess, M., Matlack, G., Meyer, K.M., Pe'er,
750 G., Radchuk, V., Voinopol-Sassu, A.J., Travis, J.M.J., 2016. Community
751 dynamics under environmental change: How can next generation mechanis-
752 tic models improve projections of species distributions? Ecological Mod-
753 elling 326, 63–74. URL: https://www.sciencedirect.com/science/article/
754 pii/S0304380015005281, doi:10.1016/j.ecolmodel.2015.11.007.

755 Sommer, P.S., Kaplan, J.O., 2017. A globally calibrated scheme for generating
756 daily meteorology from monthly statistics: Global-wgen (gwgen) v1.0. Geosci-
757 entific Model Development 10, 3771–3791. URL: https://gmd.copernicus.org/
758 articles/10/3771/2017/, doi:10.5194/gmd-10-3771-2017.

759 Trotsiuk, V., Hartig, F., Cailleret, M., Babst, F., Forrester, D.I., Baltensweiler, A.,
760 Buchmann, N., Bugmann, H., Gessler, A., Gharun, M., Minunno, F., Rigling,
761 A., Rohner, B., Stillhard, J., Thürig, E., Waldner, P., Ferretti, M., Eugster,
762 W., Schaub, M., 2020. Assessing the response of forest productivity to climate

extremes in Switzerland using model–data fusion. Global Change Biology 26, 2463–2476. URL: https://onlinelibrary-wiley-com.inee.bib.cnrs.fr/doi/abs/10.1111/gcb.15011, doi:10.1111/gcb.15011.

Tóth, B., Weynants, M., Pásztor, L., Hengl, T., 2017. 3D soil hydraulic database of Europe at 250 m resolution. Hydrological Processes 31, 2662–2666. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.11203, doi:10.1002/hyp.11203.

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. URL: https://www.science.org/doi/10.1126/science.aad8466, doi:10.1126/science.aad8466.

Wang, E., Engel, T., 1998. Simulation of phenological development of wheat crops. Agricultural Systems 58, 1–24. URL: https://www.sciencedirect.com/science/article/pii/S0308521X98000286, doi:10.1016/S0308-521X(98)00028-6.

Zhang, Y., Pichon, L., Roux, S., Pellegrino, A., Simonneau, T., Tisseyre, B., 2024. Why make inverse modeling and which methods to use in agriculture? A review. Computers and Electronics in Agriculture 217, 108624. URL: https://www.sciencedirect.com/science/article/pii/S0168169924000152, doi:10.1016/j.compag.2024.108624.