Evaluating phenological models for the prediction of leaf-out dates in six temperate tree species across central Europe

David Basler[*]

Institute of Botany, University of Basel, Schönbeinstrasse 6, CH-4056 Basel, Switzerland

ABSTRACT

Inter-annual variation in climate is reflected by changes in the timing of phenology. Over the last decades a considerable number of models have been developed in order to explain the inter-annual variation of spring phenology in trees. Contrary to empirical models, "process-based" models aim at simulating physiological processes in order to yield more realistic predictions of growing season onset dates. Despite the increasing knowledge on the environmental controls of seasonal dormancy in trees, the detailed action and interaction of the involved environmental drivers (chilling, photoperiod and warm temperature) remains to be elucidated. This study aims at a uniform comparison of a wide range of existing models (and new recombinations), on a multitude of long-term observation series in six tree species across central Europe, using extensive cross-validation. Even though the assessed models differ in the phases of dormancy and environmental drivers accounted for, they yielded a surprisingly similar quality of prediction of leaf unfolding dates. Depending on the species, the lowest average prediction errors for leaf unfolding (*RMSE*) ranged from 7 to 9 days for the dataset pooled across sites and years and from 4 to 6 days for site-specific predictions, in absence of any obvious geographical pattern. Simple models, that feature ecodormancy release only, performed similar or better than more complex models, which additionally include endodormancy release through chilling temperatures. Model parameterisation tended to converge towards similar behaviour and models with many parameters tended to overfit on the 40 year time-series of leaf unfolding. Additionally, all models tended to underestimate the inter-annual variation of leaf unfolding and failed to predict very early or late dates of leaf unfolding in certain years. The transfer of site-specific parameters to other sites was associated with an almost doubling of the average prediction error, independent of distance and climatic similarity between the calibration and validation sites. The findings challenge the accurate implementation of the physiological processes controlling spring phenology in the models and highlight shortcomings associated with model parameterisation on observational time-series only.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Plant phenology studies the seasonal and visible phenomena of plant development related to weather. The timing of phenological events reflects a combination of internal (genetic) settings and environmental influences. Given its significance for avoiding late spring freezing damage and impacts of early autumnal cold events, a well-timed phenology is crucial for plant survival. In addition, the control of synchronous flowering among individuals assures sexual reproduction. The phenological events defining the onset and end of the growing season are of special interest, since they are setting the length of the growing season, and

thereby controlling range limits of species (Chuine and Beaubien, 2001). During the dormant period, buds pass through three distinct states of dormancy (Lang et al., 1987): (1) paradormancy, a state of specific bud dormancy maintained due to physiological factors outside the bud but inside the plant (e.g., apical dominance), (2) endodormancy, state of inactivity mediated by factor inside the bud and (3) ecodormancy, a state of inactivity imposed by unfavourable environmental conditions at otherwise full preparedness for advancing seasonal development. The transitions between the different phases of dormancy are gradual and species-specific (Perry, 1971). In tree species adapted to cool climates, dormancy is induced by the shortening of day-length in autumn, perceived in leaves, and modulated by concurrent temperatures. Moderate sub-zero temperatures are then inducing endodormancy, which is generally released in late winter or early spring, after sufficiently long exposure to cool, but non-freezing temperatures in

the range of 2–7 °C ("chilling"; Coville, 1920; Doorenbos, 1953; Battey, 2000). Many species need an additional weather independent photoperiod signal to effectively advance the transition from endodormancy to the following ecodormancy. During this stage, actual weather (largely temperature) controls bud development and bud burst (environmental "forcing").

Phenology gained much attention during the last decades, once its implications in the climate change discussion became acknowledged. Ever since phenological data was collected, the phenological events were related to climate and simple models were built to calculate the timing of phenological phases, especially in agro-ecosystems. Nowadays, the applications of phenological models range from reconstruction and quality assessment of phenological time-series, spatial extrapolation of observations and even to species-specific predictions of phenology, and thus species performance, in future climate. Thus, phenology plays an important role in species distribution models (e.g., Chuine and Beaubien, 2001) or dynamic global vegetation models (e.g., Krinner et al., 2005). Through the intimate linkage with the length of the growing season and thus, net primary production, phenology plays also an important role in carbon cycle models at ecosystem and global scale (Richardson et al., 2013). For temperate and boreal tree species, numerous models have been developed to simulate the events of spring phenology, such as bud burst or leaf unfolding, whereas only few models attempted to simulate the autumnal phases of phenology, such as leaf colouration and leaf fall (White et al., 1997; Delpierre et al., 2009). The more mechanistic models commonly outperform simple correlative statistical models for phenology, which often use linear correlations to spring temperature only (but see Olsson and Jönsson, 2014). These "process-based" models are also able to reflect the non-linear responses of phenology to the various environmental drivers. In simple phenological models, the bud development towards bud-burst is basically defined as a response to concurrent temperature, mostly by adopting the concept of accumulated temperature over a certain threshold (degree days). However, the shortcomings of this simple approach (accounting for the release of ecodormancy only) and the increasing knowledge of the underlying physiological processes motivated the development of numerous advanced models for spring phenology, which account also for chilling and photoperiod influences. Most recently, even the complex interactions of all three drivers of spring phenology, chilling signals, photoperiod, and actual thermal forcing were integrated into a single model (Caffarra et al., 2011). However, with the increasing number of factors, complexity of models increases dramatically and parameterisation becomes increasingly difficult. At first, the statistical fitting of parameters was difficult and often led to unstable parameter estimates (Kramer, 1994), thus Hänninen (1995) compared 96 model formulations using parameters derived from literature. Later, efficient optimisation methods (Chuine et al., 1998) and appropriate methods for the statistical estimation of prediction errors, such as bootstrapping (Häkkinen, 1999) or cross-validation (Chuine et al., 1999) led to further improvement of model parameterisation and evaluation. No single model structure was found to predict spring phenology across different species, so the best predictive models are still species-specific (Hunter and Lechowicz, 1992; Chuine et al., 1998; Schaber and Badeck, 2003) and different model structures may perform equally well for a given species (Schaber and Badeck, 2003). A recent uncertainty analysis for a set of phenological models using data from Harvard forest revealed, that prediction errors are largely a result of the uncertain nature and strength of the actual drivers (model structure), and to a lesser extent due to model parameterisation (Migliavacca et al., 2012). The more recently developed models have been tested on rather limited datasets for only a few species. The current study collectively analyses the performance of current process-based phenology model structures for three aspects (1) generalisation, (2)

site-specific accuracy and (3) spatial transferability, using a large and consistent phenology data set covering 40 year of observation on a multitude of sites throughout Europe for 6 temperate tree species. The assessed models differ with respect to the mechanisms they account for (dormancy induction, endodormancy release and ecodormancy release), the employed drivers (forcing temperature, chilling temperature and photoperiod), and the specific responses for different species. This study aims at improving the understanding of the capabilities and uncertainties of these models, and disclosing some pitfalls in modelling the spring phenology of temperate and boreal forest trees.

2. Material and methods

2.1. Models types

The "process-based" phenology-models published so far, and included here, simulate the environmental influence on bud development, until a critical developmental threshold for bud burst or leaf unfolding is reached. Parameters common to most models are a starting date, after which the specific environmental drivers affect bud development, and one or more parameters controlling the rate of response to environmental drivers.

The models are grouped according to their scope of operation into three categories: (1) models explaining ecodormancy release only, (2) models explaining the release of endo- and ecodormancy and (3) models explaining the whole transition from dormancy induction until bud burst. Further, I classified models by the environmental drivers they are accounting for: chilling temperature, photoperiod and forcing temperature (Table 1). Models were implemented according to the original publication (Table 1, supplementary Table S1); however, I fitted the starting date rather than using an arbitrary date (such as 1 January). For each parameter, upper and lower limit was defined within a wide, but (biologically) reasonable range (see supplementary Table S2).

2.1.1. Models accounting for ecodormancy release only.

These are the oldest models, dating back to de Réaumur (1735), accounting for thermal forcing in spring only. These "Thermal Time" models (Wang, 1960; Cannell and Smith, 1983; Hunter and Lechowicz, 1992; Chuine et al., 1999) are using degree days as forcing units. A modification of this model type, hereafter named the "sigmoid Thermal Time model" (Hänninen, 1990; Kramer, 1994) uses a sigmoid, rather than linear, forcing function (see supplementary Table S1). Although photoperiod is well known to influence phenology of crops (e.g., Masle et al., 1989; Siebert and Ewert, 2012) and late successional tree species (Caffarra and Donnelly, 2010; Körner and Basler, 2010; Basler and Körner, 2014), few models yet include photoperiod as explicit driver of spring phenology (the fixed starting date of most models may, however, imply a strong photoperiod threshold). In the Photothermal-time model developed for crops (Masle et al., 1989) and successfully applied to trees by Črepinšek et al. (2006), photoperiod has been included as an additional factor influencing the rate of forcing. Recently, a study investigating the shortcomings of the classic Thermal Time model again suggested the inclusion of photoperiod as explicit driver and thereby proposed an extension of the Photothermal-time model by an additional exponential constant (M1 model; Blümel and Chmielewski, 2012).

2.1.2. Models accounting for endo- and ecodormancy release.

The chilling requirement, indicating to the plant that winter has passed, plays an additional role in dormancy release of temperate and boreal trees. In current models, the response to chilling temperatures is implemented either as a triangular function of temperature (defined by minimal chilling temperature, optimal chilling

Table 1
Overview of the phenological models for leaf unfolding included in this study. The models are grouped by implemented processes and drivers: chilling temperatures (C), forcing temperatures (F) and photoperiod (P). Temperature responses not separable in chilling or forcing are indicated with T.

Model name	Drivers	No. parameters	Comments/references
Empirical			
NULL		1	Mean date of leaf unfolding
Linear	T	2	Simple linear regression with mean spring temperature
Ecodormancy release only			
Thermal Time[a]	F	3(4)	de Réaumur (1735), Wang (1960), Cannell and Smith (1983) Chuine et al. (1999), Hunter and Lechowicz (1992) Hänninen (1990), Kramer (1994)
Photothermal-time[a]	PF	3(4)	Masle et al. (1989), Črepinšek et al. (2006)
M1[a]	PF	4(5)	Blümel and Chmielewski (2012)
Endo- and eco dormancy release			
Alternating	CF	5	Cannel and Smith (1983), Murray et al. (1989)
Sequential[b]	CF	8	Hänninen (1990), Kramer (1994)
Parallel[b]	CF	9	Landsberg (1974), Hänninen (1990), Kramer (1994)
Unified	CF	9	Chuine (2000)
Sequential M1[b]	CPF	9	Combination of Sequential model with M1 model
Parallel M1[b]	CPF	10	Combination of Parallel model with M1 model
Unified M1	CPF	10	Combination of Unified model with M1 model
DORMPHOT spring	CPF	8	DORMPHOT model without dormancy induction
PIM 1–12	PT	10	Schaber and Badeck (2003)
Dormancy induction, endo- and ecodormancy release			
Four-Phase[b]	CF	12	Vegis (1964), Hänninen (1990)
Deepening-Rest[b]	CF	10	Kobayashi and Fuchigami (1983)
DORMPHOT	CPF	11	Caffarra et al. (2011)

[a] These models were also calibrated using a sigmoid temperature response (Hänninen, 1990; Kramer, 1994), adding one parameter.
[b] These models were also calibrated using a bell-shaped chilling response (Chuine, 2000).

temperature and maximal chilling temperature; Hänninen, 1990; Kramer, 1994) or a bell shaped curve (Chuine, 2000; see equation in supplementary Table 1). Depending on the model's assumption, the chilling requirement has to be fulfilled either before forcing temperatures are accumulating, such as in the Sequential model (Hänninen, 1990; Kramer, 1994), or chilling and forcing may act simultaneously, as implemented in the Parallel model (Landsberg, 1974; Hänninen, 1990; Kramer, 1994). The either sequential or parallel integrations of chilling into models have later been combined in the Unified Model (Chuine, 2000). However, the flexible structure of this model has a cost in terms of higher number of parameters to be defined. For this study, the three formerly mentioned models were extended with the addition of a photoperiod response, as photoperiod is known to control the phenology of certain species: In the Parallel-, Sequential- and Unified model, the (sigmoid) response to forcing temperature was extended by the photoperiod response factor of the M1 model (see equation in supplementary Table S1).

Again another formulation for including chilling fulfilment has been introduced by Murray et al. (1989) as the Alternating model, where a specific day may either add to the accumulation of forcing or to the accumulation of chilling, depending on the actual daily mean temperature. I did not include further chilling models designed mainly for specific fruit trees, such as the Chill-Hours Model (summed hours with $T < 7.2\,°C$; Weinberger, 1950), the Utah Model (temperature-weighted accumulation of chilling hours; Richardson et al., 1974), the Positive Utah Model (Lindsley-Noakes et al., 1995), and the very complex "Dynamic model" for chilling accumulation (Fishman et al., 1987a,b), as these models would require hourly input data not available for this study.

A different approach at phenological modelling accounts for molecular (e.g., hormonal) regulation of bud dormancy and was introduced by Schaber and Badeck (2003) as the "Promoter–Inhibitor models" (PIM): these models do not explicitly separate endodormancy and ecodormancy but rather describe dormancy release as a continuum under control of the balance of (virtual) promoters and inhibitors. Promoter-accumulation and inhibitor-decomposition rates are modulated by different combinations of temperature and/or photoperiod.

2.1.3. Models for the whole dormant period

Some more sophisticated models include (temperature-induced) dormancy induction prior to the release of dormancy through chilling and warm temperatures, such as the Four-Phase Model (Vegis, 1964; Hänninen, 1990) or the Deepening-Rest model (Kobayashi and Fuchigami, 1983). The most complex model of this type, the DormPhot model (Caffarra et al., 2011) designed for *Betula*, integrates even the complex interactions of photoperiod with dormancy induction, chilling and thermal forcing.

2.1.4. NULL model and linear regression

All models were compared to the NULL model that assumes a fixed mean date of leaf unfolding over all years. Additionally, I also provide the results of a simple linear regression model of leaf unfolding dates against mean spring temperature: the relevant period was determined by choosing the best linear correlation of leaf unfolding dates with mean temperatures of either a single month (January–May) or of a continuous combination of these months.

2.2. Phenological data

The phenological time-series used for this study was extracted from the PEP725 Pan European Phenology Database (Data set accessed 2013-01-07 at http://www.zamg.ac.at/pep725/). Out of the many tree species covered by this dataset, I selected 6 species with abundant temporal and spatial coverage: *Aesculus hippocastanum* L., *Betula pendula* Roth, *Fagus sylvatica* L., *Quercus robur* L., *Larix decidua* Mill. and *Picea abies* (L.) H. Karst. (in the following species are referred to by their genus). These species represent contrasting characteristics in terms of successional status (early vs. late successional), leaf traits (broad leaved vs. coniferous) and canopy duration (deciduous vs. evergreen). As phenological spring phase, leaf unfolding (BBCH code 11) was selected for the broad-leafed species and needle elongation (BBCH 10) for conifers. In each species, all data covering the observation period of 1970–2009 was used to apply a robust mean outlier detection using the 30-day rule (Schaber and Badeck, 2002) to exclude the most obvious flaws. Only stations where the full 40 years were covered after outlier
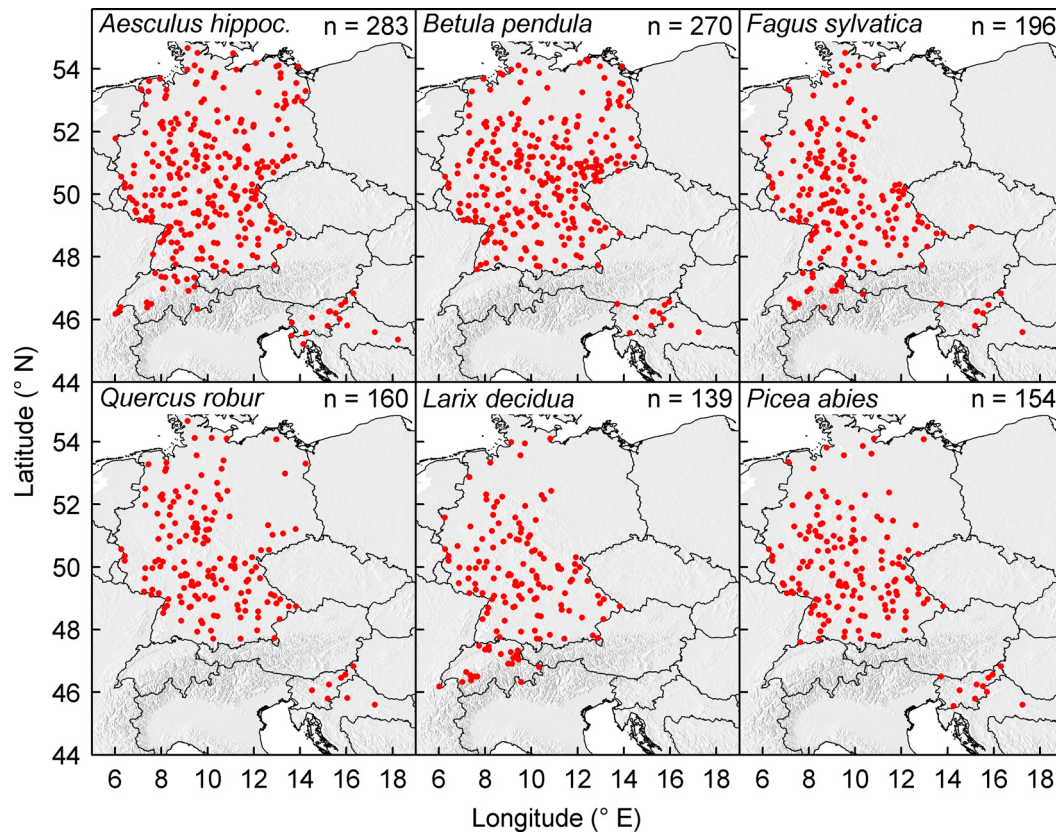
Fig. 1. Locations of the phenological observation sites used in this study (leaf unfolding data 1970–2009). Data was provided by the Pan European Phenological Network (PEP).

removal were used in the study. Time-series fulfilling these requirements were located in Germany, Switzerland, Austria, Slovenia and Croatia (Fig. 1), spanning an elevation range from sea level to 1440 m a.s.l. Mean leaf unfolding was earliest in *Larix*, followed by *Aesculus* and *Betula*, while a later mean leaf unfolding was found in the late successional species *Fagus*, *Quercus* and *Picea* (Table 2). The inter-annual range of the overall spring phenology per site was around 5 weeks.

2.3. Climate data

As temperature reference, the gridded daily mean temperatures provided in the E-OBS Dataset version 7.0 (Haylock et al., 2008) was used. The dataset provides European daily mean temperatures since 1950 on a 0.25° regular grid (~28 km). Temperatures at the phenological observation sites were calculated by using the corresponding grid cell's temperature, corrected for the observation site's elevation by a linear lapse rate. The lapse rates were calculated daily, using a linear regression of the temperatures of the 25 surrounding grid cells against their elevation. The applied method provides reasonable estimates of local temperature when

tested against actual weather station data (obtained from 157 weather stations across Switzerland and Germany during the 40 years period: mean $R^2 = 0.98 \pm 0.06$, *RMSE* $1.0 \pm 0.6\,°C$; Data provided by the German Weather Service DWD and MeteoSwiss) and outperformed the use of monthly- or annual averaged lapse rates. Photoperiod was calculated as a function of latitude and day-of-year using basic trigonometry (sunrise to sunset, without twilight; http://www.gandraxa.com/length_of_day.xml).

2.4. Model calibration and validation

First, to assess the general performance of the models, I fitted the models for each species to the pooled dataset of all years and sites, and used a 10-fold cross-validation (the resulting errors are hereafter referred to as "pooled calibration" and "pooled validation"). Second, to assess the site-specific performance of the models, each model was fitted over the whole 40 years ("site-specific calibration") as well as evaluated using leave-one-out cross-validation over the 40 year observation period per individual site ("site-specific validation"). Finally, to characterise the spatial transferability of site-specific model parameters, I used the

Table 2
Properties of the phenological datasets used in this study. Mean (±SD) day of leaf unfolding (BBCH 11), inter-annual range and trend during the observed period 1970–2009.

Species	Sites	Data points	Mean leaf unfolding (DOY)	Maximum inter-annual range	Mean trend 1970–2009 (d y^{-1})	Mean R^2 1970–2009
Aesculus hippocastanum	283	11,320	109.8 ± 7.5	37.3 ± 6.4	−0.32 ± 0.20	0.19 ± 0.2
Betula pendula	270	10,800	109.9 ± 6.5	38.0 ± 6.1	−0.29 ± 0.16	0.14 ± 0.1
Fagus sylvatica	196	7840	118.4 ± 6.1	33.2 ± 6.4	−0.29 ± 0.17	0.23 ± 0.2
Quercus robur	160	6400	124.6 ± 6.7	35.8 ± 6.5	−0.38 ± 0.18	0.29 ± 0.2
Larix decidua[a]	139	5560	105.7 ± 8.3	43.9 ± 7.3	−0.31 ± 0.22	0.13 ± 0.1
Picea abies[a]	154	6160	128.0 ± 6.7	36.1 ± 8.0	−0.36 ± 0.19	0.27 ± 0.2

[a] Needle elongation (BBCH 10).

parameters calibrated on each site ("site specific calibration", see above) to predict the phenology of all remaining sites ("external site validation") for each model.

2.5. Software

The model evaluation framework used here was implemented using the programming language C. Best parameters were estimated by minimising the root mean squared error (*RMSE*) between observations and prediction, applying an adaption of the Adaptive Simulated Annealing code (ASA; Ingber, 1993). Simulated annealing (Metropolis et al., 1953) has been used previously for the parameterisation of phenology models and has been shown to reliably find the global minimum error (Chuine et al., 1998; Schaber and Badeck, 2003; Caffarra et al., 2011). For a single calibration of a model, the simulated-annealing procedure was repeated 20 times each (using random initial parameters) to find the best fitting parameters. Given that the parameter estimation is computationally intensive, I parallelised the code using OpenMPI (www.openmpi.com) and the calculations were done on the high-performance computing (HPC)-cluster of the University of Basel, using up to 40 processors in parallel. Including the extensive cross- and spatial-validation of the models, a total approximately 1.8 million model calibrations and 54 million validations on different (sub-)datasets were calculated for the 36 models (excluding the NULL model).

2.6. Statistics

Model root mean squared errors (*RMSE*) were tested for normal distribution using the Shapiro–Wilk test. Model performance was compared using an ANOVA on the log transformed *RMSE* of the individual sites using site as random factor. A post-hoc Tukey HSD test was then used to identify performance differences between models. As a measure for climatic similarity between sites, the mean Euclidian distances between the daily mean temperatures from September up to and including May over the 40 years observation period was calculated. Model residuals of all sites and years were checked for correlation with basic geographical variables (latitude, longitude and elevation), mean spring temperature (January–May) and the timing of leaf phenology (mean and variation). All statistical analyses and figures were done using R 2.15.0 (R Development Core Team, 2010)

3. Results

3.1. Overall phenology

The estimation of a single, best-fit set of parameters for each model to match phenology across all sites resulted in a mean pooled calibration error (*RMSE*) of around 7–9 days, depending on species: lowest calibration errors were obtained in *Betula* (6.6 days), highest in *Larix* (8.6 days) while *Fagus*, *Quercus*, *Picea* and *Aesculus* took intermediate positions (6.9, 7.2, 7.5 and 7.9 days, respectively). Due to the large size of the dataset, the pooled validation error (as estimated by a 10-fold cross-validation) was very similar to the pooled calibration error and varied only slightly across the individual cross-validation folds. As expected, all models outperformed the NULL model (9.9–13.6 days); however no single model was distinctly superior. In contrast, no significantly different performance (*RMSE*) was observed between the majorities of the models (Fig. 2, "pooled validation"). The residuals of the different models were highly correlated (mean $r = 0.97 \pm 0.02$), except for the null models (mean $r = 0.67 \pm 0.05$).

Species-specific differences were found for the time period providing the lowest calibration and validation error in the simple linear regression of leaf unfolding dates with mean temperatures: the mean of March and April temperatures was best for *Aesculus*, *Betula* and *Fagus*, whereas the mean of April and May temperatures was best for *Quercus*, the mean of February–April temperatures for *Larix* and finally the mean of March–May temperatures for *Picea*.

3.2. Site-specific phenology

Fitting the models to the individual 40 year time-series of each site resulted in a decreased mean *RMSE* across sites in all models compared to pooled validation error (Fig. 2). The lowest mean site-specific validation *RMSE*s ranged from 4 to 5 days, with only minor differences among species (Fig. 3). While the models accounted for up to 70% of the variation in leaf unfolding during model calibration, only up to 40% of the variation was explained by the models in the cross-validation. Site-specific model calibration excludes the variation between sites (between ecotypes or between individual trees), at the cost of less stable predictions, as indicated by the considerable difference between calibration and validation error, which was roughly 12% larger than the calibration error across species and models (Fig. 2). Interestingly, models with only few parameters, based on forcing temperature and photoperiod (M1, Photothermaltime) or on forcing temperature alone (Thermal Time model), resulted in a smaller within-site cross-validation error than models including a chilling requirement (Fig. 3). Contrary, the models including the chilling requirement achieved a smaller calibration error when fitted over the full 40 year time-series, likely due to overfitting of these highly parameterised models. No model achieved outstanding performance at all sites, as indicated by the ranking of the models according to their mean cross-validation error (Fig. 3). Most models fit reasonably well on the majority of sites and worse on only a few sites, resulting in a positively skewed, log-normal distribution of the *RMSE* across sites. Within a site, the residuals were normally distributed. No tight correlation of the model performance with geographical parameters of the sites (latitude, longitude and elevation) or site mean spring temperature (January–May) was present in any model (R^2 always <0.06, *slope* always <0.3). However, the prediction error of all models was correlated with the yearly anomaly of phenology i.e. models predictions were worse in years where the timing of leaf-out diverged further from the long-term mean (Fig. 4). This trend was strongest in *Fagus* (mean *slope* = -0.58 ± 0.05, mean $R^2 = 0.37 \pm 0.05$; excluding the NULL model) and *Picea* (*slope* = -0.52 ± 0.02, $R^2 = 0.35 \pm 0.04$), intermediate in *Quercus* (*slope* = -0.49 ± 0.02 and $R^2 = 0.33 \pm 0.03$) and *Aesculus* (*slope* = -0.45 ± 0.02, $R^2 = 0.29 \pm 0.03$), and slightly less pronounced in *Larix* (*slope* = -0.42 ± 0.02, $R^2 = 0.27 \pm 0.03$) and *Betula* (*slope* = -0.35 ± 0.02, $R^2 = 0.21 \pm 0.03$).

No significant differences (significance level 0.05) were found between model using the different chilling (triangular vs. bell shaped) and forcing functions (linear above base temperature vs. sigmoid temperature response), except in *Fagus* and *Picea*, where the Sequential model applying the bell shaped chilling functions achieved a significantly lower error than the variant applying the triangular response function. The addition of photoperiod to the classical endodormancy release models (Parallel, Sequential and Unified) led to slightly (not significantly) lower mean prediction errors for leaf unfolding in all species, especially in the case of the Parallel M1 model. Of the models describing the whole dormancy period, the DormPhot model always yielded the lowest prediction errors and was significantly different from other models of the same category in *Aesculus* and *Betula*, but not in the four other species.

Interestingly, the start day of the models simulating the ecodormancy phase only (without chilling) was always fitted best between mid-January and mid-February, indicating that most variation of the actual bud burst dates under current climate may be (statistically) related to the temperature and photoperiod
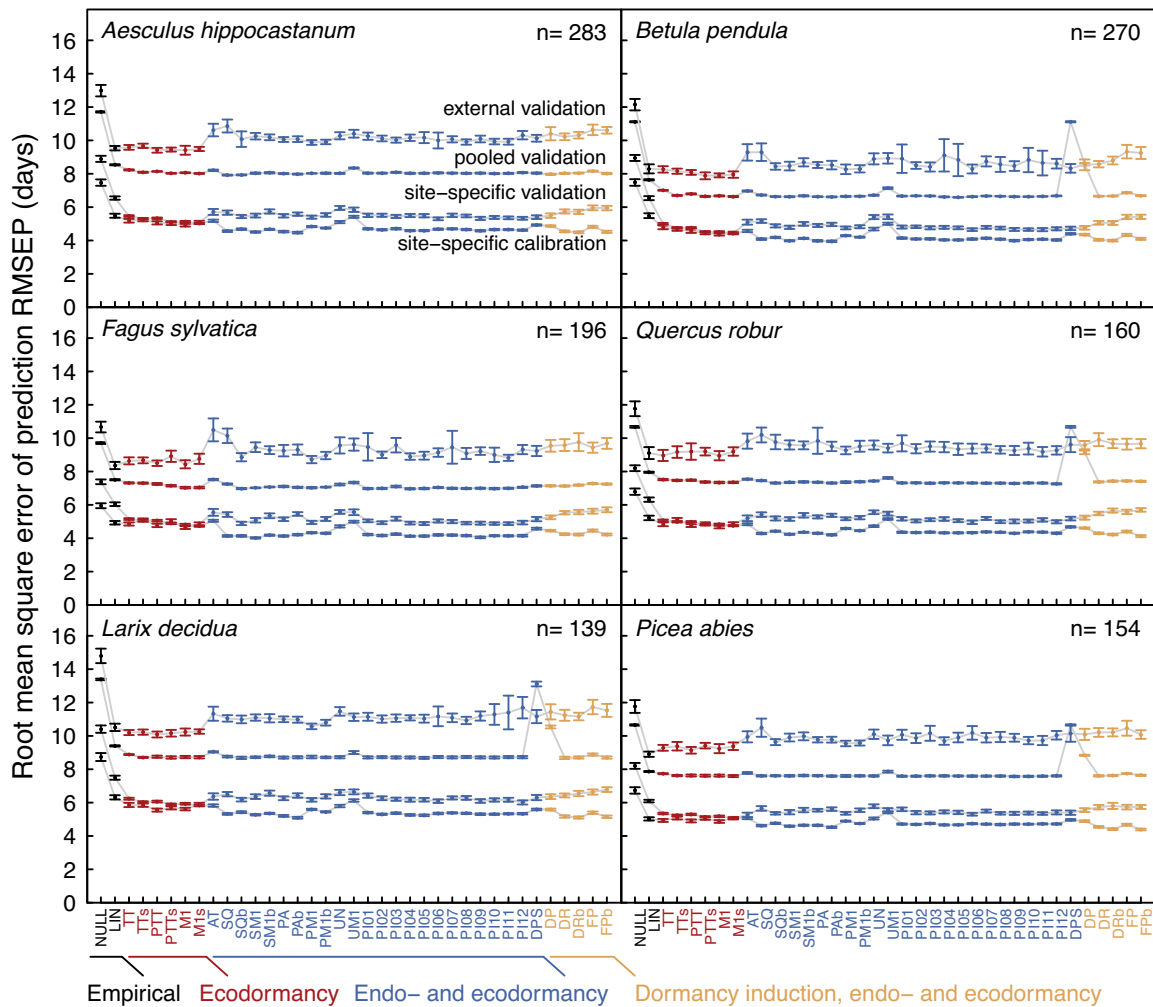
Fig. 2. Mean (±SE) *RMSE* of predicted leaf unfolding dates of 37 models in 6 species for pooled validation across all years and sites (40 years × *n* sites; 10-fold cross validation), site-specific calibration/validation (40 years per site; leave-one-out cross-validation) and, external validation of site-specific parameters. Model abbreviations: NULL: Mean date of leaf unfolding, LIN: Simple linear regression against mean temperature of specified months (*Aesculus, Betula* and *Fagus*: March and April; *Quercus*: April and May; *Larix*: February to April; *Picea*: March to May), TT: Thermal Time, PTT: Photothermal-time, M1: M1 model, AT: Alternating, SQ: Sequential, SM1: Sequential M1, PA: Parallel, PM1: Parallel M1, UN: Unified, UM1: Unified M1, PM01- PM12: Promotor–Inhibitor, DPS: DormPhot spring, DP: DormPhot, DR: Deepening-Rest FP: Four-Phase; 's' denotes the variant using a sigmoid, rather than linear forcing function; 'b' denotes the variant using a bell-shaped, rather than triangular, chilling function.

conditions during early spring. Similarly, the calibration of the chilling response in endodormancy-release models often yielded rather late starting dates for chilling accumulation (early to mid-winter).

3.3. External validation

Applying the best site-specific parameter estimates of the different models to predict phenology of the remaining sites resulted in an almost two-fold increase of the overall validation error compared to the site-specific calibration error (1.8-fold increase compared to the site-specific validation error, 1.25-fold compared to the pooled validation error; Fig. 2). On some sites however, the specific parameters resulted even in a lower *RMSE* than on the site used for parameter calibration. The lowest mean *RMSE* of the external-validation across sites ranged from 7.9 ± 15 days in *Betula* (M1 model), 8.3 ± 1.8 days in *Fagus* (M1 model), 9.3 ± 1.9 days in *Aesculus* (M1 model), 9 ± 1.9 days in *Picea* (Photothermal-time model) to 10 ± 12.0 days in *Larix* (Photothermal-time model). However, the best models performed only 29% (*Fagus*) to 35% (*Betula*) better than the NULL model (10.8 ± 14.2 days; Fig. 5). Again, the mean external-validation *RMSE*s of the different models were

almost similar, as was formerly found for the site-specific cross-validation *RMSE*. The external-validation performance was neither reasonably correlated with the geographical distance from the calibration site to the validation site in any model (R^2 always <0.03, *slope* always <0.01) nor with the mean Euclidean distance value of climate similarity among observation sites (R^2 always <0.05, *slope* always <0.02).

The model evaluation error increased with the increasing deviation of the evaluations site's mean phenology from the overall mean phenology. Surprisingly, the simple linear regression was among the models resulting in the lowest transfer error in 4 out the 6 species and yielded even the lowest absolute mean validation error in *Fagus* and *Picea*.

4. Discussion

The comparison of 35 "process-based" phenological models for leaf unfolding in 6 tree species using a long-term observation dataset (40 years) revealed a surprisingly small effect of model structure on the quality of prediction under a current climate. In the following the observed patterns and their implications for phenology modelling are discussed.
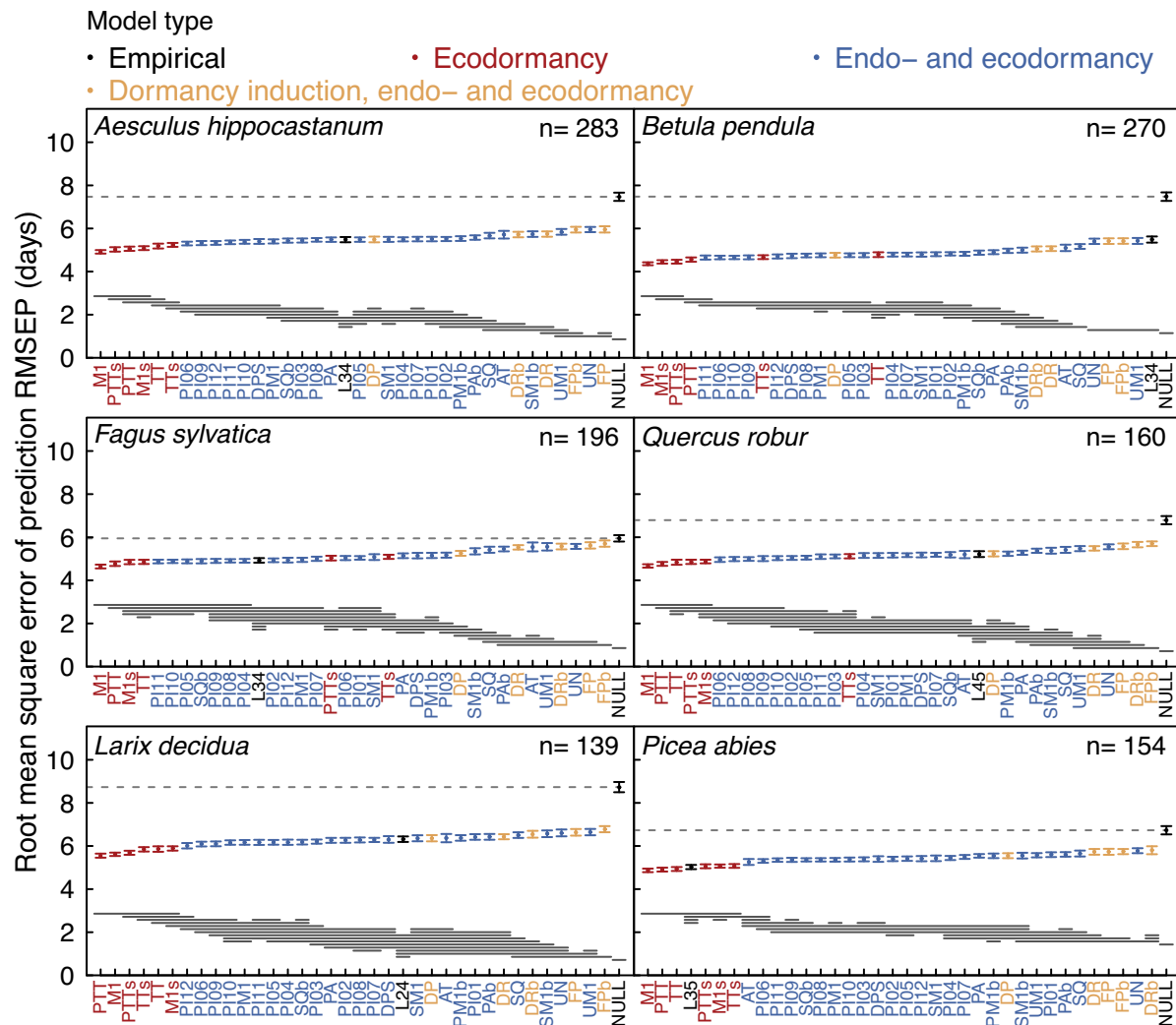
Fig. 3. Mean (±SE) *RMSE* of the predicted leaf unfolding date from the site-specific leave-one-out cross-validation of 37 models on 40 years of leaf unfolding observations in 6 species. Horizontal grey lines under the bars connect non-significantly different models (ANOVA followed by Tukey HSD; $p < 0.05$). The relative distance between model *RMSE* and *RMSE* of the NULL model (=total variation, dashed line) represents the explained variance of the models (R^2). Models were sorted by increasing *RMSE*. For model abbreviations, see Fig. 2.

4.1. Model performance and structure

4.1.1. Model prediction errors

In general, no model could accurately predict the phenology of leaf unfolding with a single set of parameters calibrated across all years and sites. Instead, a broad range of models yielded an equally large error with highly correlated residuals. A mean error for leaf unfolding dates of more than a week is rather insufficient for most applications, as, for example, a variation of leaf unfolding of ±10 days has been suggested to induce a variation of ±5.0% of annual GPP and ±2% annual ET using global circulation models (Migliavacca et al., 2012). Site-specific calibration decreased the mean prediction error to around 4–5 days, similar to that of previous studies using fewer models and data (e.g., Schaber and Badeck, 2003).

4.1.2. Spatial extrapolation

The increased mean prediction error present when applying site-specific parameter sets to predict the phenology of other sites (external validation; Fig. 2), indicates a limited spatial extrapolation potential of process based models. Indeed, for all 6 species assessed here, the prediction quality of a simple linear regression

model was almost similar to that of the process-based models, a pattern also observed for two species by Olsson and Jönsson (2014). The generally high prediction error for the external validation support the conclusions of a recent study investigating leaf unfolding of *Betula* along urbanisation gradients in Germany (using the DormPhot model), that the possibility to use a space-for-time substitution in phenological time-series is limited (Jochner et al., 2013). This claim is especially true as the increase in the error of prediction on the validation site was neither correlated with the distance from the calibration site nor with climate similarity of the sites (in terms of temperature). Thus, the observed increase of the prediction error is likely based on several confounded factors: (1) the individual variation of phenology among trees, (2) possible ecotypic/provenance differentiation of trees across sites or (3) the overfitting of a highly parameterised model on the small site-specific dataset (40 observations of leaf unfolding per site).

Leaf phenology is known to exhibit considerable variation even between individuals of a given population. For example, most of the variation of *Pinus sylvestris* L. in Scotland occurs within, and not between populations (Salmela et al., 2013). As most of the time-series used here, represent the phenology of individual trees (depending on the observation protocol of the original
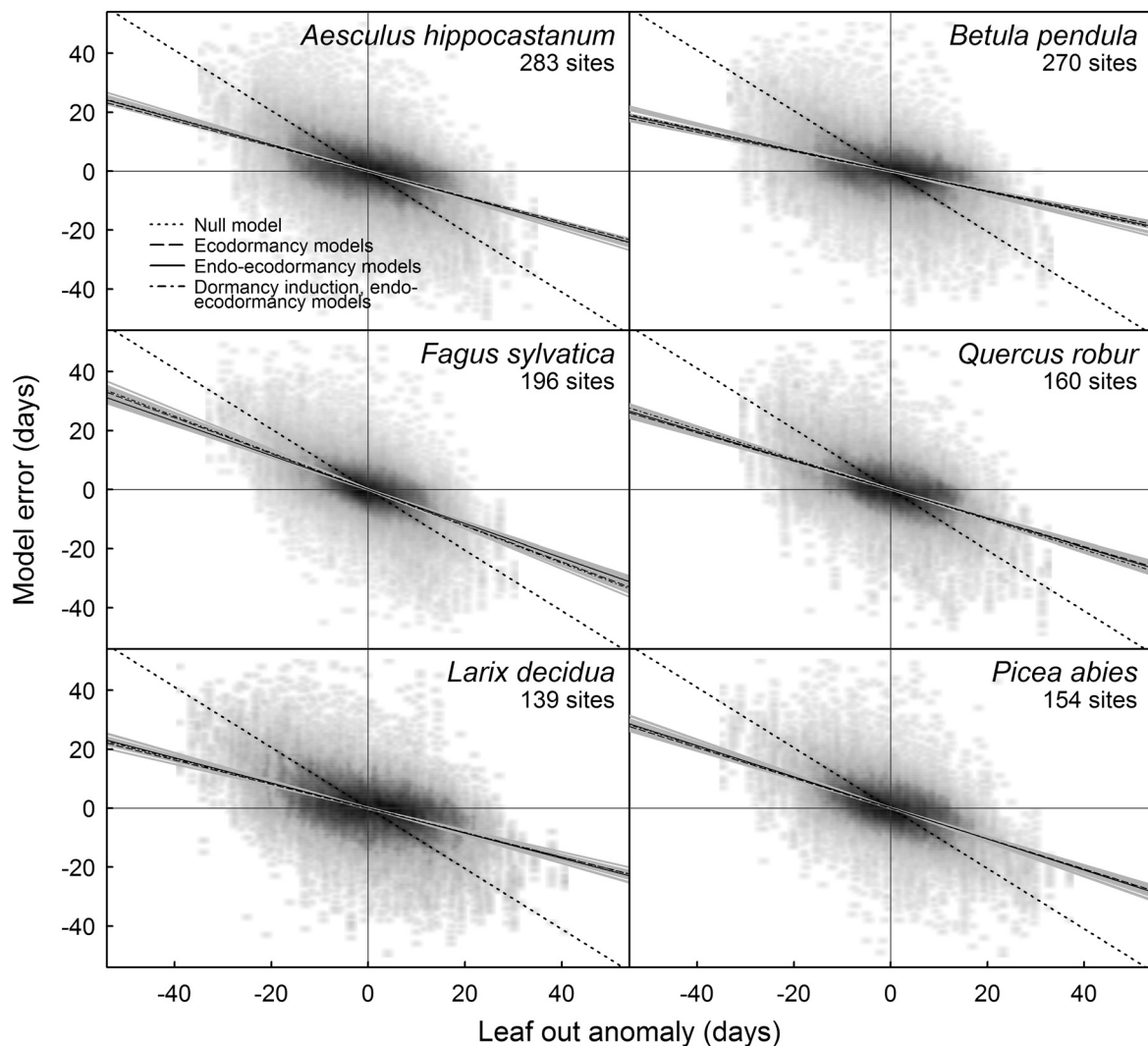
Fig. 4. Prediction errors (residuals) of 36 phenological models in site-specific cross-validations (40 year × *n* sites) as influenced by the phenological anomaly per site in 6 tree species. Black lines indicate the mean trends across models categories. Individual model trends are indicated with light grey lines. Due to the large number of data points a smoothed grayscale density representation of the scatterplot is shown.

collector/weather service), it is unknown to which extent the individual tree is representing the mean phenology of a site.

A provenance differentiation of phenology, often related to climatic gradients, has been evidenced in common gardens for the timing of bud burst (e.g., Kriebel and Wang, 1962; von Wuehlisch et al., 1995) and has been attributed to different chilling and thermal forcing requirement among provenances, for example, a variation in the required temperature sums for bud burst (Gunderson et al., 2012; Vitasse et al., 2013). Some ecotypic differentiation has also been evidenced for the photoperiodic induction of bud set (Wareing, 1956; Vaartaja, 1959; Kriebel and Wang, 1962; von Wuehlisch et al., 1995; Thomas and Vince-Prue, 1997) and bud burst in spring (Linkosalo and Lechowicz, 2006; Basler and Körner, 2012, 2014). However, the direction of the response of such genetic clines depends on species and the genetic effect on the timing of bud burst is often smaller than the environmental effects in situ (Vitasse et al., 2013). Conducting "hypothetical transplants", several previous modelling studies, concluded that local adaptation plays a subordinate role in predicting the flowering phenology of several, early successional lowland tree species (Chuine et al., 1999, 2000a,b).

4.1.3. Modelling endodormancy release

The similar performance of different models in all validation procedures conducted here might have been caused by the general spring-temperature accumulation principle common to all these models (even though the temperature-response functions differ among models). An analysis of the calibrated parameters and the high correlation among predictions of different models indicate, that most models conform to a model dominated by the accumulation of thermal forcing, while the additional parameters (e.g., for chilling accumulation) modulated the residual variation only. Indeed, certain parameter combinations may even prune some features from the complex models, reducing them to simple thermal time models (Linkosalo et al., 2008).

The finding that simple ecodormancy release models produce slightly better estimates of leaf phenology under current conditions than models with additional endodormancy release through chilling, is in line with several previous modelling studies (Häkkinen et al., 1998; Hannerz, 1999; Linkosalo et al., 2000, 2008; Leinonen and Kramer, 2002; Hänninen and Kramer, 2007; Granhus et al., 2009). A valid calibration of the chilling responses during winter is limited to sites and species where the chilling fulfilment is currently an issue and where a lack of chilling is also reflected in delayed leaf
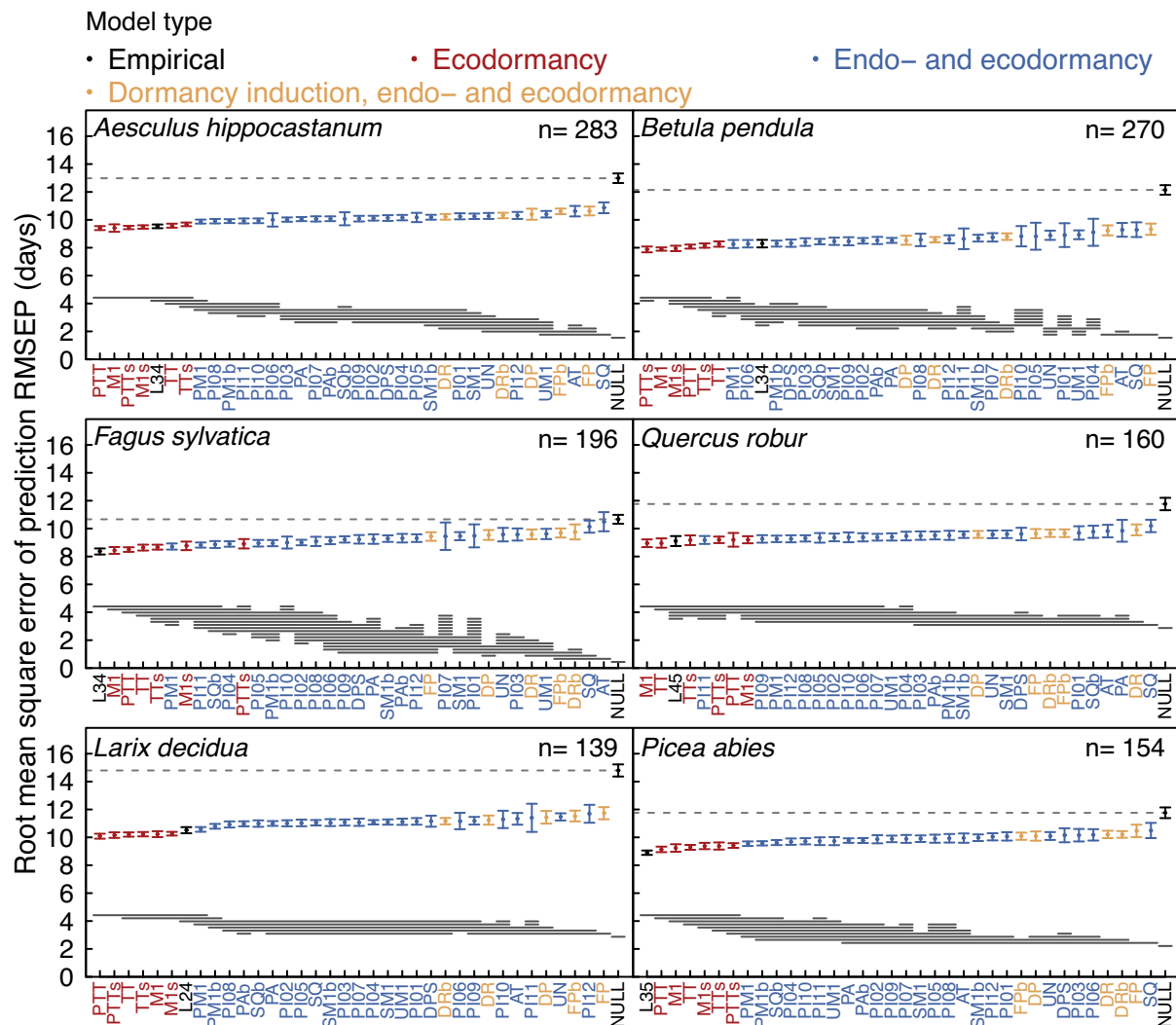
Fig. 5. Mean (±SE) *RMSE* of the predicted leaf unfolding date of 37 models in 6 species from the external site validation based on parameters calibrated on 40 years leaf unfolding observations per site. Horizontal grey lines under bars connect non-significantly different models (ANOVA followed by Tukey HSD; *p* < 0.05). The relative distance between model *RMSE* and *RMSE* of the NULL model (=total variation, dashed line) represents the explained variance of the models (*R*²). Models were sorted by increasing *RMSE*. For model abbreviations, see Fig. 2.

unfolding: on sites where the chilling requirement is fulfilled early in winter, buds remain in a state of ecodormancy for much of the remaining winter (as growth is suspended due to low temperature or short photoperiods). Hence, the chilling parameters are unlikely to be estimated unambiguously using fitting procedures, especially since the actual chilling requirements and effective ranges of chilling temperatures are yet scarcely known for forest tree species (Battey, 2000) and thus wide parameter ranges have to be used for model calibration.

The calibration of models which include dormancy induction (DormPhot, Deepening-Rest and Four-Phase model), suffer from the sparsely available observational data on the relevant autumnal phases (bud set and dormancy induction). The bulk of autumnal observations consist of partially coincidental, but more striking events such as leaf colouring and leaf fall which have also been used as starting dates in phenological modelling studies (Schaber and Badeck, 2003). However, these events do not represent the physiological state of bud dormancy induction and are strongly modified by of a few cold nights in autumn (but see Delpierre et al., 2009). Statistical calibration of these parameters restricted to previously fixed parameter ranges is thus often the only solution to calibrate such models.

4.1.4. Photoperiod responses during ecodormancy release

Models that included photoperiod during the ecodormancy release phase were generally among the models with the best performance in all 6 species. Linkosalo et al. (2006) speculated that the lower performance of models including chilling, compared to simple forcing models, may be caused by a yet inadequately formulated temperature response during the early phase of ecodormancy in late winter, soon after the fulfilment of the chilling requirement. The general trend to late starting dates in phenological models suggests either a negligible effect of mid-winter temperatures on the date of leaf unfolding under current conditions (as is also indicated by the good fit of the simple linear regressions), indicating either that the state of dormancy may affect the shape of the temperature response to chilling (Campbell and Sugano, 1975; Hänninen, 1990) or forcing temperatures (Junttila and Hänninen, 2012), or the presence of a strong external threshold, such as a possible photoperiod threshold. The trend to late starting dates may further result from the intrinsic model structure, such as analysed for the Thermal Time model (Blümel and Chmielewski, 2012). The inclusion of photoperiod as driver during ecodormancy release, as in the Photothermal-time or M1 model or also in combination with precedent endodormancy release, may slightly indeed

improve estimates of leaf unfolding dates and lead to more realistic parameters for the 6 tree species included here, and thus, should be considered as a reliable starting point for further model development.

4.1.5. Underestimated inter-annual variation

All evaluated models underestimated the inter-annual variation of leaf unfolding dates. The significant decrease of predictions quality for very early/late years indicates that models do not fully represent the true response to the environmental drivers. The underestimation of inter-annual variation was slightly less pronounced in *Betula* and *Aesculus* (Fig. 4), the two species, where the inter-annual range of leaf-out dates is larger (Table 1), suggesting that the sparse coverage of extreme responses in long-term observational data may bias model calibration and prevent the generalisation of the resulting parameter set. Consequently, either a dataset containing a wider range of responses or a model calibration scheme that puts more weight to extreme responses might be needed to improve model parameterisation. These results support Hänninen's (1995) conclusions, based on a comparison of leaf unfolding dates in field and in warming experiments, that to predict phenology in a future climate, the use of complementary data from experiments should be considered to increase data coverage for the calibration of the model parameters. However, in most cases the amount of data resulting from experiments does not allow for direct calibration of complex models and should rather be used to complement existing observational time-series or to restrict parameter ranges, also because data from experiments may not be representative for observations in situ (Wolkovich et al., 2012). For example, the phenology of tree seedlings, as often used in experiments, does not represent the responses of adult trees due to ontogenic differences (Vitasse, 2013). Thus, to study dormancy release in warming and photoperiod experiments, cuttings from adult trees should preferably be used (Vitasse and Basler, 2014; Primack et al., 2015), whereas rooted cuttings or grafts from adult trees may further be used to investigate the induction of dormancy or responses to chilling temperatures in mid-winter.

Another possibility to extend the range of observed responses is to assess phenological data of transplant gardens e.g., International Phenological Gardens (Schnelle and Volkert, 1974; Chmielewski and Rötzer, 2001; Primack and Miller-Rushing, 2009), where identical clones of trees were transplanted to different climates and long time-series of phenological observations are available. In fact, such data has been used, for example to show that the transferability of model parameter estimates within a genotype is high (Chuine et al., 2000a), and to calibrate the very complex DormPhot model for *Betula* (Caffarra et al., 2011). Unfortunately, most data from experiments is limited to only few provenances and may thus, even be less suited for spatial extrapolation than in-situ observations.

4.2. Methodical considerations

4.2.1. Temperature data

The quality of the driving temperature data may have influenced model calibration and thus, the validation among sites. The gridded climate dataset used for this study is missing the exact local temperature conditions actually perceived by trees (despite the lapse-rate correction applied). Yet, even precise weather station data from the vicinity of the phenological observation site may not reflect the actual temperature trees or buds experience, yet the situation is certainly better for tall trees than for low stature vegetation (Kollas et al., 2014). Furthermore, for practical reasons, e.g. data availability, data handling and computation time, only daily mean temperature, rather than higher resolution temperature data, have been applied to calibrate and run the models. This neglects the influence of the diurnal amplitude of the temperature, known to affect phenology in several species (Campbell and Sugano, 1979; Erez and Couvillon, 1987; Myking, 1997; Partanen et al., 1998; Saxe et al., 2001).

4.2.2. Phenological time-series

The phenological data used for the parameterisation and validation of the models were obtained from several national phenological networks across Europe, which used historically different observation protocols. The data collected by many different observers are likely not free from flaws, although the robust error detection scheme (Schaber and Badeck, 2002) applied to the time-series, should have exclude the most unlikely data from the dataset. Unfortunately, no measure of the accuracy of the remaining individual observations is available.

5. Conclusions

The results obtained here show that it is not possible to prioritise a single mechanistic model as possible candidate for the actual mechanism of dormancy release. Yet, the relatively similar performance of all 35 process based model assessed here, indicates that the model have the structural flexibility to reproduce the weather-induced inter-annual variation of spring phenology, even though their intrinsic mechanistic assumption or parameterisation may not reflect actual physiological responses that control dormancy release. The transition from an empirically fitted model to a model that accurately describes the physiological processes is gradual and largely depends on the level of prior knowledge to reasonably restrict parameter ranges. There are yet no clear physiological or molecular markers that would allow to clearly separate and model each phase separately, e.g., the fulfilment of the chilling requirement (Cooke et al., 2012). For most time-series in central Europe, simple models (simulating the ecodormancy release only) yield reasonable prediction for the concurrent climate when calibrated with long phenological time-series. The simulation of endodormancy release or even of the whole period of dormancy in a phenological model is challenged by at least three, partially linked aspects: first, the accurate mechanistic understanding of dormancy induction and release, second their correct representation in models, and third, model calibration (overfitting) and data availability. The understanding of the underlying physiological mechanisms is essential, when making predictions for the future climates beyond current observational data coverage. Thus, for making such predictions I suggest running an ensemble of selected simple (ecodormancy) and more complex (endo-ecodormancy) models, until the underlining processes are fully revealed and accurately represented in a model.

com) and the data providers in the ECA&D project (http://eca.knmi.nl).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.agrformet.2015.11.007.

References

Basler, D., Körner, C., 2012. Photoperiod sensitivity of bud burst in 14 temperate forest tree species. Agric. For. Meteorol. 165 (0), 73–81.

Basler, D., Körner, C., 2014. Photoperiod and temperature responses of bud swelling and bud burst in four temperate forest tree species. Tree Physiol. 34, 377–388.

Battey, N.H., 2000. Aspects of seasonality. J. Exp. Bot. 51 (352), 1769–1780.

Blümel, K., Chmielewski, F.M., 2012. Shortcomings of classical phenological forcing models and a way to overcome them. Agric. For. Meteorol. 164 (0), 10–19.

Caffarra, A., Donnelly, A., 2010. The ecological significance of phenology in four different tree species: effects of light and temperature on bud burst. Int. J. Biometeorol. 55 (5), 711–721.

Caffarra, A., Donnelly, A., Chuine, I., 2011. Modelling the timing of Betula pubescens budburst. II. Integrating complex effects of photoperiod into process-based models. Clim. Res. 46 (2), 159–170.

Campbell, R.K., Sugano, A.I., 1975. Phenology of bud burst in Douglas-fir related to provenance, photoperiod, chilling, and flushing temperature. Bot. Gaz. 136 (3), 290–298.

Campbell, R.K., Sugano, A.I., 1979. Genecology of bud-burst phenology in Douglas-fir: response to flushing temperature and chilling. Bot. Gaz. 140 (2), 223–231.

Cannell, M.G.R., Smith, R.I., 1983. Thermal time, chill days and prediction of budburst in Picea sitchensis. J. Appl. Ecol. 20 (3), 951–963.

Chmielewski, F.M., Rötzer, T., 2001. Response of tree phenology to climate change across Europe. Agric. For. Meteorol. 108 (2), 101–112.

Chuine, I., 2000. A unified model for budburst of trees. J. Theor. Biol. 207 (3), 337–347.

Chuine, I., Beaubien, E.G., 2001. Phenology is a major determinant of tree species range. Ecol. Lett. 4 (5), 500–510.

Chuine, I., Belmonte, J., Mignot, A., 2000a. A modelling analysis of the genetic variation of phenology between tree populations. J. Ecol. 88 (4), 561–570.

Chuine, I., Cambon, G., Comtois, P., 2000b. Scaling phenology from the local to the regional level: advances from species-specific phenological models. Glob. Change Biol. 6 (8), 943–952.

Chuine, I., Cour, P., Rousseau, D.D., 1998. Fitting models predicting dates of flowering of temperate-zone trees using simulated annealing. Plant Cell Environ. 21 (5), 455–466.

Chuine, I., Cour, P., Rousseau, D.D., 1999. Selecting models to predict the timing of flowering of temperate trees: implications for tree phenology modelling. Plant Cell Environ. 22 (1), 1–13.

Cooke, J.E.K., Eriksson, M.E., Junttila, O., 2012. The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. Plant Cell Environ. 35 (10), 1707–1728.

Coville, F.C., 1920. The influence of cold in stimulating the growth of plants. Proc. Natl. Acad. Sci. USA 6, 434–435.

Črepinšek, Z., Kajfež-Bogataj, L., Bergant, K., 2006. Modelling of weather variability effect on fitophenology. Ecol. Model. 194 (1–3), 256–265.

de Réaumur, R.A.F., 1735. Observations du thermomètres, faites a Paris pendant l'année 1735, comparées avec celles qui ont été faites sous la ligne, a l'isle de France, a Alger et quelques unes de nos isles de l'Amérique. Acad. R. Sci. 545, 545–576.

Delpierre, N., Dufrene, E., Soudani, K., Ulrich, E., Cecchini, S., Boe, J., Francois, C., 2009. Modelling interannual and spatial variability of leaf senescence for three deciduous tree species in France. Agric. For. Meteorol. 149 (6–7), 938–948.

Doorenbos, J., 1953. Review of the literature on dormancy in buds of woody plants. Medelingen Landbouwhogeschool Wageningen/Nederland 53, 1–24.

Erez, A., Couvillon, G.A., 1987. Characterization of the influence of moderate temperatures on rest completion in Peach. J. Am. Soc. Hortic. Sci. 112 (4), 677–680.

Fishman, S., Erez, A., Couvillon, G.A., 1987a. The temperature-dependence of dormancy breaking in plants: computer-simulation of processes studied under controlled temperatures. J. Theor. Biol. 126 (3), 309–321.

Fishman, S., Erez, A., Couvillon, G.A., 1987b. The temperature-dependence of dormancy breaking in plants: mathematical-analysis of a 2-step model involving a cooperative transition. J. Theor. Biol. 124 (4), 473–483.

Granhus, A., Floistad, I.S., Sogaard, G., 2009. Bud burst timing in Picea abies seedlings as affected by temperature during dormancy induction and mild spells during chilling. Tree Physiol. 29 (4), 497–503.

Gunderson, C.A., Edwards, N.T., Walker, A.V., O'Hara, K.H., Campion, C.M., Hanson, P.J., 2012. Forest phenology and a warmer climate: growing season extension in relation to climatic provenance. Glob. Change Biol. 18, 2008–2025.

Häkkinen, R., 1999. Statistical evaluation of bud development theories: application to bud burst of Betula pendula leaves. Tree Physiol. 19 (9), 613–618.

Häkkinen, R., Linkosalo, T., Hari, P., 1998. Effects of dormancy and environmental factors on timing of bud burst in Betula pendula. Tree Physiol. 18 (10), 707–712.

Hannerz, M., 1999. Evaluation of temperature models for predicting bud burst in Norway spruce. Can. J. For. Res. 29 (1), 9–19.

Hänninen, H., 1990. Modelling bud dormancy release in trees from cool and temperate regions. Acta For. Fennica 213, 1–47.

Hänninen, H., 1995. Effects of climatic-change on trees from cool and temperate regions: an ecophysiological approach to modeling of bud burst phenology. Can. J. Bot. 73 (2), 183–199.

Hänninen, H., Kramer, K., 2007. A framework for modelling the annual cycle of trees in boreal and temperate regions. Silva Fenn. 41 (1), 167–205.

Haylock, M.R., Hofstra, N., Tank, A.M.G.K., Klok, E.J., Jones, P.D., New, M., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J. Geophys. Res. Atmos. 113 (D20119), 1–12.

Hunter, A.F., Lechowicz, M.J., 1992. Predicting the timing of budburst in temperate trees. J. Appl. Ecol. 29 (3), 597–604.

Ingber, L., 1993. Adaptive Simulated Annealing (ASA) Global Optimization C-code. Caltech Alumni Association, Pasadena, CA.

Jochner, S., Caffarra, A., Menzel, A., 2013. Can spatial data substitute temporal data in phenological modelling? A survey using birch flowering. Tree Physiol. 33 (12), 1256–1268.

Junttila, O., Hänninen, H., 2012. The minimum temperature for budburst in Betula depends on the state of dormancy. Tree Physiol. 32 (3), 337–345.

Kobayashi, K.D., Fuchigami, L.H., 1983. Modeling bud development during the quiescent phase in red-osier Dogwood (Cornus Sericea L.). Agric. Meteorol. 28 (1), 75–84.

Kollas, C., Randin, C.F., Vitasse, Y., Körner, C., 2014. How accurately can minimum temperatures at the cold limits of tree species be extrapolated from weather station data? Agric. For. Meteorol. 184 (15), 257–266.

Körner, C., Basler, D., 2010. Phenology under global warming. Science 327 (5972), 1461–1462.

Kramer, K., 1994. Selecting a model to predict the onset of growth of Fagus sylvatica. J. Appl. Ecol. 31 (1), 172–181.

Kriebel, H.B., Wang, C.W., 1962. The interaction between provenance and degree of chilling in bud-break of sugar maple. Silvae Genet. 11 (5/6), 125–130.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., Prentice, I.C., 2005. A dynamic global vegetation model for studies of the coupled atmosphere–biosphere system. Glob. Biogeochem. Cycles 19 (1), GB1015.

Landsberg, J.J., 1974. Apple fruit bud development and growth: analysis and an empirical model. Ann. Bot. 38 (158), 1013–1023.

Lang, G.A., Early, J.D., Martin, G.C., Darnell, R.L., 1987. Endo-, para-, and ecodormancy: physiological terminology and classification for dormancy research. HortScience 22 (3), 371–377.

Leinonen, I., Kramer, K., 2002. Applications of phenological models to predict the future carbon sequestration potential of boreal forests. Clim. Change 55 (1–2), 99–113.

Lindsley-Noakes, G.C., Louw, M., Allan, P., 1995. Estimating daily positive Utah Chill Units using daily minimum and maximum temperatures. J. S. Afr. Soc. Hortic. Sci. 5, 19–23.

Linkosalo, T., Carter, T.R., Häkkinen, R., Hari, P., 2000. Predicting spring phenology and frost damage risk of Betula spp. under climatic warming: a comparison of two models. Tree Physiol. 20 (17), 1175–1182.

Linkosalo, T., Häkkinen, R., Hänninen, H., 2006. Models of the spring phenology of boreal and temperate trees: is there something missing? Tree Physiol. 26 (9), 1165–1172.

Linkosalo, T., Lechowicz, M.J., 2006. Twilight far-red treatment advances leaf bud burst of silver birch (Betula pendula Roth). Tree Physiol. 26, 1249–1256.

Linkosalo, T., Lappalainen, H.K., Hari, P., 2008. A comparison of phenological models of leaf bud burst and flowering of boreal trees using independent observations. Tree Physiol. 28, 1873–1882.

Masle, J., Doussinault, G., Farquhar, G.D., Sun, B., 1989. Foliar stage in wheat correlates better to photothermal time than to thermal time. Plant Cell Environ. 12 (3), 235–247.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21 (6), 1087–1092.

Migliavacca, M., Sonnentag, O., Keenan, T.F., Cescatti, A., O'Keefe, J., Richardson, A.D., 2012. On the uncertainty of phenological responses to climate change, and implications for a terrestrial biosphere model. Biogeosciences 9 (6), 2063–2083.

Murray, M.B., Cannell, M.G.R., Smith, R.I., 1989. Date of budburst of 15 tree species in Britain following climatic warming. J. Appl. Ecol. 26 (2), 693–700.

Myking, T., 1997. Effects of constant and fluctuating temperature on time to budburst in Betula pubescens and its relation to bud respiration. Trees-Struct. Funct. 12 (2), 107–112.

Partanen, J., Koski, V., Hänninen, H., 1998. Effects of photoperiod and temperature on the timing of bud burst in Norway spruce (Picea abies). Tree Physiol. 18 (12), 811–816.

Perry, T.O., 1971. Dormancy of trees in winter. Science 171 (3966), 29–36.

Primack, R.B., Laube, J., Gallinat, A.S., Menzel, A., 2015. From observations to experiments in phenology research: investigating climate change impacts on trees and shrubs using dormant twigs. Ann. Bot.

Primack, R.B., Miller-Rushing, A.J., 2009. The role of botanical gardens in climate change research. New Phytol. 182 (2), 303–313.

Olsson, C., Jönsson, A.M., 2014. Process-based models not always better than empirical models for simulating budburst of Norway spruce and birch in Europe. Glob. Change Biol. 20, 3492–3507.

R Development Core Team, 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Richardson, A.D., Keenana, T.F., Migliavacca, M., Ryu, Y., Sonnentag, O., Toomey, M., 2013. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. Agric. For. Meteorol. 169 (0), 156–173.

Richardson, E.A., Seeley, S.D., Walker, D.R., 1974. A model for estimating the completion of rest for 'Redhaven' and 'Elberta' peach trees. HortScience 9, 331–333.

Salmela, M.J., Cavers, S., Cottrell, J.E., Iason, G.R., Ennos, R.A., 2013. Spring phenology shows genetic variation among and within populations in seedlings of Scots pine (*Pinus sylvestris* L.) in the Scottish Highlands. Plant Ecol. Divers. 6 (3–4), 523–536.

Saxe, H., Cannell, M.G.R., Johnsen, B., Ryan, M.G., Vourlitis, G., 2001. Tree and forest functioning in response to global warming. New Phytol. 149 (3), 369–399.

Schaber, J., Badeck, F.W., 2002. Evaluation of methods for the combination of phenological time series and outlier detection. Tree Physiol. 22 (14), 973–982.

Schaber, J., Badeck, F.W., 2003. Physiology-based phenology models for forest tree species in Germany. Int. J. Biometeorol. 47 (4), 193–201.

Schnelle, F., Volkert, E., 1974. International Phenological Gardens in Europe: the basic network for international phenological observations. In: Lieth, H. (Ed.), Phenology and Seasonality Modeling. Ecological Studies. Springer, Berlin, Heidelberg, pp. 383–387.

Siebert, S., Ewert, F., 2012. Spatio-temporal patterns of phenological development in Germany in relation to temperature and day length. Agric. For. Meteorol. 152, 44–57.

Thomas, B., Vince-Prue, D., 1997. Photoperiodism in Plants. Academic Press, London, pp. 428.

Vaartaja, O., 1959. Evidence of photoperiodic ecotypes in trees. Ecol. Monogr. 29 (2), 91–111.

Vegis, A., 1964. Dormancy in higher plants. Annu. Rev. Plant Phys. 15, 185–224.

Vitasse, Y., 2013. Ontogenic changes rather than difference in temperature cause understory trees to leaf out earlier. New Phytol. 198 (1), 149–155.

Vitasse, Y., Basler, D., 2014. Is the use of cuttings a good proxy to explore phenological responses of temperate forests in warming and photoperiod experiments? Tree Physiol. 34 (2), 174–183.

Vitasse, Y., Hoch, G., Randin, C.F., Lenz, A., Kollas, C., Scheepens, J.F., Körner, C., 2013. Elevational adaptation and plasticity in seedling phenology of temperate deciduous tree species. Oecologia 171 (3), 663–678.

von Wuehlisch, G., Krusche, D., Muhs, H.J., 1995. Variation in temperature sum requirement for flushing of beech provenances. Silvae Genet. 44 (5–6), 343–346.

Wang, J.Y., 1960. A critique of the heat unit approach to plant-response studies. Ecology 41 (4), 785–790.

Wareing, P.F., 1956. Photoperiodism in woody plants. Annu. Rev. Plant Phys. 7, 191–214.

Weinberger, J.H., 1950. Chilling requirements of peach varieties. Proc. Am. Soc. Hortic. Sci. 56 (DEC), 122–128.

White, M.A., Thornton, P.E., Running, S.W., 1997. A continental phenology model for monitoring vegetation responses to interannual climatic variability. Glob. Biogeochem. Cycles 11 (2), 217–234.

Wolkovich, E.M., Cook, B.I., Allen, J.M., Crimmins, T.M., Betancourt, J.L., Travers, S.E., Pau, S., Regetz, J., Davies, T.J., Kraft, N.J.B., Ault, T.R., Bolmgren, K., Mazer, S.J., McCabe, G.J., McGill, B.J., Parmesan, C., Salamin, N., Schwartz, M.D., Cleland, E.E., 2012. Warming experiments underpredict plant phenological responses to climate change. Nature 485 (7399), 494–497.