Title: Representing explicit budburst and senescence processes for evergreen conifers in global models

Running title: A coniferous phenology model for global predictions

Authors : Marc Peaucelle[1], Philippe Ciais[1], Fabienne Maignan[1], Manuel Nicolas[2], Sébastien Cecchini[2], Nicolas Viovy[1]

[1] Laboratoire des Sciences du Climat et de l'Environnement, CEA CNRS UVSQ, Gif-sur-Yvette, France
[2] Office National des Forêts, Département Recherche Développement et Innovation, Fontainebleau, France

Marc Peaucelle: marc.peaucelle@lsce.ipsl.fr (corresponding author, phone:+33 169 088 739 /fax:+33 169 087 716)
Philippe Ciais: philippe.ciais@lsce.ipsl.fr
Fabienne Maignan : fabienne.maignan@lsce.ipsl.fr
Manuel Nicolas : manuel.nicolas@onf.fr
Sébastien Cecchini : sebastien.cecchini@onf.fr
Nicolas Viovy: nicolas.viovy@lsce.ipsl.fr

Key words: tree phenology, needle budburst, needle senescence, evergreen needleleaves, ecosystem model, RENECOFOR

Type of paper: Primary research article

Abstract:

Global ecosystem models lack an explicit representation of budburst and senescence for evergreen conifers despite their primordial role in the carbon cycle. In this study we evaluated eight different budburst models, combining forcing, chilling and photoperiod, for their ability to describe spring budburst, and one model of needle senescence for temperate evergreen coniferous forests. The models' parameters were optimized against field observations from a national forest monitoring network in France. The best fitting budburst model was determined according to a new metrics which accounts for both temporal and spatial variabilities of budburst events across sites. The best model could reproduce observed budburst dates both at the site scale (±5 days) and at regional scale (±12 days). We also showed that the budburst models parameterized at site scale lose some predictive capability when applied at coarser spatial resolution, e.g., in grid-based simulations. The selected budburst model was then coupled to a senescence function defined from needle survivorship observations in order to describe the full phenology cycle of coniferous forests. Implemented in the process-driven ecosystem model ORCHIDEE, this new conifer phenology module represented accurately the intra and inter-annual dynamics of leaf area index at both the local and regional scales when compared against MODIS remote sensing observations. A sensitivity analysis showed only a small impact of the new budburst model on the timing of the seasonal cycle of photosynthesis (GPP). Yet, due to the faster renewal of needles compared to the standard version of ORCHIDEE, we simulated an increase in the GPP by on average 15% over France, while the simulated needle turnover was doubled. Compared to 1970-2000, projections indicated an advancement of the budburst date of 10.3±2.8 and 12.3±4.1 days in average over the period 2060-2100 with the best forcing and chilling-forcing models respectively. Our study suggests that including an explicit simulation of needle budburst and senescence for evergreen conifers in global terrestrial ecosystem models may significantly impact future projections of carbon budgets.

1. **Introduction**

The phenology of conifers is strongly correlated with local climate (Steiner, 1980; Worrall, 1983; Burr *et al.*, 1989; Leinonen & Hänninen 2002; Hänninen *et al.*, 2007). A number of previous studies concluded that the growing-season length of conifer forests will extend with climate warming and rising CO_2 concentration, thus leading to significant modifications of biogeochemical processes being controlled by phenology (Murray *et al.*, 1994; Polgar & Primack, 2011; Gunderson *et al.*, 2012; Migliavacca *et al.*, 2012; Richardson *et al.*, 2013). For both deciduous and evergreen species, phenology is commonly divided into three different phases: bud dormancy, bud flush and senescence. The timing of these events partly controls the seasonal cycle of leaf area index and gross primary productivity (GPP; Chen *et al.*, 2016). It also impacts albedo, evapotranspiration, and litter inputs to the soil, of which the latter affects soil respiration (Richardson *et al.*, 2013). On longer time scales, phenology also impacts the competitiveness of a species and its spatial distribution (Baldocchi *et al.*, 2001; Chuine & Beaubien, 2001; Polgar & Primack, 2011). These impacts make it essential to represent phenological events accurately in ecosystem models, both in space and in time, if we seek to improve the simulation of the future role of vegetation in carbon, water and nutrient cycling and its feedbacks on climate. Richardson *et al.* (2012) pointed out to shortcomings in the representation of phenological processes in global vegetation models. Almost all of the 14 vegetation models with different phenology parameterizations that they compared overestimated the length of the growing season and consequently the GPP for temperate and boreal forests.

The state of the understanding of complex molecular pathway processes of dormancy and budburst mechanisms is insufficient to allow a fully mechanistic simulation in global models (Rohde & Bhalerao, 2007; Yakovlev *et al.*, 2008; Rinne *et al.*, 2011; Cooke *et al.*, 2012). Thus, empirical models are used for estimating the response of budburst to temperature. Several conceptual models using temperature to determine the date of budburst have been proposed: they fall into two broad categories. The first category assumes that budburst occurs after a threshold of degree-days has

been reached during a specific period (forcing). The second class of models assumes that budburst requires both a chilling period during winter followed by a forcing from increasing temperature. We investigated models belonging to both these categories. For conifers, photoperiod, in combination with temperature, has also been proposed as a controlling variable, particularly for boreal regions where the chilling requirement can be quickly reached and photoperiod acts as a safety limitation to prevent a too early budburst and plant exposure to frost (Richardson *et al.*, 2013; Way and Montgomery, 2015).

Empirical models for budburst are generally derived from local meteorological data and observed budburst timing, and mostly focus on deciduous species. The direct use of a model calibrated on a site for gridded simulations over a region can be a source of errors, for example because of altitudinal differences not resolved at a given grid horizontal resolution or because the whole range of temperatures was not taken into account during the calibration (Olsson *et al.*, 2013). This problem calls for a multi-site calibration of budburst models with data drawn from a wide area, typical of that used in the grid-based applications of ecosystem models.

Regarding the mechanisms involved, past studies did not highlight phenological differences between deciduous and evergreen conifers for budburst. However, compared to deciduous species that shed their leaves in autumn, evergreen conifers keep most of their needles over the year. Needle lifespan can span from 2-3 years (e.g. *Pinus Sylvestris*) to more than 10 years (e.g. *Picea abies*) for evergreen conifers. Needle senescence has been less studied (Estrella & Menzel, 2006; Delpierre *et al.*, 2009) than budburst. Some authors observed a peak of senescence during autumn (Sampson *et al.*, 2003; Kivimäenpää & Sutinen, 2007; Wang & Chen, 2012), however integrated over all needle cohorts, needle senescence can be seen as a continuous process in evergreen species. Reich *et al.* (2014) showed the role of needle longevity (related to nitrogen content) and the impacts of needle senescence on carbon cycling in boreal forest, but very few studies investigated the regulation of needle yellowing and turnover (renewal rate of needles), and none of them proposed any mechanistic model for needle senescence (Muukkonen, 2005; Kayama *et al.*, 2007).

While current global vegetation models roughly simulate phenological events for deciduous species, these processes are still lacking for evergreen species for which the common approach is to represent phenology implicitly through leaf biomass variations.

In this study, we tackle the following objectives:

- To calibrate empirical budburst models for temperate needleleaved species in order to reproduce field observations collected in forest monitoring plots for a range of contrasting climate conditions.

- To evaluate the accuracy of these models when used at low spatial resolution (0.25 and 0.5°) typical of global models and to test the simulation results against independent remote sensing observations.

- To implement a model for needle senescence on the basis of litterfall observations.

- To evaluate the potential impact of these model developments in a global vegetation model (ORCHIDEE - Organizing Carbon and Hydrology In Dynamic Ecosystems; Krinner *et al.*, 2005) on the representation of forest canopies and the associated carbon balance simulated for temperate needleleaved forests in France.

2. **Materials and methods**

We used budburst observations and litterfall samples collected from 1997 to 2011 over the 51 forest plots of the French RENECOFOR (REseau National de suivi à long terme des ECOsystèmes FORestiers) network covering the six main coniferous species in France (number of plots in parentheses): *Pseudotsuga mensiezii* (Douglas fir; 6), *Picea abies* (Norway spruce; 11), *Pinus nigra* (Corsican pine; 2), *Pinus pinaster* (Maritime pine; 7), *Pinus sylvestris* (Scots pine; 14) and *Abies alba* (Silver fir; 11). The parameters of eight different budburst models were calibrated against the RENECOFOR site observations to get the best value for a performance metrics defined specifically to account for both temporal and spatial variabilities. We selected the models that best described budburst for the temperate evergreen needleleaf plant functional type (PFT) as a whole, but also separately for each species. In addition, a new senescence model based on needle age was developed and calibrated based on a literature review of needle survivorship observations. The new phenology module (budburst+senescence) was then incorporated into the process-based model, ORCHIDEE, and evaluated spatially against leaf area index estimated from remote sensing observations. The flow chart of the model calibration and evaluation is given in Fig. 1.

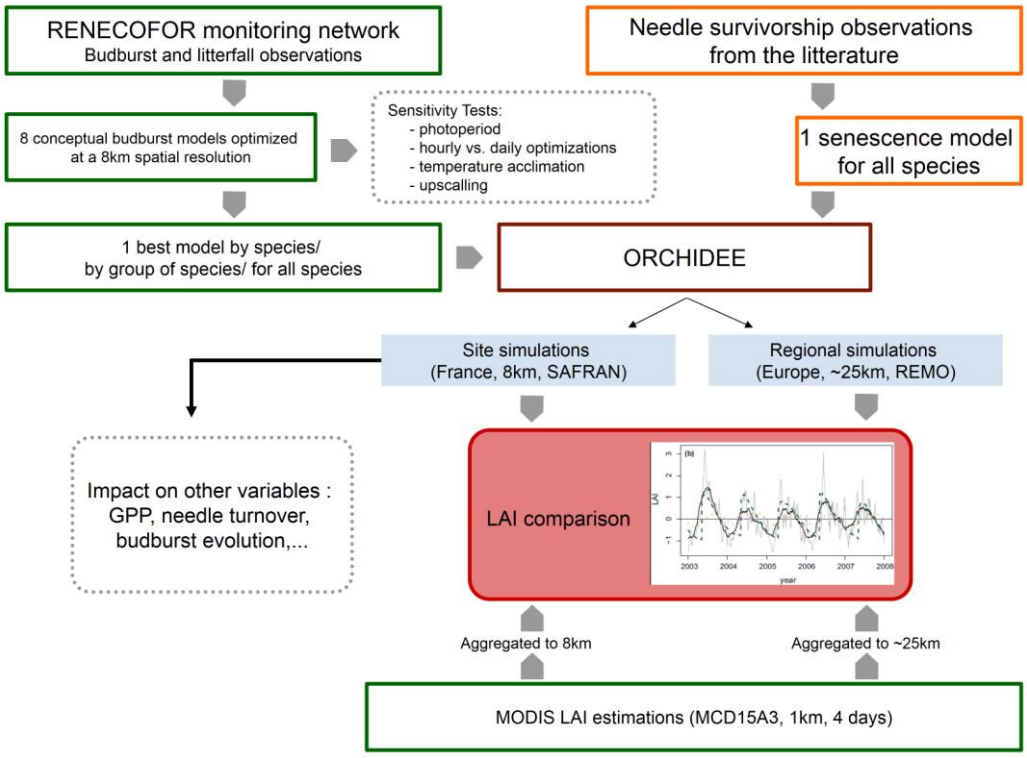
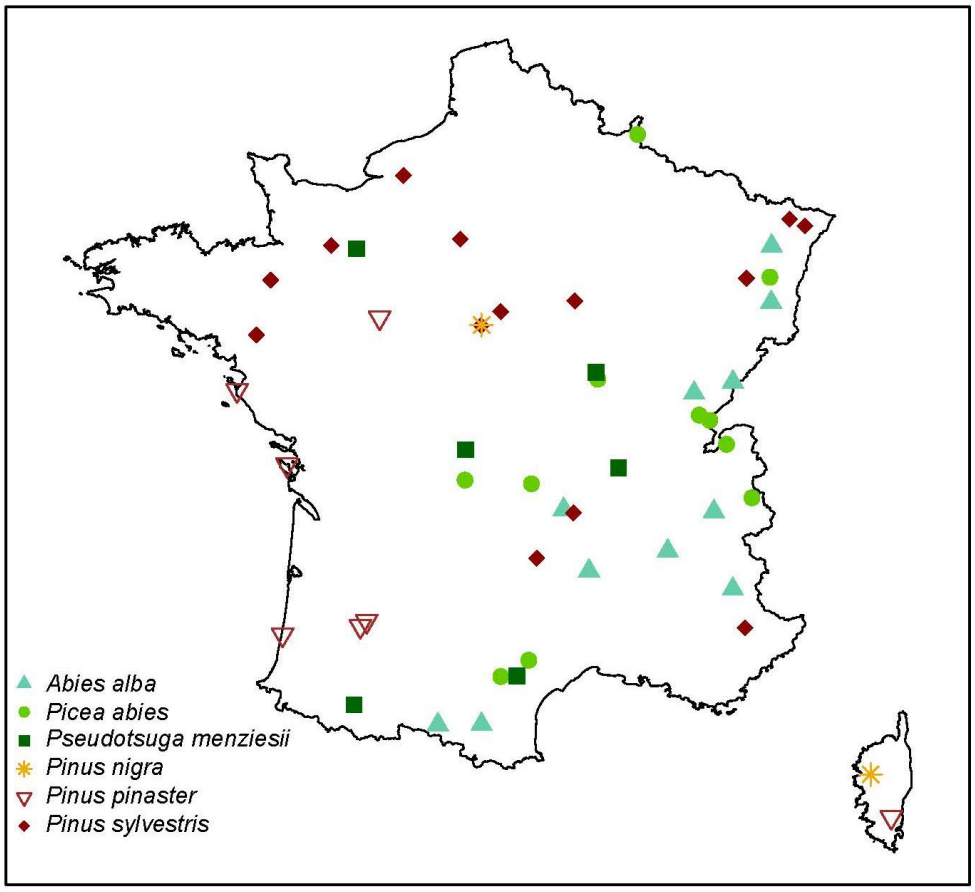


Fig. 1: Flow chart of the model calibration and evaluation

146
147 2.1. *Data from the RENECOFOR forest monitoring network*

148
149 Visual observations of the timing of budburst were collected in 51 plots located in public forests
150 and stratified according to the major commercial tree species grown in France. Those plots are part
151 of the French national long-term forest monitoring network (RENECOFOR, part of the ICP –
152 International and Co-operative Program - Forest Level II program (http://icp-forests.net/page/level-
153 ii)), which covers a total of 102 permanent plots (51 coniferous) monitored since 1992 (Fig. 2).



154
155 **Fig. 2**: Distribution throughout France of the 51 RENECOFOR permanent plots dominated by
156 evergreen coniferous tree species.
157

158 Plots are in average 70y old (in 1994) with a range from 23 to 181y old. At each location, 36 trees
159 were chosen for phenological observations (Lebourgeois *et al.*, 2010), and observations were
160 performed at least every week. For budburst, two different dates were measured. The first one is the
161 day of the year when 10% of the trees have open buds for at least 20% of the crown (BD1). The
162 second date corresponds to the day of the year when 90% of the trees have open buds for at least

20% of the crown (BD9). Observations were performed each year from 1997 to 2011. We approximated a mean bud flushing date (50%; BD5) for budburst model parameterization defined as:

$$BD5 = BD1 + \frac{BD9 - BD1}{2} \quad (1)$$

The final dataset contains 605 site-years of observations. Litterfall was collected seasonally from 1995 to 2007 at all RENECOFOR plots using litter traps. The detailed litter sampling methodology is described in the Supplementary material (Appendix SA).

2.2. *Budburst models*

We tested two types of model for mono-cyclical budburst events, based on a temperature forcing (e.g., degree-days) during spring, or based on "chilling-forcing", i.e., with a chilling during winter and a forcing period during spring. We used the model M1 as reference (called the static or null model) in which budburst equals the median of the budburst dates observed across all sites (Table 1). A model with predictive value (spatially or temporally) should have better performance than M1.

Species	**Mean BD5 (SD)**	**Min BD5**	**Max BD5**	**n**
Abies alba	137.5 (11.5)	109.0	175.5	148
Picea abies	136.8 (11.5)	107.5	169.5	132
Pseudotsuga menziesii	130.3 (10.9)	102.5	151.0	72
Pinus nigra	127.4 (14.7)	106.0	163.5	20
Pinus pinaster	119.6 (18.2)	81.5	164.5	77
Pinus sylvestris	129.5 (13.3)	97.0	175.0	156

Table 1: Mean, minimum and maximum observed budburst date (DoY) in the RENECOFOR network over the period 1997-2011. n = number of site-year observations

Eight models (M2-M9) were taken from the literature and tested (Cannell & Smith, 1983; Hänninen, 1990; Kramer, 1994; Chuine *et al.*, 1998; Chuine, 2000; Harrington *et al.*, 2010; Vitasse *et al.*, 2011). The equations and parameters of each model are detailed in Appendix SB. All models

185 are representative of central hypotheses in budburst modeling and all of them except M9 were
186 already applied at regional scales. The models were selected to represent different concepts, but also
187 for their number of parameters to be small enough to allow inclusion in a global vegetation model.
188 The list of models tested has two spring warming forcing models and six chilling-forcing models
189 (Table 2).
190 In addition, we performed sensitivity tests to investigate the role of photoperiod, temperature
191 acclimation of parameters and the use of hourly or daily meteorological data (see Appendix SA for
192 the results of the sensitivity tests).
193

	Model	Model name and reference	Number of parameters (without t_o, t_c and $t_{c,end}$)
Spring forcing models	**M1**	Reference model : median value of budburst dates observed in the field	0
	M2	*Thermal time model*: M2 – Kramer, 1994	1
	M3	*UniForc*: M3 – Chuine *et al.*, 1998	2
Chilling-forcing models	**M4**	*Sequential model*: M4 – Hänninen, 1990; Kramer, 1994	5
	M5	*Parallel model* M5 – Cannell & Smith, 1983; Kramer, 1994	6
	M6	*Unified model* M6 – Chuine, 2000	5
	M7	*Alternating model* M7 – Kramer, 1994;	2
	M8	*Logistic alternating model* M8 – Vitasse *et al.*, 2011	3
	M9	*Harrington model* M9 – Harrington *et al.* 2010	4
	Photoperiod	The daily forcing temperature calculated in models M2-M8 is weighted by the day length following Blümel & Chmielewski (2012)	+ 1

194
195 **Table 2:** Names and references of the eight budburst thermal models optimized in this study and the
196 corresponding number of optimized parameters when the starting date for chilling (t_c), forcing (t_o)
197 and the ending date for chilling (t_{cend}) are fixed. See Appendix SB for a full description of the model
198 parameters and their equations. Note that critical temperature thresholds are not optimized in this
199 study but are estimated from the observed budburst dates. For example in M2 for which classical
200 optimization studies optimized both the base temperature (*Tb*) and the critical forcing threshold
201 (*Fcrit*) for budburst, here only *Tb* is optimized, *Fcrit* being the median *Fcrit* simulated at each
202 observed budburst date with M2.

2.3. *Budburst model selection criterion*

The purpose of this study being to calibrate a budburst model for regional simulations, we needed a specific performance metric that characterizes the ability of this model to capture both spatial and temporal budburst gradients. Most studies have used root mean square error (RMSE, Eqn 2) or linear regression between simulated and observed budburst dates to select the best phenological model (Vitasse *et al.*, 2011; Olsson *et al.*, 2013). We argue that two performance criteria are desirable: a representation of both spatial and temporal extremes of budburst across a region, and a representation with minimal systematic spatial or temporal bias. A single metric is insufficient to account for these two criteria. To select the best set of parameters for each model, we thus propose a combined metric accounting for three key aspects of the model: the ability to reproduce extreme values, the average bias to observations and the effect of outliers. This new criterion is the Euclidean distance to optima (DIST, Eqn 3) of six weighted different metrics across sites and years. This new metrics maximizes model performances by catching both spatial and temporal variabilities of budburst and by reducing the chance to converge to local minimum during calibration. This combined metric has two components of the RMSE to limit outliers, the squared bias (BS, Eqn 4) and the squared difference between standard deviation (SDSD, Eqn 5; Kobayashi & Salam, 2000) normalized by the mean natural difference observed between sites $SDSD_{nat}$ and BS_{nat}; the Spearman rank correlation coefficient (R_s), the linear regression slope (lm_s; Eqn 6) and its associated coefficient of determination (lm_{R2}) between observed and modeled budburst, which captures extreme values; finally the average bias (AB, Eqn 7), and temporal and spatial biases between modeled and observed budburst dates. For the purpose of large scale simulations, more weight was given to capture bias and extremes than outliers (Eqn 3), with the sum of weighting factors equaling 1. With this metrics, the best performance is achieved when DIST=0. The different components of DIST are:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - y_t)^2} \quad (2)$$

$$DIST = \sqrt{\left(0.05\left(\frac{SDSD}{SDSD_{nat}}\right)^2 + 0.05\left(\frac{BS}{BS_{nat}}\right)^2 + 0.2(R_s - 1)^2 + 0.1(lm_s - 1)^2 + 0.1(lm_{R2} - 1)^2 + 0.5AB^2\right)} \quad (3)$$

$$BS = (\bar{x} - \bar{y})^2 \quad (4)$$

$$SDSD = (SD_x - SD_y)^2 \quad (5)$$

$$y = cste + lm_s x \quad (6)$$

$$AB = \frac{1}{n}\sum_{t=1}^{n}\frac{(y_t - x_t)}{x_t} \quad (7)$$

with *x* and *y* the observed and simulated budburst date, *n* the number of observations *t*, and SD the standard deviation.

$SDSD_{nat}$ and BS_{nat} were defined as the SDSD between sites in similar conditions and represent variability in the observed data. Pairs of sites being within 0.5° (~55 km) maximum distance, with a mean annual temperature difference less than 0.5 °C, are considered as having "similar conditions". Here, species similarity was not specified as a "similar condition", in order to have enough observations. Thus, we hypothesized that $SDSD_{nat}$ and BS_{nat} are caused by species differences but also non-resolved biotic or edaphic factors (local adaptation, age, soil effect, etc...), and define the smallest value that an optimized model should approach when considering all conifers species. The $RMSE_{nat}$ was estimated from nine sites with "similar conditions" to 7.6±3.5 days, $SDSD_{nat}$ to 3.16±5.09 days² and BS_{nat} to 41.76±38.46 days².

The optimization of the parameters of models based on cross-site spatial variability only (DIST-S averaging budburst years across site) resulted in a different best model to the one based on the temporal variability only (DIST-T averaging budburst dates across all the sites each year) or considering both the spatial and temporal variability (DIST-ST). Thus, we optimized below the parameters of the eight budburst models described in Table 2 and we selected as "best predictive

model" the one corresponding to the minimum value of the DIST-ST metrics.

2.4. *Budburst model optimization*

The parameters of each model were optimized to minimize the value of DIST-ST against a subset of the RENECOFOR observations (optimization dataset) consisting of 455 sites-years randomly selected from the full dataset, with at least one observation per site. The remaining 150 observations were used as cross-validation data. Note that models have different numbers of parameters (Table 2), i.e., different degrees of freedom. In addition to the cross-validation, the overall model accuracy was assessed by coupling DIST-ST results to the Akaike's information criterion corrected for sample size (AICc) in order to select the best predictive (DIST-ST) and parsimonious (AICc) model. Parsimonious models were selected by calculating the AICc difference (ΔAIC) between AICc and the minimal AICc obtained among all models. Thus, the higher is ΔAIC, the less parsimonious is a model. Models with ΔAIC higher than 10 were excluded (Burnham & Anderson, 2003).

We optimized models with a generalized simulated annealing algorithm (R package genSA; Chuine, 2000; Xiang *et al.*, 2013) considering parameters 1) per species, 2) grouping pines (*Pinus pinaster*, *Pinus sylvestris* and *Pinus nigra*) versus firs and spruces (grouping *Abies alba*, *Picea abies* and *Pseudotsuga mensiezii*, hereafter 'fir') into two groups according to Peaucelle *et al.* (2016), and 3) pooling all conifer species together. Models were also fitted site by site to assess possible emerging relationships between local parameter values and environmental conditions (Appendix SA). For chilling-forcing models (M6-M9), an exponential relationship between chilling units and forcing units is commonly used to estimate budburst. However, this exponential relationship is potentially an artifact (Chuine, 2000) and is not observed for all species. We thus decided to compare relationships fitted by exponential or by linear functions (Appendices SA & SB). Different optimizations were performed by fixing or by optimizing the starting date of forcing ($t_{o;}$ 1st of January) and chilling ($t_{c;}$ 1st of November) in order to assess the models' robustness with

fewer degrees of freedom. For M6 to M8 we assessed the impact of optimizing the end date for chilling accumulation ($t_{c,end}$), thus representing the fulfillment of the chilling requirement, otherwise chilling is summed until budburst (Vitasse *et al.* 2011).

2.5. *Meteorological dataset*

We used the SAFRAN (Système d'Analyse Fournissant des Renseignements Adaptés à la Nivologie) meteorological data (Vidal *et al.*, 2010) for model optimization and for ORCHIDEE site-scale simulations. This dataset produced by the Centre National de Recherches Météorologiques (CNRM) provides hourly weather data over France at a spatial resolution of 0.07° (8 km). At 0.07° resolution, each of the 51 coniferous forest sites is located in an independent grid cell. Once the best set of parameters was retrieved for each model, we assessed the effect of the spatial scale of climate data by applying the same models with SAFRAN data aggregated at 0.25° (~28 km) and 0.5° (~55 km) resolution respectively. At 0.25° resolution, 49 grid cells contained at least one site, and at 0.5°, 43 grid cells contained at least one site. All temperatures were corrected in a simple way for local altitude following Eqn 8 (U.S. Standard Atmosphere, 1976, Olsson & Jönsson, 2015):

$$T_{obs} = T_{saf} + 6.4\left(A_{saf} - A_{obs}\right) \quad (8)$$

where T_{obs} (°C) is the mean observed temperature of the site, T_{saf} (°C) the mean temperature of the site from SAFRAN dataset, A_{saf} (km) the mean altitude of the SAFRAN cell and A_{obs} (km) the altitude of the site.

2.6. *Senescence model*

The senescence model is not a stand-alone model as is the case for the model of budburst dates, but rather a modification of the ORCHIDEE original phenology described in Krinner *et al.* (2005). Compared to budburst models that are functions of environmental conditions, the senescence model

299 is based on needle age. The original version of ORCHIDEE includes two types of senescence for
300 needles. Firstly, a base rate of leaf mortality is applied each day (Krinner *et al.*, 2005). It represents
301 the probability for needles to fall independently of needle age or meteorological conditions.
302 Secondly, senescence is triggered when needle age (calculated for four cohorts as in Section A1 of
303 Krinner *et al.*, 2005) reaches a pre-defined longevity parameter for each PFT. As no phenological
304 process is explicitly defined in the default model, needle age is implicitly estimated from needle
305 biomass with the assumption that newly assimilated biomass through photosynthesis is used to
306 create new needles at the beginning of the year.
307 We did not find any suitable needle senescence model for coniferous species in the literature. We
308 thus decided to fit a senescence function against field observations of needle survival probability
309 from different studies (all studies and species are listed in Table SC1, Appendix SC). We retrieved
310 45 needle survivorship curves (determines the probability of needles to survive (0-1) over time
311 according to their age) from the literature and used these to calibrate a logistic function given by:

$$S(t) = \frac{1}{1 + e^{(4\mu(\lambda - t) + 2)}} \quad (9)$$

312 Where *S(t)* is the survivorship probability of a needle, *t* the needle age (days), and μ (days^{-1}) and λ
313 (days) parameters to be fitted on literature observations.
314
315 2.7. *Modification of the ORCHIDEE model*
316 The inclusion of an explicit phenology for evergreen conifers in ORCHIDEE (Krinner *et al.* 2005)
317 needed the modification of the original model. The needle maximum age parameter for evergreen
318 conifers, fixed at 910 days (average lifespan of *Pinus* needles) in the standard version of
319 ORCHIDEE was modified to depend on species. This maximum age can vary considerably, going
320 from 2 years in pine species to more than 10 years for *Abies alba* and *Picea abies* (Peaucelle *et al.*,
321 2016). In ORCHIDEE, the V_{cmax} parameter (maximal rate of the RUBISCO carboxylation activity
322 in µmol m^{-2} s^{-1}) increases with needle age, reaching a maximum value when the relative age of the
323 needle (the ratio of the needle age to its maximum) is 0.03 and then linearly decreasing to its

minimum value when the relative age reaches 0.5. This function describing the evolution of V_{cmax} roughly represents species with short-lived needles such as pines (Niinemets, 2002). Observations show a rapid decrease of V_{cmax} after only 1 year even in high needle-longevity species (Porté & Loustau, 1998; Niinemets, 2002; Warren, 2006). We thus adapted this relationship prescribing V_{cmax} to reach its maximum 3 months after formation and then starting to decrease linearly after one year, until reaching 0.5 V_{cmax} at the maximal needle age of the species. In following simulations, we used two different needle maximum age, 1275 and 2340 days for pines and spruces/firs species, respectively (Peaucelle *et al.*, 2016).

Given the senescence function in Eqn 9, we also modified the way carbon is distributed in the crown by representing each cohort of needle (*i.e.* groups of needles developed the same year). All the biomass gained during the current year is placed in the youngest needle cohort. Other cohorts do not receive new biomass, but lose needles according to the senescence function from Eqn 9. To exclude simulations where budburst never occurs, we imposed a maximum needle onset at day 182 (1 July), the latest observed budburst in our dataset.

2.8. *Validation of the phenology models*

Simulated budburst date was evaluated against observed data from the optimization and the cross-validation dataset. We also investigated the ability of each model to predict spatial and temporal variations in budburst across sites by comparing median modeled and observed budburst dates at each site, and by looking at the interannual variability in the timing of budburst at each site.

Because of the high cross-site variability, an evaluation of the senescence model could not be performed against litterfall observations, which also depend on non-modeled factors such as stand health, stand age, density, species composition and management events. Thus, observed and modeled litterfall were simply compared for information, and we validated the senescence model through the indirect comparison with satellite-derived leaf area index (Wang *et al.*, 2004). We compared LAI simulated at a spatial resolution of 0.07° with ORCHIDEE forced by SAFRAN

against MODIS MCD15A3 LAI (1 km, 4-day frequency; Myneni *et al.*, 2002). This local comparison between modeled and satellite LAI was performed on a few grid cells where the coverage of two representative coniferous species (*Picea abies* and *Pinus sylvestris*) exceeds 80% at 1 km, based on the European tree species map of Brus *et al.* (2012). The correlation coefficient between modeled and estimated LAI was used to assess modeled LAI seasonality.

We also performed grid-based simulations for coniferous forests in Europe at a 0.25° spatial resolution to compare simulated and satellite LAI at a larger scale over the period 2000-2007. Initial conditions and forcing data used for simulations are detailed in the Supplementary Material (Appendix SA).

Finally, we ran the model over France at a resolution of 0.07° for the period 1970-2100 to assess patterns in budburst timing in future decades. The climate forcing (daily data) was from the A2 scenario of ARPEGE v4 model downscaled and bias corrected by Pagé *et al.* (2008). In these simulations of the future, land cover was imposed from the IGBP map (Loveland & Belward, 1997) and soil depth and texture (used to derive wilting points and field capacities and thence to give plant water stress) from the FAO dataset (Vetter *et al.*, 2008).

3. **Results**

3.1. *RENECOFOR budburst and litterfall observations*

On average 50% of buds flushed for the sites of Fig. 2 within a range of 14 days around the 12 May (day 132; Table 1). The earliest budburst was recorded on 22 March 2007 for *Pinus pinaster* in "Les Landes" forest, in southwest France. The latest budburst dates were observed in mountainous regions for *Abies alba* and *Pinus sylvestris*, 24 June 1999 and 24 June 2008, respectively. On average, coniferous stands dropped 2336.5 kg ha^{-1} yr^{-1} (dry matter) of their needles as litter from 1997 to 2007, which represents 65% of the total annual litterfall for all compartments on average (all stands, only considering the dominant species; 49% considering secondary species) and more than 70% for *Abies alba*, *Picea abies* and *Pinus pinaster* (Table 3). We could see large differences

in the mass of needles lost per year among species. Species with minimum losses were *Abies alba* (mean over all sites was 1892.0 kg ha^{-1} yr^{-1}) and *Pinus sylvestris* (1859.5 kg ha^{-1} yr^{-1}) whereas the maximum was observed for *Pinus pinaster* (3175.4 kg ha^{-1} yr^{-1}).

Species	**Leaves** kg ha^{-1} yr^{-1}	**CV_L** %	**Prop.** %	**Total** kg ha^{-1} yr^{-1}	**CV_T** %	**n**
A.alb	1892.2	37.2	58.0	3263.2	33.0	145
P.abi	2692.6	30.6	72.8	3698.6	30.0	121
P.men	2036.7	31.8	84.5	2411.3	30.4	66
P.neg	2447.2	24.3	51.6	4740.4	18.4	21
P.pin	3175.4	30.3	76.1	4172.4	28.0	81
P.syl	1859.5	34.6	49.6	3750.4	30.0	171
All	2336.5	20.4	49.2	3580.4	21.4	605

Table 3: Litterfall mass (kg dry matter ha^{-1} yr^{-1}) measured in the RENECOFOR network over the period 1997-2007 for each compartment. CV corresponds to the coefficient of variation for each compartment (leaves, and total) and Prop. corresponds to the proportion of the compartment compared to the total litterfall. P.men=*Pseudotsuga menziesii*, P.abi=*Picea abies*, P.neg=*Pinus negra*, P.pin=*Pinus pinaster*, P.syl=*Pinus sylvestris*, A.alb=*Abies alba*.

3.2. *Budburst models comparison and selection of a best model*

Best models (parsimonious and predictive) retained for each species are listed in Table 4 and the corresponding model parameters are given in Table SC2 (Appendix SC). Figure SC1 shows the DIST-ST evaluation metric after parameter optimization for each model forced by daily temperatures. For all species together, the best model (DIST-ST criterion) is the simple spring forcing model M3 (DIST-ST=0.25, RMSE=12.5 days, Fig. 3) with a starting date fixed to 1 January. The most parsimonious is the chilling-forcing model M7 (DIST-ST=0.39). With M3, both spatial (DIST-S=0.32, RMSE=10.7 days) and temporal (DIST-T=0.17, RMSE=7.3 days) variability was well reproduced for the validation dataset. The DIST-ST values obtained with models M2, M6, M7, M8 and M9 are close to this best model, with DIST-ST of 0.35, 0.26, 0.39, 0.4 and 0.29, respectively, but only models M7, M8 and M9 are considered parsimonious according to ΔAIC. The sequential (M4) and parallel models (M5) could not reproduce observations properly (DIST-ST=1.58 to 2.56, RMSE=19.0 to 28.0 days).

Species	Model	DIST-ST
All	M3	0.25
Firs	M7	0.47
Pines	M7	0.46
Abies alba	M3	0.56
Picea abies	M7	1.25
Pseudotsuga mensiezii	M9	0.32
Pinus pinaster	M3	1.15
Pinus sylvestris	M7	0.45

Table 4: Best models retained for each species according to both the predictive power considering spatial and temporal variability (DIST-ST) and the parsimony (with the lowest number of parameters). Models were optimized against daily temperatures. See Appendix SB for a detailed description of each model.

For groups of species, in the case of firs, the best model is the chilling-forcing model M7 (DIST-ST=0.47, RMSE=11.9 days). The temporal variability (DIST-T=0.45) representation was equivalent to the spatial variability for M7 for firs (DIST-S=0.48). Models M2, M3, M6, M8 and M9 have performances close to M7 but only M3, M7 and M8 have a $\Delta AIC<10$. For pines, all optimized models produced better DIST-ST values than the null model M1, but higher RMSE. The best model for pines is again the chilling-forcing model M7 (DIST-ST=0.46, RMSE=16.3 days), while the best parsimonious is M2 (DIST-ST=0.63). Model M7 better represented the spatial pattern of observed budburst (DIST-S=0.20, RMSE=10.2 days) than for temporal variability (DIST_T=0.62, RMSE=13.5 days).

For individual species, the most parsimonious model is also M7. In the case of *Abies alba* the best DIST-ST is obtained with M6 (DIST-ST=0.48) while the best parsimonious model is the model M3 (DIST-ST=0.56). For *Picea abies*, none of the models reproduced accurately the observations, M7 having the best score and parsimony with DIST-ST=1.25 (RMSE=21.0 days). For *Pseudotsuga mensiezii* the best model was M9 (DIST-ST=0.32, RMSE=10.7 days), which concords with the results of Harrington et al. (2010) for this species. However, for both *Pseudotsuga mensiezii* and

Pinus pinaster the most parsimonious model led to high DIST-ST values, with 2.19 and 2.64 respectively. For *Pinus pinaster*, the best DIST-ST was obtained with M8 (DIST-ST=0.33, RMSE=14.0 days), but was not selected as parsimonious. The best model was thus M3 with DIST-ST=1.15. Finally, the best DIST-ST for *Pinus sylvestris* was obtained with M4 (DIST-ST=0.38, RMSE=12.0 days) while the most parsimonious model is again M7.

Lower performances of the optimization at the species level compared to groups of species can be explained by the smaller training dataset available for parameter calibration. We argue that models calibrated with all species or groups of species should thus be more robust and more suitable for large scale simulations.

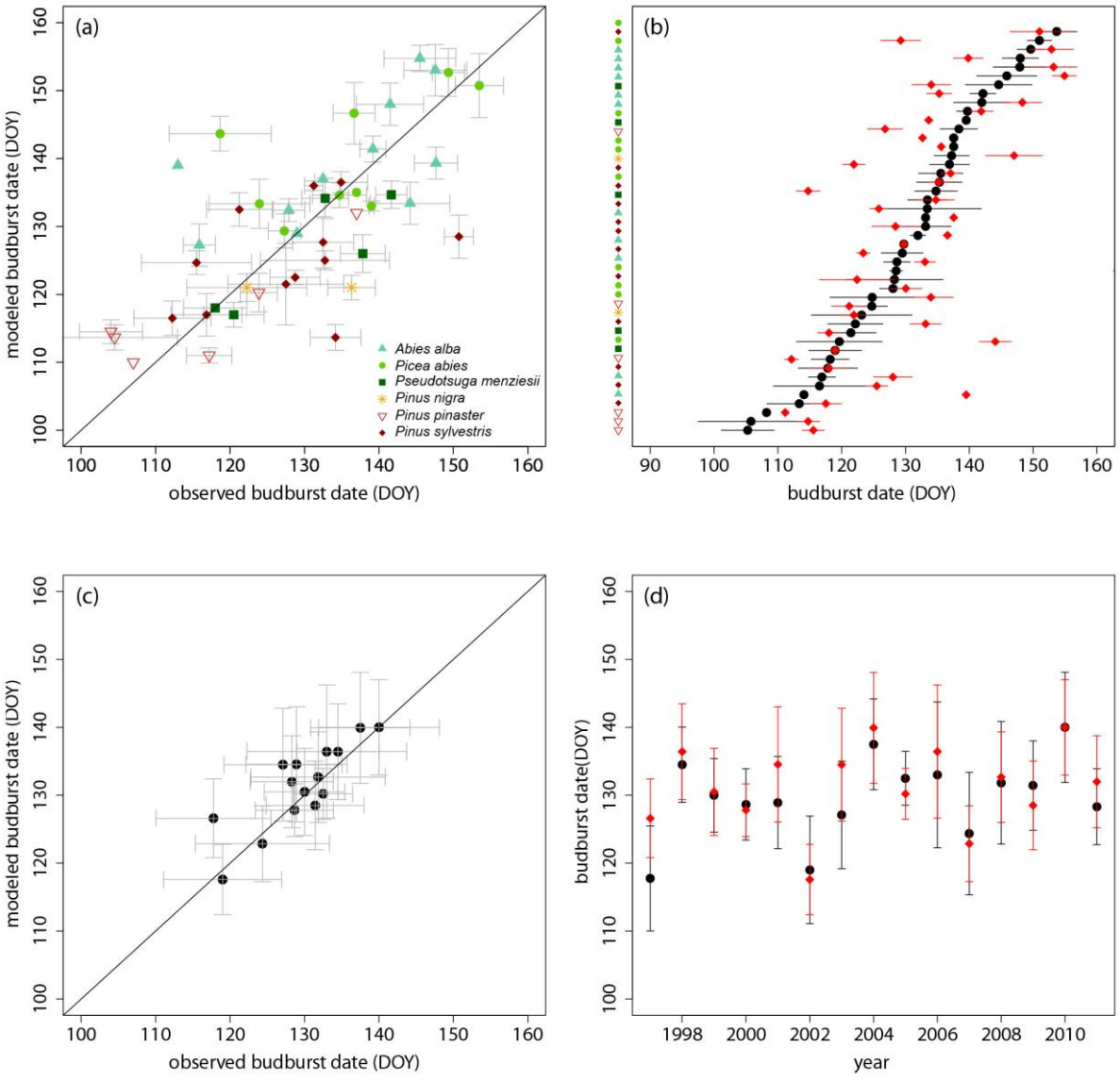


Fig. 3: Results for each site and each year for the best model with all species together (model M3). Representation of the validation dataset (150 obs.). (a) and (b) correspond to mean dates by site (Error bars correspond to the inter-annual variability), (c) and (d) correspond to mean dates by year (all sites together, error bars correspond to the inter-site variability). The y-axis of (b) represents each site for which the dominant species is represented by symbols listed in (a). For (b) and (d), black dots correspond to mean observations, red diamonds correspond to mean modeled budburst dates.

Results with or without optimizing the starting date for temperature accumulation (t_o, t_c) have similar model performance (Fig. SC1). However, we can see different performances of the same model depending on the species. For example, the optimization of model M6 with fixed t_c led to better DIST-ST for *Picea abies* and worse DIST-ST for *Pseudotsuga mensiezii* compared to M6 with optimized t_c. Thus, we preferentially selected models with a fixed starting date for large scale simulations when optimization results were equivalent. The same conclusion applies to optimizations with varying t_{cend} (end of chilling accumulation).

For all species, pines and firs, we found quiet similar performances for both forcing and chilling-forcing models. Note that chilling-forcing models may still be more physiologically realistic for future predictions where warmer winters may exacerbate the effects of incomplete fulfillment of chilling, or for applications in cold regions where chilling should be more important than in France and western Europe. We selected M3 (forcing) and M7 (chilling-forcing) for inclusion in ORCHIDEE.

3.3. Model performances from site-scale to grid-based resolution

We checked for model robustness at lower spatial resolutions, representative of typical forcing data for global vegetation models. At 0.25° resolution, most of DIST-ST values were higher than at a resolution of 0.07° and the best models differed. For all species, the best DIST-ST increased from 0.25 (M3-t_o) to 0.43 (M8+t_o+t_c). But the best parsimonious model was still M7 with a DIST-ST value of 0.49. Some models were no longer able to work correctly and DIST-ST values diverged

(DIST-ST= 132.2 for *Pinus sylvestris*-M8). At 0.5° the most parsimonious model remained M7 but DIST-ST increased from 0.49 to 0.77. If we compare the evolution for the best models M3 and M7 at 0.07° and 0.5°, we can observe that the result is more degraded for M3 (DIST-ST increased from 0.25 to 1.77) than for M7 (DIST-ST from 0.39 to 0.77), but is still a much better performance than the null model M1 with DIST-ST=9.17.

At a lower spatial resolution, some models could not be used. This was the case for models with a fixed threshold for chilling accumulation (M4, M5 and M6). By averaging temperatures, the critical threshold for chilling accumulation of these models was never reached at some sites and consequently forcing temperatures could never accumulate. In the rare cases when the model succeeded in estimating a budburst date, we could see that the performance was lower than the null model M1 (DIST>10). The implications of using models derived at the site scale for low-resolution prediction are further addressed in the discussion section.

3.4. *Senescence model parameters*

The minimum and maximum needle lifespan in Eq. 9 retrieved from literature studies were 4 and 15 years, respectively. Independently of environmental factors, species or tree health, needle survivorship follows almost the same pattern in each study: the needle biomass turnover is relatively low during the first years of the needle life and then rapidly increases over time. The logistic relationship we fitted on those data was strongly correlated to the needle lifespan of the tree (Eqn 10, Eqn 11) with R^2=0.93 and 0.94 for parameters μ and λ, respectively (Fig. SC3, SC4; Appendix SC).

$$ln(\mu) = 1.35 - 1.33ln\left(Needle_{age}\right); R^2 = 0.93 \quad (10)$$

$$\lambda = -0.82 + 0.43Needle_{age}; R^2 = 0.94 \quad (11)$$

with $Needle_{age}$ the maximal needle age.

The strong correlation between senescence parameters and the maximal needle age thus allows the use of one unique relationship for all species.

3.5. *Comparison against satellite data*

We first compared LAI simulated at site scale for two representative species (*Picea abies* (Fig. 4a) and *Pinus sylvestris* (Fig. 4b)) with MODIS satellite observations over the MODIS pixel of 1 km containing each site. Simulations were performed here with budburst results obtained for the pines and fir/spruces groups. For both models M3 (best forcing model) and M7 (best chilling-forcing model), results are equivalent. The amplitude of the LAI cycle with the improved phenology (1.4±0.1 for *Pinus sylvestris*; 1.2±0.1 for *Picea abies* - model M7) was closer to the amplitude of observed LAI (1.7±0.6; 1.9±0.7) than with the standard version of the model which does not have an explicit needle budburst equation (0.4±0.1; 0.4±0.1, respectively). For the 15 sites compared with MODIS LAI for each species, the mean correlation coefficient between modeled and estimated LAI improved from 0.45 ±0.2 to 0.77±0.1 for *Picea abies* and from 0.47±0.2 to 0.71±0.1 for *Pinus sylvestris*.

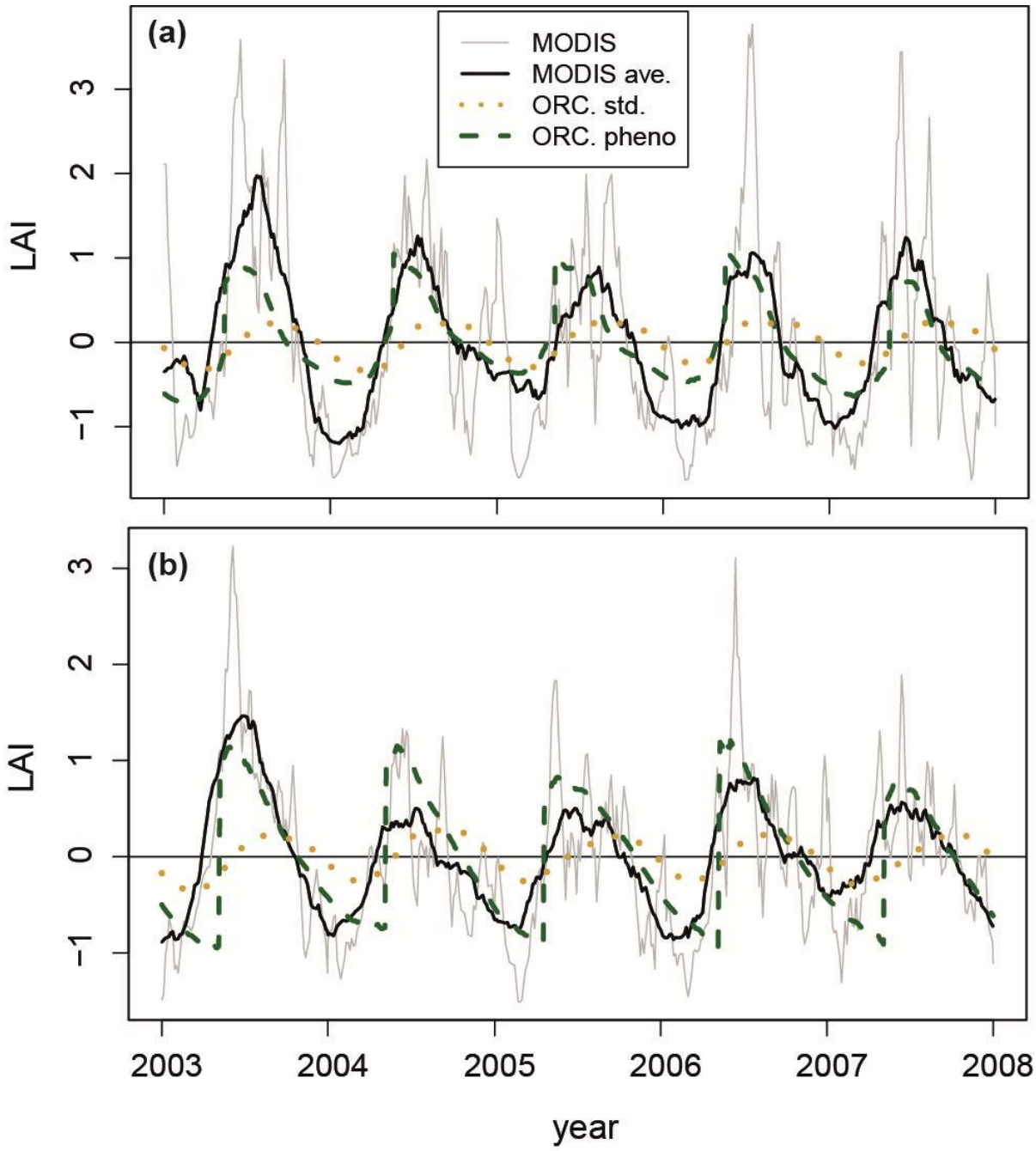


Fig. 4: Comparison between observed (MODIS) and simulated (model M7) LAI dynamics with ORCHIDEE for a) a *Picea abies* stand (lat=50.16º, long=5.46º) and b) a *Pinus sylvestris* stand (lat=49.25º, long=8.06º). All data are centered on the average observed (or simulated) LAI value (2003-2008). Gray and black lines represent the observed MODIS LAI and the moving average over a 30-days window, respectively. The orange dotted line represents the simulated LAI with the standard version of ORCHIDEE without phenological processes. The green dashed line represents the simulated LAI with ORCHIDEE including budburst and senescence processes.

Figure 5 shows the correlation coefficient between satellite-observed and simulated LAI at the European scale with the best forcing model (M3 optimized with all species) and the senescence model. The spatial correlation between modeled and satellite LAI improved by 0.24 (from

R=0.48±0.3 to 0.72±0.2) over Europe even if the calibration was performed only over France. We could observe the same improvement with the best parsimonious chilling-forcing model M7 (R=0.69±2). Moreover, we observed that the modeled budburst with M3 reached the imposed limit of budburst date in very high latitudes and altitudes. On the contrary, the chilling-forcing model M7 was better able to predict the LAI seasonality at high altitudes and high latitude, because it never reached the imposed budburst date in these areas.

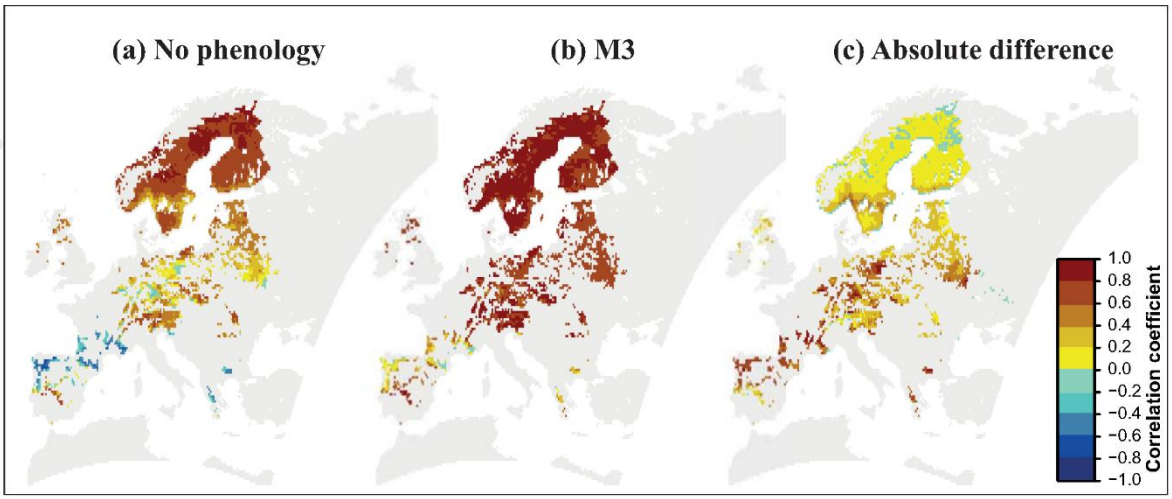
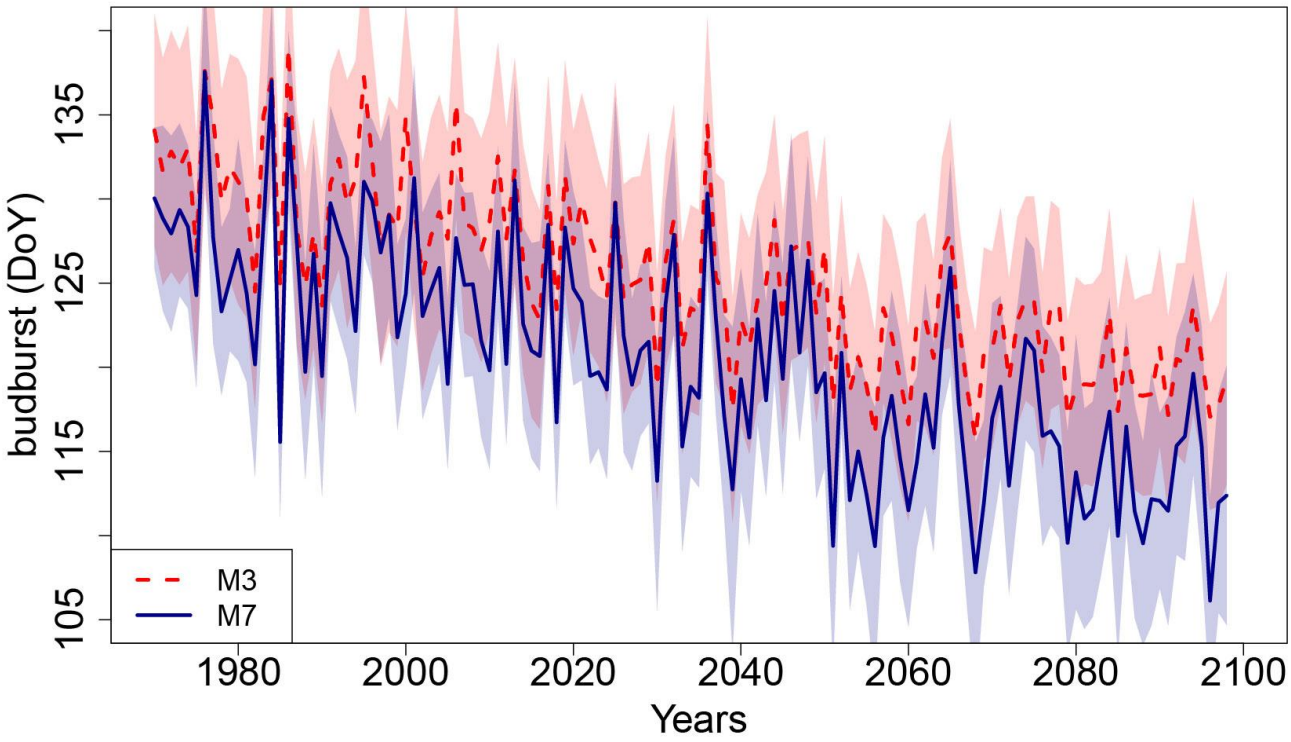


Fig. 5: Spatial representation of the correlation coefficient between MODIS estimated LAI and ORCHIDEE simulated LAI at a 0.25° spatial resolution averaged over the period 2000-2007: a) in the standard configuration without explicit phenological processes, or b) with the budburst model M3 and the senescence model. Map c) gives the absolute difference (b - a). Only the dominant species used in this study are represented (*Abies alba, Picea abies, Pinus pinaster, Pinus sylvestris, Pseudotsuga mensiezii*). The correlation is calculated for pixels with a minimal coniferous fraction cover of 20%.

3.6. *Budburst evolution from 1970 to 2100*

Figure 6 shows the simulated evolution of the mean budburst date over France from 1970 to 2100 with the best model M3 and the most parsimonious forcing-chilling model M7. For the two models, we simulated an earlier needle unfolding over time (mean slope=-0.126±0.01 days yr^{-1}, R^2 =0.57±0.06). Compared to 1970-2000, projections indicated an advancement of the budburst date of coniferous species of 10.3±2.8 and 12.3±4.1 days in average over the period 2060-2100 with M3 and M7, respectively. However, model M7 exhibited higher variability in the prediction of needle unfolding over time and an earlier budburst date on average compared to M3 (4.8±0.8 days) over

542 the period 1970-2100.



543
544 **Fig. 6** : Temporal evolution of the mean budburst dates (DOY) simulated over France from 1970 to
545 2100 for models M3 and M7 (all species together). For each line, the corresponding colored area
546 indicates one s.d. either side of the mean.
547

548

549 3.7. *Impact of the new conifer phenology model on GPP*

550 We assessed the impact of the new phenology model on the simulated GPP. We found in our
551 simulations an increase of GPP by 15±1% when compared to the standard version of ORCHIDEE.
552 For example, GPP increased from an average 5.5±0.2 g C m^{-2} d^{-1} to 6.3±0.2 g C m^{-2} d^{-1} for spruce
553 stands over France during the historical period (2000-2010). GPP increase was mainly induced by a
554 modification of the canopy composition. Due to the production of new needles each year and
555 senescence removing old needles, needle cohorts forming the canopy are younger (see Materials
556 and methods) in the new model (845.3±55.5 days for spruces versus 1740±18.8 days in the standard
557 version of ORCHIDEE). New needles having higher photosynthetic rates, this results into a higher
558 simulated GPP. Because old needles are already present at the beginning of the growing season (i.e.
559 photosynthetic activity starts as soon as climate conditions are favorable) and because new needles

reach their maximal activity only 3 months after unfolding, simulated GPP was not sensitive to changes in budburst date. By imposing the needle onset (from day 90 to 160), we observed a mean GPP difference of 0.002 g C m^{-2} d^{-1}, which sums up to only 0.8±0.1 g C m^{-2} over the whole year for each day difference in budburst.

With the new senescence model instead of the standard parameterization, we found an increase of the needle turnover from 0.16±0.002 to 0.43±0.025 g C m^{-2} d^{-1} for spruce. Consequently, litterfall for *Picea abies* stands without explicit senescence has an average of 1155.0±12.6 kg ha^{-1} yr^{-1} over France while we simulated losses of 3146.5±182.7 kg ha^{-1} yr^{-1} with explicit senescence, which is closer to observations for this species (2692.6 kg ha^{-1} yr^{-1}); this is a relative difference of -57% and +17%, respectively, between observed and simulated litterfalls.

4. **Discussion**

4.1. *Uncertainties in model validation*

Phenology is a central function in stands of conifers. In this study we optimized different phenological models against *in situ* budburst observations for six coniferous species. We showed that most models reproduced budburst dates with a precision of ±12 days across France. However we highlighted the need to consider both spatial and temporal variability when calibrating a phenological model. Figure SC5 (Appendix SC) illustrates the differences in results for the two spring forcing models M2 and M3 considering either the simple RMSE or our more comprehensive DIST metrics to select the best model after calibration. With the model M2 the best optimization based on RMSE does not allow to simulate budburst occurring before day 130 (Fig. SC5a, b), while with the model M3 (Fig. SC5c, d) the best optimization based on RMSE resulted in a high variability of simulated budburst. Even when considering both temporal and spatial variability of budburst with the DIST metrics, our results highlighted that, while some models managed to reproduce the observed budbursts (for instance M7 for firs, section 3.2), the same model calibrated

on different dataset generally reproduce better the spatial variability than the temporal variability of budburst. The DIST metrics proposed in this study is a first attempt to take into account both temporal and spatial variability of budburst. The combination of multiple metrics is promising to improve model calibration but more investigation is needed to improve this metrics, like the number of component we need to consider and their weight.

This implementation of better phenology models for evergreen conifers in the global model ORCHIDEE shows an improved ability to reproduce the seasonal LAI dynamics observed from MODIS both at the site and regional scale, despite representativeness differences between one site and a MODIS pixel, and the contribution of understory vegetation to the MODIS signals. Previous studies have shown a good correspondence between field-measured LAI and MODIS products (Jensen *et al.*, 2011; Rautiainen *et al.*, 2012). However, needle senescence is locally dependent of stand health, age and disturbances. We indirectly validated our senescence model against MODIS LAI data, despite uncertainties of this product. Comparisons with litterfall observations further allowed to show that our senescence model represent an improvement. By comparing site measurements of carbon fluxes and MODIS products, Verma *et al.* (2014) highlighted the uncertainties linked to the heterogeneity of the vegetation at a larger spatial resolution. In this study we visually checked for canopy openness at the site scale, however LAI from MODIS products integrates contributions from both the dominant species and the understory vegetation, which will result in a bias when comparing PFT specific simulations and observed amplitude and dynamics of the LAI (Wang et al., 2004; Jensen et al., 2011; Rautiainen et al., 2011; Rautiainen & Heiskanen, 2013). Moreover, LAI is related to stand age and health (Pokorný et al., 2013), which were not taken into account in our study.

The senescence model proposed in this study defines a continuous process over the needle lifespan. Moreover, we fixed the length of the period of needle accretion in ORCHIDEE, which can result in biases in LAI at the beginning of the growing season. We argue that an optimization of the senescence model parameters against carbon fluxes and remote sensing observations could further

improve the senescence model.

4.2. *Relevance of site-calibrated models for gridded simulations*

All tested models in this study could be optimized with good accuracy at the site scale (RMSE=6.0±3.4 days for M3, 4.3±2.7 days for M7) or across sites (RMSE=12.5 days with all sites). We also show that forcing models performed equally compared to chilling-forcing models at the regional scale. Olsson & Jönsson (2014) indicated that simple models with few parameters are more accurate over larger regions in general. In our study, most phenological models were more efficient than the null model (M1), i.e., setting a fixed date equal to the median observed one, even at a degraded spatial resolution. However, some models were not able to reproduce budburst events when calibrated using large-scale temperature forcing data, especially models M4 and M5.

Modelers should be cautious when applying empirical models fitted at site scale for large-scale predictions for two aspects: 1) If a model developed for site scale studies does not work at larger resolutions, it means either that the model is not generic enough: the sites were too specific or some processes are missing, like adaptation or acclimation for example; 2) The spatial aggregation of temperatures can smooth and modify the response of chilling and forcing. Modelers have to check that models calibrated on sites, and thus dependent of site conditions, are still able to reproduce average responses of budburst globally and not only in limited environmental conditions. Thus the model validity should be assessed at different spatial and temporal scales. The metrics developed in our study lowers the weight of outliers in the calibration and thus limits this effect.

Here, the best forcing model M3 has more degraded performance than the best chilling-forcing model M7 when applied at 0.25° and 0.5°, even if M7 exhibits a lower performance than M3 when calibrated at 0.07°. This result suggests that the model M7 may be more suitable than M3 to be used in a global model such as ORCHIDEE.

We also found that the best model calibrated with *in situ* observations is not necessarily the best model to reproduce the seasonality of the satellite LAI cycle. Our results suggest that looking at the

whole phenological cycle, in addition to *in situ* observations, could be a better way of discriminating between budburst models intended to be used in global vegetation models. In a recent study, Gamon *et al.* (2016) demonstrated a method to track photosynthetic phenology in evergreen conifers using a remotely sensed reflectance chlorophyll/carotenoid index (CCI). This new indicator could allow an accurate calibration of phenological processes for conifers. However, because of the land cover heterogeneity over Europe, a calibration of phenological models based only on satellite observations would also benefit in the future from a very high temporal and spatial resolution dataset (Delegido *et al.*, 2011; Verrelst *et al.*, 2012; Klosterman *et al.* 2014).

4.3. *Impact of phenology for large-scale simulations of GPP and LAI*

The new phenological processes incorporated into ORCHIDEE led to a better representation of the seasonal cycle of LAI both at site and regional scale (Figs 5, 6). Simulated LAI was however not sensitive to the timing of needle onset, while the production of new needles and the use of an explicit representation of needle senescence had a strong impact on simulated variables. In our simulations, all the new needle biomass was allocated to younger needles with high photosynthetic efficiency, thus leading to simulated GPP being higher than in the original model. The explicit representation of the senescence also led to a higher needle turnover, and litterfall. A higher litterfall rate will strongly impact soil carbon pools and heterotrophic respiration. Here, the estimated amount of needle-fall was in the range of values observed in French stands, while the standard version of ORCHIDEE underestimated the amount of litter. However, the lack of information about the living needle biomass at each site did not allow an accurate comparison of simulated and observed litterfall. Needle-fall is closely related to stand age, stand health, climate and disturbances (Balster & Marshall, 2000; Choi *et al.*, 2006; Reich *et al.*, 2014). The relation defined in this study does not take into account all the factors influencing needle-fall and more investigation is needed.

The strong correlation between the senescence parameters and the maximal age of needles (Eq. 7,

8) makes our proposed senescence model relevant for a large variety of evergreen coniferous species and may be generalized to other evergreen species in other biomes. With the generalization of models with varying traits (Pavlick *et al.*, 2012, Verheijen *et al.*, 2015), we argue that our senescence model could be easily implemented with trade-offs concerning the maximal age of needles, as for example relationships between the needle lifespan and mean annual temperatures, recently implemented in the CABLE model by Reich *et al.* (2014). As for budburst, we can expect a change in the senescence rate with global changes. The inclusions of trade-offs between needle longevity and climate in the ORCHIDEE model as it was done by Reich *et al.* (2014) will be a first step in understanding the impact of such changes on the carbon balance of forest ecosystems.

For evergreen species, a significant amount of leaves/needles is already present at the beginning of spring. The presence of old needles thus allows the recovery of carbon and water fluxes when temperatures become favorable — this explains the low sensitivity of simulated fluxes to budburst date. However, even if the needle onset had few impacts in this study, the implementation of an explicit budburst model was shown to play a key role in other mechanisms such as ozone sensitivity of needles (Watanabe *et al.* 2010; Verbeke, 2015), frost risk (Hänninen, 2006; Man *et al.*, 2015), biogenic emissions (Richardson *et al.*, 2013) or vegetation dynamics (Lu *et al.*, 2016). The mechanisms presented in this study could be extended to improve the representation of other evergreen species in global models.

5. Conclusion

Phenology plays a central role in bio-geochemical cycles in conifers stands. In this study we optimized different phenological models against budburst observations of six conifers species. We show that all models managed to reproduce needle emergence at the site scale with good accuracy (±5d). At the national scale, most models reproduce budburst dates with a precision of 12 days. However, we highlight the need to consider both spatial and temporal variability when calibrating phenological models.

Different budburst models performed equally independently of their complexity and the process

693 they include (forcing, chilling, photoperiod) and models calibrated at a fine spatial resolution were
694 not able to predict budburst dates when applied at coarse resolutions typical of the grid of global
695 models. This suggests that common models developed for site scale experiments might be
696 inadequate for large scale simulations. This first attempt in implementing an explicit phenological
697 model for evergreen conifer PFTs for large scale simulations managed to reproduce the observed
698 LAI dynamics both at the site and regional scale. A sensitivity analysis highlighted that the new
699 phenology module has a significant impact on the simulated carbon fluxes. We showed that needle
700 onset will be ±11 days earlier in 2060-2100 compared to 1970-2000 and more analysis are needed
701 to quantify the effect of evergreen conifers phenology on the projected carbon budget. The findings
702 will help future research to better improve current and future predictions of carbon, water, nutrient
703 and heat cycles using ecosystem model.
704
705 **Acknowledgment**
706 This work was granted access to the HPC resources of TGCC under the allocation 2015-6328 made
707 by GENCI. The authors would like to acknowledge the financial support from the European
708 Research Council Synergy grant ERC-SyG-2013-610028 IMBALANCE-P. Many thanks go to all
709 the people who contributed to setting up and maintaining the RENECOFOR network, to collecting
710 data and to ensuring its quality. RENECOFOR is part of the ICP Forests monitoring program and
711 has been continuously supported by French public funds (Office National des Forêts, Ministry of
712 Agriculture, Ministry of Ecology, ADEME) and by the European Union from 1991 until 2006.
713

References

Atmosphere, U.S., 1976. US standard atmosphere. National Oceanic and Atmospheric Administration.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K.T., Pilegaard, K., Schmid, H.P., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001. FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem–Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. Bull. Am. Meteorol. Soc. 82, 2415–2434. doi:10.1175/1520-0477

Balster, N.J., Marshall, J.D., 2000. Decreased Needle Longevity of Fertilized Douglas-Fir and Grand Fir in the Northern Rockies. Tree Physiol. 20, 1191–1197. doi:10.1093/treephys/20.17.1191

Blümel, K., Chmielewski, F.-M., 2012. Shortcomings of classical phenological forcing models and a way to overcome them. Agric. For. Meteorol. 164, 10–19. doi:10.1016/j.agrformet.2012.05.001

Brus, D.J., Hengeveld, G.M., Walvoort, D.J.J., Goedhart, P.W., Heidema, A.H., Nabuurs, G.J., Gunia, K., 2012. Statistical mapping of tree species over Europe. Eur. J. For. Res. 131, 145–157. doi:10.1007/s10342-011-0513-5

Burnham, K.P., Anderson, D.R., 2003. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer Science & Business Media.

Burr, K.E., Tinus, R.W., Wallner, S.J., King, R.M., 1989. Relationships among cold hardiness, root growth potential and bud dormancy in three conifers. Tree Physiol.

Cannell, M.G.R., Smith, R.I., 1983. Thermal Time, Chill Days and Prediction of Budburst in Picea sitchensis. J. Appl. Ecol. 20, 951–963. doi:10.2307/2403139

Chen, M., Melaas, E.K., Gray, J.M., Friedl, M.A., Richardson, A.D., 2016. A new seasonal-deciduous spring phenology submodel in the Community Land Model 4.5: impacts on carbon and water cycling under future climate scenarios. Glob. Change Biol. 22, 3675–3688. doi:10.1111/gcb.13326

Choi, D.S., Kayama, M., Jin, H.O., Lee, C.H., Izuta, T., Koike, T., 2006. Growth and photosynthetic responses of two pine species (Pinus koraiensis and Pinus rigida) in a polluted industrial region in Korea. Environ. Pollut. 139, 421–432. doi:10.1016/j.envpol.2005.06.006

Chuine, I., 2000. A unified model for budburst of trees. J. Theor. Biol. 207, 337–347. doi:10.1006/jtbi.2000.2178

Chuine, I., Beaubien, E.G., 2001. Phenology is a major determinant of tree species range. Ecol. Lett. 4, 500–510. doi:10.1046/j.1461-0248.2001.00261.x

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, 455–466. doi:10.1046/j.1365-3040.1998.00299.x

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, 1707–1728. doi:10.1111/j.1365-3040.2012.02552.x

753 Delegido, J., Verrelst, J., Alonso, L., Moreno, J., 2011. Evaluation of Sentinel-2 Red-Edge Bands
754 for Empirical Estimation of Green LAI and Chlorophyll Content. Sensors 11, 7063–7081.
755 doi:10.3390/s110707063

756 Delpierre, N., Dufrêne, E., Soudani, K., Ulrich, E., Cecchini, S., Boé, J., François, C., 2009.
757 Modelling interannual and spatial variability of leaf senescence for three deciduous tree species in
758 France. Agric. For. Meteorol. 149, 938–948. doi:10.1016/j.agrformet.2008.11.014

759 Estrella, N., Menzel, A., 2006. Responses of leaf colouring in four deciduous tree species to climate
760 and weather in Germany. Clim. Res. 32, 253.

761 Gamon, J.A., Huemmrich, K.F., Wong, C.Y.S., Ensminger, I., Garrity, S., Hollinger, D.Y.,
762 Noormets, A., Peñuelas, J., 2016. A remotely sensed pigment index reveals photosynthetic
763 phenology in evergreen conifers. Proc. Natl. Acad. Sci. 113, 13087–13092.
764 doi:10.1073/pnas.1606162113

765 Gunderson, C.A., Edwards, N.T., Walker, A.V., O'Hara, K.H., Campion, C.M., Hanson, P.J., 2012.
766 Forest phenology and a warmer climate – growing season extension in relation to climatic
767 provenance. Glob. Change Biol. 18, 2008–2025. doi:10.1111/j.1365-2486.2011.02632.x

768 Hänninen, H., 2006. Climate warming and the risk of frost damage to boreal forest trees:
769 identification of critical ecophysiological traits. Tree Physiol. 26, 889–898.

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

772 Hänninen, H., Slaney, M., Linder, S., 2007. Dormancy release of Norway spruce under climatic
773 warming: testing ecophysiological models of bud burst with a whole-tree chamber experiment. Tree
774 Physiol. 27, 291–300. doi:10.1093/treephys/27.2.291

775 Harrington, C.A., Gould, P.J., St.Clair, J.B., 2010. Modeling the effects of winter environment on
776 dormancy release of Douglas-fir. For. Ecol. Manag., doi:10.1016/j.foreco.2009.06.018

777 Jensen, J.L.R., Humes, K.S., Hudak, A.T., Vierling, L.A., Delmelle, E., 2011. Evaluation of the
778 MODIS LAI product using independent lidar-derived LAI: A case study in mixed conifer forest.
779 Remote Sens. Environ. 115, 3625–3639. doi:10.1016/j.rse.2011.08.023

780 Kayama, M., Kitaoka, S., Wang, W., Choi, D., Koike, T., 2007. Needle longevity, photosynthetic
781 rate and nitrogen concentration of eight spruce taxa planted in northern Japan. Tree Physiol. 27,
782 1585–1593.

783 Kivimäenpää, M., Sutinen, S., 2007. Microscopic structure of Scots pine (Pinus sylvestris (L.))
784 needles during ageing and autumnal senescence. Trees 21, 645–659. doi:10.1007/s00468-007-0157-
785 8

786 Klosterman, S.T., Hufkens, K., Gray, J.M., Melaas, E., Sonnentag, O., Lavine, I., Mitchell, L.,
787 Norman, R., Friedl, M.A., Richardson, A.D., 2014. Evaluating remote sensing of deciduous forest
788 phenology at multiple spatial scales using PhenoCam imagery. Biogeosciences 11, 4305–4320.
789 doi:10.5194/bg-11-4305-2014

790 Kobayashi, K., Salam, M.U., 2000. Comparing simulated and measured values using mean squared
791 deviation and its components. Agron. J. 92, 345–352.

792 Kramer, K., 1994. Selecting a Model to Predict the Onset of Growth of Fagus sylvatica. J. Appl.
793 Ecol. 31, 172–181. doi:10.2307/2404609

Krinner, G., Viovy, N., Noblet-Ducoudré, N. de, 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, 33. doi:200510.1029/2003GB002199

Lebourgeois, F., Pierrat, J.-C., Perez, V., Piedallu, C., Cecchini, S., Ulrich, E., 2010. Simulating phenological shifts in French temperate forests under two climatic change scenarios and four driving global circulation models. Int. J. Biometeorol. 54, 563–581. doi:10.1007/s00484-010-0305-5

Leinonen, I., Hänninen, H., 2002. Adaptation of the timing of bud burst of Norway spruce to temperate and boreal climates. Silva Fenn.

Loveland, T.R., Belward, A.S., 1997. The IGBP-DIS global 1km land cover data set, DISCover: First results. Int. J. Remote Sens. 18, 3289–3295. doi:10.1080/014311697217099

Lu, X., Wang, Y.-P., Wright, I.J., Reich, P.B., Shi, Z., Dai, Y., 2016. Incorporation of plant traits in a land surface model helps explain the global biogeographical distribution of major forest functional types. Glob. Ecol. Biogeogr. n/a-n/a. doi:10.1111/geb.12535

Man, R., Colombo, S., Lu, P., Dang, Q.-L., 2015. Effects of winter warming on cold hardiness and spring budbreak of four boreal conifers. Botany 94, 117–126. doi:10.1139/cjb-2015-0181

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, 2063–2083. doi:10.5194/bg-9-2063-2012

Murray, M.B., Smith, R.I., Leith, I.D., Fowler, D., Lee, H.S.J., Friend, A.D., Jarvis, P.G., 1994. Effects of elevated CO2, nutrition and climatic warming on bud phenology in Sitka spruce (Picea sitchensis) and their impact on the risk of frost damage. Tree Physiol. 14, 691–706. doi:10.1093/treephys/14.7-8-9.691

Muukkonen, P., Lehtonen, A., 2004. Needle and branch biomass turnover rates of Norway spruce (Picea abies). Can. J. For. Res. 34, 2517–2527. doi:10.1139/x04-133

Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R., Lotsch, A., Friedl, M., Morisette, J.T., Votava, P., Nemani, R.R., Running, S.W., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sens. Environ., 83, 214–231. doi:10.1016/S0034-4257(02)00074-3

Niinemets, Ü., 2002. Stomatal Conductance Alone Does Not Explain the Decline in Foliar Photosynthetic Rates with Increasing Tree Age and Size in Picea Abies and Pinus Sylvestris. Tree Physiol. 22, 515–535. doi:10.1093/treephys/22.8.515

Olsson, C., Bolmgren, K., Lindström, J., Jönsson, A.M., 2013. Performance of tree phenology models along a bioclimatic gradient in Sweden. Ecol. Model. 266, 103–117. doi:10.1016/j.ecolmodel.2013.06.026

Olsson, C., Jönsson, A.M., 2015. Budburst model performance: The effect of the spatial resolution of temperature data sets. Agric. For. Meteorol. 200, 302–312. doi:10.1016/j.agrformet.2014.10.003

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. doi:10.1111/gcb.12593

Pagé, C., Terray, L., Boé, J., 2008. Projections climatiques à échelle fine sur la France pour le 21ème siècle: les scénarii SCRATCH08. Clim. Model. Glob. Change CERFACS.

Pavlick, R., Drewry, D., Bohn, K., Reu, B., Kleidon, A., 2012. The Jena Diversity-Dynamic Global Vegetation Model (JeDi-DGVM): a diverse approach to representing terrestrial biogeography and biogeochemistry based on plant functional trade-offs. Biogeosciences Discuss 9, 4627–4726.

Peaucelle, M., Bellassen, V., Ciais, P., Peñuelas, J., Viovy, N., 2016. A new approach to optimal discretization of plant functional types in a process-based ecosystem model with forest management: a case study for temperate conifers. Glob. Ecol. Biogeogr. doi:10.1111/geb.12557

Pokornỳ, R., Stojnič, S., others, 2013. Leaf area index of Norway spruce stands in relation to age and defoliation. Beskydy 5, 173–180.

Polgar, C.A., Primack, R.B., 2011. Leaf-out phenology of temperate woody plants: from trees to ecosystems. New Phytol. 191, 926–941. doi:10.1111/j.1469-8137.2011.03803.x

Porté, A., Loustau, D., 1998. Variability of the Photosynthetic Characteristics of Mature Needles Within the Crown of a 25-Year-Old Pinus Pinaster. Tree Physiol. 18, 223–232. doi:10.1093/treephys/18.4.223

Rautiainen, M., Heiskanen, J., 2013. Seasonal contribution of understory vegetation to the reflectance of a boreal landscape at different spatial scales. IEEE Geosci. Remote Sens. Lett. 10, 923–927.

Rautiainen, M., Mõttus, M., Heiskanen, J., Akujärvi, A., Majasalmi, T., Stenberg, P., 2011. Seasonal reflectance dynamics of common understory types in a northern European boreal forest. Remote Sens. Environ. 115, 3020–3028. doi:10.1016/j.rse.2011.06.005

Reich, P.B., Rich, R.L., Lu, X., Wang, Y.-P., Oleksyn, J., 2014. Biogeographic variation in evergreen conifer needle longevity and impacts on boreal forest carbon cycle projections. Proc. Natl. Acad. Sci. 111, 13703–13708. doi:10.1073/pnas.1216054110

Richardson, A.D., Anderson, R.S., Arain, M.A., Barr, A.G., Bohrer, G., Chen, G., Chen, J.M., Ciais, P., Davis, K.J., Desai, A.R., Dietze, M.C., Dragoni, D., Garrity, S.R., Gough, C.M., Grant, R., Hollinger, D.Y., Margolis, H.A., McCaughey, H., Migliavacca, M., Monson, R.K., Munger, J.W., Poulter, B., Raczka, B.M., Ricciuto, D.M., Sahoo, A.K., Schaefer, K., Tian, H., Vargas, R., Verbeeck, H., Xiao, J., Xue, Y., 2012. Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis. Glob. Change Biol. 18, 566–584. doi:10.1111/j.1365-2486.2011.02562.x

Richardson, A.D., Keenan, 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, 156–173. doi:10.1016/j.agrformet.2012.09.012

Rinne, P.L.H., Welling, A., Vahala, J., Ripel, L., Ruonala, R., Kangasjärvi, J., Schoot, C. van der, 2011. Chilling of Dormant Buds Hyperinduces FLOWERING LOCUS T and Recruits GA-Inducible 1,3-β-Glucanases to Reopen Signal Conduits and Release Dormancy in Populus. Plant Cell 23, 130–146. doi:10.1105/tpc.110.081307

Rohde, A., Bhalerao, R.P., 2007. Plant dormancy in the perennial context. Trends Plant Sci. 12, 217–223. doi:10.1016/j.tplants.2007.03.012

874 Sampson, D.A., Albaugh, T.J., Johnsen, K.H., Allen, H.L., Zarnoch, S.J., 2003. Monthly leaf area
875 index estimates from point-in-time measurements and needle phenology for Pinus taeda. Can. J.
876 For. Res. 33, 2477–2490. doi:10.1139/x03-166

877 Steiner, K.C., 1980. Patterns of variation in bud-burst timing among populations in several Pinus
878 species. Silvae Genet.

879 Verbeke, T., 2015. Développement et quantification des impacts de l'ozone sur la biosphère
880 continentale dans un modèle global de végétation. PhD thesis. Saint Quentin en yveline, Saint
881 Quentin en yveline, France.

882 Verheijen, L.M., Aerts, R., Brovkin, V., Cavender-Bares, J., Cornelissen, J.H.C., Kattge, J., van
883 Bodegom, P.M., 2015. Inclusion of ecologically based trait variation in plant functional types
884 reduces the projected land carbon sink in an earth system model. Glob. Change Biol. 21, 3074–
885 3086. doi:10.1111/gcb.12871

886 Verma, M., Friedl, M., Richardson, A., Kiely, G., Cescatti, A., Law, B., Wohlfahrt, G., Gielen, B.,
887 Roupsard, O., Moors, E., others, 2013. Remote sensing of annual terrestrial gross primary
888 productivity from MODIS: an assessment using the FLUXNET La Thuile dataset. Biogeosciences
889 BG 10, 11627–11669.

890 Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G., Moreno, J., 2012.
891 Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for
892 Sentinel-2 and -3. Remote Sens. Environ. 118, 127–139. doi:10.1016/j.rse.2011.11.002

893 Vetter, M., Churkina, G., Jung, M., Reichstein, M., Zaehle, S., Bondeau, A., Chen, Y., Ciais, P.,
894 Feser, F., Freibauer, A., Geyer, R., Jones, C., Papale, D., Tenhunen, J., Tomelleri, E., Trusilova, K.,
895 Viovy, N., Heimann, M., 2008. Analyzing the causes and spatial pattern of the European 2003
896 carbon flux anomaly using seven models. Biogeosciences 5, 561–583.

897 Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., Soubeyroux, J.-M., 2010. A 50-year high-
898 resolution atmospheric reanalysis over France with the Safran system. Int. J. Climatol. 30, 1627–
899 1644. doi:10.1002/joc.2003

900 Vitasse, Y., François, C., Delpierre, N., Dufrêne, E., Kremer, A., Chuine, I., Delzon, S., 2011.
901 Assessing the effects of climate change on the phenology of European temperate trees. Agric. For.
902 Meteorol. 151, 969–980. doi:10.1016/j.agrformet.2011.03.003

903 Wang, R., Chen, J.M., 2012. Seasonal leaf area index variations derived from needle growth and
904 fall measurements in two eastern white pine (Pinus Strobes L.) stands, in: 2012 IEEE 4th
905 International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications.
906 pp. 413–417. doi:10.1109/PMA.2012.6524866

907 Wang, Y., Woodcock, C.E., Buermann, W., Stenberg, P., Voipio, P., Smolander, H., Häme, T.,
908 Tian, Y., Hu, J., Knyazikhin, Y., Myneni, R.B., 2004. Evaluation of the MODIS LAI algorithm at a
909 coniferous forest site in Finland. Remote Sens. Environ. 91, 114–127.
910 doi:10.1016/j.rse.2004.02.007

911 Warren, C., 2006. Why does photosynthesis decrease with needle age in Pinus pinaster? Trees -
912 Struct. Funct. 20, 157–164. doi:10.1007/s00468-005-0021-7

913 Watanabe, M., Matsuo, N., Yamaguchi, M., Matsumura, H., Kohno, Y., Izuta, T., 2010. Risk
914 assessment of ozone impact on the carbon absorption of Japanese representative conifers. Eur. J.
915 For. Res. 129, 421–430. doi:10.1007/s10342-009-0316-0

Way, D. A., & Montgomery, R. A. (2015). Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant, Cell & Environment*, *38*(9), 1725-1736.

Worrall, J., 1983. Temperature–bud-burst relationship in amabilis and subalpine fir provenance tests replicated at different elevations. Silvae Genet.

Xiang, Y., Gubian, S., Suomela, B., Hoeng, J., 2013. Generalized simulated annealing for global optimization: the GenSA Package. R J. 5.

Yakovlev, I.A., Asante, D.K.A., Fossdal, C.G., Partanen, J., Junttila, O., Johnsen, O., 2008. Dehydrins expression related to timing of bud burst in Norway spruce. Planta 228, 459–472. doi:10.1007/s00425-008-0750-0

Supporting information captions

Appendix SA: Litter sampling methodology, initial conditions and forcing data used for ORCHIDEE simulations, sensitivity analysis

Appendix SB: Description, equations and parameters for the eight budburst models

Appendix SC: Supplementary figures and tables