Biogeosciences

On the uncertainty of phenological responses to climate change, and implications for a terrestrial biosphere model

M. Migliavacca[1], O. Sonnentag[2,3], T. F. Keenan[2], A. Cescatti[1], J. O'Keefe[4], and A. D. Richardson[2]

[1]European Commission, DG-Joint Research Centre, Institute for Environment and Sustainability,
Climate Risk Management Unit, Ispra (VA), Italy
[2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
[3]Département de Géographie, Université de Montréal, Montréal, QC H2V 2B8, Canada
[4]Harvard Forest, Petersham, MA 01366, USA

Correspondence to: M. Migliavacca (mirco.migliavacca@jrc.ec.europa.eu)

Abstract. Phenology, the timing of recurring life cycle events, controls numerous land surface feedbacks to the climate system through the regulation of exchanges of carbon, water and energy between the biosphere and atmosphere.

Terrestrial biosphere models, however, are known to have systematic errors in the simulation of spring phenology, which potentially could propagate to uncertainty in modeled responses to future climate change. Here, we used the Harvard Forest phenology record to investigate and characterize sources of uncertainty in predicting phenology, and the subsequent impacts on model forecasts of carbon and water cycling. Using a model-data fusion approach, we combined information from 20 yr of phenological observations of 11 North American woody species, with 12 leaf bud-burst models that varied in complexity.

Akaike's Information Criterion indicated support for spring warming models with photoperiod limitations and, to a lesser extent, models that included chilling requirements.

We assessed three different sources of uncertainty in phenological forecasts: parameter uncertainty, model uncertainty, and driver uncertainty. The latter was characterized running the models to 2099 using 2 different IPCC climate scenarios (A1fi vs. B1, i.e. high CO_2 emissions vs. low CO_2 emissions scenario). Parameter uncertainty was the smallest (average 95 % Confidence Interval – CI: 2.4 days century^{-1} for scenario B1 and 4.5 days century^{-1} for A1fi), whereas driver uncertainty was the largest (up to 8.4 days century^{-1} in the simulated trends). The uncertainty related to model structure is also large and the predicted bud-burst trends as well as the shape of the smoothed projections varied among models (± 7.7 days century^{-1} for A1fi, ± 3.6 days century^{-1} for B1). The forecast sensitivity of bud-burst to temperature (i.e. days bud-burst advanced per degree of warming) varied between 2.2 days $^{\circ}C^{-1}$ and 5.2 days $^{\circ}C^{-1}$ depending on model structure.

We quantified the impact of uncertainties in bud-burst forecasts on simulated photosynthetic CO_2 uptake and evapotranspiration (ET) using a process-based terrestrial biosphere model. Uncertainty in phenology model structure led to uncertainty in the description of forest seasonality, which accumulated to uncertainty in annual model estimates of gross primary productivity (GPP) and ET of 9.6 % and 2.9 %, respectively. A sensitivity analysis shows that a variation of ± 10 days in bud-burst dates led to a variation of ± 5.0 % for annual GPP and about ± 2.0 % for ET.

For phenology models, differences among future climate scenarios (i.e. driver) represent the largest source of uncertainty, followed by uncertainties related to model structure, and finally, related to model parameterization. The uncertainties we have quantified will affect the description of the seasonality of ecosystem processes and in particular the simulation of carbon uptake by forest ecosystems, with a larger impact of uncertainties related to phenology model structure, followed by uncertainties related to phenological model parameterization.

1   Introduction

Phenology is the study of the timing of recurrent biological events and the causes of their temporal change in response to biotic and abiotic forces (Lieth and Radford, 1971). As variability in the timing of phenology is tightly coupled to variability in climate, phenology can be considered as an important indicator of climate change (e.g. IPCC, 2007; Menzel et al., 2006). Numerous studies have documented the impacts of climate change on plant and tree phenology. For instance, the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC) reported an overall trend towards earlier spring phenological events (e.g. bud-burst, leaf unfolding, flowering and pollen release) between 2 and 5 days decade^{-1} (Rosenzweig et al., 2007). Menzel et al. (2006) estimated an average advance of spring phenology in Europe of 2.5 days decade^{-1} while Schwartz et al. (2006) similarly showed earlier bud-burst of 1.1 days decade^{-1} across most temperate Northern Hemisphere land regions over the 1955–2002 period. Jeong et al. (2011) reported several different start of season trends at global and regional scales and suggested a reduction of the rate of advancement of the start of the season over the period 2000–2008 (0.2 days) compared to the period 1982–1999 (5.2 days).

Although these studies highlighted that spring phenology has responded to recent climate change, large uncertainties remain as to how phenology will respond to projected future climate change. There are several conflicting reports in the literature about the relative roles of different environmental factors such as chilling (i.e. the exposure to cool temperatures that is required before dormancy can be broken) and photoperiod in controlling tree phenology. Körner and Basler (2010) have argued, photoperiod will constrain the degree to which future warming causes continued advances in spring phenology. Other studies (e.g. Chuine et al., 2010; Morin et al., 2009; Vitasse et al., 2011) suggest that photoperiod has not been shown to be more dominant than temperature when predicting bud-burst or flowering. These latter studies argue that temperature has a more dominant role than photoperiod during both the endodormancy phase (i.e. an inactive phase caused by conditions or factors within a plant or seed itself) and the ecodormancy phase (i.e. the cessation of growth induced by environmental factors) that controls spring phenology. Another study (Schleip et al., 2008) suggests the importance of adequately weighting the temperature forcing to determine the effective temperature which controls each phenophase. These different hypotheses lead to different possible structures for phenology models, but the associated uncertainties in forecasts of phenological responses to climate change have yet to be quantified.

Uncertainty in model projections can be classified in three categories: uncertainty due to (1) model parameters; (2) model structure; and (3) model drivers. The evaluation of phenological model parameter uncertainty is necessary in order to estimate the uncertainty in performance of a particular model structure and for parameter optimization; model structural uncertainty stems from different model assumptions and formulations, with different processes described differently by each model; model driver uncertainty is due to uncertainty in future climate scenarios used for ecological forecasts (Cook et al., 2010).

Model-data fusion (e.g. Keenan et al., 2011; Wang et al., 2009) provides a promising approach for assessing uncertainties in ecological forecasting. Also referred to as model-data integration or data assimilation, model-data fusion relies on the combination of models with observational constraints through an optimization approach (e.g. simulating annealing, quasi-newton methods, etc.). In this way, model parameters, model states and their respective uncertainties can be estimated, conditional on the data (here: consistent phenological observations by a human observer). With a model-data fusion approach it is possible to objectively incorporate data, along with associated uncertainties, allowing for a full characterization of posterior distributions of model parameters. In this way, confidence estimates of model projections can be obtained, both for current climate conditions and for future climate change scenarios (e.g. Keenan et al., 2012a).

Phenology, and in particular bud-burst, controls numerous land surface feedbacks to the climate systems and ecological interactions through the regulation of exchanges of carbon, water and energy between the biosphere and atmosphere (e.g. Richardson et al., 2009, 2010; Baldocchi, 2008; Morisette et al., 2009; Fitzjarrald et al., 2001). Uncertainty in the prediction of spring phenology can, therefore, feed-forward to generate uncertainty in estimates of carbon and water cycling from terrestrial biosphere models.

Several studies have shown the sensitivity of different biogeochemical and terrestrial biosphere models to bud-burst and other phenological transitions (e.g. Randerson et al., 2009; Levis and Bonan, 2004; White et al., 2000; Migliavacca et al., 2009). Furthermore, recent multi-model synthesis studies have shown that spring phenology is poorly simulated by many terrestrial biosphere models (Richardson et al., 2012; Keenan et al., 2012b). The results show large biases in model estimates of carbon cycling due to errors in modeled phenology (Richardson et al., 2012), and identify phenology model errors as a systematic cause of poor model performance for interannual variability in terrestrial carbon cycling (Keenan et al., 2012b). Hence, modeled future carbon, water and energy fluxes, as well as many biosphere-climate interactions projected by terrestrial biosphere models, might be subject to uncertainty due to the uncertain representation of phenological responses to climate change.

Herein we present a phenological forecasting study using phenological data for 11 different North American tree species observed at Harvard Forest over the last 20 yr (Richardson and O'Keefe, 2009). We characterize the sources of uncertainty in bud-burst models under present climate conditions, and also for forecasts for the 21st century.
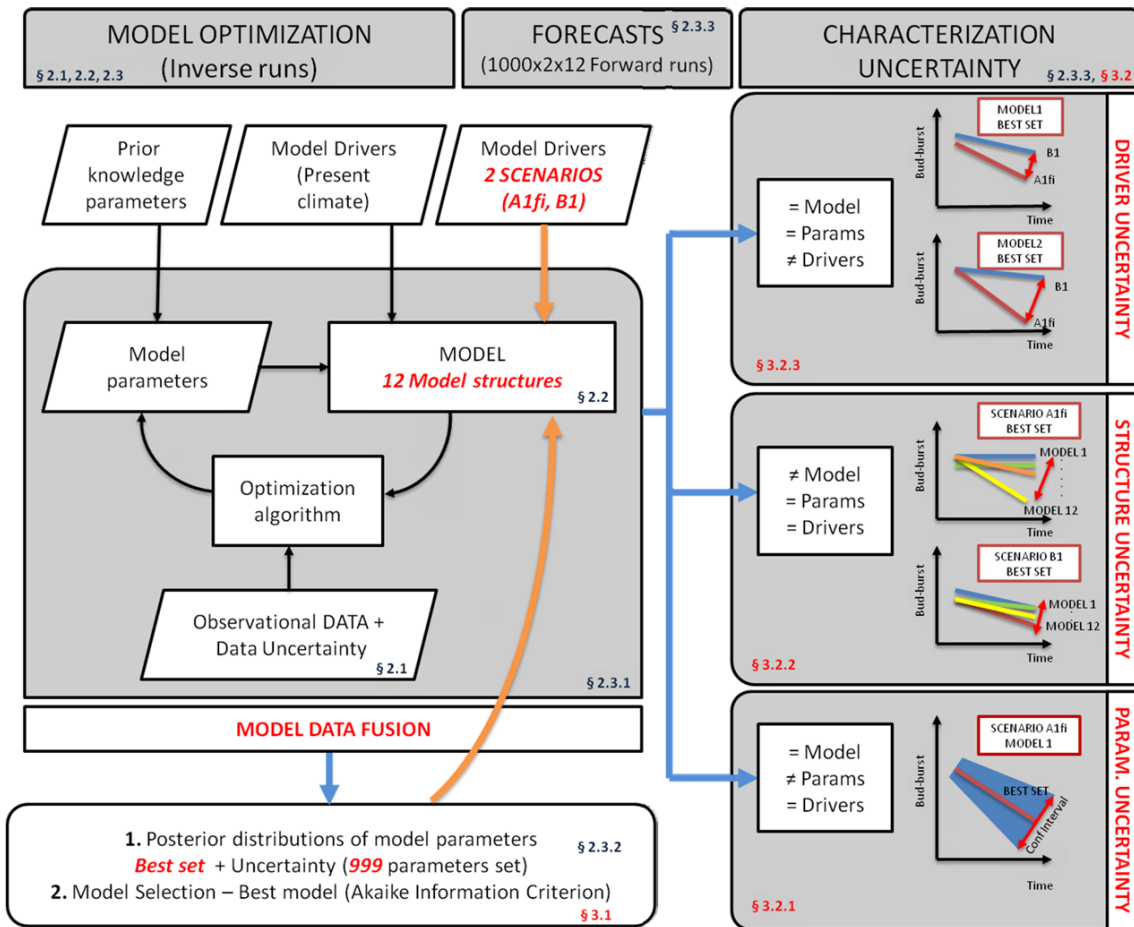
Fig. 1. Schematic representation of the model-data fusion approach for model optimization, the forecat mechanism and the characterization of the three sources of uncertainty. In each box is reported the paragraph where the description of methods (blue characters) and results (red characters) can be found. The schematic representation of the effects of the different uncertainties on bud-burst projections is also reported.

In addition, we quantify the impacts of these uncertainties for modeling forest-atmosphere fluxes of carbon and water. The two main questions behind our analysis are:

1. How big are the different sources of uncertainty in phenological forecasts?

2. How do these uncertainties affect the prediction of photosynthetic CO_2 uptake and evapotranspiration as described by a process-oriented terrestrial biosphere model?

To answer these questions we combine phenological observations collected at Harvard Forest, with 12 different phenological models, using a model-data fusion approach. With this analysis we characterize the uncertainty of model parameters and model structure. We then project model estimates of phenology forward, along with the associated parameter uncertainties, using statistically downscaled climate projections (Delworth et al., 2006; Hayhoe et al., 2007) for two different IPCC climate change scenarios (A1fi, or high CO_2

emissions scenario, and B1 or low CO_2 emissions scenario). This allows us to explore how the uncertainty characterized using current phenological observations is propagated in the future, and to quantify how model structural and parameter-based uncertainty interacts with uncertainty in climate scenarios. The schematic representation of the different steps of the analysis, model-data fusion approach for model optimization and the forecast mechanism is reported in Fig. 1.

To answer the second question we analyze the impact of the uncertainty of future bud-burst in a widely used terrestrial biosphere model (Boreal Ecosystems Productivity Simulator; BEPS, Ju et al., 2006). We evaluate the differences between gross primary productivity (GPP) and evapotranspiration (ET) as simulated by BEPS with the native phenological model and forced by the bud-burst forecasts obtained with the best model formulation selected according to data. Finally, we evaluate the sensitivity of GPP and ET to different levels of uncertainty in bud-burst (± 10, ± 1 day).

Table 1. List of the 11 species used in the analysis, species identifier (Species ID), species Latin and common name, average bud-burst date (BB), best model according to data (Best Model), coefficient of determination (R^2), root mean square error (RMSE, day), Sen's slope trend estimated for scenario A1fi and scenario B1 with the best model, with the Akaike Weigthed average (Average wAICc) and the standard deviation. All trends are expressed in day century^{-1}.

Species ID	Latin name	Common name	BB	Best Model	R^2	RMSE	A1fi Best	A1fi wAICc	A1fi sd	B1 Best	B1 wAICc	B1 sd
ACRU	*Acer rubrum*	Red maple	126.2	Alt-CF1	0.77	2.20	−8.7	−8.7	7.2	−3.9	−3.9	3.5
ACSA	*Acer saccharum*	Sugar maple	121.5	Par2-CF2	0.82	2.09	−7.1	−8.7	5.7	−3.5	−4.3	2.9
BEAL	*Betula alleghaniensis*	Yellow birch	126.2	Alt-CF1	0.63	3.44	−9.1	−8.8	8.5	−3.9	−3.8	4.5
BELE	*Betula lenta*	Black birch	129.5	SW-CF2	0.32	4.91	−5.0	−6.9	15.7	−1.7	−3.0	7.2
BEPA	*Betula Papyrifera*	Paper birch	122.9	Alt-CF1	0.56	3.31	−8.7	−6.4	7.7	−3.8	−2.8	3.6
FAGR	*Fagus grandifolia*	Beech	128.3	SW-CF1	0.67	2.22	−7.3	−5.3	8.1	−3.5	−2.4	4.2
FRAM	*Fraxinus americana*	White ash	130.2	SW-CF2	0.67	2.73	−5.6	−5.6	7.5	−2.3	−2.3	3.7
PRSE	*Prunus serotina*	Black cherry	111.2	SW-CF2	0.74	2.59	−9.3	−9.4	13.4	−5.4	−5.4	2.5
QUAL	*Quercus alba*	White oak	133.3	SW-CF2	0.73	3.21	−14.7	−14.0	3.9	−7	−6.7	1.6
QURU	*Quercus rubra*	Red oak	127.1	SW-CF2	0.74	2.94	−12.6	−10.6	5.5	−5.6	−4.7	3.0
QUVE	*Quercus velutina*	Black oak	130.6	SW-CF2	0.78	2.63	−14.9	−15.2	4.5	−6.8	−7.3	2.0

2 Materials and methods

2.1 Site description and phenological observations

The Harvard Forest (42.54° N, 72.18° W, el. 220 to 410 m a.s.l.) site used in this study is located in central Massachusetts, about 100 km west of Boston, USA. The climate is classified as humid-continental, with a mean July temperature of 20 °C and mean January temperature of 7 °C. Mean annual precipitation is 1100 mm, and is distributed evenly across the seasons. The species composition in Harvard Forest is dominated by transition hardwoods: red oak (*Quercus rubra*), red maple (*Acer rubrum*) black oak (*Quercus velutina*), white oak (*Quercus alba*) and yellow birch (*Betula alleghaniensis*). Conifers include eastern hemlock (*Tsuga canadensis*), red pine (*Pinus resinosa*) and white pine (*Pinus strobus*). Since 1990, springtime phenology observations have been made at 3–7 day intervals. Leaf development was monitored on three or more individuals (a total of 39 permanently marked trees or shrubs) of 11 woody species (Table 1). Phenological observations used here are available online (http://harvardforest.fas.harvard.edu).

In the present analysis we focus on bud-burst dates from 1990 to 2011. We define bud-burst as the date when 50 % of all buds on an individual tree have recognizable leaves emerging (Richardson et al., 2009). Our analysis uses temperature and photoperiod as drivers of phenology. Mean daily air temperatures are computed from the maximum and minimum daily temperatures recorded for the period of study at the Shaler (1964–2002) and Fisher (2001–2011) meteorological stations. Photoperiod is computed by using a standard equation based on latitude and day of year (Monteith and Unsworth, 1990).

2.2 Phenological models

A large number of different models exist for the simulation of bud-burst for different species (e.g. Chuine et al., 1999; Schaber and Badeck, 2003; Morin et al., 2009). The application of different models against different datasets of the same species, however, gave contrasting results about which modeling approach is best (Hunter and Lechowicz, 1992; Chuine et al., 1998, 1999). The models provide a context for interpreting observed interannual and inter-specific variability in phenology, and to assess which model structure is best supported by the available data.

The models used in this study are largely based on those presented by Chuine et al. (1999) and updated in Richardson and O'Keefe (2009). We define two main categories of models (Table 2). The model categories differ in their assumptions of how warm and cold temperatures control developmental processes (Fig. 2). In the spring warming models, temperatures above a base temperature accumulate until a threshold (in degree-days) is reached, thus triggering budburst. In the chilling models, cold weather also plays a role. In the sequential chilling model, a chilling threshold must be reached before warming is effective; in the alternating and parallel chilling models, an increase in the amount of chilling experienced reduces the amount of warming that is required. For all models, the rates of forcing (or spring warming) and chilling are calculated based on the threshold approach (CF1) and on the Sarvas function (CF2) (Sarvas, 1972; Chuine et al., 1999). In CF1, the rate of chilling and forcing for day of year (t) are accumulated if the daily air temperature ($x(t)$) is lower or higher than a specific threshold, respectively. In the approach CF2, rates of chilling (R_c) and forcing (R_f) accumulation are both specified as nonlinear functions of daily $x(t)$ according to the triangular function reported in the Sarvas model (Sarvas, 1972). More specifically, in CF2, chilling is accumulated according to Eqs. (1, 2, 3):

Table 2. List of phenology models fit to Harvard Forest phenology data, model identifier (Model ID), models class (chilling models or spring warming models) and fit parameters (number and variables symbols). Spring warming, sequential, alternating and parallel model structures are described in text and in Fig. 2. CF1 and CF2 refer to different functional forms for forcing and chilling rates, as described in text. t_1: time step at which accumulation of chilling units begins (not used in spring warming models; fit parameter in all other models). C^*: chilling state at which transition from rest to quiescence occurs (fit parameter in sequential and parallel 1 models; not used in other models). t_2: time step at which accumulation of forcing units begins (fit parameter in spring warming models; equal to t_1 in alternating and parallel 2 models; date when cumulative chilling (S_c) is equal to C^* in sequential and parallel1 models). F^*: forcing state at which transition from quiescence to bud-burst occurs (fit parameter in spring warming and sequential models; function of S_c in alternating, parallel 1 and parallel 2 models). T_c: critical temperature for chilling function $R_c(t)$ (not used in spring warming models; fit parameter in all other models). T_f critical temperature for forcing function $R_f(t)$ (fit parameter in all models). a, b model constants ($a > 0$, $b < 0$) relating F^* to S_c, i.e. $F^* = a \exp(bS_c(t))$ at $t = y$. y is the predicted bud-burst date (a, b not used in spring warming, alternating or sequential models; fit parameter in parallel models).

| Model name	Model ID	Model Class	Fit Parameters
Spring warming CF1	SW-CF1	Spring Warming	3 (t_2, T_f, F^*)
Spring warming CF2	SW-CF2	Spring Warming	2 (t_2, F^*)
Alternating CF1	Alt-CF1	Chilling	4 (t_1, T_f, a, b); $t_2 = t_1$; $T_c = T_f$
Spring warming CF1 t0 fixed	SW-CF1t0	Spring Warming	2 (T_f, F^*); $t_2 = 1$ January
Spring warming CF2 t0 fixed	SW-CF2t0	Spring Warming	1 (F^*); $t_2 = 1$ January
Alternating CF1 t0 fixed	Alt CF1t0	Chilling	3 (T_f, a, b); $t_2 = t_1 = 1$ January; $T_c = T_f$
Sequential CF1	Seq-CF1	Chilling	5 (t_1, T_f, T_c, C^*, F^*)
Sequential CF2	Seq-CF2	Chilling	4 (t_1, T_c, C^*, F^*)
Parallel1 CF1	Par1-CF1	Chilling	6 (t_1, T_c, T_f, C^*, a, b)
Parallel1 CF2	Par1-CF2	Chilling	5 (t_1, T_c, T_f, C^*, F^*)
Parallel2 CF1	Par2-CF1	Chilling	5 (t_1, T_c, C^*, a, b); $t_2 = t_1$
Parallel2 CF2	Par2-CF2	Chilling	4 (t_1, T_c, a, b); $t_2 = t_1$
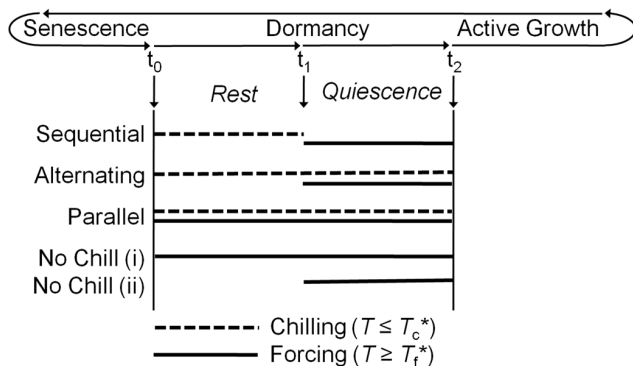
Fig. 2. Model representation of the seasonal cycle of terrestrial ecosystems, from senescence through dormancy (a period of rest, followed by quiescence) and then active growth. In temperate and boreal systems, the transition through dormancy to active growth has been described using a variety of approaches. Threshold points t_0 and t_1 may be triggered either by photoperiod or temperature (i.e. various forms of chilling: sequential, alternating or parallel, as indicated). At threshold point t_2, bud-burst is triggered when accumulated forcing reaches a critical state. The manner in which chilling and forcing accumulates varies among models. Chilling models describe phenology as a combination of temperature forcing, photoperiod limitation and chilling limitation; Spring warming models describe phenology as function of temperature forcing and photoperiod limitation (Table 2 for a list of model). Modified from Chuine (2000).

$$R_c = 0 \quad \text{if} \quad x(t) \leq -3.4 \quad \text{or} \quad x(t) \geq 10.4 \tag{1}$$

$$R_c = \frac{x(t) + 3.4}{T_c + 3.4} \quad \text{if} \quad -3.4 < x(t) \leq T_c \tag{2}$$

$$R_c = \frac{x(t) - 10.4}{T_c - 10.4} \quad \text{if} \quad T_c < x(t) \leq 10.4 \tag{3}$$

where T_c is the species-dependent temperature threshold for chilling accumulation. R_f is a sigmoid function of $x(t)$. Forcing is accumulated when $x(t) > 0$ as in Eq. (4):

$$R_f = \frac{28.4}{1 + e^{-0.185(x(t) - 18.4)}} \tag{4}$$

Either in spring warming or chilling models, photoperiod can control the point in time (i.e. a day-length threshold) at which chilling and forcing begin to have an effect.

For example, parameter t_2 in Table 2 controls the date at which forcing and/or chilling begins to be accumulated in some models. We also compared versions of the models with no photoperiod control; for these (models denoted by the suffix t_0; see Table 2), parameter t_2 is fixed to 1 January, which is the value for onset of degree-day accumulation most commonly used in other studies.

2.3 Statistical analysis

2.3.1 Model parameters and uncertainty estimations

Model parameters (listed in Table 2) were estimated for each species and model using bud-burst observations from

Table 3. ΔAICc values for a range of different models (see text and Tables 1, 2 for additional information) fit to Harvard Forest bud-burst data. Species ID are as given in Table 1. The best model, based on Akaike's Information Criterion corrected for small samples (AICc) has ΔAICc = 0 and is indicated by bold type.

Species ID	SW-CF1	SW-CF2	Alt-CF1	SW-CF1t0	SW-CF2t0	Alt-CF1t0	Seq-CF2	Seq-CF1	Par-CF2	Par-CF1	Par2-CF2	Par2-CF1
ACRU	11.49	16.24	**0.00**	28.93	31.59	15.48	35.49	14.51	39.07	13.20	14.51	19.73
ACSA	0.82	3.52	1.43	24.29	21.99	7.81	17.50	7.94	21.21	7.40	**0.00**	30.07
BEAL	9.19	7.35	**0.00**	20.16	22.41	11.41	25.03	13.36	28.65	13.32	11.85	19.04
BELE	2.93	**0.00**	0.12	15.78	15.82	5.66	19.46	11.23	23.08	10.79	51.97	12.20
BEPA	4.67	3.83	**0.00**	18.75	17.81	6.74	19.46	3.92	23.08	6.76	1.38	10.88
FAGR	**0.00**	7.25	6.00	27.34	31.55	10.43	29.89	11.25	32.27	7.40	1.60	11.60
FRAM	3.37	**0.00**	4.13	27.10	28.35	13.59	29.97	13.42	33.28	14.05	16.33	25.54
PRSE	3.56	**0.00**	3.47	22.44	23.23	7.63	12.56	11.51	12.88	7.37	64.97	17.84
QUAL	8.26	**0.00**	4.13	7.12	9.03	2.38	7.16	14.73	10.88	4.48	7.97	7.52
QURU	5.93	**0.00**	5.96	20.40	20.63	5.16	15.33	6.89	18.71	14.37	1.84	14.74
QUVE	3.91	**0.00**	6.10	14.29	15.03	7.71	1.50	8.70	4.94	7.81	4.39	14.46
MEDIAN	3.91	**0.00**	3.47	20.40	21.99	7.71	19.46	11.25	23.08	7.81	7.97	14.74
Best Model	1	**6**	3	0	0	0	0	0	0	0	1	0
ΔAICc < 2	2	6	5	0	0	0	1	0	0	0	4	0
ΔAICc < 6	8	8	10	0	0	4	1	1	1	1	5	0
Model Ranking	3	1	2	10	11	4	9	7	12	5	6	8

Harvard Forest as constraints. This allowed for the characterization of species-specific biological responses to environmental cues.

Model optimization and uncertainty analysis was performed using a model-data fusion framework based on simulated annealing-type routines and Monte Carlo techniques (Metropolis et al., 1953), as described by Richardson et al. (2010). The cost function selected for this purpose was the sum of squared error between observed and modeled data. Once the best parameter set was identified, the parameter space was further explored until 1000 parameter sets that gave statistically equivalent fits to the data were accepted. A specific parameter set was accepted if it passed a χ^2 test (at 95 % confidence level) for acceptance/rejection, after the normalization of variance based on the minimum of the cost function (Franks et al., 1999).

The resulting posterior distributions defined the parameter space within which approximately equally good agreement between data and model simulations can be obtained. Uncertainty estimates for model parameters and model predictions were thus provided directly by the model-data fusion framework, conditional on the data and the cost function.

By running an ensemble of models, with parameter sets selected from the posterior distributions, we can characterize the uncertainty in model predictions (both under current and future climate scenarios) that is attributed to parameter uncertainty.

2.3.2 Model selection and evaluation

We used the Akaike Information Criterion (AIC), a method based on information theory (Akaike, 1973; Anderson et al., 2000), for model selection purposes. AIC is a measure of the trade-off between the goodness-of-fit (model explanatory power) and the model complexity (number of parameters).

The small sample corrected criterion, AICc (Burnham and Anderson, 2002), is calculated as in Eq. (5):

$$\text{AICc} = n \log \sigma^2 + 2p + \frac{2p(p+1)}{n-p-1}, \tag{5}$$

where n is the number of samples (i.e. observation years), p is the number of model parameters and σ^2 is the residual sum of square divided by n. The competing model formulations proposed in Table 2 can be ranked according to AICc, where the model with the lowest AICc is considered best supported by the data, and most likely to be the "true" model. Candidate models can be compared directly by calculating the difference in AICc scores with the best model (ΔAICc). If ΔAICc is lower than 2, the two models are approximately equivalent. If ΔAICc > 6, then the inferior model is about 20 times less likely to be the true model (Table 3).

Finally, for each species, we tested the best model selected by AIC using phenological observations from 2010 and 2011, which had not been used for model calibration. This evaluation was conducted computing the R^2, the slope of the linear regression analysis (observed vs. modeled), as well as the root mean square error (RMSE) between observed and modeled bud-burst dates (Pineiro et al., 2008).

2.3.3 Phenological forecast and propagation of model uncertainty

With the aim of assessing the potential effects of climate change on phenology, we ran the models from 1960 to 2099 using climate projections for Harvard Forest. These were generated by Hayhoe et al. (2007) using the NOAA Geophysical Fluid Dynamics Laboratory (GFDL) CM2 global coupled climate model (Delworth et al., 2006), statistically downscaled to approximately 10 km spatial resolution. The CM2 model was run using two scenarios of CO_2 and other
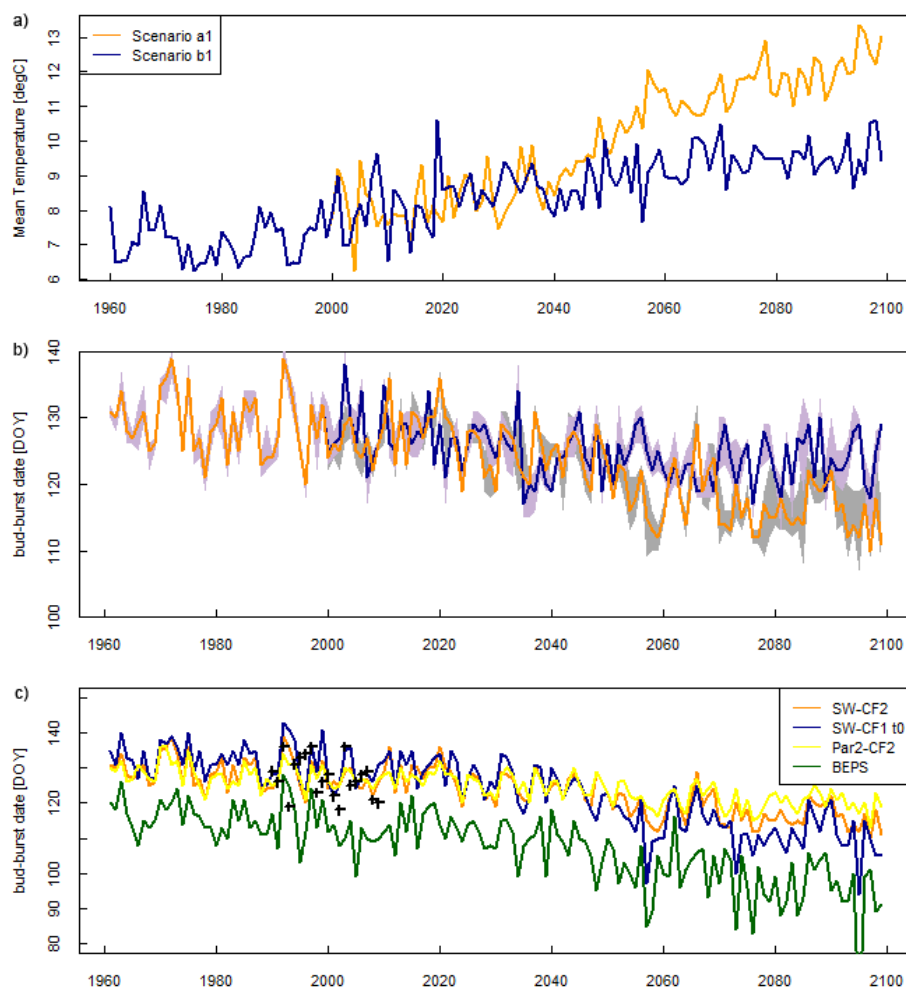
Fig. 3. (a) Time-series of mean annual temperature projections (1960–2099) under the scenario A1fi (orange line) and scenario B1 (blue line); **(b)** time series of bud-burst dates (BB) modeled for red oak (Quercus rubra) with the best model selected (Table 1). Blue line represents the BB projected under the scenario B1, orange line represents the BB projected under scenario A1fi. Gray area represents the uncertainty for the scenario A1fi, while light purple area represents the uncertainty for scenario B1; **(c)** time series of bud-burst date (BB) modeled for red oak with SW-CF2 (best model), SW-CF2t0 (as SW-CF2 but without photoperiod limitation), Par2-CF2 (second best model) and with the internal phenological routine in BEPS (green line). Gray crosses represent the phenological observations collected at Harvard Forest. Models in Fig. 2c are forced by temperatures from the scenario A1fi.

greenhouse gas emissions: the IPCC Special Report on Emission Scenarios (SRES) higher (A1fi) and lower (B1) emission scenarios (Nakicenovic et al., 2000). Compared to a 1960–1990 baseline of 7.1 °C mean annual temperature and 1100 mm annual precipitation, corresponding values at the end of simulation (mean 2070–2099) are 12.0 °C and 1270 mm for the A1fi scenario and 9.5 °C and 1240 mm for the B1 scenario (Fig. 3a).

Uncertainty was propagated by running the models forward with the ensemble of parameter sets that passed the χ^2 test at 95 % confidence (Sect. 2.3.1), yielding for each year a range of model predictions (Fig. 3b).

The uncertainty in model parameters was analyzed by evaluating the propagation of the uncertainty for each sce-

nario and model structure independently. The ratio of the mean width of the 95 % confidence interval (CI), computed for the last and the first decade of the projections, was used to quantify the degree to which parameter uncertainty changes between current and future climate conditions.

The uncertainty related to the model structure was characterized by analyzing the average smoothed projected trend across model formulations and climatic scenarios. The smoothed bud-burst projection was extracted from the time series by using a local polynomial regression fitting (Cleveland and Devlin, 1988). The interannual variability was computed as the standard deviation of the residual between the forecasted bud-burst and its smoothed time-series.

The driver uncertainty related to future climate forcing was analyzed by computing, for the best model selected for each species, the trends in advancing bud-burst under the two different climatic scenarios. Trends were characterized by using the Sen's slope estimator and the nonparametric Mann-Kendall test (Helsel and Hirsch, 2002). The Mann-Kendall test (Mann, 1945; Kendall, 1976) and the Sen's slope estimator (Sen, 1968) are non-parametric procedures for trend testing and estimation of trend magnitude of a univariate time series, respectively. Non-parametric methods are preferred to parametric methods (i.e. regression analysis: regression slope and test) because no assumption is made regarding the probability distribution of data. The second advantage of nonparametric methods, and in particular for the estimation of trend magnitudes, is their robustness to outliers or to abrupt breaks due to inhomogeneous time series (Helsel and Hirsch, 2002).

2.4　Process-based terrestrial biosphere model

2.4.1　Description of Boreal Ecosystem Productivity Simulator

The Boreal Ecosystem Productivity Simulator (BEPS) was originally developed to simulate carbon and water fluxes at daily time steps in a remote sensing framework (Liu et al., 1997, 2002). Derivatives of this original version have been developed and tested at various boreal forest and peatland ecosystems (e.g. Govind et al., 2009, 2011; Sonnentag et al., 2008; Ju and Chen, 2005; Ju et al., 2006). Spring phenological events such as bud-burst are modeled as functions of air temperature (Ju et al., 2006), with a model structure classifiable as spring warming without photoperiod limitation. After bud-burst (accumulated growing-degree days above $5\,°C$ reach $75\,°C$), leaf area index keeps increasing linearly up to growing-degree days of $500\,°C$ when prescribed maximum growing-eason leaf area index for understory and overstory are reached.

For this study, we used the version of BEPS described by Ju et al. (2006), parameterizing the soil-vegetation continuum to consist of five soil and two vegetation layers, with site-specific information for Harvard Forest from the literature (Urbanski et al., 2007) and the Harvard Forest data archive. Spanning the period 1960–2099, the model was driven by half-hourly meteorological forcing data including incoming shortwave radiation, air temperature, relative humidity, wind speed and precipitation.

For the future climate projection, we used downscaled data (Hayhoe et al., 2007) from the regionalized projection of the GFDL-CM global coupled climate-land model Delworth et al. (2006) driven with scenario A1fi (IPCC, 2007). For each run, the model was initialized following the procedure outlined in Ju et al. (2006).

2.4.2　Modeling strategy

Several model runs have been conducted over the period 1960–2099 with BEPS and forced by different bud-burst forecasts. Due to the computational demand of running BEPS with the full posterior distribution of uncertainty from each model, we designed specific experiments to test both the BEPS native phenology model, and the sensitivity of photosynthetic CO_2 uptake and ET estimated by BEPS to errors in simulated bud-burst dates. The different runs are described below:

– Run 1: BEPS with bud-burst simulated by the native phenological routine (reference run).

– Run 2: BEPS forced by bud-burst dates simulated by the best model formulation, selected according to the AICc as described in Sect. 2.3.1 and the optimized parameters set.

– Run 3: BEPS forced by 4 different bud-burst dates. The bud-burst forcing time series has been computed by adding to the bud-burst simulated with the reference run ± 1 and ± 10 days. These runs are hereafter referred as $BEPS_{+1}$. $BEPS_{-1}$; $BEPS_{+10}$ and $BEPS_{-10}$.

The differences in annual GPP and ET between Runs 1 and 2 allow for the quantification of uncertainty in photosynthetic CO_2 uptake and ET associated with the BEPS native bud-burst sub-model. Run 3 allows for the characterization of the sensitivity of photosynthetic CO_2 uptake and ET in BEPS to variations (shifts) in bud-burst dates. In other words, after the characterization of the uncertainty around individual future bud-burst dates, we look at the effect of constant "extra" days in spring along the simulation period in terms of carbon and water fluxes described by BEPS. Hereafter we referred to Run 3 as the "Sensitivity Runs".

3　Results

3.1　Performance evaluation of phenological models and future bud-burst trends

The AICc values show that models belonging to the class spring warming with photoperiod control tended to be better supported by the data than competing model structures (Table 3). In particular, for seven species the simple spring warming models (SW-CF2 plus SW-CF1) are selected as best (Tables 1 and 3). For several species, however, AICc gives support for chilling models. Among chilling models, alternating (Alt) models are more often selected (3 times as best and 5 times with $\Delta AICc < 2$), while sequential (Seq) models are selected just once. Spring warming models without photoperiod control are never selected as the best model. Moreover, Table 3 shows that the SW-CF2 model has the lowest
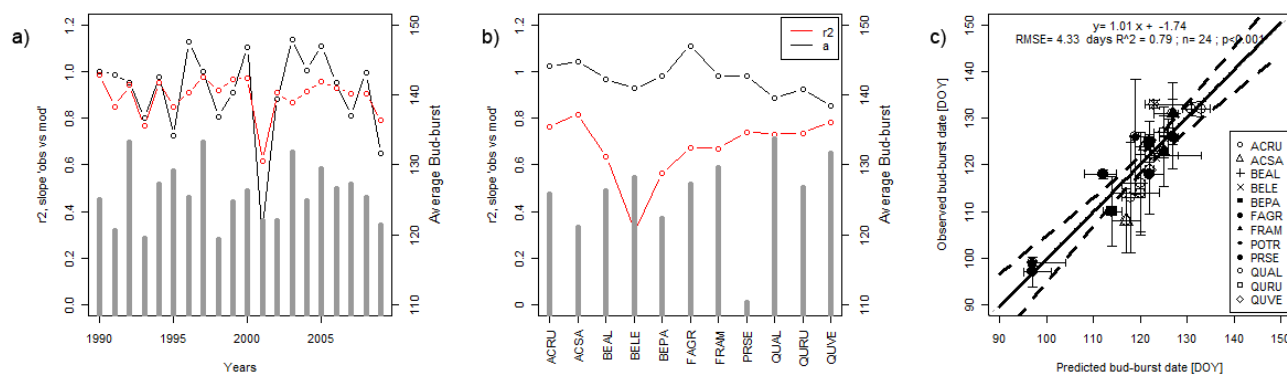
Fig. 4. Coefficient of determination (R^2) and slope of the linear regression between observed and predicted bud-burst with the best model selected (Table 1) for each year **(a)**, for each species **(b)**. Vertical lines represent the average bud-burst day of the year in each year **(a)** and for each species **(b)**. Gray bars represent the average bud-burst dates across species **(a)** and the average bud-burst for each species **(b)**. **(c)** Predicted versus observed bud-burst dates for the years 2010–2011 simulated for the 11 species and the best model selected as described in Table 1. Error bars for the predictions represent the propagated uncertainty of model parameters while for the observed dates is the minimum and maximum bud-burst dates observed for each individual.

AICc across all species, followed by Alt-CF1. The time series of bud-burst dates (1960–2099) simulated for red oak with the best model selected (SW-CF2) for both scenarios are shown in Fig. 3b while simulations with different model structures are reported in Fig. 3c.

The R^2 and the slope of the linear regression between observed and predicted bud-burst dates, with the model selected as best for the period 1990–2009 (i.e. years used for model calibration) and for each species, are reported in Fig. 4a and b, respectively. Figure 4a shows that the best model selected for each species is able to explain the variability of bud-burst across species for each year. Figure 4b shows that the best model selected is able to explain the variability of bud-burst for each species. Figure 4c shows that these calibrated models were able to successfully predict 2010 and 2011 bud-burst dates ($R^2 = 0.79$; RMSE $= 4.3$ day). This represents a strong test of the models, as 2010 had an anomalously early start to the growing season.

Among the significant trends computed with the Sen's slope, we have found an average advance in bud-burst across species of 9.4 (1 sd: ± 3.4, minimum: 5.0, maximum: 14.9) day century^{-1} and 4.3 (1 sd: ± 1.7, minimum: 1.7, maximum: 7.0) day century^{-1} for scenario A1fi and B1, respectively.

3.2 Uncertainty of phenological forecasts

3.2.1 Uncertainty of model parameters

The impact of uncertainty in model parameters was evaluated by running the models forward with 1000 realizations of model parameter sets accepted by the χ^2 test. As an example, the parameter uncertainty for red oak is depicted with colored areas in Fig. 3b. The uncertainty in individual years results in an uncertainty in the projected bud-burst trends. A summary of the resulting uncertainty in the magnitude

of the projected bud-burst trends simulated for each species with the best model selected as in Table 1 is given in Fig. 5. The width of boxplots represents the uncertainty in projected trends for each species. The uncertainty in future trends varied across species and is larger for simulations conducted under the scenario A1fi. The average uncertainty, computed as the average of the 95 % CI, is 2.4 day century^{-1} (ranging from ± 0.7 day century^{-1} for ACRU to ± 4.1 day century^{-1} for BELE) for scenario B1 and 4.5 day century^{-1} (ranging from ± 0.7 day century^{-1} for ACRU to ± 9.2 day century^{-1} for BELE) for A1fi.

The average ratio of the mean width of the 95 % confidence interval, computed for the last and the first decade of the projections with the best models, is 1.2 (1 sd: ± 0.2) for both scenarios. For many models and many species, uncertainty at the end of the simulation is similar to that for the present climate. For models without photoperiod limitation (more often for SW-CF1t0 and SW-CF2t0), and for some chilling models (Par and Par2 models), the uncertainty at the end of the simulation doubles (ratio between 2.0 and 5.3) and, in particular under scenario A1fi, there is a large increase in uncertainty.

3.2.2 Uncertainty of model structure

The smoothed trends computed for a selection of models, averaged across species, for the two scenarios are reported in Fig. 6. The average of the model ensembles (gray line in Fig. 6) and the average computed with the Akaike's weights (black line in Fig. 6) according to Turkheimer et al. (2003) are also shown. Weighted averages and the standard deviation of trends computed across the model structures are reported in Table 1. The average standard deviation of the projected trends is a measure of the uncertainty in trends related
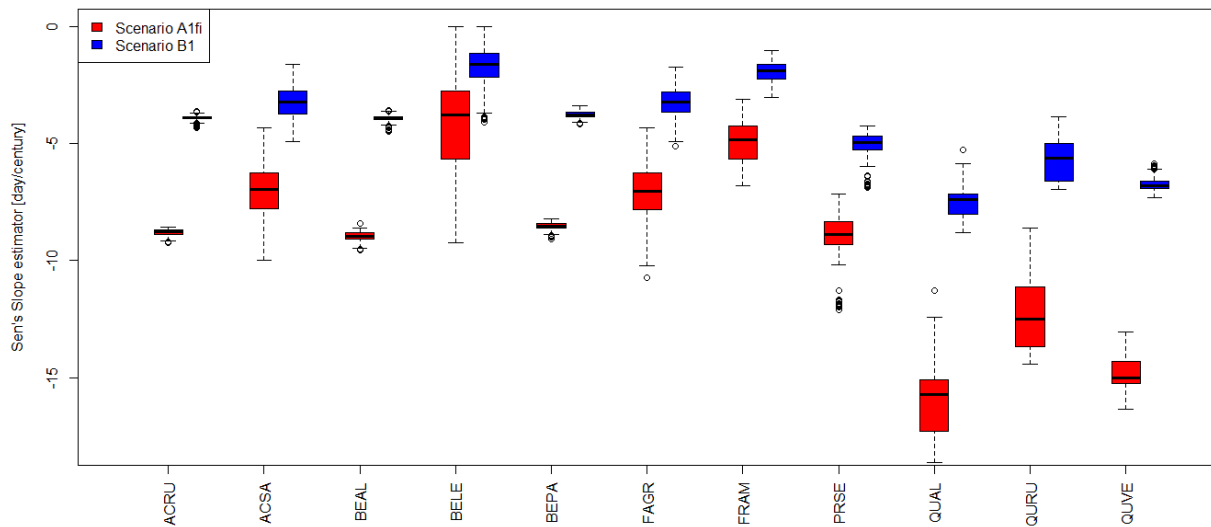
Fig. 5. Box-plot of the trends computed with the Sen's slope estimator for each species. Bud-burst dates (BB) projections were modeled by using the best model selected and the parameters ensemble (1000 parameter set) for each species (Table 1). Red boxes represent the predictions under the scenario A1fi, while blue boxes represent the predictions under the scenario B1.
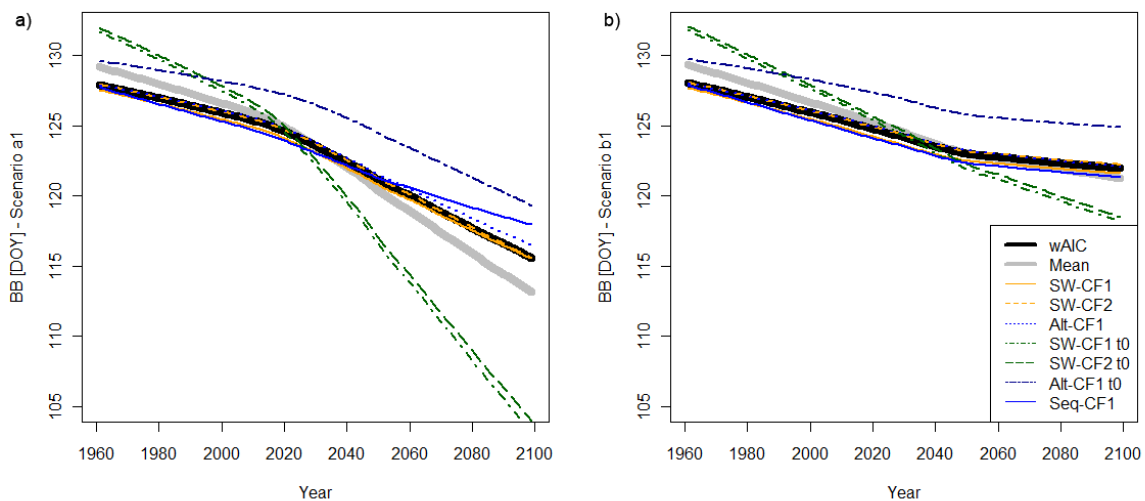

Fig. 6. Smoothed bud-burst (BB) projected time series under the scenario A1fi **(a)** and the scenario B1 **(b)**. Different lines represent different models listed in Table 1. Orange lines represent spring warming models with photoperiod limitations; green lines represent spring warming models without photoperiod limitations (starting dates fixed at 1 January); blue lines represent some chilling class models with (alternating, Alt) and without photoperiod limitation (Alt-CF1t0 and sequential, Seq-CF1). The bud-burst projection are smoothed by using a local polynomial regression fitting (Cleveland and Devlin, 1988). Black line represents the Akaike weighted average of model ensembles wile gray line represents the average.

to model structure and varies between $\pm 3.6\,\mathrm{day\,century^{-1}}$ for B1 and $\pm 7.7\,\mathrm{day\,century^{-1}}$ for A1fi.

Models more frequently selected as the best (i.e. spring warming limited by photoperiod and alternating models) predict a response to future warming in the middle of the model ensemble. Spring warming models without photoperiod limitations show a strong trend toward early bud-burst, particularly relevant under scenario A1fi, that affect also the arithmetic average of multi-model ensemble (gray line in Fig. 6).

By 2099, differences in forecast bud-burst date across models could reach about 10 days for the A1fi scenario. Chilling models and spring warming models with photoperiodic limitation show less steep projected trends. The trends simulated with these two classes of models level-off around 2060 for scenario B1 while for scenario A1fi a slow reduction in the slope is observed over time. For scenario B1, the leveling off reflects the projected decrease of the rate of temperature warming (Fig. 3a) while for scenario A1fi the interaction
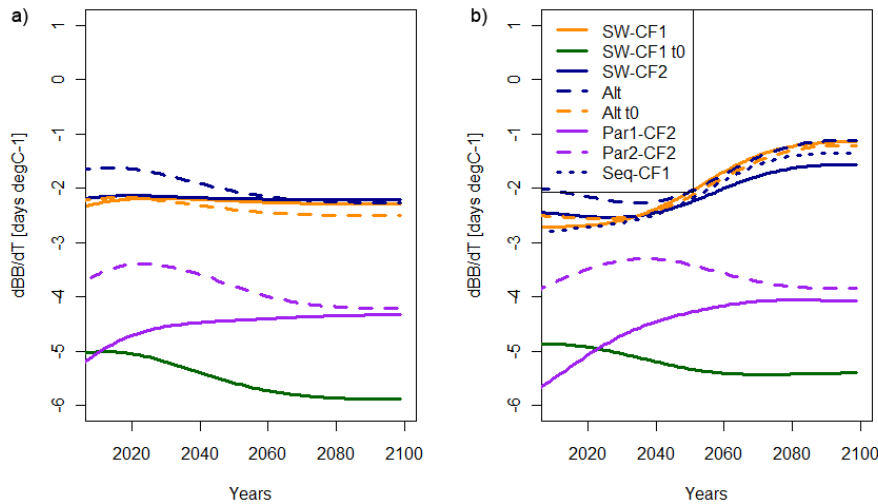
Fig. 7. Future sensitivity of bud-burst to temperature (dBB/dT) as projected by different phenological models under the scenario A1fi **(a)** and the scenario B1 **(b)**. (dBB/dT) is computed as the ratio between the smoothed bud-burst projections for each model and the smoothed temperature. Different lines represent some of the most representative models listed in Table 2.
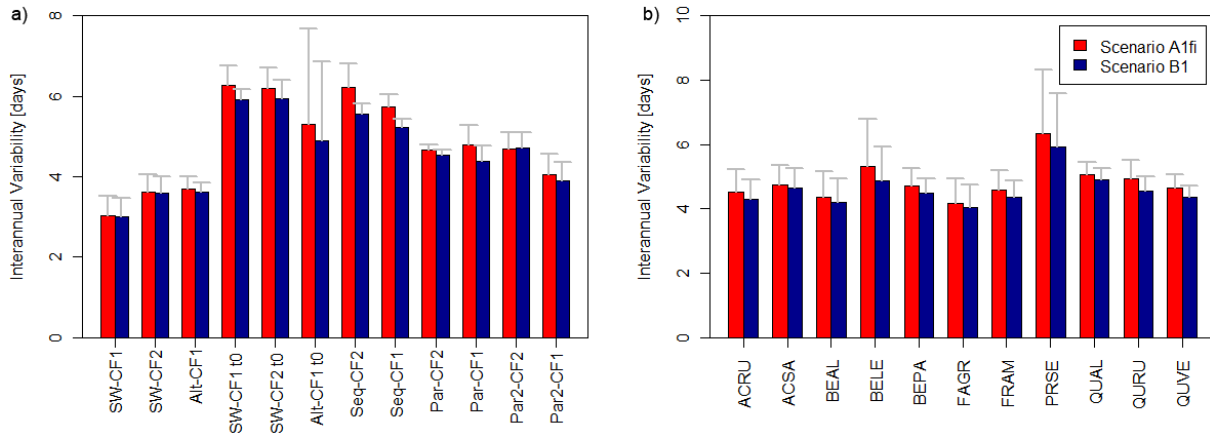

Fig. 8. Histogram of the interannual variability predicted by all model structures reported in Table 2 (left panel) and for all the species reported in Table 1 (right panel). Red histograms represent the interannual variability under scenario A1fi, blue histograms represent the interannual variability under scenario B1. The standard deviation (1 sd) of the interannual variability is also reported.

between model structure, parameters and warm temperatures (see Discussion) contributed to reduce the rate of bud-burst advancing.

The future sensitivity of bud-burst to temperature is variable across model structures as depicted in the time series in Fig. 7. We observe a stronger sensitivity for spring warming models without photoperiod limitation (e.g. SW-CF2t0: -5.3 day $°C^{-1}$ for A1fi, -5.0 day $°C^{-1}$ for B1) compared with spring warming models limited by photoperiod (SW-CF2: -2.4 day $°C^{-1}$ for A1fi, -2.2 day $°C^{-1}$ for B1) and chilling models (e.g. Alt, -2.2 day $°C^{-1}$ for both scenarios). For these models (often selected as the best) and for both scenarios, sensitivity to temperature is stable or decreases in time and spans from -2.4 to -2.9 day $°C^{-1}$ in the 1960s to values ranging from -2.3 to -2.4 day $°C^{-1}$ by the end of

simulation. Under the scenario B1, the average reduction of the bud-burst sensitivity to temperature at the end of simulation for the best models selected is about 8.3 % compared to the sensitivity at the beginning of the simulation. Models without photoperiod or chilling limitation show an increase of the sensitivity of bud-burst to temperature by the end of simulation reaching about -6.0 day $°C^{-1}$ while parallel models showed a quite variable and higher sensitivity to temperature.

The interannual variability predicted by all models and by all species over the entire simulation period is reported in Fig. 8. The blue histograms represent the year-to-year variations under climate scenario B1 and the red histograms under the scenario A1fi. For chilling models and spring warming models without photoperiod limitations, the future

Table 4. Summary of BEPS simulation protocol. Bud-burst forcing represents the bud-burst dates used for model runs. GPP is the mean Gross Primary Productivity, ET is the mean evapotranspiration while GPP and ET sensitivity represent the average difference to the reference run for annual GPP and annual ET expressed as variations of $gC\,m^{-2}$ and cm of water per 1 day of variation of bud-burst. The reference run is the run conducted with BEPS and with bud-burst simulated with its internal phenological routine. In squared parentheses the minimum and maximum values for each variable are reported.

Run ID	Bud-burst forcing	GPP $[gC\,m^{-2}\,yr^{-1}]$	ET $[cm\,yr^{-1}]$	GPP sensitivity run- reference run $[gC\,m^{-2}\,day^{-1}]$	ET sensitivity run- reference run $[cm\,day^{-1}]$
BEPS [Reference Run]	BEPS Native phenology	1569.34 [1295.15/1771.29]	46.53 [40.27/54.57]	Reference	Reference
BEPS$_{SW-CF2}$	SW-CF2	1433.30 [1167.22/1680.88]	45.13 [39.98/52.35]	8.03 [−1.97/19.16]	0.08 [−0.01/0.25]
BEPS$_{-10}$	BEPS Native phenology − 10 days	1636.65 [1343.45/1855.01]	46.94 [40.69/54.72]	6.73 [2.85/10.52]	0.04 [−0.02/0.11]
BEPS$_{-1}$	BEPS Native phenology − 1 days	1578.78 [1301.54/1784.08]	46.59 [40.31/54.65]	9.44 [3.89/21.94]	0.07 [−0.04/0.25]
BEPS$_{+1}$	BEPS Native phenology + 1 days	1561.52 [1291.42/1757.64]	46.44 [40.21/54.56]	−8.40 [−21.69/−2.19]	−0.08 [−0.25/0.03]
BEPS$_{+10}$	BEPS Native phenology + 10 days	1488.87 [1214.80/1702.30]	45.76 [39.80/53.64]	−8.05 [−13.71/−0.97]	–0.08 [−0.19/0.02]

interannual variability predicted is larger than the one predicted by spring warming models with photoperiod limitation and alternating models. Within different species we do not observe large differences, except for black cherry. Interannual variability under the scenario A1fi is slightly larger as shown by the differences in red and blue histograms in Fig. 8. However, differences within the 2 scenarios are not statistically significant.

3.2.3 Uncertainty of model drivers

The uncertainty of model drivers is related to the uncertainty in future climate scenarios used for ecological forecasts and here we analyzed the differences in phenological forecasts obtained using the A1fi and B1 scenarios.

The uncertainty of trends for each species was assessed by computing the Sen's slope estimator for each of the 1000 projected bud-burst time series, for the best model for each species (Fig. 5). The differences between blue and red boxes in Fig. 5 represent differences in mean trends between A1fi and B1 scenario. Projected trends computed for the best model selected for each species for the two scenarios are reported in Table 1. A larger uncertainty in bud-burst trends was evident for scenario A1fi compared to scenario B1.

Among the significant trends computed with the Sen's slope, we have found an average advance in bud-burst across species of 9.4 (1 sd: ±3.4, minimum: 5.0, maximum: 14.9) day century^{-1} and 4.3 (1 sd: ±1.7, minimum: 1.7, maximum: 7.0) day century^{-1} for scenario A1fi and B1, respectively.

3.3 Uncertainty in gross primary productivity and evapotranspiration simulated with BEPS

We used the BEPS model to simulate photosynthetic CO_2 uptake and ET at Harvard Forest. The runs conducted with BEPS are in good agreement with GPP estimates reported in Urbanski et al. (2007). For the period 1997–2004 the mean annual GPP estimated from eddy covariance measurements is $14.0\,MgC\,Ha^{-1}$, the GPP simulated with BEPS$_{SW-CF2}$ is $14.9\,MgC\,Ha^{-1}$ while the BEPS reference run is $16.0\,MgC\,Ha^{-1}$.

We give a summary of ET and GPP from the different BEPS model runs in Table 4. On average the BEPS native phenology model simulates an earlier bud-burst of about 17 days (up to a maximum of 59 days) compared to that simulated with the best model formulation identified above (i.e. SW-CF2). An example of the time series of bud-burst simulated with BEPS, SW-CF2 and SW-CF2t0 are shown in Fig. 3c. Differences between the reference run and BEPS$_{SW-CF2}$ represent the impact of uncertainty in model structure on photosynthetic CO_2 uptake and ET modeled. We observe an average overestimation of the reference run of 136.0 (±59.2, 1 sd) $gC\,m^{-2}\,day^{-1}$ for GPP (±9.6 % of the annual GPP) and 1.38 (±0.77, 1 sd) $cm\,yr^{-1}$ for ET. The correlation between the differences in fluxes (ΔGPP and ΔET) and in bud-burst (ΔBB) computed between the reference run and BEPS$_{SW-CF2}$ is stronger for GPP ($r = -0.85\ p < 0.001$, slope $= -6.49\,gC\,m^{-2}\,day^{-1}$) than for ET ($r = -0.55$, $p < 0.001$, slope $= -0.055\,cm\,day^{-1}$). Figure 9 shows that ΔET and ΔBB are strongly negatively correlated ($r = -0.82$,
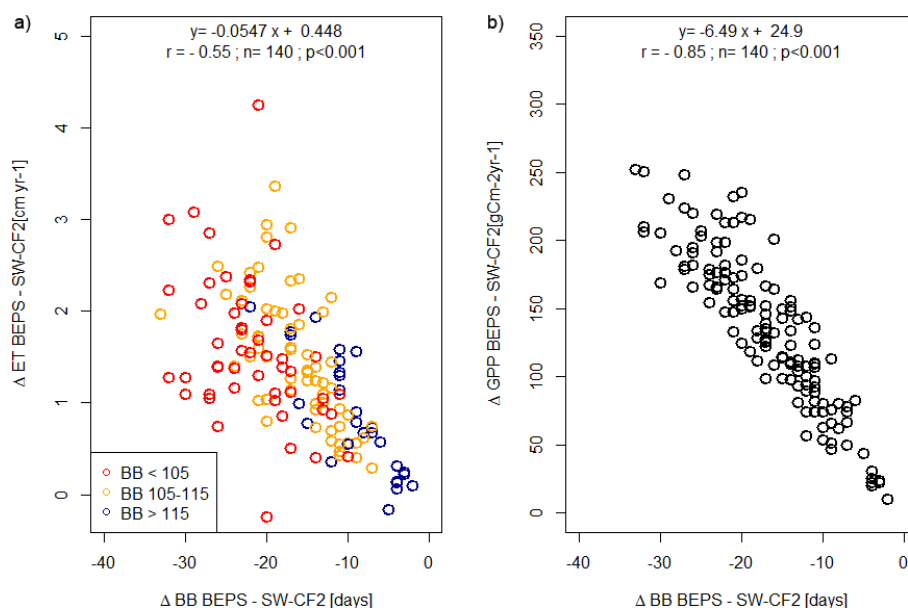
Fig. 9. Relationships between differences in BB (ΔBB) simulated with the best model (SW-CF2 as described in Table 1) and the internal phenological routine of BEPS and the differences in annual ET **(a)** and GPP **(b)** simulated with reference run and BEPS$_{\text{SW-CF2}}$ (Table 4). In **(a)** circles are represented with different colors according to the BB date simulated with the internal routine of BEPS.

$p < 0.001$) when the bud-burst occurs late (bud-burst > 115) while they are poorly correlated ($r = -0.38$, $p < 0.01$) when the bud-burst occurs early (bud-burst < 115). This highlights that year-to-year variations in GPP are well correlated to year-to-year variations of bud-burst dates, while year-to-year variations of ET depend on variations in bud-burst in years with late bud-burst but not in years with early bud-burst, suggesting the role of other meteorological factors controlling ET late in the season when bud-burst occurs early.

The results of the sensitivity runs (i.e. differences between the reference run and BEPS$_{-1}$, BEPS$_{+1}$, BEPS$_{-10}$ and BEPS$_{+10}$) are reported in Table 4 . These represent the sensitivity of annual GPP and ET to uncertainty in bud-burst dates due to model parameter uncertainty (reported as sensitivity of fluxes to ± 1 day of bud-burst uncertainty). The time series of the residuals (ΔGPP, ΔET) to the reference run are reported in Fig. 10. We observe an average increase in annual GPP of 4.3 % (67.3 ± 16.4 (1 sd) gC m^{-2}) for 10 days of advance in bud-burst and a decrease -5.1 % (-80.5 ± 20.3 (1 sd) gC m^{-2}) for 10 days of delay in bud-burst. The sensitivity runs confirm that ET responds less to variations of bud-burst than does GPP: ET varies by 0.92 % and -1.6 % for a change in bud-burst of -10 and $+10$ days, respectively. The average sensitivity of annual GPP to ± 1 day change in bud-burst simulated by BEPS is 8.15 gC m^{-2} day^{-1}, while for annual ET is 0.07 cm day^{-1} (variation of annual fluxes per 1 day of variation of bud-burst). The sensitivity of ET to ± 10 days of variations in bud-burst is asymmetric (green and gray boxes in Fig. 10) and larger for a $+10$ days change

than a -10 days change. Moreover, year-to-year variability in ΔET is larger compared to ΔGPP.

The within-year sensitivity of ΔGPP and ΔET to variations in ΔBB is reported in Fig. 11a, b and confirms that the sensitivity of GPP is similar both for early and late bud-burst, while the sensitivity of ET is less pronounced for the run BEPS$_{-10}$ and in particular for the years with earlier bud-burst (blue line in Fig. 11b). Within-year sensitivity of the average soil moisture in summer to ΔBB and the within-year sensitivity of the percentage change in summer soil moisture to ΔBB are shown in Fig. 11c, d. The average soil moisture in summer is lower for years with earlier bud-burst, and the sensitivity of summertime soil moisture to variations in bud-burst dates is lower for years with earlier bud-burst.

Table 5 summarizes the correlation and the slopes of the linear regression computed between ΔGPP and ΔET simulated with the different sensitivity runs and different meteorological and environmental conditions. Year-to-year variations in ΔGPP are correlated with the bud-burst dates, spring and mean annual temperature and the sensitivity (i.e. slope of the linear regression analysis) is stronger for the run BEPS$_{-10}$. The interannual variability of ΔET is poorly correlated with environmental conditions except for temperature and snow depth for the run BEPS$_{-10}$. The interannual variability of summer ΔET is more correlated with the main meteorological variables (Table 5) compared to annual ΔET.
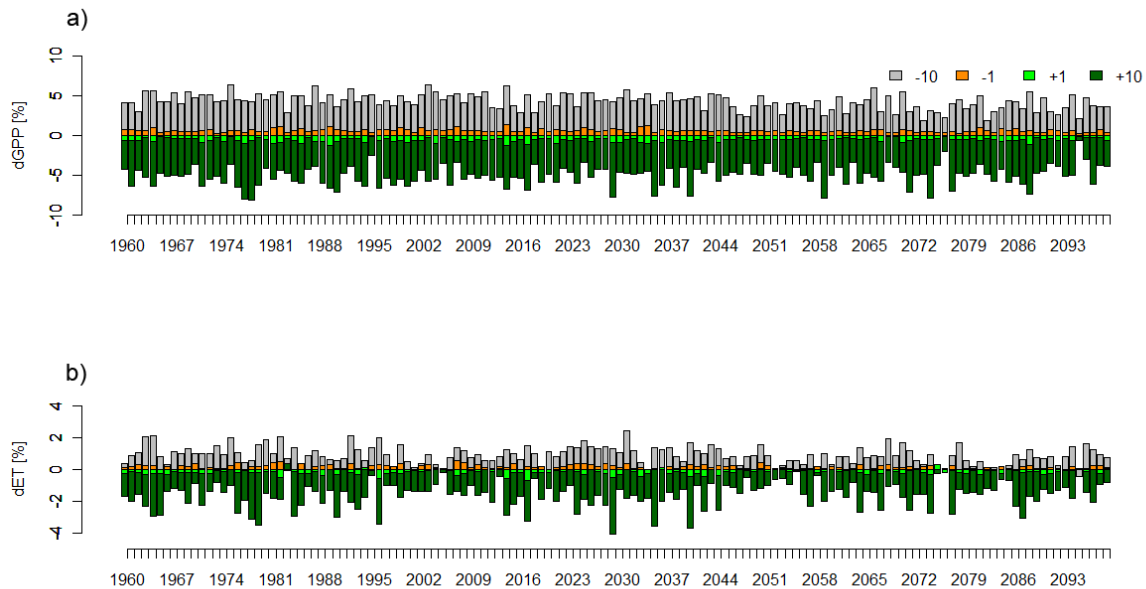
Fig. 10. Barplot of the percentage of variations of annual cumulated GPP **(a)** and ET **(b)** simulated with the runs $BEPS_{-10}$, $BEPS_{-1}$, $BEPS_{+1}$ and $BEPS_{+10}$ (Table 4) with respect to the reference run.

4 Discussion

4.1 Sources of uncertainty of phenological forecasts

Our analysis has characterized different sources of uncertainty in phenological forecasts. Here we have explicitly accounted for uncertainty in estimated model parameters (parameter uncertainty), uncertainty due to model structure (model uncertainty) and uncertainty due to the forcing drivers (driver uncertainty).

Our analysis suggests that, once model parameters are optimized according to the data, parameter uncertainty is the smallest (average 95 % CI: ± 2.4 day century^{-1} for scenario B1 and ± 4.5 day century^{-1} for A1fi) and the most straightforward to quantify by using a model-data fusion framework (model-data-optimization routine). Driver uncertainty is the largest, up to 8.4 day century^{-1} of difference in trends computed with the two scenarios and the best models, and the most complicated to quantify given the uncertainty in future climate projections. Model uncertainty is also large and comparable with driver uncertainty: the predicted budburst trends, as well as the shape of the smoothed projections (Fig. 6), showed large variability across models (1sd: ± 3.6 day century^{-1} for B1 and ± 7.7 day century^{-1} for A1fi).

With the best model selected for each species and its optimized parameters set, an average trend towards earlier spring by 4.4 day (scenario B1) and 8.8 day (scenario A1fi) over the coming century is predicted.

Although the parameter uncertainty is smaller than other sources of uncertainty, we observe that in some species (such as yellow birch, white oak, red oak and sugar maple) forecast bud-burst trends vary greatly depending on the parameter sets used in some models.

Although, forecast uncertainty due to parameter uncertainty scales with time does not increase for all the model structures. The increase of uncertainty by the end of simulation is model-dependent and it is larger for those model structures less supported by data (e.g. spring warming without photoperiodic limitation and parallel models). When these models are used for long-term forecast of spring phenology, the non-stationary of uncertainty in time should be carefully considered.

The above considerations suggest an interaction between the parameter uncertainty and the driver uncertainty. More specifically, we found higher uncertainty for the high CO_2 emissions scenario (Fig. 5). The expansion of the propagated uncertainty under the A1fi scenario indicates the importance of reducing the uncertainty in model parameters (which could be accomplished with longer time series of observations) and, more importantly, in reducing driver uncertainty (i.e. uncertainty in climate scenario).

Driver uncertainty is related to uncertainty in future climate scenarios used to run the phenological models and it is quantified as the difference between runs of the same model forced by different drivers. As shown in Fig. 6, the differences in bud-burst simulated by different models at the end of the simulation period is larger with the scenario A1fi (Fig. 6a) than with scenario B1 (Fig. 6b). In particular, models less supported by the AICc are more sensitive to changes in driver scenarios (e.g. up to 2 weeks for spring warming class models without photoperiod limitation and for some versions of parallel and sequential models). Therefore, accurate model
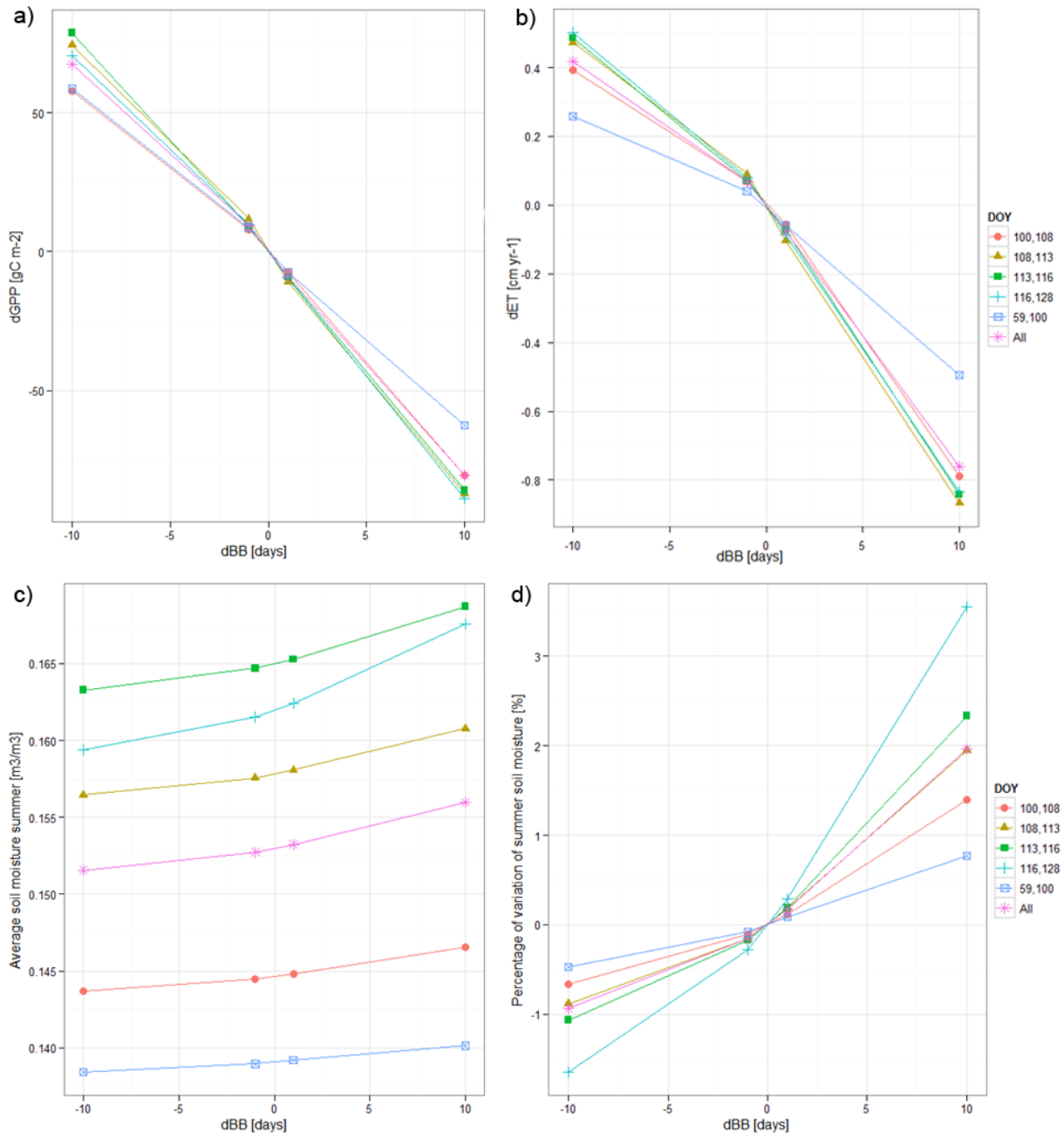
Fig. 11. (a) Within-years sensitivity of GPP (ΔGPP) to variations of BB (ΔBB); **(b)** within-years sensitivity of ΔET to ΔBB; **(c)** within-years sensitivity of the average soil moisture in summer to ΔBB; **(d)** within-years sensitivity of the percentage of variations of summer soil moisture to ΔBB. Different lines represent the within-years sensitivity. Years are grouped according to the distribution of the day of the year (DOY) of bud-burst (i.e. from years with early bud-burst to years with late bud-burst as reported in the legend). The purple line with (*) represents the within-years sensitivity computed using all data.

selection could also help to reduce uncertainty related to model drivers.

The mean temperature increase expected with scenario A1fi is about 3 times larger than that projected for scenario B1. However, some models predict an adjustment of the apparent sensitivity of bud-burst to temperature over the next

century (Fig. 7). In general, there is a reduction in the amount that bud-burst advances for a given increase in temperature. We note that this should not be seen as an adaptation of trees to climate change. Rather, this reflects photoperiod and/or chilling constraints (according to the species), which effectively limit the degree to which the bud-burst can advance

Table 5. Slope of the linear regression analysis computed between ΔGPP, ΔET and summertime ΔET (from June to August) and different meteorological variables. In parentheses the Pearson's correlation coefficient is reported. The differences of GPP (ΔGPP) and ET (ΔET) are computed as the differences in annual GPP [gC m^{-2} yr^{-1}] and ET [cm yr^{-1}] between the runs BEPS$_{-10}$, BEPS$_{-1}$, BEPS$_{+1}$, BEPS$_{+10}$ and the reference run (Table 3). Tann is the mean annual temperature, Tspring is the springtime mean temperature (i.e. from March to May), Tsoil is the mean soil temperature, Prec [mm] is the annual cumulated precipitation and Snow [cm] is the average snow depth. Bold characters represent statistical significant correlations ($p < 0.001$).

ΔGPP	BEPS$_{-10}$	BEPS$_{-1}$	BEPS$_{+1}$	BEPS$_{+10}$
BB	**0.70 (0.45)**	0.07 (0.20)	−0.08 (−0.23)	**−0.85 (−0.44)**
Tann	**−4.76 (−0.57)**	**−0.57 (−0.33)**	**0.63 (0.31)**	**3.85 (0.38)**
Tspring	**−5.46 (−0.59)**	**−0.75 (−0.38)**	**0.71 (0.32)**	**4.37 (0.38)**
Tsoil	**−5.36 (−0.50)**	**−0.70 (−0.31)**	0.67 (0.26)	**4.14 (0.31)**
Prec	−0.02 (−0.19)	−0.003 (−0.16)	−0.01 (0.26)	0.02 (0.17)
Snow	**1.09 (0.36)**	−0.06 (0.10)	−0.13 (−0.19)	−0.76 (−0.20)

ΔET	BEPS$_{-10}$	BEPS$_{-1}$	BEPS$_{+1}$	BEPS$_{+10}$
BB	**0.010 (0.41)**	0.001(0.29)	−0.001 (−0.18)	**−0.011 (−0.30)**
Tann	**−0.042 (−0.33)**	−0.008 (−0.29)	0.005 (0.19)	0.040 (0.21)
Tspring	**−0.048 (−0.33)**	−0.008 (0.28)	0.006 (0.21)	**0.049 (0.23)**
Tsoil	−0.032 (−0.19)	−0.007 (−0.23)	0.005 (0.14)	0.029 (0.12)
Prec	2.14×10^{-4} (−0.14)	-5.13×10^{-6} (−0.02)	-2.78×10^{-5} (−0.08)	-1.94×10^{-4} (−0.08)
Snow	**0.020 (0.43)**	0.002 (0.25)	−0.002 (−0.21)	**−0.020 (−0.28)**

ΔET (JJA)	BEPS$_{-10}$	BEPS$_{-1}$	BEPS$_{+1}$	BEPS$_{+10}$
BB	**−0.011 (0.59)**	**−0.002 (0.53)**	**0.002 (0.47)**	**0.019 (0.63)**
Tann	**0.056 (0.60)**	**0.009 (0.54)**	**−0.009 (0.51)**	**−0.087 (0.57)**
Tspring	**0.097 (0.59)**	**0.011 (0.55)**	**−0.010 (0.48)**	**−0.096 (0.55)**
Tsoil	**0.062 (0.51)**	**0.011 (0.49)**	**−0.011 (0.44)**	**−0.097 (0.48)**
Prec	$\mathbf{3.83 \times 10^{-4}}$ **(0.33)**	$\mathbf{6.29 \times 10^{-5}}$ **(0.30)**	$\mathbf{-7.58 \times 10^{-5}}$ **(0.33)**	$\mathbf{-6.24 \times 10^{-4}}$ **(0.33)**
Snow	**−0.012 (0.34)**	−0.001 (0.23)	0.002 (0.27)	**0.017 (0.29)**

in spring. For instance, considering the average trends computed for all species, spring warming models without photoperiod limitation (e.g. SW-CF2t0) predict an earlier bud-burst compared to spring warming models (e.g. SW-CF2) of 11.5 day century^{-1} for scenario A1 and 5.7 day century^{-1} for B1.

Here, we focused on the uncertainty related to the socio-economic scenario used (A1fi vs B1). However, regarding climate data, other uncertainties can also be explored. Cook et al. (2010), showed the impact of the temporal and spatial resolution of climate scenario in phenological modeling, as well as the uncertainty related to the use of different general circulation models (GCM). This uncertainty could be assessed by analyzing the ensemble spread obtained by running one phenological model with climate drivers from different GCMs.

The model uncertainty is comparable to driver uncertainty, and both the predicted bud-burst trends and the shape of the smoothed projections show a large variability across models (Fig. 6). In this regard, the importance of photoperiod limitation is highlighted by (1) the large difference (up to 10 days by the end of our simulations) between spring warming models with and without photoperiod limitation, and by (2) the

similarity of the sensitivity of bud-burst to temperature for models that feature photoperiod limitation.

For some chilling class models (e.g. parallel models), we observe that in some years the predicted bud-burst dates were highly uncertain and unrealistic (e.g. no bud-burst predicted). As noted in an earlier study (Morin et al., 2009), this effect might be related to interactions between parameter uncertainty and model structural uncertainty. Thus, in anomalous years (i.e. very warm winter), for certain extreme parameter sets, the winter chilling requirements were not fulfilled and models failed to predict bud-burst in spring (i.e. no bud-burst predicted). This occurred more often for scenario A1fi than scenario B1, highlighting the need to reduce this source of uncertainty by either reformulating the model structure or better constraining model parameters by using more observations (longer time series or broader geographic coverage).

Several model structures and parameterizations resulted in predicted bud-burst dates that are not sensitive to warm temperatures during winter and early spring, which limits the degree to which modeled bud-burst can advance under future climate scenarios. Up to a certain level of projected warming in climate, different model structures predict a different absolute value of bud-burst but a similar behavior of the shape and trend of the projections. In these conditions, spring

temperature is the main driver, while photoperiod and chilling are less limiting. For high CO_2 emissions scenarios, photoperiod and chilling become important limiting factors leading to a different shape of projected spring phenology. Our model selection ranking (spring warming with photoperiodic limitation have been selected more frequently than chilling models) supports the position of Körner and Basler (2010) in the current debate about the interactions between warm temperatures in spring, cold temperatures during dormancy and photoperiod, and the degree at which they will control phenology under future climate warming. However, it should be considered that although this study is conducted over a relatively long time series, it includes only one site, which may limit our ability to effectively constrain the photoperiod parameter. Moreover, one of the major limitations related to currently available observational datasets is that these might be too short to span a wide range of climatic conditions needed to constrain more complex models (Richardson et al., 2012). For this reason, data from warming experiments or common garden experiments (e.g. Morin et al., 2010), as well as new long-term datasets of consistent phenological observations from across a wide geographic range and climatic conditions, could be of great value for constraining model parameterization (i.e. to reduce parameters uncertainty) and for testing and developing phenology models (i.e. to reduce model uncertainty).

Finally, the uncertainty described above and multi-model trends reported in Fig. 6 suggest that when constructing a multi-model ensemble, not all model structures should be weighted equally. More weight should be given to those models that are better supported by the data, and less weight to those that are not supported by the data. Akaike weights (Turkheimer et al., 2003) can be used for this purpose.

4.2 Impacts of uncertainty in bud-burst forecasts on photosynthetic CO_2 uptake and evapotranspiration modeling

The uncertainties of phenological forecasts might have important implications for the land surface energy budget and for carbon and water cycling modeling because of the sensitivity of biogeochemical models to bud-burst dates (Levis and Bonan, 2004; White et al., 2000). The prognostic routine in commonly used terrestrial biosphere models (e.g. Thornton et al., 2002; White et al., 1997; Lawrence et al., 2011; Ju et al., 2006) is often based only on the heat sums or growing degree-day assumption (i.e. roughly similar to models SW-CF1t0 and SW2-CF2t0), and model parameters are not optimized against phenological observations. In our analysis, simple model structures such as these were not well supported by the data (with the addition of photoperiod or chilling, model performance was greatly improved). This might lead to a poor description of spring phenology either in terms of bud-burst or seasonal development of leaf area index (Richardson et al., 2012; Lawrence et al., 2011). Hence,

this source of error likely propagates to the seasonality of carbon, water and energy fluxes projected by land surface models (Richardson et al., 2012; Keenan et al., 2012b).

In our modeling exercise with BEPS we explored the effects on evapotranspiration and photosynthetic CO_2 uptake due to (1) the uncertainty in phenological model routines not optimized embedded in the model (i.e. differences between the reference run and BEPS$_{SW-CF2}$) and to (2) the uncertainty in model parameters (i.e. sensitivity runs: BEPS$_{-10}$, BEPS$_{-1}$, BEPS$_{+1}$, BEPS$_{+10}$).

Model uncertainty (i.e. differences in GPP and ET obtained with the reference run and the BEPS$_{SW-CF2}$) lead to biases of 9.6 % for annual GPP and 2.9 % for annual ET, indicating the importance of improving the model's default phenological routines.

Beside the uncertainty related to model structure, phenological forecasts for future climate scenarios are uncertain because of the uncertainty in phenological model parameters. We have quantified that parameter uncertainty can be considered, conservatively, as up to ± 10 days. We explored the impact of this uncertainty on terrestrial biosphere models through a sensitivity analysis and showed (Table 4) that a variation of ± 10 and ± 1 day in bud-burst dates lead to a variation of about ± 5 % in annual GPP (8.15 gC m^{-2} day^{-1}, i.e. a sensitivity of annual GPP to 1 day of variation in bud-burst) and about ± 2 % in annual ET (0.07 cm^{-1} day^{-1}, i.e. sensitivity of annual ET to 1 day of variation in bud-burst).

The sensitivity of GPP to bud-burst uncertainty assessed in this study is comparable with recent empirical estimates: Richardson et al. (2010) found a relationships between interannual phenological anomalies in spring onset and interannual CO_2 uptake anomalies ranging from 1.20 (se: ± 6.20) to 20.2 (se: ± 2.90) gC m^{-2} day^{-1} (supplementry Table 5 in Richardson et al., 2010); Richardson et al. (2009a), indicated that a one-day advance in spring onset date in two different North American forests was associated with an increase in GPP of 12.30 (se: ± 2.50) gC m^{-2} day^{-1}. Although the impact on GPP of uncertainty in model parameters is lower than uncertainty due to phenological model structure, this is of the same magnitude (or slightly lower) of the sensitivity of GPP to anomalies in bud-burst as reported in the abovementioned empirical studies. This emphasizes the importance of correctly modeling phenological transition dates in order to successfully predict annual CO_2 uptake and its interannual variability, as well as other biosphere-atmosphere interactions and climate system feedbacks.

As shown in Fig. 11a, b the sensitivity of GPP to bud-burst is relatively constant while the sensitivity of ET to variations in bud-burst decreases substantially for the runs BEPS$_{-10}$, in particular in years with earlier bud-burst. We observe that an earlier bud-burst of 10 days in spring is not always associated with an increase in ET (Fig. 11b, c, d), perhaps because with early bud-burst the soil water reservoir is inevitably depleted by late summer. As shown in Fig. 11c, according to BEPS, earlier bud-burst is associated with lower

average soil moisture in summer. Moreover, the sensitivity of summer soil moisture to variations of ± 10 days in bud-burst (Fig. 11d) is very small in years with early bud-burst, and increases in years with late bud-burst. Similar patterns have been reported in boreal forest stands (Kljun et al., 2006) and in a broadleaved forest in Central Europe (Leuzinger et al., 2005) where an earlier spring phenology increased transpiration rates, leaving less moisture in the soil in summer.

BEPS runs suggest that at Harvard Forest, the photosynthetic CO_2 uptake is more sensitive than the ET to uncertainty in phenology (Figs. 9, 10 and 11). Regarding water cycling, Lawrence and Slingo (2004) and White et al. (1999), showed that the three primary components of ET, namely, transpiration, soil evaporation and canopy evaporation respond in a opposite direction to changes in phenology. For instance, a certain amount of soil evaporation can occur independently of the seasonal development of leaves in years with late bud-burst. Or, while bud-burst and spring leaf development enhances transpiration, it may limit surface soil evaporation because of the lower amount of transmitted solar radiation below the canopy. Moreover, in years with early bud-burst, the enhancement of ET in spring might lead to lower soil moisture in summer limiting the increase of transpiration in summer and then buffering the sensitivity of total ET to bud-burst. However, these considerations are based on modeling studies and empirical studies would be useful to better understand the sensitivity of ET to spring phenology.

5 Conclusions

The evaluation of different phenology models with a model-data fusion approach and the long-term Harvard Forest phenology record indicate support for spring warming models with photoperiod limitation and to a lesser extent chilling models with alternating description of chilling accumulation. Models without the explicit description of the photoperiod limitation (or where accumulation of degree-days is arbitrarily assumed to begin on 1 January) are the least supported by the data.

We characterize and quantify the three different sources of uncertainty in phenological forecasts: parameter uncertainty, model uncertainty, and driver uncertainty. Parameter uncertainty is the smallest, and most readily quantified whereas driver uncertainty is larger and the most unquantifiable. Model structure uncertainty is comparable with driver uncertainty, but can be quantified by combining models with different complexity and observational data. The uncertainty in model structure affects the temperature sensitivity of bud-burst simulated by phenology models.

Phenology regulates many ecosystem feedbacks to climate, and here we quantify the impact of uncertainties in bud-burst forecasts to simulated carbon and water cycling. We conclude that the structure and the optimization of parameters of the phenological routine is an important source

of uncertainty for terrestrial biosphere models. The photosynthetic CO_2 uptake is more sensitive than the ET to uncertainty in phenology, and the impact of uncertainty in phenological model parameters on carbon uptake is of the same magnitude of the sensitivity of productivity to year-to-year variations of bud-burst. For ET, the impact of this uncertainty is more relevant in years with late bud-burst given the asymmetric response of water cycling to variations of bud-burst observed with our model.

To reduce model uncertainties it is necessary to continuously strive to re-evaluate model predictions against new observational datasets. Of particular value would be data from diverse populations (i.e. including genetic variability) growing under divergent environmental conditions. To push models to their limits, these data should include variations in both temperature (data from transplant or warming experiments would be particularly valuable) and, to the fullest extent possible, photoperiod. These data will permit the development of better phenological routines, and the reduction of uncertainties in forecasts of ecosystem responses to climate change that are mediated by phenology.

References

Akaike, H.: Information theory and an extension of the maximum likelihood principle, in: Proceedings of the Second International Symposium on Information Theory, edited by: Petrov, B. N. and Csaki, F., Akademiai Kiado, Budapest, 267–281 (Reproduced in Kotz, S. and Johnson, N. L. (Eds.), 2003), Breakthroughs in Statistics, vol. I, Foundations and Basic Theory, Springer-Verlag, New York, 610–624, 1973.

Anderson, D. R., Burnham, K. P., and Thompson, W. L.: Null hypothesis testing: Problems, prevalence, and an alternative, J. Wildlife Manage., 64, 912–923, 2000.

Baldocchi, D.: Breathing of the terrestrial biosphere: Lessons learned from a global network of carbon dioxide flux measurement systems, Aust. J. Bot., 56, 1–26, 2008.

Burnham, K. P. and Anderson, D. E.: Model selection and multimodel inference: A practical information-theoretic approach, 2nd Edn., Springer-Verlag, New York, 2002.

Chuine, I. A.: Unified model for tree phenology, J. Theor. Biol., 707, 337–347, 2000.

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

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

Chuine, I., Morin, X., and Bugmann, H.: Warming, photoperiods, and tree phenology, Science, 329, 277–278, 2010.

Cleveland, W. and Devlin, S.: Locally-weighted regression: An approach to regression analysis by local fitting, J. Am. Stat. Assoc., 83, 596–610, 1988.

Cook, B., Terando, A., and Steiner, A.: Ecological forecasting under climatic data uncertainty: A case study in phenological modeling, Environ. Res. Lett., 5, 044014, doi:10.1088/1748-9326/5/4/044014, 2010.

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., Cooke, W. F., Dixon, K. W., Dunne, J., Dunne, K. A., Durachta, J. W., Findell, K. L., Ginoux, P., Gnanadesikan, A., Gordon, C. T., Griffies, S. M., Gudgel, R., Harrison, M. J., Held, I. M., Hemler, R. S., Horowitz, L. W., Klein, S. A., Knutson, T. R., Kushner, P. J., Langenhorst, A. R., Lee, H.-C., Lin, S.-J., Lu, J., Malyshev, S. L., Milly, P. C. D., Ramaswamy, V., Russell, J., Schwarzkopf, M. D., Shevliakova, E., Sirutis, J. J., Spelman, M. J., Stern, W. F., Winton, M., Wittenberg, A. T., Wyman, B., Zeng, F., and Zhang, R.: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics, J. Climate, 19, 643–674, 2006.

Fitzjarrald, D. R., Acevedo, O. C., and Moore, K. E.: Climatic consequences of leaf presence in the eastern United States, J. Climate, 14, 598–614, 2001.

Franks, S. W., Beven, K. J., and Gash, J. H. C.: Multi-objective conditioning of a simple SVAT model, Hydrol. Earth Syst. Sci., 3, 477–488, doi:10.5194/hess-3-477-1999, 1999.

Govind, A., Chen, J. M., Margolis, H., Ju, W., Sonnentag, O., and Giasson, M. A.: A spatially explicit hydro-ecological modeling framework (BEPS-Terrainlab v2.0): Model description and test in a boreal ecosystem in eastern North America, J. Hydrol., 367, 200–216, 2009.

Govind, A., Chen, J. M., Bernier, P., Margolis, H., Guindon, L., and Beaudoin, A.: Spatially distributed modeling of the long-term carbon balance of a boreal landscape, Ecol. Modell., 222, 2780–2795, 2011.

Hayhoe, K., Wake, C., Anderson, B., Liang, X.-Z., Maurer, E., Zhu, J., Bradbury, J., DeGaetano, A., Stoner, M., and Wuebbles, D.: Regional climate change projections for the Northeast USA, Mitigation and Adaptation Strategies for Global Change, 13, 425–436, 2007.

Helsel, D. and Hirsch, R.: Statistical methods in water resources, in: Techniques of water resources investigations, edited by: Survey, U. S. G., 1–522, 2002.

Hunter, A. F. and Lechowicz, M. J.: Foliage quality changes during canopy development of some northern trees, Oecologia, 89, 316–323, 1992.

IPCC: Climate change 2007: Synthesis report, IPCC, Geneva, Switzerland, 104 pp., 2007.

Jeong, S.-J., Ho, C.-H., Gim, H.-J., and Brown, M. E.: Phenology shifts at start vs. end of growing season in temperate vegetation

over the northern hemisphere for the period 1982–2008, Glob. Change Biol., 17, 2385–2399, 2011.

Ju, W. and Chen, J. M.: Distribution of soil carbon stocks in Canada's forests and wetlands simulated based on drainage class, topography and remotely sensed vegetation parameters, Hydrol. Process., 19, 77–94, 2005.

Ju, W., Chen, J. M., Black, T. A., Barr, A. G., Liu, J., and Chen, B.: Modelling multi-year coupled carbon and water fluxes in a boreal aspen forest, Agr. Forest Meteorol., 140, 136–151, 2006.

Keenan, T. F., Carbone, M., Reichstein, M., and Richardson, A. D.: The model-data fusion pitfall: Assuming certainty in an uncertain world, Oecologia, 167, 587–597, 2011.

Keenan, T. F., Davidson, E., Moffat, A., Munger, J. W., and Richardson, A. D.: Using model-data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling, Glob. Change Biol., doi:10.1111/j.1365-2486.2012.02684.x, in press, 2012a.

Keenan, T. F., Baker, I., Barr, A., Ciais, P., Davis, K., Dietze, M., Dragoni, D., Gough, C. M., Grant, R., Hollinger, D., Hufkens, K., Poulter, B., McCaughey, H., Rackza, B., Ryu, Y., Schaefer, K., Tian, H., Verbeeck, H., Zhao, M., and Richardson, A. D.: Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO_2 exchange, Glob. Change Biol., 18, 1971–1987, doi:10.1111/j.1365-2486.2012.02678.x, 2012b.

Kendall, M. G.: Rank Correlation Methods, 4th Edn., Griffin, 1976.

Kljun, N., Black, T. A., Griffis, T. J., Barr, A. G., Gaumont-Guay, D., Morgenstern, K., McCaughey, J. H., and Nesic, Z.: Response of net ecosystem productivity of three boreal forest stands to drought, Ecosystems, 9, 1128–1144, 2006.

Körner, C. and Basler, D.: Warming, photoperiods, and tree phenology response, Science, 329, 278–278, 2010.

Lawrence, D. M. and Slingo, J. M.: An annual cycle of vegetation in a GCM. Part I: Implementation and impact on evaporation, Clim. Dynam., 22, 87–105, 2004.

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thorton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Skaguchi, K., Bonan, G. B., and Slater, A. G.: Parameterization improvements and functional and structural advances in version 4 of the Community Land Model, J. Adv. Model. Earth Syst., 3, 27 pp., doi:10.1029/2011MS000045, 2011.

Levis, S. and Bonan, G. B.: Simulating springtime temperature patterns in the community atmosphere model coupled to the community land model using prognostic leaf area, J. Climate, 17, 4531–4540, 2004.

Leuzinger, S., Zotz, G., Asshoff, R., and Korner, C.: Responses of deciduous forest trees to severe drought in Central Europe, Tree Physiol., 25, 641–650, 2005.

Lieth, H. and Radford, J. S.: Phenology, resource management, and synagraphic computer mapping, Bioscience, 21, 62–70, 1971.

Liu, J., Chen, J. M., Cihlar, J., and Park, W. M.: A process-based boreal ecosystem productivity simulator using remote sensing inputs, Remote Sens. Environ., 62, 158–175, 1997.

Liu, J., Chen, J. M., Cihlar, J., and Chen, W.: Net primary productivity mapped for Canada at 1-km resolution, Global Ecol. Biogeogr., 11, 115–129, 2002.

Menzel, A., Sparks, T. H., Estrella, N., Koch, E., Aasa, A., Ahas, R., Alm-Kubler, K., Bissolli, P., Braslavska, O., Briede, A., Chmielewski, F. M., Crepinsek, Z., Curnel, Y., Dahl, A., Defila, C., Donnelly, A., Filella, Y., Jatcza, K., Mage, F., Mestre, A.,

Nordli, O., Penuelas, J., Pirinen, P., Remisova, V., Scheifinger, H., Striz, M., Susnik, A., Van Vliet, A. J. H., Wielgolaski, F. E., Zach, S., and Zust, A.: European phenological response to climate change matches the warming pattern, Glob. Change Biol., 12, 1969–1976, 2006.

Mann, H. B.: Nonparametric tests against trend, Econometrica, 13, 245–259, 1945.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E.: Equations of state calculations by fast computing machines, J. Chem. Phys., 21, 1087–092, 1953.

Migliavacca, M., Meroni, M., Busetto, L., Colombo, R., Zenone, T., Matteucci, G., Manca, G., and Seufert, G.: Modeling gross primary production of agro-forestry ecosystems by assimilation of satellite-derived information in a process-based model, Sensors, 9, 922–942, 2009.

Monteith, J. L. and Unsworth, M. H.: Principles of environmental physics, edited by: Arnold, E., London, 291 pp., 1990.

Morin, X., Lechowicz, M. J., Augspurger, C., O'Keefe, J., Viner, D., and Chuine, I.: Leaf phenology in 22 North American tree species during the 21st century, Glob. Change Biol., 15, 961–975, 2009.

Morin, X., Roy, J., Sonie, L., and Chuine, I.: Changes in leaf phenology of three European oak species in response to experimental climate change, New Phytol., 186, 900–910, 2010.

Morisette, J. T., Richardson, A. D., Knapp, A. K., Fisher, J. I., Graham, E. A., Abatzoglou, J., Wilson, B. E., Breshears, D. D., Henebry, G. M., Hanes, J. M., and Liang, L.: Tracking the rhythm of the seasons in the face of global change: Phenological research in the 21st century, Front. Ecol. Environ., 7, 253–260, 2009.

Nakićenović, N., Alcamo, J., Davis, D., de Vries, B., Fenhann, J., Gaffin, S., Gregory, K., Grubler, A., Jung, T. Y., Kram, T., Lebre La Rovere, E., Michaelis, L., Mori, S., Morita, T., Pepper, W., Pitcher, H., Price, L., Riahi, K., Roehrl, A., Rogner, H., Sankovski, A., Schlesinger, M., Shukla, P., Smith, S., Swart, R., van Rooijen, S., Victor, N., and Dadi, Z.: Special Report on Emissions Scenarios, Cambridge University Press, Cambridge, 2000.

Piñeiro, G., Perelman, S., Guerschman, J. P., and Paruelo, J. M.: How to evaluate models: Observed vs. predicted or predicted vs. observed?, Ecol. Modell., 216, 316–322, 2008.

Randerson, J. T., Hoffman, F. M., Thornton, P. E., Mahowald, N. M., Lindsay, K., Lee, Y. H., Nevison, C. D., Doney, S. C., Bonan, G., Stockli, R., Covey, C., Running, S. W., and Fung, I. Y.: Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models, Glob. Change Biol., 15, 2462–2484, 2009.

Richardson, A. D. and O'Keefe, J.: Phenological differences between understory and overstory: A case study using the long-term Harvard Forest records, in: Phenology of ecosystem processes, edited by: Noormets, A., Springer Science + Business, New York, 87–117, 2009.

Richardson, A. D., Hollinger, D. Y., Dail, D. B., Lee, J. T., Munger, J. W., and O'Keefe, J.: Influence of spring phenology on seasonal and annual carbon balance in two contrasting New England forests, Tree Physiol., 29, 321–331, 2009.

Richardson, A. D., Williams, M., Hollinger, D., Moore, D., Dail, D., Davidson, E., Scott, N., Evans, R., Hughes, H., Lee, J., Rodrigues, C., and Savage, K.: Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints, Oecologia, 164, 25–40, 2010a.

Richardson, A. D., Black, T. A., Ciais, P., Delbart, N., Friedl, M. A., Gobron, N., Hollinger, D. Y., Kutsch, W. L., Longdoz, B., Luyssaert, S., Migliavacca, M., Montagnani, L., Munger, J. W., Moors, E., Piao, S., Rebmann, C., Reichstein, M., Saigusa, N., Tomelleri, E., Vargas, R., and Varlagin, A.: Influence of spring and autumn phenological transitions on forest ecosystem productivity, Philos. T. Roy. Soc. B, 365, 3227–3246, 2010b.

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., and Xue, Y.: Terrestrial biosphere models need better representation of vegetation phenology: Results from the North American Carbon Program Site Synthesis, Glob. Change Biol., 18, 566-584, 2012.

Rosenzweig, C., Casassa, G., Karoly, D. J., Imeson, A., Liu, C., Menzel, A., Rawlins, S., Root, T. L., Seguin, B., and Tryjanowski, P.: Assessment of observed changes and responses in natural and managed systems. Climate Change 2007: Impacts, Adaptation and Vulnerability, Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J., and Hanson, C. E., Cambridge University Press, Cambridge, UK, 79–131, 2007.

Sarvas, R.: Investigations on the annual cycle of development on forest trees active period, Communicationes Instituti Forestalis Fenniae, 76–110, 1972.

Sen, P. K.: Estimates of the regression coefficient based on Kendall's tau, J. Am. Stat. Assoc., 63, 1379–1389, 1968.

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

Schleip, C., Menzel, A., and Dose, V.: Norway spruce (Picea abies): Bayesian analysis of the relationship between temperature and bud burst, Agr. Forest Meteorol., 148, 631–643, 2008.

Schwartz, M. D., Ahas, R., and Aasa, A.: Onset of spring starting earlier across the northern hemisphere, Glob. Change Biol., 12, 343–351, 2006.

Sonnentag, O., Chen, J. M., Roulet, N. T., Ju, W., and Govind, A.: Spatially explicit simulation of peatland hydrology and carbon dioxide exchage: Influence of mesoscale topography, J. Geophys. Res.-Biogeo., 113, G02005, doi:10.1029/2007JG000605, 2008.

Thornton, P. E., Law, B. E., Gholz, H. L., Clark, K. L., Falge, E., Ellsworth, D. S., Golstein, A. H., Monson, R. K., Hollinger, D., Falk, M., Chen, J., and Sparks, J. P.: Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests, Agr. Forest Meteorol., 113, 185–222, 2002.

Turkheimer, F. E., Hinz, R., and Cunningham, V. J.: On the undecidability among kinetic models: from model selection to model averaging, J. Cerebr. Blood F. Met., 23, 490–498, 2003.

Urbanski, S., Barford, C., Wofsy, S., Kucharik, C., Pyle, E., Budney, J., McKain, K., Fitzjarrald, D., Czikowsky, M., and Munger, J. W.: Factors controlling CO_2 exchange on timescales from hourly to decadal at Harvard Forest, J. Geophys. Res.-Biogeo., 112, G02020, doi:10.1029/2006JG000293, 2007.

Vitasse, Y., Francois, C., Delpierre, N., Dufrene, E., Kremer, A., Chuine, I., and Delzon, S.: Assessing the effects of climate change on the phenology of European temperate trees, Agr. Forest Meteorol., 151, 969–980, 2011.

Wang, Y. P., Trudinger, C. M., and Enting, I. G.: A review of applications of model-data fusion to studies of terrestrial carbon fluxes at different scales, Agr. Forest Meteorol., 149, 1829–1842, 2009.

White, M. A., Thornton, P. E., and Running, S. W.: A continental phenology model for monitoring vegetation responses to interannual climatic variability, Global Biogeochem. Cy., 11, 217–234, 1997.

White, M. A., Running, S. W., and Thornton, P. E.: The impact of growing-season length variability on carbon assimilation and evapotranspiration over 88 years in the eastern US deciduous forest, Int. J. Biometeorol., 42, 139–145, 1999.

White, M. A., Thornton, P. E., Running, S. W., and Nemani, R. R.: Parameterization and sensitivity analysis of the BIOME-BGC terrestrial ecosystem model: Net primary production controls, Earth Interact., 4, 1–85, 2000.